

Counts, Proportions, Interactions – A View on Categorical Data

Martin Theus, AT&T Labs Research, Florham Park, NJ
Adalbert FX Wilhelm, University of Augsburg, Germany

OUTLINE

Counts

No Success Story without Failure

The Key to Success

Proportions

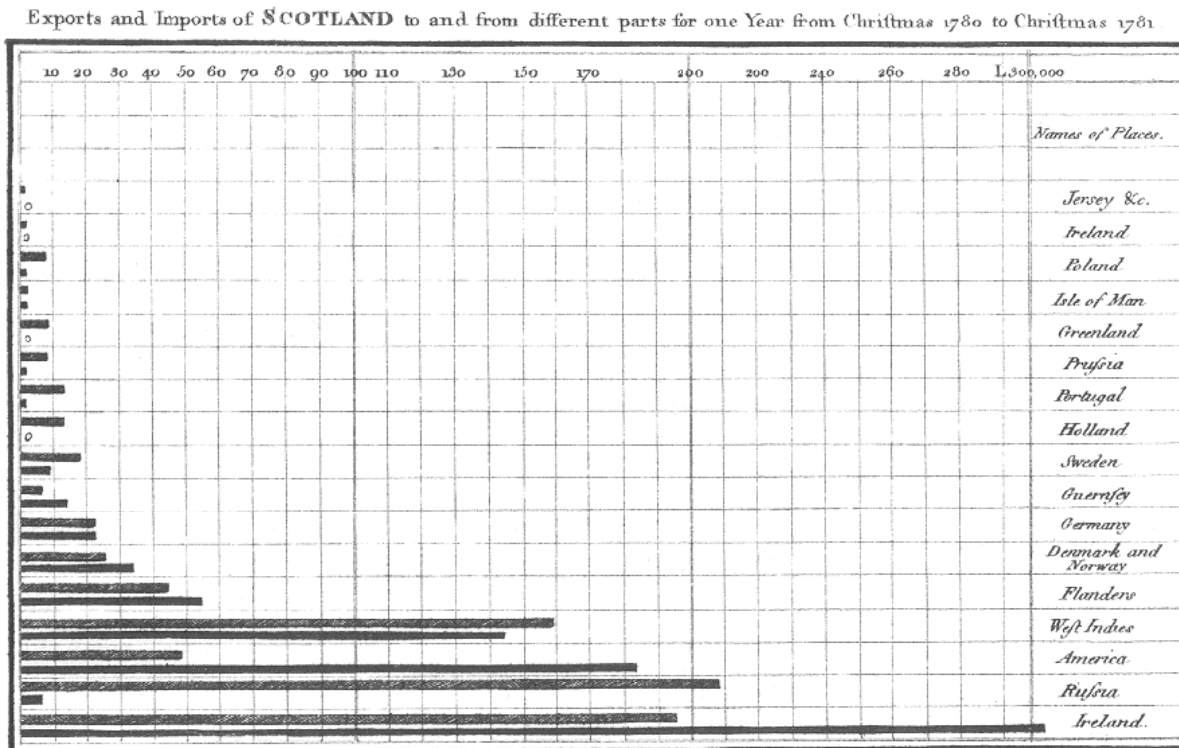
Understanding Black Boxes

Interactions

Counts, Proportions, Interactions – A View on Categorical Data

Martin Theus, AT&T Labs Research, Florham Park, NJ
Adalbert FX Wilhelm, University of Augsburg, Germany

William Playfair (1786)

