Problem Set 1

Working in collaboration with: Christine Malinowski, Kathryn Cuff, Nic Dobbins
Estimated time: 20 hours

Exercise 1.1 *Calculate the median profit for the companies in the US and the median profit for the companies in the UK, France and Germany.*

Median Profits of Companies in the Forbes 2000, by Country

Country	**Median Profit** (in billions)
United States	0.24
United Kingdom	0.205
France	0.19
Germany	0.23
UK, France Germany (combined)	0.21

Exercise 1.2 *Find all German companies with negative profits.*

German Companies with Negative Profits

Company	**Profits** (in billions)
Allianz Worldwide	-1.23
Deutsche Telekom	-25.83
E.ON	-0.73
HVB-HypoVereinsbank	-0.87
Commerzbank	-0.31
Infineon Technologies	-0.51
BHW Holding	-0.38
Bankgesellschaft Berlin	-0.74
W&W-Wustenrot	-0.08
mg technologies	-0.13
Nurnberger Beteiligungs	-0.03
SPAR Handels	-0.40
Mobilcom	-3.62

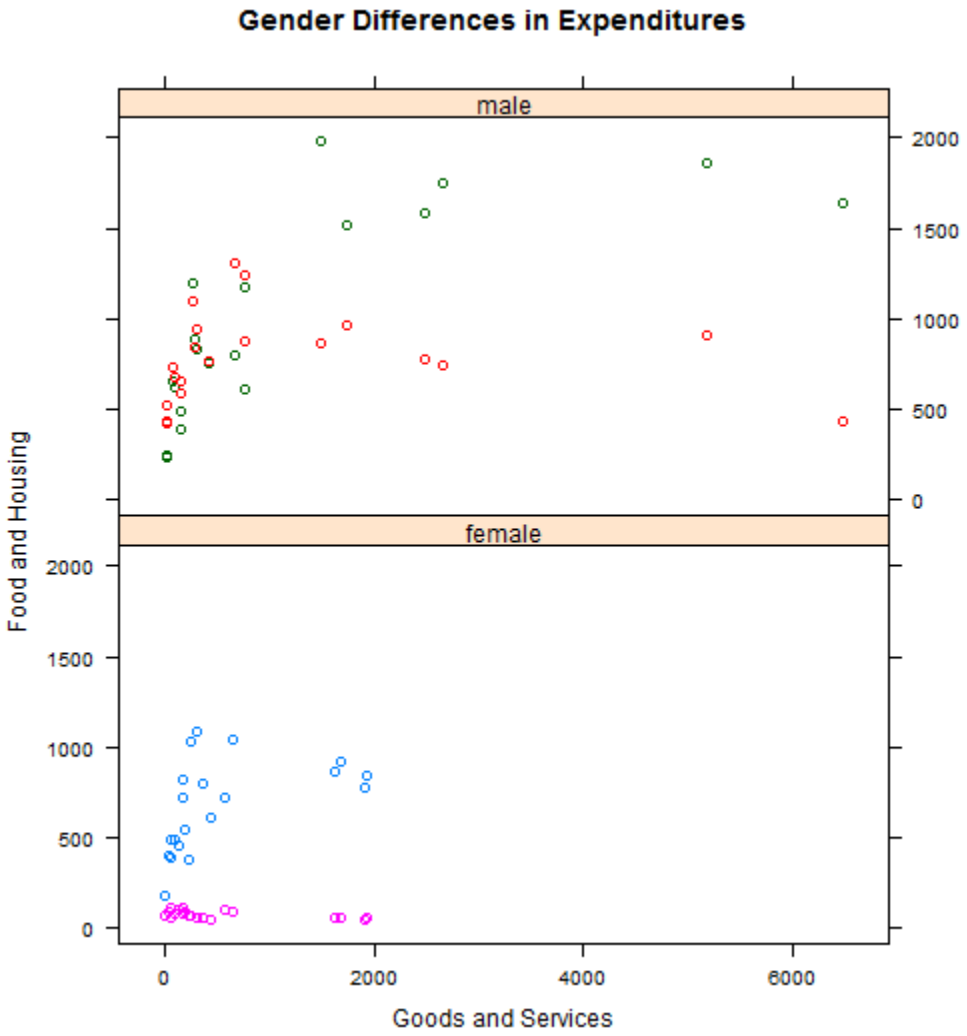
Exercise 1.3 *To which business category do most of the Bermuda island companies belong?*

