Re-Educating Digitisation
Non-Visitor Research

Co-funded by the European Union

RED – Workshop III
Non-Visitor Research I

- I. Introduction -

Author
Prof. Dr. Tibor Kliment

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

The programme!

Topics of the Non-Visitor Research Workshop I

Section 1

I. **Introduction:**
   - Installing SPSS on every device

II. **Introduction into SPSS and Statistical Basics II**
   - Frequencies, means, measures of dispersion, crosstabs

Section 2

III. **Introduction into Data Analyses with PSPP**
   - Downloading and Installing PSPP
   - Introduction into the programme

IV. **Data analyses I of the *See You Sound Festival visitor study***
   - Introduction into strategies of data analysis
   - Hands on exercises into data analyses and data presentation
   - Findings, discussing and interpreting results

Section 3

V. **Data analyses II of the *See You Sound Festival visitor study***
   - Hands on data analyses and data presentation
   - Findings and interpreting results

VI. **Non-Visitor Studies: Introduction I**

VII. **Preparing the Non-Visitor Study in Barcelona**
   - Set up of the study, how to conduct interviews, developing questionaire, financing the study
   - Further time schedules

Benefits: What is the outcome for your in this week:

- More in depth knowledge about doing **data analysis**, **drawing conclusions and taking measures** from that

- Knowledge about **visitors`** structures, motifs, needs, why they come and why not, know how satisfied they are

- Ways to find out **how to make your program more visitor oriented**

- How to save money because of more **loyal visitors**; how to make more turnover from ticketing and funding

- How to make better and more **cost-efficient media campaigning**

- Knowledge about a variety of **other research methods for different purposes**, their pros and cons

- Knowledge about what **non-visitor studies** are for, their possibilities and challenges

- Know how to secure **representativity** in visitor- and non-visitor studies

END OF SECTION !

OUTCOME

Introduction into more advanced quantitative analyses with practical exercises

- **Introduction into other methods of research**

- **Non-Visitor studies: Introduction and differences to visitor-analyses**

- **Knowing about sampling techniques and securing representativity in your studies**

- **Preparing a non-visitor study in Barcelona**

Workshop IV (Barcelona)

I. **Conducting and analysing a non-visitor study in Barcelona**

II. **Working with modern marketing concepts (Life-Style, Personas, introduction into STP-Model etc.)**

III. **Introduction into PSPP**

What is the outcome for you:

After experiencing this workshop:

- **You know how to develop professional questionaires**

- **You know about specialized data analyses programs**

- **You can do basic statistical analyses**

- **You know how to interpret and present the results**

- **You know how to implement the results into your festival business…**

After experiencing all workshops:

- **You are able to plan and manage the entire research process and you know how to choose the right research design**

- **You know how to combine marketing concepts (in the area of pricing, distribution, communication, product development etc.) with visitor research**

- **All together: Your audience development strategy is lifted onto another level, you have more success with your festival and your life will be happier**

RED Re-Educating Digitisation

Rheinische Hochschule Köln

Benefits: What may be the outcome for your in this week:

- **Knowledge about constructing questionaires and doing data analysis**
- **Knowledge about your visitor structure**
- **Knowledge about visitors` motifs, needs, why they come and why not, know how satisfied they are**
- **Ways to make your program more visitor oriented**
- **Make better and more efficient media campaigning**
- **Gaining more visitors and attain new target groups**
- **Save money because of getting more loyal visitors and less expensive campaigning**
- **Make more turnover from ticketing and funding**

RED
Non-Visitor Research I

- II. Introduction into Statistics and SPSS II –
Modifying Data, Measures of centrality and dispersion, building crosstabs

Author
Prof. Dr. Tibor Kliment

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

Short repetition:
Data modifying - Recoding, computing, counting

Recoding variables

Recoding can be used in different situations:

- To categorize a variable (for example, you may want to use people's age to create different age groups, to merge categories, to change string into numerical data)

- Want to change the order of values when displaying data or doing analyses with two or more variables

In SPSS, there are three basic options for recoding variables:

- Recode into d**ifferent variables**

- Recode into s**ame variables**

- **DO IF - Syntax**

Each of these options allows you to re-categorize an existing variable. Recode into different Variables and DO IF Syntac create a new variable without modifiying the original variable, while Recode into same Variable will permantenly overwrite the original variable.

In general, it is best to recode a variable always into a different variable so that you never alter the original data and can easily access the original data if you need to make different changes later on.

Recode Variables – Some general rules

- Often the number of categories is selected in order to have catchy and **meaningful category limits.**

- In an example with net income, the category limits follow each other in steps of 500; we then have 9 categories instead of 8.

- One should **avoid cutting up areas** in which characteristic values appear in clusters by a class boundary or in which the majority of cases is in

- Do not let them appear in clusters at the **edge of a class**.

- The categories must be **clearly defined and mutually exclusive** (disjunctive), i.e. each measured value can only be put into one category.
No duplicate values for limits (e.g. 0 - under 500 €, 500 to under 1,000 €, 1,000 to under 2,000 etc.)

Recode Variables – Some general rules

- The categories must be **exhaustive**, i.e. each measured value must be assignable to a single category.

- One should form **intervals of equal width**, with round numbers as interval centres or with round numbers as interval limits (e.g. income 1,500 to 2,000 €).

- **Avoid classes that are open to one side**. If, however, the minimum and maximum are unknown or the cell populations are too small, the lowest class must remain open at the bottom and the highest class must remain open at the top (e.g. top category of the variable net income: "4.000 and more").

- Here, however, the question is how the class centers are to be determined for further calculations.
  Here you use the **midpoint of the class specified**, for open classes use the boundary given

- **Build classes so that in each class there are enough cases for further analysis (i.e. building crosstabs)**
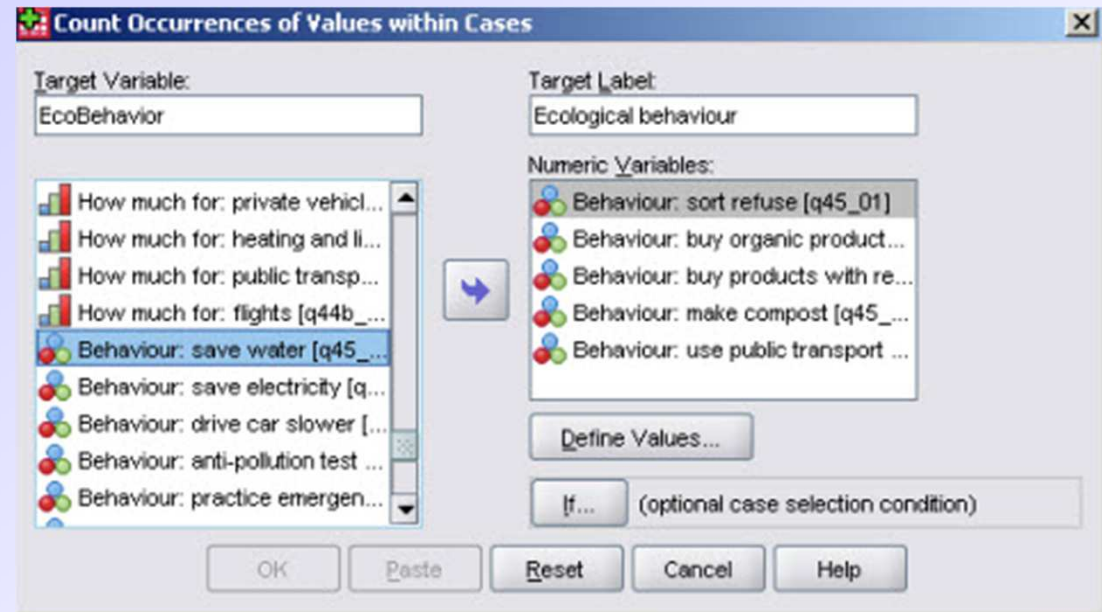
Generate a new variable by compute

- Generating variables in SPSS is simple, especially if you want to generate a new variable from an already existing variable.

- Compute is generally used, when you want to combine two or more variables

- Reseachers often generate new variables that are copies of current ones if they want to change or recode the date, while also keeping the original data so it is not lost.

- There is variety of possibilities of creating new variables .

Count Values – First Step

- This command creates a variable that contains, how often one or more certain values occur in a number of variables.

- Use '**Transform**' and click on '**Count Values within Cases**' to call up the count dialog.

- Indicate a target variable that will receive the counts and specify a label for it.

- Consider which values are to be counted (you can count more than one value)

Count Occurrences of Values within Cases

Target Variable:
EcoBehavior

Target Label:
Ecological behaviour

How much for: private vehicl...
How much for: heating and li...
How much for: public transp...
How much for: flights [q44b_...
Behaviour: save water [q45_...
Behaviour: save electricity [q...
Behaviour: drive car slower [...
Behaviour: anti-pollution test ...
Behaviour: practice emergen...

Numeric Variables:
Behaviour: sort refuse [q45_01]
Behaviour: buy organic product...
Behaviour: buy products with re...
Behaviour: make compost [q45_...
Behaviour: use public transport ...

Define Values...

If... (optional case selection condition)

OK   Paste   Reset   Cancel   Help

Count Values – Second Step

- Select the value(s) you want to count. You can count specific values, ranges of values, but also Missing values and SYSMIS. Add all values you want to count. (In our example we are counting 1s in all the variables)

- Note: counting never produces missing values

- We try the example with V7.1 to V7.4 with value 4
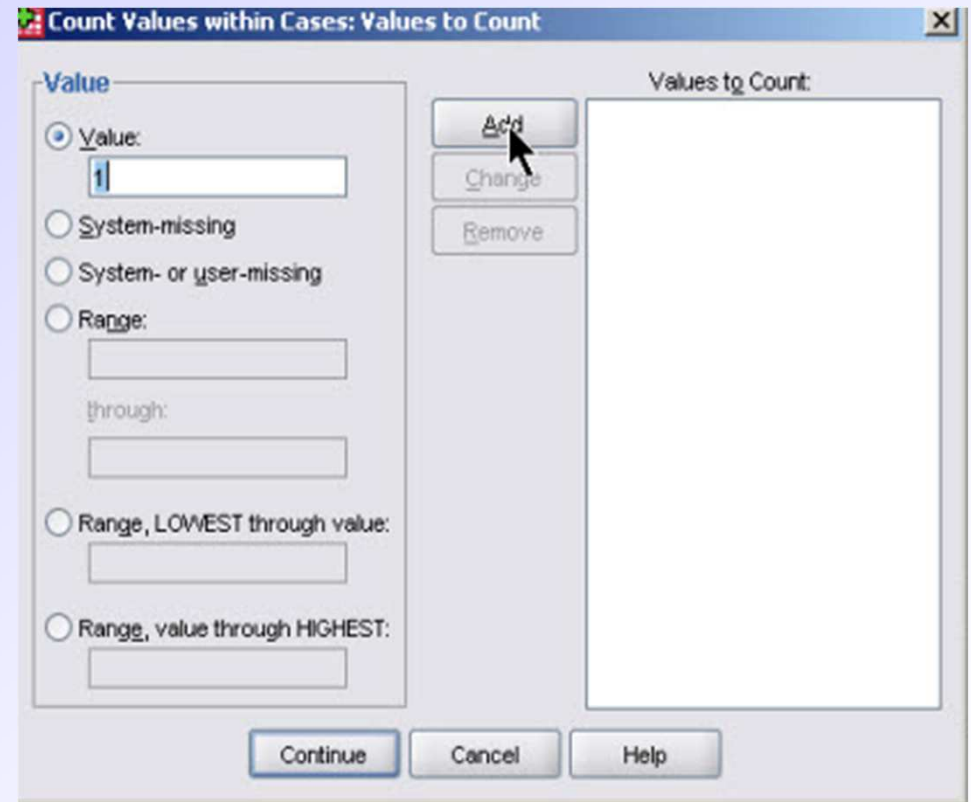
Exercises: Load the SYS - Data file

1. **RECODE Var C "Residence" into Residence_2 (1 -> 1 In Turino, 2-5->2 Outside Turino). Make a frequency command.**

2. **RECODE V2 into a new variable "NOFV" / "Number of festivals visited". Under the condition that V1 has the value 1 -> assign value=1!**

3. **Build 2 classes** for ***by total_visits_2*** *in the past 5 years. Recode 1,2 ->1 (rare visitor) and 3-5->2 (regular visitors)*

4. **Compute a new variable "NOA" / "Number of Activties", that encompasses the frequency of all visits to the festival by adding V11.11 – V11.14.**

5. **Create** an new variable **"BML" / "Broadness of Music Liked"** that counts the number of music genres liked per respondent (only value 3).
   Do a **COUNT** command for that purpose (Value = 1).
   Do a **frequency command** with **BML** and interpret the result.

Some basic statistical analyses

Univariate analyses: Doing Frequencies (distribution table)

- Gives you an idea about the distribution of one variable in a short glace
- Most **common and important mode of data analyses** in SPSS
- Shows you absolute figures, percentages, valid percentages (excludes missings) and cumulative percentages

The data analysis – Frequencys: important core variables

Frequencies command can be used also to make

- **to determine quartiles, percentiles,**

- **measures of central tendency (mean, median, and mode),**

- **measures of dispersion (range, standard deviation, variance, minimum and maximum),**

- **(measures of kurtosis and skewness, and create histograms.)**

A short introduction into statistical measures

- While frequency counts provide information about the entire distribution of a variable, statistical measures provide information about specific properties of a distribution.

- Measures serve to **condense information**, because they provide compressed information about the characteristic properties of a distribution with only one **single number**.

- Statistical measures are to describe a univariate distribution.

They are divided into two groups:

- **Location measures** (also: measures of central tendency, centrality values)

- **Measures of dispersion** (also: dispersion measures)

What are central statistical measures?

Measures of centrality	Measures of dispersion
Mode ('X'$_M$) (read X across M)	Range (R)
Median ('X'$_Z$) (read X across Z)	(mean) quartile distance (QA/QMA), abs. Deviation
Arithmetic mean ('X') (read X across)	Variance (s^2), Standard Deviation (s)
Geometric mean ('X'$_G$) (read X across G)	
Harmonic mean ('X'$_H$) (read X across H)	

A brief introduction into some statistical measures

- While **location measures** provide information about the **centrality of a distribution**, i.e. reflect the **typical** (mode), the **central** (median) or the **average** (arithmetic mean) value of a distribution,

- The **dispersion measures** indicate the extent to which the data of a distribution deviate from these "typical" values.
  In this way, they measure the degree of heterogenity of a distribution and show how well or how poorly the location measures represent a distribution:

    - If we have a relatively **homogeneous distribution**, i.e. only very few measured values deviate from the centrality values (= low dispersion), then the **centrality values are very good representatives** of the distribution.

    - If instead we have a very **heterogeneous distribution**, i.e. the measured values deviate quite strongly from the centrality values (= high dispersion), then the **centrality values do not represent the distribution** very well.

What is the Mode ($´X´_M$)? (Read X across m)?

The 'MO' mode $´X´_M$

- The simplest positional measure is the mode or modal value
- To determine the modal value, one asks the simple question *"Which value occurs most frequently?"* It is applicable from nominal measurement level!

- The mode is the value that occurs most frequently in a data set
- There can be multiple modes in a record or there can be no mode at all
- If serval values share the greatest frequency of occurrence, each of them is a mode
- The frequencies producure reports only the smallest of such multiple modes

Examples of the Mode

X_i	h_i	
1	1	
2	**12**	
3	10	
4	5	
5	2	
7	1	
Total	n = 31	

Here h = 12 with x_i = 2; $\grave{X}_M$=2
Since we only have one modal value here, it is a unimodal distribution.
There may not be a clear modal distribution!

Caution: Do not confuse with the h_i value; for categorical (Nominal, ordinal) characteristics, the mode is the most frequent measured value x_i . For cardinal characteristics, it is the measured value itself.

The concept of mode is consistently applicable to nominal, ordinal as well as cardinal/scale features. For classified data, one speaks of modal class instead of mode.

Robustness: The value of the mode is not influenced by outliers or extreme values for cardinal characteristics

What is the „Mean" Value (´X´)? (Read X across)

- Mean is the sum of all data divided by the number of data

- There are different types of means, including **the Arithmetic Mean (AM)**, the **Geometric Mean (GM)** and the **Harmonic Mean (HM)**

- **-> For our purposes in most cases the AM is important, sometimes also the GM**

- Formula for the **Arithmetic Mean**: $´X´ = (x_1 + x_2 + x_3 + ... + x_n) / n$

Examples of the Arithmetic Mean

- Data set: 1,2,3,4,5

- Number of data (n): 5

- Arithmetic mean (AM): (1+2+3+4+5) / 5 = 3

x_i	h_i	f_i
1	1	0,03
2	12	0,39
3	10	0,32
4	5	0,16
5	2	0,06
7	1	0,03
'x`=22/7=3,66	n = 31	

In case of a frequency distribution the following formula is used:

$$`AM` = \frac{(1 \times 1) + (2 \times 12) + (3 \times 10) + (4 \times 5) + (5 \times 2) + 7 \times 1)}{31} = 2,97$$

$$`AM` = 1 \times 0{,}03 + 2 \times 0{,}38 + 3 \times 0{,}32 + 4 \times 0{,}16 + 5 \times 0{,}06 + 7 \times 0{,}03 = 2{,}97$$

The arithmetic mean is generally calculated for multiple frequencies as follows (the h_i or f_i are the weights):

$$`AM` = \frac{h_1 \cdot x_1 + h_2 \cdot x_2 + h_3 \cdot x_3 + \ldots + h_n \cdot x_n}{n} = 1/n \sum_{i=1} h_i \cdot x_i = \sum_{i=1} f_i \cdot x_i$$

What is the „Median" ($\acute{X}\acute{}_Z$)?

- The median is the value that lies in the middle of a sorted data set

- The median $'X'_Z$ is the value that halves a series of measurements ordered by size. It is applicable from ordinal measurement level to scale.

- The median $'X'_Z$ is that characteristic value of a characteristic X which 50 % of all characteristic values fall below or at most reach and which 50 % of all characteristic values exceed or at least reach.

- If the data set has an **odd number of values**, the median is simply the value in the middle

- If the data set has an **even number of values**, the median is the artithmetic mean of the two values in the middle

- The median is a measure of central tendency not sensitive to outlying values (unlike the mean, which can be affected by a few extremly high or low values).

- Is well suited for skewed data, i.e. little susceptible to outliers and extreme values

- This requires an ordinal or metric (cardinal) scale for determination and interpretation

Example of the Median (odd number of values)

a) Median calculation for <u>odd number of</u> cases (`M'$_{ZO}$)

The median here is *the measured value of the middle case of an ordered series of measured values.* Therefore, first order the data according to size, as in this ordered master list:

1,2,2,2,3,3,3,4,4,4,4,5,5,7,7 ; n = 15 (= odd number)

The median - the middle case - is calculated with the following formula:

$(n + 1) / 2$

Here: $(15 + 1) / 2 = 8$

Attention: Not 8, but the *measured value of the 8th. case* is the median;

Therefore M_{ZO} = 4. As can also be seen from the cumulated absolute frequencies, the 16th case belongs to measured value 3.

Examples of the Median (even number of values)

a) Median calculation for *even number of cases* (`M'$_{ZE}$)

For an even number of cases, the median is the halved value of the middle two cases. First, the middle two cases are calculated with:
$X_{n/2}$ = 1st. middle case and $X_{n/2 +1}$ = 2nd. middle case

Ex.: 1,2,2,2,2,3,3,3,4,4,4,4,5,5,7,7 ; n = 16 (= even number)

16 / 2 = 8th. case (= reading 3); (16 / 2) +1 = 9th. case (= reading 4)
'M'$_e$ = ½ (3 + 4) = 3.5

The median is now nothing other than the *Arithmetic Mean* of the middle two measured values:
In general, the formula looks like this: 'M'$_{ZE}$ = 1/2 $(X_{(n/2)} + X)_{(n/2)+1}$

Note again that the two middle cases [n/2 or (n/2)+1] are determined, which must then be replaced by their corresponding measured values in the case of non-cardinal data. For cardinal data, it is the value itselve.

How to calculate the Median of a set of data in SPSS

1. Click 'Analyses' → 'Descriptives Statistics' → 'Frequencies'

2. Move the variable for which you wish to calculate the median into the right-hand column

3. Click the 'Statistics button', select 'Median' under 'Central Tendency' and then press 'Continue'

4. Click 'OK' to perform the calculation
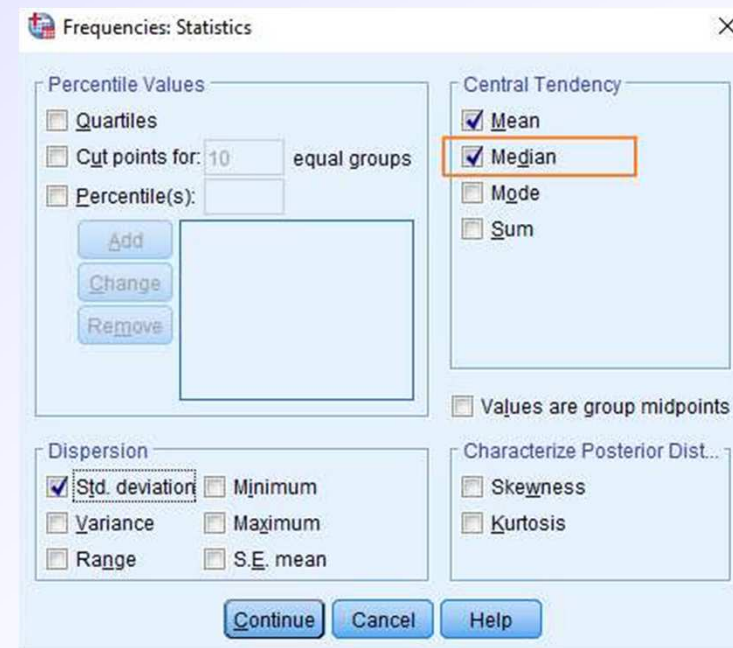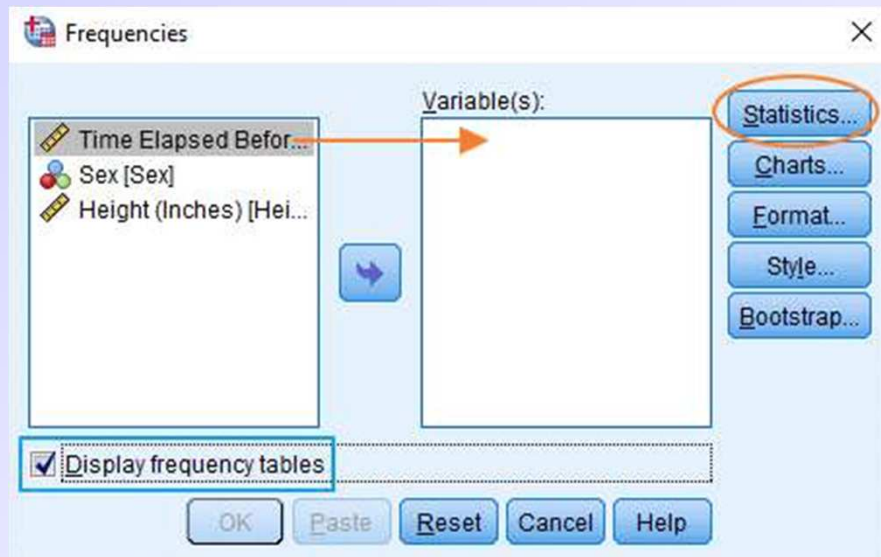
Measures of centrality in comparison

Positive Skew: Mode, Median, Mean
Symmetrical Distribution: Mean, Median, Mode
Negative Skew: Mean, Median, Mode
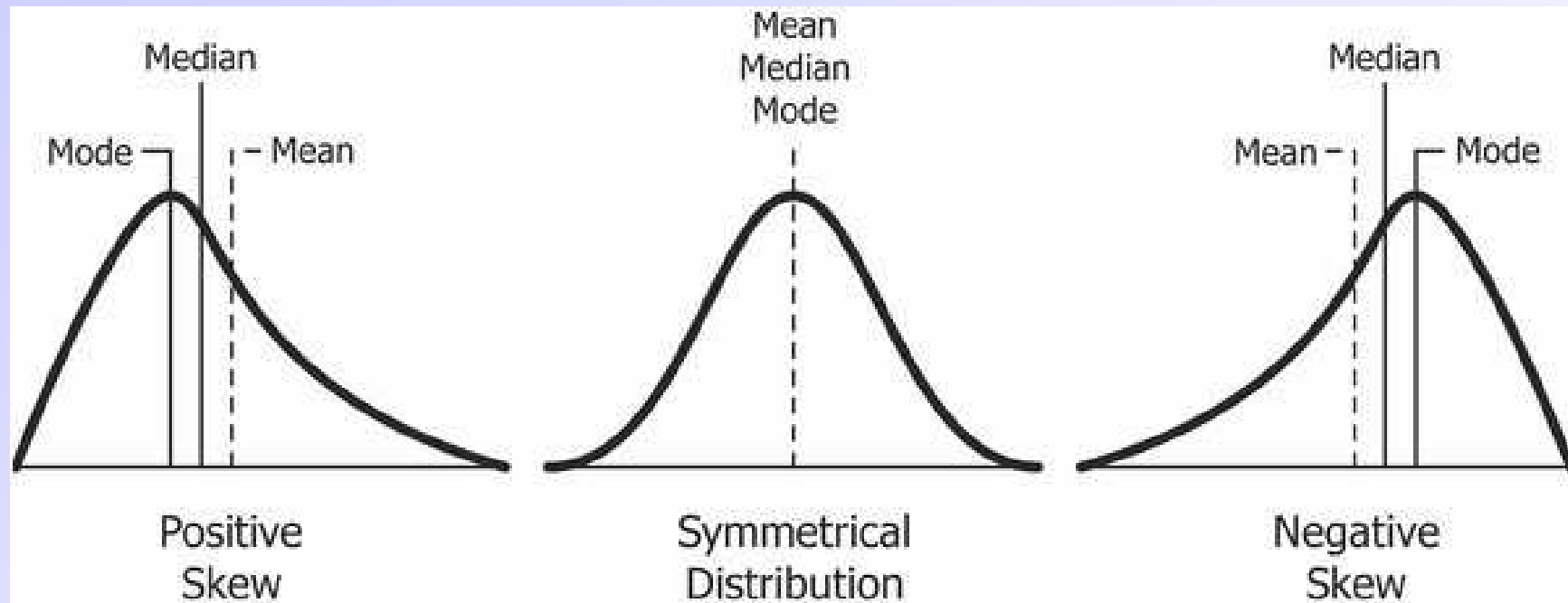
Exercise:

Calculate all three measures of centrality for :

- **AGE**
- **NOFV "Number of festivals visited",**
- *V12 "Recommendation of festival",*
- **NOA "Number of activties",**
- **"BML" / "Broadness of Music Liked"**

How to compare means in subgroups

a) Simple comparisons

```
SPSS command:
ANALYZE / COMPARE MEANS / MEANS
Put the metric variable into dependent, the independent
variable is the group variable.
TABLES=NOA BY A_G3
  /CELLS MEAN COUNT STDDEV.
```

b) Comparison with statistical test on signifikance

```
SPSS command:
ANALYZE / COMPARE MEANS / ONEWAY ANOVA
Set under options "Descriptives" and put the metric variable
into dependent, the independent variable is the group
variable. Every result with sig. bigger than ,05 is
significant!
ONEWAY V2 BY A_G3
  /STATISTICS DESCRIPTIVES
  /MISSING ANALYSIS.
```

Exercise:

Calculate all three measures of centrality in SYS-Data file for :

- **AGE**
- **NOFV "Number of festivals visited",**
- ***V12 "Recommendation of festival",***
- **NOA "Number of activities",**
- **"BML" / "Broadness of Music Liked"**

- **Within the subgroups of**
- **AGE_3**
- **Gender**
- **Residence_2**

Check for significance of the differences!

What are statistical measures and measures of dispersion?
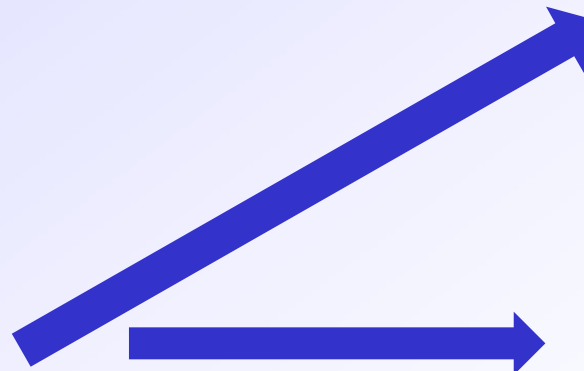
SPSS has two primary options for calculating statistics:

- **Descriptives:** For basic statistics – mean, median, range and standard deviation
- **Frequencies:** Additional options – quartiles, percentiles and more dispersion statistics
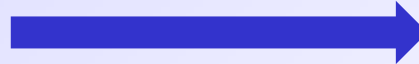
Statistical measures → **Descriptives**

Measures of dispersion → **Frequencies**

Descriptives

- The Descriptives procedure displays univariate summary statistics for several variables in a single table and calculates standardized values (z scores).

- Variables can be ordered by the size of their means (in ascending or descending order), alphabetically, or by the order in which you select the variables (the default)
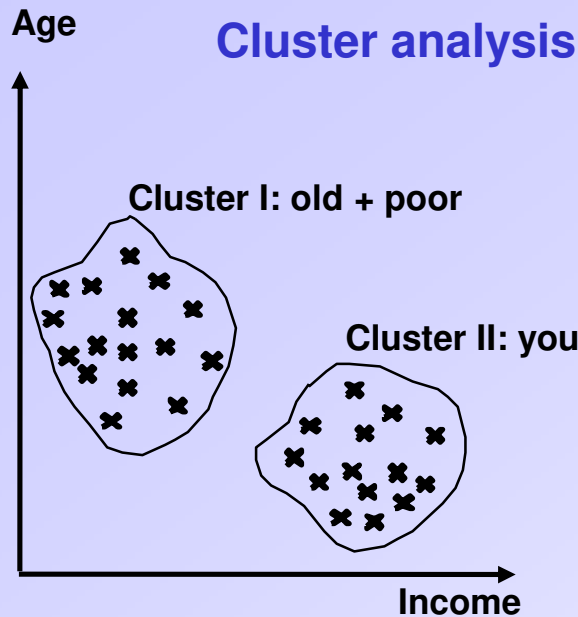
Types of Measures of Dispersion

- While mean values represent typical values of a sample, measures of dispersion are supposed to indicate whether the characteristic values are close to the mean value or at a greater or lesser distance from it.

- They "qualify" the quality of the position measure

- However, distances can only be measured for cardinal characteristics. Therefore, measures of dispersion are useful to define especially for cardinal characteristics - for ordinal characteristics also the quartile distance.

There are several types of measures of dispersion, including:

- Variance (s^2)

- Standard deviation (s)

- Interquartile range (IQR)

Meaning of Dispersion

Cluster analysis

Age

Cluster I: old + poor

Cluster II: young + rich

Income

Discriminant analysis

Visitor loyalty

Not loyal visitors

loyal visitors

Age income, gender, place of living
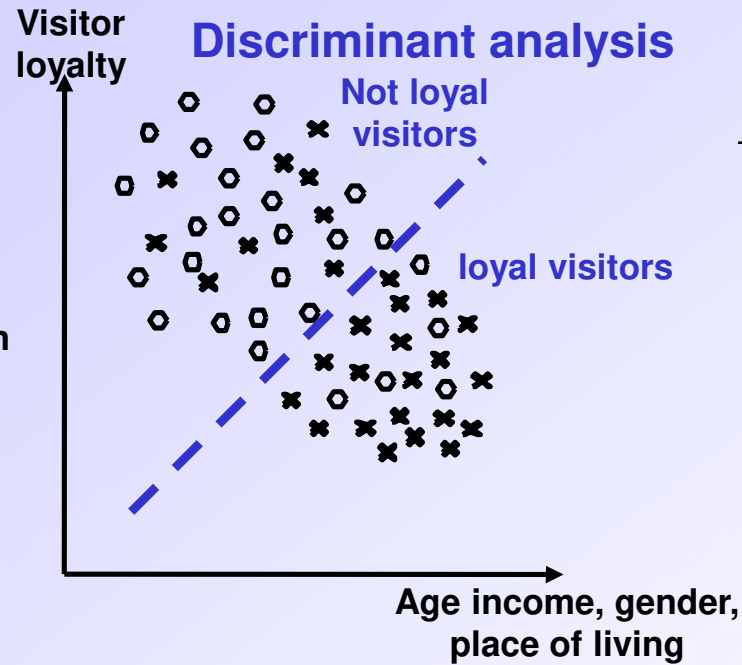
Factor analysis

Variables	F1	F2	F3	F4
1	.1	.2	**.7**	.2
2	.2	**.6**	.2 .	0
3	**.4**	**.5**	.2	**.7**
...				
13				

**Factor 1= Var.3 +...
Factor 2= Var. 2+ 3 +...
etc.**

Analysis of variance

	Region I	Region II	Region III
Jan-April	Type A	Type B	Type C
May-Aug.	Type B	Type C	Type A
Sept.-Dec.	Type C	Type A	Type B

Regression analysis

Number of visits

Visitor satisfaction
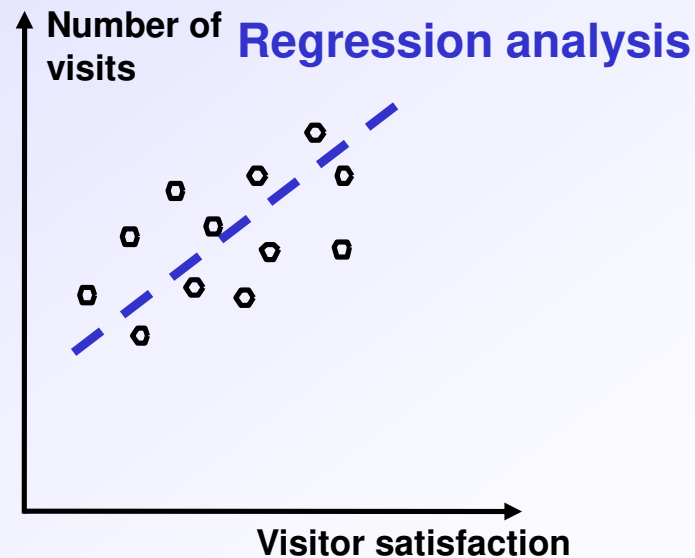
Multidimensional Scaling: Film Festivals

SYS Turino

Venice

Berlin

Locarno

Cannes

Shanghai

Karlsbad
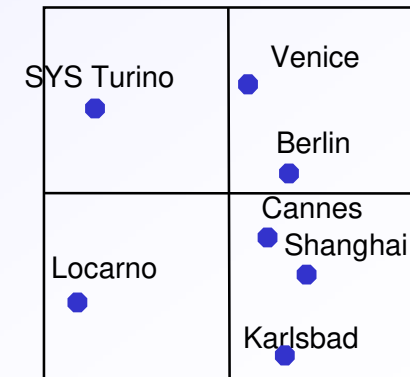
What is Variance?

- The variance is a measure of how far each value in the data set deviates from the mean value
- It shows us how "broadly" or "narrowly" a data set is distributed
- Formula for the variance: $s^2 = \sum(x_i - \acute{X}\acute{})^2 / n$
- Represents the basis for further calculations, e.g. in regression or analysis of variance
- **A high variance shows us, that the mean value does not represent the data very well**
- It is only applicable for scale data, not for ordinal or nominal data

Example:

- Data set: (1+,2+,3+,4+,5)/5
- Mean value ($\acute{X}\grave{}$) = 3
- Variance s^2: $[(1 - 3)^2 + (2 - 3)^2 + (3 - 3)^2 + (4 - 3)^2 + (5 - 3)^2] / 5 = 2$

What is Standard Deviation?

- The standard deviation is the square root of the variance
- It gives the unit of the measurement that corresponds to the data referred (e.g. centimeter, kilometer, euro, whereas variance has no dimension)
- Formula for the standard deviation: $s = \sqrt{s^2}$
- For normally distributed values, approx. 67% of all measured values lie in the intervall described by the mean +- 1 of the standard deviation
- Standard deviation describes the average deviation from the mean
- **A high „s" shows, that the mean value does not represent the data very well**
- It is only applicable for scale data, not for ordinal or nominal data

- Data set: 1,2,3,4,5
- Mean value (´X´): 3
- Variance (s^2): 2 (as described above)
- Standard deviation (s): $\sqrt{2} = 1.41$

What is the "Interquartile Range" (IQR)?

- The interquartile range is the difference between the 75th and 25th percentile of a data set; it encompasses **the middle 50% of the sample**

- An interquartile range is a measure of where 50% of the respondents/values lie

- The interquartile range formula is the first quartile subtracted from the third quartile:

$$IQR = Q_3 - Q_1$$

- The IQA is also robust to outliers and strong deviations

- Data set: 1,2,5,6,7,9,12,15,18,19,27, n: 11

- Q_2: Median: 9

- Q_1: 5 (median in the lower half of the data)

- Q_3: 18 (median of the upper half of the data)

- Interquartile Range: 18 – 5 = 13

Co-funded by
the European Union

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

Exercise:

Calculate all measures of dispersion for :

- **AGE**
- **NOFV "Number of festivals visited",**
- ***V12 "Recommendation of festival",***
- **NOA "Number of activties",**
- **"BML" / "Broadness of Music Liked"**

The data analysis – Box-Plot

- The boxplot (also box diagram or box graph) combines different scatter measures and position measures, and displays them graphically.

