# Basic Inferencial Data Analysis

Vijayeshswari Palakurthi

6/5/2020

**Data and packages:** Let's load the "ToothGrowth" data (available in the datasets package) and some relevant library packages used in this project

```
library(datasets)
library(ggplot2)
library(reshape2)
library(dplyr)
data(ToothGrowth)
```

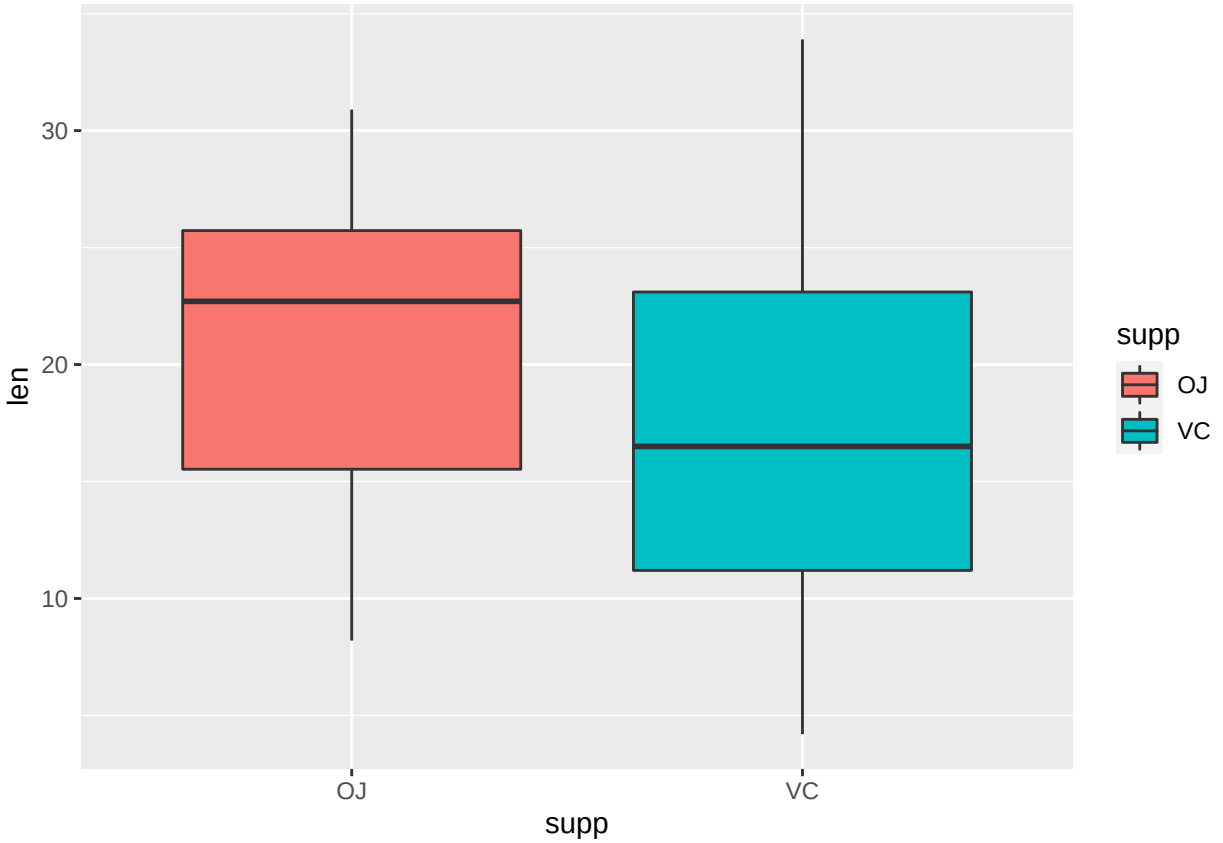
Here are some explanations from the help page of the package:

**Description:** The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice or ascorbic acid (a form of vitamin C and coded as VC).