Most Populous Business Category, by County

Country	**Category** (number of companies)
Bermuda	Insurance (10)

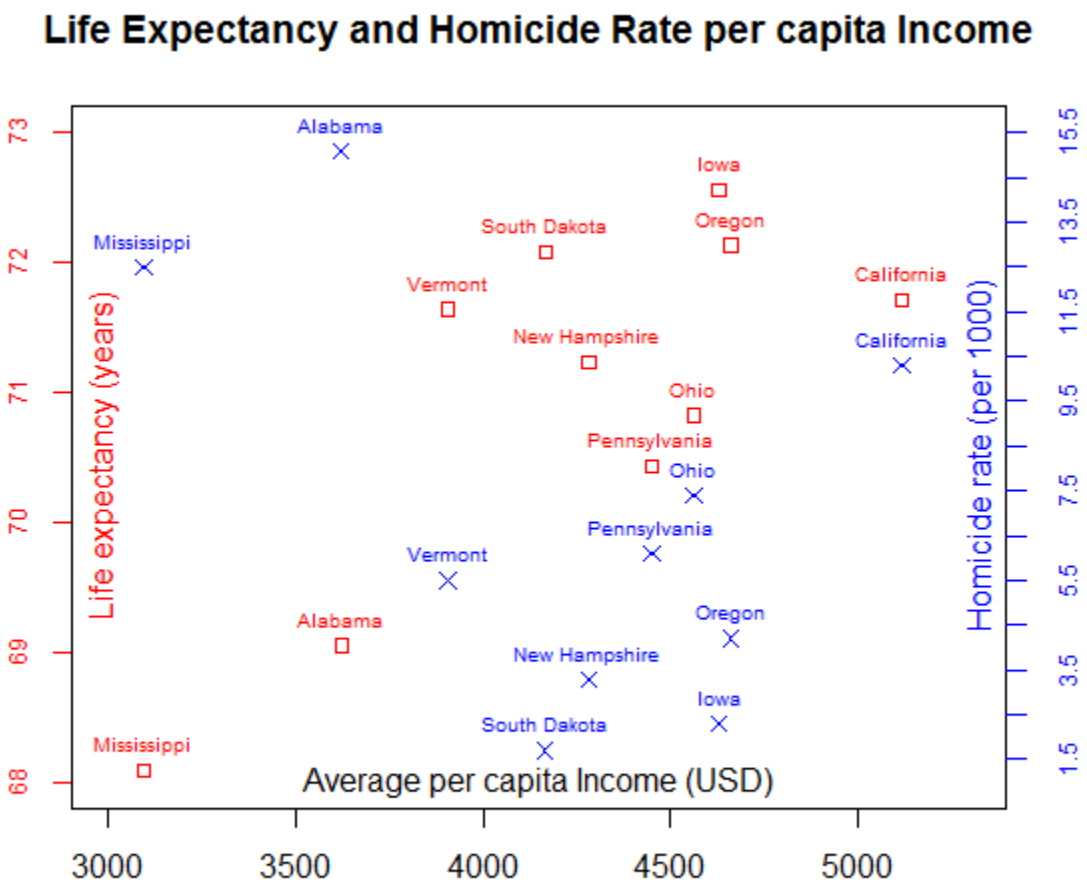
Exercise 1.4 *For the 50 companies in the Forbes data set with the highest profits, plot sales against assets (or some suitable transformation of each variable), labeling each point with the appropriate country name which may need to be abbreviated (using abbreviate) to avoid making the plot look too 'messy'.*

Exercise 2.1 *The aim of the survey (Table 2.3) was to investigate how the division of household expenditure between the four commodity groups depends on total expenditure and to find out whether this relationship differs for men and women. Use appropriate graphical methods to answer these questions and state your conclusions.*



