DataVizA Tutorial: Plotting two variables

Department of Econometrics and Business Statistics, Monash University

Tutorial 4

Log Scales

1. Why might we display data using a log scale?
2. Using the Swiss Exports data plot a scatterplot of exports in January 1988 against exports in December 2018.
3. Do the same plot but using log scales.
4. Compare these two plots
5. How do you understand the warning message that occurs when the log log scale is used?

Credit default data

The dataset *credit.rds* contains demographic information and repayment history for individuals who may have either defaulted on their credit card payments. The variable of interest is *default* which is equal to 1 for customers who fail to pay their credit card bill and 0 otherwise. More details on the dataset can be found here

6. Using box plots, explore whether the distribution of age is different for the default group and non-default group. Interpret your result.
7. Using box plots, explore whether the distribution of the credit limit (LIMIT_BAL) is different for the default group and non-default group. Interpret your result.
8. Suppose you work for the credit card company. In a business meeting a colleague suggests that the credit card limit could be useful in predcting when a customer defaults.
9. Suppose you work for the credit card company. In a business meeting a colleague looks at the plot and suggests that low credit card limits are causing defaults. They suggest to raise the credit card limit of all customers. Is this a good idea?