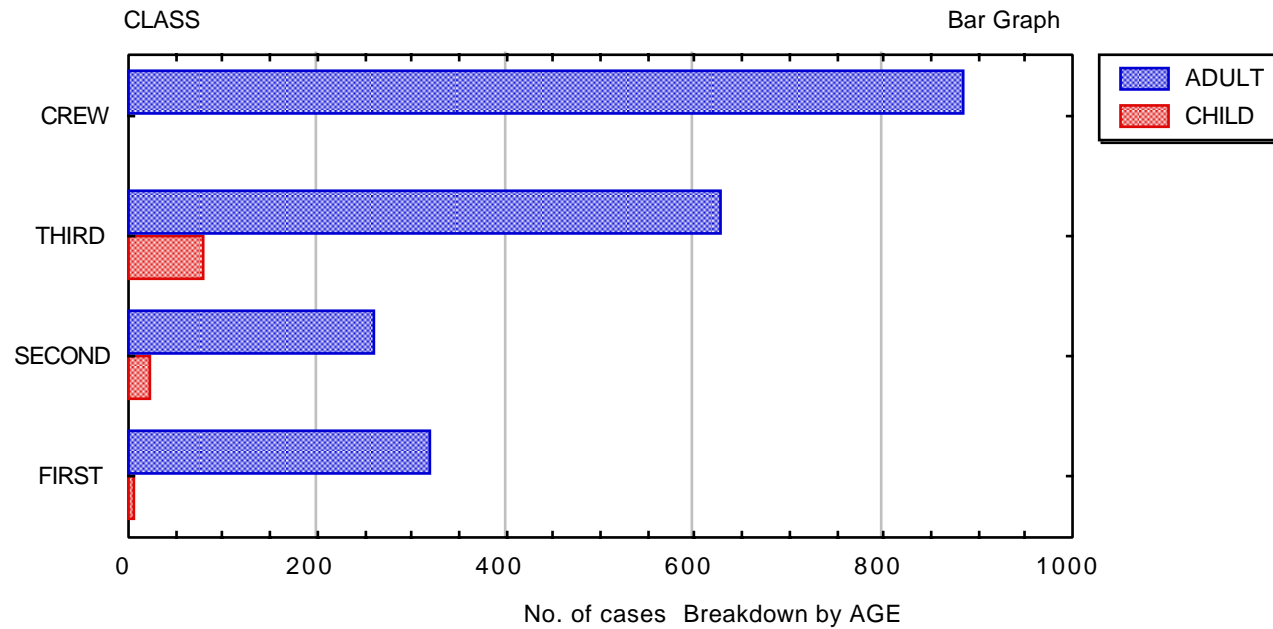


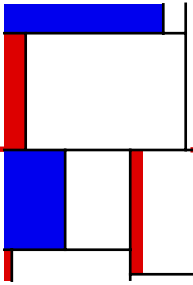
The Upright divisions are Ten Thousand Pounds each. The Black Lines are Exports the Ribbed lines Imports
Published as the Act direct June 7th 1786 by W^m Playfair *Printed by J. Baskin, Strand, London.*

Counts, Proportions, Interactions – A View on Categorical Data

Martin Theus, AT&T Labs Research, Florham Park, NJ
Adalbert FX Wilhelm, University of Augsburg, Germany