- A boxplot shows all values of the five-point summary (minimum, first quartile, median, third quartile and maximum)

- These metrics are very robust to outliers and deviations from the normal distribution

- Boxplot therefore provide quick insights into the distribution of your data regardless of how it is distributed

- It is used to display the distribution of an ordinal or metric variable

- Boxplot also mark possible outliers
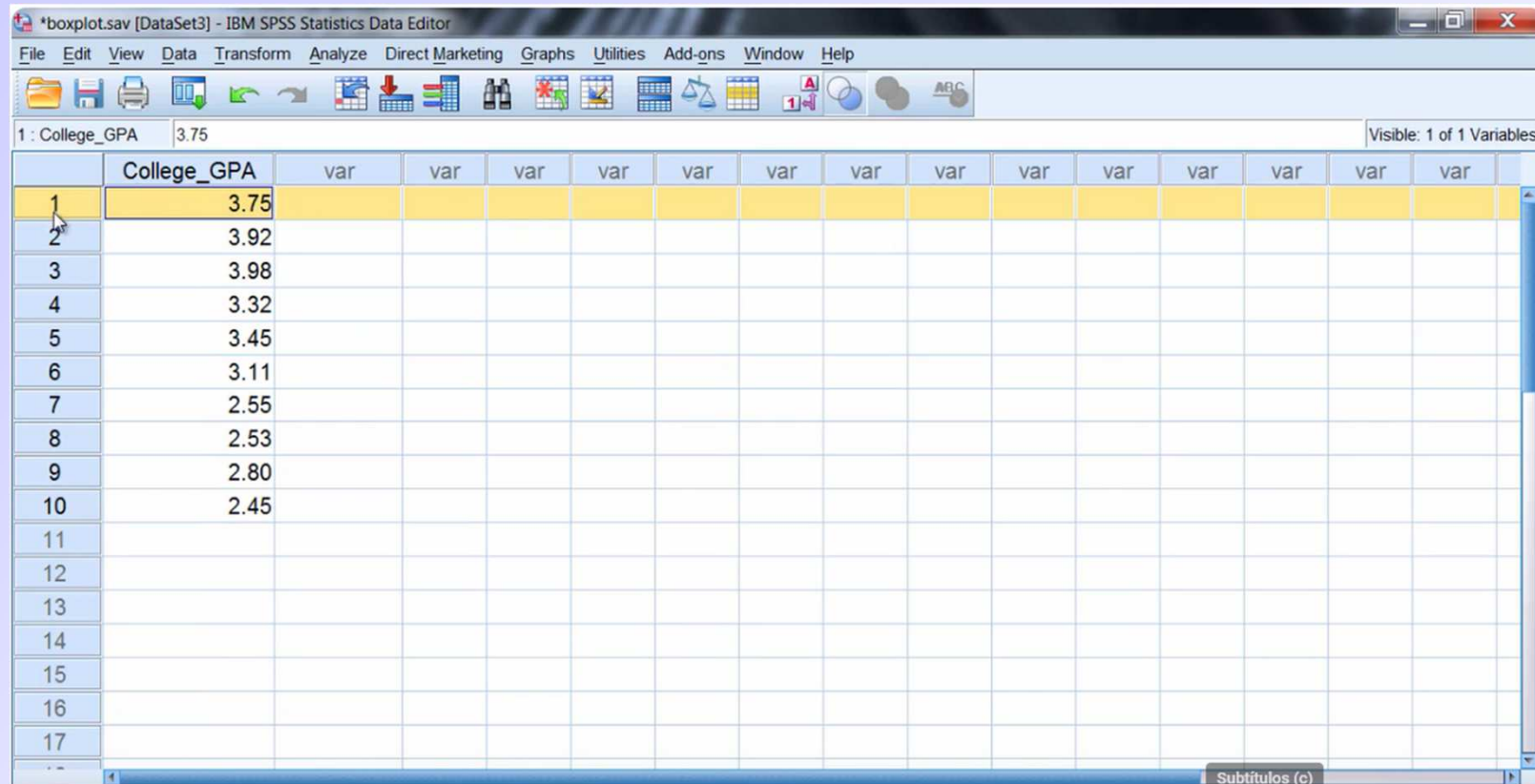
The data analysis – Box-Plot

How to create and interpret a boxplot

The data analysis – Box-Plot

How to create and interpret a boxplot (1/2)

Click the link below to open the video:

https://www.youtube.com/watch?v=UsQTcFMa1_Y&ab_channel=QuantitativeSpecialists

The Data Analysis – Box-Plot
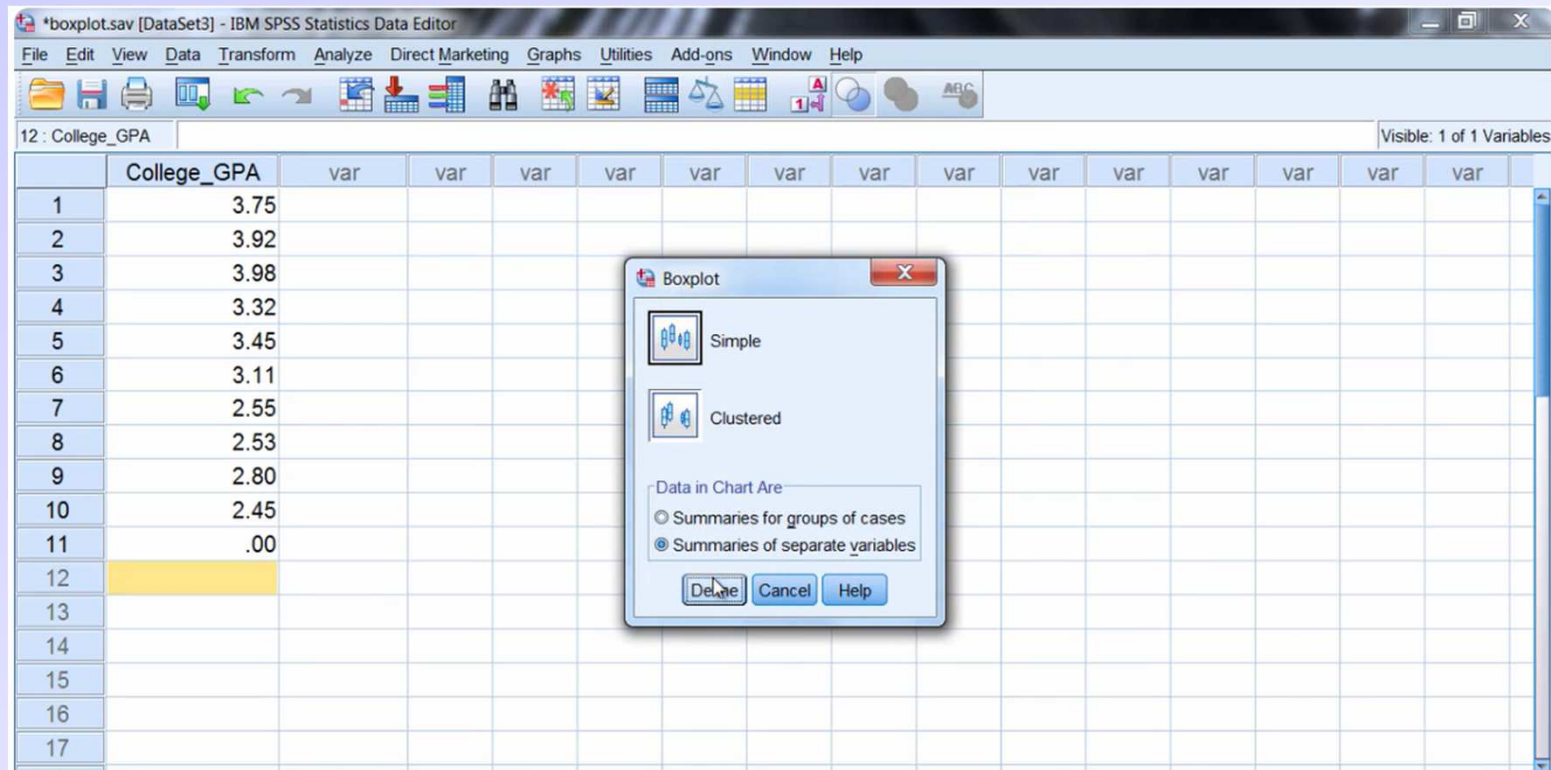
How to create and interpret a boxplot (2/2)

Clic on the picture to watch the video or copy and paste the following link in your browser:
https://www.youtube.com/watch?v=X4hSSu5oTj4&ab_channel=QuantitativeSpecialists

Exercises:

Calculate a boxplot for for :

- **AGE**
- **NOFV "Number of festivals visited",**
- *V12 "Recommendation of festival",*
- **NOA "Number of activties",**
- **BML / "Broadness of Music Liked"**

Caculate a boxplot for NOFV, V12, NOA and BML by age_G3 groups

Interpret the results!

Introduction into statistical Bivariat Analyses: Crosstabs

Definition:

- A crosstab is a special type of table that relates one variable to another variables or more variables. The crosstab shows us how often certain combinations of variables occur.

- It is only applicable to categorical variables.

Why are crosstabs important for data analysis?

- Can help you visualize the **differences between subgroups**

- Shows you **possible relationships between two (or more) variables**

- Makes it easier to interpret data, which is especially benefical if you have limited knowledge of statistical analysis

Crosstabs – rows, columns and table dimension

- Crosstabs are a simple and very efficient tool do **realize relationships between variables**

- Statistical knowledge is no prerequisite for doing analyses

- The dimensions of the crosstab refer to the **number of rows** and **columns in the table** (the "total" rows/column are not included.)

- The table dimensions are reported as R x C, where R is the number of categories for the row variable and C is the number of categories for the column variable.

- Additionally, a "square" crosstab is one in which the row and column variables have the same number of categories.
  Tables of dimensions 2x2, 3x3, 4x4 etc. are all square crosstabs.

Crosstabs – Understanding Row, Column and Total Percent

A typical 2x2 crosstab has the following construction:

- The letters a,b,c and d represent what are called 'cell counts'.
- a = number of observations corresponding to row 1 AND column 1.
- b = number of observations corresponding to row 1 AND column 2.
- c = number of observations corresponding to row 2 AND column 1.
- d = number of observations corresponding to row 2 AND column 2.

	Column 1	Column 2	Row totals
Row 1	a	b	$a+b$
Row 2	c	d	$c+d$
Column totals	$a+c$	$b+d$	$a+b+c+d$

Crosstabs – Understanding Row, Column and Total Percent

By adding a, b, c and d, we can determine the total number of observations in each category and in the table overall.

- **Row sum** of row 1 (i.e., total number of observations in row 1): a + b
- **Row sum** of row 2 (i.e., total number of observations in row 2): c + d
- **Column sum** of column 1 (i.e., total number of observations in column 1): a + c
- **Column sum** of column 2 (i.e., total number of observations in column 2): b + d
- **Total sum** (i.e., total number of observations in the table): n = a + b + c + d

The row sums and column sums are referred to as *'marginal frequencies'*.

When you are describing the composition of your sample, it is often useful to refer to the proportion of the row or column that fell within a particular catefory. This can be achieved by computing the 'row percentages' or 'column percentage'

Crosstabs – Understanding Row, Column and Total Percent

Notice that when computing **row percentage**, the denominators for cells a,b,c, d are determined by the row sums (here, a + b and a + d). The percentage in the 'row totals' column here must equal 100%.

	Column 1	Column 2	Row totals
Row 1	a	b	$a + b$
Row 1 %	$a / (a + b)$	$b / (a + b)$	$(a + b) / (a+b) = 100\%$
Row 2	c	d	$c + d$
Row 2 %	$c / (c + d)$	$d / (c + d)$	$(c + d) / (c + d) = 100\%$
Column totals	$a + c$	$b + d$	$a + b + c + d$
% of total	$(a + c) / (a + b + c + d)$	$(b + d) / (a + b + c + d)$	$(a + b + c + d) / (a + b + c + d) = 100\%$

Crosstabs – Understanding Row, Column and Total Percent

Notice that when **total percentages** are computed, the denominators for all of the computations are equal to the **number of oberservations** in the table, i.e. a + b + c + d.

	Column 1	Column 2	Row totals
Row 1	a	b	$a + b$
% of total	$a/(a+b+c+d)$	$b/(a+b+c+d)$	$(a+b)/(a+b+c+d)$
Row 2	c	d	$c + d$
% of total	$c/(a+b+c+d)$	$d/(a+b+c+d)$	$(c+d)/(a+b+c+d)$
Column totals	$a + c$	$b + d$	$a + b + c + d$
% of total	$(a+c)/(a+b+c+d)$	$(b+d)/(a+b+c+d)$	$(a+b+c+d)/(a+b+c+d) = 100\%$

Components of a crosstab

Example: Summarizing the relationships between Variables:

- Let's use different aspects of the Crosstabs procedure to investigate the relationship between class rank and living on campus

- There are several variables relating to this question:

    - Rank: Class rank – Freshmen, Sophomore, Junior, Senior

    - LiveOnCampus: Do you live on campus? – Yes/No

    - State: Are you an in-state or out-of-state student? – In State, Out of state

Crosstabs – Understanding Row, Column and Total Percent

Simple Crosstabs: Output

- The second table Class Rank: Do you live on campus? (Crosstabulation) contains the crosstab
- We can quickly observe information about the interaction of these two variables:
- Many more **freshman** (1st. year student) lived on-campus (100) than off-campus (37)
- About an equal number of **sophomore** (students in 2nd. year) lived off-campus (42) versus on-campus (48)
- Far more **juniors** (3rd. year students) lives off-campus (90) than on-campus (8)
- Only one (1) **senior** (4th. year students) lives on-campus, the rest lived off-campus (62)

Class rank * Do you live on campus? Crosstabulation

Count

		Do you live on campus?		Total
		Off-campus	On-campus	
Class rank	Freshman	37	100	137
	Sophomore	42	48	90
	Junior	90	8	98
	Senior	62	1	63
Total		231	157	388

Crosstabs – Row, Column and Total percentage

Output – Row percentage:

- If the row variable is 'RankUpperUnder' and the column variable is 'LiveOnCampus', then the row percentages will tell us what percentage of the upperclassmen (year 3 + 4) or what percentage of the underclassmen (first 2 years) live on campus

- Variable RankUpperUnder will determine the denominator of the percentage computation

Class Rank * Do you live on campus? Crosstabulation

			Do you live on campus?		Total
			Off-campus	On-campus	
Class Rank	Underclassman	Count	79	148	227
		% within Class Rank	34.8%	65.2%	100.0%
	Upperclassman	Count	152	9	161
		% within Class Rank	94.4%	5.6%	100.0%
Total		Count	231	157	388
		% within Class Rank	59.5%	40.5%	100.0%

Crosstabs – Row, Column and Total percentage

Output – Column percents:

- If the row variable is *'Rank Upper / Under'* and the column variable is *'Live On Campus'*, then the column percentage will tell us what percentage of the individuals who live on campus are upper (Junior/Senior) or underclassmen (Fresh/Sopho)

- Variable *'LiveOnCampus'* will determine the denominator of the percentage camputation

Crosstabs – Row, Column and Total percentage

Output – Row percentage:

- if the row variable is '*Rank Upper Under*' and the column variable is *'Live On Campus'*, then the **row percentages** will tell us what percentage of the upperclassmen or what percentage of the underclassmen live on campus
- Variable RankUpperUnder will determine the denominator of the percentage computation

Class Rank * Do you live on campus? Crosstabulation

			Do you live on campus?		Total
			Off-campus	On-campus	
Class Rank	Underclassman	Count	79	148	227
		% within Class Rank	34.8%	65.2%	100.0%
	Upperclassman	Count	152	9	161
		% within Class Rank	94.4%	5.6%	100.0%
Total		Count	231	157	388
		% within Class Rank	59.5%	40.5%	100.0%

Crosstabs – Row, Column and Total percentage

Output – Column percentage:

- if the row variable is *'Rank Upper Under'* and the column variable is *'Live On Campus'*, then the **column percentage** will tell us what percentage of the individuals who live on campus are upper or underclassmen
- Variable *'Live On Campus'* will determine the denominator of the percentage camputation

Class Rank * Do you live on campus? Crosstabulation

			Do you live on campus? Off-campus	Do you live on campus? On-campus	Total
Class Rank	Underclassman	Count	79	148	227
		% within Do you live on campus?	34.2%	94.3%	58.5%
	Upperclassman	Count	152	9	161
		% within Do you live on campus?	65.8%	5.7%	41.5%
Total		Count	231	157	388
		% within Do you live on campus?	100.0%	100.0%	100.0%

Crosstabs – Row, Column and Total percentage

Output – Total percentage:

- If the row variable is '*Rank Upper / Under*' and the column variable is 'LiveOnCampus', then the **total percentage** tells us what proportion of the total is within each combination of '*Rank Upper Under*' and '*Live On Campus*'

- The overall table size determines the denominator of the percentage computations

Class Rank * Do you live on campus? Crosstabulation

			Do you live on campus?		Total
			Off-campus	On-campus	
Class Rank	Underclassman	Count	79	148	227
		% of Total	20.4%	38.1%	58.5%
	Upperclassman	Count	152	9	161
		% of Total	39.2%	2.3%	41.5%
Total		Count	231	157	388
		% of Total	59.5%	40.5%	100.0%

Some general rules when doing crosstabs

Have the following in mind:

- **Follow the KISS – Rule (keep ithe table simple and stupid)**

- **Verify the table first by taking a look at the absolute numbers (total of each row or column should be 80 minimum)**

- **Think before, what you want to see from the table**

- **Show only one kind of result in one table: Row percent <u>or</u> column percent <u>or</u> total percent!**

- **Show the percents *preferably always in column percent***

- **If you think, there may be an independent and a dependent variable, put the independent variable always in the *head* of the table**

Bad example: Avoid this!

Attending film festivals * Age Crosstabulation

			Age 18 - 34	Age 35 - 49	Age 50 - 64	Age 65 +	Total
Attending film festivals	Frequent	Count	9	4	3	1	17
		% within Attending film festivals	52,9%	23,5%	17,6%	5,9%	100,0%
		% within Age	5,1%	3,2%	2,3%	,8%	3,1%
		% of Total	1,6%	,7%	,5%	,2%	3,1%
	Occasionally	Count	25	18	5	5	53
		% within Attending film festivals	47,2%	34,0%	9,4%	9,4%	100,0%
		% within Age	14,2%	14,5%	3,8%	4,1%	9,6%
		% of Total	4,5%	3,3%	,9%	,9%	9,6%
	Rare	Count	38	28	26	12	104
		% within Attending film festivals	36,5%	26,9%	25,0%	11,5%	100,0%
		% within Age	21,6%	22,6%	19,7%	9,9%	18,8%
		% of Total	6,9%	5,1%	4,7%	2,2%	18,8%
	Never	Count	104	74	98	103	379
		% within Attending film festivals	27,4%	19,5%	25,9%	27,2%	100,0%
		% within Age	59,1%	59,7%	74,2%	85,1%	68,5%
		% of Total	18,8%	13,4%	17,7%	18,6%	68,5%
Total		Count	176	124	132	121	553
		% within Attending film festivals	31,8%	22,4%	23,9%	21,9%	100,0%
		% within Age	100,0%	100,0%	100,0%	100,0%	100,0%
		% of Total	31,8%	22,4%	23,9%	21,9%	100,0%

Better example!

1. First or repeated visit? * B. Gender Crosstabulation

% within B. Gender

		B. Gender		
		Female	**Male**	**Total**
1. First or repeated visit?	**first time**	**42%**	**59 %**	**49 %**
	repeated	**58 %**	**41 %**	**52 %**
Total		**100 %**	**100 %**	**100 %**

Doing Crosstabs with other data analyses software

Creating crosstabs using excel

Step 1: Enter the Dataset into

Excel

	A	B	C	D	E
1	**Team**	**Position**	**Points**		
2	A	Guard	12		
3	A	Guard	19		
4	A	Forward	22		
5	A	Forward	24		
6	A	Forward	17		
7	A	Center	29		
8	B	Guard	32		
9	B	Guard	33		
10	B	Guard	19		
11	B	Forward	9		
12	B	Center	8		
13	B	Center	14		
14					
15					
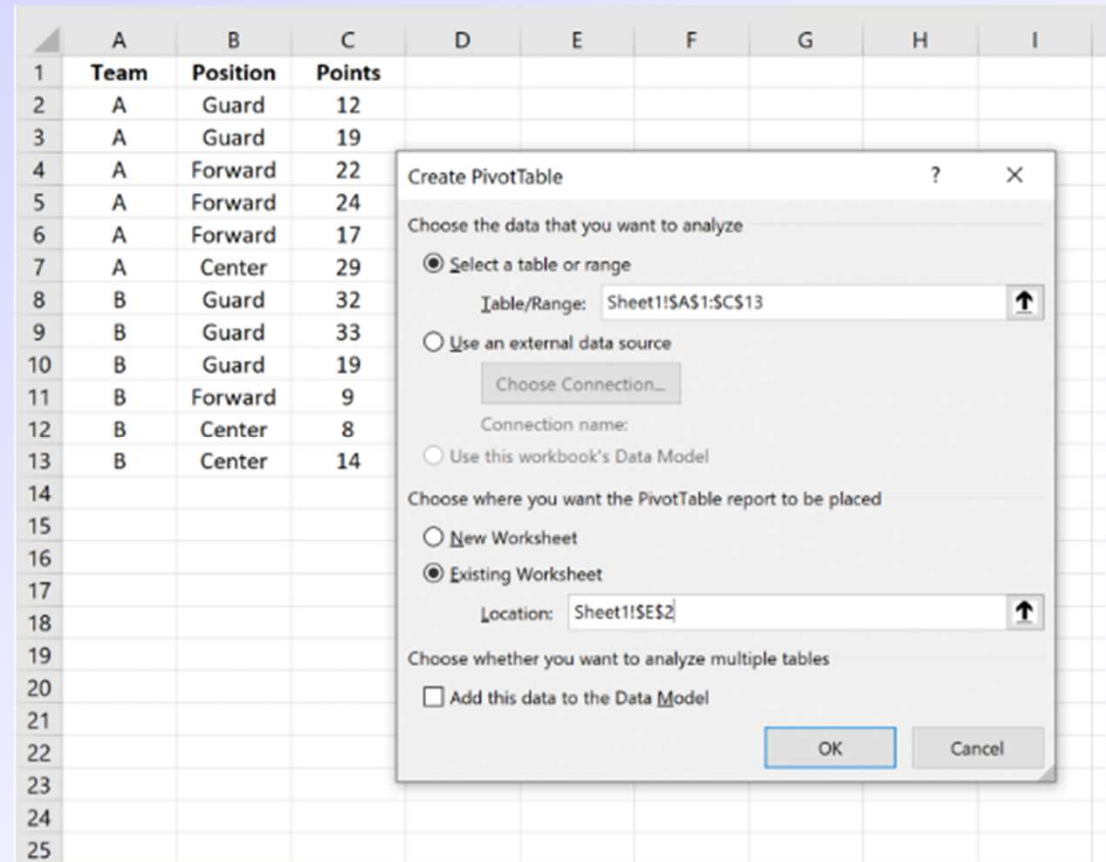
Step 2: Create the Crosstab

Click the **Insert** tab along the top

line and then click the **PivotTable**

button

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

Creating crosstabs using excel

In the new window that appears, select the range that contains the data as the **Table**/**Range** and choose any cell you'd like in the **Existing Worksheet** to place the crosstab. We'll choose cell **E2**

	A	B	C
1	Team	Position	Points
2	A	Guard	12
3	A	Guard	19
4	A	Forward	22
5	A	Forward	24
6	A	Forward	17
7	A	Center	29
8	B	Guard	32
9	B	Guard	33
10	B	Guard	19
11	B	Forward	9
12	B	Center	8
13	B	Center	14

Create PivotTable ? ✕

Choose the data that you want to analyze
- ⦿ Select a table or range
  - Table/Range: Sheet1!$A$1:$C$13
- ○ Use an external data source
  - Choose Connection...
  - Connection name:
- ○ Use this workbook's Data Model

Choose where you want the PivotTable report to be placed
- ○ New Worksheet
- ⦿ Existing Worksheet
  - Location: Sheet1!$E$2

Choose whether you want to analyze multiple tables
- ☐ Add this data to the Data Model

OK   Cancel

Creating crosstabs using excel

Step 3: Populate the Crosstab with Values

Once you click OK, a new window on the right side of the screen will appear.

Drag the **Team** variable to the **Rows** area, the **Position** variable to the **Columns** area, then the **Position** variable again to the **Values** ares as follows:
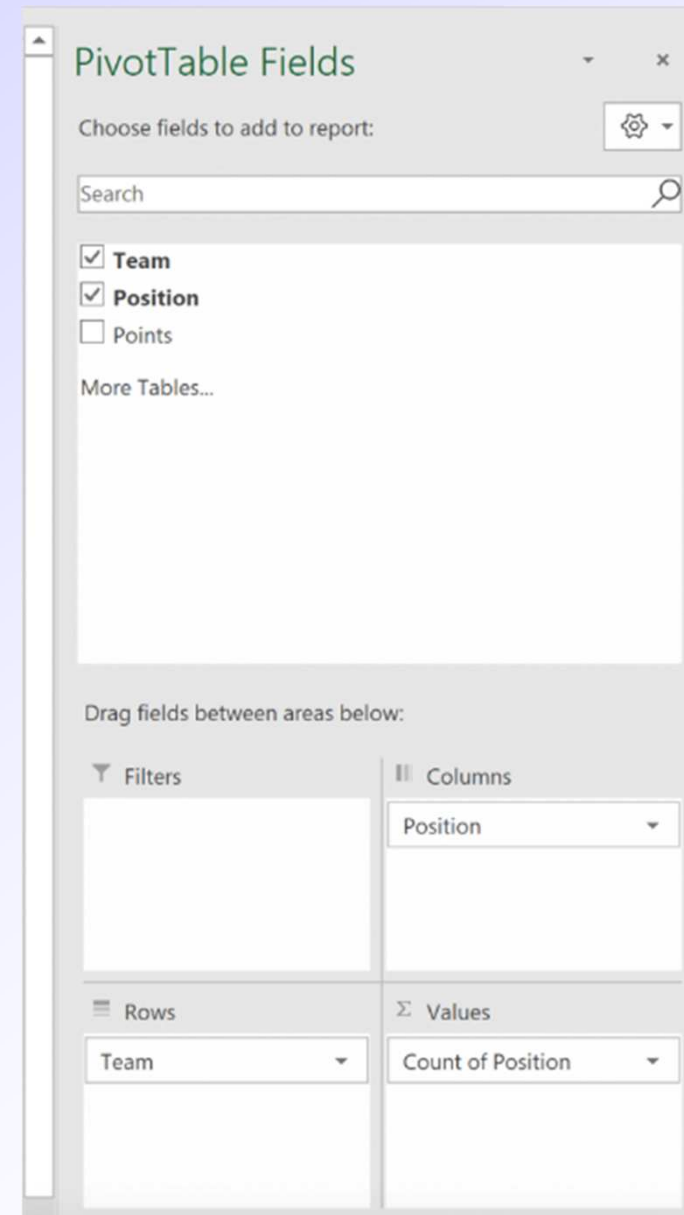
Creating crosstabs using excel

Once you do so, the following crosstab will appear in the cell that you specified:

	A	B	C	D	E	F	G	H	I
1	Team	Position	Points						
2	A	Guard	12		Count of Position	Column Labels			
3	A	Guard	19		Row Labels	Center	Forward	Guard	Grand Total
4	A	Forward	22		A	1	3	2	6
5	A	Forward	24		B	2	1	3	6
6	A	Forward	17		Grand Total	3	4	5	12
7	A	Center	29						
8	B	Guard	32						
9	B	Guard	33						
10	B	Guard	19						
11	B	Forward	9						
12	B	Center	8						
13	B	Center	14						
14									

Creating crosstabs using excel

Step 4: Interpret the Crosstab

Row totals:

A total of 6 players are on team A and a total of 6 players are on team B

Column Totals:

- A total of 3 players have a position of Center
- A total of 4 players have a position of Foward
- A total of 5 players have a position of Guard

Individual Cells:

- 1 player has a position of Center on team A
- 3 players have a position of Foward on team A
- 2 players have a position of Guard on team A
- 2 players have a position of Center on team B
- 1 player has a position of Forward on team B
- 3 players have a position of Guard on team B

Creating crosstabs using Google Sheets

Step 1: Enter the following dataset into Google Sheets that shows information for variour basketball players

	A	B	C	D	E
1	Team	Position	Points		
2	A	Guard	12		
3	A	Guard	19		
4	A	Forward	22		
5	A	Forward	24		
6	A	Forward	17		
7	A	Center	29		
8	B	Guard	32		
9	B	Guard	33		
10	B	Guard	19		
11	B	Forward	9		
12	B	Center	8		
13	B	Center	14		
14					

Creating crosstabs Google Sheets

Step 2: Click the "insert" tab along the top line and then click "Pivot table" from the dropdown menu
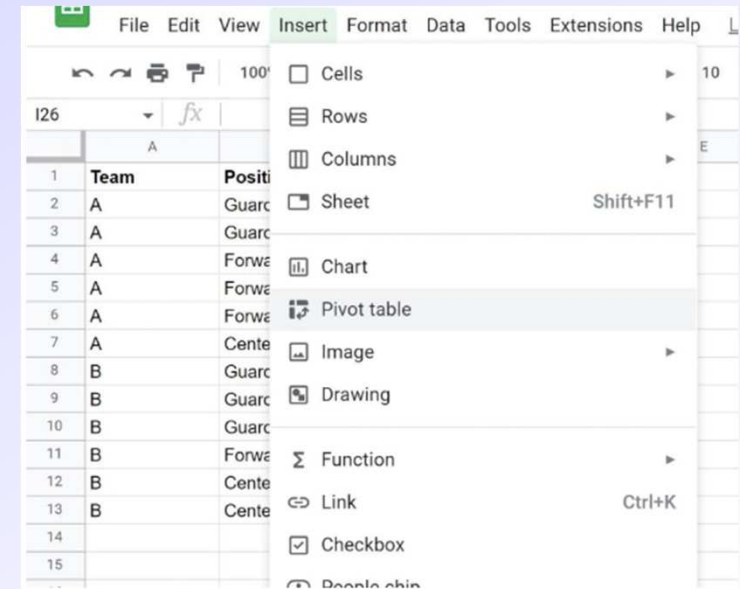
In the new window that appears, enter **Sheet1!A1:C13** as the Data range and **Sheet1!E1** as the Insert to location, then click **Create**:
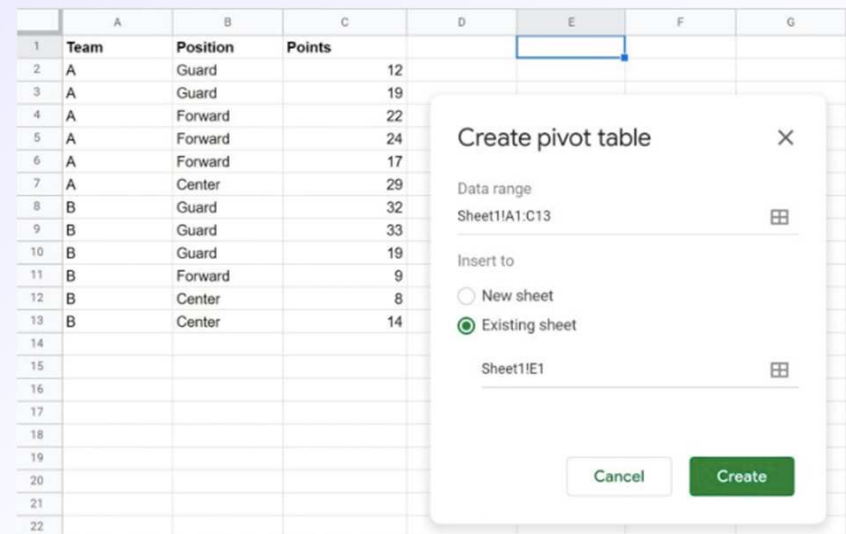
Creating crosstabs Google Sheets

Step 3: Once you click **Create**, a new Pivot table editor panel will appear on the right side of the screen.

Choose Team for the **Rows**, Position for the Columns and Points for the **Values**:

Once you do so, the following crosstab will appear in the cell that you specified:

Co-funded by
the European Union

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln
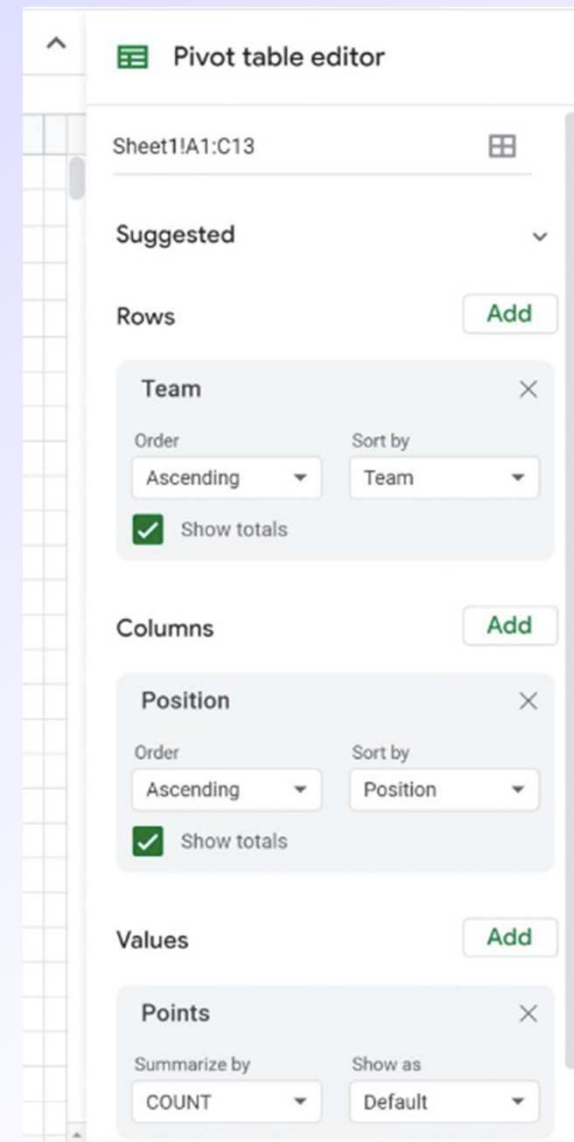
Creating crosstabs Google Sheets

COUNT of Point	Position			
Team	Center	Forward	Guard	Grand Total
A	1	3	2	6
B	2	1	3	6
Grand Total	3	4	5	12

Step 4: Interpret the Crosstab

Row Grand Total:
• A total of 6 players are on team A and a total of 6 players are on team B
Column Grand Totals:
• A total of 3 players have a position of Center; A total of 4 players have a position of Forward and a total of 5 players a position of Guard
Individual Cells:
• 1 player has a position of Center on team A, 3 player have a position of Forward on team A and two players have a position of Guard on team A.
• 2 players have a position of Center on team B, 1 player has a position of Forward on team B and 3 player have a position of Guard on team B.

Saving the file

- The file should always be saved in order to save the work that has been done to date

- In order to save the data you have to be in the data view window

- Click 'File' and after that's 'Save as'

- You can save the file in different forms by clicking 'save as type'

- The saved data should always end with 'sav' in order to open it in SPSS again

- To finish saving the file click 'save'

Before doing any kind of analysis beware of the following:

- **Do a frequency command and check for values that should not be there**

- **Remove resp. correct the wrong values (into same/other variable)**

- **Build the classes or compute new variables**

- **Check if everything is correct by frequency command**

- **Save the data file under a new name (i.e. under the current date)**

Recoding variables – Recode into *Different Variable*

1. From the menu, choose '**Transform**' → '**Recode into Different Variables**'. The 'Recode into different variables' dialog box will appear.

2. Select the variable you want to recode

3. In the '**Output Variable**' area, enter the name for the new variable and click '**Change**'

4. Click '**Old and New Value**' to specify how to recode values.

5. Specifiy an old value and a new value. Click '**Add**' to place the specification into the **Old → New** list. In this example, the age variable is recoded into four age groups (below 20,21 to 40, 41 to 60,61 and older.)

6. Click '**Continue**' and return to the previous dialog box.

7. Click '**OK**'.

Recording Variables
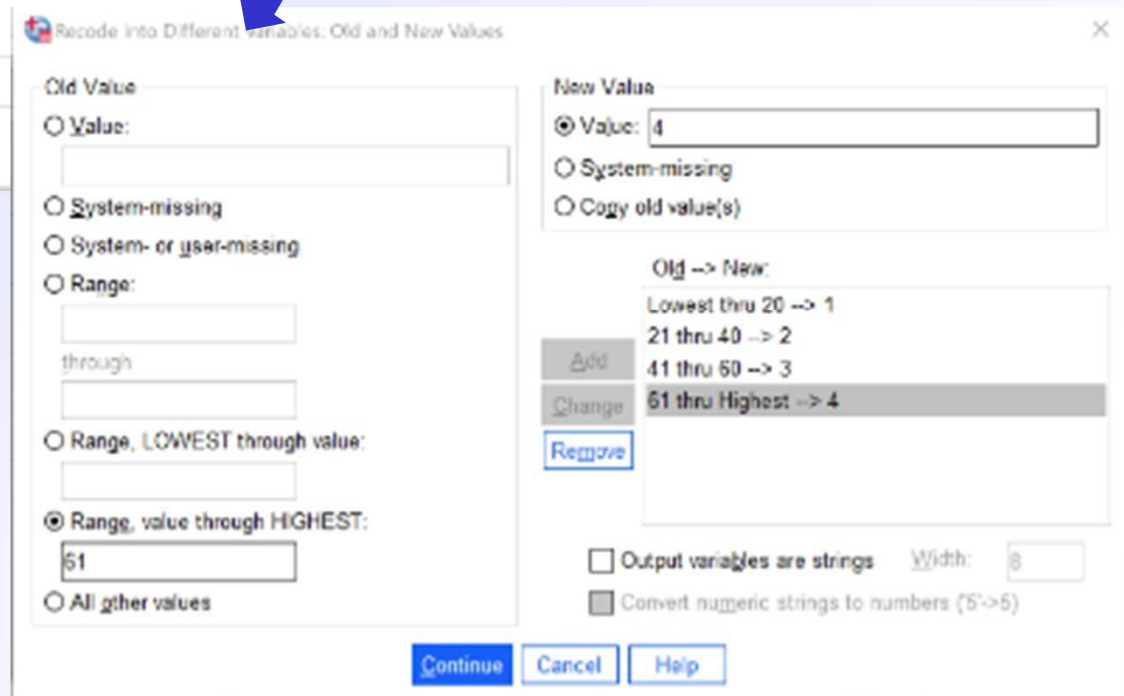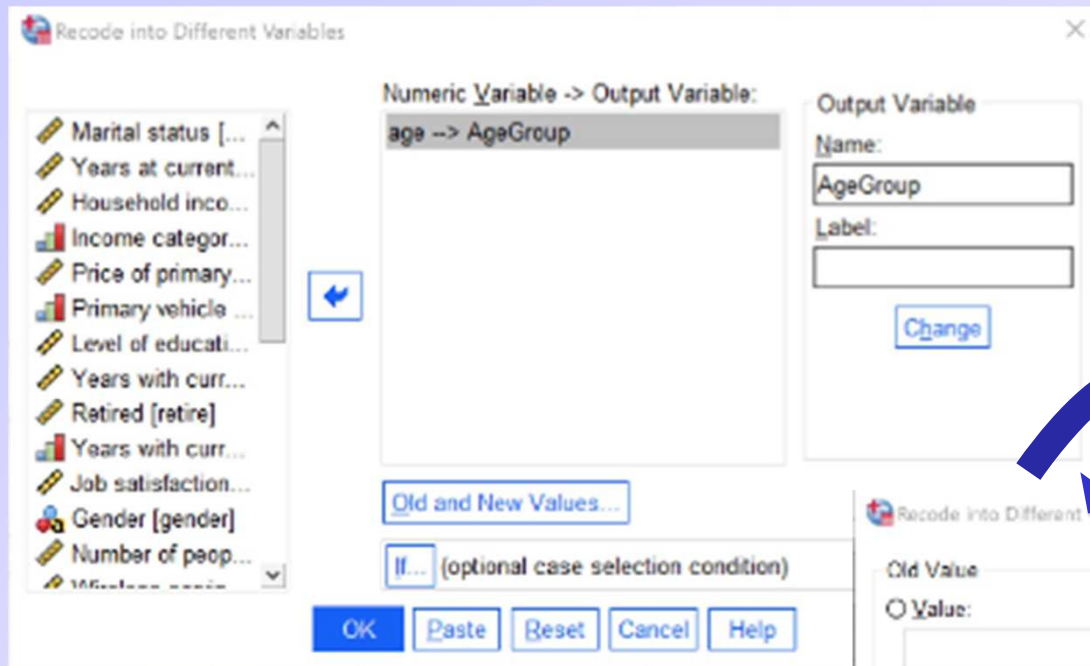
Recode Variables – Recode into *Same Variable*

- Works the same way as 'Recode into different variables' (transform → Recode intro Same Variables), except for that any changes made will permanently alter the original variable.

- The original values will be replaced by the recoded values

- In general, it is good practice not to recode into the same variable because it overwrite the original variable.

- If you ever need to use the variable in its original form (or wanted to check your steps), that information would be lost

Recode Variables – DO IF Syntax

- Performs similarly to the recode procedures, but allows for more control over specifying numeric ranges.

- If you want to perform an recode under certain conditions, or if you want to perform a recoding based on more than one variable, you'll need to use 'DO IF-ELSE IF Syntax'.

- This is given, if you want to recode a variable under certain conditions

EXERCISE CHECKING DATA AND RECODING/COMPUTING VARIABLES

LOAD THE SYS-KOELN_DATA FILE:

Go into the Variable view resp. Data View

- **Check and correct the <u>measure</u> assignments of all variables**
- **Check the variables GOING to the, GENDER and AGE for wrong values**
- **Remove the wrong values from the data file <u>manually</u> and by <u>using a recode order</u> (making them to missings)**
- **Recode E into a new variable "A" with "age" with 3 groups**
- **Compute an Index of number of liked music genres V9.1 to V9.12 by summing up all values "2" and saving them in a new variable**
- **Compute a count command in order to find out, how many persons do not do a single activity from V7.1 to V7.4**
- **Find out, how many sources of awareness the common vistor has. Hwo many were not reached by a single channel?**

EXERCISES DOING FREQUENCIES AND COMPUTING MEASURES OF CENTRALITY (MC) AND DISPERSION (MD)

FILM_KOELN_

- **Do frequencies command for all variables and interpret the results**

- **Compute applicable measures of centrality and dispersion**

- **Recode V4 into a new variable V4_Metric by choosing the middle points of the classes of V4. Compute all MC and MD.**

- **Try to compute AM within Age-Group variable.**
  Apply the command COMPARE Means from the menue analyze.

EXERCISES DOING CROSSTABS

- **Do crosstabs of variables V2…, V3 and V4 by gender and age_grouped. Interpret the results.**

RED
Non-Visitor Research I

- III. Introduction into Data Analysis
with PSPP -

Author:
Prof. Dr. Tibor Kliment

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

How to install PSPP

- PSPP can be downloaded for free from the Free Software Foundation. The official web site is GNU PSPP.

