Basic Data Analysis of the TootGrowth data

Summary

The objective of this project is to analyze the ToothGrowth data in the R datasets packagem, including testing hypothesis about the relationships between the variables.

Data Processing

```
# Loading the data and packages
library(lattice)
library(datasets)
data(ToothGrowth)

# Attaching the data
attach(ToothGrowth)
```

To see how the data is distributed and what are the variables, it was used head(), str() and unique() functions

```
# Inspectioning the data
head(ToothGrowth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

To facilitate the manipulation of the dose and supp variables, both were converted to factors

```
dose_f <- factor(dose, levels = c(0.5, 1, 2),
                 labels = c("0.5 mg/day","1 mg/day","2 mg/day"))
supp_f <- factor(ToothGrowth$supp, levels = c("OJ", "VC"),
                 labels = c("Orange Juice","Vitamine C"))
```

Exploratory Analysis

Summaries

It was made a summary of the mean, variance and standard deviation for Growth in length by concentracion of the supplement

```
summary_conc <- data.frame("Mean" = tapply(ToothGrowth$len, dose_f, mean),
                           "Variance"=tapply(ToothGrowth$len, dose_f, var),
                           "SD" = tapply(ToothGrowth$len, dose_f, sd))
print(summary_conc)
```

```
##              Mean Variance       SD
## 0.5 mg/day 10.605 20.24787 4.499763
## 1 mg/day   19.735 19.49608 4.415436
## 2 mg/day   26.100 14.24421 3.774150
```

It was made a summary of the mean, variance and standard deviation for Growth in length by type of the supplement

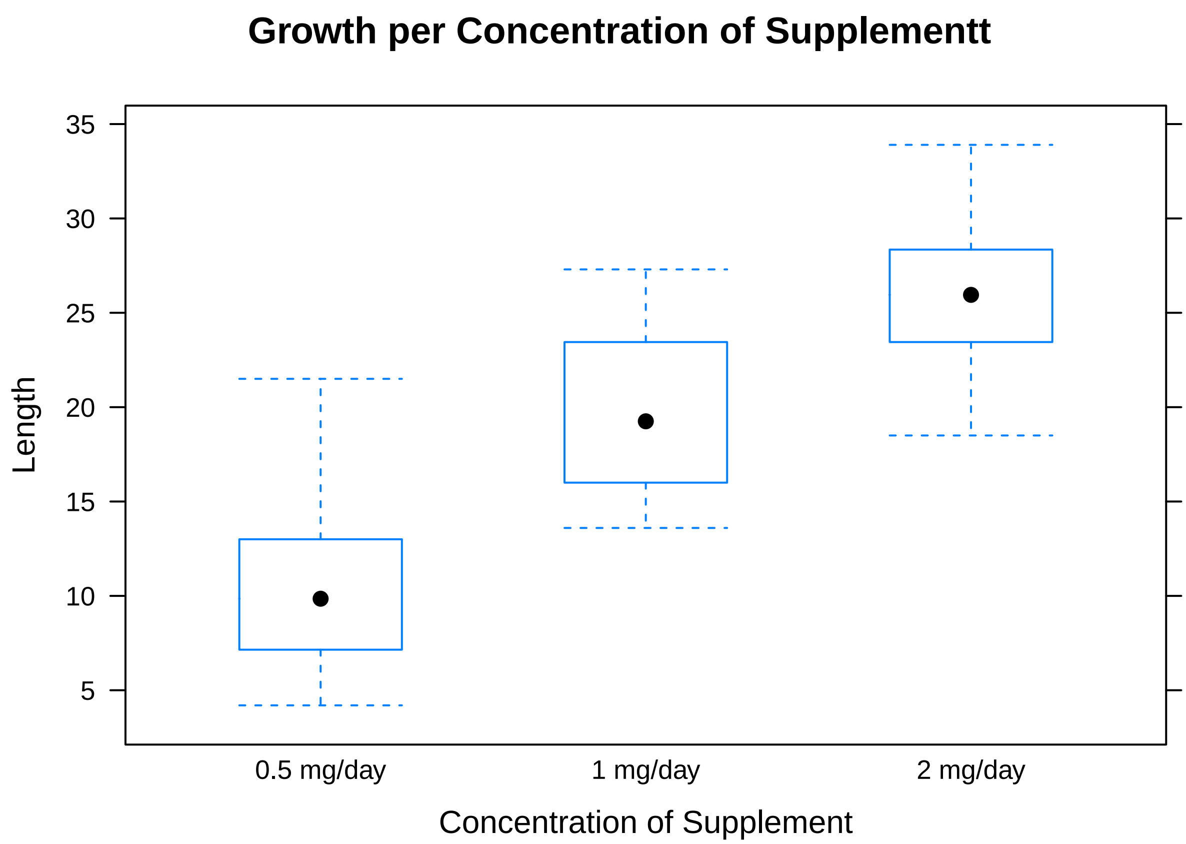
```
summary_conc <- data.frame("Mean" = tapply(ToothGrowth$len, dose_f, mean),
                           "Variance"  = tapply(ToothGrowth$len, dose_f, var),
                           "SD" = tapply(ToothGrowth$len, dose_f, sd))
print(summary_conc)
```