1998 Statistics Package:

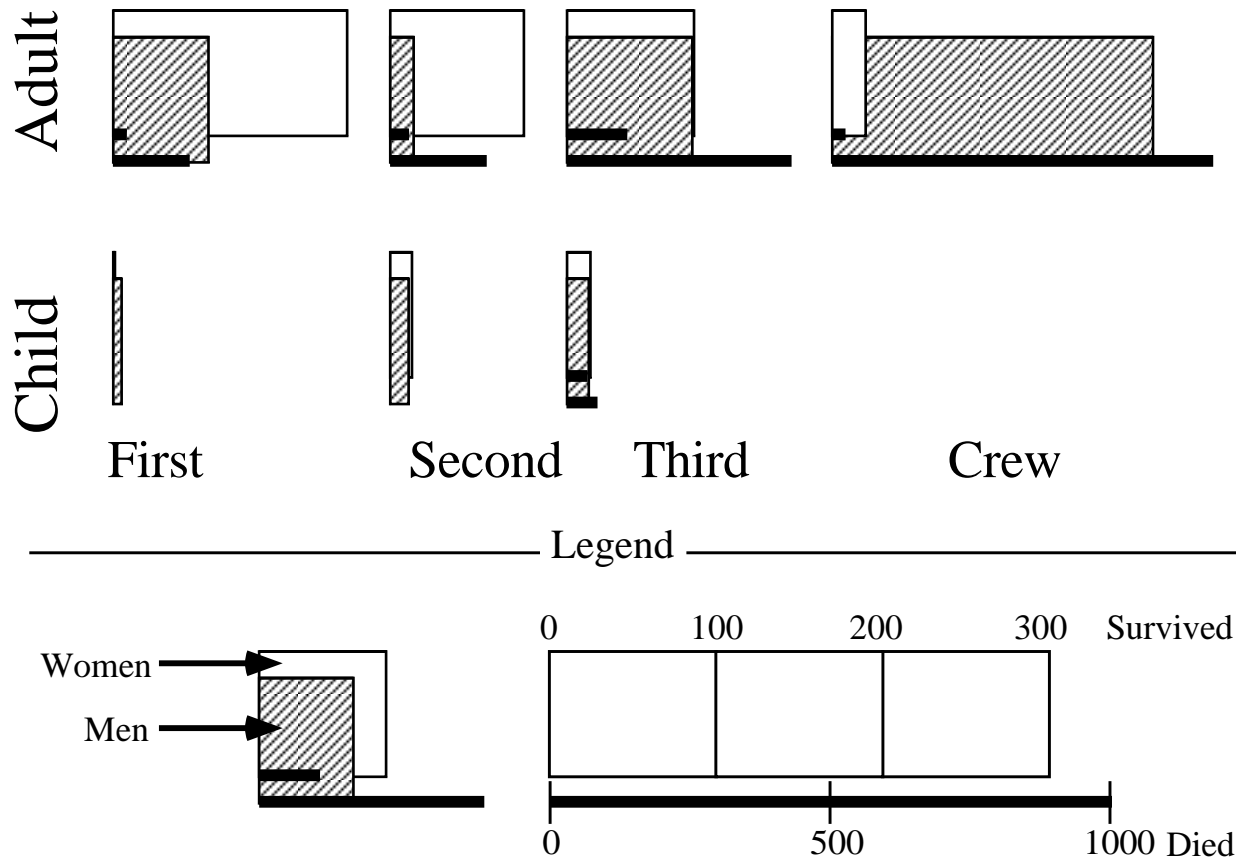


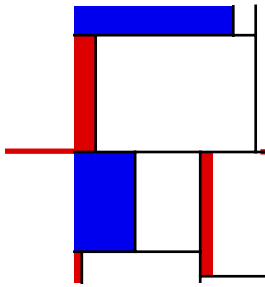


Counts, Proportions, Interactions – A View on Categorical Data

Martin Theus, AT&T Labs Research, Florham Park, NJ
Adalbert FX Wilhelm, University of Augsburg, Germany

Bertin (1967)

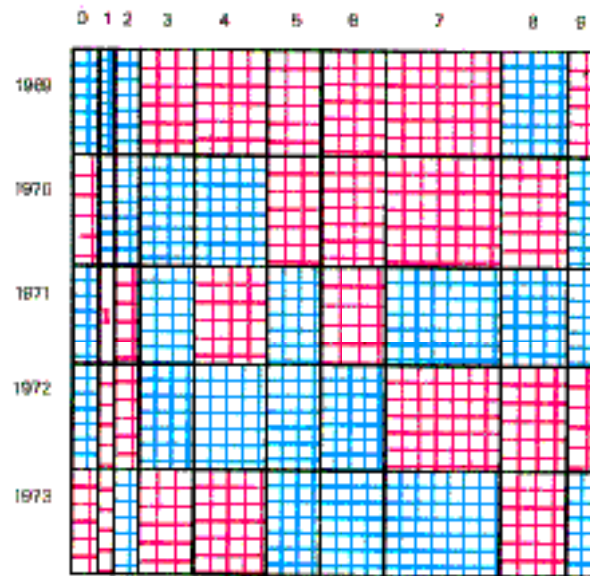


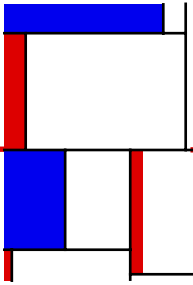


Counts, Proportions, Interactions – A View on Categorical Data

Martin Theus, AT&T Labs Research, Florham Park, NJ
Adalbert FX Wilhelm, University of Augsburg, Germany

Riedwyl & Schuepbach (1983)

