Navigating Data

Martin Theus
Augsburg University

www.rosuda.org

Outline

- Selections
- Scales
- Multiples

Martin Theus
Augsburg University

Introduction

- Statistics deals with the analysis of data
- Statistical graphics, i.e. data visualization is the only way to look at data
- Static graphics are sometimes sufficient for a presentation of final results, but the exploration of data needs interaction with the graphics
- These interactions make up more than 80% of the software development efforts
- The main interactions are
  (1) **selecting** data
  (2) looking at data at different **scales**
  (3) creating **multiple** views of the data

Martin Theus
Augsburg University

Selections (1)

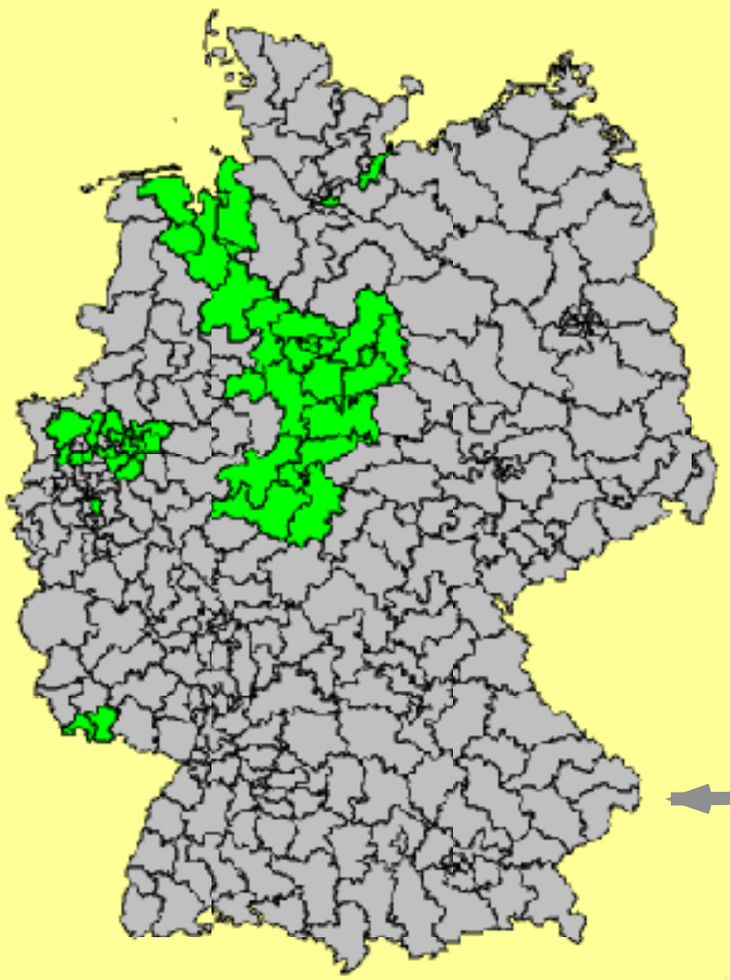
- Selecting subgroups is most powerful when the subgroup is highlighted in all other graphics
- Selection and linked highlighting corresponds to Subsetting and Conditioning
- Simple Selection can only select simple groups
- More sophisticated selections need special graphical interface
- Selection Sequences:
  - (1) give visual feedback on the selected data
  - (2) allow to combine selections to complex queries