```
##              Mean Variance       SD
## 0.5 mg/day 10.605 20.24787 4.499763
## 1 mg/day   19.735 19.49608 4.415436
## 2 mg/day   26.100 14.24421 3.774150
```

Plots

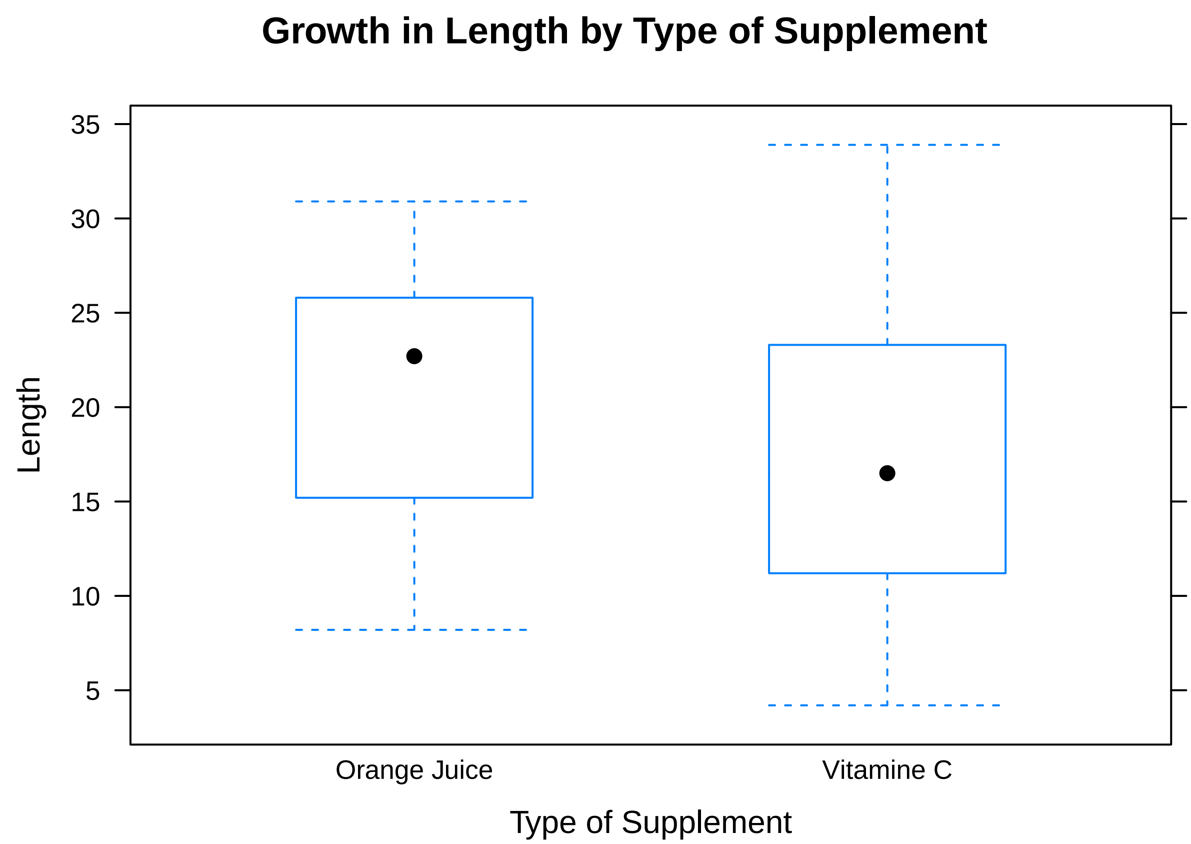
It was made a boxplot of the Growth in length by concentracion of the supplement