- Before downloading, it's worthwhile to browse the official site and learn more about PSPP

https://www.gnu.org/software/pspp/get.html

Objectives

- What is PSPP?

- Advantages of PSPP

- How to install PSPP

- Basic structure

- How to import other format files

- Recode data

- Processing data

- How to analysie data

- How to Save Data

Advantages of PSPP

- Free download and no subscription fees

- Compatible with SPSS data files

- Compatible with SPSS scripts

- Multiplatform compatible

- Faster than SPSS

What is PSPP?

- PSPP is a tool for statistical analysis of sampled data.

- It reads the data, analyzes the data according to commands provided, and writes the results to a listing file, to the standard output or to a window of the graphical display.

- The language accepted by PSPP is similar to those accepted by SPSS statistical products.

- PSPP produces tables and charts as output, which it can produce in several formats; currently, ASCII, PostScript, PDF, HTML, DocBook and TeX are supported.

- Similar features to SPSS: Layout, menu commands, scripts

- Data file and script compatibility with SPSS

How to install PSPP

Windows

- You can install pspp on Windows 64 Bit via the pspp-1.6.2-install.exe installer.

  The windows installers for nightly builds are untested nightly builds.
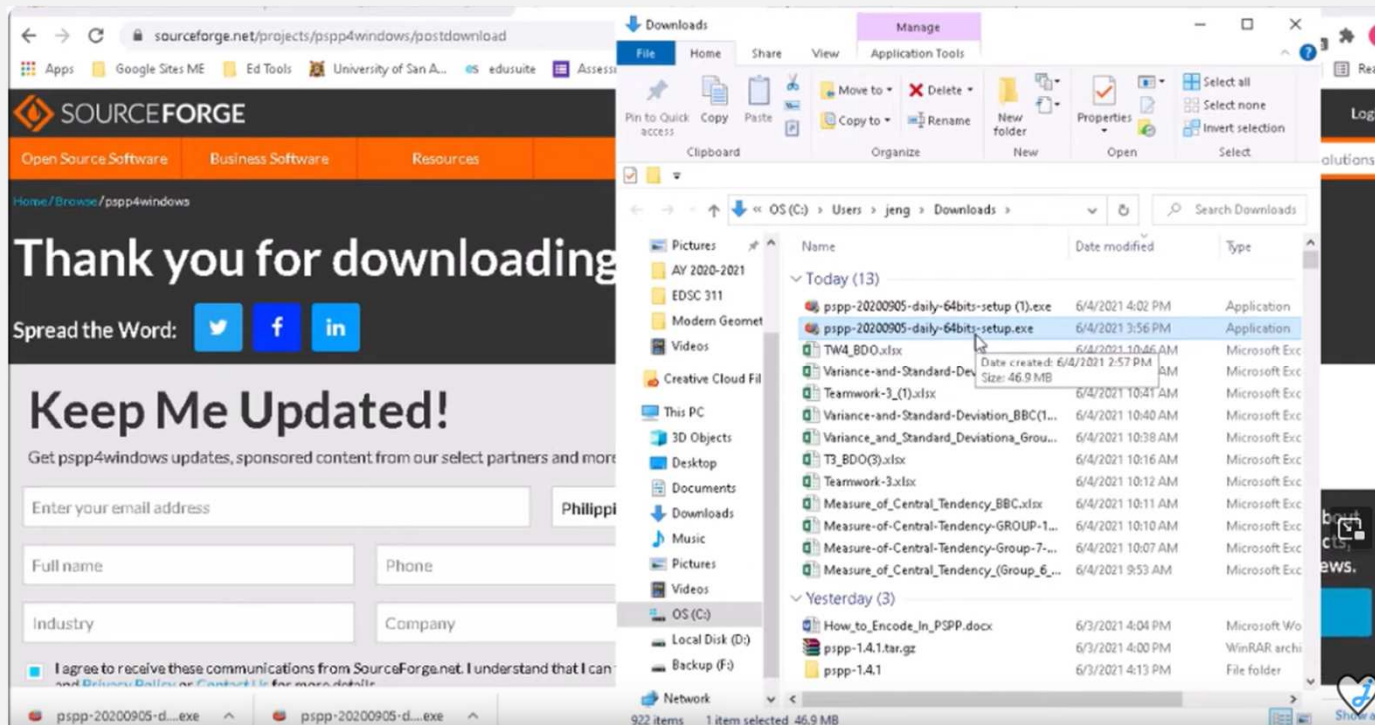
https://www.youtube.com/watch?v=XnbvstmkNas&ab_channel=JengAmor

How to install PSPP

MacOS

- There are three ways to install PSPP on MacOS:

  - The Application bundle provides the guick version via a normal installer. If you do not know what XCode is - go for this one.

  - MacPorts. Run port install pspp-devel to get the latest and most featureful version of PSPP or port install pspp to get an older but possibly better tested version (more information).

  - Homebrew. The PSPP brew tap also includes a description how to install the stable or the development version of PSPP.

How to install PSPP – Application bundle

- Click on "Application bundle" you will be forwarded to another side with the download links.

-

Co-funded by
the European Union

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln
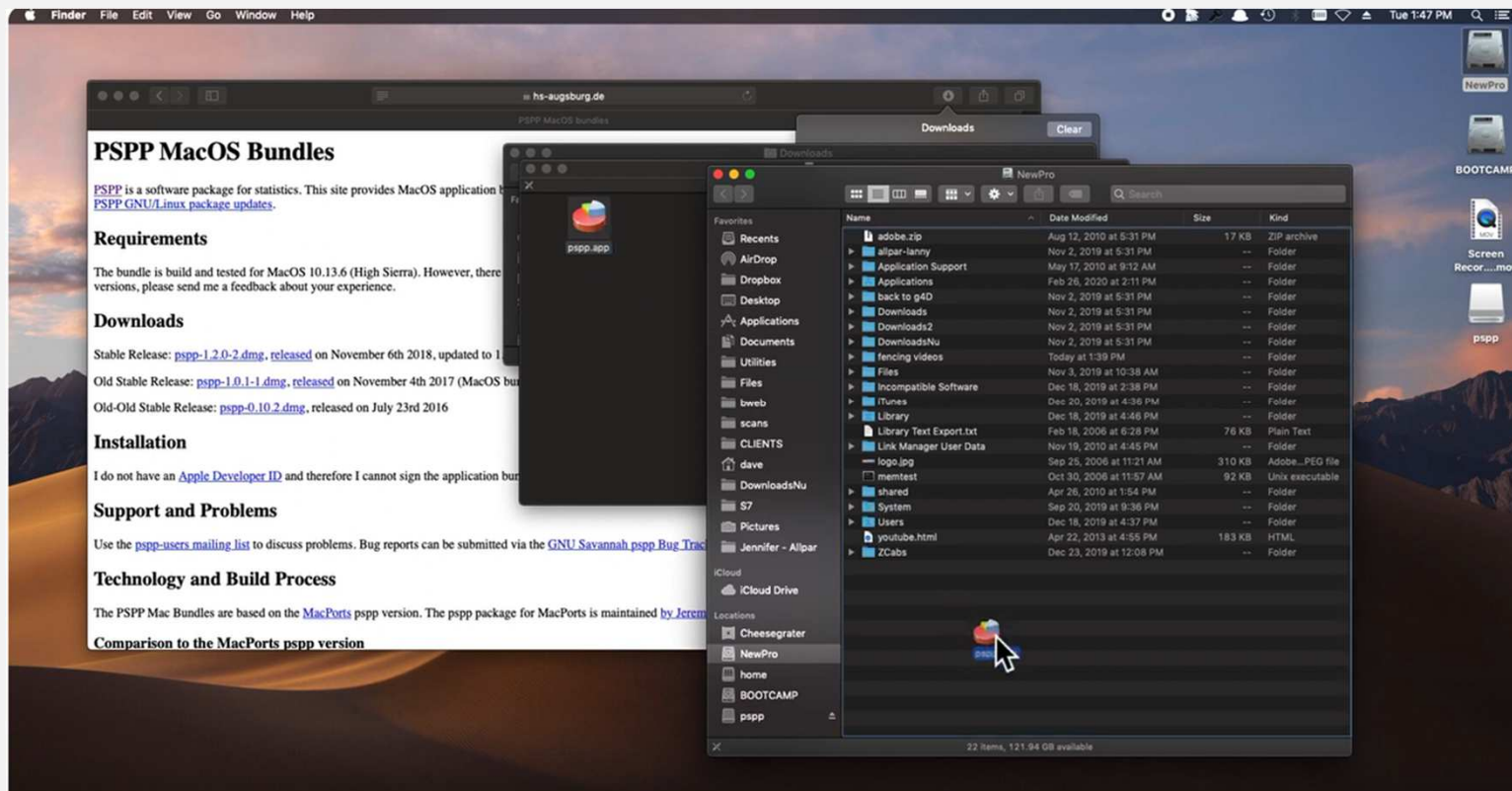
PSPP User interface

- There are two different

  windows:

  - Data Editor
    - Data view
    - Variable View
  - Output Window
  - Syntax editor

Data Editor

- First window you will see when you start the programm

- Provides a convenient, spreadsheet-like method for creating and editing data files.

- You find the Data view and Variable view at the bottom left (Similar to SPSS)
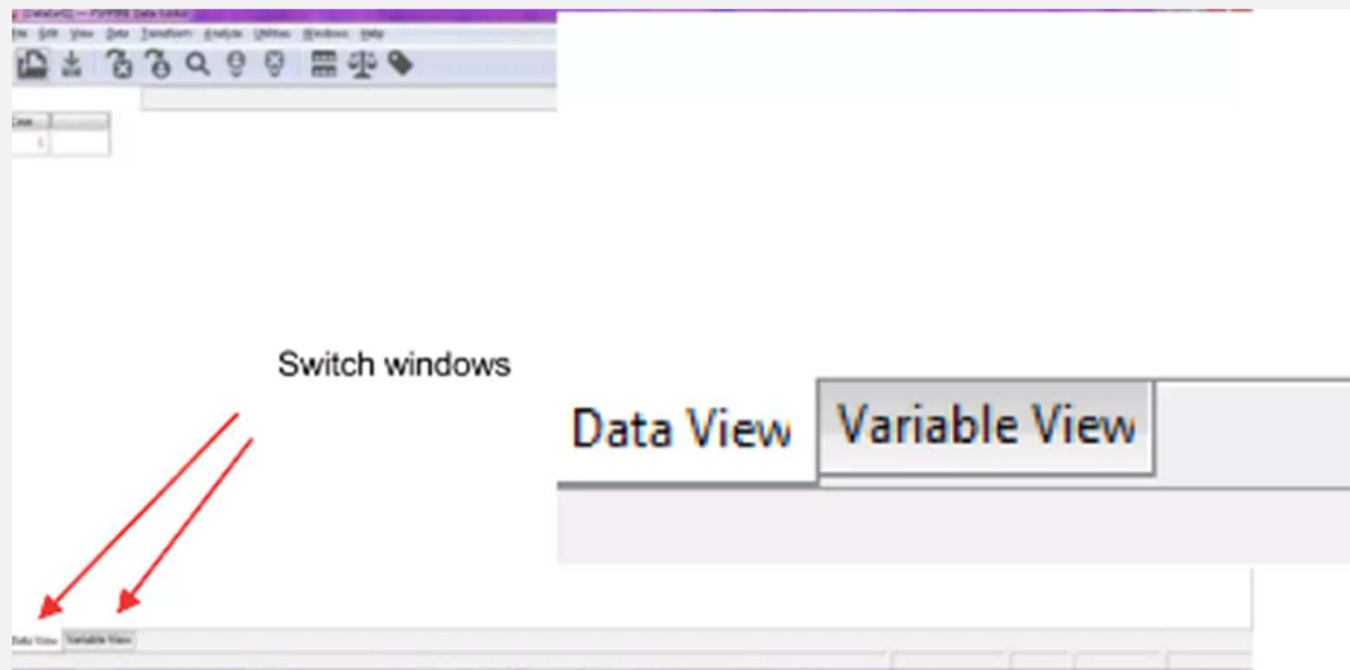
Data View

- Displays the actual data values or defined value labels

- At the top left side you will find the Toolbar

Data View

- Rows are cases: Each row represents a case or an observation. For example, each individual respondent to a questionnaire is a case.

- Columns are variables: Each column represents a variable or characteristic that is being measured. For example, each item on a questionnaire is a variable

- Cells contain values: Each cell contains a single value of a variable for a case. The cell is where the case and the variable intersect. Cells contain only data values.

Variable view

- Displays variable definition information, including:
  - defined variable
  - value labels,
  - data type (string, date or numeric)
  - Measurement level (nominal, ordinal, or scale) and
  - user-defined missing values

Variable view

- Contains descriptions of the attributes of each variable in the data file

- Rows are variables

- Columns are variable attributes

- You can add or delete variable and modify attributes of variables, including the following attributes:

  - Variable name

  - Data type

  - Number of digits of characters

  - Number of decimal places

  - Descriptive variable and value labels.

  - User-defined missing values

  - Column width

  - Measurement level

Variable name

- Each variable name must be unique – duplication is now allowed
- Variable names can be up to 64 bytes long, and the first character must be a letter or one of the characters @, #, or $. Subsequent characters can be any combination of letters and numbers
- Variable names cannot contain spaces. Can keep space using underscores
- Reserved keywords cannot be used as variable names.
- Reserved keywords are: ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO, and WITH

Variable type

- Variable type specifies the data type for each variable. By default, all new variables are assumed to be numeric. You can use Variable Type to change the data type.
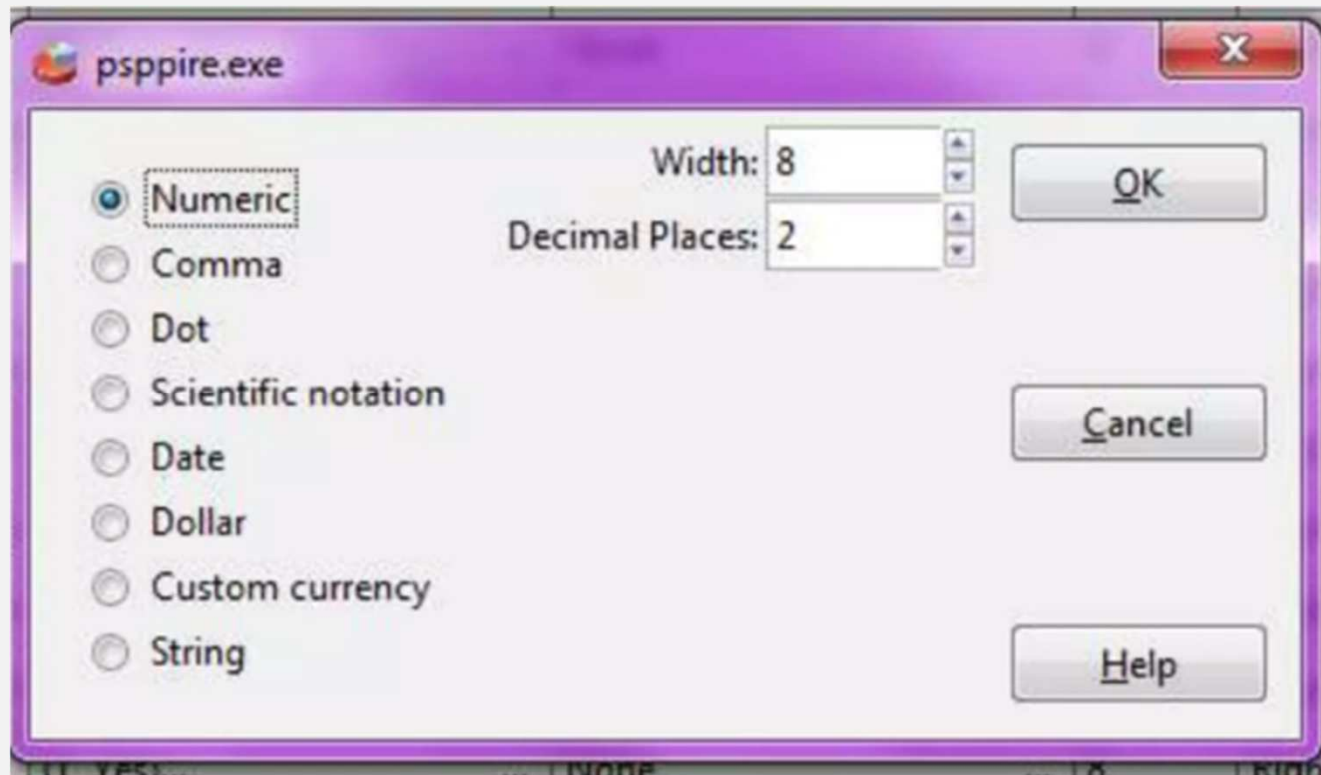
Variable labels and missung values

Variable labels:

- Can assign descriptive variable labels up to 256 characters (128 characters in double-byte langueages).

- Can contain spaces and reserved characters that are not allowed in variable names

Missing values:

- Defines specified data values as user-missing

- Example: You might want to distinguish between data that are missing because a respondent refused to answer and data that are missing because the question didn't apply to that respondent

- Data values that are specified as user-missing are flagged for special treatment and are excluded from most calculations

Variable labels

- You can assign descriptive value labels for each value of a variable.

- Particularly useful if your data file uses numeric codes to represent non-numeric categories

- Example: Codes of 1 and 2 for male and female

- Value labels are saved with the data file. You don't need to redefine value labels each time you open a data file
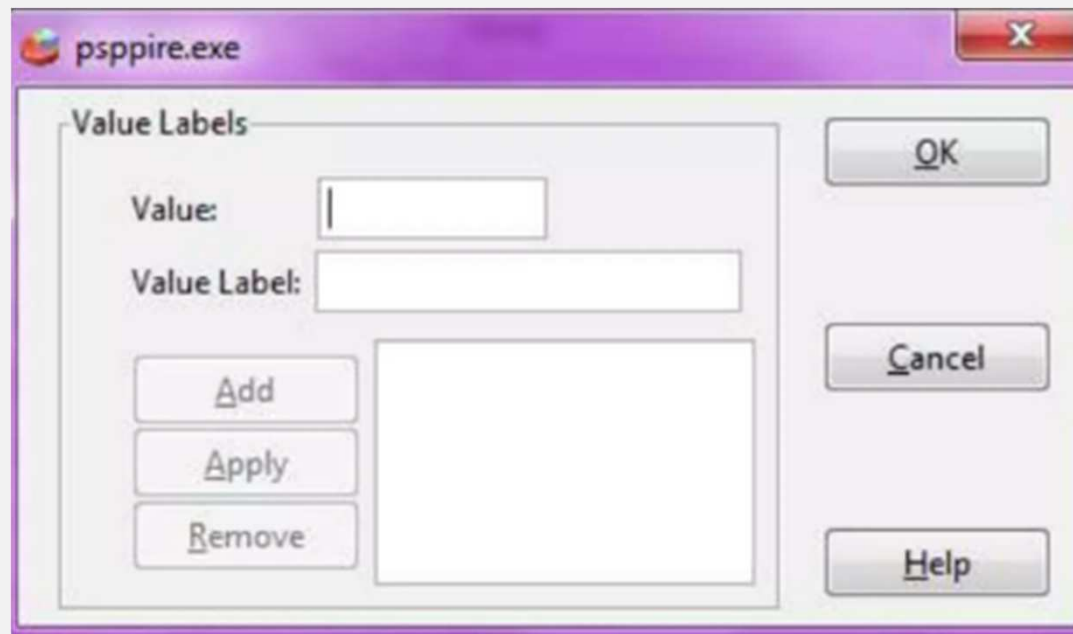
Variable measures

- Same measurements as in SPSS

- Nominal

- Ordinal

- Scale

Importing a Data File

- If a data file is already available, the data does not need to be manually entered.

- It is a good idea to use preexisting data files rather than entering data by hand.

- PSPP can open pre-existing data files with the "File" → "Important data" command.

- A common file format is comma separated values (.csv) in which each field is separated by a comma. Importing one of these files starts a wizard that steps you through the importation process. Part of this importing process might be to define the variable names and data types of each field.
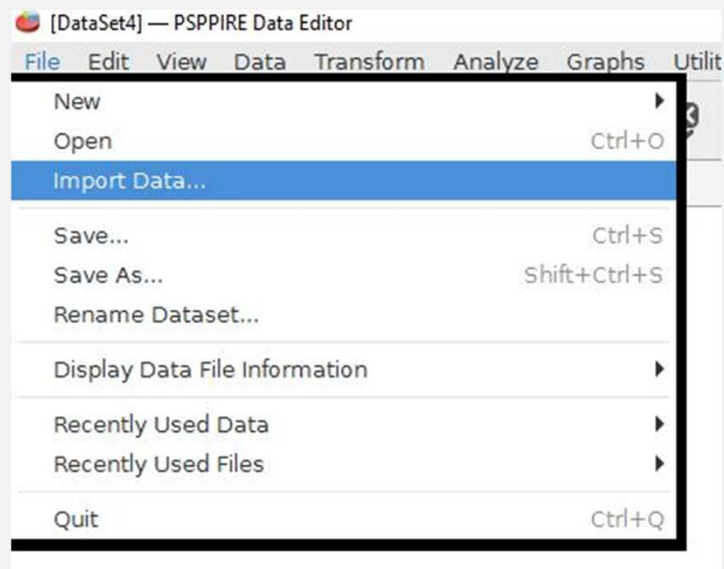
Navigating and Editing a Data File

The computer screen can only display a small amount of a large data file. Navigating through a large amount of data can be awkwardly. Fortunately, the tool bar has helpful features for navigating and editing large data files.
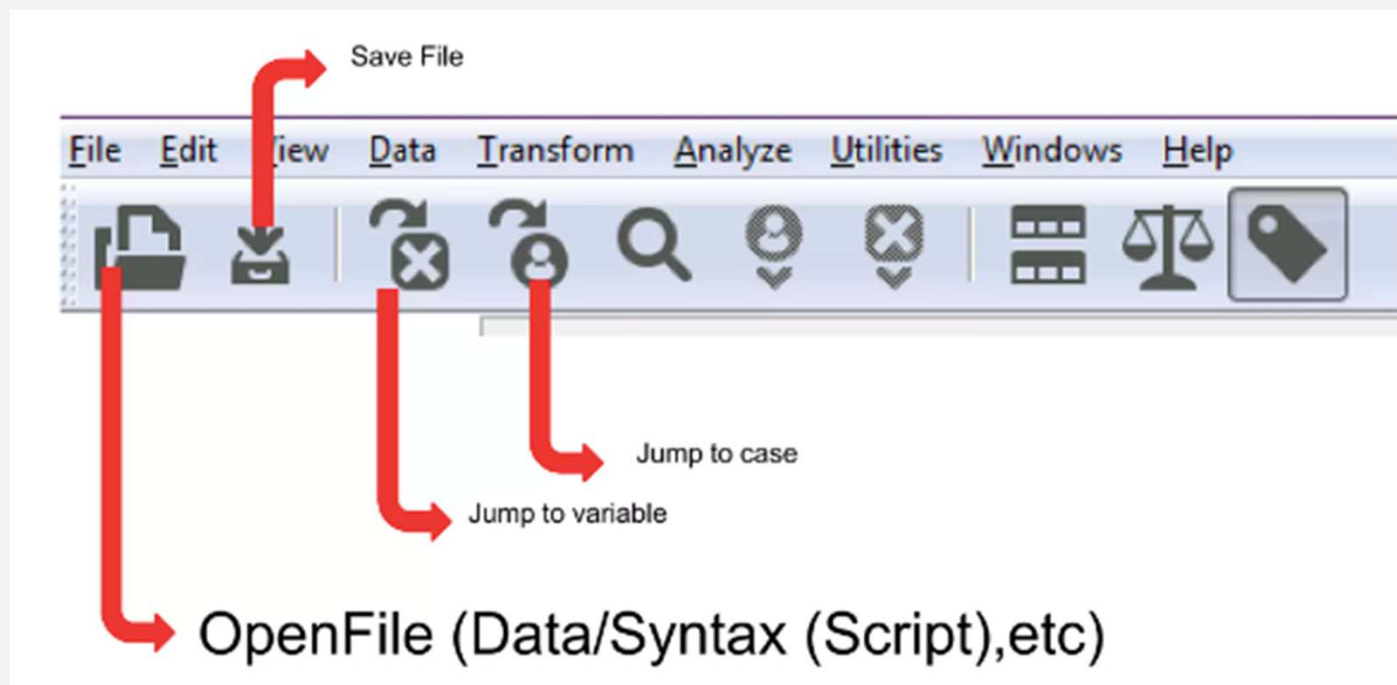
Navigating and Editing a Data File

From left to right:

1. Jump to variable: This will move the display horozontally from left to right to a variable of interest

2. Jump to case: This will move the display downward to a particular case.

3. Search for values

4. Create a new case at the current position: This button inserts a new row for a new case

5. Crete a new variable at the current position: This button inserts a new column for a new variable

6. Split file

7. Weight file

8. Show/hide value labels: This will toggle the values/labels for categorical variables.

Navigating and Editing a Data File

If we want to insert or delete a case or variable:

- right-clicking on the case variable for a specific row will raise a pop-up menu with the option to insert or delete a case at that position.

- A similar feature is avaiable for the columns.

Sorting Data

- Sometimes it is useful to sort the cases into groups.

- Example: You might want to sort the festival participants from the most to the least visits to identify how many people where there each screening.

- PSPP has several ways to accomplish this goal.

- The Data menu has a "Sort Cases" command.

Co-funded by
the European Union

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln
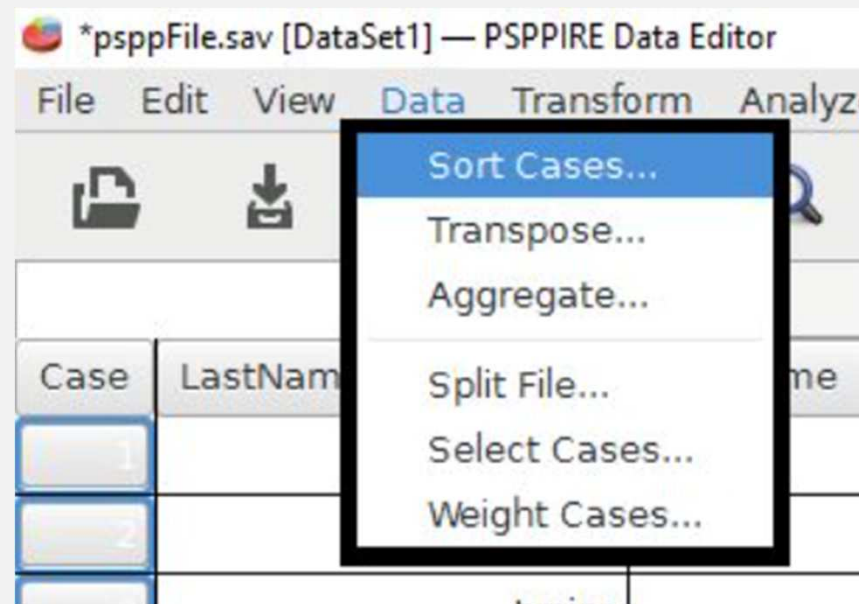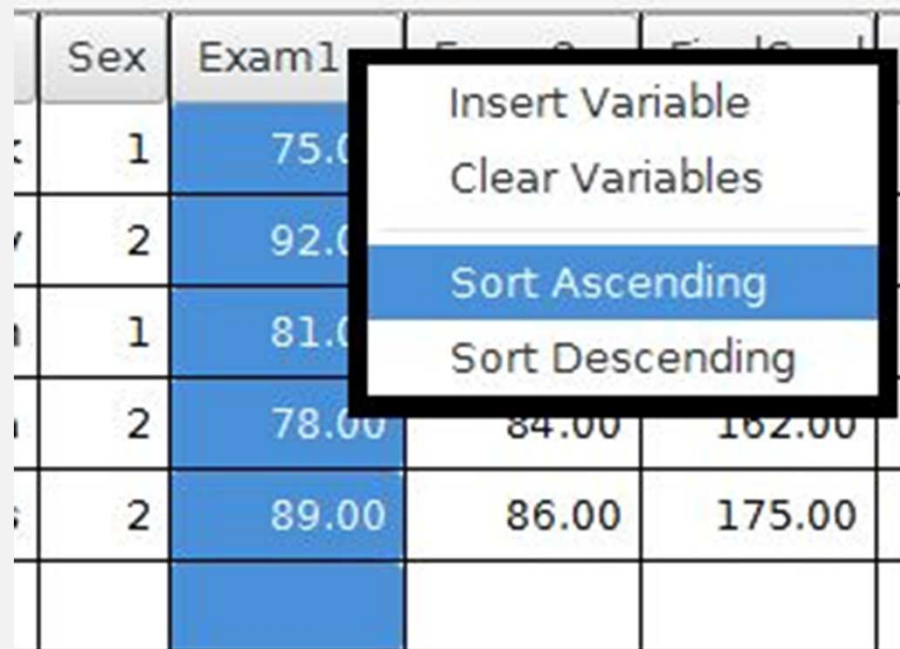
Sorting Data

- The "Sort Cases" dialog box has a list of variables on the left and a blank **"Sort by**" field on the right.

- Choose the variables that you want to use by moving them to the "Sort by" field.

- The order is important, with the first field being the most important.

- The sort order **"Ascending**" option will organize the cases from the lowest to the highest.

- The **"Descending**" option is from the highest to lowest.

- Click the **"okay**" button when you are finished
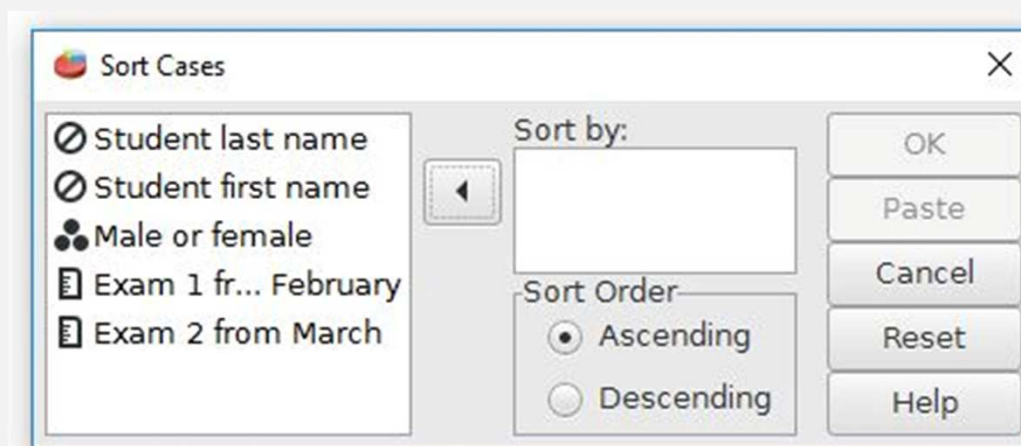
Sort by Column Heading

- A quick way to perform a simple sort is to right click on the column heading of a variable.

- The Pop-up menu has options for Sort Ascending or Descending.

- This sort is performed on a single variable

- Sorting for two or more variables, would require using the menu option from **"Data"** ➔ **"Sort Cases".**

Dialog for Choosing Variables

- PSPP doesn't know which variables you want to use when you want to do an analysis or command. You must often tell PSPP excatly which variables the analysis must be done upon.

- PSPP typically gives you a list of variables on the left in a dialog box.

- You must choose the variables that you want to use by moving them to a field in the center or right of the dialog box that is often called "**Variables**".

- In the example on the next page it is the **"Sort By"** field.

Dialog for Choosing Variables

- The variables can be chosen in two ways:

  - One way is to click on the variable name on the left, then drag it to the empty variable field on the right.

  - Another method is to highlight the variable that you want, then use the little arrow button between the fields to move the variable.
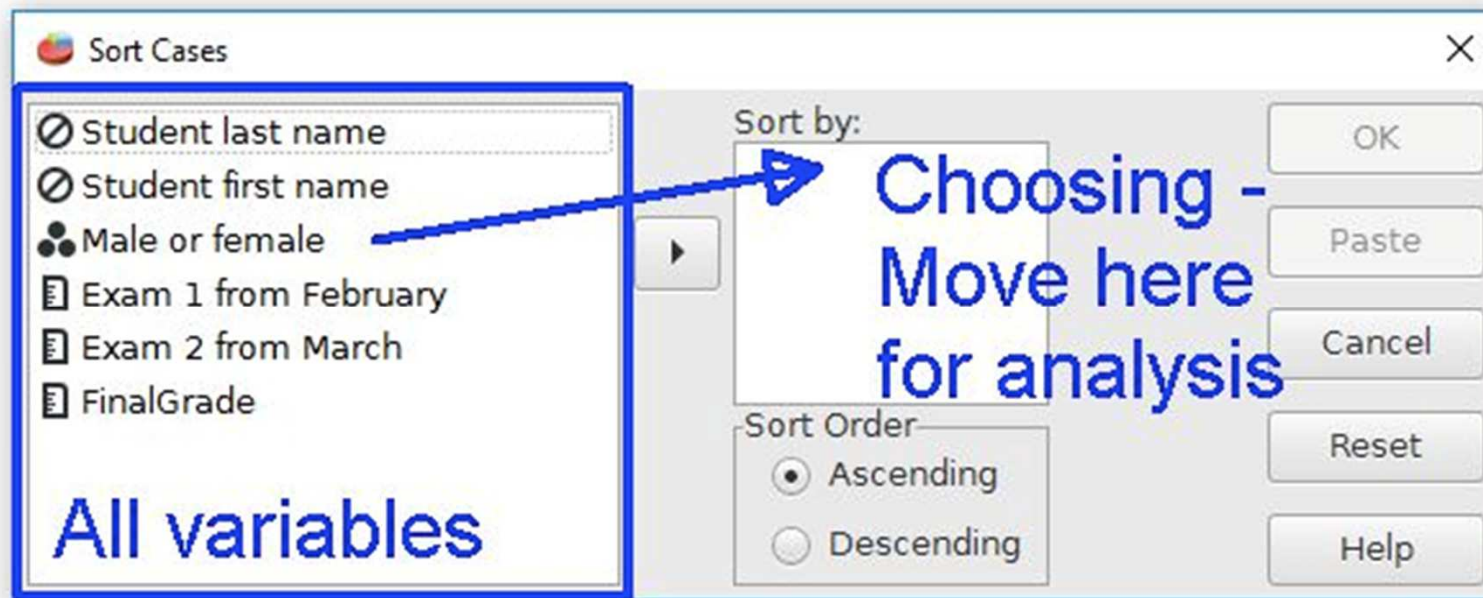
Recode variables

- Two options
  - Recode into Same Variables
  - Recode into Different Variables
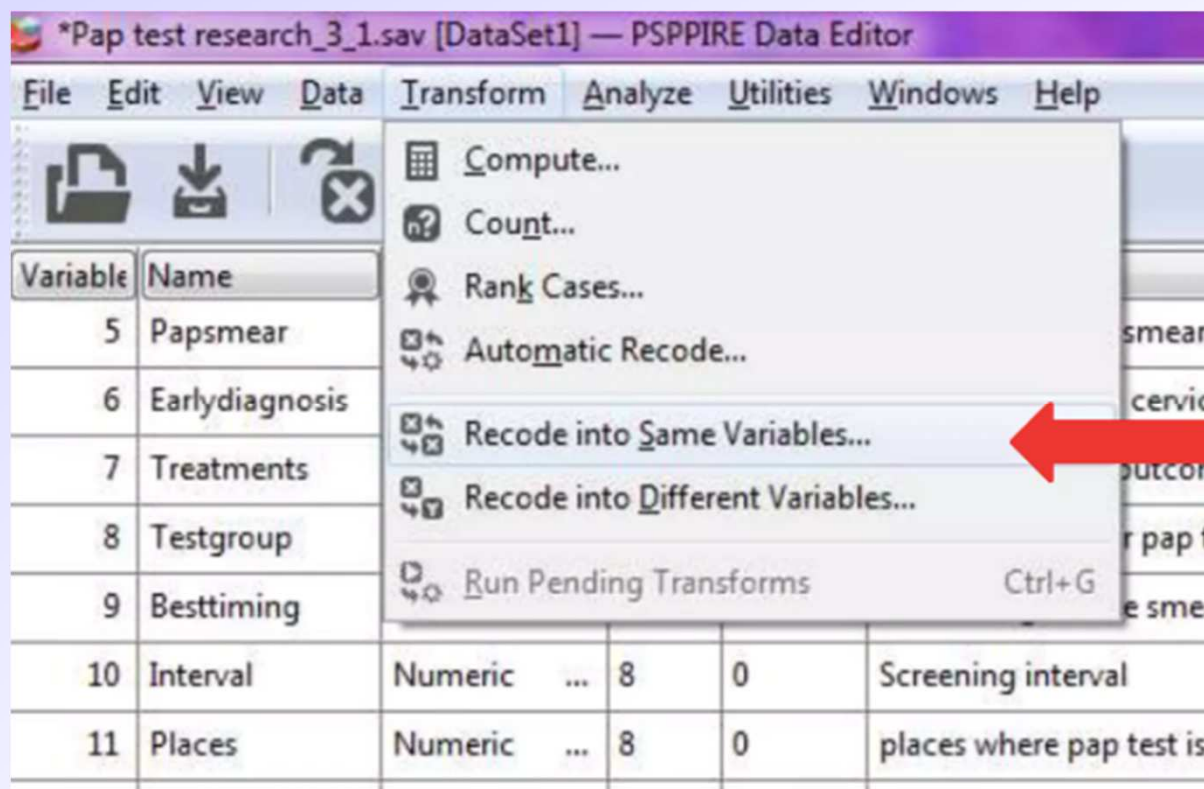
Recode variables – Same Variables

- Recoding for the same variable will change the values of the variable to a new coding scheme

- Example: The original file uses a code of 1 = male and 2 = female. Let's say that we want to change this coding scheme to 0 = male and 1 = female. To make this change select **"Transform" → "Recode into Same Variables"**.

Co-funded by
the European Union

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln
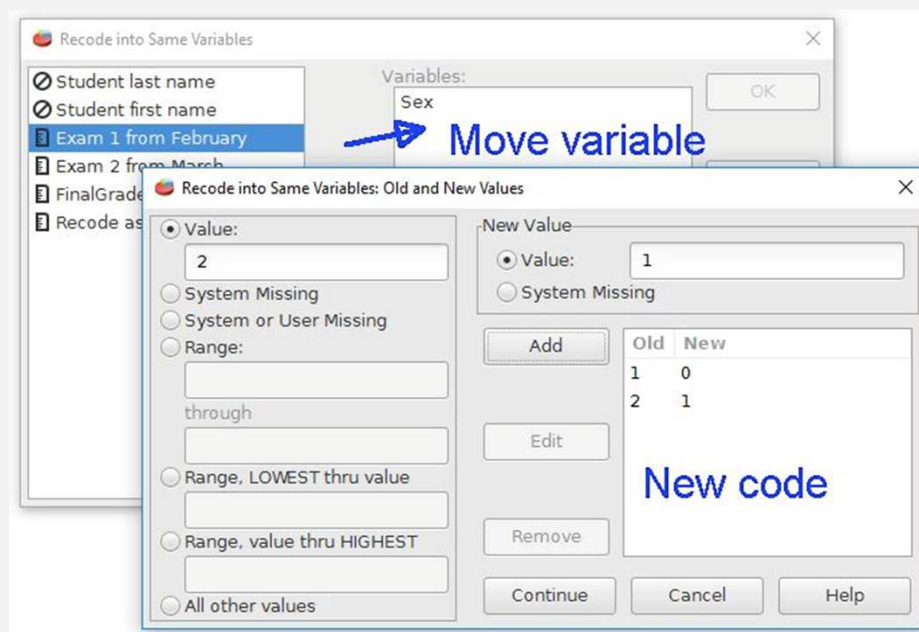
Recode variables – Same Variables

- The first step is to select the variable to be recoded and move it to the Variables field.

- Next, enter the new coding scheme. This works like the Recode into Different Variables command. This simple example will convert all 1 values to 0 and all 2 values to 1. Click on the "Continue" and "Ok" buttons to make this desired coding change.

- Making this change will replace the 1 and 2 values with 0 and 1 values. Be aware that the original data will be lost. Make sure that this new coding scheme is what you really want before performing this command.