**Format:** A data frame with 60 observations on 3 variables. [,1] len numeric Tooth length [,2] supp factor Supplement type (VC or OJ). [,3] dose numeric Dose in milligrams/day

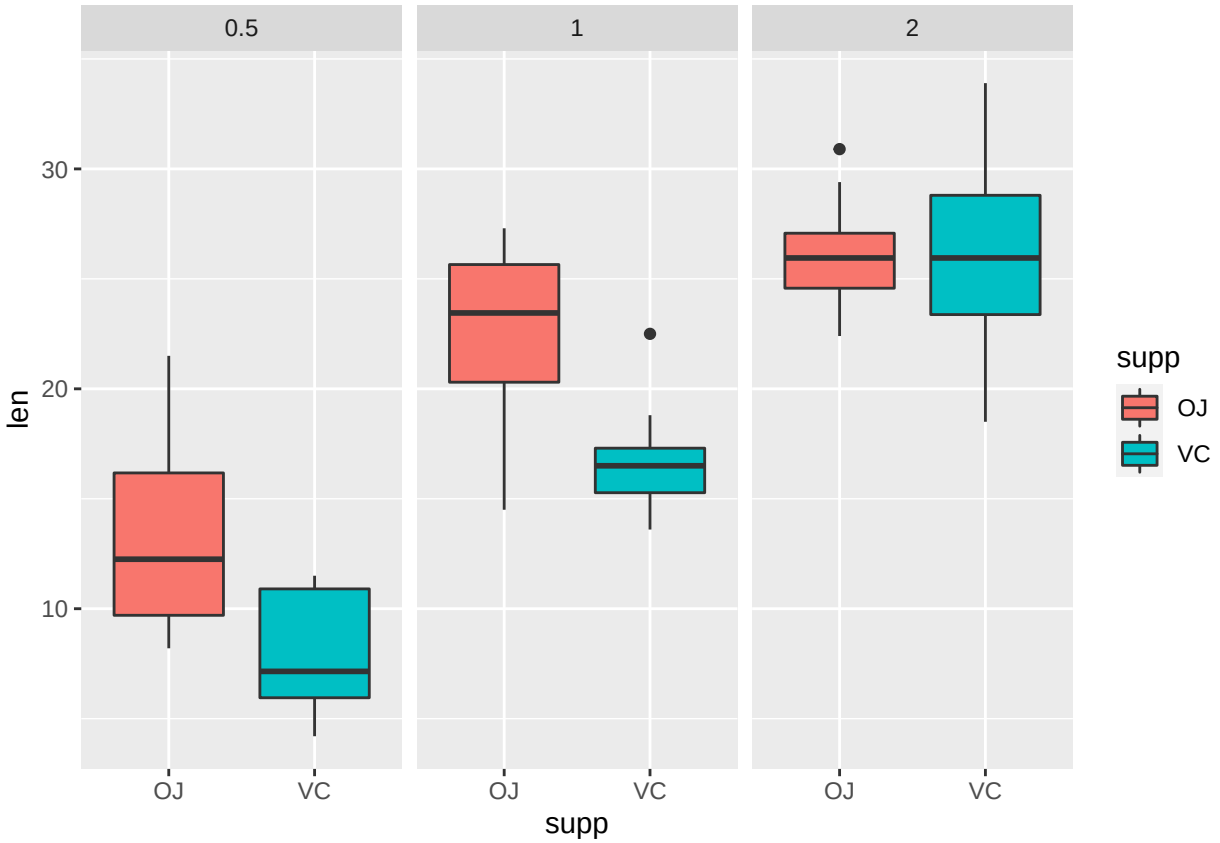
**Exploratory analysis:** Let's create a boxplot of the growth vs. the delivery method.

```
g = ggplot(ToothGrowth, aes(supp, len,fill =supp))
g =g    +geom_boxplot ()
g
```




This is a boxplot using the 2 variables dose and delivery method:

```
g = ggplot(ToothGrowth, aes(supp, len,fill =supp))
g =g    +geom_boxplot ()
g =g    +facet_grid (. ~ dose)
g
```




**Some first observations:** * The growth seems to increase with the dose * The growth seems to be more important when the guinea pig is fed orange juice instead of vitamin C, but this effect seems to dissipate as the dose increases.

**Tooth growth comparison:** We'll use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. As the number of observations is small, let's use Gossett's T tests for our analysis.

**Conclusion:** The difference between the means is significant between the guinea pigs receiving a 0.5mg/day dose and a 2 mg/day dose. In the 2 other cases, the hypothesis between a difference in the means can't be confirmed when using a 95% confidence interval.