

Übung zur Vorlesung  
**Informationsvisualisierung**

Emanuel von Zezschwitz  
Ludwig-Maximilians-Universität München  
Wintersemester 2012/2013

# **Solution**

## **Exercise 2**

# Exercise 2 - 1 a)

- What's a heatmap?
  - Visualizes frequencies/size of a value using a color palette (metaphor: “cold” to “warm”)
  - Often, colors are calculated based on the maximum and minimum in the dataset

# Exercise 2 - 1 b)

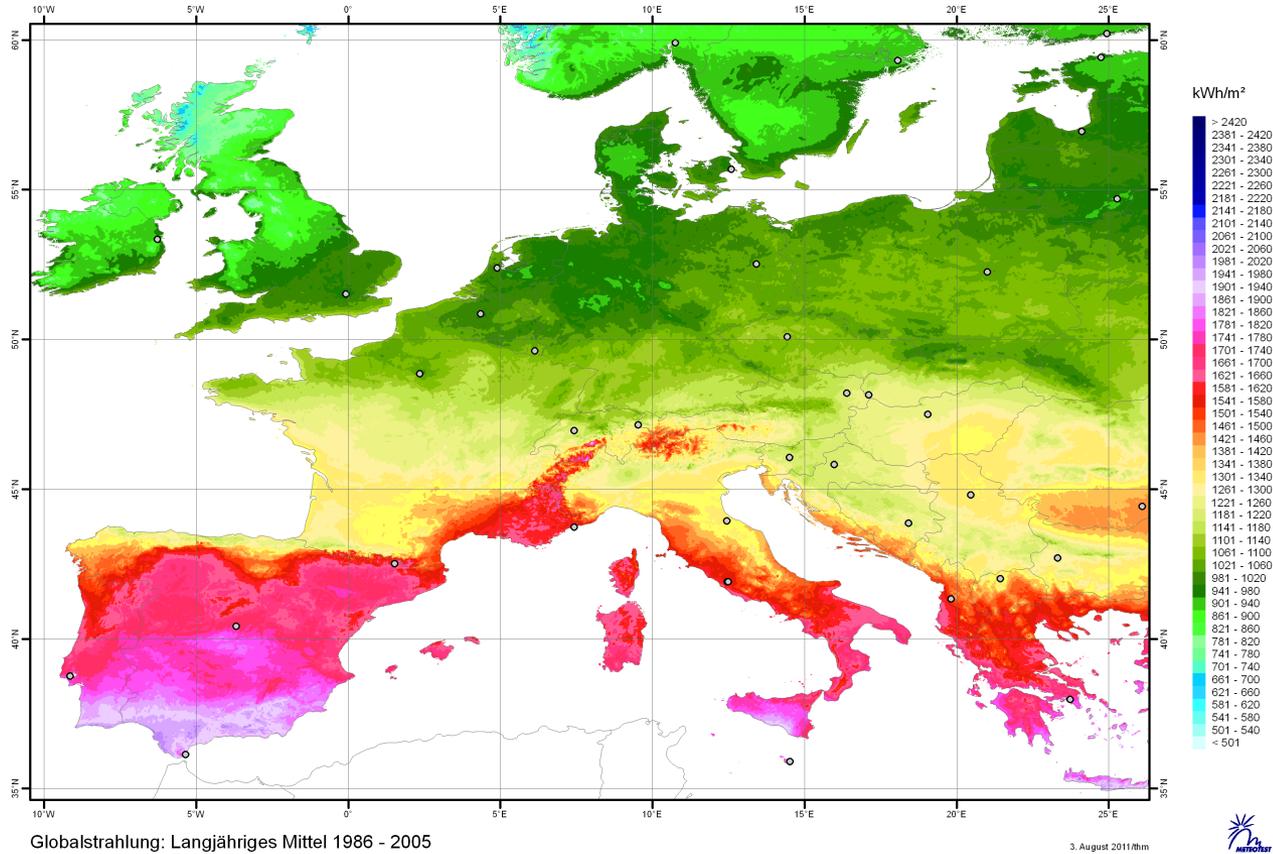
1. infrared camera





# Exercise 2 - 1 b)

## 3. geo localization



# Exercise 2 - 2

## **The Beatles, 1963: „I Want to Hold Your Hand“:**

Oh, yeah, I tell you something  
I think you'll understand  
When I say that something's  
I wanna hold your hand (3x)

Oh, please say to me  
And/You'll let me be your man  
And, please, say to me  
You'll let me hold your hand ...

