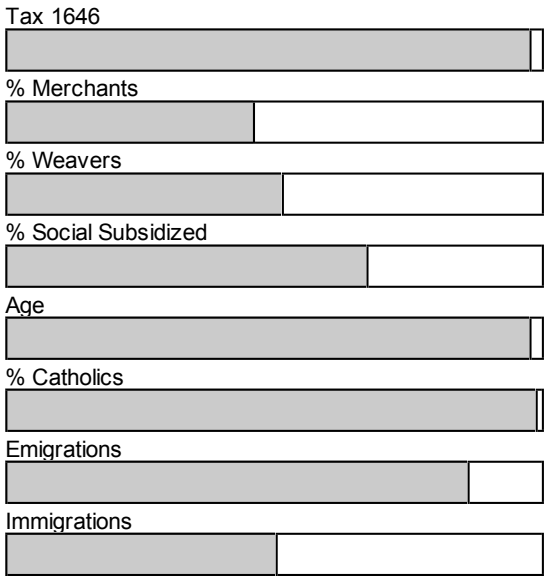
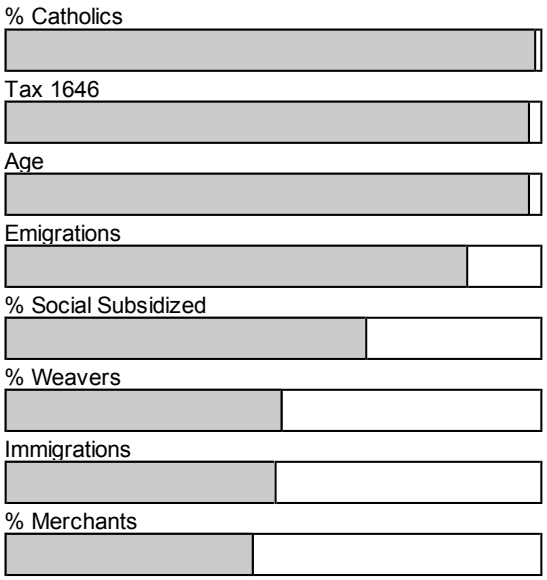
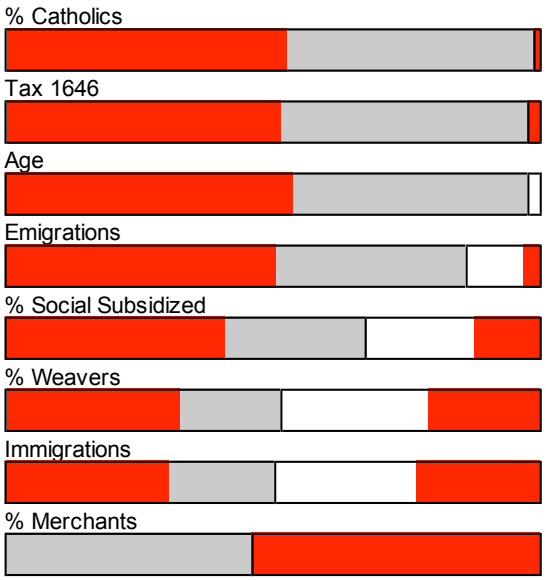
Missing Data: Missing Value Plot

- The missing value plot is as simple as effective (in an interactive environment)
- For each variable the amount of non-missing data and missing data is represented by a bar each which add up to 100%
- Simple selections of the bars can already tell us much about the structure of the missingness