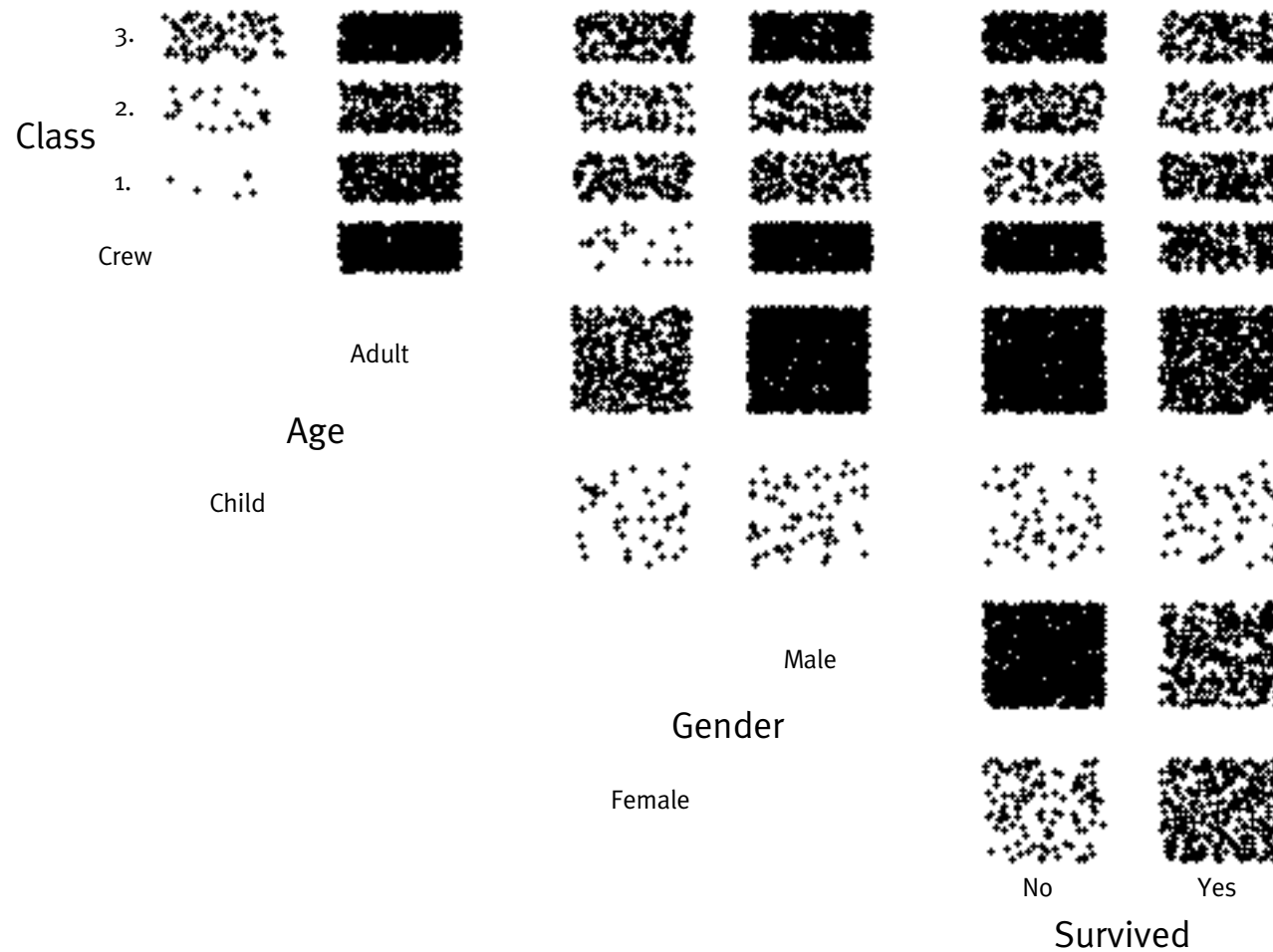




Counts, Proportions, Interactions – A View on Categorical Data

Martin Theus, AT&T Labs Research, Florham Park, NJ
Adalbert FX Wilhelm, University of Augsburg, Germany

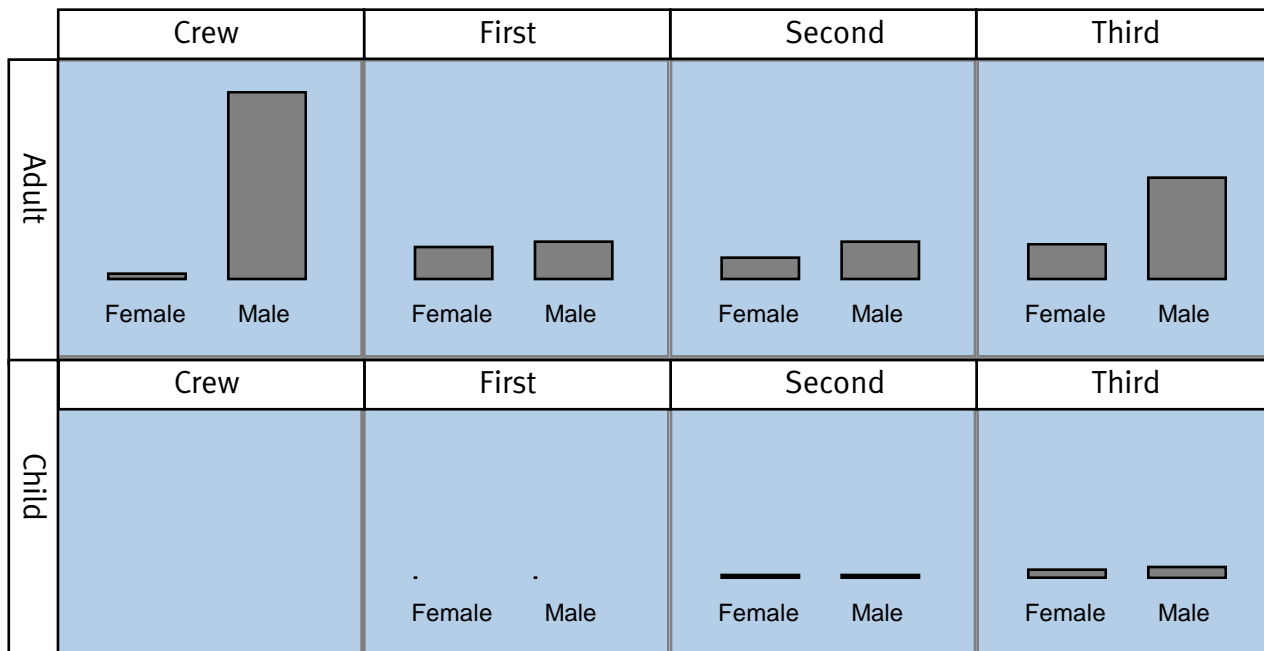
Ostermann & Nagel (1992)

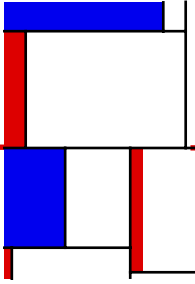


Counts, Proportions, Interactions – A View on Categorical Data

Martin Theus, AT&T Labs Research, Florham Park, NJ
Adalbert FX Wilhelm, University of Augsburg, Germany

Cleveland (1993)





Counts, Proportions, Interactions – A View on Categorical Data

Martin Theus, AT&T Labs Research, Florham Park, NJ
Adalbert FX Wilhelm, University of Augsburg, Germany

DEMANDS ON STATISTICAL GRAPHS

- **Generalizability**

A design of a graph should be generalizable to more than just the number of variables it was initially designed for.