And when I touch you, I feel happy inside  
It's such a feeling that my love  
I can't hide, I can't hide

Yeah, you got that something  
I think you'll understand  
When I say that something's  
I wanna hold your hand ...

And when I touch you, I feel happy inside  
It's such a feeling that my love  
I can't hide...

Yeah, you got that something  
I think you'll understand  
When I feel that something  
I wanna hold your hand ...

## **The Beatles, 1964: „Komm gib mir deine Hand“:**

O komm doch, komm zu mir  
Du nimmst mir den Verstand  
O komm doch, komm zu mir  
Komm gib mir deine Hand (3x)

O du bist so schön  
Schön wie ein Diamant  
Ich will mir dir gehen  
Komm gib mir deine Hand ...

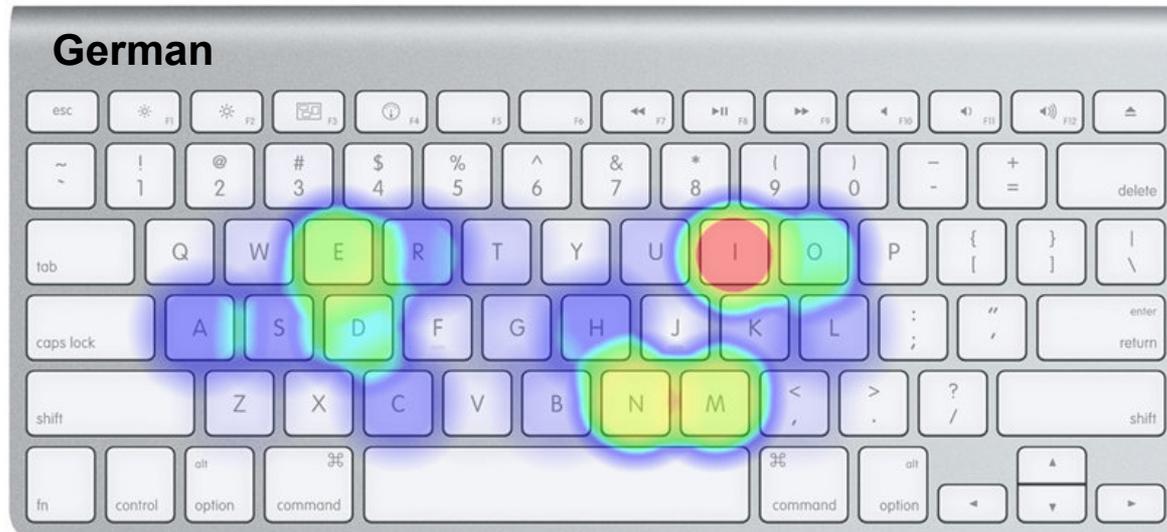
In deinen Armen bin ich glücklich und froh  
Das war noch nie bei einer anderen einmal so  
Einmal so, einmal so

O du bist so schön  
Schön wie ein Diamant  
Ich will mir dir gehen  
Komm gib mir deine Hand ...

In deinen Armen bin ich glücklich und froh  
Das war noch nie bei einer anderen einmal so  
Einmal so, einmal so

O komm doch, komm zu mir  
Du nimmst mir den Verstand  
O komm doch, komm zu mir  
Komm gib mir deine Hand...