Recode into Same Variables

- The **Recode into Same Variables** dialog box allows you to reassign the values of existing variables or collapse ranges of existing values into new values
  - Example: you could collapse salaries into salay range categories.
- You can recode numeric and string variables.
- If you select multiple variables, they must all be the same type.
- You cannot recode numeric and string variables together

Recode into Different Variables

- The Recode into Different Variables dialog box allows you to reassign the values of existing variables or collaspe range of existing values into new values for a new variable

- Example: You could collapse salaries into a new variable containing salary-range categories.

- You can recode numeric and string variables.

- You can recode numeric variable into string variables and vice versa.

- If you select multiple variables, they must all be the same type. You cannot recode numeric and string variables together.

Recode into Different Variables

https://www.youtube.com/watch?v=qwiXgfznU50&ab_channel=GregoryFulkerson

Descriptive Statistics

- The starting point for most statistical analyses is a description of the data.

- The aim is to summarize the data into a big-picture or bird's-eye view of what the typical participant is like.

- The next set of PSPP commands will focus on simple descriptive statistics such as measures of central tendency, variability, and graphs.

- The critical difference between PSPP and spreadsheets is that PSPP does not use formulas in the cells of the data view.

- Data analysis with PSPP involves choosing the analyses from the drop-down Analyze menu. IBM, the maker of SPSS, has claimed that this approach is superior to spreadsheets because it eliminates the possibility of using incorrect formulas.

Descriptive Statistics

- Most of the descriptive statistics that we will be using are located in the Descriptive Statistics folder of the Analyze drop-down menu.

- The basic process will be to choose the analysis, then select the variables and the details of the analysis in a dialog box. The results of the analysis will be displayed in the output window.
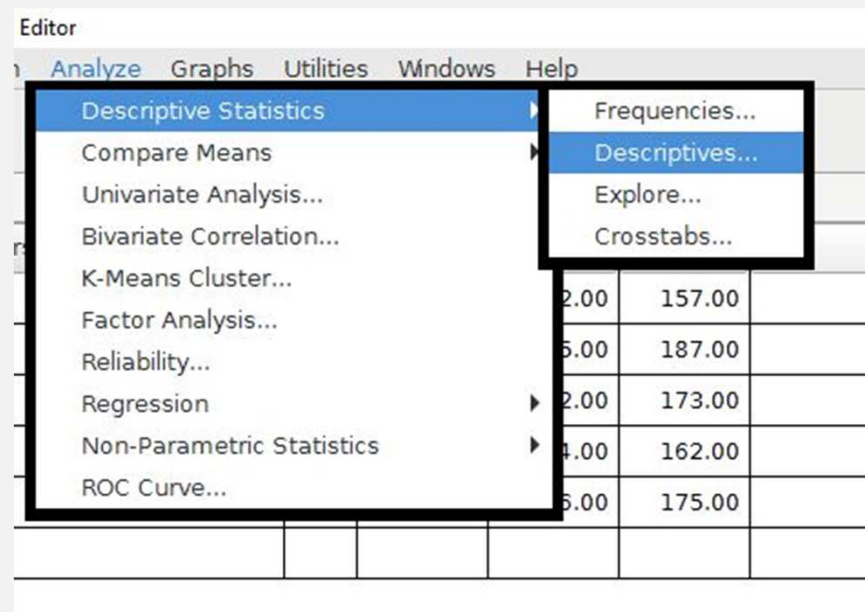
The Frequencies Command

- The Frequencies command can create a wide range of descriptive statistics: Frequency tables, central tendency, and even graphs. This tool is especially useful for determining the frequencies and percentages of particular scores.

- Begin by selecting **"Analyze"** → **"Descriptive Statistics"** → **"Frequencies"**. The dialog box will look like this:
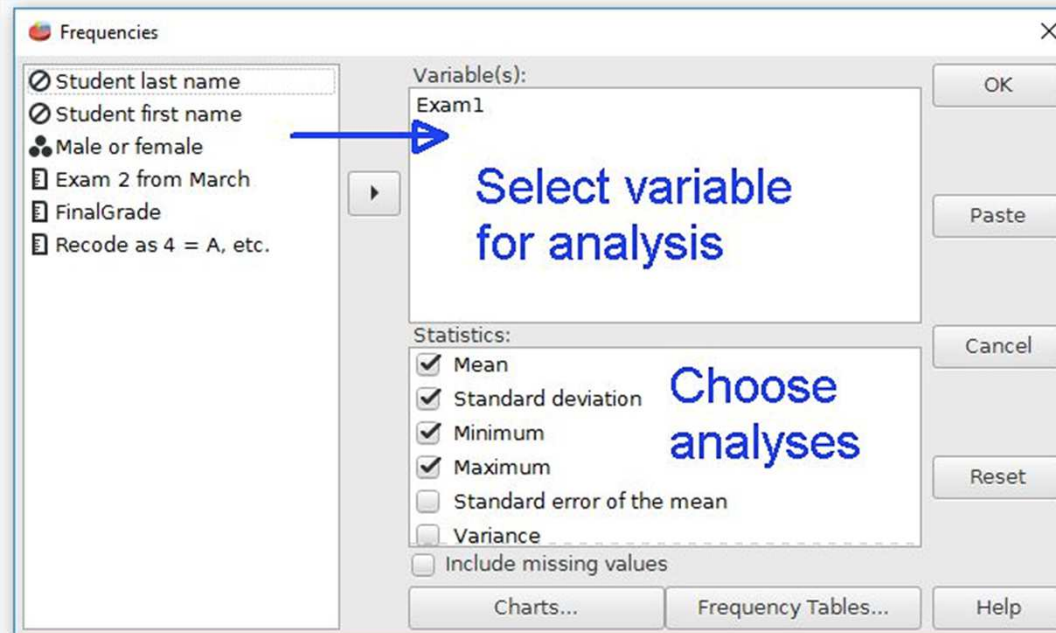
The Frequencies Command

- The first step: choose the variable that will be analyzed. Move this variable from the list on the left to the Variable(s) field. We are using Exam1 for our example.

- The next step is to choose the desired descriptive statistics by checking the ones that you want.

- The Charts button has some options for choosing histograms, bar charts, and pie charts.

- The Frequencies Table button has options for organizing the frequencies list, such as ascending or descending order.

- Clicking the "OK" button will perform the analysis.

The Frequencies Command

- Switch to the output viewer to see the results of the analysis. The upper part of the output has the commands that are used by PSPP. The first table has each score with the frequency, percentage, valid percentage, and cumulative percentage. The frequencies count how many time a particular score occurs. Each score in our example occurs one time.

- Valid percentage is a percentage that ignores cells with missing data. The next table has some basic descriptive statistics. Cumulative percentage adds all the scores at or below a particular value.

- The last part is a pie chart that was produced by selecting the pie chart option.

- Note that the label for the Exam1 variable - "Exam 1 from February" - has been used to label all of these outputs. This shows why entering an informative label for each variable is important when the data files are created.

The Frequencies Command

```
FREQUENCIES
    /VARIABLES= Exam1
    /FORMAT=AVALUE TABLE
    /PIECHART=.
```

Exam 1 from February

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	75.00	1	20.00	20.00	20.00
	78.00	1	20.00	20.00	40.00
	81.00	1	20.00	20.00	60.00
	89.00	1	20.00	20.00	80.00
	92.00	1	20.00	20.00	100.00
	Total	5	100.0	100.0	

Exam 1 from February

N	Valid	5
	Missing	0
Mean		83.00
Std Dev		7.25
Minimum		75.00
Maximum		92.00

Exam 1 from February
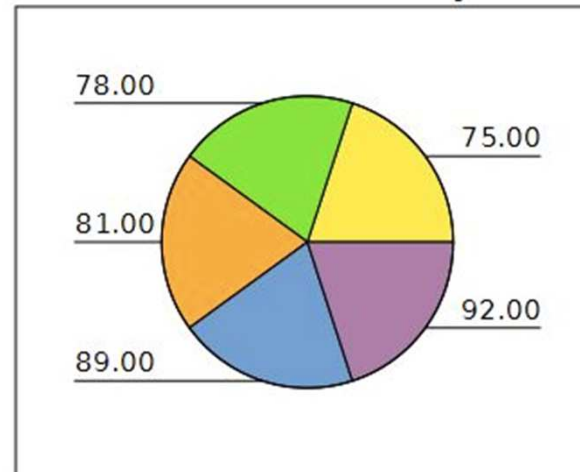
Descriptives Command

- The Descriptives command is a good all-purpose option for basic describes statistics, such as measures of central tendency and variability. It is similar to the Frequencies command but lacks the frequencies table.

- This command is available from **"Analyze"** → **"Descriptive Statistics"** → **"Descriptives"**.

- Choose the variables first by moving them to the Variables column. Next, the desired analyses can be selected by checking the boxes. Some advanced options are available at the bottom.

Descriptives Command

- Choose the variables first by moving them to the Variables column.

- Next, the desired analyses can be selected by checking the boxes. Some advanced
  options are available at the bottom.
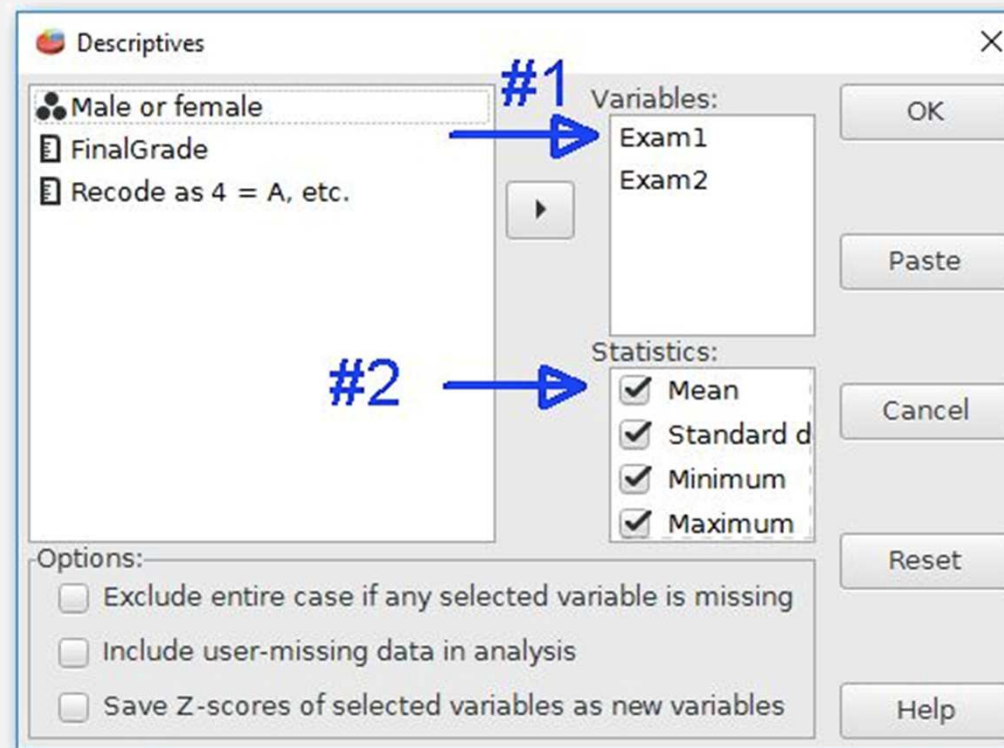
Descriptives Command

- Here is the output for the Exam1 and Exam2 variables with the Descriptives command. This output uses the default analyses:

```
DESCRIPTIVES

DESCRIPTIVES
  /VARIABLES= Exam1 Exam2.
```

Valid cases = 5; cases with missing value(s) = 0.

Variable	N	Mean	Std Dev	Minimum	Maximum
Exam 1 from February	5	83.00	7.25	75.00	92.00
Exam 2 from March	5	87.80	5.50	82.00	95.00

- Descriptives command does not calculate the mode or the median measures of central tendency.
- These statistics are only available from the the Analyze > Descriptive Statistics > Frequencies command. The best guess for this odd omission is that the mean is the most commonly used form of central tendency.

Calculating statistics in this manner with PSPP is so much easier than using spreadsheets. It's a breeze once you understand the basic features of how it works.

Descriptives Command

- Descriptives command does not calculate the mode or the median measures of central tendency.

- These statistics are only available from the the **"Analyze"** → **"Descriptive Statistics"** → **"Frequencies" command**.

- The best guess for this odd omission is that the mean is the most commonly used form of central tendency.

- Calculating statistics in this manner with PSPP is so much easier than using spreadsheets. It's a breeze once you understand the basic features of how it works.

The Crosstabs Command

- In some research situations we might want to compare the frequency of people across two simultaneous categories.

- Example: Compare the frequencies of male and female who like do go to the cinema or who likely to watch movies at home. The two category variables would yield at least four possible groups, such as male Cinema visitors, male stream watchers, female cinema visiitors, female stream watchers.

- The data are nominal scale data, which means that we are counting the number of perople in each category

- This double classification scheme goes beyond the simple listing of frequencies that was done with the Frequencies command.

- A Crosstabs analysis will require two categorical variables that are appropriately coded to represent group membership.

The Crosstabs Command

- The Crosstabs command is available from "**Analyze**" → "**Descriptive statistics**" → "**Crosstabs**".

- For illustration, we will use the Sex and Political Party variables.

Co-funded by
the European Union

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln
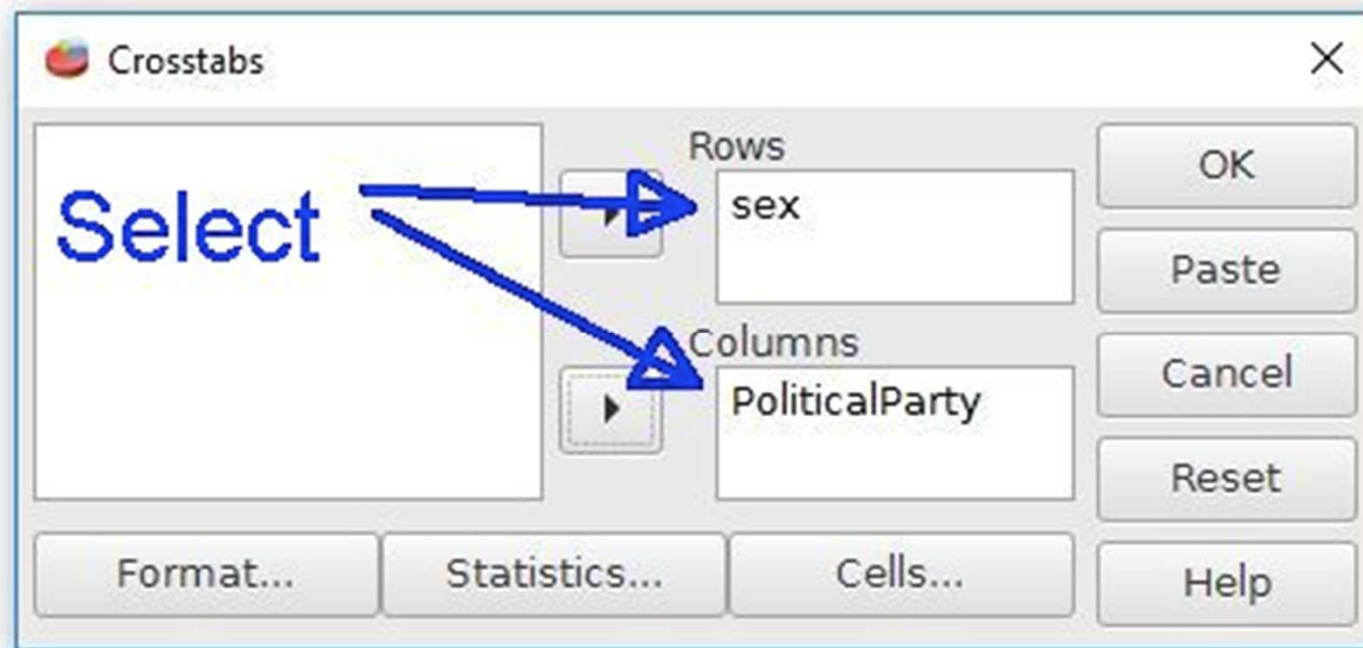
The Crosstabs Command

- The analysis will need at least one categorical variable for the rows and at least one for the columns. The variables of Sex and PoliticalParty in this example will need to be moved right to the Row and Column fields to select them for the analysis.

- The Format and Statistics buttons have some additional options for the analyses.

- The output will look like this:

```
CROSSTABS

CROSSTABS
    /TABLES= sex BY    PoliticalParty
    /FORMAT=AVALUE TABLES PIVOT
    /STATISTICS=CHISQ
    /CELLS=COUNT ROW COLUMN TOTAL.
```

Summary.

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
sex * PoliticalParty	30	100.0%	0	0.0%	30	100.0%

sex * PoliticalParty [count, row %, column %, total %].

sex	PoliticalParty		Total
	Democrat	Republican	
male	5.00	10.00	15.00
	33.33%	66.67%	100.00%
	31.25%	71.43%	50.00%
	16.67%	33.33%	50.00%
female	11.00	4.00	15.00
	73.33%	26.67%	100.00%
	68.75%	28.57%	50.00%
	36.67%	13.33%	50.00%
Total	16.00	14.00	30.00
	53.33%	46.67%	100.00%
	100.00%	100.00%	100.00%
	53.33%	46.67%	100.00%

Chi-square tests.

Statistic	Value	df	Asymp. Sig. (2-tailed)	Exact Sig. (2-tailed)	Exact Sig. (1-tailed)
Pearson Chi-Square	4.82	1	.028		
Likelihood Ratio	4.96	1	.026		
Fisher's Exact Test				.037	.033
Continuity Correction	3.35	1	.067		
Linear-by-Linear Association	4.66	1	.031		
N of Valid Cases	30				

Means for Categorical Variables
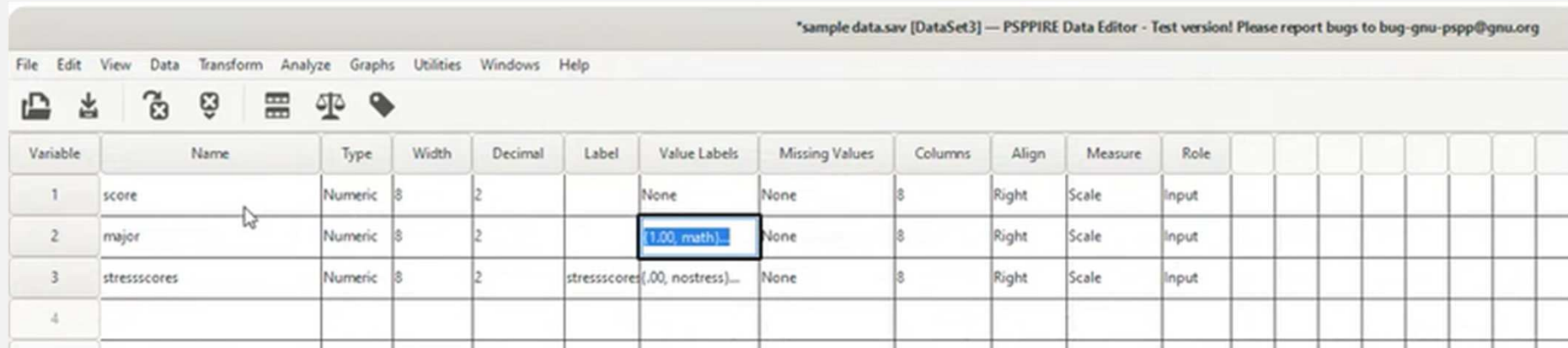
- A common situation in statistics is to have two or more groups that will be compared, such as an experimental group and a control group. Descriptive statistics such as means and standard deviations will need to be determined for each group as part of the comparison.

- One way to accomplish this goal is to use the Means command "Analyze" → **"Compare Means" → "Means"**.

- The Compare Means group has several inferential tests that can generate descriptive statistics in addition to the inferential statistics.

Means for Categorical Variables

https://www.youtube.com/watch?v=XdC8vWiEKb4&ab_channel=DrJ.

Graphs: Histograms

- Graphs are an important way to visually communicate statistics. PSPP can make basic graphs through commands located in the Graphs menu.

- Click on **"Graphs"** → **"Histogram"** to open up the histrogram window.

- For Histograms, the variables must be chosen from the list of variables.

- Most of the graphing options are rather simple and are adequate for making basic graphs. Advanced graphing features might require other software with more advanced graphing capabilities.
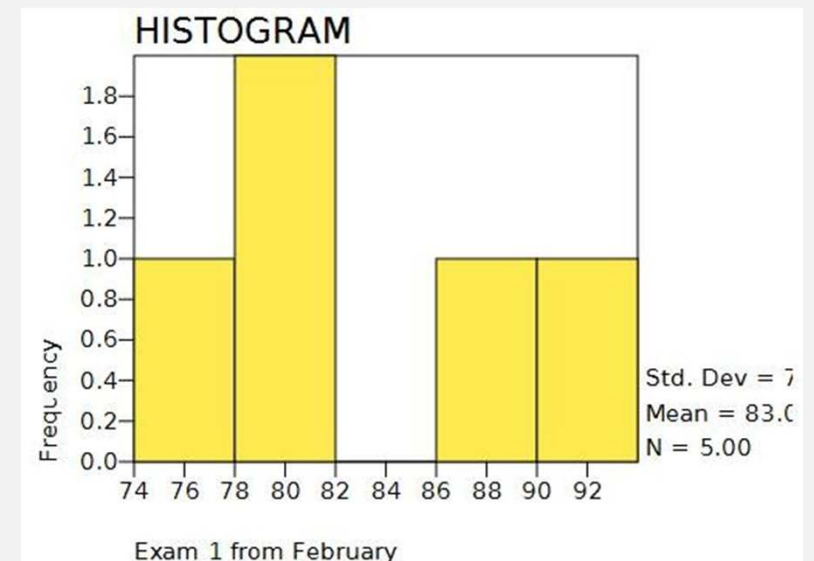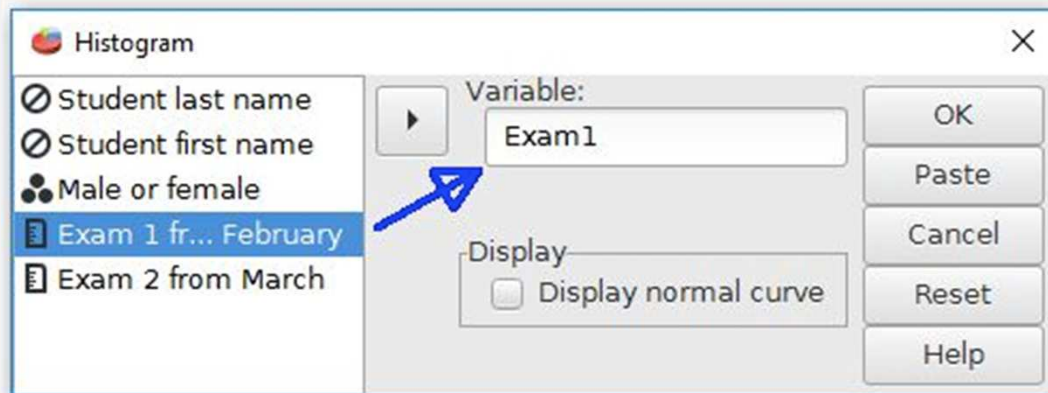
Graphs: Bar Graphs

- The Bar Graph command has more options to choose from than the Histogram graphs.

- A continuously scaled variable must go into the "bars represent" field.

- There are several options to choose from for the bars, with means possibly being the most common analysis.

- The categorical variable that will represent the bars goes into the "category axis" field
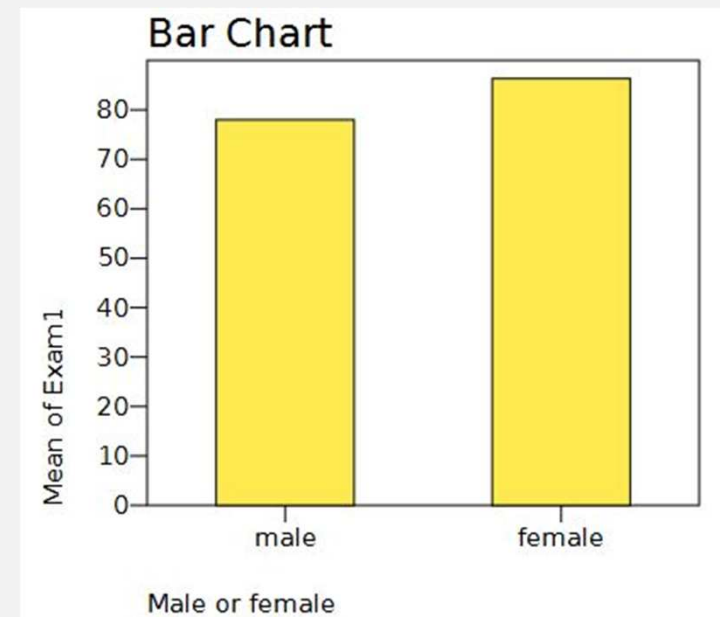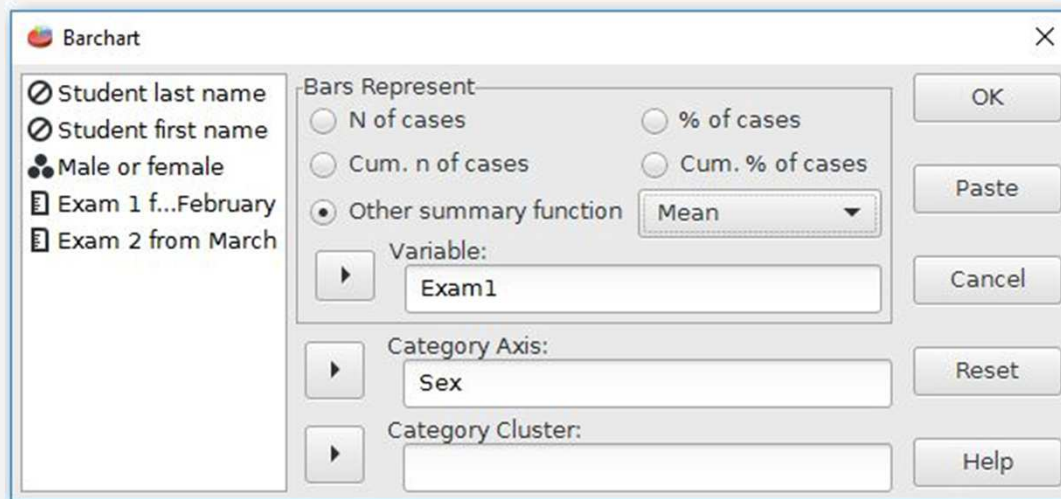
Graphs: Pie Charts

- PSPP can make pie charts, histograms, and bar charts from the Frequencies command (Analyze > Descriptive Statistics > Frequencies).

- Click on the **"charts"** button of the Frequencies dialog box to select these options
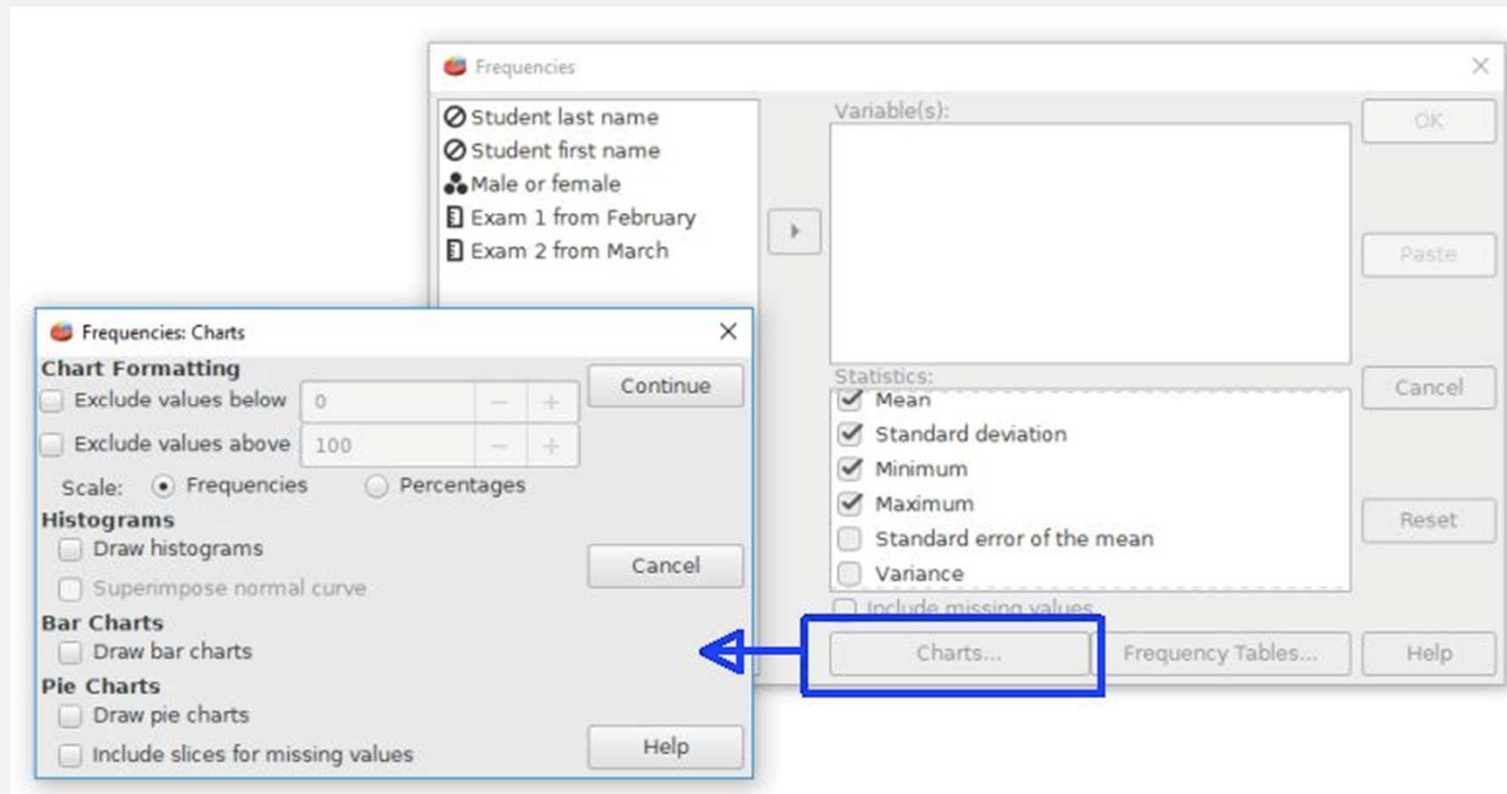
Graphs: Pie Charts

- A basic pie chart from the Frequencies command will look like this:
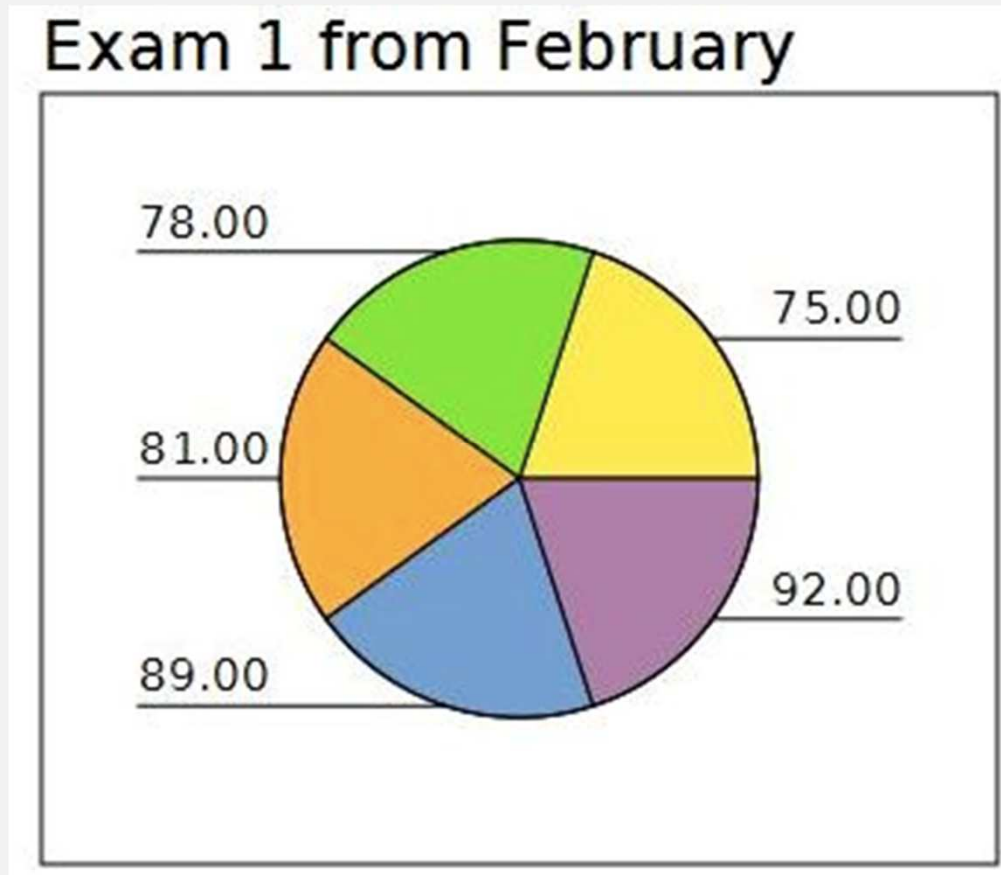
Saving Data

- The file can be saved in several ways:

- Keystroke: Use the Control-S key sequence

- Tool bar button: Click the save button, which looks like this:

- Menu option: Select the "File" menu, then "Save" from the drop-down options

- The save dialog box will show various storage devices and folders on you computer

- Click on „"Save" button (Bottom right corner of the dialog box) to complete the saving.

- Save your work frequently to prevent the possibility of data loss.
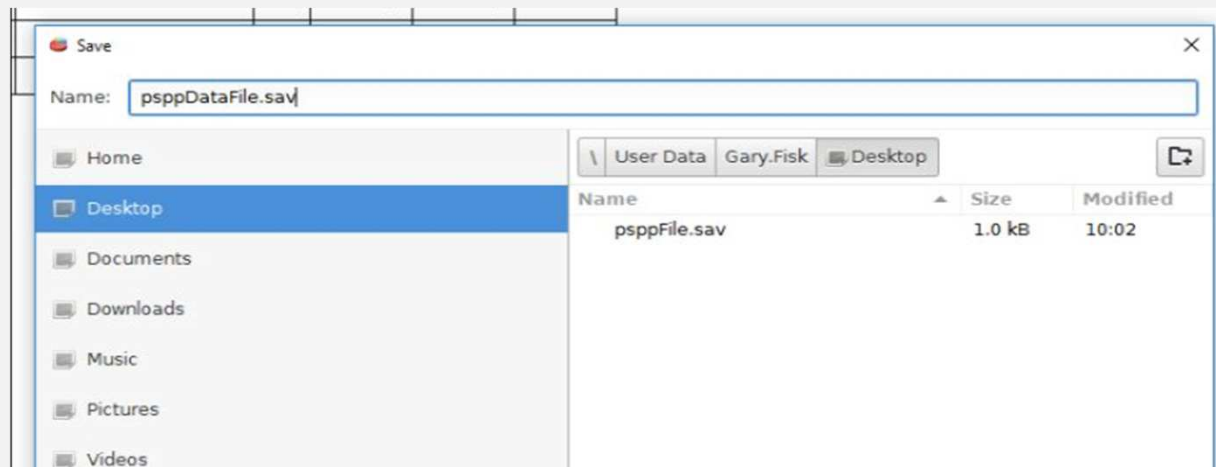
The Output Viewer

- The operation of saving a file opens up a new PSPP window. This is the output viewer, which is the second important part of PSPP

- Output viewer displays the results from statistical analyses.

- Operations are listed on the left

- The larger main part on the right shows the results of a data analysis.

- The Output viewer shows important sometimes. In the picture is shows us where the file was saved.
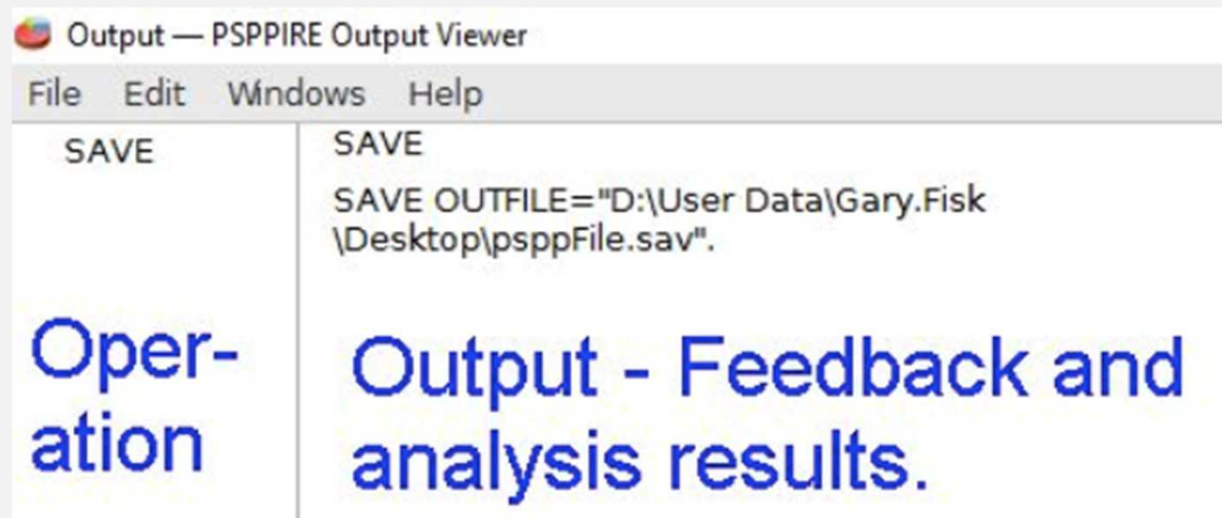
The Output Viewer

- The output viewer is independent of the data view.

- The analysis created by the output must be printed or saved sparately from the data in the data view.

- The output viewer file menu has options for Print and Export. The Export feature is like saving. This will be particularly useful if you want to use the analysis results in another softwareprogram, such as including it in a report

The Output Viewer

- Just type the file name followed by the file extension that you want.

- The most useful options will be text (*.txt, no formatting) or OpenDocument (*.odt, formating).

- OpenDocument is a file format that most word processors, such as Microsoft Word, can open.

- There is no Word option in PSPP because Microsoft Word is a proprietary software program with restrictive licensing.

Infer file type from extension

PDF (*.pdf)

HTML (*.html)

OpenDocument (*.odt)

Text (*.txt)

Text [plain] (*.txt)

PostScript (*.ps)

Comma-Separated Values (*.csv)

RED
Non-Visitor Research I

- Data Analysis of SYS-Survey I+II –

Author:
Prof. Dr. Tibor Kliment

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

Some general recommendations about data evaluation and presentation of results

It is important to formulate the right questions prior to the analysis

Only those who ask questions will get answers from the data later...
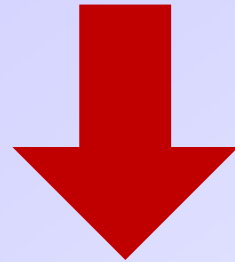
Definition of interest...

- How can we imagine our visitors? Who comes, why, who doesn't? What are possible/new target groups for us?

- How can we make our communication more effective and effecient?

- Are our distribution channels the right ones? How is our content offer received?

- What are issues that require permanent monitoring?

- Do we lack services?
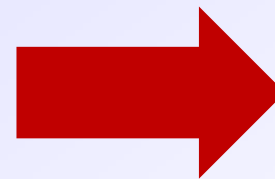
Defining the interest in knowledge in the later data analysis...

To be discussed today...
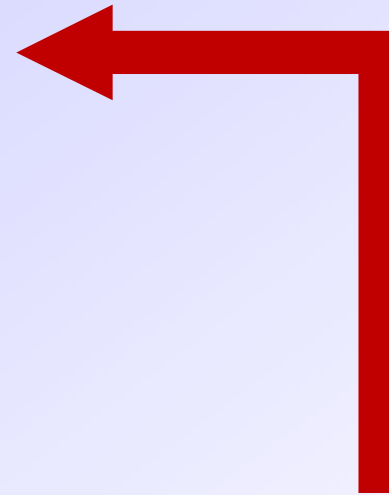
How do I evaluate my data?