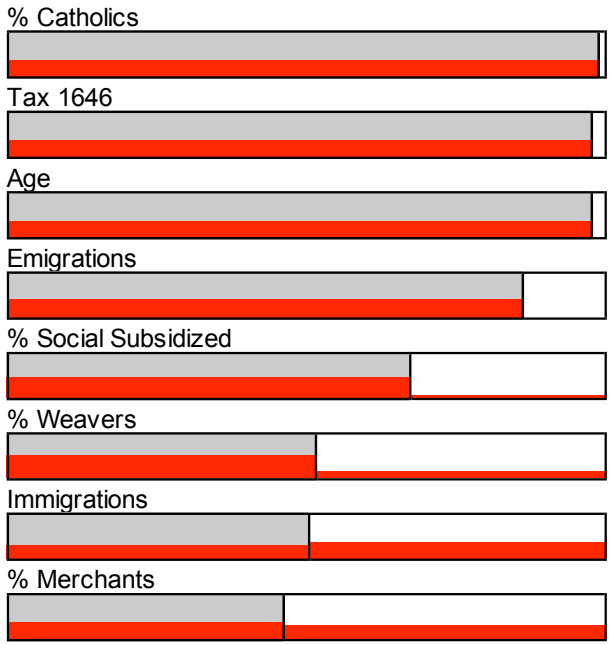


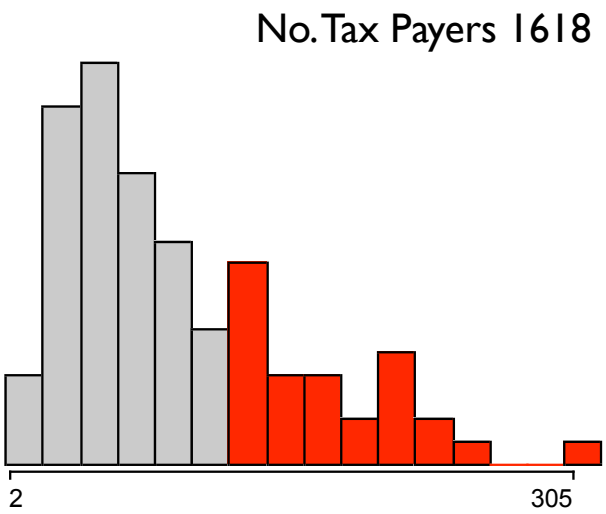




Missing Data: Missing At Random

- There are several concepts of describing the structure of missingness in a dataset – MCAR can hardly ever be found in real data
- The Missing At Random property is given for a certain subset of variables when the MCAR property holds true on this subset
- We usually want the MAR property to hold true for the variables in a model
- The missing value plot in conjunction with, e.g., a histogram can show the presence of this property very easily

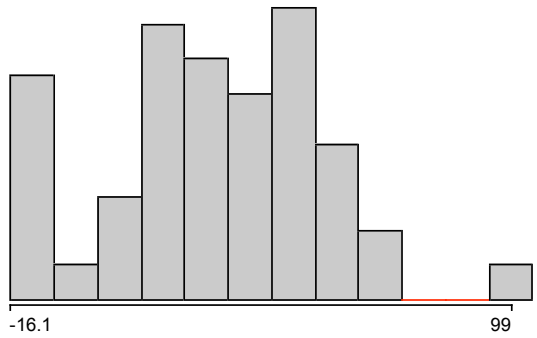




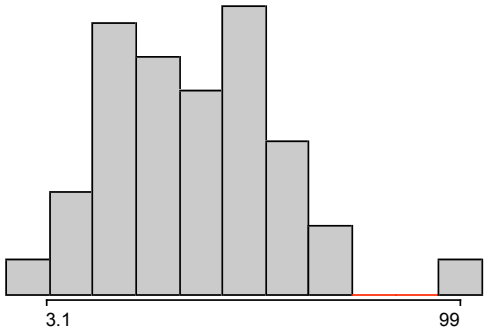
Missing Data: Enhancing Plots

- When working with graphics, the handling of missing values is relatively easy
- In a barchart we just add an extra category that shows the missing values
- Other plots are not that obvious to extend but with a little care, misinterpretations can be avoided
- Example: Histogram

