Some things done by hand you should not do include

- editing tables and figures manually
- downloading data manually without documenting what you do.
- pointing and clicking on a screen
- getting tables from others.

Reproducibility is valued because

- it is necessary to ensure that the analysis has been correctly carried out and the conclusions are justified.
- so that changes to the analysis can be investigated.
- it can lead to fewer errors.

Reproducibility does not ensure that the analysis has been done correctly.

Exploratory Data Analysis

We will discuss

1. Correlation
2. Principal Components analysis

In R we will be using ~~detoff~~ data frames.

A data frame is like a matrix for ~~hodi~~ holding data, the rows correspond to observations and the columns are covariates for a given observation. The columns can have different data types e.g. real, integers, logical (TRUE/FALSE) charaters "ABC...F". Another type is factor, which represents categorical data (e.g. sex (M/F), insurance group, country, smoking/non-smoking)

You can create your own data frame, download them from the web, or use built-in data frames already in R.

A data frame whose values are separated by spaces can be uploaded using

read.table ("<filename>.txt")

A data frame whose values are separated by commas can be uploaded using

read.csv ("<filename>.csv")

CSV stands for comma separated value