It is advantageous to formulate <u>concrete </u>questions or hypotheses for the values <u>before looking at the data for the first time</u>!

– **Set precise expectations for the data, if possible also for the subgroups:**

  - *Why is the value the way it is?*
  - *Why is it not higher or lower?*
  - *Why is it unchanged, even though we...*
  - *Why has the value (not) changed over time?*
  - *Why is the value different for a comparable institution?*
  - *What does that mean for us? Etc. ...*

– **Think about what would be "good", "acceptable" or "unsatisfactory" for you/the institution**

– **Question the data found: Try to find explanations, especially if data turn out differently than expected**

– **Note contextual conditions that may have affected the results during the interview phase**

How do I arrive at my analysis questions, hypotheses or expectations?

For example, through:

- The institution has a certain **self-image** as mission, vision, mission, brand, etc.

- External **target agreements** with policy-makers, sponsors etc.

- **Service expansions or improvement measures** were undertaken in certain areas and are to be examined in terms of effects

- **Evaluate regular** innovations (special exhibitions, new programme, etc.)

- Effects of **externally induced changes** (e.g. sponsors, local government)

- **External shocks** with massive effects in all areas ("corona")

- Visitor research as a **constant "voice of the audience"**, commentator, analyst, forecaster or early warning!

For example:

- *We currently have a great selection of movies. Have we had more first-time visitors as a result? What about our regular visitors? Our visitors have become older? What is the reason for that?*

- *The new season has begun: Why is visitor satisfaction with the current programme worse than last year? How is the relaunched website received by users?*

- *We have strengthened the service staff, improved orientation in the venue and upgraded gastronomy. Is this reflected in visitor satisfaction?*

- *We spend a lot of money on advertising and PR without knowing exactly whether it is going into the right channels. Should we better go to other media? And which ones?*

- *We want a younger audience: Why is our audience older than comparable festivals?*

- *The ticket prices had to be increased recently. Is a different audience coming into the house now?*

- *...*

Implementation of the results

The last step, the implementation of the results, should already be considered when formulating and limiting the objectives of the study.

Goals must be formulated "SMART
(cf. Klein 2011: Kultur-Marketing, Munich):

- ✓ Specific: Is the target content clearly defined?
- ✓ Measurable: How can you tell if the goal has really been achieved?
- ✓ Achievable: Is the achievement of the goal actually feasible? Is it within the sphere of influence of the persons responsible? Is the goal challenging but not overwhelming?
- ✓ Realistic: Is it feasible?
- ✓ Time-scaled/terminated

Example:
- **Objective:** Target group younger people between 18 and 29 years
- **Time reference:** One calendar year
- **Example target:** The range of offers and services for younger people increases. The share of younger people within the next calendar year is therefore to increase by 25%
- **Area:** Young people from city of the festival

Analysing the SYS - Data

DEVELOPING ANALYSIS STRATEGIES FOR SYS SURVEY

Steps for analysis:

1. **Develop questions for analysing the data. Concentrate on one research area out of four:**
   a) Audience structure (demografics, needs, wants)
   b) Audience behaviour on the festival
   c) Motivation and satisfaction in the visit
   d) Media- /Communication and ticketing

2. **Analyse the relevant data to answer your questions: Do frequencies, build crosstabs, use means, divide results into subgroups …**

3. **Draw practical conclusions from that findings; what are the learnings / measures to take?**

4. **Write down your most important findings, conclusions and prepare a little ppt-presentation (about 10-15 min.) with bullet points, tables or grafics**

a) Audience structure (demografics, needs, wants)

1. *What are the important demografics (age, gender, education, region) of the audience?*

2. **What are important demografic differences between <u>first time and recurring visitors</u>? Is the ration between the two subgroups satisfying? What do you conclude from that?**

3. **Who will come back again next time, who will not?**

4. **What are possible causes and hurdles that prevent visitors from coming back next time? What measures are to take?**

5. **What do you conclude from the audience composition (who comes with whom)?**

6. **Try to sketch out the major target groups of the festival. What criteria do you choose? How would you address these target groups? (improving certain offers, more targeted communication, …)**

b) Visiting behaviour on the festival

1. **Which music genres are most important, which less? Would you make changes with regard to the music and selection of movies?**

2. **How satisfied are the visitors? Are there differences between different social groups? What does this mean to you?**

3. **Which general motifs (Var. 10) are the most important for the visitors to come? What do you conclude from that?**

4. **Is there a connection between number of visits and visitor satisfaction?**

5. **What is the recommendation rate NPI among different groups? What measures do you derive from that?**

d) Media- /Communication/ticketing

1. **What is the reach of the campaign? Which visitors were well reached, which not?**

2. **Which media functioned well in the campaign? Which media are cost-effective, which not? Which media would you skip next time?**

3. **When would you start the campaign and ticket sales?**

4. **What do you do with the catalogue in future?**

5. **What is the image of the festival? Is it different between Torino and outwards, heavy and light visitors?**

6. **What are leisure activities of the audience? What can you conclude from that for improving communication?**

7.   **Check, whether ticket price is an issue! For whom? What can you do?**

Strategy of analysis of communication channels and „Sources of Interest"

Is a central question for the evaluation of the festival's communication activities: Fist clarify your communication aims!
Then:

1. Which channels are most effective and cost-efficient?

2. Which target groups use which channels?

3. By using statistical analysis you can find out: Which channels have the biggest impact on campaign awareness, sympathy, visiting the festival

4. Which channels in future should be sorted out?

But be aware of the following problems:

1. **The answers result from a combination of the respondent's media use and the intensity of the festival's communication efforts:**
   - Is a perceived communication activity due to the intensive media use of the respondent, or
   - to the activity of the institution? Or both?

   Conversely, is a non-perceived communication

   – **due to the lack of respondent´s use of the channel referred, or**
   – **is it due to too little activity by the festival? Or both?**

     -> These questions can be tested by questioning also the general media usage and the money spend on the channels

2. **Perception of the advertising messages also depend strongly on the interest of the respondent (self-enhancing circle)**

On the problem of assessing the success of communication measures:

	Communication activity of the institution on the channel given	**No communication activity of the institution on the channel**
Use of the channel by the respondent	• OK • But: Problem of the attribution of the success of the measure Note underreporting due to unaided query form	• If activities recognised by the respondent due to confusion or memory distortion possible
No use of the channel by the respondent	• No effect due to wrong choice of medium or too little activity of the institution	Irrelevant

In general, use media for your campaign that are heavily used bei the target group and which do not cause scattering losses!

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

Assessing reach and cost-effectiveness of the media used in a campaign

Definition: „*Net Reach*" (NR) - Number of visitors reached at least once (or more often) by the campaign or by specific media

- **NR of the campaign or specific media within the *population of the city*? -> Unclear…**

- **NR of the *media campaign among visitors*:**
  Total (n) or in % (Reach all media / all visitors)

- **NR of *specific media* in comparison:**
  Total (n) or in % (Reach one medium / all visitors)

-

- **Assessing cost-effectiveness of all media used in a campaign:**
  „*Thousand User Price*": (Cost for all ads / visitors reached) x 1.000
  (you may also find the *Thousand Contact Price*, which is not applicable here)

- **Assessing cost-effectiveness of specific media used in a campaign:**
  „Thousand user price": (Cost for ads media X / visitors reached by media X) x 1.000

➢ **For making comparisons do this for all media and skip those with bad cost-effectiveness**
➢ **But consider also qualitative factors: Image, impact, usage, purpose of usage, possibilities to convey messages etc.**

Process of dealing with open ended questions

Get an overview about the amount and variety of answers
Decide, how detailed you want to do the analysis:

1. **Type I: Simple analysis**

2. **Copy and paste the answers** into a WORD document

3. **Build chapters** by different topics (gerenal remarks, pricing, service, movies ....)

4. **Sort the answers**

5. **Go through the answers and simply read them**

or

1. **Type II: Build a coding scheme**

2. Extend the existing data file, build **more variables**

3. **Assign values** to all answers

4. **Enter the values** into the data file

5. **Do frequencies and check the results**

6. **Do a quantitative analyses with the newly created variables like any other variable in the data set**

Some simple clues for creating tables

- Do not be contended with basic counts, but always work with **means of centrality, combined variables resp. tables** and correlation measures

- Always select **column percentages in** the tables and **stick to them in** the whole presentation

- **Take independent variables** for the **header columns**, put **dependent variable in the rows**

- Select relevant **target groups/characteristics for subgroup comparisons**, and tabulate all questions in the questionnaire over it

- Always note a **total column** as the reference of the subgroup comparisons

- Indicate in all tabels or grafics what **respondent base** the answers refer to (regard filter questions!)

- Do not further analyse subgroups below **80 cases,** and simply ensure a reader-friendly presentation

Structure of tabular presentations

Options for the selection of tabulation groups/features

Demographic characteristics:

- Age, gender, education, occupation....
- Place of residence

Other features
- Periods, time
- Other institutions etc.)

Behavioural descriptive characteristics:

- Visit frequency, duration,
- First/repeat visit, etc.
- Use of mediation services offers etc.

Attitudes, needs and motives:

- Visit motives
- attitudes
- satisfaction recommendation willingness to recommend
- Lifestyle typology

Example table volume -> Tables_Independent-Theaterscene-Cologne

Some simple tips for creating presentations

- Presentations force to **focus on content** (topic, area, specific questions)
- Graphical representations can **illustrate** complex issues in a **simple way**
- The visualisation of facts or contexts causes them to be remembered **better and longer**
- An argument supported by graphs or charts is more suggestive than tables or texts, and often difficult to refute in argumentation
- They offer **possibilities for manipulation** (choice of axis divisions, reference base, pictograms, etc.)
- Observe the same presentation principles as above for the tables
- Ensure that the design is **graphically "attractive"** (preferably not from SPSS but by PPT or similar).

Example -> Presentation-Non-Visitors_Free-Theaters-Cologne

END OF SECTION !

Building Indicators

- *Indicators:* Observable measures that support a question ("Make it measurable„)

- In order to make the concepts to be analysed measurable, they must be translated into suitable parameters (indicators) (= operationalisation)

- *Operationalisation:* Instructions with the help of which certain *indicators* are assigned to a question -> develop an appropriate definition (nominal, realdef.)

- It is important to ensure that the indicators are simple and clear, plausible and based on what is necessary to measure

- You need theories, assumptions or good arguments to establish the connection between the question and the indicators.

- *Concept of multiple indicators:* Each construct should be operationalised by several indicators!

- Example: Analysis of „visitor motives", "advertising impact", „success of a library"

Operate - depending on your resources - a reporting system appropriate to the addressees. E.G:

- ✓ Data documentation in the form of table volumes that can be called up for reference by anyone from anywhere

- ✓ Topic-specific individual reports, overarching annual reports (incl. summaries)

- ✓ Regular presentations with selected results for specific occasions, topics areas in the institution or topics, (temporal density depending on sample sizes)

- ✓ **Try new ways of doing things:**
  -> E.g. no data diggers in reports but visualise research results in videos

RED
Non-Visitor Research II

- Introduction into Non-Visitor Studies -

Author
Prof. Dr. Tibor Kliment

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

Non-visitors as a problem and potential for cultural institutions

Different perspectives: economic, social and facility-related

Different views on cultural participation - economic dimension

- **Economic interest of the cultural institution or the cultural administration**

- **Typical target groups are exhausted**

- **"Get the venue full"**

- **Optimise capacity utilisation**

- **Sell more tickets / more turnover**

- **Get away from an elite image**

- **Maintain public funding**

Looking at Cultural Participation - Political Dimension

- **Political interest of the cultural institution or cultural administration**

- **Make "Culture for all" accessible**

- **Audience as a reflection of society**

- **"No one must visit cultural institutions, but no one is locked out!**

- **Reducing social inequality**

- **Include minority groups into society**

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

Visitor research as a thematic field in cultural institutions

- **Audience research is indispensable today**

- **Information on the current status and changes of current and potential visitors is essential for the further development, professionalisation and success of one's own work.**

- **Many institutions today now have good knowledge about their visitors**

- **There is little or no knowledge about groups that visit facilities only very rarely or not at all**

- **Explicit non-visitor studies are rare, as hardly any are conducted by the institutions**

- **Question: Why is that? Is non-visitor research "more difficult"?**

- **Is it more difficult to implement the findings?**

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

Visitor research as a thematic field in cultural institutions

- **Question: Why is that? Is non-visitor research "more difficult"? Is it more difficult to implement the findings?**

- **....?**

Differences between visitor and non-visitor research:

Targets
- Visitor research focuses on visitors/users/audiences who **are in the institution**
- Non-visitor research focus people who are **not** or not now or not any longer **in the institution**

Sampling
- Visitor-studies are conducted *within the* institution
- Non-visitor studies are *outside* the institution

Subjects
- Visitor studies focus on the visitors´ experience at the visit
- Non-visitor studies focus other topics (images, level of awareness, knowledge of the institution, perceptions from outside, structural factors)

Aims
- Visitor studies aim at *making visitors coming more often und become more satisfied*
- Non-visitor studies aim at bringing people first time to the *threshold of the venue*

Both
- Want to *reduce barriers* to the visit and become more attractive
- Non-visitors can also be analysed within visitor research (for example when subscriptions are cancelled, visitors that are rarely coming)

Cultural institutions should ask themselves a number of questions:

- **Who are the non-visitors exactly?**

- **What are their interests, motives, behaviours, social structures, milieus**

- **What stops them from visiting? Barrieres**

- **How could they be enthused for the own offers? Which "pull factors„ are there?**

- **What methods can be used to find out more about non-visitors?**

Non-visitors - an approach to the terminology

- **Exact description depends on exactly how these non-visitors are captured**

- **Belonging to the group of visitors or non-visitors depends strongly on, for example,**
  - which cultural offer counts (culture in general, genre, institution)
  - which form of visit is chosen
  - which time horizon is chosen

- **If the concept is narrowed to high culture - proportion of non-visitors far greater than with a broader concept of culture**

- **What is counted as a specific visit? Which places? Analogue-digital? Duration? Frequency in the past?**

- **Important: Definition from how many/how few visits in a certain period of time / in a certain facility if someone is understood as a non-visitor or as an occasional visitor.**

- **Examples of visitor or non-visitor definitions...**

Non-visitors - an approach to the terminology

Best-known population surveys in Germany: **KulturBarometer**

- Non-visitors: visit culture **less often than once a year**

- Occasional visitors: **Visit** facilities **less often than once a month, at least once a year**

- **Result of the study**: 50% non-visitors, approx. 35-40% occasional visitors, approx. 5-10% frequent visitors.

Study commissioned by the German Stage Association counts, for example:

- Non-attendance**: No** theatre performance for **more than three years** and **no more than one musical or festival performance per year.**

- **Consequence: Regular musical visitors of a multi-speciality theatre thus fall under non-visitors**

Visitors to pubblicy funded
cultural institutions in Germany

(Renz 2016)

- Core visitors
- Occasional visitors
- Never visitors

Cultural behaviour in the EU

	Not in the last twelve months	1-2 times	3-5 times	More than 5 times	Don't know
Watched or listened to a cultural programme on TV or on the radio	27% / 21%	16% / 17%	15% / 15%	41% / 46%	1% / 1%
Read a book	32% / 28%	19% / 20%	12% / 14%	37% / 37%	0% / 1%
Been to the cinema	48% / 48%	23% / 20%	16% / 14%	13% / 17%	0% / 1%
Visited a historical monument or site (palaces, castles, churches, gardens, etc.)	48% / 45%	27% / 29%	13% / 13%	12% / 12%	0% / 1%
Visited a museum or gallery	62% / 58%	23% / 25%	8% / 9%	6% / 7%	1% / 1%
Been to a concert	65% / 62%	24% / 24%	7% / 8%	4% / 5%	0% / 1%
Visited a public library	68% / 64%	11% / 13%	6% / 6%	14% / 16%	1% / 1%
Been to the theatre	72% / 68%	20% / 21%	5% / 7%	3% / 4%	0% / 0%
Seen a ballet, a dance performance or an opera	81% / 82%	13% / 13%	3% / 3%	2% / 2%	1% / 0%

EB79.2 Apr-May 2013
EB67.1 Feb-Mar. 2007

Not in the last twelve months — 1-2 times — 3-5 times — More than 5 times — Don't know — EU27

How frequently do Europeans listen to music?
(AS A % OF THE EU POPULATION)

Frequency	Value
Every day	61.3
Several times a week	20
Once a week	4.2
1 to 3 times a month	2.3
Less often	5.8
Never	6
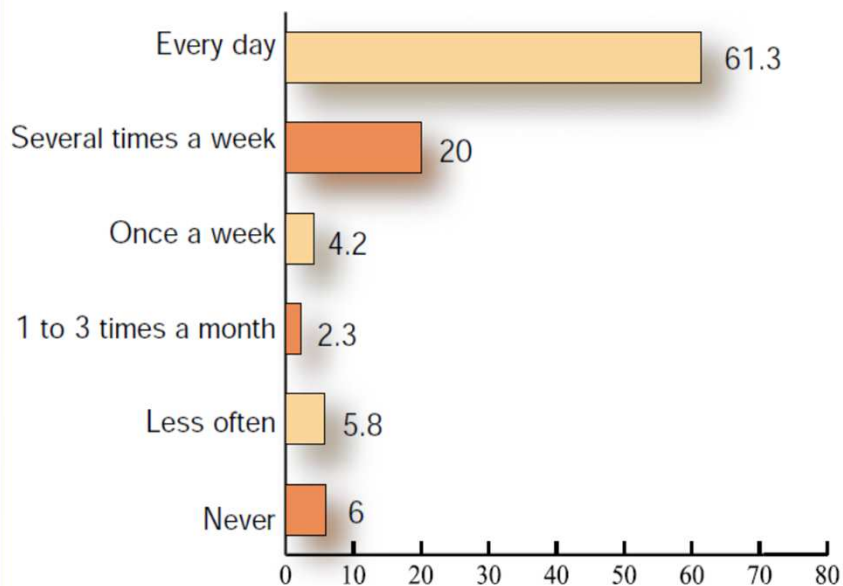
base: 16162

Type of music listened to by Europeans
(AS A % OF THE POPULATION LISTENING TO MUSIC)

Type	Value
Rock, pop	55.1
Easy listening	32.1
Folk, traditional	29.9
Classical music	28
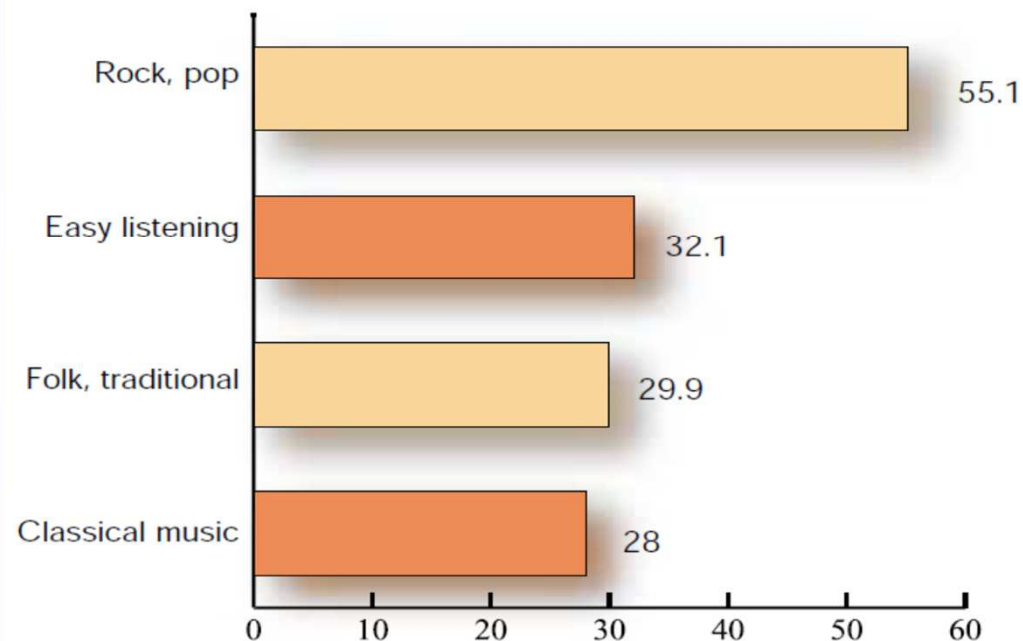
The majority of European citizens (55.1%) listen to rock and pop music. This tendency is mostly found in Denmark (69.9%), France (69.1%) and Belgium (64.8%). On the other hand, the lowest rate of people listening to this kind of music is found in Austria, Finland, Portugal and Greece (43.8%, 42.9%, 36.7% and 24.6% respectively).

One third of the respondends (29.9%) also mentioned folk and traditional music. This is the kind of music that is most listened to in Portugal, Greece and Austria (65.1%, 62.5%, 44.8% respectively of the interviewees). We can also see that in Sweden and Spain, traditional music is the second type of music most listened to by respectively. Classical music is in Luxembourg (45.8%), Sweden (41.3%) and the United Kingdom (40.9%) most widely spread.

Non-visitors - an approach to the terminology

- **Geographical, socio-demographic and/or socio-economic factors have different relevance in the decision-making processes and for the behaviour of visitors.**

- **Conclusion: Even if the definition of the term appears simple at first glance, it can vary greatly.**

Reasons for not visiting cultural institutions I

Reasons that lie *in the person of the non-visitor*:

- **Lack of motivation**
- **Socialisation, no contact with culture through parents, school, peer group**
- **Low education, offer is unknown or incomprehensible or exhausting**
- **Cultural foreignness (you don't know how to behave there, visitors are different from you)**
- **Other personal interests and social environment (one goes somewhere else, no relevance for one's own life, circle of friends prefers other activities)**
- **Low income**
- **No time (often pretextual)**
- **No interest**
- **Bad experiences in the past („No More Visitors")**

    → These institutional barriers can only be removed in the long term!

Reasons for not visiting cultural institutions II

There is a basic motivation to visit, but there are hurdles in the offer that work against it:

- **No or few cultural offers available (for example in the countryside)**

- **Many offers in the modern leisure industrie that compete with traditional cultural offers**

- **Specific barriers of the institution (poor quality/bad service, poor experience, inconvenient performance times)**

- **Barriers due to lack of information (wrong media channels, not knowing where info is available, inappropriate forms of presentation, target group is not reached)**

- **Physical access barriers (long distances, no parking, no public transport, no accessibility for handicapped)**

- **Barriers due to too high costs, lack of time, problem of scheduling**

- **Barriers due to other personal reasons (no escort, family circumstances, objective lack of time, poor physical constitution…)**

    → Removing hurdles is within the control of the institution

Problem

- **Subjective barriers are more effective than those that come from the offer or the institution**

- **These barriers are hard to overcome by the institution**

Cultural Institutions face permanent challenges due to lack of age diversity

Generation effect

Interest arises through socialisation and lasts throughout life (i.e. generation Y/ Y/ Z, millenials...)
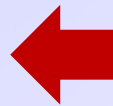
Period effect

Certain events in time shape interests and needs (for ex. Corona pandemia, fiancial depression, upcoming of internet etc.)

Age / life - cycle effect

Interest begins with entry into certain age

Reasons for not visiting cultural institutions III

1. Get a clear picture of which barriers actually come into play for which target group; barriers are strongly group-specific

2. Merely lowering the barriers does not yet lead automatically to more cultural visits
   -> Target groups must be informed accordingly

3. People with an affinity for culture have different visitor barriers compared to (rather) non-culturally affine population groups.

4. Non-culturally affine population groups primarily have social or subjective barriers

5. Frequent reasons for not coming, regardless of cultural affinity, are too high admission prices, lack of time and few cultural offerings in the vicinity.

   → Often, however, only pretextual reasons for other prioritisation or social and subjective barriers.

How can non-visitors be inspired for the own offers?

- **First consideration: Are they really non-visitors? Or are they rather occasional visitors?**

- **From a cultural marketing perspective, *potential visitors who* can at least be classified as *occasional visitors* are of particular interest -> where can they be found?**

- **From a cultural participation perspective, people who rarely or never attend cultural events are of particular interest; also they are the biggest group → are much more difficult to reach, even if barriers are removed.**

- **Classic communication channels (e.g. website, print material etc..) are mostly suitable for people with an affinity to culture**

- **They are not suitable for non-culturally affine people due to poor accessibility, difficulties in understanding, media tailored to the core audience and a general lack of interest.**

Audience Development

- **Only chance of success with long-term anchoring of audience development in the institution**

- **Just keeping the offer and marketing it new/better/differently is not enough**

- **For those with an affinity to culture, the focus is on learning new things, while for the more infrequent visitors, other aspects, such as social interaction, fun or relaxation, are more important**

→ **Reorientation of the offer to approach non-culturally affine population: More fun and event orientation required**

→ **Leave the classic places / venues, go out on the street and perform so-called third places in the city (adult education centre, libraries, ...).**

- **Classical audience development has its limits with a population that does not take advantage of any cultural offerings at all**

How to find out more about non-visitors?

Depends on what exactly is to be found out
→ **No "patent remedy"**

There are different ways to approach the topic:

- **Research: What studies and literature are available and what measures have already been implemented and evaluated?**

- ***Continuous visitor studies*: Enables potential to be identified through comparisons between frequent, occasional and infrequent visitors; conclusions onto potential visitors possible**

- **Collect information about non-visitors outside the own house: However, it must be checked beforehand which locations are to be involved, as the non-visitors are to be recorded representatively for the entirety.**
  The aim is to record all non-visitors/visitors so that highly calculable visitor potentials can be determined.

(How) did Covid-19 get the audience structure changed?

The age structure in the audience - first sharp falls, then short-term recovery....

Proportion of over 60s in the total audience

Berlin
(ICTf 2023)
n = 66.344
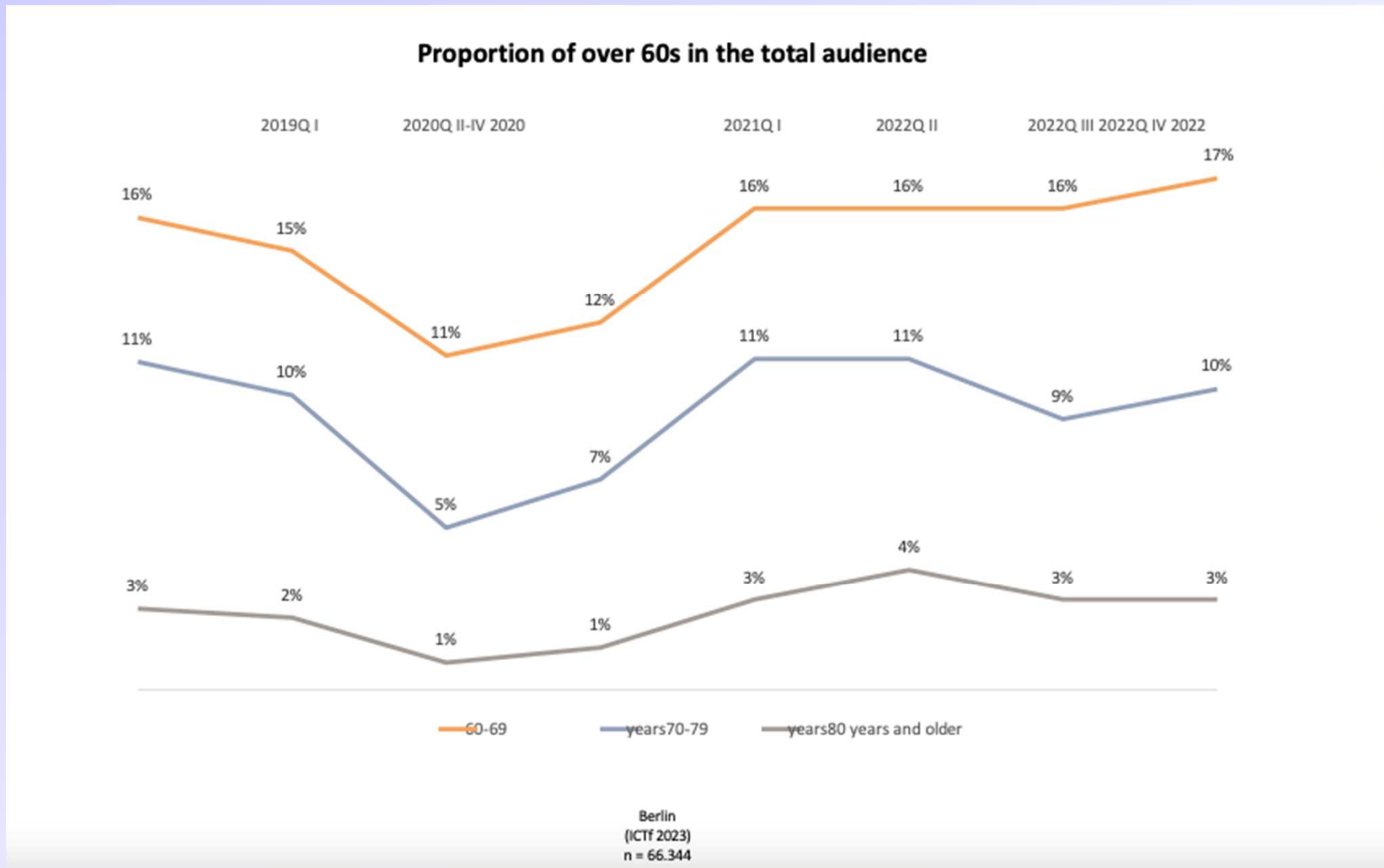
Local visitors do not (yet) compensate for tourists who stay away.

Place of residence of the respondents

| | 2019Q I | 2020Q II-IV 2020 | | 2021Q I | 2022Q II | 2022Q III | 2022Q IV 2022 |

Berlin — Germany (outside Berlin) — Abroad
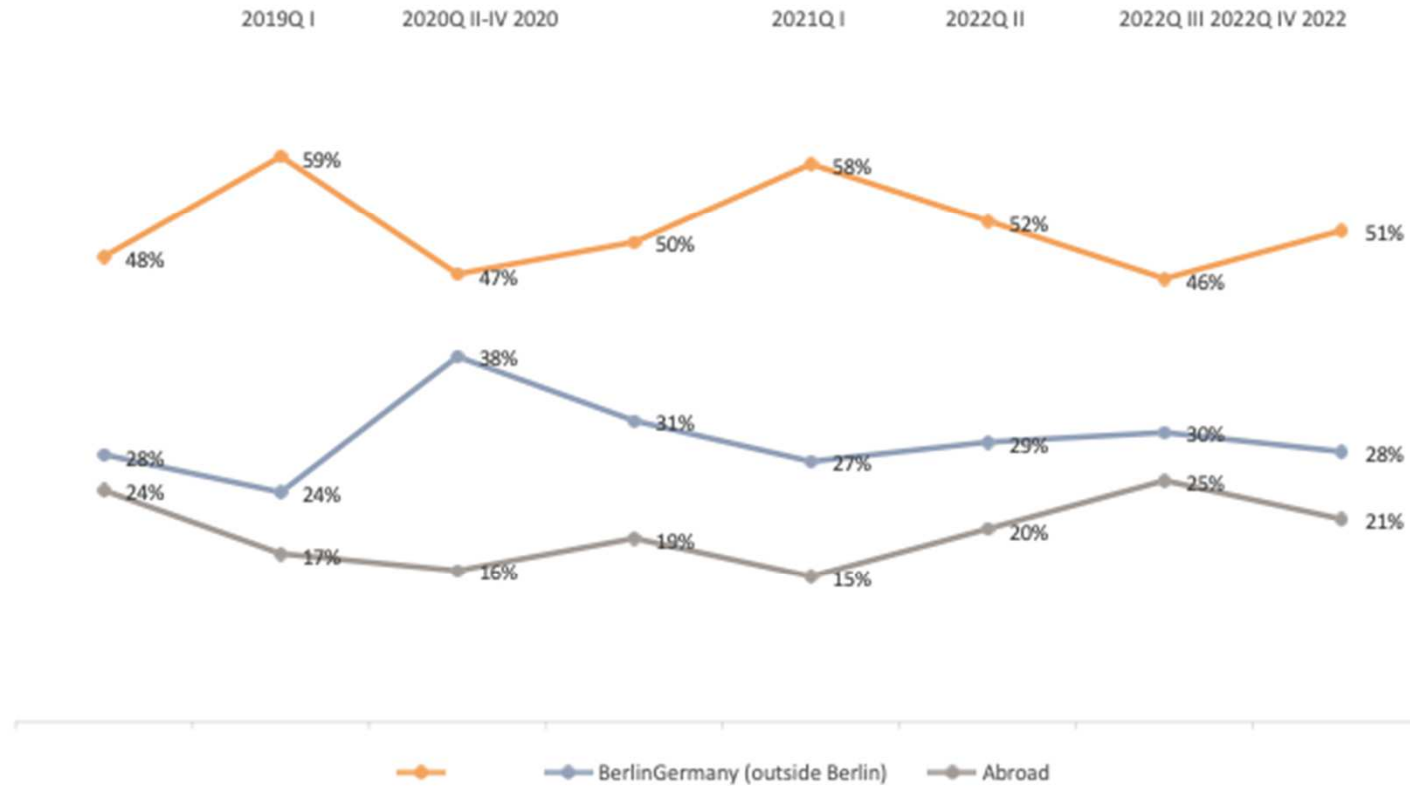
Berlin
(ICTf 2023)
n = 66.344

The pandemic has increased social inequality in the Audience further aggravated

Changes in the proportions of low modernity/closed biography lifestyles in the audience.

Conservative Upscale
- 2019: 7,3%
- Q II-IV 2020: 4,0%
- 2021: 5,0%
- Q I 2022: 4,6%
- Q II 2022: 5,3%
- Q III 2022: 4,9%
- Q IV 2022: 6,6%

Conventionalists
- 2019: 4,6%
- Q II-IV 2020: 2,1%
- 2021: 4,7%
- Q I 2022: 2,7%
- Q II 2022: 3,0%
- Q III 2022: 3,3%
- Q IV 2022: 3,3%

Down to earth traditional
- 2019: 2,9%
- Q II-IV 2020: 0,7%
- 2021: 2,7%
- Q I 2022: 1,1%
- Q II 2022: 1,0%
- Q III 2022: 1,3%
- Q IV 2022: 1,2%

Legend:
- 2019
- Q II-IV 2020
- 2021
- Q I 2022
- Q II 2022
- Q III 2022
- Q IV 2022

Berlin
(ICTf 2023)
n = 34.329

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

Barriers that prevent visits

Barriers that prevent visits

Problem

- Subjective barriers are more effective than from the offer outgoing barriers!

Problem 2

Cultural behaviour in Urban Romania

	Not in the last 12 months	1-2 times	3-5 times	More than 5 times
Watched or listened to a cultural programme on TV or on the radio (M)	30.8	18.2	15.2	33.3
Read a book	40.5	22.8	11.8	22.5
Visited a historical monument/site	41.5	27.7	14.1	13.8
Been to a concert	49.3	28.1	11.8	9.2
Been to cinema	51.4	22	12.8	12.4
Visited a museum or gallery	54.9	27	7.9	7.5
Seen a ballet, a dance performance or an opera	62.1	24.3	7	.2
Been to the theatre	67.6	18.6	6.9	.4
Visited a public library	68.5	11.3	.9	10.5

Legend:
- Not in the last 12 months
- 1-2 times
- 3-5 times
- More than 5 times

Co-funded by the European Union

RED Re-Educating Digitisation

Rheinische Hochschule Köln

Cultural behaviour in Urban Romania

	Lack of interest	Lack of time	too expensive	Lack of information	Limited offer or low quality
Visited a public library	25	53.1			
Seen a ballet, a dance...	22.5	54.3			
Been to the theatre	20.4	51.5			
Visited a museum or gallery	18.4	52.1			
Been to cinema	18.3	55.6			
Read a book	17.9	58.5			
Watched or listened to a...	17.1	53.9			
Been to a concert	15.6	51			

END OF SECTION !

RED – Workshop
Non-Visitor Research II

- Sampling and Representativity
in Non-Visitor Studies -

Author:
Prof. Dr. Tibor Kliment

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

Why sampling: Advantages and disadvantages

- Full surveys are often difficult or impossible to conduct with large populations
- Sampling is much more cost-effective
- Sampling allows for more complex research designs
- Sampling is more accurate
- Sampling is faster

But:

- Estimates based on samples contain errors
- The precision with which the population parameters can be estimated
- Sampling Error depends on the sample quality and the size of the sample

Sample

☐ **Sample: Subset of the population selected either purely at random *or* according to specific criteria (e.g. random selection).**

☐ **The aim is to estimate parameters (e.g. means, proportions, percentages) from the results in the sample.**

☐ **On the basis of the sample, statements are to be made about the underlying population (extrapolation).**

☐ **If the population cannot be specified, a representative sample is not possible. Nor is it clear what the results refer to!**

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

Def. Population

- All units (e.g. inhabitants of a city, visitors of a festival etc.) that come into question for the study
- We distinguish between **drawing population:**
  The drawing population, also known as the "sampling frame" or "sampling population," represents the entire group or set from which a sample is drawn. It encompasses all the potential individuals, entities, or elements that could be part of the study. The drawing population serves as the initial pool from which a sample is selected to ensure that the sample is representative of the broader population.
- And **survey population:**
  The survey population, sometimes referred to as the "target population" or "study population," is a subset of the drawing population. It consists of the specific group of individuals or entities that the researcher or survey aims to study and collect data from.

- **The population must be described as precisely as possible in terms of subject matter, time and location,** e.g.

- *... all adults who have their permanent residence in Barcelona in fall 2023*; all visitors who have attended screenings of *See-You-Sound festival in Tourin in 2023*.

- If all members of the population are interviewed, this is called a full survey.

"Under-" and "Overcoverage" in a sample

Population Failures

Refusals

Inference population

Undercoverage
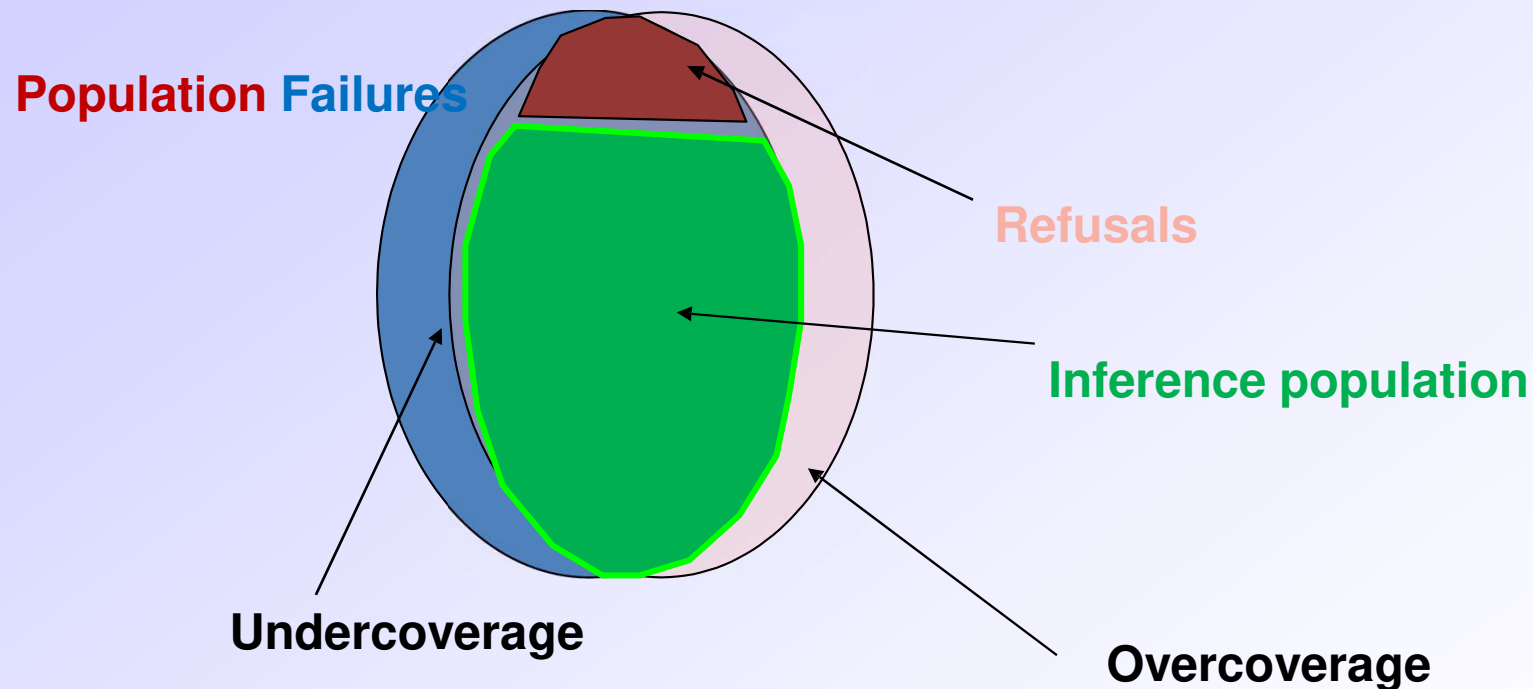
Overcoverage

Example: All people who are resident in Cologne
Undercoverage: People in hospital, students in dormitories etc.
(=> Residents' Registration Office sample)
Overcoverage: Visitors, commuters

Representativeness - requirements for the selection procedure

Three criteria must be met for a sample to be „representative":

1. The population is **known** and **definable**

2. Elements of the sample are **defined**

3. The selection procedure is **systematic** and **nameable**

Note: There is *no general representativeness* of a sample, but only *representativeness in relation to something specific* !

