Übung zur Vorlesung

Informationsvisualisierung

Alexander De Luca, Emanuel von Zezschwitz
Ludwig-Maximilians-Universität München
Wintersemester 2011/2012

Evaluating Infovis

…and stuff

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)

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 [1]

- Nominal
- Ordinal

non-parametric

- Interval
- Ratio

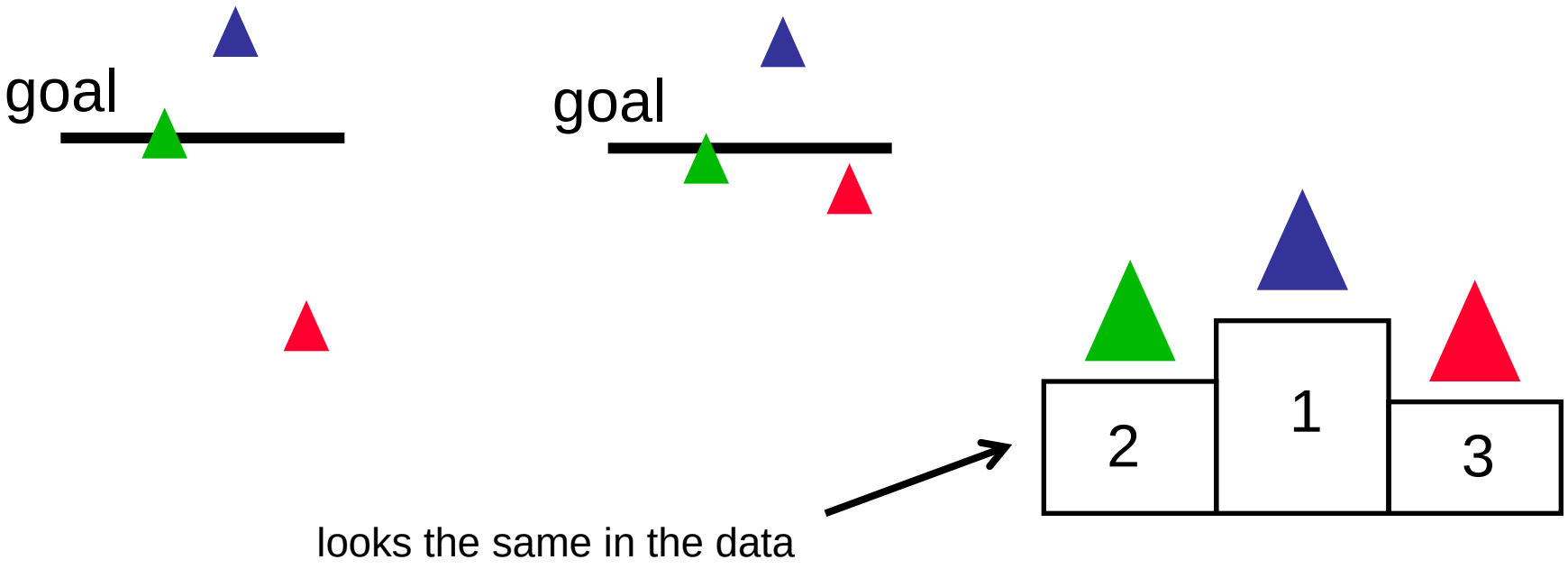
parametric

more information

Ordinal vs. Interval

- ordinal provides an order
- doesn‘t tell anything about the differences

- example: triangle race



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 indepent 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 [2] 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 [2]