# Exercise 2 - 2

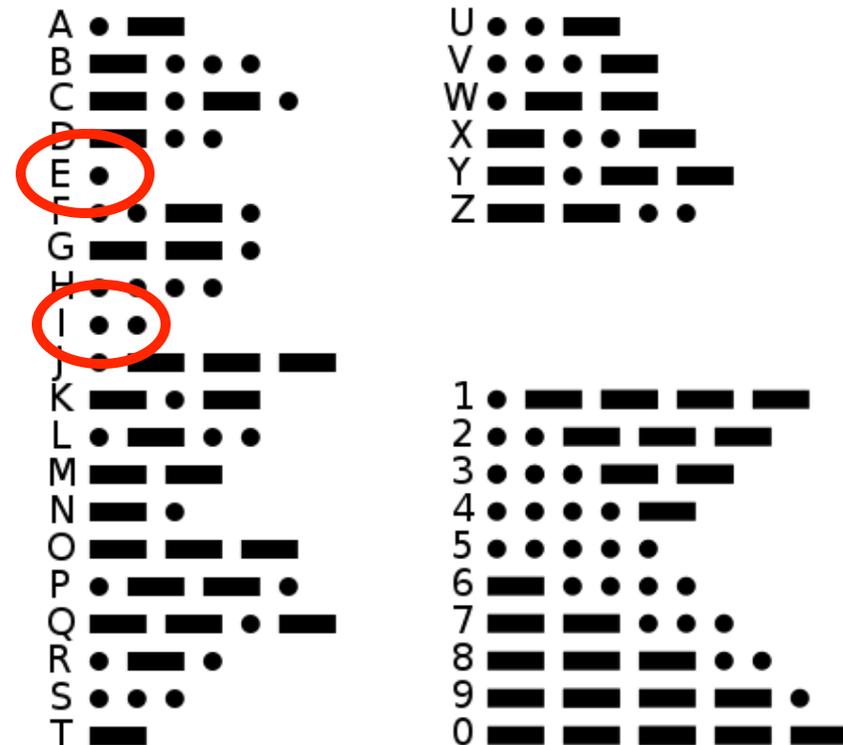


# Exercise 2 - 2

- Information visualization should allow for efficient data exploration
- Heatmaps show immediately (preattentively) which letters are used most frequently in the respective texts
- E, I, N, M dominate those German texts
- Entropy encodings like the Morse alphabet exploit such properties

## International Morse Code

1. A dash is equal to three dots.
2. The space between parts of the same letter is equal to one dot.
3. The space between two letters is equal to three dots.
4. The space between two words is equal to seven dots.



# Visualizing Multivariate Data

# Multivariate Data?

→ Data based on more than one variable per sampling unit.

Example: Weather data (Munich)

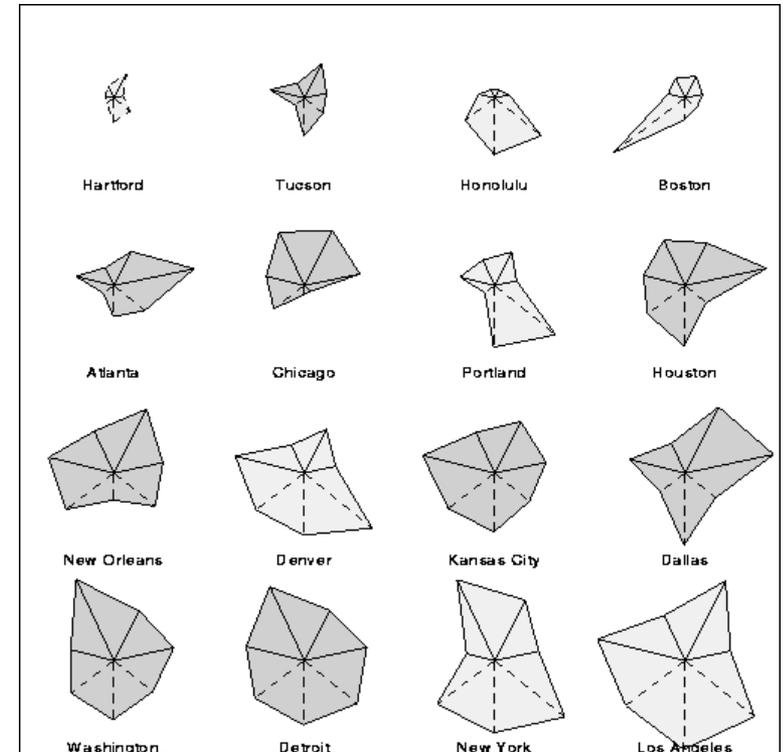
	Temperatur °C		Niederschlag		relative Feuchte	Sonne h/Tag
	max. Ø	min. Ø	mm	Tage		
Jan	1,6	-5,1	53	16	83	2
Feb	3,6	-4	52	15	83	2,7
Mär	8,1	-0,8	56	13	77	4,1
Apr	12,6	2,6	75	14	72	5,1
Mai	17,4	6,8	107	15	73	6,4
Jun	20,5	10,2	131	16	73	6,8
Jul	22,8	12,1	116	16	73	7,6
Aug	22,3	11,8	116	15	75	6,9
Sep	19,1	8,9	79	13	78	5,6
Okt	13,6	4,4	57	12	82	4,2
Nov	6,9	-0,1	64	14	86	2,2
Dez	2,6	-3,7	60	14	86	1,6

