Statistical Inference Course Project part 2

Louis

9/20/2020

Coursera Statistical Inference Course Project

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

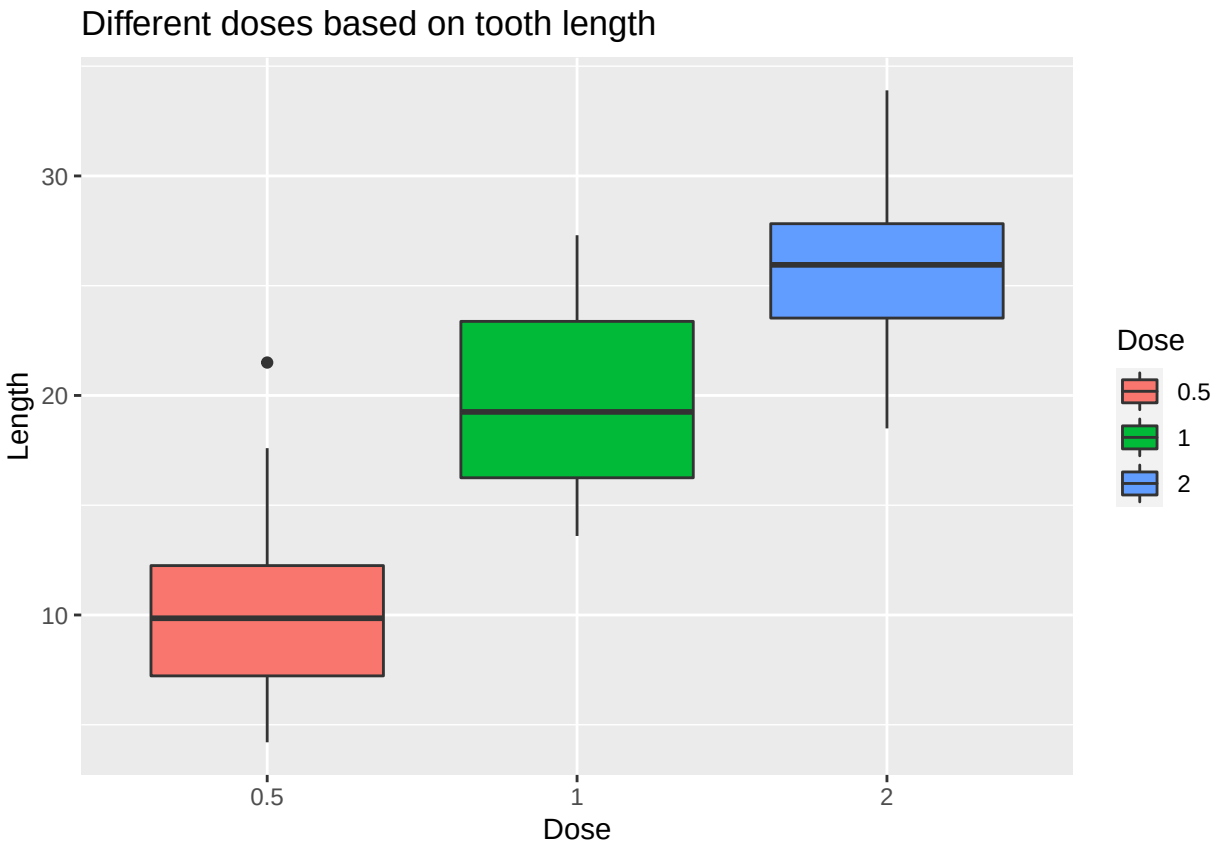
```
library(knitr)
```

Part 2 : Sorting out the Tooth Growth Data

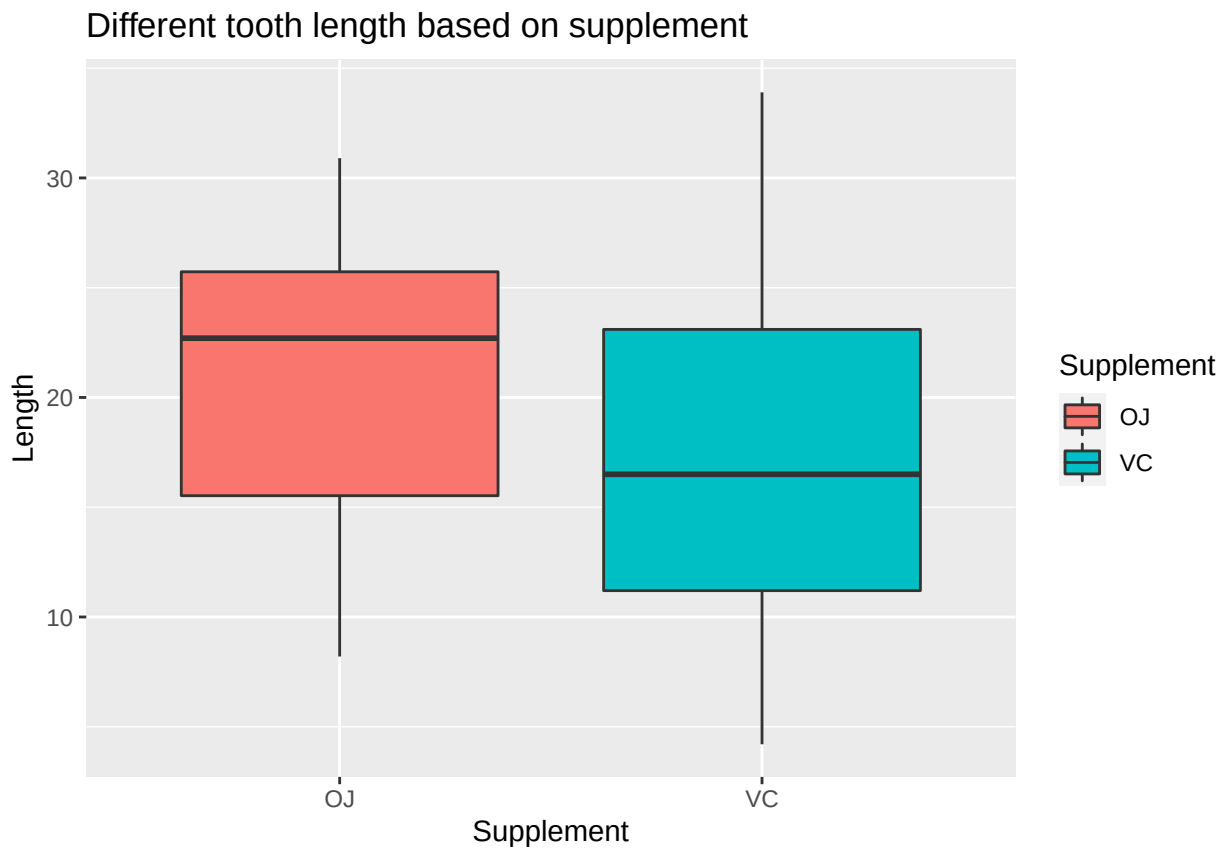
```
# Load the Data
data <- read.csv("ToothGrowth.csv")
# Renaming the columns for more easier understanding
data <- rename( data,Number = X, Length = len, Supplement = supp, Dose = dose)
# Due to there being an select type of Doses, change it to a factor
data$Dose <- as.factor(data$Dose)
```