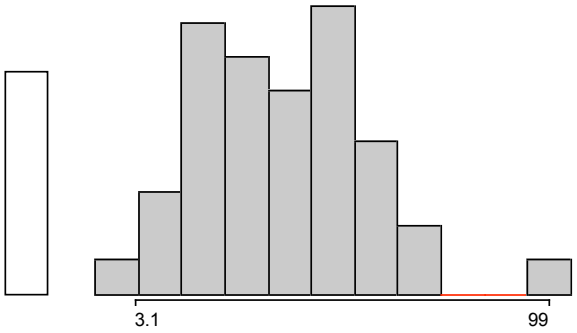
a) add an extra value



b) ignore missings



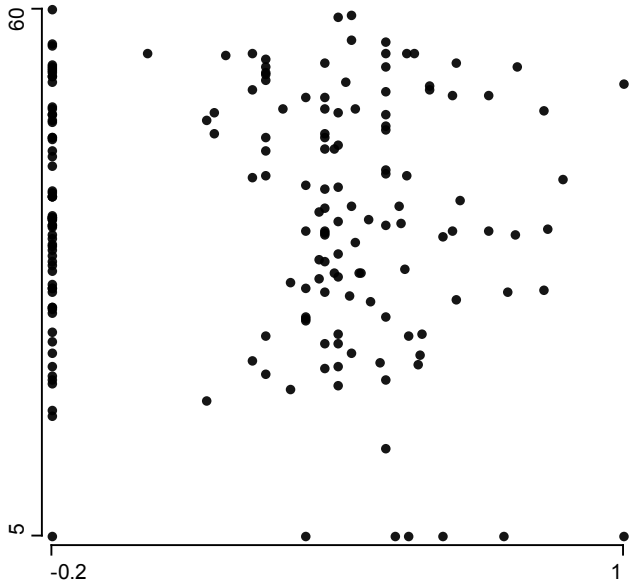
c) show a missing group



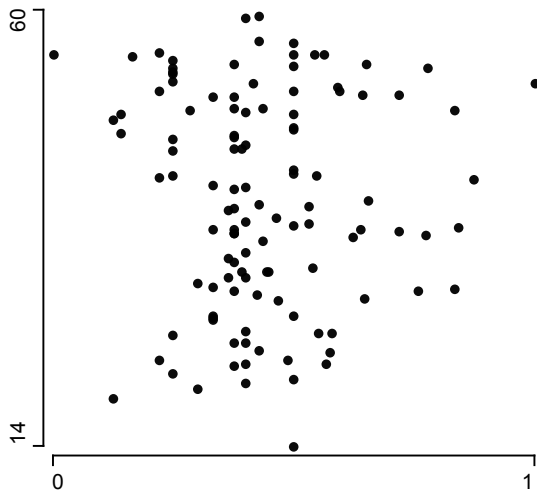
Enhancing Plots: Scatterplots

- Scatterplots show two variables at a time, thus we have a good chance to plot at least one dimension
- Again, we can choose from three ways to handle the missings
- Although variants a) and c) look very similar, the explicit positioning of the missing values outside the plot window makes any misinterpretation far less likely

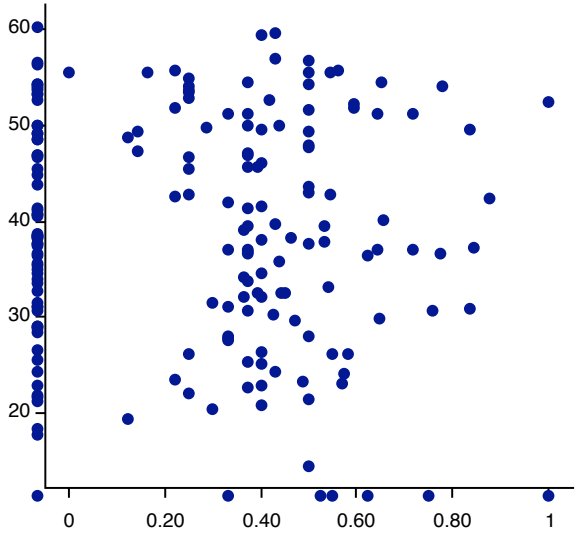
a) add an extra value



b) ignore missings

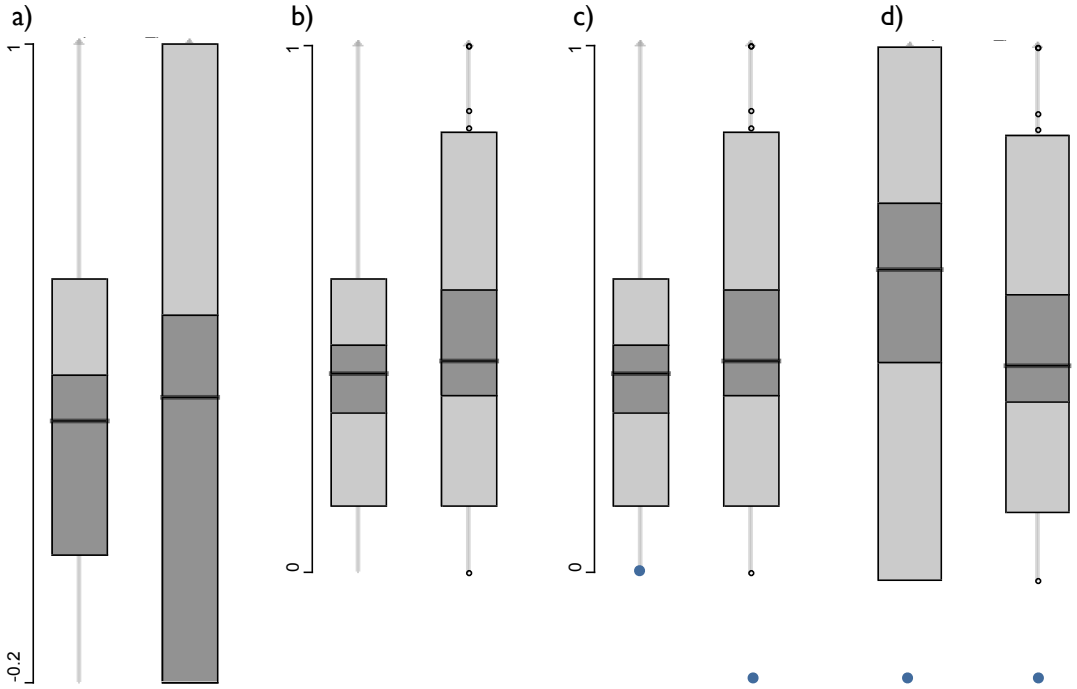


c) show a missing group



Enhancing Plots: Boxplots

- Incorporating missing values in boxplots offers quite some degrees of freedom to end up in misinterpretations as we do only see the summaries of the data
- We can essentially compare four different versions
  a) add missings to the data (e.g. minimum-0.2range)
  b) ignore missings
  c) add missings to the plot (e.g. minimum-0.2range) but don't alter the boxplot
  d) same as c) but with individual scale



Enhancing Plots: Parallel Coordinate Plots

- Missing values in parallel coordinate plots result in broken lines
- If values of an observation are missing for both neighboring axes, there needs to be some marker that represents this ‘orphan’ value