-> *"For whom or what is the sample representative?"*

Exercise: How representative are the respective samples?

1. You run a film-festival and do a **visitor survey** on your website with every 10th. visitor entering the site

2. You do a **visitor study** by sending questionaires by e-mail to the subscribers

3. You do a **visitor study** by laying out the questionaires somewhere in the venue

4. You post a **non-visitor questionaire** to your instagram account, followers who then fill out the questionaire

5. You do a **non-visitor study** and draw a sample from the City Residents' Registration Office and send a questionnaire link to the addresses

6. You do a **non-visitor study** in the biggest shopping mall of your city

7. You do **non-visitor interviews** in front of two theaters and three museums

Requirements for the selection procedure: How large should a sample be?

The answer is not easy to answer, but some clues apply:

1. **How homogeneous or homogenous is the population?**
   ->The more inhomogeneous the population, the larger the sample must be

2. **How detailed do I want to evaluate subgroups? -> See table below**

3. **Problems arise when analyzing to small subgroups**

N=300	Non visitors	visitors	
16-29	50	50	**100**
30-45	50	50	**100**
46-60	50	50	**100**
	150	**150**	**300**

4. **How big should the certainty of the decision be in a representation conclusion?**

5. **How much time do I have for data collection?**

Attention: What does not matter: The size of the underlying population, venue visitors etc!

Interval sizes in representation inference

Random sample-related margins level of significance at 5%

Sample size

Half width of the confidence interval

Messwert: Erwartungswert in % in der (Teil)Stichprobe

Legend (Sample size):
- 200 Fälle
- 300 Fälle
- 400 Fälle
- 500 Fälle
- 600 Fälle
- 700 Fälle
- 800 Fälle
- 900 Fälle
- 1.000 Fälle
- 1.500 Fälle
- 2.000 Fälle
- 2.500 Fälle
- 3.000 Fälle
- 3.500 Fälle
- 4.000 Fälle
- 5.000 Fälle
- 10.000 Fälle
- 15.000 Fälle
- 20.000 Fälle
- 25.000 Fälle
- 30.000 Fälle

https://www.calculator.net/sample-size-calculator.html?type=2&cl2=95&ss2=100&pc2=60&ps2=100000&x=89&y=16#findci

Co-funded by the European Union

RED Re-Educating Digitisation

Rheinische Hochschule Köln

Selection procedure

So the central question is: How can it be ensured that the parameters of a population are estimated as unbiased as possible?

Through different selection procedures:

- **Arbitrary** (non-controlled) selection (e.g. pedestrian survey, snowball sample, shopping area, sites of interest etc.)

- **Not-random** / **deliberate** / purposive selection (e.g. quota sampling)

- **Random selection** (condition for the application of *inferential statistics*)

Non-random selections

Arbitrary selection

Conscious selection
(deliberate selection at discretion)

**Quota selection
(stratified random selection)**

Selection of "typical" or "extreme" cases".
(mostly when doing qualitative interviews with open questions)

Selection according to concentration principle
(selection only of the „most important" cases)

**"Snowball procedure„
Mostly when doing qualitative int. with open questions.**

Arbitrary selection

- **E.g.: Pedestrian surveys by reporters of passers-by in the shopping zone, in front of music club, a theatre etc.**
  - **The representativeness of the sample is not given**
  - **Also the numerical basis is unclear, makes extrapolations impossible**
  - **The results are therefore usually worthless in a scientific sense (but may still informative for *market research*)**

Non-random sampling

-> **The selection of the interviewees**

Of big importance in this kind of sampling is the *quota selection:*

By specifying quotas (i.e. proportions with which certain characteristic values should be present in the sample), the aim is to achieve a representative sample.

- **Selection according to concentration principle**

- **Selection of *typical cases* (middle: typical moviegoer or music fan; extremes: rare moviegoers vs. often moviegoers)**

- **Important technique, when you do open, qualitative interviews**

Conscious selection: "Quota selection"

- When random selection is not practical (too expensive, too time-consuming, no list available etc.).
  Often used in commercial market research institutes.

- In this process, there is given a quota plan how high the proportion of people with certain socio-demographic characteristics must be among the respondents (often gender, age, education, sometimes also multi-level quota [*combination of characteristics*].

- Prerequisite: Distribution of these characteristics in the population is known and can be reproduced in the sample.

Example: Quota selection for Barcelona

AGE GROUPS

Age	TOTAL	Women	Men		Age	TOTAL	Women	Men
TOTAL	1.660.435	870.269	790.166		TOTAL	1.452.000	0,52	0,48
0-15 years	208.185	100.979	107.206		0-15 years			
16-24 years	148.153	73.084	75.069		16-24 years	0,13		
25-39 years	378.152	189.871	188.281		25-39 years	0,34		
40-64 years	574.874	295.741	279.133		40-64 years	0,52		
65 years or older	351.071	210.594	140.477		65 years or older			

LEVEL OF EDUCATION (FROM 16 YEARS)

Level of education	TOTAL	Women	Men		Level of education	TOTAL	Women	Men
TOTAL	1.452.250	769.290	682.960		TOTAL	1.452.250	769.290	682.96
Without education	19.617	14.431	5.186		Without education	0,01		
Primary school	208.204	116.310	91.894		Primary school	0,14		
Secundary school	320.320	160.449	159.871		Secundary school	0,22		
High school	365.741	182.227	183.514		High school	0,25		
Universitiy or superior studies	515.186	284.719	230.467		Universitiy or superior studies	0,36		
No answer	23.182	11.154	12.028		No answer	0,01		

Problems of quota selection

- The selected characteristics (e.g. age, gender) were the quota is built on, **have to be known in advance**. The appropiate statistics about the distribution of the variables must be known and available before.

- They are oriented towards **easy feasibility** and are often only marginally related to what is actually of interest (e.g. brand-related attitudes, media use, cultural behaviour).

- **Arbitrariness of the interviewers**, who are often left to choose respondents by themselves (and then tend to select cooperative or known persons). Thus risk of uncontrollable bias.
  -> Can be compensated for by a heterogeneous, large number of interviewers.

- **People are contacted until the quotas are fulfilled**. It is mainly those who are easy to reach who are in the sample (bias).

- Also: Since there is no random selection, there is **no inferential statistics possible!** (but is however still done in practice)

- *However, it may deliver somewhat reliable results in practice!*

„Snowball procedure": Selection of specific groups

☐ Usefull, if the population is difficult to reach or find, or the characteristic of interest is not "visible" or registered, or if people hardly cooperate.
Ex.: The „film industry experts", "marketeers in film distributer companies" etc.

☐ *Procedure:* Identification & interview with a starting person who after the interview can name other people

☐ Recommended people are referred to by the person that named them, are interviewed and can make further recommendations after the interview.

☐ **No representative selection of respondents by this technique possible!**

Random selections

Simple probability selection

Complex probability selection

Card item selection

Area selection

Cluster Sampling (cluster selection)

Multilevel selection

Stratified selection

Proportionally layered

Disproportionately stratified

Simple random selec. Random-digit dialing

Systematic random selection

Simple random selection

Systematic random selection

Pure random selection

Systematic random selection

Random sampling

Sampling is *random if* each unit of the sample population has an **equal** or a ***predictable chance of being included*** in the sample.

Advantage of random sampling:

- Due to the random process, one can specify exactly the probality by which elements are selected and how sample results deviate from those of the population.

- You do **not need information about properties** of the underlying population (lack box), properties are automatically reproduced, als relationships among those.

- ***Only this method allows the inferential statistical*** conclusion from the sample to the population, in which error probabilities and confidence intervals can be calculated and determined!

- *-> Sampling error calculation*

Random sampling

a) Simple random sampling

-> All elements are selected in a single process

1) Pure random sample

All sample elements are determined independently of each other at random.

2) Systematic random sampling

Only <u>the first</u> sample element is determined at random, all other elements are selected systematically. Sample elements are dependent on each other. However, this selection is random based, if there is no underlying structure under.
This method is very usefull for **visitors studies** in the venue.
Used most often at all in visitor studies.

Random sampling

b) Complex random sampling

1) Stratified random sample

The elements of the population are divided into strata (or groups) with respect to the characteristic of interest. Separate random samples are then drawn from these strata. E.g. places, city districts, households, persons. Aim is to reduce variance.

2) Cluster sampling

The selection is made from aggregates of study units (clusters). For instance different venues for movie screenings

3) Multi-stage sampling

One after the other, a series of simple random samples is carried out.

Multilevel random selection

Stage 1:
Community selection
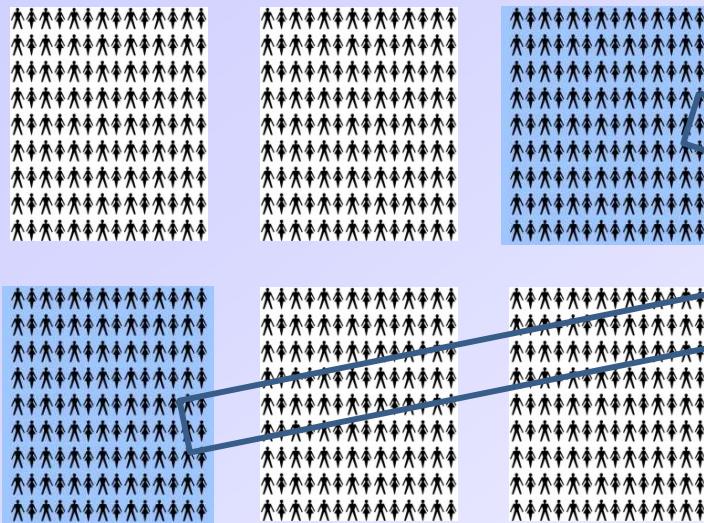
Stage 2:
Household selection
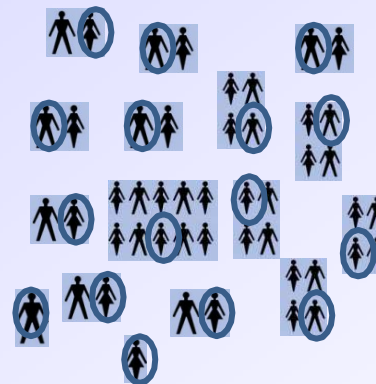
Stage 3:
Person selection

o **Different selection probabilities can be corrected by weighting**

o **In m-person households, the selection probability is 1/m, weighting by reciprocal value m/1. Need to know household size!**

o **Selection in the household by „*Swedish-Key*" or „*Next Birthday*"- Method!**

Co-funded by
the European Union

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

Selection for *random samples*

1. **List selection: List of all units e.g. list of inhabitants of registration offices of municipalities, list of all telephone entries.**

2. **RDD method: Generation of telephone numbers by random digit dialling or randomised last digit.**

3. **Area selection: Random route method**

   a. **Random selection of areas (municipalities)**

   b. **Selection of start addresses**

   c. **Walk-through instructions for selecting an address (e.g. first street on the right, next on the left, third house, second floor)**

5. **Probability selection within the household: Birthday method (person in household who had last birthday), Sweden key**

The Random Route Procedure
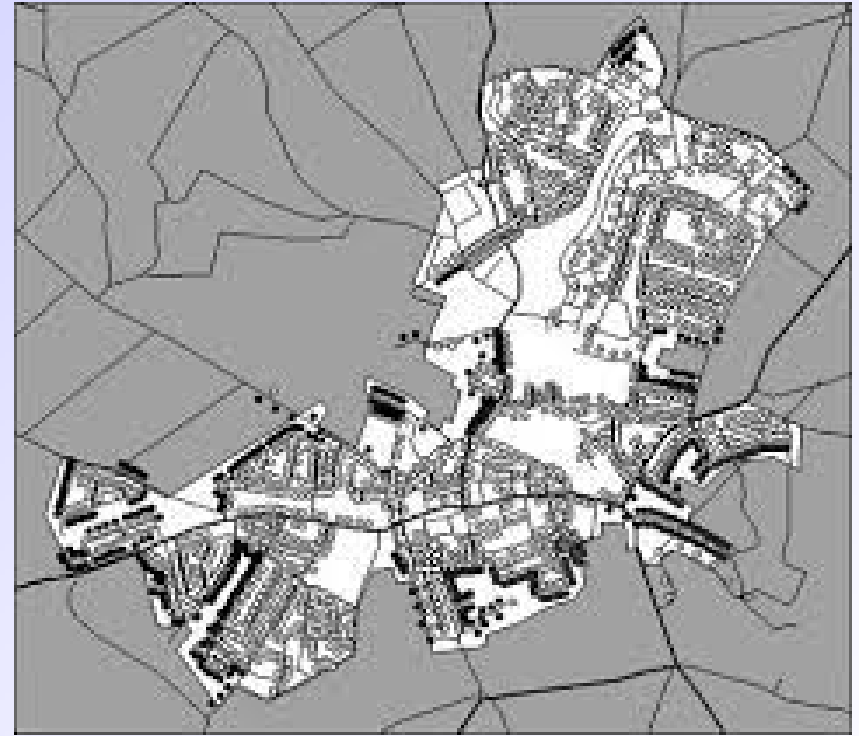
Start address 1 Start address 2 Start address 3 Start address 4 Start address 5

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln
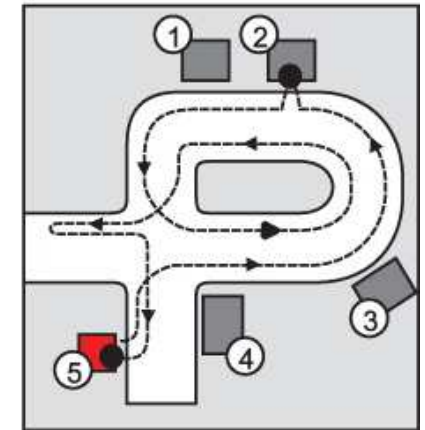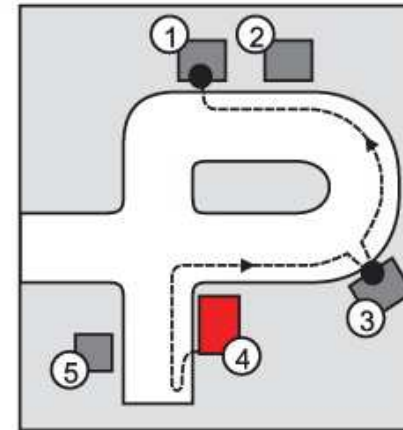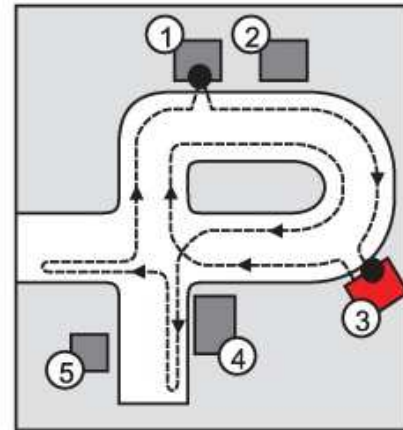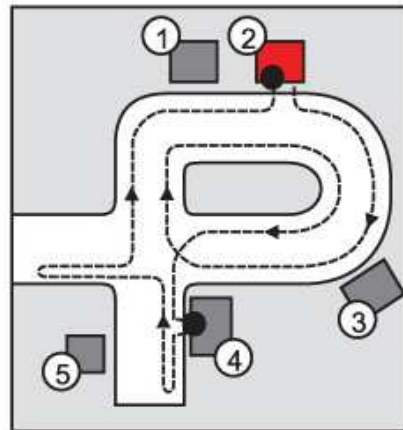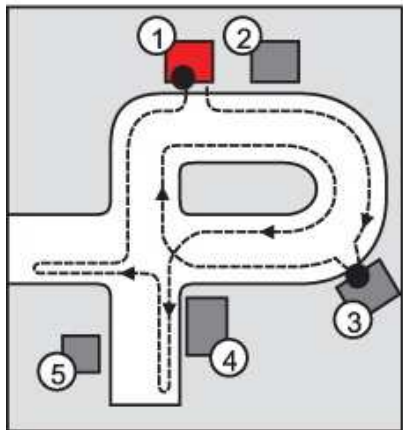
Practical problems of random sampling

- ***Interviewer fabrication***
  - for Random Route: fullfilling to quota

- ***Nonresponse problem***
  - Response rates often < 20
  - Nonresponse usually not random
    - Non-interviewees: sick people, language problems, ...
    - Hard to reach: Unusual working hours, mobile persons
    - Refusers: Distrust, political attitude, ...
    - Item non-response: For income about 25%.

- **Problem of selective sampling (e.g. middle class bias)**

- **Incomplete registers or lists**

Examples of *failures in sampling*

- **Addresses not found**
- **Persons interviewed who do not belong to the target population**
- **Unavailability of a contact person / the target person**
- **No telephone connection/mobile phone/internet**
- **Illness of the target person**
- **Refusal of the contact person / the target person**
- **Communication problems**
- **Termination of the interview**
- **Error or fraud by interviewer**
- **Data acquisition and processing errors**

...

When exactly is low response rate/selective sampling problematic and when is it not?

Problematic when...

- I would like to **infer the population from my sample**,
  e.g. determine the average income

- would like to quantify certain groups, e.g. the proportion of regular movie-goers or target groups in the population

Not problematic if...

- I want to test *certain aspects*
  e.g. check the usability of a website, check impact of ads

- I conduct a "qualitative" study with a few in-depth interviews (more on this later).

Representativeness and *failures*

- The response rate alone is not very meaningful, more important are **systematic failures**

- It depends largely on the topic of the questionaire and the sampling method whether a default is considered a neutral default or a systematic default

- Special efforts must be made to reach refusers and hard-to-find people. Distortions of the results are most likely to be expected here.

- Even if certain socio-demographic characteristics (such as gender, age, etc.) of the sample correspond to those of the population they may not say something about the distribution of other characteristics.

- Thus, they cannot prove representativeness with **regard to the characteristics and constructs studied.**

- Quota samples in particular also have this problem.

Basic tips for fieldwork in a venue (after visit sampling):

■ Participation in a survey should be made as easy and pleasant as possible for the respondents. If you do a **after-visit survey** make sure people have time for a survey (e.g. after the performance but before going to the cloakroom)

■ Make sure a **quiet place** is chosen for the interview and that sufficient seating and writing facilities are available. To motivate respondents, small gifts are useful.

■ In order to provide a representative sample of respondents you may distribute the interviews over **all days of the festival and over different times per day**. Make a precise **timetable** and continuously review it to ensure that the sample is conducted in a targeted manner.

If extra (external) staff is deployed for the field work, they should be **trained before** and consulted regularly on the progress and possible ambiguities

■ The response rate of surveys (i.e. the number of completed questionnaires) cannot always be calculated exactly in advance. It depends strongly on the number of visits and the motivation of the respondents to participate. Buffers should always be planned in order to be able to extend the survey period if needed.

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

Finally:

1. **What is the best way to create a sample from *visitors*?**

   – **Draw a systematic sample (every xth.-visitor is approached in the venue)**
   – **Draw a random sample from a data-list or make a full survey (subscribers or former visitors etc.)**
   – **Distribute questionaires on every seat in the performance hall**

2. **What is the best way to create a sample from non-visitors?**

   – **Get a sample of adresses from the city´s registration office -> probably the easiest way but requires some work resp. personell**
   – **If there is: Use an access panel of a market reseach institute**
   – **Try to get funds and let the field work be done by a professional market research institute**
   – **If noting else is possible: Approach visitors in different places of the city and points of interest (theatres, museums, library ...) and do a weighting procedere afterwards**

Practical task:

Now we strengthen representativity of our Barcelona survey by doing a weighting process

SPSS

- **Create a new variable „weight" = 1 by compute command**

- **Do a frequency command of *gender* and *age* (as to age: build age-groups) and *professional status***

- **Compare the data from the sample with the statistical figures from the city of Barcelona**

- **Create the weight variable by weight = 1**

- **Do „IF procedure" for all values of gender and age:**

    IF … weight = target value / actual value

- **Switch the weight variable on**

- **Make a frequency command and check if the new values like they should be**

Practical task:

Now we strengthen representativity of our Barcelona survey by doing a weighting process: Adjust the sample to reality.

PSPP

- **Make a frequency of age and gender (for that build age groups)**

- **Create a new variable „weight"; what does weight mean?**

- **COMPUTE Weight = 1.**

- **Look into the table figures from Barcelona; make a "shall through is"**

- **Do „IF procedure" for all values of gender and age:**

    IF (variable x = …) weight = … .

    Do this for all values of variable x…

- **Make a frequency command of weight**

- **Switch the weight variable on**

- **Make a frequency command and check if the new figures are like they should be**

RED
Non-Visitor Research II

- Data Analysis of IN-EDIT Non-Visitor Survey –

Author:
Prof. Dr. Tibor Kliment

Co-funded by
the European Union

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

Some general recommendations about data evaluation and presentation of results

It is important to formulate the right questions prior to the analysis

Only those who ask questions will get answers from the data later...
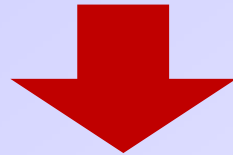
Definition of interest...

– How can we imagine our non-/almost visitors? Who comes, why, who doesn't?

– What are possible/new target groups for us? What's their size/potential?

– What is social structure of our targets?

– How can we attract new target? What are their needs, motifs and barriers?

– How can we make our communication more effective and efficient?
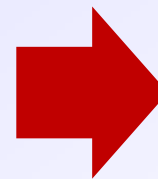
Defining the interest in knowledge in the later data analysis...

To be discussed today...
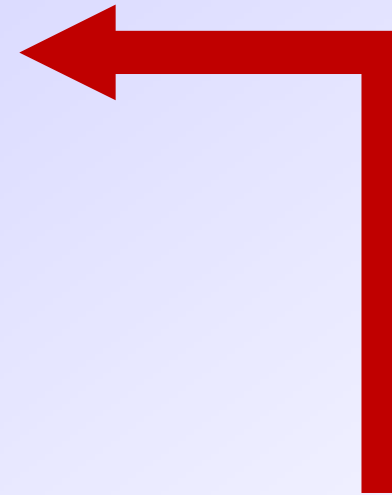
Analysing the Barcelona - Data

DEVELOPING ANALYSIS STRATEGIES FOR IN-EDIT SURVEY

First steps:

1. **Let´s clean the data, check for missing values, assign proper variable descriptions**

2. **The better representativity by weighting the data by the figures from Barcelona**

3. **Build categories for relevant variables (price, age, years last visit)**

4. **First step to analyse the data: Do frequencies, build crosstabs, use means, divide results into subgroups …**

5. **Draw practical conclusions from that findings; what are your learnings / measures to take?**

6. **Write down your most important findings**

A) Overview about major findings

1. *Go through the data and get an overview. What strikes you, what did you not expect?*

2. *What is the leisure behaviour of people in Barcelona? How important are they going to festivals, cinemas? Make a ranking according to importance (consider only top two categories).*
   Divide leisure activities among different groups (age, gender, education).

3. *Find out, whether people that visit film-festivals are similar or different from visitors of cinemas or concerts or music festivals.*

4. *Find out how many people do not know In-Edit, simply know it or have even been there? Divide among different groups (age, gender, education). Make tables for doing comparisons and calculate mean values.*

5. *Find out, when visitors came to In-Edit last time. Make comparisons by age gender, education and see whether you find differences.*

B) Demografics and IN-EDIT FESTIVAL

1. *What is the image of IN-EDIT? Is it different among different "knowers" and other subgroups (have visited, only known by media etc.)*

2. *How many people are interested in visiting IN-EDIT?*

3. *Describe the different groups that are interested or not by doing tables using demografics (age, gender and education).*

4. *What is the image of IN-EDIT? Find out, whether the image is different among different "knowers" (have visited, only known by media etc.)*

5. *What do people expect from culture events? Make a ranking and divide by age, gender and profession.*

C) Needs, barriers, interests

1. *Find out, what the most important barriers to a* IN-EDIT *visit are. For whom is what important?*

2. *What may attract people when visiting cultural events/IN-EDIT? Differentiate between social groups and vising interest.*

3. *Investigate into willingness to pay. Build groups of prices. Combine prices by demografics.*

4. *Find out about media use concerning cultural events. Make tables by age gender, education.*

5. *Combine media use by visiting interest. Use people that have interest different channels?*

Some general hints for creating tables

- Do not be contended with basic counts, but always work with **means of centrality, combined variables resp. tables** and correlation measures

- Always select **column percentages in** the tables and **stick to them in** the whole presentation

- **Take independent variables** for the **header columns**, put **dependent variable in the rows**

- Select relevant **target groups/characteristics for subgroup comparisons** and tabulate all questions in the questionaire over it

- Always note a **total column** as the reference of the subgroup comparisons

- Indicate in all tabels or grafics what **respondent base** the answers refer to (regard filter questions!)

- Do not further analyse subgroups below **80 cases,** and simply ensure a reader-friendly presentation

END OF SECTION !

RED
Non-Visitor Research II

- STP-Modell, Pricing and Life-Style Analysis –

Prof. Dr. Tibor Kliment
Barcelona, October/November 2024

Overview

1. **Introduction to the STP-Model**
2. **Segmentation: Understanding Your Market**
3. **Targeting: Finding Your Ideal Customers**
4. **Positioning: Creating a Unique Brand**
5. **Example from a study in Cologne**
6. **Collective exercise with Barcelona study**

What is the STP-Modell?

STP in marketing stands for *segmentation*, *targeting and positioning*. It is a three-step formula that helps companies and institutions segment their target audience, target the right buyers and position their products for maximum impact.

Before the STP model existed, the focus in marketing was mainly on intensive promotion of products, but without knowing the market or the target group.

However, this approach was only moderately successful for companies. In search of a better strategy, Philip Kotler invented the STP model in 1969.

But have in mind: STP moves rather the customer into the center (and not so much the product...)

In order to identify and address the existing and potential visitors (almost-visitors), the following procedure is used:

1. In the first step, the (potential) visitor market is divided into segments

2. Then the segments are quantitatively evaluated and selected

3. Finally, proposals are developed to reach the segments in the best possible way

This marketing strategy answers two key questions: firstly, which visitors to "address" (segmenting and targeting) and how to create "value" for these segments (differentiation and positioning).

This process includes *segmenting*, *targeting* and *positioning,* and is known as the *STP strategy*.

Based on this, a target group-specific marketing mix is used to communicate the benefits to the previously defined target groups.

The meaning, purpose and benefits of STP marketing

STP puts the customer at the centre. Without STP marketing campaigns would be too generic and fail to capture the interest of consumers.

The STP model provides a solid formula to help identify the most valuable customer and meet customer expectations. The STP marketing model reap several benefits:

- **Personalised marketing:**
  It is simply impossible to target every person in the market through a marketing activity. The STP model allows you to narrow down who you are communicating with.

- **Segmentation:**
  Divides the target audience into different groups and adapts the communication to the needs of each group. This allows you to tailor your messages to specific groups of people, which would not be possible with an extremely large, heterogeneous target group.

- **Optimised marketing mix:**
  By segmenting your target group, you can focus on the channels that are used by your target group. Instead of spending time and money on channels that don't generate conversions, the STP model allows you to focus on the channels that are most likely to generate ROI.

- **More effective product innovation:**
  When you know your target audience, you can develop offers that best meet their needs. The STP model helps you find the right targets (also for research) so you can optimize and better market your product.

- **Identify the most lucrative customers and clients:**
  Not every person is the ideal customer. If your tired of chasing less promising visitors that don't bring visits - or don't spend much - the STP model can improve both the quality and quantity of customer lifetime value.

- **Reduce marketing costs:**
  Targeting a general audience is expensive and time-consuming - and usually doesn't result in many conversions. With the STP model, on the other hand, you can select the channels, products and messages that work best for their target audience. This focus saves you time and money, and also maximises your returns.

- **More accurate marketing data:**
  Audience segmentation provides reliable, relevant data about your ideal customers. You can make data-driven decisions that yield positive results.

- **Increase market share:**
  Customers respond positively to personalised, targeted messages and products. Altogether the STP strategy is to improve marketing performance, increase visitor numbers and satisfaction, and increase cost efficiency.

Segmentation

Segmentation means dividing the market into distinct groups

The „S" in STP stands for segmentation. Divide the market into different groups of customers.

- It is always tempting to target all segments at the same time, but unfortunately it is often not effective.

- It is better to identify the optimal segments for your objective and limit your marketing to that audience.

Structure of segmentations

Options for the selection of tabulation groups/features

Demographic characteristics:

- Age, gender, education, occupation....
- Place of residence

Other features
- Periods, time
- Other institutions etc.)

Behavioural descriptive characteristics:

- Visit frequency, duration,
- First/repeat visit, etc.
- Use of mediation services offers etc.

Attitudes, needs and motives:

- Visit motives
- attitudes
- satisfaction recommendation willingness to recommend
- Lifestyle typology

One can segment your audience based on several factors:

1. **Behaviour:** Visiting behaviour or purchase patterns - use data from market research, previous purchases, website visits, reviews etc. to segment your audience.

2. **Demographics:** Segment customers by age, education, occupation, economic status, gender, income, household size.

3. **Geography:** Group your clientele by country, state or city. Also, you can segment according to where they live in a city, suburb or rural area.

4. **Psychographics:** Psychographics are more difficult to quantify. They include attitudes, lifestyle and (music-)interests. Have become very important today.

While there are more ways to segment your audience than these four - such as by life stage, product related typologies etc. the four criteria listed above are the most commonly used.

Usually, you apply severall or all criteria one after the other:
Always start with 1. behaviour and then go to 2. demografics and then use 3. psychografics

You can always segment your audience according to your individual needs.

The SINUS Milieus as Consumer Typology

© Sinus Sociovision 2008
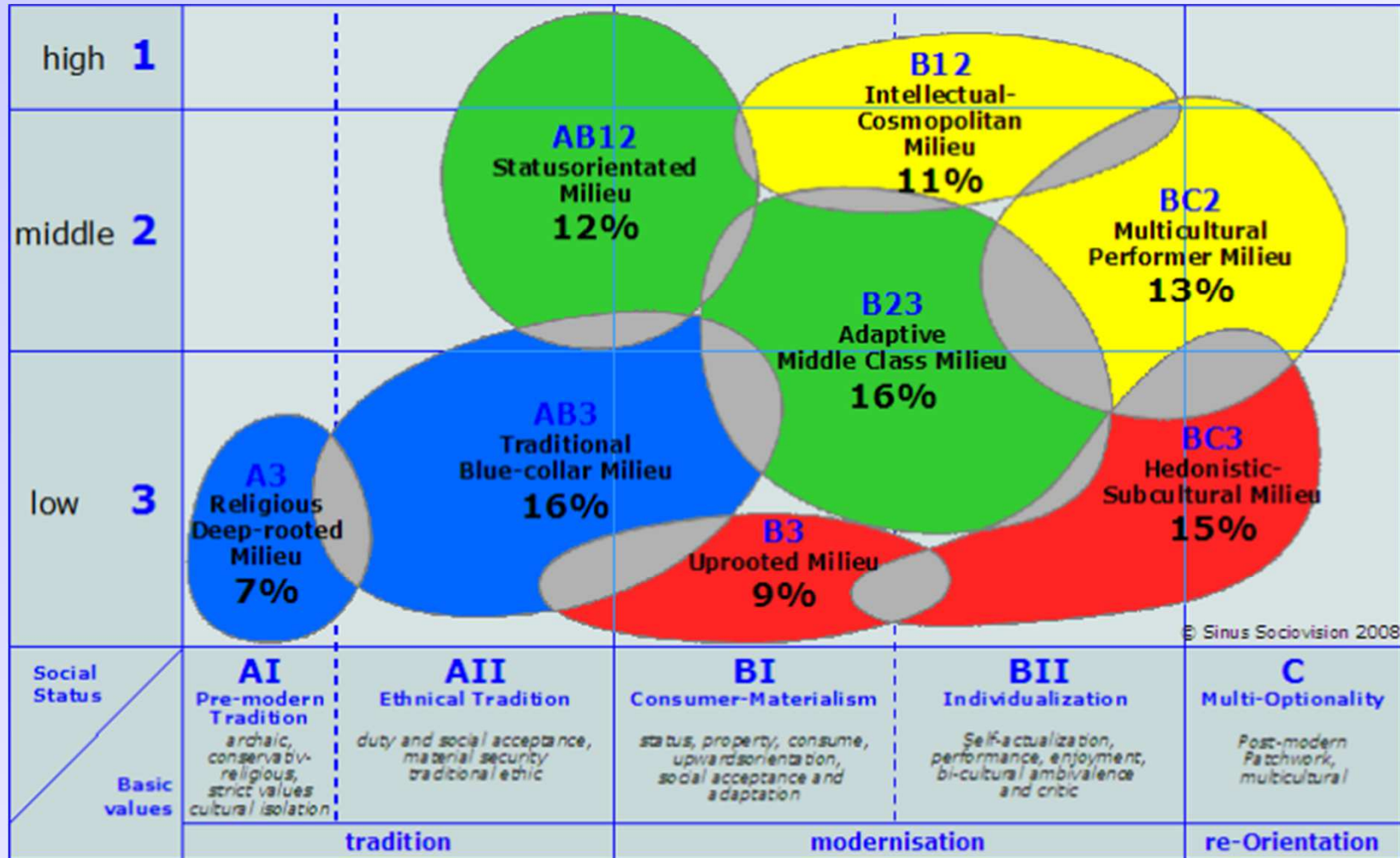
The „Limbic-Map" as a Consumer Typology

© Dr. Hans-Georg Häusel, Gruppe Nymphenburg Consult AG

Lifestyles in Berlin compared to the Federal Republic of Germany

According to Otte, Gunnar: Further Development of the Lifestyle Typology, Version 2019. Mainz.

Targeting

The „T" in STP stands for Targeting. Select from the market the most intesting groups of customers.

- The aim of targeting is to examine the existing segments and determine which of them are most likely to complete a purchase or make a visit.

- Limiting your targeting to only few segments at a time may seem to limit your options, but it allows you to measure your marketing efforts effectively.

- Furthermore, by choosing the right segments, you will increase interest, visits and satisfaction among your customers.

Formal requirements for the target group definition:

- Relevance of criteria, discriminatory power

- Internal Homogeneity

- External Heterogenity

- Size

- Money

- Availability

- Stability over time

- Accessibility

- Good product fit

Positioning

The "P" in STP stands for positioning. Positioning in this context refers to how your customers shall see your product.

Def. „Positioning":

All measures aimed at inserting a product in the subjective perception of the buyers in such a way that it stands out from the competing offers, and is preferred to them. (Kroeber-Riehl)

In this step of the STP model, you use the insights gained from segmentation and targeting to present your product to your chosen audience.

First, consider the occasion of positioning: Is it...

- a first positioning (new brand),
- actualisation of an existing positioning,
- re-positioning of an existing product, and
- positioning through actualisation

Pepsi-Cola

- At the time of the infamous competition between Pepsi-Cola and its competitor Coca-Cola in the 1980s, Pepsi used segmentation to target its ideal audience. The company identified the following three market segments:
    - People who were loyal to the Coca-Cola brand, people who liked Pepsi, but still preferred Coca-Cola, people who consumed both brands equally.
    - Pepsi focused on the third segment because these people were the most open-minded. When Coca-Cola launched "New Coke" in 1985 - Pepsi was there when the loyalty of Coca-Cola faithful wavered. Targeting this segment of formerly loyal Coca-Cola customers brought Pepsi a 14% increase in sales.

Apple

- Apple targeted a segment of affluent individuals interested in design, performance and luxury. The company marketed itself as a premium brand by using symbolic and experiential positioning. Apple' focused on innovation and exclusivity, creating a high-priced technology brand.
- Today, Apple users remain extremely loyal to the technology company and its highly valued brand.

McDonald's

- MCD targets middle- and low-income consumers. With slogans like "I'm Lovin' It" and "We Love To See You Smile", McDonald's positions itself as an open, family-friendly restaurant chain for people who want to consume fast food at low cost.
- As an international brand, McDonald's makes intensive use of STP. For example, McDonald's takes into account the local cuisine of a country and creates local food offerings based on the taste preferences of its target group, such as the Canadian national dish "Poutine" for Canada.

When doing a positioning, it is essential to take into account:

1. You start with the *product*
2. Then consider the *positioning object*
3. Then consider the competitors

Properties/desires of the target persons

Positions of the competing objects

Properties of the positioning object
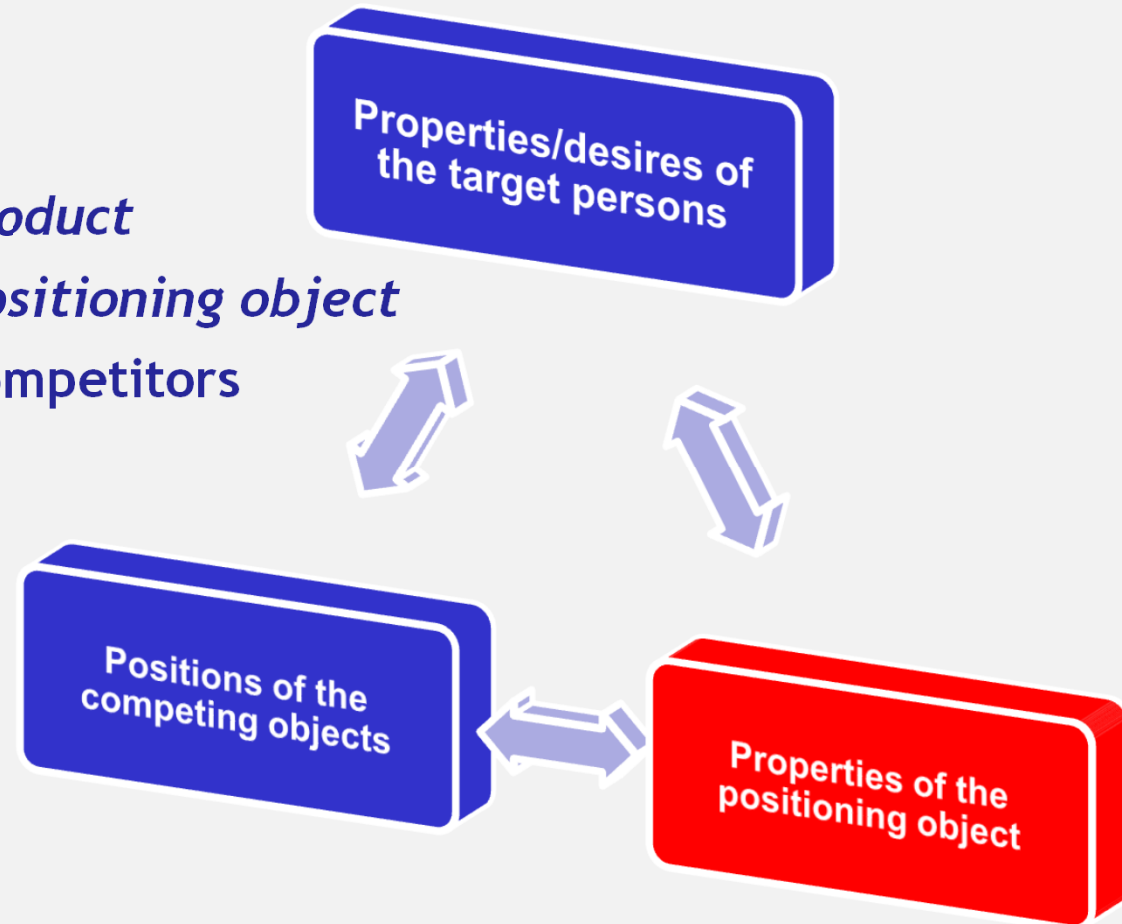
Central questions are very simple:

- What product/service does you offer?

