

Prof. Antony Unwin
Dept of Computer-Oriented Statistics and Data Analysis
Institute for Mathematics
University of Augsburg
<http://stats.math.uni-augsburg.de/>

Stochastik IV – Graphical Data Analysis

Exercise Sheet 3: Graphics for categorical variables

Tutorial: Tuesday, 8th November, 2011, 10.00 - 11.30 Uhr, Room 3029

1. Alzheimer's

Three groups of patients (one with Alzheimer's, one with other forms of dementia and a control group with other diagnoses) were studied. Counts are given in the dataset *alzheimer* in the package *coin*. Prepare a graphical summary of plots of each of the three variables, *smoking*, *disease* and *gender*, in a single window. Are the disease groups very different in size? Are there more men or women in the study? How would you describe the distribution of the smoking variable? Do you think the smoking data are likely to be reliable?

2. Slot machines

According to the R help entry this dataset (*vlt* in package *DAAG*) was collected from the three windows of a video lottery terminal in playing the game 'Double Diamond'. There are seven possible symbols that may appear in each window. Draw equally scaled barcharts to see if the distributions of frequencies are the same for each window. Describe any important features.

3. Multiple Sclerosis

The dataset *MSP* in *vcd* provides information on the diagnoses of two neurologists on two groups of patients in a three dimensional array. How do the distributions of the ratings of the neurologists compare? How would you describe their rating patterns? Draw two barcharts with common scaling. Before drawing the barcharts reorder the categories into a sensible ordinal scale instead of the default alphabetic order. (Hint: Initially using the command `melt(MSP)` from the package *reshape2* will put the dataset into a form that is easier to work with.)

4. Occupational Mobility

According to the R help page the *Yamaguchi87* dataset in *vcdExtra* has become a classic for models comparing two-way mobility tables. How do the distributions of occupations of the sons in the three countries compare? How do the distributions of the sons' and fathers' occupations in the UK compare? Are you surprised by the results or are they what you would have expected?

Extra: see next page

Graphic of first year student numbers in mathematics at Augsburg

Abbildung 1: Anfängerzahlen nach Studiengang und Geschlecht

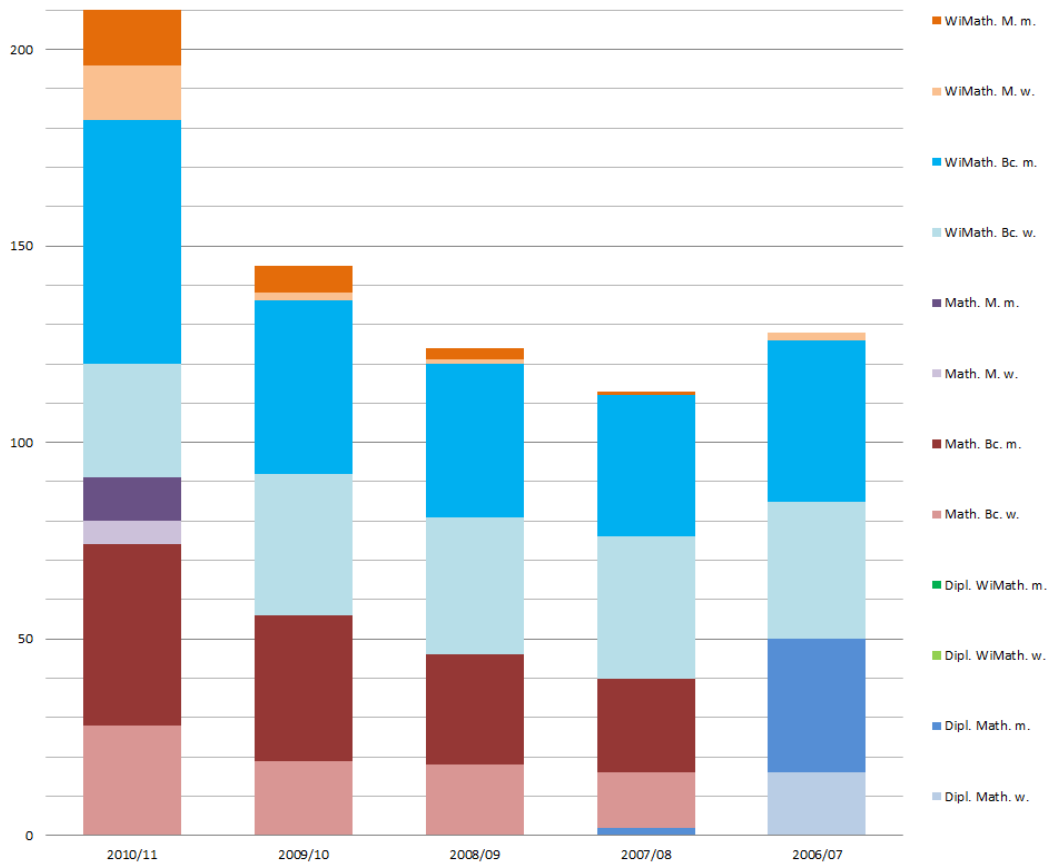


Figure 1: First year student numbers by year and gender for mathematics at Augsburg University

What information can you find in this graphic? Is there potentially other information that could be found in the data? What alternative graphic displays might be worth considering?