# Glyphs

- Small-sized visual symbols
- Variables are encoded as properties of glyph
- Each case is represented by a single glyph
  
- **Main Limitation:** Have to be learned
- Not suitable for large data sets.

# Star Glyphs

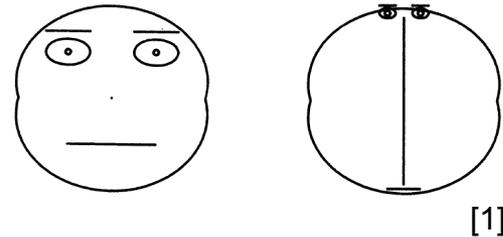
- aka web diagram, spider chart, star diagram ...
- Radial axes representing the variables
- Allows for comparison based on the shape of the resulting object



- Limitations:
  - Works for small data sets only [5]
  - Hard to compare fine differences in spoke lengths
  - Thus better suited for identifying outliers

# Chernoff Faces

- Theory
  - Humans are able to recognize small changes in facial characteristics
  - Data is encoded by stylized faces using up to 18 characteristics



- Limitations
  - Extreme values negatively influence the impression of a face and the recognition of other values [1]
  - Experiments [2] reveal that recognition of Chernoff faces is a serial process and thus there is no significant advantage over other iconic visualization

# Your Turn

Student	Course	Relationship	MMI 1	Infovis	DBS	Theor.	DS
1	MI	yes	2	1	2	4	2
2	Inf	no	2	2	1	2	1
3	KuM	yes	2	1	4	5	4
4	KuM	no	1	2	3	3	2
5	MI	no	1	1	2	2	2
6	Inf	yes	2	1	2	4	3
7	Inf	no	3	2	1	2	2
8	MI	yes	2	1	2	3	3
9	KuM	yes	1	2	3	4	5

 Nominal

 Ordinal

# **Some Background on Data Types & Evaluation**

# **Types of Data**

# Qualitative vs. Quantitative Data

- deals with descriptions
  - data can be observed but not measured
  - colors, textures, smells, tastes, etc.
  - Qualitative -> Quality
- deals with numbers
  - data which can be measured
  - length, height, area, volume, speed, costs etc.
  - Quantitative -> Quantity

## Oil Painting

### Qualitative data:

- blue/green color, gold frame
- smells old and musty
- texture shows brush strokes of oil paint
- peaceful scene of the country

## Oil Painting

### Quantitative data:

- picture is 40 cm by 60 cm
- with frame 45 cm by 65 cm
- weighs 4 kilogramm
- costs 300€

# Types of Data

From [3]