```
bwplot(len~dose_f,
   ylab="Length", xlab="Concentration of Supplement",
   main="Growth per Concentration of Supplementt")
```



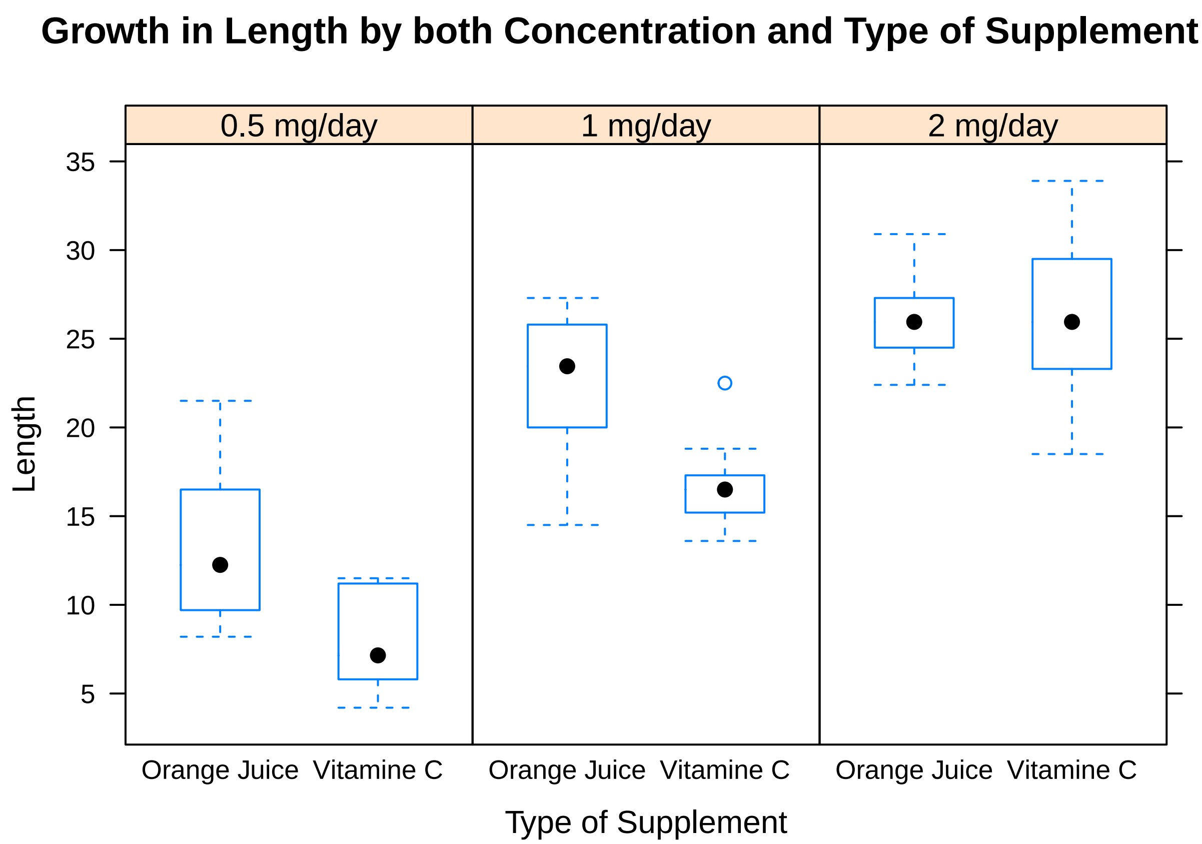
It was made a boxplot of the Growth in length by type of the supplement

```
bwplot(len~supp_f,
   ylab="Length", xlab="Type of Supplement",
   main="Growth in Length by Type of Supplement")
```



It was made a boxplot of the Growth in length by both concentration and type of the supplement

```
bwplot(len~supp_f|dose_f,
   ylab="Length", xlab="Type of Supplement",
   main="Growth in Length by both Concentration and Type of Supplement", layout=c(3,1))
```



Hypothesis Testing

Hypothesis Testing for the difference in Growth in Length by the type of supplement

H0 = There is no difference in Growth in Length by type of supplement H1 = There is a difference in Growth in Length by type of supplement

```
# Testing
t.test(formula = len ~ supp_f, paired = FALSE, data = ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp_f
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group Orange Juice   mean in group Vitamine C
##                   20.66333                   16.96333
```

As the p-value is more than 0.05, we conclude that with a 95% confidence, there is no difference in Growth in Length by type of supplement

Hypothesis Testing for the difference in Growth in Length by the concentration of supplement

Although, it could be tested the differences in length between each pair of concentration of supplement, the plots previously made suggest that the differen between the 0.5 and 2.0 concentration is great enough that is not necessary in this case, so it would be tested only the differences between the other 2 pairs.

_ Hypotthesis Testing for the difference in Growth in length between the 0.5 and 1.0 concentrations

H0 = There is no difference in Growth in Length between the 0.5 and 1.0 concentrations H1 = There is a difference in Growth in Length between the 0.5 and 1.0 concentrations

```
diff_0.5_1.0 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
t.test(len ~ dose, paired = F, var.equal = F, data = diff_0.5_1.0)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

As the p-value is less than 0.05, it can not be refused the null hypothesis, se we conclude that with a 95% confidence, there is a difference in Growth in Length between the 0.5 and 1.0 concentrations of the supplement

_ Hypotthesis Testing for the difference in Growth in length between the 0.5 and 1.0 concentrations

H0 = There is no difference in Growth in Length between the 1.0 and 2.0 concentrations H1 = There is a difference in Growth in Length between the 1.0 and 2.0 concentrations

```
diff_1.0_2.0 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))
t.test(len ~ dose, paired = F, var.equal = F, data = diff_1.0_2.0)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

As the p-value is less than 0.05, it can not be refused the null hypothesis, se we conclude that with a 95% confidence, there is a difference in Growth in Length between the 1.0 and 2.0 concentrations of the supplement

Conclusion

_ The growth in length is associated with an increased concentration of the supplement, however the type of supplement does not seem to produce a significant difference in the length.