E.g. Scatterplot

- **Consistency**

Data measured on the same scale should be plotted by the same method

E.g. counts by areas,
points on a continuous scale

- **Extensability**

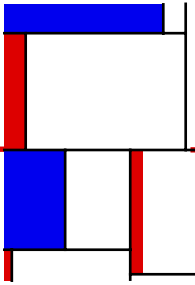
The basic design of a plot should allow the extension of the plot for different purposes

E.g. highlighting and coloring of subgroups, superposing residuals or other modeling information.

- **Interactivity**

The functionality of a plot produced by a modern statistics package should reach beyond a simple 'drawing'

- Plots should be linked and show highlighting of selected data.
- The user should be able to interrogate for information.
- The parameterization of a plot should be easy to change dynamically.

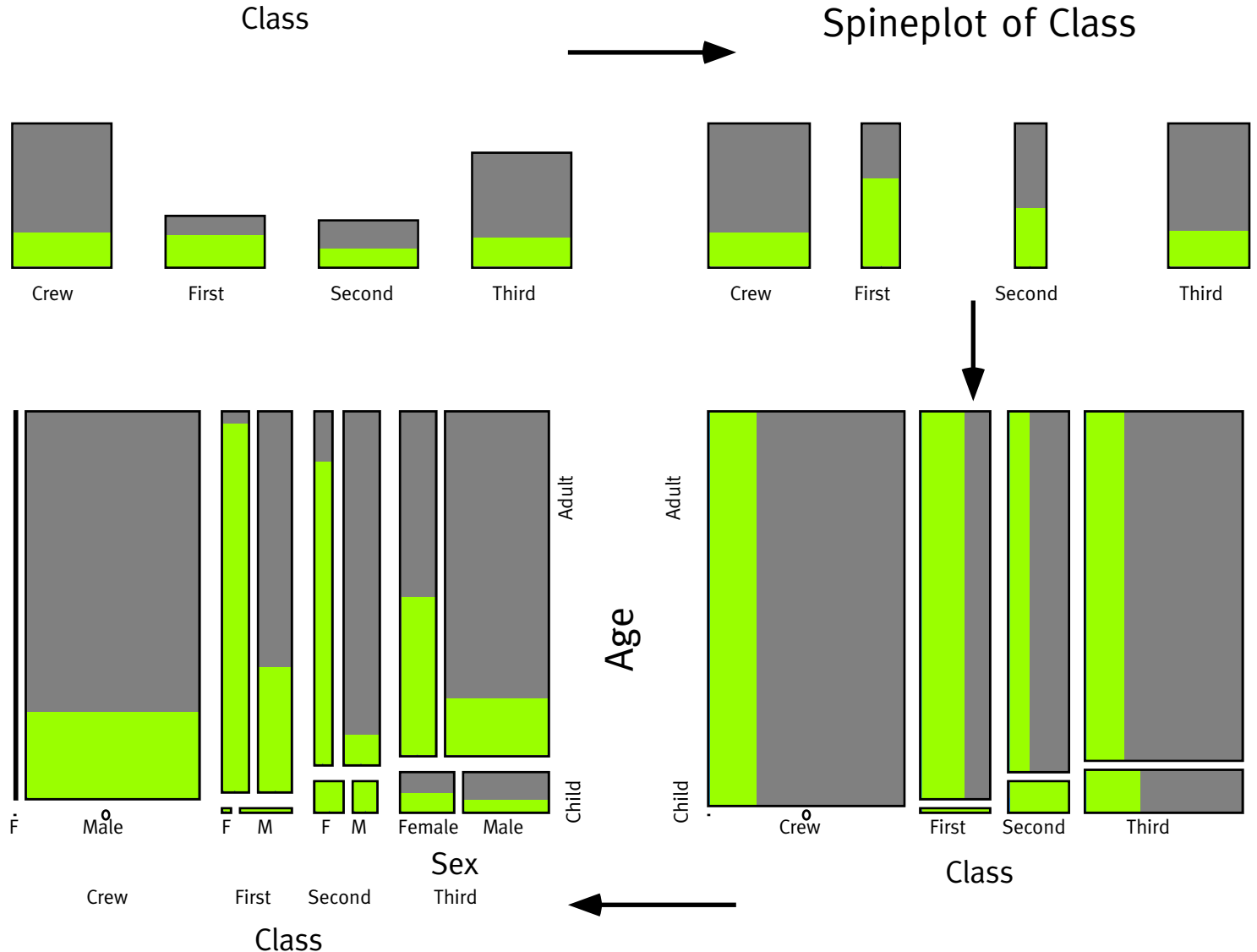


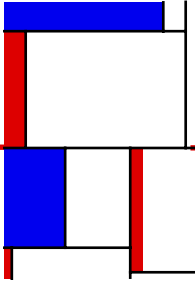
Counts, Proportions, Interactions – A View on Categorical Data