- Nominal
- Ordinal
- Interval
- Ratio

**non-parametric**

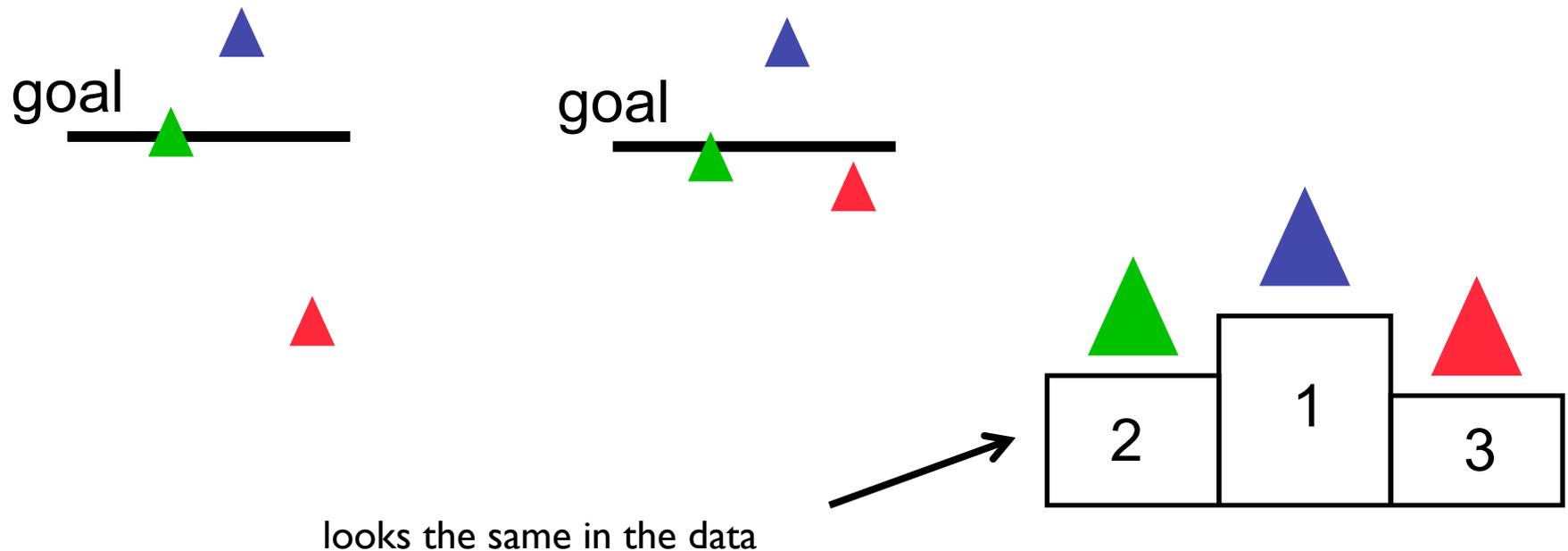
**parametric**



more information

# Ordinal vs. Interval

- ordinal provides an order
- doesn't tell anything about the differences
- example: triangle race



# Evaluating InfoVis

# Cause and Effect

Goal: Find causal links between variables



Precondition: Cause has to precede effect

How to infer causality:

- Two controlled conditions
  - Cause is present (experimental condition)
  - Cause is absent (control condition)

# Hypotheses

- Prediction of the result:
  - „how will the *indepent variables* effect the *dependent variables*?“
- Hypotheses must be formulated before running the study
  - By doing the experiment, the hypotheses is either proved or disaproved



# Variables

- Independent variables: „*What do I change?*“
  - Manipulated by the experimenter
  - Conditions under which the tasks are performed
  - The number of different values is called **level**, e.g.
    - Traffic light can be *red, yellow* or *green* (3 levels)
- Dependent variables: „*What do I observe?*“
  - Affected by the independent variables
  - Measured in the user study
  - Dependent variables should only depend on the independent variables

# Study Designs

## Basic approaches

- Observational: „observe what naturally happens“
- Experimental: „manipulate some aspects“

## Design types

- Within subject („repeated measures“)
  - Each subject is exposed to all conditions
  - The order of conditions must be randomized to avoid ordering effects
- Between groups („independent measures“)
  - Separate groups (participants) for each condition
  - Careful selection of groups is essential
- Hybrid („mixed“) designs

# Participants

- Should be representative for the target group
- Avoid bias (e.g. not only men, students)
- Choose the right sample size
  
- Choose domain experts [4] if possible (especially in infovis)
  - More realistic results and tasks
  - Busy people with few time
  - Hard to get a big enough sample size