- 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

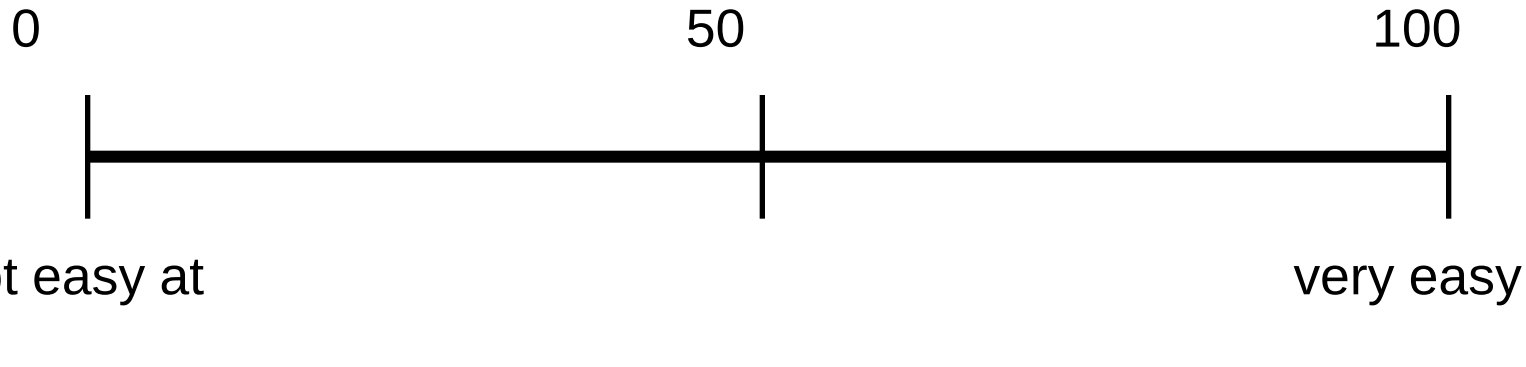
1. fully agree
2. agree
3. disagree
4. totally disagree

* 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?