Martin Theus, AT&T Labs Research, Florham Park, NJ
Adalbert FX Wilhelm, University of Augsburg, Germany

MOSAIC PLOT (Titanic Data)

- Class
- Age
- Sex
- Survived





Counts, Proportions, Interactions – A View on Categorical Data

Martin Theus, AT&T Labs Research, Florham Park, NJ
Adalbert FX Wilhelm, University of Augsburg, Germany

INTERACTIVE MOSAIC PLOTS

General Interactivity

- Selection ✓
- Highlighting ✓

Specific Interactivity

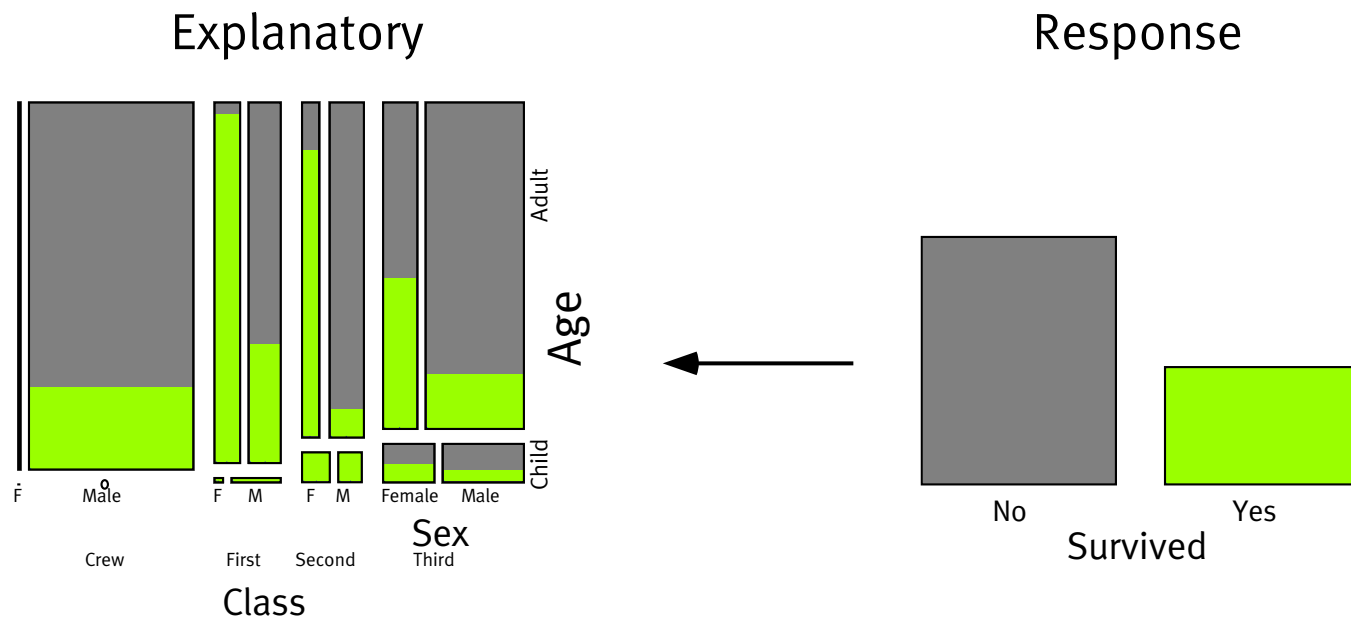
- Interrogation, since labels would jam most plots
- Reordering of the variables
- Reordering of the categories

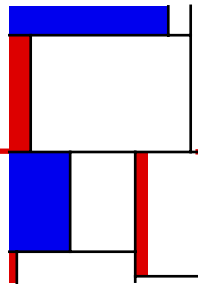
Counts, Proportions, Interactions – A View on Categorical Data

Martin Theus, AT&T Labs Research, Florham Park, NJ
Adalbert FX Wilhelm, University of Augsburg, Germany

VISUALIZING RESPONSE MODELS

- What is the influence of a particular subgroup on the response variable?
Titanic again