Martin Theus
Augsburg University

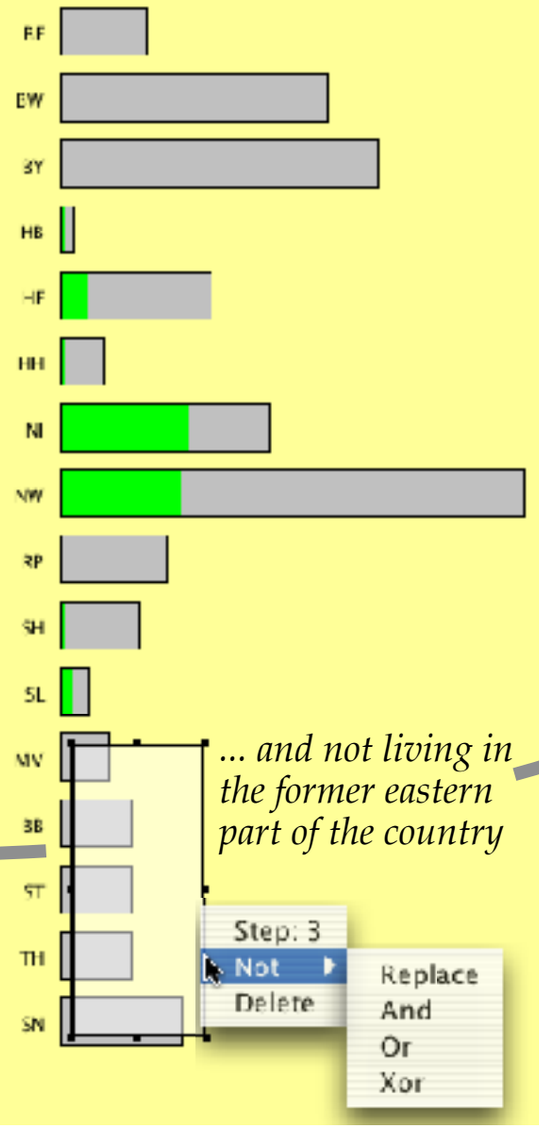
Selections (2)

- Example: German Election Data

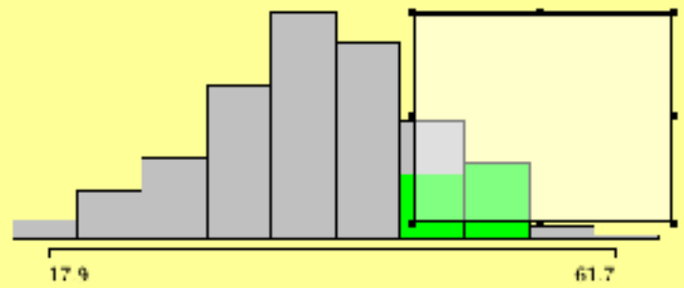
Electoral Districts



States



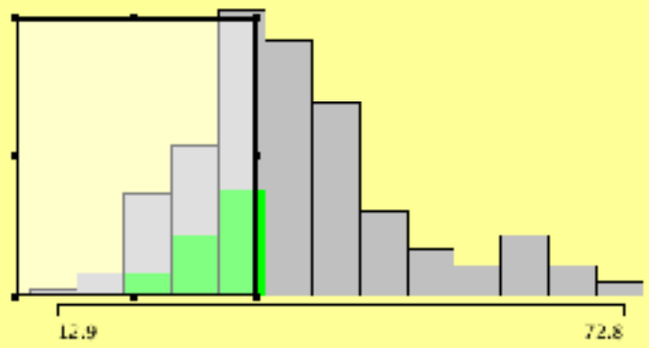
% Votes for the left wing party



Select all districts with more than 45% for the left wing party ...

... and all districts with less than 35% for the conservative party

% Votes for the conservative party



... and not living in the former eastern part of the country

Martin Theus
Augsburg University

Selections (3)

- Graphical selections of subgroups must be translated into database queries
- German Election Example:

```
SELECT districts
  FROM election
 WHERE (((SPD > 45)
        AND CDU < 35)
        AND NOT state
              IN (MV, BB, ST, TH, SN))
```

Martin Theus
Augsburg University

Scales (1)

- A change of scale is often just called "zoom"
- To achieve "Overview first, zoom and filter, then details on demand" (Shneiderman), we need far more!
- Zooming comes in various flavors:

  (1) **Classical Zooming**

  Simple change of coordinate system

  (2) **Logical Zooming**

  Aggregation level and graphical representation change

  (3) **Censored Zooming**

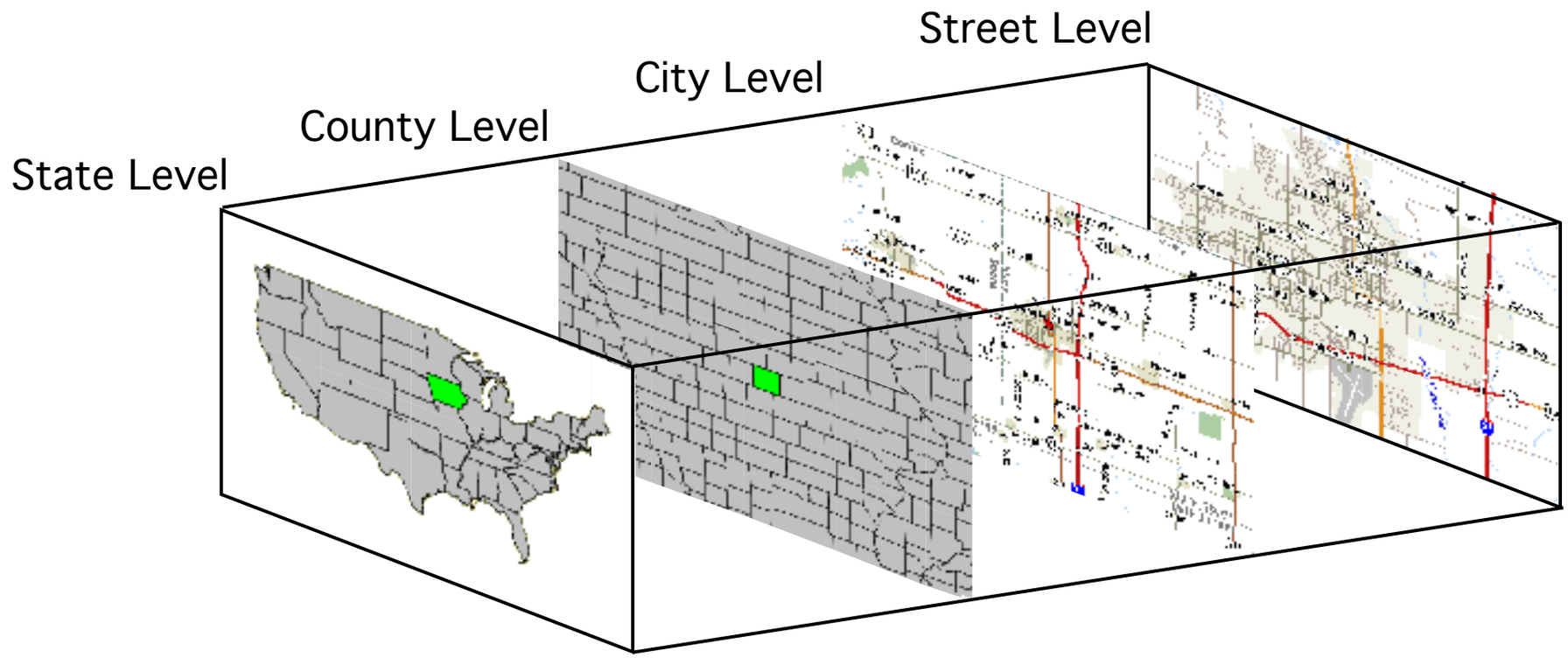
  Zooming of objects, size is limited to a fixed range

Martin Theus
Augsburg University

Scales (2)

- Classical Example:

  Logical zooming in a geographical context:



Martin Theus
Augsburg University

Scales (3)

- **More advanced example:**
  - **Zooming in scatterplots**

    Start with the whole data set in a binned scatterplot and switch to point display when looking at less points

  - **Zooming in mosaic plots**

    Keep the number of bins "constant" and zoom in on conditioned groups

- **In any case:**

  Zooming must be hierarchic, i.e. a "zoom back" steps back to a formerly visited view!

Martin Theus
Augsburg University

Multiples (1)