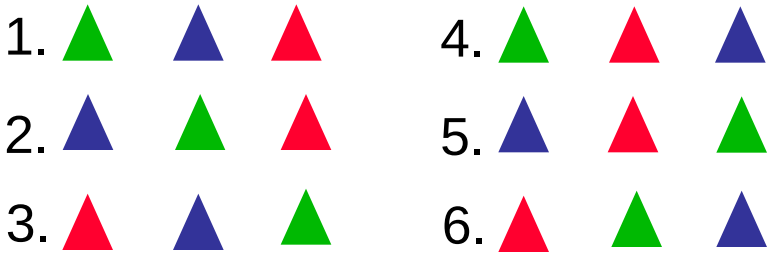


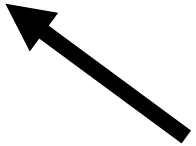
Learning Effect

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

better

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





counterbalanced

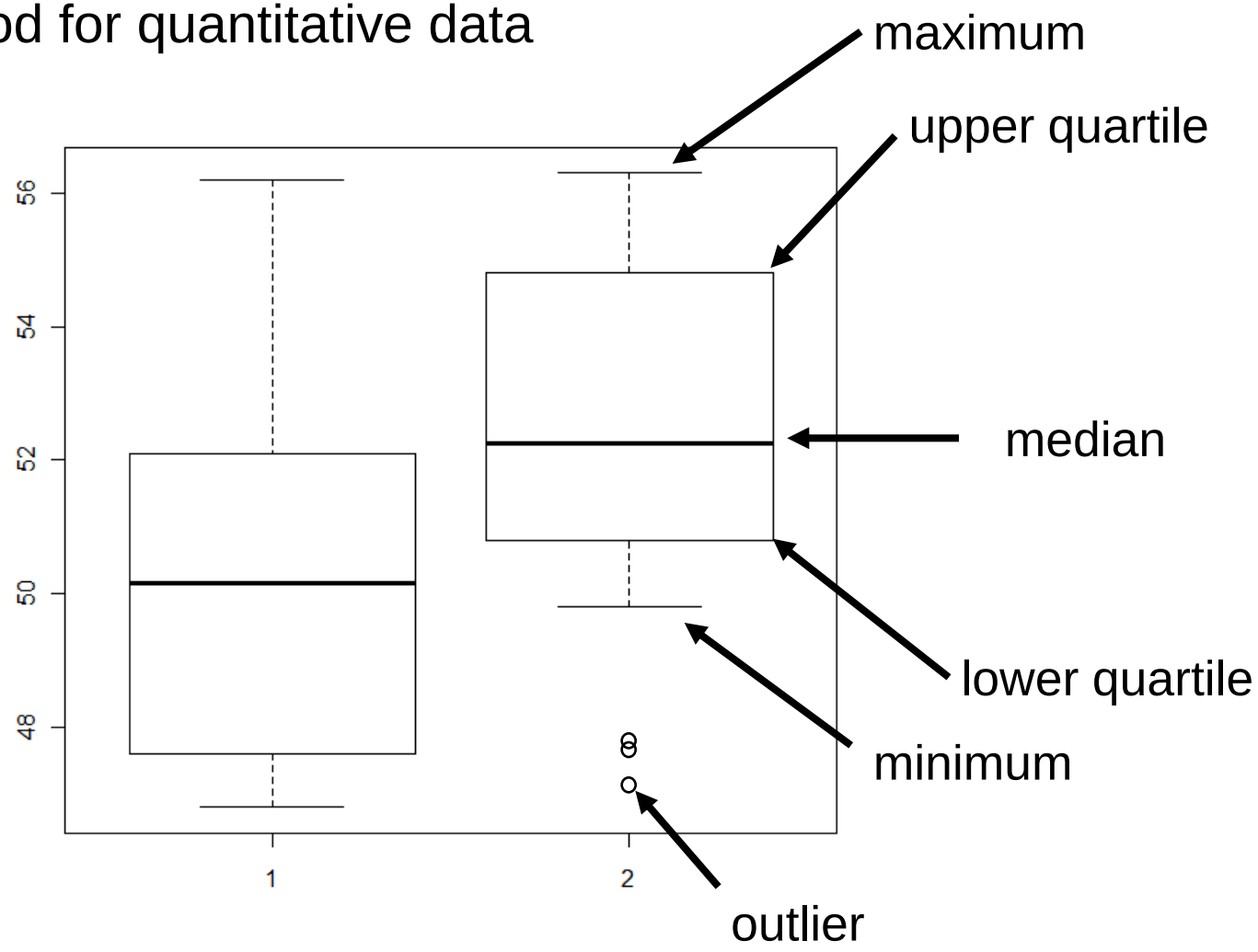
Analysis

- 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

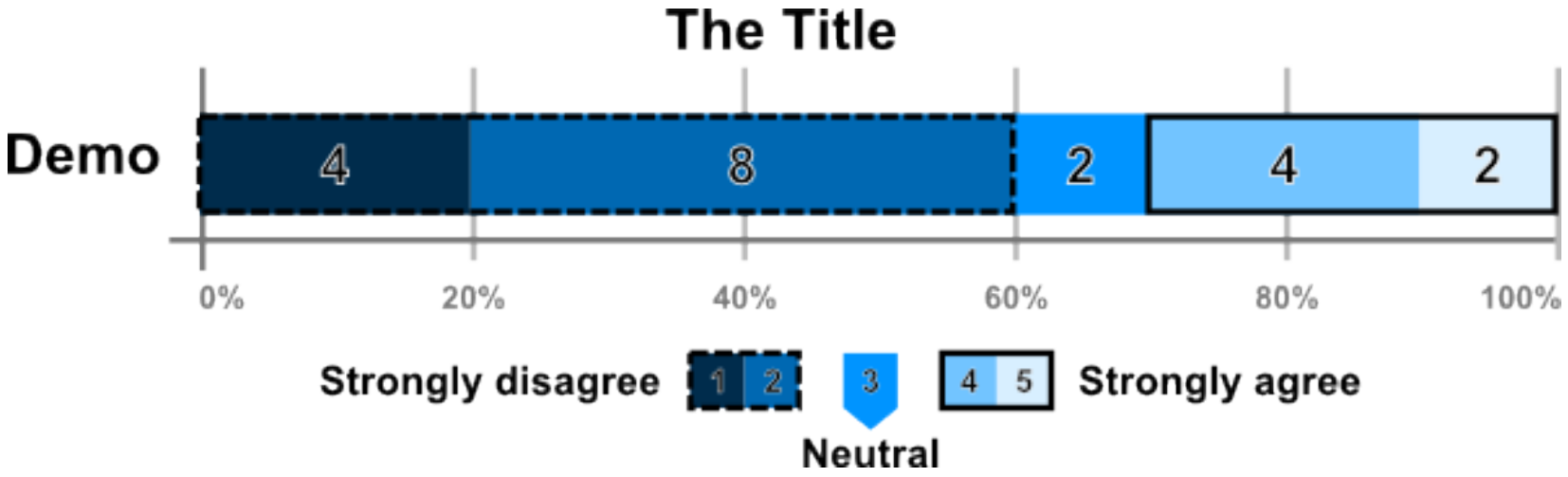
- good for quantitative data



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. Field, A., Hole, G. How to Design and Report Experiments. (book)
2. 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.