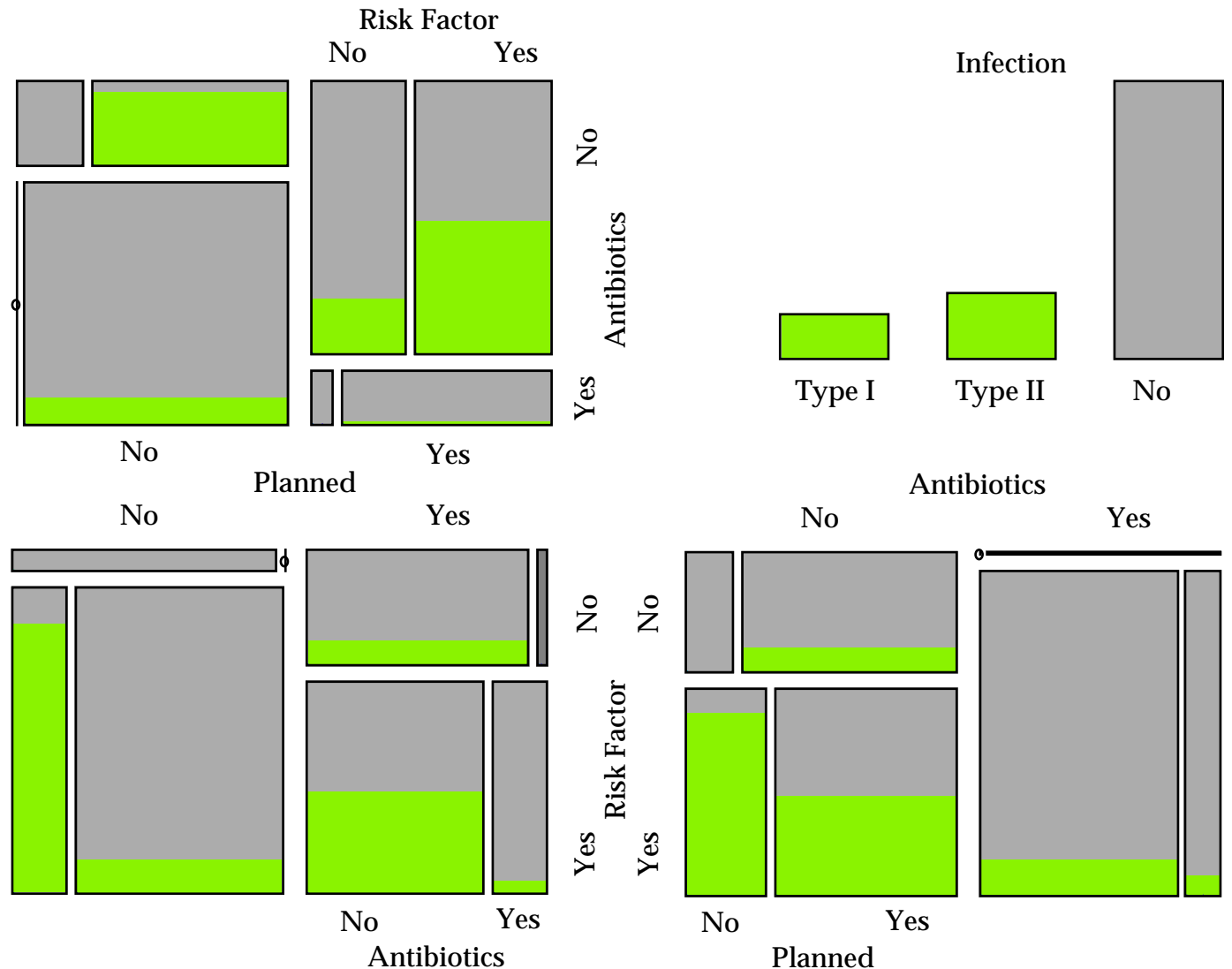
Counts, Proportions, Interactions – A View on Categorical Data

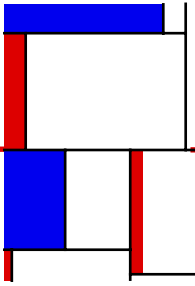
Martin Theus, AT&T Labs Research, Florham Park, NJ
 Adalbert FX Wilhelm, University of Augsburg, Germany

CASE STUDY

(Caesarean Birth)