# Principles

- The results of the experiment should be

## **1.Valid**

- Measurements are accurate and due to manipulations (**internal validity**)
- Findings are representative and not only valid in the experiment setting (**external validity**)

## **2.Reliable**

- Consistency of measurement
- A persons score doing the same test under the same conditions twice must be similar

## **3.Generalizable**

- Results should be valid for all people
- Test users must be representative

# Infovis Specifics [4]

- Find out:
  - If the visualization supports the user in the information task
  - How to improve the visualization to better support them
- Participants:
  - Domain experts if possible
- Data sets
  - Usually extremely large sets
  - Don't just choose a subset
- Time:
  - It is not unusual for a task to take weeks or months
  - Hard to reproduce this in an experiment
- Tool status:
  - Hard to provide a fully functional tool rather than a prototype

# Likert Scales

- used to „measure“ opinions
- participants give ratings
- **Attention:** there is a huge discussion going on whether likert scale data is ordinal (non-parametric) or interval (parametric)\*

## centered

1. fully agree
2. agree
3. neutral
4. disagree
5. totally disagree

## uncentered

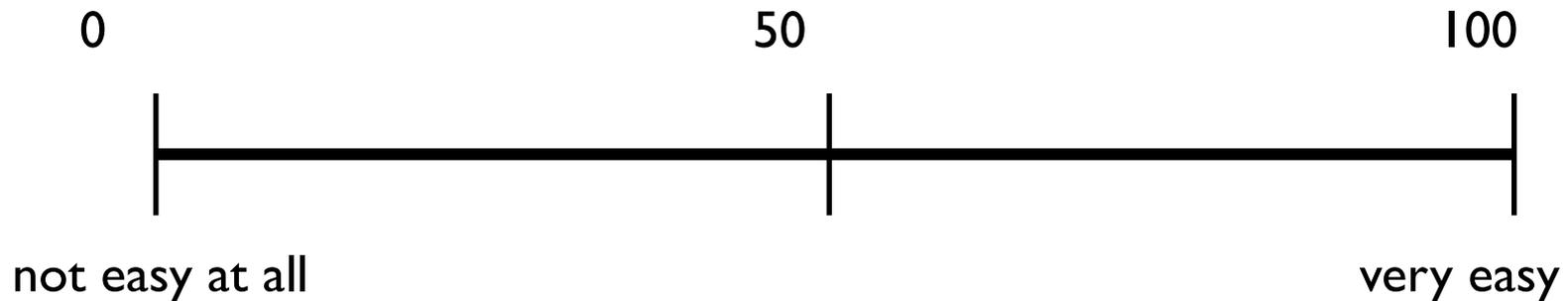
1. fully agree
2. agree
3. disagree
4. totally disagree

\* Computer scientists believe it is ordinal. Please read the following blog entry for information and implications:  
<http://cacm.acm.org/blogs/blog-cacm/107125-stats-were-doing-it-wrong/fulltext>

# Visual-Analog Rating Scales

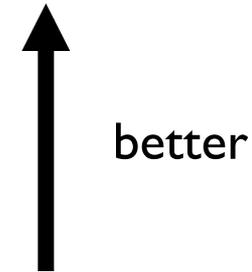
- no categories
- advantage: users cannot remember their response

How easy to use was the prototype?

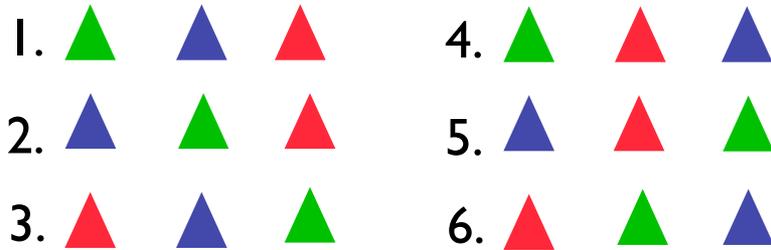


# Learning Effect

- people get better over time
- to avoid influences on the experiment:
  - use perfect counterbalancing if possible
  - Latin square designs
  - randomization
  - other designs



Example: One variable with 3 levels.  $3! = 6$  arrangements.