Plotting the data to make more sense of it

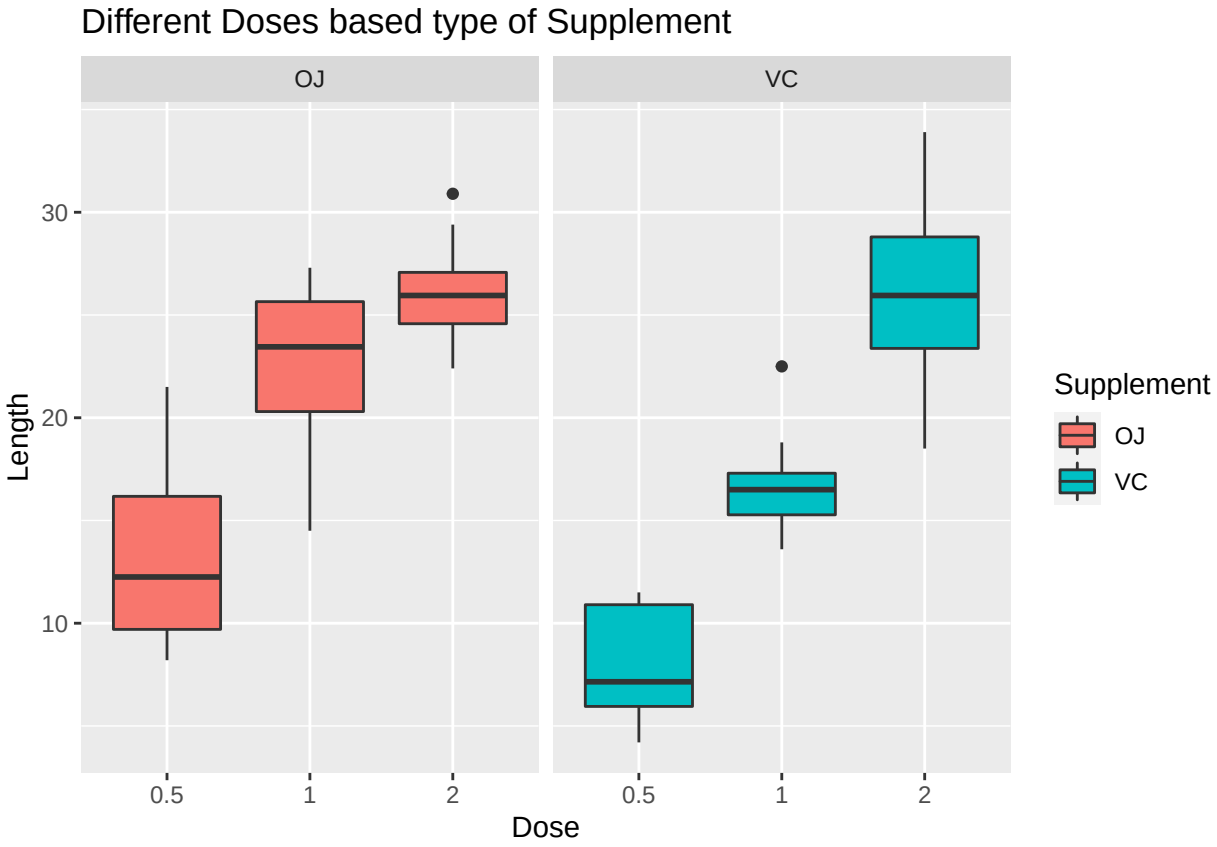
```
theplot1 <- ggplot(data = data, aes(x = Dose, y = Length))
theplot1 + geom_boxplot(aes(fill = Dose)) +
        labs(title = "Different doses based on tooth length")
```




```
theplot2 <- ggplot(data = data, aes(x = Supplement, y = Length))
theplot2 + geom_boxplot(aes(fill = Supplement)) +
        labs(title = "Different tooth length based on supplement")
```




```
theplot3 <- ggplot(data = data, aes(Dose, Length))
theplot3 <- theplot3 + geom_boxplot(aes(fill = Supplement)) +
        facet_grid(~ Supplement)
theplot3 + labs(title = "Different Doses based type of Supplement")
```



P-values of the data:

If p-value is bigger than level of significance(0.05) then there is a correlation between the data.

```
# Comparing Length and Supplement
t.test(Length~Supplement, data=data)
```

```
##
##  Welch Two Sample t-test
##
## data:  Length by Supplement
## 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 OJ mean in group VC
##         20.66333         16.96333
```

```r
# There is no correlation with different supplements and tooth length

# Comparing length with 0.5mg and 2mg doses
t.test(data$Length[data$Dose == 0.5],data$Length[data$Dose == 2], data=data)
```

```
##
##  Welch Two Sample t-test
##
## data:  data$Length[data$Dose == 0.5] and data$Length[data$Dose == 2]
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean of x mean of y
##    10.605    26.100
```

```r
# There is a correlation between tooth length and dose(0.5mg and 2mg)

# Comparing length with 0.5mg and 1mg doses
t.test(data$Length[data$Dose == 0.5],data$Length[data$Dose == 1], data=data)
```

```
##
##  Welch Two Sample t-test
##
## data:  data$Length[data$Dose == 0.5] and data$Length[data$Dose == 1]
## 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 of x mean of y
##    10.605    19.735
```

```r
# There is a correlation between tooth length and dose(0.5mg and 1mg)

# Comparing length with 1mg and 2mg doses
t.test(data$Length[data$Dose == 1],data$Length[data$Dose == 2], data=data)
```

```
##
##  Welch Two Sample t-test
##
## data:  data$Length[data$Dose == 1] and data$Length[data$Dose == 2]
## 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 of x mean of y
##    19.735    26.100
```

```
# There is a correlation between tooth length and dose(1mg and 2mg)
```

Conclusion:

There is a direct correlation between the dose and the tooth length.

We can assume that with an larger tooth length, the dose will increase as well.

There is no correlation with the type of supplement and tooth length.