- What is it, what can it do?

- Where can you get it? What does it cost?

- What is the product range/produc programme structure?

- Where are the best future prospects?

- Which offers have the highest unique selling proposition?

- Which product features are the most sustainable?

- Does the product/service have opportunities or problems?

- Essential to think „out of the box":
  What are other possible product features, that relate to other additional benefits ? Social, prestige etc. or other funcional benefits…

- How do you create create visitor value?

Criteria of successful positioning strategies:

1. Relevance of the feature
2. Concentration on a few features (max. 2!)
3. Discriminatory ability of the features
4. Durability and future orientation
5. Continuity of the position
6. Flexibility in appearance

The most common positioning mistakes:

1. Restriction to positioning too close to the product
2. Too objective positioning
3. Reactive planning
4. Lack of continuity in positioning
5. Lack of systematics in implementation

When positioning, the following must be taken into account:

Properties/desires of the target persons

Positions of the competing objects

Properties of the positioning object
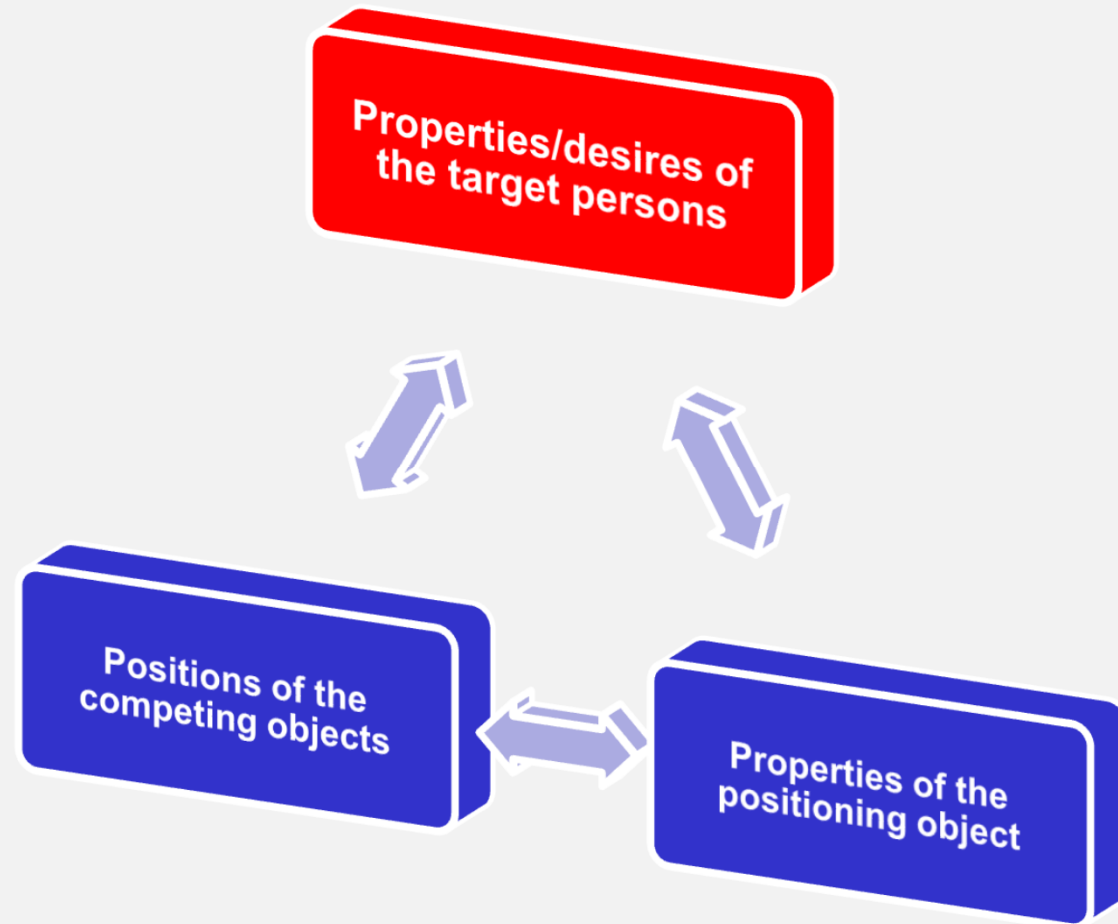
Important questions are:

- Could the target group properly be classified according to behaviour, demographic, psychologic criteria?

- What factual and emotional terms can be used to describe the target group (e.g. industry visitors or private visitors, professional audience or lay-people, trend-conscious or conservative etc.?

- Who are the decision-makers? Who decides for visits, with which activity, in which phases of the decision-making process (customer journey)?

- How and where can the target group best be reached?

- Want to form personas? Create detailed customer personas to understand your target audience better. Personas are fictional representations of your ideal customers, including their age, gender, interests, pain points, and buying behavior. This helps in crafting personalized marketing messages.

Example I: Positioning of a *running shoe* based on the means-end chain

Property	→	Benefit component	→	Value attitude

| Concrete Property (e.g.: with Heel support) | → | Abstract Property (e.g.: lie well at the foot) | → | Functional Benefit-component (e.g.: run faster) | → | Social or psychic Benefit-component (e.g.: I am relaxed after the running | → | Instrumental Value attitude (e.g.: I am physically fit) | → | Terminal Value attitude (e.g.: „Do sth. for my health") |
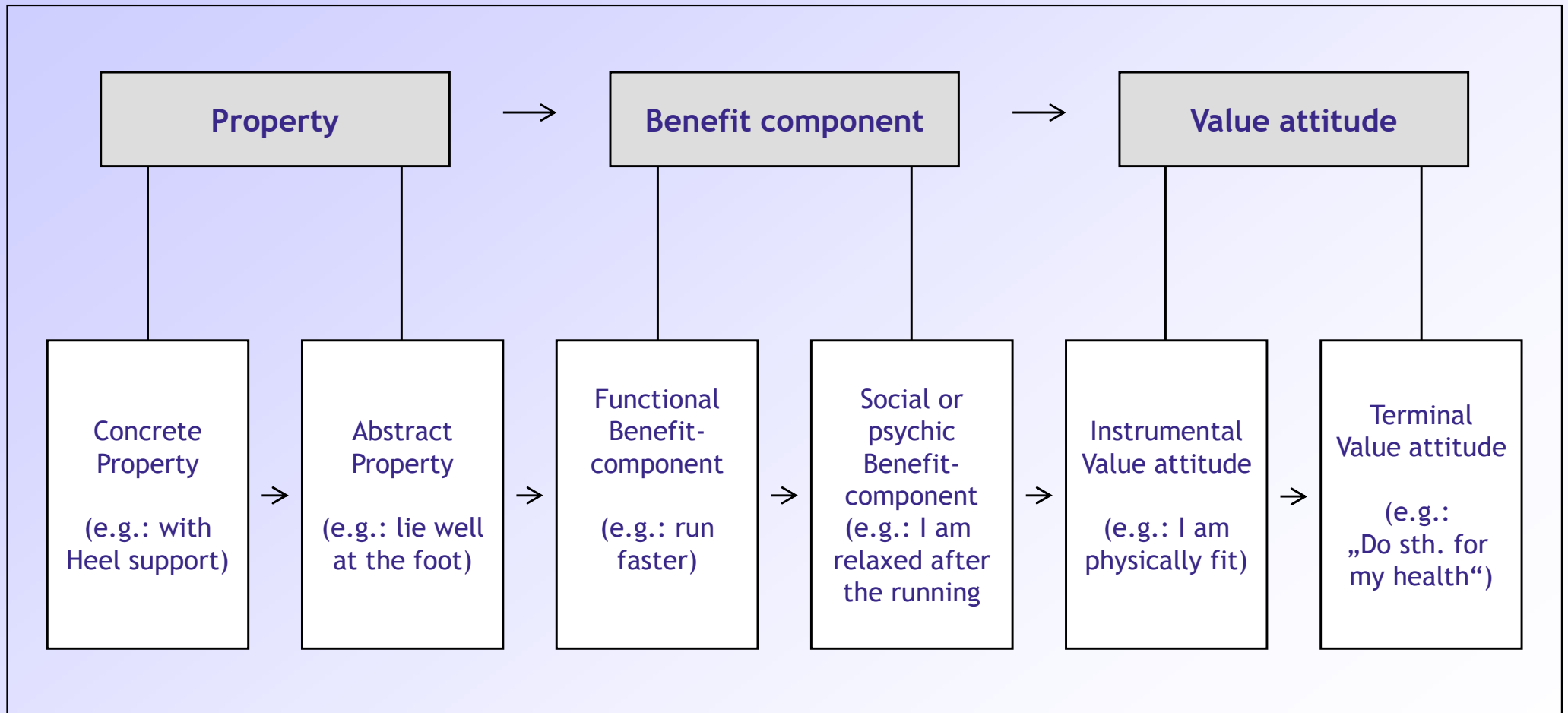
Example I: Positioning of a museum exhibition based on the means-end chain

Property	→	Benefit component	→	Value attitude

Concrete Property

(e.g.: "Numerous known images"

→

Abstract Property

(e.g.: "Fascinating, unique exhibition".

→

Functional Benefit-component

(e.g.: "Learn something new")

→

Social or psychic Benefit-component
(e.g.: "Can come together with others„, „entertainment with ambition")

→

Instrumental Value attitude

(e.g.: "Am culturally the height of the times")

→

Terminal Value attitude

(e.g.: "More Self-esteem")

Communicate your positioning to your customers!

- The positioning concept is abstract and only in the heads of the marketeers.
- It is necessary to make the positioning tangible and experiencably for the customer. Therefore a copy strategy is needed.

Copy Strategy

The copy strategy translates the abstract positioning concept into the actual advertising by which advertisers/creatives can operate. It is the specific customer approach. The copy strategy is defined for the medium for the long term. It consists of:

Consumer benefit

Reason why

Tonality
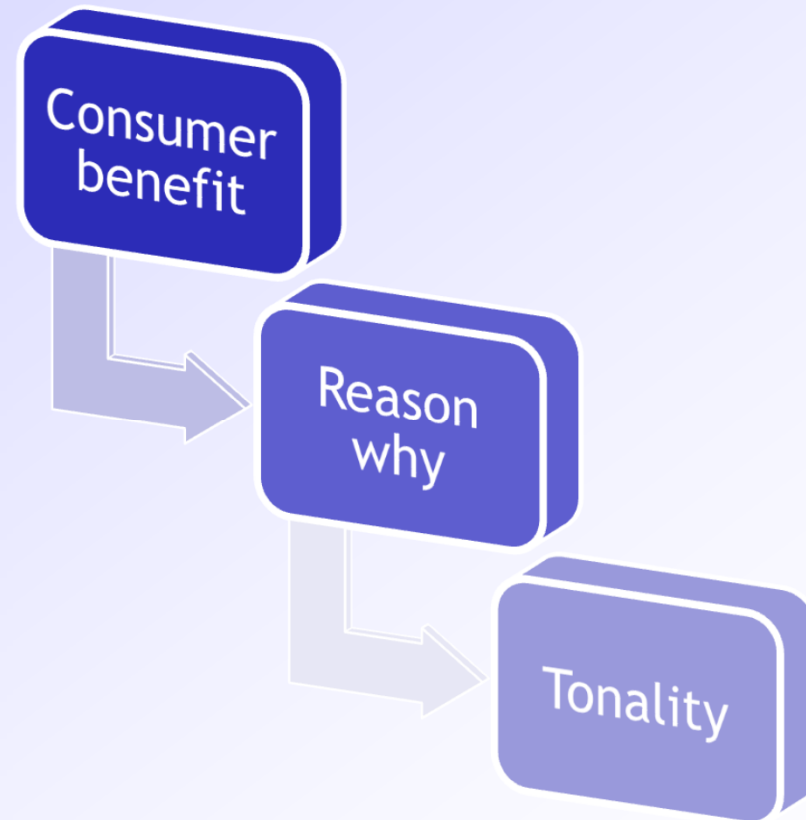
Consumer Benefit

The main value proposition. If the value proposition is truly unique, that is best: a unique selling proposition (USP) - but these are rare in today's markets with broad offerings
- Basic benefits, the basic benefit usually does not offer an approach to differentiation, only the additional benefit.
- Additional benefits (social, symbolic, aesthetic etc.)

USP / UAP

If the consumer benefit is an additional substantial benefit that is not offered by any other company, this is a real USP.

Since genuine USPs are rare, attempts are made to convey this claim to uniqueness in advertising → Unique Advertising Propostion

Demonstration possibilities of the consumer benefit (Ex.)
- Emotional orientation, creating worlds of experience
- Orientation towards mission statements/lifestyles
- Orientation towards the brand/product personality

"Reason Why"

- **Explanation and justification that the promised benefit actually exists. The Reason Why underpins the promise. It is to gain credibility.**

- **The benefit justification makes the advertising message more credible and strengthens the advertising effect.**

<u>Reasons are given by:</u>

    - Guarantee statement
    - Test results (e.g. Stiftung Warentest)
    - Special ingredients
    - Testimonials (celebrities)
    - Tradition of the festival
    - Special manufacturing processes
    - New research results
    - Special competence etc.
- ...

„Flair" or „Tonality" of the message

The flair determines the special atmosphere that should surround the central advertising message.

The tonality outlines the tone in which you want to address the target group: personal "you", rational, emotional, humorous, serious, conservative, discreet, loud / striking, etc.; you could decide here to address the customers humorously but discreetly and to use the first name.

A distinction is made between a psychological and a semiotic dimension:

Psychological dimension:
Conveys non-thematic information which is derived from the intended image. Central to this are the unconscious qualities of appeal:
- Evoke memories
- Arouse expectations
- Trigger emotions
- Create experiences

Semiotic dimension
The semiotic dimension is realised through verbal, visual, typographical and auditory stylistic devices.

Examples of different advertising styles:

- *A slice of life; ex.: Rahma family, dancing fans on a concert*
- *Lifestyle advertising; e.g.: Enjoyment of beer as a lifestyle in an upmarket ambience.*
- *Dream world; e.g.: O2 underwater world*
- *Mood pictures; e.g.: Marlboro - Country*
- *Consistent use of certain melodies (musical advertising); e.g.: T-Mobile melody*
- *Symbolic figures; e.g.: Master Propper, The White Knight*
- *Highlighting competence; e.g.: I as a dentist's wife…*
- *Highlighting lasting quality and lasting competence, e.g. VW Beetle: it runs and runs and runs....*
- *Scientific evidence; e.g.: Blend-a-med research*
- *Advertising with celebrities (testimonial advertising); e.g.: Manuel Neuer promotes shaving water*

The kind of purchase and the choice of an ad strategy strongly depends on customers involvement

➢Assumption: Involvement (concern, commitment, interest, ego-involvement) has a strong influence on stimulus processing. Involvement research tries to explain differences in retention and behaviour with the same stimuli.
People process messages in different ways depending on their involvement.
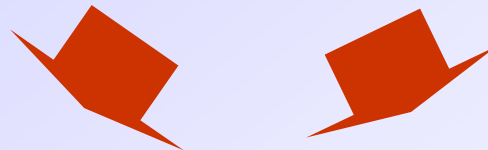
Highly involved individuals process information intensively, arguments are weighted up, information is compared with existing knowledge and subjected to critical evaluation (central route of information processing)

Low-involved individuals perceive the message only casually, and process information by orienting themselves to peripheral stimuli of the message (attractiveness of the communicator, number of elements, credibility of the source, etc.) (peripheral route of information processing)

Positioning approaches

a) Structural
- **Brand/product strategy**
- **Competitive**

b) Content positioning:
- **Emotion**
- **Involvement**

Kind of involvement

	Rational	**Emotional**
Low	Needs are trivial, low information interest, no emotions, product is knows, purchase by habit, very high ad pressure ⬇ **Communication Design: „Actualisation"** z.B. Dole-Bananas	Needs are trivial, low-risk, pure emotional-oriented purchase for self-reward, external conditions important, high ad pressure ⬇ **Communication Design: „Emotional"** z.B. Ads for *Milka chocolate*
High	Riskoriented, rationally guided purchase, less campaign pressure needed ⬇ **Communication Design: „Informative"** z.B. Life-insurances, stocks, credit card	Needs exist, can be adressed, high information interest, images are important, less campaign pressure needed ⬇ **Communication Design: „Emotional and informative"** z.B. *Thomas Cook* travels

Strength of involvement

3

	High Involvement Communication	Low Involvement Communication
Advertising aim	Convince	Impress
Content of message	Everything that is important	Say something
Lenght of message	Exhaustively	Short
Change of attitude	By rational arguments	By affective aspects
Medium of communication	Language	Music, pictures sound
Frequency	Rare	Frequent
Timing	Moment before purchase	Always

Effects of the level of involvement on consumer behaviour

High-involvement characteristic	Low-involvement characteristic
•Active information search	•Passive information intake
•Active engagement	•"Let it pass"
•High processing depth	•Low processing depth
•Less persuasion ("sovereign consumer")	•High persuasion ("secret seduction")
•Comparative evaluation before purchase	•Valuation at most after purchase
•Attention to many features	•Attention to a few features
•Few acceptable alternatives	•Many acceptable alternatives
•Much social influence	•Little social influence
•Goal "Optimisation	•Target "no problems"
•Brand loyalty through conviction	•Brand loyalty through habit
•Strongly anchored, intense attitude	•Low anchored, flat
•High memory capacity	•Low memory capacity

When positioning, the following must be taken into account:

Properties/desires of the target persons

Positions of the competing objects

Properties of the positioning object
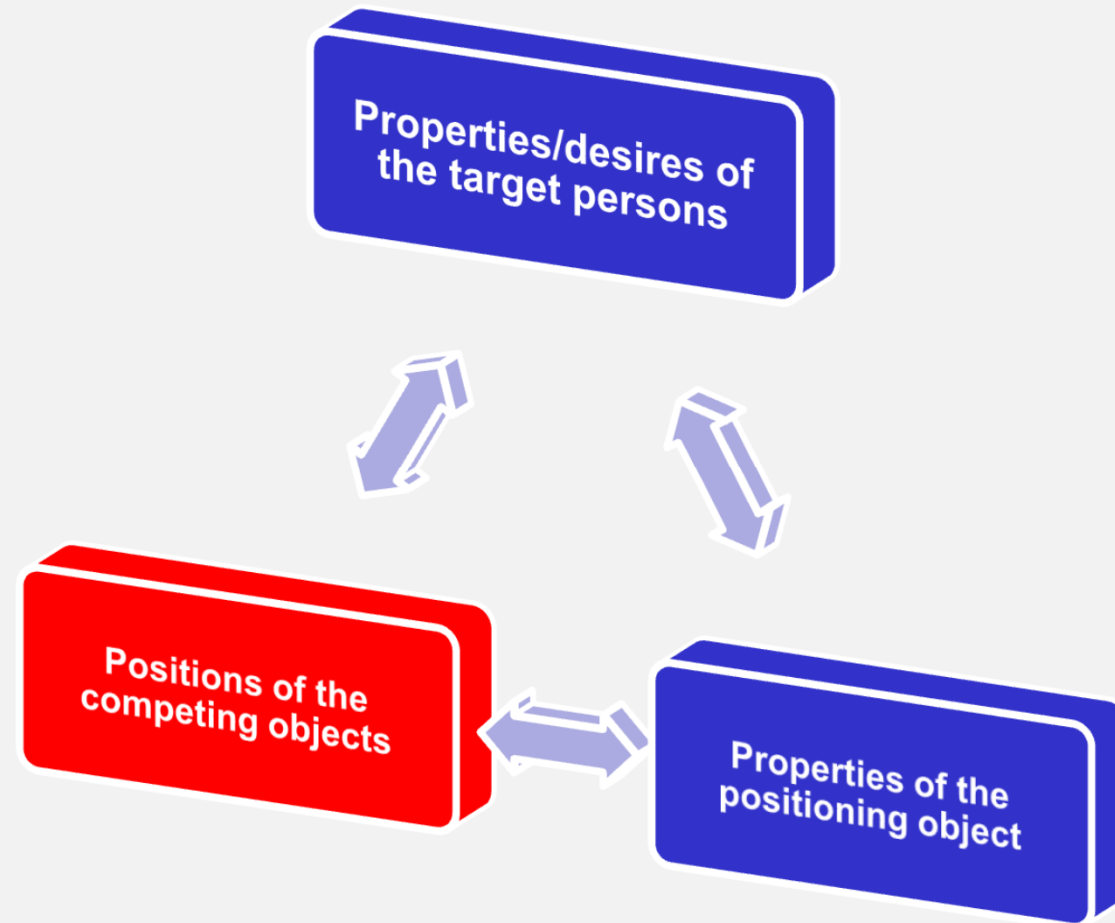
Core questions at this step are:

Description of the competitive situation

- Which competitors are in direct comparison? Focus on most important competitors!
- How does the competition try to distinguish and position itself?
- What brand promises does it make?
- What are important competitors outside your own branch?

Description of the market

- What is the market situation?
- How is the market structured?
- What dynamics, developments and long-term trends are discernible?

Classic positioning strategies:

1. **Positioning in a "large" segment (may mean positioning in the same space as the competition)**

2. **Positioning at the interface of several segments**

3. **Positioning in a market niche (manifest/latent)**

4. **Unique position on a dimension outside the previous perceptual space**
   The aim is to occupy a new dimension of characteristics in a unique way, which is unknown to the customer and important for his purchase decision (so-called "genuine comparative competitive advantage")
   Ex.: APPLE, IKEA, MAGNUM ice cream
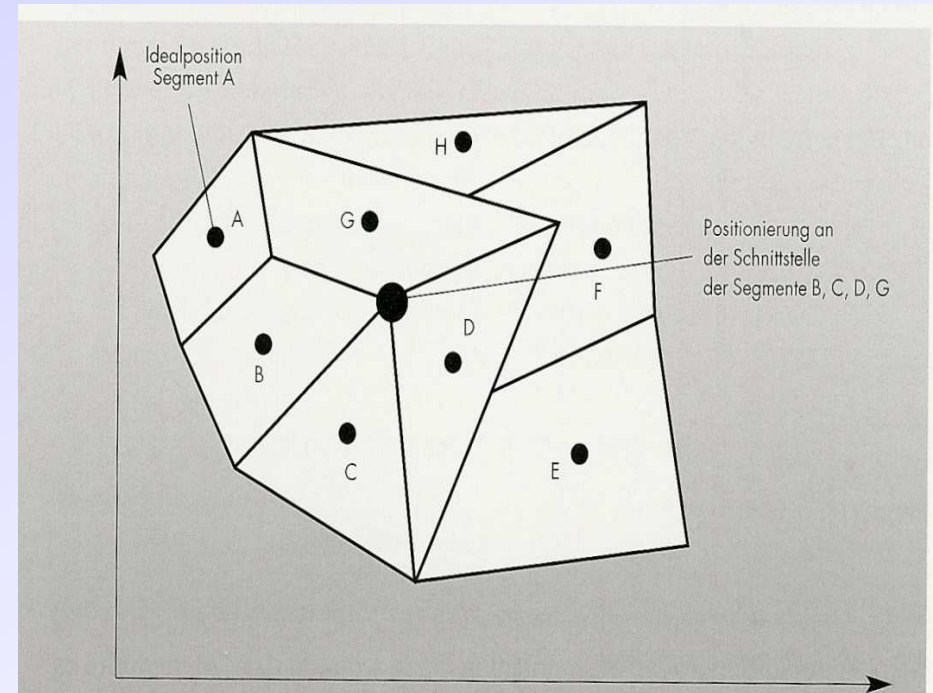
Copy - Analysis

The copy analysis is intended to convey which positions the competitors are striving for the market and which copy strategies they are pursuing on their advertising media ("brand identity").

Take a look at their copy-strategies by looking at their advertising!

Core components of a copy analysis are:

- Consumer Benefit
- Reason Why
- Tonality
- Target group

The following aspects are particularly interesting:

- Headline
- subline
- Claim (benefit) and slogan (concise advertising slogan)
- Key visuals
- CD constants (colour, logo, typo, etc.)

Example: STP Modell for the independent theatre scene in Cologne

In the following, the relevant target groups for the independent theatre scene in Cologne will be identified according to the STP method, evaluated according to attractiveness and a positioning in terms of content will be proposed. Subsequently, the possibilities for addressing the defined target groups through the media will be explored.

In concrete terms, the procedure is as follows:

1. Definition of two target groups : creating a combination of the characteristics of interest in visiting and knowledge of independent theatres. Evaluation of the resulting segments.

2. Description of the selected target groups according to their demographic structure, their cultural leisure behaviour, perceived barriers to visiting and possible incentives to visit.

3. Selection of media and communication channels.

4. Local location of the target groups in the Cologne city area.

5. Suggestions for the design of the offer or communicative approach.

1. target group selection and evaluation (in % of the population aged 18 and over in Cologne)

Table percentages: All percentages in the table add up to 100%. Basis: 908,000 Cologne residents aged 18 and over		Interest in visiting the independent theatres		
		very	something	less / not at all
Knowledge of independent theatres	Know them from a personal visit	27% "Repeat Visitor"	21% "Repeat visitors"	5% "Non-more-visitors"
	Heard of them, but not yet been there	3% "Almost-visitor	12% "Almost-visitor	9% "Non-visitor"
	Don't know her at all	3% "Almost-visitor	4% "Almost-visitor	16% "Non-visitor

Potentials: Repeat visitors: 48% or 435.000 people as existing visitors
       Almost-visitors: 22% or 200.000 persons additional potential
       Non-(more) visitors: 30% or 263.000 persons as non-visitors

➢ **The greatest potential is repeat visitors with an interest in visiting, who are relatively easy to attract on the basis of visits given in the past.**
➢ **The next largest target group is the almost-visitors with a maximum of 200.000 visitors.**
➢ **Non-more or non-visitors are eliminated as potential with approx. 30%.**

Visitor potential according to Cologne districts (Basis: All respondents, in %)

	Chorweiler	Innenstadt	Nippes	Kalk	Ehrenfeld	Porz	Lindenthal	Rodenkirchen	Mülheim
Non-(more) visitors	33%	26%	29%	34%	29%	41%	25%	27%	31%
Almost-visitor	25%	22%	19%	22%	16%	19%	24%	21%	26%
(Repeat) Visitors	42%	53%	51%	44%	55%	40%	51%	52%	44%

■ (Repeat) Visitors　■ Almost-visitor　■ Non-(more) visitors
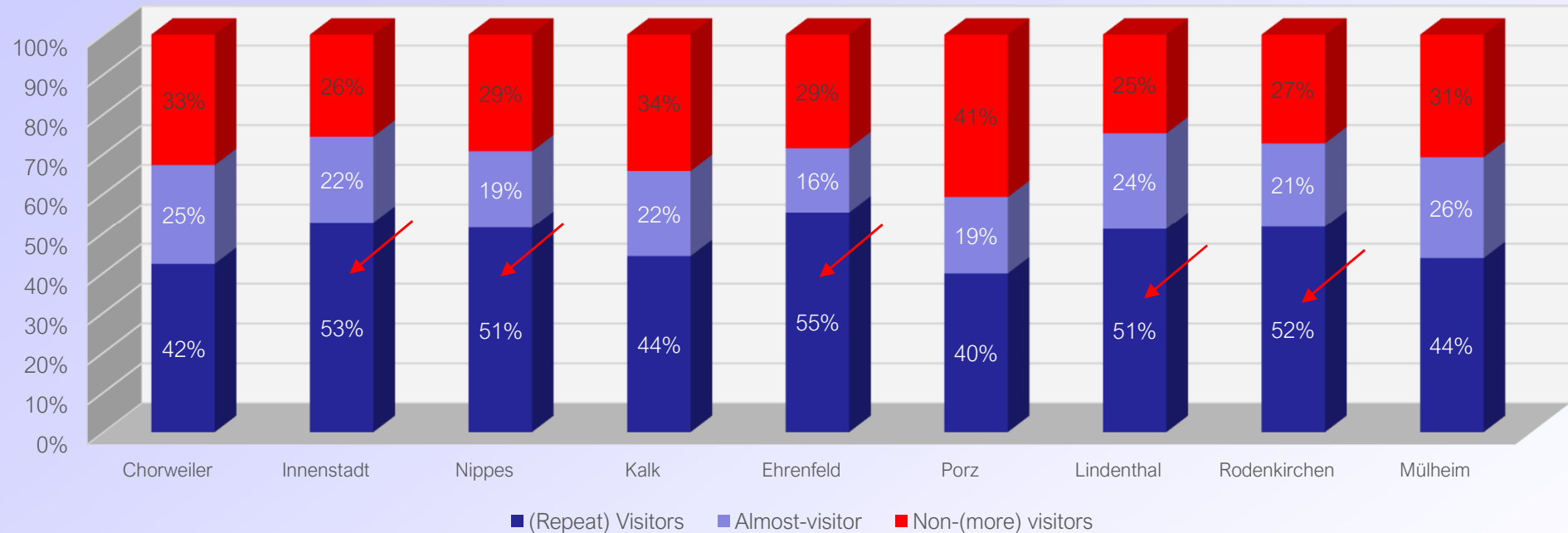
- The illustration shows the partly quite different distribution of the target group potential among the different districts. The most interesting districts are Ehrenfeld and Innenstadt, followed by Rodenkirchen, Lindenthal and Nippes. Here the share of (repeat) visitors is the largest. At the same time, the potential of almost-visitors is already relatively limited here. Porz, Chorweiler and Kalk, on the other hand, are at the lower end of the spectrum. The potential of almost-visitors is greatest in Mülheim. Chorweiler is only of limited significance due to low case numbers.

- The potential analysis thus follows the structures shown so far: these are the same districts that are ahead in terms of awareness of the independent theatres and the visits made there. These districts are also likely to have the most visitors to the independent scene in the future.
  In addition, these districts are particularly suitable for locally oriented advertising activities, such as stationary outdoor advertising or the display of printed advertising materials in suitable locations.

Target group selection

The target group analysis identifies as central target groups the...

1. "Repeat visitors" to the independent scene: These already know the scene from their own experience and document an interest in visiting here. This group makes up about 48% of the adult population in Cologne. In absolute numbers, this is about 430,000 people. Added to this are the …

2. "Almost-visitors" to the independent scene: They are characterised by the fact that they only know the independent providers from the media, from hearsay or not at all, but nevertheless have an interest in visiting them. This group comprises 22% of Cologne's population or about 200,000 people. This is a considerable, as yet untapped potential.

The strategies for the two groups differ in that the loyalty of repeat visitors to the theatres should be intensified and the frequency of visits increased ("retention and loyalty marketing"). This seems all the more important as the group is unusually large with 48% on the one hand, and on the other hand the loyalty of these visitors to the independent providers is in part relatively weak.

In the case of almost-visitors, the aim is to lead them to the independent venues or groups for the first time through suitable measures (attraction marketing). There, the offer must be convincing in itself.

People who are not interested in visiting the independent theatres - regardless of whether they already know them or not - are excluded as target groups. The chance of winning over this group seems too small for numerous reasons (poorer image of the independent scene, lower expectations, no recognisable incentives, etc.).

Target group description (in % of subgroups)

Note: Characteristics with an asterisk are statistically significant for distinguishing the target groups		Repeat visitors	Almost-visitor	Non-visitors
Education*	Until secondary school leaving certificate	29%	**48%**	55%
	A-Levels/Fachabitur	25%	29%	23%
	Studies (Univ./FH)	**45%**	23%	22%
Gender	Male	47%	40%	55%
	Female	53%	**60%**	45%
Age (mean value in years)		**48,4**	**43,3**	48,2
Visits to the Städt. Theatres in the last 12 months*		1,84	0,83	0,53
Theatre visits several times a year		48%	21%	11%
Willing to pay for a ticket at the independent theatre		**37,5 €**	31,1 €	-
Frequency of visits (at least several times/year)	Opera	**13%**	8%	4%
	Museum	**60%**	**42%**	25%
	Classical concerts	32%	18%	14%
	Rock-Pop Concerts	29%	36%	22%
Images of the Free Th.	Flagship for the city*	**91%**	55%	30%
(agree fully/mostly)	Indispensable for culture in Cologne*.	**94%**	55%	31%
	More important than opera and drama*	**59%**	32%	16%
	Important for the integration of cultures in Cologne*.	**87%**	53%	29%
	I like them.	**94%**	55%	28%

- The group of repeat visitors tends to be more highly educated (academically), somewhat older, active theatre-goers per year and attends the Städt. Theatres. The independent providers enjoy an outstanding image here. Repeat visitors are generally very culturally active.
- The almost-visitors, on the other hand, have a significantly lower level of education and are younger. Women predominate. They also rate independent theatres positively, but are more distanced than repeat visitors. Their willingness to pay is somewhat lower. They also have an affinity for culture, although they are somewhat less involved in high-cultural activities.
- In principle, visitors and almost-visitors are structurally largely similar.

Positioning I

Note: Characteristics with an asterisk are statistically significant for distinguishing the target groups	Repeat visitors	Almost-visitor	Non-visitors
Sector interests (Shares for "*I would visit*")			
• **Cabaret/Comedy**	81%	81%	-%
• **Classical spoken theatre**	76%	60%	-%
• Dance	60%	57%	-%
• Modern, experimental theatre	55%	47%	-%
• Folk or dialect theatre	63%	**65%**	-%
• Children- Youth Theatre	51%	48%	-%
• **Variety, Circus, Artistry**	71%	**75%**	-%
Obstacles to <u>theatre attendance</u> in general (Percentages for "*fully/mostly agree*")			
• **Would prefer to entertain/relax when going out*.**	**40%**	**56%**	69%
• Going to the theatre is exhausting*	12%	**24%**	35%
• Acquaintances would look at me funny*	6%	14%	17%
• No accompaniment to the visit*	19%	25%	27%
• Financial reasons*	23%	30%	32%
• Must plan theatre visit elaborately*	23%	**36%**	31%
• **Too little free time***	**36%**	**47%**	34%
• Inconvenient playing times*	30%	**41%**	31%

- Repeat visitors have a wide range of interests, with cabaret and classical theatre leading the way. Theatre. The interests of the almost-visitors are similar to those of the repeat visitors, but on the whole more conventional. Modern, experimental theatre and classical spoken theatre are less popular, but folk theatre, dialect theatre and variety theatre are more popular. Cabaret/comedy is at the top of the list. But there is also greater interest in dance.
- The almost-visitors in particular want entertainment and relaxation more often when going out. They also experience more time problems compared to going to the theatre and find it more time-consuming to organise.
- Conclusion: The positioning and approach can be the same for both groups, with a stronger emphasis on the entertainment moments, the comfort in the theatre and organisational facilitation when visiting the theatre for the almost-visitors.

RED Re-Educating Digitisation

Rheinische Hochschule Köln

Positioning II

Note: Characteristics with an asterisk are statistically significant for distinguishing the target groups	Repeat visitors	Almost-visitor	Non-visitors
Obstacles to attending an independent theatre *(Percentages for "strongly agree/highly agree")*			
• **Performances are boring***	10%	**33%**	
• **Obtaining information is cumbersome***	**19%**	**32%**	29%
• No offers in my language/culture*	5%	17%	15%
• Have had bad experiences there	8%	-	-
• Performances are incomprehensible*	9%	19%	26%
• Programme offer is unattractive*.	14%	29%	42%
• Artistic level is not good*	10%	19%	23%
• **I'd rather go to the Schauspiel or the Cologne Opera*.**	**22%**	**38%**	22%
• Unattractive ambience*	12%	21%	28%
• Atmosphere is stiff and formal*	9%	18%	23%
• Poor comfort and service there*	13%	19%	19%
• **Theatre ticket is expensive***	**25%**	**35%**	44%
• Ticket purchase is cumbersome*	14%	**26%**	19%
Incentives to visit *(proportions for "very/quite important")*			
• **More stay comfort**	32%	**44%**	0%
• **More entertaining performances**	**44%**	**58%**	0%
• More comprehensible performances	31%	**43%**	0%
• **More demanding performances**	**36%**	46%	0%
• **Easier ticket purchase**	**33%**	43%	0%
• Barrier-free access	**38%**	39%	0%

The mentions here are primarily image-driven

- Repeat visitors cite few obstacles overall. The almost-visitors, on the other hand, more often consider the performances of independent theatres to be boring (33%) and their programme offerings to be unattractive (29%). They consider obtaining tickets and information to be cumbersome (32%), and a third of respondents find tickets expensive. Almost 40% also prefer to go to the theatre or the opera.
- When it comes to incentives to visit, visitors most often mention more entertaining performances, followed by more aspirations. Ticket purchase is also inconvenient for 1/3. Among the almost-visitors, more comfort of stay is in front, followed by easier ticket purchases. They want more popular performances that are more entertaining, more comprehensible and more demanding. The almost-visitors are definitely interested in culture, but favour the Städt. Theatres (38%).

Pricing strategy

Pricing strategy

The pricing strategy is closely related to all aspect of segmentation, targeting and positioning. In the following some basic strategies will be presented.

1. **Cost Analysis:** Begin by understanding your costs, including production, overhead, marketing, and distribution expenses. Foundation for setting a profitable price point (*cost-plus pricing*).

2. **Market:** Analyze **competitor pricing** as well to gauge where you fit in the market.

3. **Value Proposition:** Align your pricing with your value proposition. If your product or service offers unique features, exceptional quality, or solves a specific problem, you may be able to justify a higher price.

4. **Market Research:** Research your target market to understand the *price sensitivity* of your customers. Are they willing to pay a premium for your product or service, or are they price-conscious?

5. **Pricing Strategies:** Consider different pricing strategies, such as *value-based pricing, penetration pricing, skimming pricing*. Choose the strategy that best suits your product, market, and business goals.

7. **Positioning:** Pricing can be a powerful tool for positioning your brand. A higher price may convey exclusivity and quality, while a lower price might emphasize affordability.

8. **Segmentation:** If your business serves multiple customer segments, consider implementing *price differentiation*. Tailor your pricing to match the willingness to pay of each segment.

9. **Psychological Pricing:** Experiment with pricing psychology techniques, such as charm pricing (ending prices with "9" or "99"), bundling, or tiered pricing. These can influence customers' perception of value.

10. **Discounts and Promotions:** Plan your discount and promotion strategy carefully. Discounts can attract new customers and boost sales, but they should not erode your profit margins.

11. **Dynamic Pricing:** For e-commerce and other dynamic markets, consider implementing dynamic pricing algorithms that adjust prices in real-time based on demand and supply factors.

Price Elasticity

- Price elasticity is a **measure of how sensitive the quantity demanded or supplied of a good or service is to its price**.

- It is calculated by dividing the percentage change in quantity demanded or supplied by the percentage change in price

- **Elastic demand** is when the quantity demanded of a service changes significantly as the price of service changes. Inelastic demand is when the quantity demanded of a service changes little as the price of the changes.

- The formula for calculating the price elasticity of demand is given by:

Price Elasticity = <u>Percentage Change in Quantity Demanded</u> = <u>($\Delta Q/Q$) x 100</u>
Percentage Change in Price ($\Delta P/P$) x 100

- Elastic demand: Luxury items, such as expensive cars, jewelry, and designer clothes. Inelastic demand: Necessities such as food, water, and electricity.

- If price elasticity is **greater than 1**, the good is elastic; **if less than 1**, it is inelastic.

- If price elasticity is exactly 1 price change leads to an equal percentage change in demand, it is known as unitary elasticity.