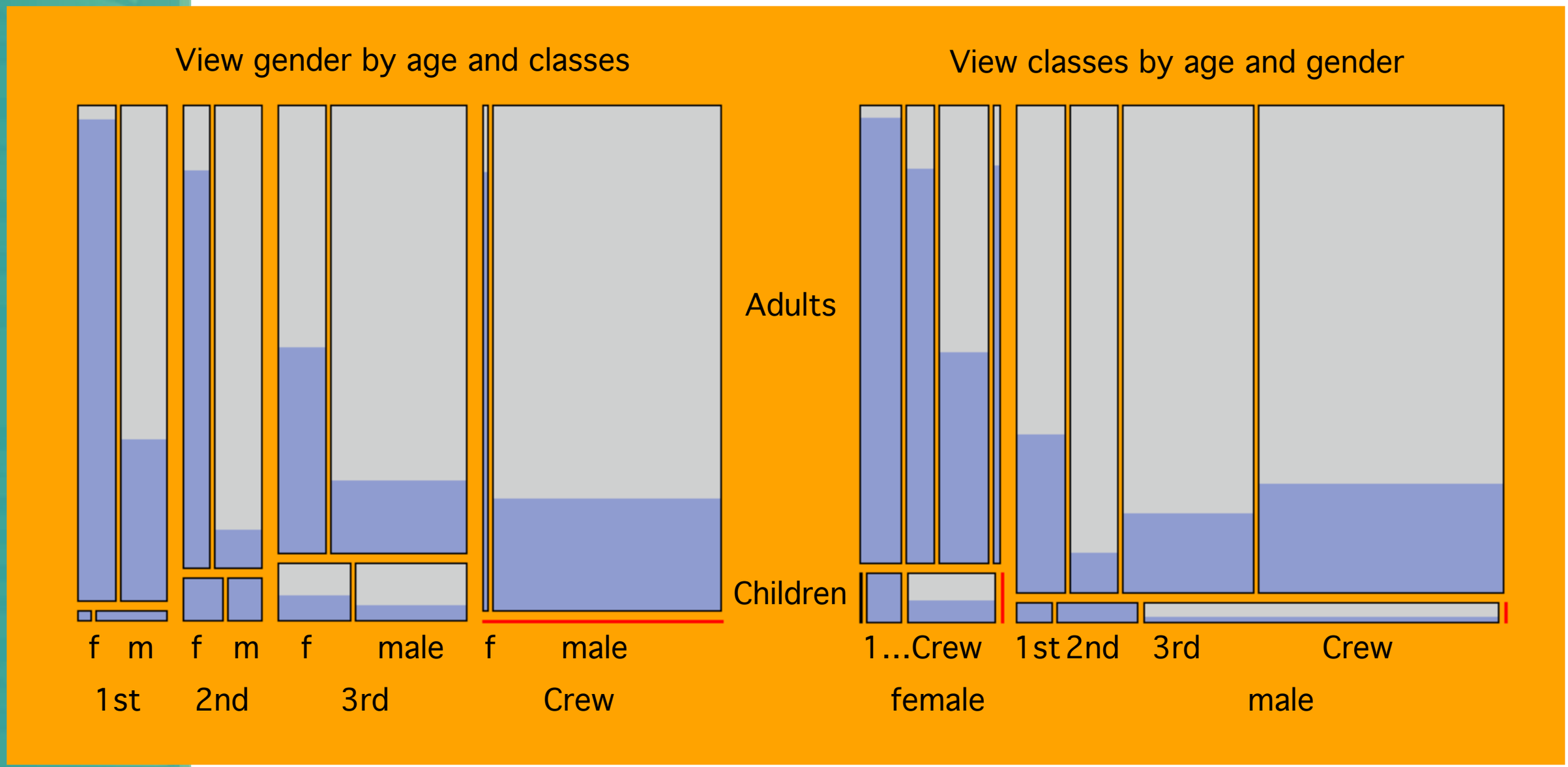
There are many reasons why we need more than just a single view at the data set/variable

- Looking at graphics at different zoom levels simultaneously
- Looking at the same variable with different plot types (e.g. barchart/spineplot, histograms/spinogram)
- Looking at an array of plots of the same type of many variables (e.g. scatterplot matrices)
- Looking at multivariate data from different perspectives (e.g. grand tour in scatterplots and mosaic plots)

Martin Theus
Augsburg University

Multiples (2)

Multiples Views on the Titanic Data:



Martin Theus
Augsburg University

Multiples (3)

Support for multiple views:

- Window System with desktop concept
- Flexible and fast change of plot parameters
- High interaction graphical interfaces

... and

- "Think beyond the default!", i.e. the exploratory nature of an analysis is very important.

Martin Theus
Augsburg University

Conclusions

- Graphical selections of subgroups are powerful and easy to interpret
- A direct connection to a database is possible
- Huge data sets need to be investigated on different levels of detail
- Simple zooming is usually not enough
- Multiple views support an exploratory data analysis strongly

- Software: www.rosuda.org/Mondrian