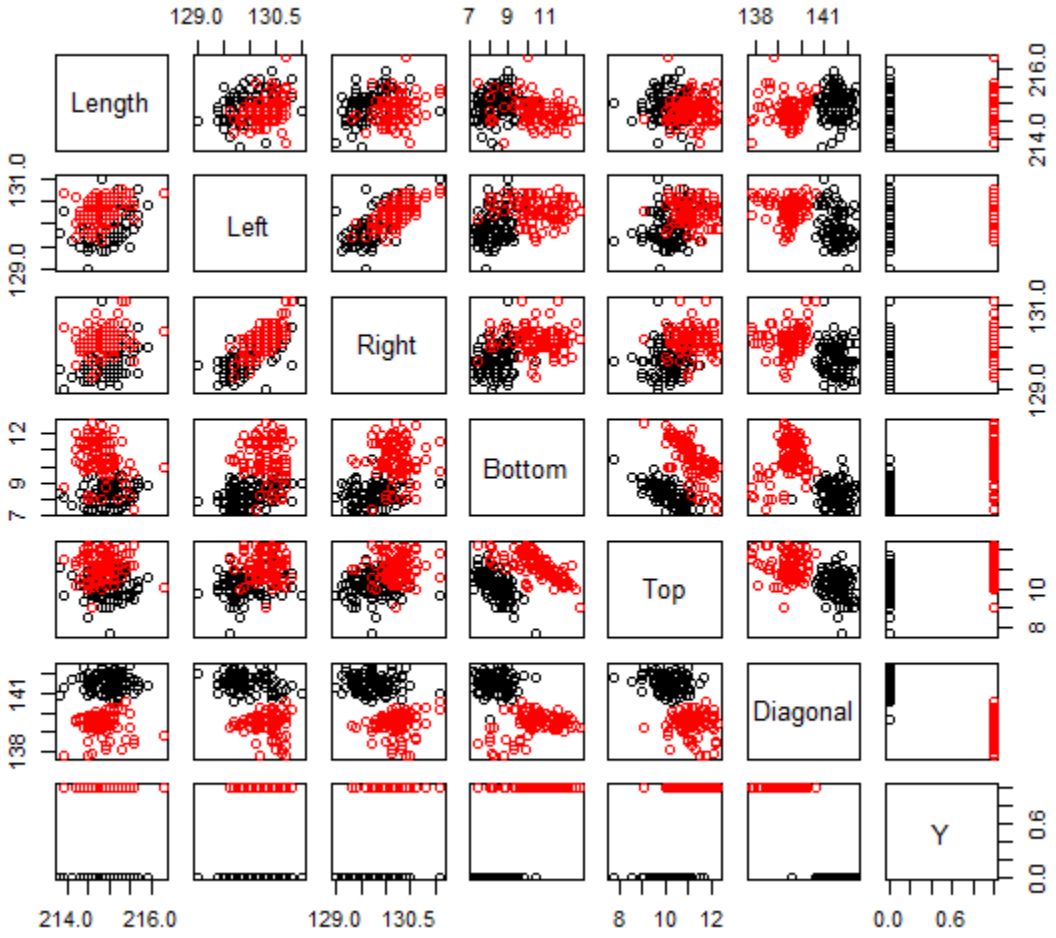
I wish I had more time with this one. I couldn't figure out how to determine which colors indicated each expenditure. It looks like males spend more on housing.

Exercise 2.3 *Instead of doing parts 1 and 2 separately, do the following: Create a single graph that contains two sets of points. The first set of points plots (in red) show the Life expectancy conditional on Per Capita income; the second plots (in blue) show the homicide rate conditional on per capita income. Label each point with the corresponding state name.*



Exercise 2.4 *Use whatever graphical techniques you think are appropriate to investigate whether there is any "pattern' or structure in the data (Table 2.6). Do you observe something suspicious?*

Note: *To clarify: "Left" is the height of the note, measured on the left edge, "Right" is the height of the note, measured on the right edge, Bottom and Top are the distance of the inner frame to the top/bottom border, respectively. Diagonal is the length of the inner diagonal. Don't get too caught up in this problem, but if you notice any unusual patterns, state what you found.*



This matrix shows all the banknote measurements plotted on top of each other, with black indicating known genuine notes and red indicating counterfeit notes. A tight grouping of the red and black points indicates little variation between the measurements and therefore would not be a good measure to determine counterfeit notes, as shown in the length vs. the left and right measurements. The bottom and top margins vs. length, left, and right show a bit more variation, which could be a stronger indication of counterfeit. The diagonal measurement, vs. all other measurements, shows the most variation. The grouping between genuine (black) and counterfeit (red) is clearly separated. This would be the strongest indicator whether a note is counterfeit or not.