## Project A: blood pressure

In a small dose-finding study of a blood pressure reducing drug, 40 high blood pressure patients (systolic blood pressure at least 150 mmHg) were randomised over 4 concentrations of the active compound (arginine) in the drug: 0, 2, 5 and 10 mg per day. The outcome is the systolic blood pressure reduction after 2 months, measured in mmHg. The data can be read as shown in the next chunck of R code: load("BloodPressure.RData")

## Data reading

Read in the dataset and analyze the skim through the data variables.

## Descriptive statistics

* Summarize your data and calculate the following: mean, median, minimum, maximum, first and third quartile (for each variable).
* For the categorical variable existing, calculate a frequency table
* Calculate the correlation coefficient between bp-reduction and dose

## Graphics

* Generate a bar chart of a categorical variable for the gender.
* Generate a bar chart graph with mean bp.reduction in males and females
* Make a histogram of a continuous variable “Dose”, “bp.reduction”.
* Make a scatterplot of 2 continuous variables *Dose* and *bp.reduction*, and add the regression lines for each gender
* Make a boxplot of age and a separate boxplots per Doses (as.factors).

## Outlier detection

* Explore the data for any existing outliers, identify them (do NOT remove them if found).
* What do you think?

## Testing for normality/ homoscedasticity

* Check the normality using two methods
* Check the homoscedasticity using two methods.
* What do you think?

1. Statistical Inference

* Calculate the 90%, 95%, 99% confidence interval for the means of bp.reduction per each Dose.
* How would you describe those inferences and what do you observe in terms of the interval width when request higher confidence (i.e. 99% C.I.)?

1. Hypothesis testing

* We hypothesis that bp.reduction is different between male vs female (in the group that received placebo, i.e. Dose = 0). Assuming normality and homoscedasticity, can you test this hypothesis using statistical hypothesis framework
* Assess whether the previous assumptions have been meet for the test.
* We hypothesis that bp.reduction is “higher” in the group receiving Dose = 10 compared to the control (Dose =0). Can you test this hypothesis assuming heteroscedasiticy
* Assess the previous test assumption
* We hypothesis that bp.reduction is different between the different doses (ignoring the gender). Can you perform comparison between the different groups, after assessing the assumptions and performing post-hoc testing.

1. Linear model

* Fit a linear regression to the data and interpret the regression coefficient (for the same hypotheses mentioned above)
* Calculate and interpret a 95% confidence interval of the regression slope
* Estimating the average blood pressure reduction for patients that would receive 3mg/day of the medication (bonus).