- ➤ **Try to find the areas of inelastic demand in your demand curve!**

Price-Demand Curve
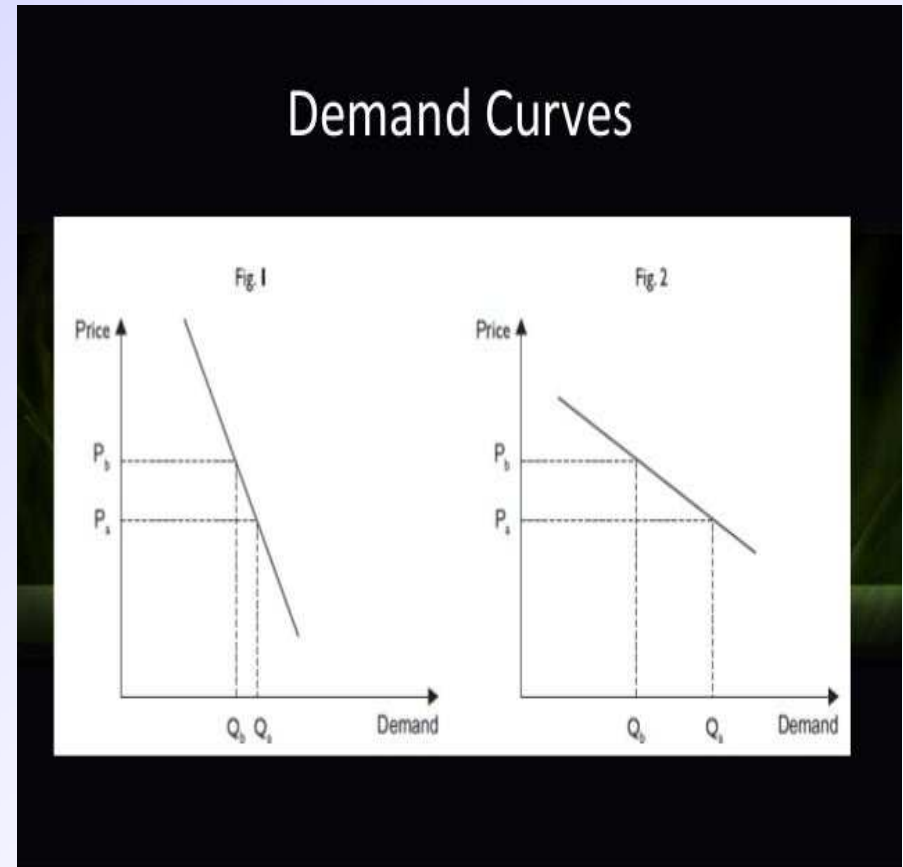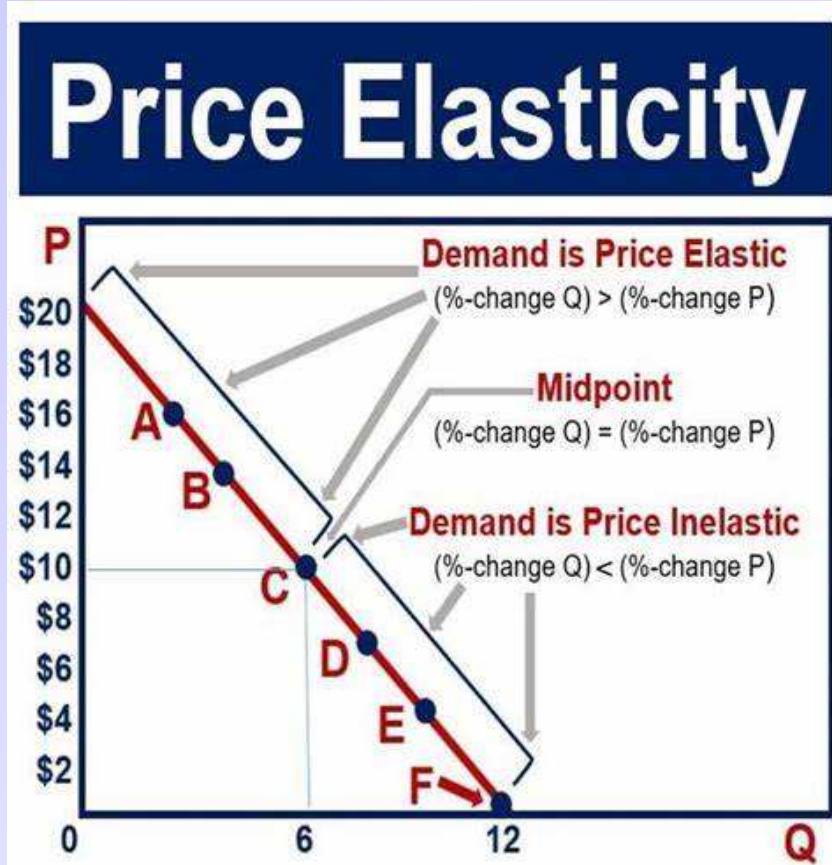
Willingness to pay for a ticket

% of respondents

100 | 99
97
88% would accept a price of max. 15 €.
88
Reduction in price by 15 € (from 25 € to 10 € resp. 60%) doubles demand from 40% to 85%
72
60
Die Besuchsinteressierten würden im Mittel bis zu 35 € pro Ticket bezahlen. Dies wären ca. 35%.Hochgerechnet wären dies maximal ca. 200Tsd. Personen bzw. ein Umsatz von 7,5 Mio. €
41
36
26
24
Changes in price do not entail changes in turnover
13 13
Increse in price incease turnover
8
6   6
4
3
1
0

<= 5 | 6 - 10 | 11 - 15 | 16 - 20 | 21 - 25 | 26 - 30 | 31 - 35 | 36 - 40 | 41 - 45 | 46 - 50 | 51 - 55 | 56 - 60 | 66 - 70 | 71 - 75 | 76 - 80 | 86 - 90 | 96 - 100 | 101+

€

The analysis of the willingness to pay of the respondents with an interest in visiting an independent theatre reveals a wide range, from about 5 € to about 100 € for a single ticket.

Special features become apparent in the process:

- In the lower price range up to about 28 € per ticket, the price sensitivity of visitors is very high. Lowering ticket prices can lead to a disproportionate increase in demand and thus to an increase in turnover.
- In the middle range of about 30 - 45 €, price and demand changes are neutral to each other. Here, price variations do not mean any changes in turnover.
- On the other hand, the price sensitivity of visitors is low from about 50 €. Price increases hardly have a negative effect on demand. Here - in combination with certain additional services - more expensive tickets could also be sold.
- Overall, a price differentiation of tickets seems to have a clear revenue-increasing effect ("ceteris paribus").

Zahlungsbereitschaft * Alter (Jahre) Crosstabulation

% within Alter (Jahre)

		Alter (Jahre)				Total
		18 - 29	30 - 44	45 - 59	60 +	
Zahlungsbereitschaft	bis 15 €	15%	13%	10%	10%	12%
	16-25 €	31%	27%	29%	22%	27%
	26-35 €	20%	23%	20%	27%	23%
	36-50 €	22%	21%	24%	21%	22%
	51+ €	13%	16%	17%	21%	17%
Total		100%	100%	100%	100%	100%

Zahlungsbereitschaft * Höchster Bildungsabschluss Crosstabulation

% within Höchster Bildungsabschluss

		Höchster Bildungsabschluss			
		bis weiterf. Schule ohne Abitur	Abitur/Fachabitur	Studium	Total
Zahlungsbereitschaft	bis 15 €	11%	15%	11%	12%
	16-25 €	23%	26%	30%	27%
	26-35 €	24%	19%	24%	22%
	36-50 €	19%	24%	23%	22%
	51+ €	24%	15%	12%	17%
Total		100%	100%	100%	100%

Zahlungsbereitschaft * Besuchshäufigkeit der freien Theater Crosstabulation

% within Besuchshäufigkeit der freien Theater

		Besuchshäufigkeit der freien Theater				Total
		kein Besuch	selten (1 - 2 Besuche)	häufiger (3 -4 Besuche)	regelmäßig (5+ Besuche)	
Zahlungsbereitschaft	bis 15 €	14%	9%	7%	8%	11%
	16-25 €	27%	25%	24%	38%	26%
	26-35 €	24%	25%	15%	8%	23%
	36-50 €	19%	26%	29%	15%	22%
	51+ €	17%	15%	24%	31%	17%
Total		100%	100%	100%	100%	100%

Investigating into prices by market research

Let´s build a Demand Curve from the Barcelona Survey!
How to:

- Remove missing values, if not done before. Make a frequency command (with ascending values) in SPSS and move the table into Excel.

- Remove rows not needed from the table, keep the row *„values"* and „cumulative percent"

- Make a calculation and create a new column (example):
      100 - Column = SUMME(100-G85) Adjust axes

- Add add other elements (formula, r2) diagramm element, select curve, Select a formula, that covers the graph best; try different curves, if possible select a linear formular/graph

- Think, where the price with maximung turnover my be situated

- Calculate turnover: Form first and second derivation of formula or insert values

- Think about price differentiation

- Find out how elastic the demand is as the price of the good or service changes.

Customer Segmentation by Life-Styles

What are „Life-Styles"?

Def.: ***Behaviour patterns*** *can be identified as lifestyles if they form a* ***coherent context****, if they are of* ***central importance*** *in a person's life and if they are* ***stable over time.*** (Otte 2015)

- Accordingly, lifestyles are primarily determined by **a person's behaviour,** and less by their value orientations, attitudes or preferences.
- However, they are an **expression of a person's inner values** or basic orientations.
- They are cross-situational evaluation schemes of an aesthetic and ethical nature.
- The focus is on forms of expression motivated by **everyday aesthetics**:
	- A high-cultural lifestyle manifests itself in frequent visits to art exhibitions, theatre performances and readings, art exhibitions, theatre performances and readings.
	- Lifestyles can also be shaped by **everyday ethics**, such as religious beliefs, political ideologies or ecological values.
- Distinction from social milieus

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

What a difference Life-Styles make

To put it bluntly: Why socio-demographics alone are only of limited help

**Example -
Socio-demographic twins:**

- Both come from Great Britain
- They are the same age
- Both have children
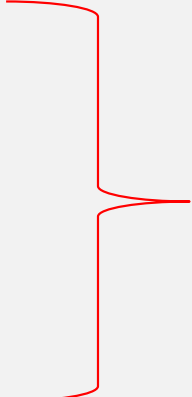- Both are divorced
- Both are very wealthy

King Charles and Ozzy Osbourne

Not the same target group because they have completely different lifestyles
Older people can have a lifestyle like young people and vice versa

Lifestyle research emerged in western societies in the 1970s and 1980s through:

– Massive expansion of wealth,
– Educational expansion,
– Increase in social mobility
– increased social security,
– Liberalisation of values and norms, and
– the increase in leisure time, and
– Increasingly competitive markets

Individual choices have increased, post-materialistic development values have gained in importance.

As a result, lifestyle research took off in western industrialised countries, and in some cases it was hyped.

Specific drivers of the development were:

• Applied market research with practical research needs and problems
• The increasing problem of not being able to group target groups sufficiently according to demographic characteristics
• In academia: Pierre Bourdieu's French sociology with the localisation of lifestyles between education and social situation became very popular

Today, there is an almost **unmanageable number of concepts** in the market of lifestyle typologies.

These either take a **holistic approach to** describing a person's life completion, such as:

- The SINUS Milieus of the SINUS Institute Heidelberg
- GfK's Euro Socio-Styles
- The semiometrics of KANTAR
- The LIMBIC Map from Häusel
- **The lifestyle typology by Gunnar Otte**

Or focus on specific aspects in certain areas of consumption or behaviour, such as:

- Age-life-cycle models
- The LoHaS model
- Media use (ARD/ZDF media user typology)
- Consumer behaviour in specific areas such as fashion and beauty, food, alcohol, cars, leisure, travel, investments, etc. ...

Unfortunately, there is no proven specific typology for cultural behaviour or even attending movies/concerts yet!

The integrative lifestyle system according to Gunnar Otte:

"Lifestyle" as a stable, regularly recurring pattern of everyday living.

Consisting of value orientations, attitudes, taste preferences and actions that are related to each other, and are mostly behaviour related.

Structural lifestyle factors: Age (periodic, generational and life cycle effects), gender, social situation (educational level, occupational position, disposable income) are also related to lifestyles.

Otte's lifestyle typology is...

- In contrast to the common approaches in lifestyle research, they are openly accessible, developed transparently and can be replicated with little survey effort,

- not designed in an exploratory way, but provided with a fixed assignment grid,

- provides a synopsis of numerous approaches to lifestyle and values research (Schulze, Bourdieu, particularly similar to SINUS Milieus),

- is not specifically focused on the culture user, but segments the population as a whole,

- Is theoretically sound.

Otte distinguishes two dimensions in his lifestyle classification, which in turn are broken down into two further sub-dimensions:

a) The first dimension can be called the *level of endowment* and breaks down into the sub-dimensions of *material standard of living* (=economic sub-dimension) and *cultural aspiration level* (=cultural sub-dimension):

- Within the *material standard of living,* lifestyles differ when considering the generosity of **housing,** the frequency and destination of **holiday trips**, the ownership of **luxury objects** (antiques, jewellery, works of art), expensive or fancy **clothes,** and the **attendance of upscale restaurants** and **exclusive events.**

- Within the *level of cultural aspiration,* lifestyles manifest themselves in the **frequency of reception of highly cultural events and works**, in a **cosmopolitan horizon of interests,** and in the **reading-based acquisition of connoisseurship and discourse skills** (e.g. with regard to culture, history, politics, science, technology).

b) The second dimension reflects the *time orientation of the lifestyle* (dimension of time) and the biographical perspective (sub-dimension of lifetime).

- In the **sub-dimension of modernity,** people vary in their willingness to **adopt** cultural **innovations and new fashions** or to **value** the **"classical" (e.**g. culture, clothing, technology) and to orient their lifestyle towards new values or old **traditions.**

- In the **sub-dimension of the biographical perspective,** people differ in their **openness to experience, desire to explore** and **search for variety on the** one hand, and the closedness of their horizon of experience on the other.

➢ **Modern forms are biographically open, innovative and experience-oriented; semi-modern forms are biographically consolidated and more strongly characterised by everyday routines; traditional forms of living are established and closed.**

What can I do with the lifestyle typology?

- It gives me a **deeper understanding of** festival attendance than socio-demographic characteristics alone can provide;

- I can **identify target groups and recognise what** visitors like and what attracts them.

- As a creative or cultural marketer, I can optimally **gear** my **offers to** these lifestyle groups

- I can **better** empathise with the lifestyles of my lifestyle groups, e.g. develop and address "personas" of my visitors on a lifestyle basis.

- I can get tips for **more effective advertising:** To whom do I direct my communication, on which channels?

- A **comparison of the distribution of lifestyles** in my audience with the distribution in the city's population provides indications of which groups are represented below average

- I get recommendations for the **content and design of the advertisement**: what advertising messages work, what the tonality should be, etc.

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

But:

Every lifestyle typology, including Otte's, is basically (only) a segmentation technique to distinguish between relevant and non-relevant target groups!

This includes the following requirements :

1. Relevance for the connection between the service and the target group typology
2. Internal **homogeneity** and external heterogeneity of grouping
3. **Discriminatory power** of the target group segmentation
4. **Findability** of the target groups in the data and **media accessibility**
5. **Concreteness** of the target group description
6. Temporal and geographical **stability** of the division
7. **Communicability of the results** and a certain level of **awareness** in the "market".

-> **In the case of Otte's typology, these requirements can be considered to be met overall**

Lifestyles in Berlin compared to the Federal Republic of Germany

		←Modernity degree →		
		Traditional/ biographical closure	**Partly modern/ Biographical consolidation**	**Modern/ Biographical openness**
↑ Equipment level ↓	**Upscale**	*Conservative-Upscales* Berlin 2019 5% **FRG 2018 5%**	*Liberal-Upscales* Berlin 2019 10% **FRG 2018 11%**	*Innovative-Reflexives* Berlin 2019 11% **FRG 2018 7%**
	Medium	*Conventionalists* **Berlin 2019 9%** **FRG 2018 12%**	*Upward mobile-middle class* **Berlin 2019 17%** **FRG 2018 17%**	*Hedonists* **Berlin 2019 10%** **FRG 2018 7%**
	Low	*Down to earth traditionals* **Berlin 2019 12%** **FRG 2018 19%**	*Home-centered* **Berlin 2019 17%** **FRG 2018 18%**	*Entertainment seakers* Berlin 2019 8% **FRG 2018 5%**

According to Otte, Gunnar: Further Development of the Lifestyle Typology, Version 2019. Mainz.

Lifestyles in Berlin compared to the Federal Republic of Germany

		←Modernity degree →		
		Traditional/ biographical closure	Partly modern/ Biographical consolidation	Modern/ Biographical openness
← Equipment level →	Upscale			
	Medium			
	Low			

According to Otte, Gunnar: Further Development of the Lifestyle Typology, Version 2019. Mainz.

The content description of the lifestyle system - Values for Berlin

Equipment Geho-ben	***Conservative-Upscale (5%)*** Tradition of the propertied bourgeoisie, conservatism, distinction through "rank", exclusivity in standard of living, classical high culture, willingness to perform and lead, religiosity.	***Liberal-Upscales (11%)*** In the tradition of the educated middle classes, liberality, professional self-realisation, high culture consumption with an "alternative" touch, a sense of authenticity, connoisseurship in consumption.	***Innovative-Reflexives (7%)*** Cultural, academic avant-garde, creativity and joy of experimentation, search for personality development, global attitude to life.
Medium	***Conventionalists (9%)*** Tradition of the petty bourgeoisie, values of duty and acceptance, security orientation, consumption of high culture with a folk touch, conservative-religious morality, domestic idyll.	***Upward mobile-middle class (17%)*** Centredness around solid professional career, family and participation in the mainstream of modern leisure culture, "averageness" and internal heterogeneity of the type through middle position.	***Hedonists (10%)*** Youth cultural style protest through fashion and music, innovative spirit, present-oriented pleasure and consumption, extraversion, urban spectacle and club culture.
Low	***Down-to-earth traditional (19%)*** Tradition of skilled work, modesty, orientation towards the practical, importance of social security, closeness to trade unions, German songs, club life.	***Home-centred (18%)*** Family-centredness and domesticity due to children and low resource availability, traditional folk festival scene and modern mass culture such as pop music and television.	***Entertainment seekers (5%)*** Experience consumption, materialistic status symbolism and out-of-home entertainment orientation, threat of declassification.
Mo-dernity	**Traditional/biographical closure**	**Partial modern/biographical consolidation**	**Modern/biographical openness**

According to Otte, Gunnar: Further Development of the Lifestyle Typology, Version 2019. Mainz.

Liberal-Upscales (11%)

In the tradition of the educated middle classes, liberality, professional self-realisation, high culture consumption with an "alternative" touch, a sense of authenticity, connoisseurship in consumption.
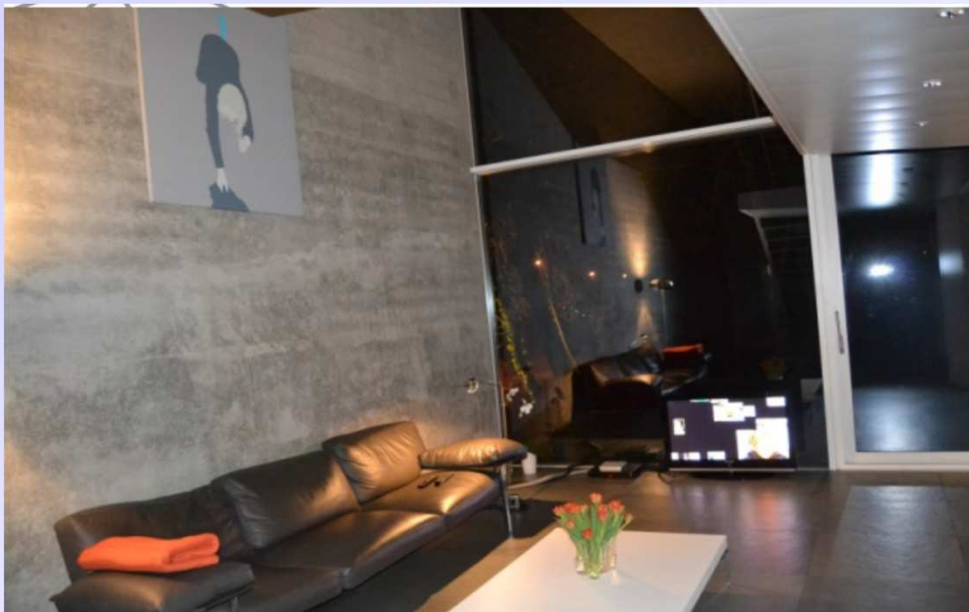
Innovative-Reflexive (7%)

Cultural, academic avantgarde, creativity and joy of experimentation, search for personality development, global attitude to life.

Co-funded by
the European Union

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

Upward mobile-middle class (17%)

Centredness around solid professional career, family and participation in the mainstream of modern leisure culture, "averageness" and internal heterogeneity of the type through middle position.

Conservative-Upscale (5%)

Tradition of the propertied bourgeoisie, conservatism, distinction through "rank", exclusivity in standard of living, classical high culture, willingness to perform and lead, religiosity.

Hedonists (7%)

Youth cultural style protest through fashion and music, innovative spirit, present-oriented pleasure and consumption, extraversion, urban spectacle and club culture.

Mostly less interesting target groups!

Conventionalists (12%)

Tradition of the petty bourgeoisie, values of duty and acceptance, security orientation, consumption of high culture with a folk touch, conservative-religious morality, domestic idyll.

Down-to-earth traditionalists (19%)

Tradition of skilled work, modesty, orientation towards the practical, importance of social security, closeness to trade unions, German songs, club life.

Home-centered (18%)

Family-centredness and domesticity due to children and low resource availability, traditional folk festival scene and modern mass culture such as pop music and television.

Entertainment seekers (5%)
Experience consumption, materialistic status symbolism and out-of-home entertainment orientation, threat of declassification.

Equipment Upscale	Conservative Upscale (5%)	Liberal Upscale (10%)	Innovative-Reflexive (15%)
	– **Theatre and drama 48 %** – **Classical concerts 45 %** – **Opera 29 %** – **Natural history/technical exp. 53 %** – **Historical. 66 %** – **Art exhibition 59 %** – **Memorial sites 52 %**	– **Theatre and drama 55 %** – **Classical concerts 47 %** – Opera/Ballet/Dance 29 % – **Natural history/technical exp. 62 %** – **Historical 66 %** – **Art exhibition 65 %** – **Memorial sites 72 %**	– **Theatre and drama 61 %** – **Classical concerts 57 %** – **Opera 38 %** – **Natural history/technical out t. 77 %** – **Historical 81 %** – **Art exhibition 78 %** – **Memorial sites 71 %**
Medium	Conventionalists (9%)	Upwardly mob. midd. class (17%)	Hedonists (10%)
	– Theatre and drama 34 % – Classical concerts 29 % – Opera 15 % – **Natural history/techn. exp. 46 %** – Historical Museums 44 % – Art exhibitions 27 % – **Memorial sites 58 %**	– **Theatre and drama 44 %** – **Classical concerts 32 %** – **Opera 14 %** – **Natural history/technical exp. 59 %** – **Historical. Museums 54 %** – **Art exhibitions 53 %** – **Memorial sites 66 %**	– **Theatre and drama 55 %** – Classical concerts 40 % – **Opera/Ballet/Dance 20** – Natural history/techn. exp. 64 % – **Historical Museums 58 %** – **Art exhibitions 58 %** – **Memorials 68%**
Low	Down-to-earth trad. (12%)	Home-centred (13%)	Entertainment seekers (8%)
	– Theatre and drama 18 % – Classical concerts 13 % – Opera 4 % – **Natural History/techn. Exp. 40 %** – Cultural-historical exhibits 38 % – Art exhibitions 24 % – **Memorial sites 48 %**	– Theatre and drama 23 % – Classical concerts 11 % – Opera/Ballet/Dance 8 % – **Natural History/Techn. Exst. 40 %** – Historical 34 % – Art exhibitions 28 % – **Memorial sites 53 %**	– Theatre and drama 31 % – Classical concerts 21 % – Opera 9 % – **Natural hist./techn. exp. 51 %** – Historical 38 % – **Art exhibitions 40 %** – **Memorial sites 36 %**
Mo-dernity	Traditional/biographical closure	Partial modern/biographical consolidation	Modern/biographical openness

Visit class. Cultural offerings in Germany (% of visitors within one year)
According to Otte, Gunnar: Further Development of the Lifestyle Typology, Version 2019. Mainz.

Finally: A typology is not an end in itself. The only decisive factor is whether or not it works well as a segmentation tool in a specific application!

What you should bear in mind when working with this typology:

1. Also tit is no *"One-Size-Fits-All"* - Model

2. Check whether the typology really **segments well for** your specific question/establishment/visitors**! Otherwise, choose other models if necessary** (according to age groups, education, motivation types, customer value, etc.).

3. Note that **the individual lifestyle types** can be **heterogeneous** or **hybrid within themselves.** Check whether your lifestyle target groups are sufficiently homogeneous.

4. Check whether these are **sufficiently large.** Refrain from analysing and processing **target group segments that are** too **small.**

5. When selecting your target group, be aware that the **boundaries** between the different types are **blurred in** reality. Therefore, do not be too selective.

6. **Do not** use **several typologies at the same time** (e.g. Otte and Falk's motivation typology).

7. Make sure you can **communicate** the **results to your addressees.**

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

END OF SECTION !

RED
Non-Visitor Research II

- Media Planning –

Author:
Prof. Dr. Tibor Kliment

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

In general, select media for your campaign that...

- have a good reach within the target group,

- do cause little scattering losses,

- have a moderate price,

- have a good reputation,

- can convey your message appropriately,

- have strong ad impact!

How can I do this?

Key Performance Indicators I:
Assessing *reach* of the media used in a campaign

Assessing „*Net Reach*" (NR) of channels - Number of visitors reached at least once by all media in the campaign or by specific media

- **NR of the *media campaign* among visitors:**
  NR = (Visitors reached by all media / Total of all visitors) x 100
  -> You can calculate this by questionaires and later data analysis

- **NR of *specific media* among visitors:**
  NR = (Total reach of one medium / total of all visitors) x 100
  -> You can calculate this by appropriate questions and data nalysis

- ➤ **Example SYS Film Festival Survey...**

18. Sources of awareness: **Personal recommendation**

	Frequency	Percent
Doesn't apply	74	58
Applies	54	**42**
Total	128	100

18. Sources of awarness: **SYS Website**

	Frequency	Percent
Doesn't apply	109	85
Applies	19	**15**
Total	128	100

18. Sources of awareness: **SYS newsletter**

	Frequency	Percent
Doesn't apply	121	95
Applies	7	**5**
Total	128	100

18. Sources of awareness: **SYS Social media**

	Frequency	Percent
Doesn't apply	91	71
Applies	37	**29**
Total	128	100

18. Sources of awareness: **Newspaper, magazine**

	Frequency	Percent
Doesn't apply	113	88
Applies	15	**12**
Total	128	100

18. How did you become aware of the SYS Festival <u>this year</u>?

Multiplel answers possible!

O$_{18.1}$ Personal recommendation (friends, acquaintances, colleagues)
O$_{18.2}$ SYS Website
O$_{18.3}$ SYS festival newsletter
O$_{18.4}$ SYS Social media channels
O$_{18.5}$ Newspaper, magazines
O$_{18.6}$ Online magazine
O$_{18.7}$ Other, namely:_____

Figures obtained from a visitor survey

Key Performance Indicators II:
Assessing *cost-effectiveness* of the media used in the campaign

Assessing cost-effectiveness: „*Thousand User Price*" (TUP) or by „*Thousand Contact Price*" (TCP).
Means, how much it costs to reach 1.000 visitors or gain 1.000 contacs
–> shows cost-effectiveness of channels and makes channels comparable

- **Assessing cost-effectiveness of all media in the campaign:**

  TUP = (Cost for all ads / total of all visitors reached) x 1.000
  TCP = (Cost for all ads / total of all contacts by the visitors) x 1.000
         -> Except in digital media, numbers of contacts are more difficult to assess

- **Assessing cost-effectiveness of specific media:**

  TUP = (Cost for ads of a channel / total of all visitors reached by the channel) x 1.000
  TCP = (Cost for all ads of a channel / total of all contacts reached by the channel) x 1.000

- ➢ **Example with SYS survey...**

RED
Re-Educating
Digitisation

Rheinische
Hochschule
Köln

18. Sources of awareness: Personal recommendation

		Frequency	Percent
Valid	Doesn't apply	74	58
	Applies	54	**42**
	Total	128	100

18. Sources of awarness: SYS Website

		Frequency	Percent
Valid	Doesn't apply	109	85
	Applies	19	**15**
	Total	128	100

18. Sources of awareness: SYS newsletter

		Frequency	Percent
Valid	Doesn't apply	121	95
	Applies	7	**5**
	Total	128	100

18. Sources of awareness: SYS Social media

		Frequency	Percent
Valid	Doesn't apply	91	71
	Applies	37	**29**
	Total	128	100

18. Sources of awareness: Newspaper, magazine

		Frequency	Percent
Valid	Doesn't apply	113	88
	Applies	15	**12**
	Total	128	100

18. How did you become aware of the SYS Festival <u>this year</u>?

Multiplel answers possible!

O$_{18.1}$ Personal recommendation (friends, acquaintances, colleagues)
O$_{18.2}$ SYS Website
O$_{18.3}$ SYS festival newsletter
O$_{18.4}$ SYS Social media channels
O$_{18.5}$ Newspaper, magazines
O$_{18.6}$ Online magazine
O$_{18.7}$ Other, namely:_____

Example for calculation TUP :
***SYS Film Festival* has had 10.000 Visitors in total**

- Newspaper:
  (3.500 € / 1.200) x 1.000 = 2.916,- €
- SYS – Website:
  (8.000 € / 1.500) x 1.000 = 5.333,- €
- SYS Newsletter:
  (2.000 € / 500) x 1.000 = 4.000,- €
- Social Media:
  (4.000 € / 2.900) x 1.000 = 1.379,- €

Key Performance Indicators III
„Affinity" (AF) of the media used in the campaign

Indication of what percentage of the medium's user base belongs to the target group

- **It tells you how efficient it is to reach your target group by a medium**

- **The *share of a target group within the user group* of a medium is the „affinity."**

    AF = (absolute reach in target group / abs. reach in total) x 100

On the other hand, affinity shows the scattering losses:

    100 – AF = Scattering losses

- **Affinity helps you to calculate media cost-efficiency more precisely!**

    Example:

Example: Affinity (AF) of media to be used in a campaign

Information media on cultural matters: Local daily press

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Rarely/never	241	24	31	31
	Occasionally	278	28	36	67
	Frequently	257	26	**33**	100
	Total	776	78	100	

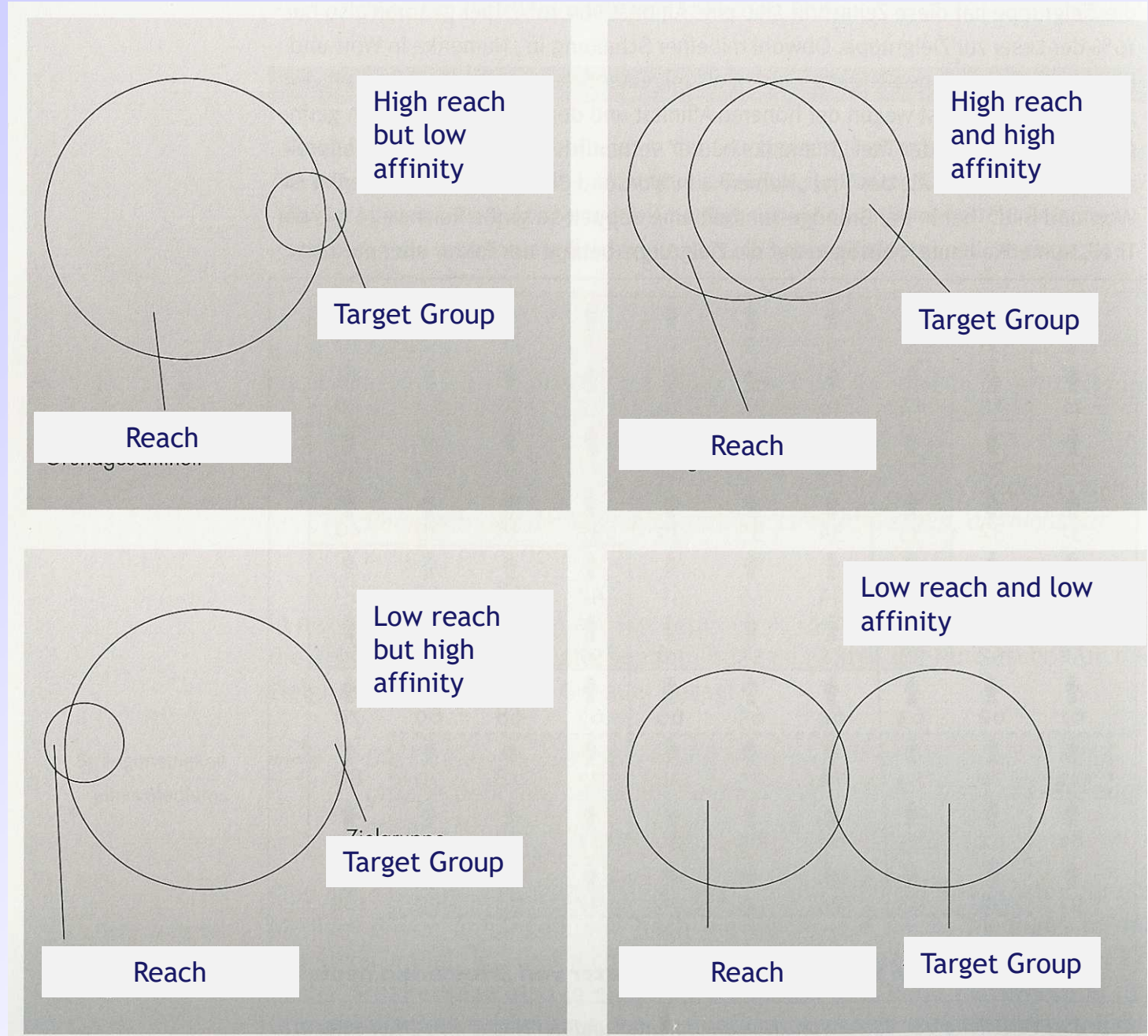Age * Information media on cultural matters: Local daily press Crosstabulation

% of Total

Example: A local daily paper in Cologne has 300.000 (frequently) readers.
A midsize ad in this daily costs 8.000 € and has a TUP of 40 €.

Let´s assume, your target group is only to 1/3 in this readership
(for example people who are interested in culture aktivities)

Result: The TUP for your targets is tripled from 40 € to 120 €!

Taken together:

Reach and affinity of an advertising medium

High reach but low affinity

Target Group

Reach

High reach and high affinity

Target Group

Reach

Low reach but high affinity

Target Group

Reach

Low reach and low affinity

Reach

Target Group

Key Performance Indicators IV: „Contact Density"

Ad pressure or „Contact Density": Measures strenght of ad impact
Tells you, how many contacts with the campaign the target group has had

1. *„Average Contact"* (AC):

   AC = Number of *contacts* with the campaign / Net Reach

   Example: The campaign yields 20.000 contacts, 5.000 people were reached
   AC = 20.000 / 5.000 = 4

2. *„Gross rating Points"* (GRP):

   GRP = AC x Net reach in target group in Percent

   Example: A campaign yields 4 AC and will reach of 70 % in target group
   GRP = 4 x 70% = 280

Net Reach, AC and GRP determine the impact and effectiveness of a campaign.

When planning the campaign, first you have to decide what contact density and what reach you want to have in your target group:

Put emphasis on reach!

• Usually, only 1-3 contacts per person are sufficient
• Cover as much as possible of your target group

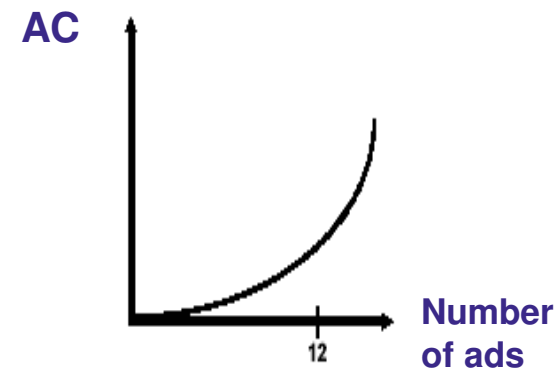Be aware that reach and frequency develop in opposite directions ->

To maximize reach, use as many different media as possible, but place only few ads in every medium.

When you want 2 contacts per visitor, then start by placing 2 ads per medium.

Example

NR

Number of ads
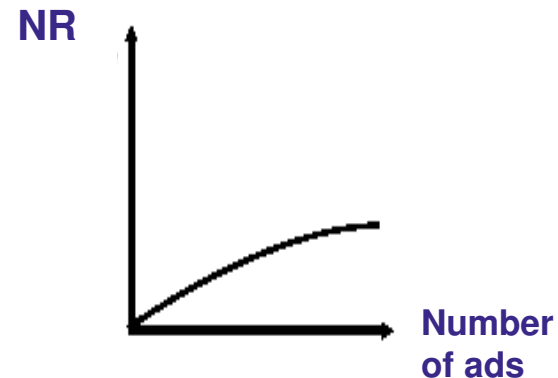
First: The more ads, the more NR

AC

Number of ads

12

Later: The more ads, the more AC

Optimizing you media selection: Calculate figures and compare „*media quality*"

- **Calculate and compare the *Net Reach*, *Thousand User Price* and *Affinity* of all media you take into account of the campaign!**

- **Make a ranking of the most cost efficient media!**

- **Compare: Are there other channels available that are more cost-efficient?**

Quality: Consider „qualitative factors" that determine ad impact:

- **Image of the medium (credibility),**
- **possibilities to create emotions and deliver information**
- **acceptance of ads in the medium by the user,**
- **intensity of contact,**
- **fit of the editorial environment**

Overview of features of different media channels:

Overview - steps for building a media plan

1. Clearly define your campaign's objectives.

2. Determine frequency (numbers of contacts/GRP) and reach: Minimum-maximum, AC (-> learning task, situation analysis, budget). Usually prefer reach to frequency (2-3 contacts).

3. Selection of all media: OoH advertising, locals, magazines, flyers, social, newsletter, search, Radio

4. Asses performance (reach, density, affinity) and costs. Are you still in the budget? If necessary develop alternatives.

5. Decide the ad distribution over time: Try to concentrate ad pressure.
For festivals a backloading campaign parallel to ticket sales usually works best.

6. Create a campaign that maintains consistent messages across all channels. Ensure synergy between online and offline.

7. Use KPIs to measure effiency of each media channel:
Monitor metrics such as Visits, PI´s, click-through rates, conversions and the data that come from market research (NR, AC, TUP) .

8. Evaluate your campaign by visitor research

Finally: Don't just rely on paid advertising!

Use the following possibilities too:

- Media partnerships (gain free ads or organic native ads)
- Sponsorships for advertising
- Try to develop organic reach, by good content quality, user activity, and search engine optimisation
- Although organic reach is often seen as free, it still requires a strategy and continuous effort.
- You may use influencer
- Fuel the Word of Mouth by giving incentives

Budget planning

In general, there are different methods of calculation a budget that lead to different results:

Corporate approaches

Advertising budget as a *percentage of specific reference variables (turnover, profit, sales volume etc.)*

> *For example choose 10% of past turnover for campaigning*

Advertising budget in absolut figures *is constant* for one period after the other

Competition Method

Budget is oriented towards the competition within the same branche
Benchmark is Share of Advertising and Share of Market.

Reference is main competitor or average of total market
Possibilities:
- SOA>SOM (aggressive competition)
- SOA=SOM (neutral)
- SOA<SOM (defensive)

Target-task method

- Orientation towards the *goals to be achieved* by advertising
- Measures are derived from advertising goals
- Calculation: reach x contacts x cost of occupancy

Example: You want to reach 25% of the target group twice (2 AC). In the target group, 100 Radio GRP cost around € 0.02 million.
Calculation: 25% NR x 4 AC = 75 GRP -> (75 GRP x 0.02 mio.) / 100 = 0,015 million

When calculating the budget, it´s best to consider all criteria:

- ✓ **Recipient-specific:**
    - ▪ **Is ad pressure strong enough?**
    - ▪ **Enough reach in your target groups?**

- ✓ **Target-specific:**
    - ▪ **Is complexity of adverting goal compatible to your budget?**
    - ▪ **Appropriate choice of media?**

- ✓ **Corporate-specific:**
    - ▪ **Is the budget reasonable in relation to turnover (and profit?) and size of your institution?**

- ✓ **Competitive:**
    - ▪ **Is your budget at least in parity with your competitors?**

- ✓ **Market-specific:**
    - ▪ **Is budget big enough for buying all the media needed?**
        - ➢ **If you can always say „yes", budget is optimal!**

END OF SECTION !