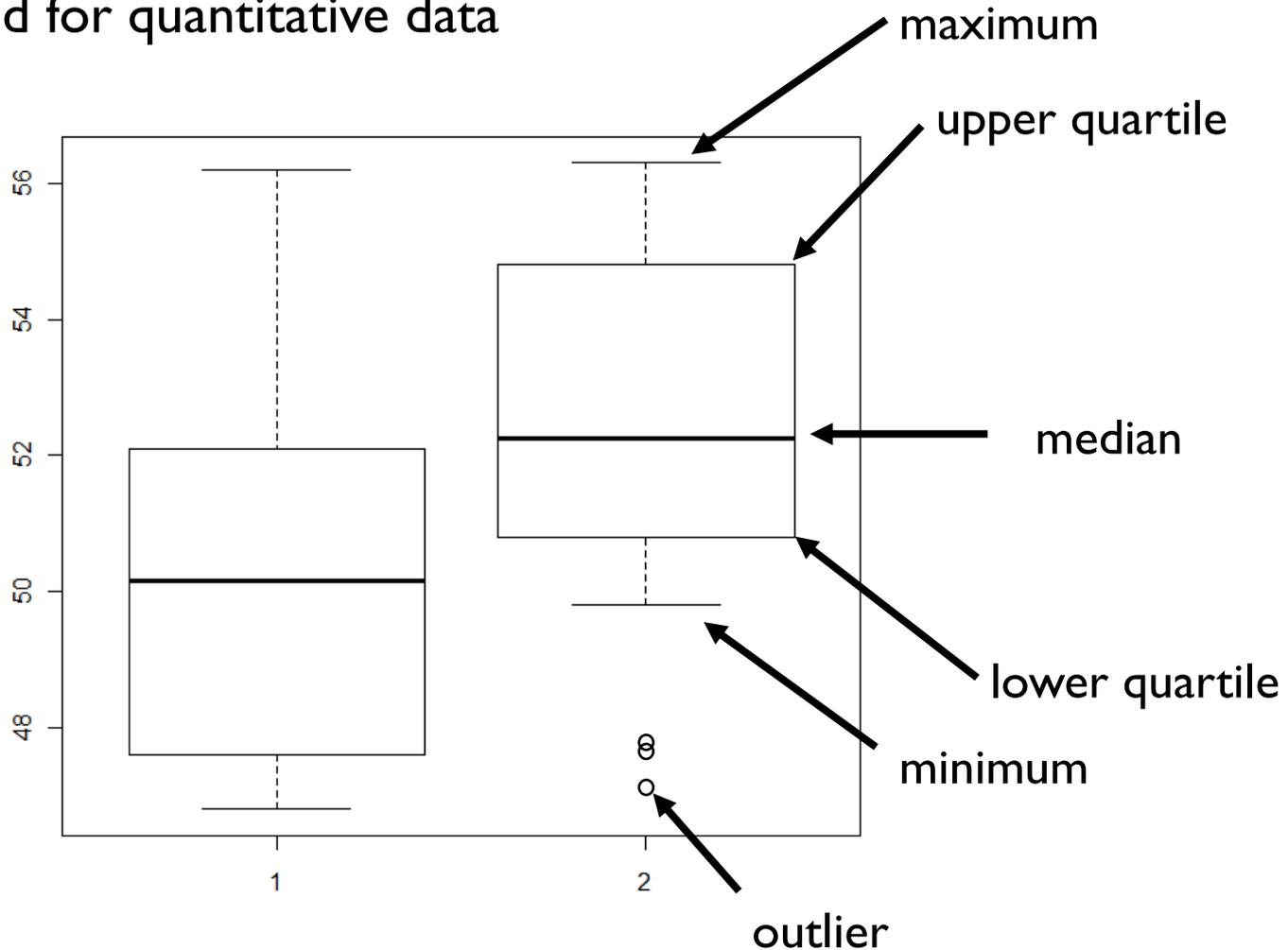


# Analysis

- Choose the right statistical tests
  - Heavily influenced by the choice of measurement tools
  - ... and the types of data used
  - Parametric tests (e.g. ANOVA, T-Test) vs. non-parametric tests (e.g. Wilcoxon, Kruskal-Wallis)
- Choose the right visualization (yes, you have to visualize the results of your visualization study ;-)