- Infection
- Risk Factor
- Planned
- Antibiotics





Counts, Proportions, Interactions – A View on Categorical Data

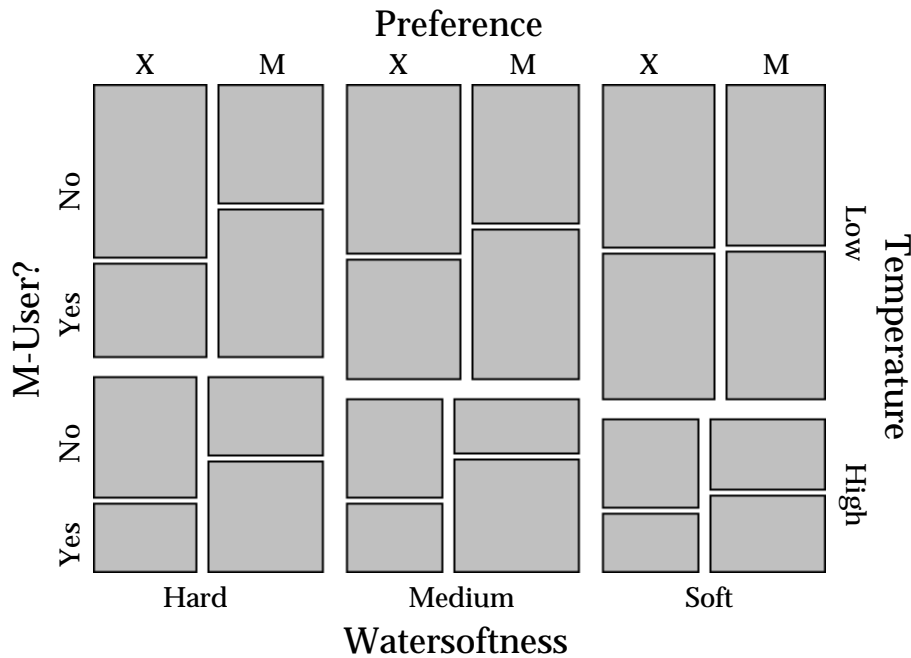
Martin Theus, AT&T Labs Research, Florham Park, NJ
 Adalbert FX Wilhelm, University of Augsburg, Germany

UNDERSTANDING BLACK BOXES

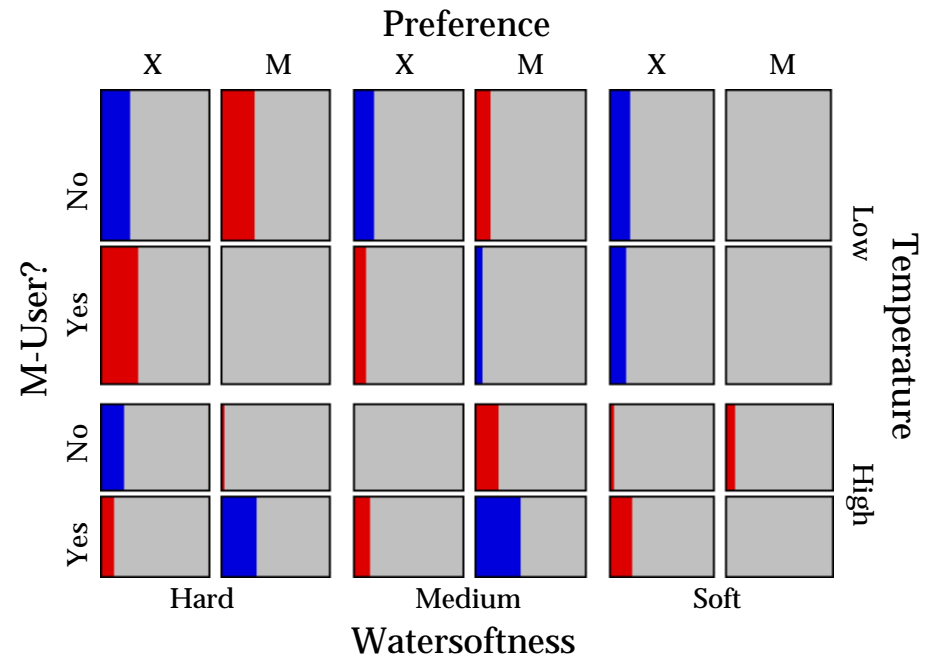
(Detergent Usage)

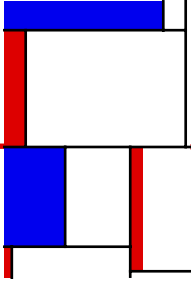
- Watersoftness
- Temperature
- Preference
- M-User?

Raw Data



Model Display (Mutual Independence) with Residual Highlighting





Counts, Proportions, Interactions – A View on Categorical Data

Martin Theus, AT&T Labs Research, Florham Park, NJ
Adalbert FX Wilhelm, University of Augsburg, Germany

CONCLUSIONS

- Only little development in statistical graphs for categorical data
- Many statistical packages hardly support categorical data
- Most plots fail to give a multidimensional insight
- Mosaic Plots are very helpful to understand categorical data
- Interactivity is a keyfeature for an efficient use of Mosaic Plots
- Linked highlighting can display the multivariate structure of datasets easily
- Empty cells often cannot be avoided in a study – graphical displays help
- The structure of empty cells might be very interesting
- Mosaic Plots can serve a modeling and diagnostics of log-linear models