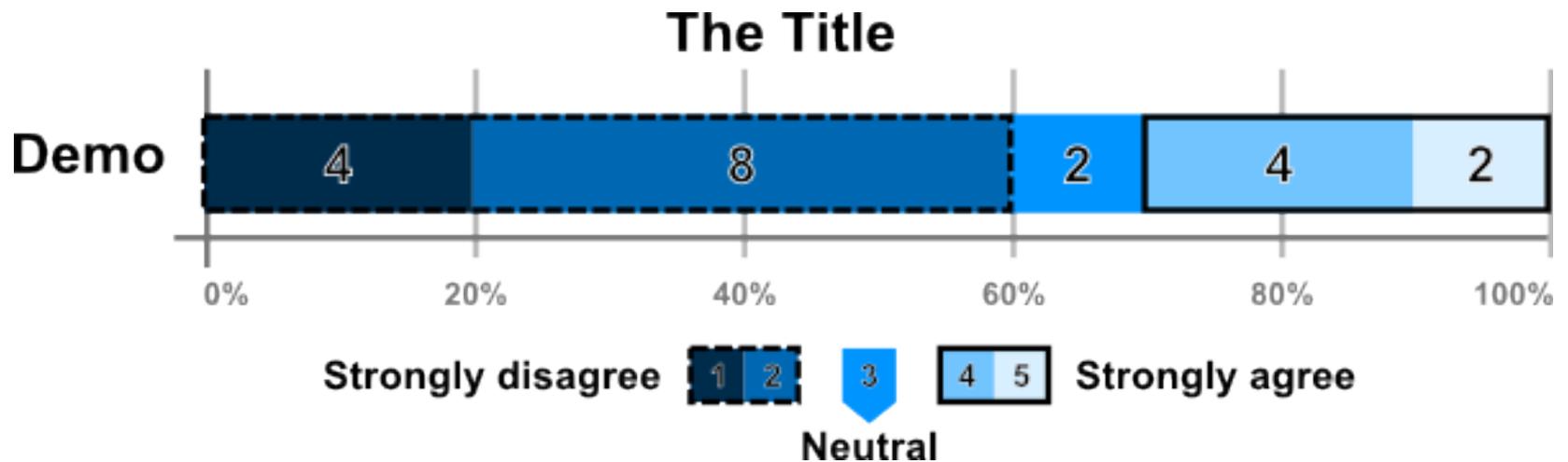
# Boxplot

- good for quantitative data



# Likert Scales?

- Don't report the mean
- If possible, report and visualize frequencies
- For example:



Visualization by Max Maurer. Script available here <http://www.paje-systems.de/likert/>

# References

1. Bernhard Flury and Hans Riedwyl. Graphical Representation of Multivariate Data by Means of Asymmetrical Faces. *Journal of the American Statistical Association*, Vol. 76, No. 376, pp. 757-765. 1981
2. Christopher J. Morris, David S. Ebert and Penny L. Rheingans, "Experimental analysis of the effectiveness of features in Chernoff faces", Proc. SPIE 3905, 12. 2000
3. Field, A., Hole, G. How to Design and Report Experiments. (book)
4. S. Carpendale. Evaluating information visualizations. In A. Kerren, J. T. Stasko, J.-D. Fekete, and C. North, editors, *Information Visualization: Human-Centered Issues and Perspectives*, LNCS 4950, pages 19–45. Springer, 2008.
5. NIST/SEMATECH (2003). Star Plot in: *e-Handbook of Statistical Methods*.
6. Siegel, J. H. & Goldwyn, R. M. & Friedman, H. P.. (1971). Pattern and Process of the Evolution of Human Septic Shock. *Surgery*. 70. 232-245.