Basic Inferential Data Analysis

Hatem Jasim Hatem

May 13, 2019

```
library(ggplot2)
```

1. Load the ToothGrowth data and perform some basic exploratory data analyses

```
data(ToothGrowth)
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 ...
```

```
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
```

```
tail(ToothGrowth)
```

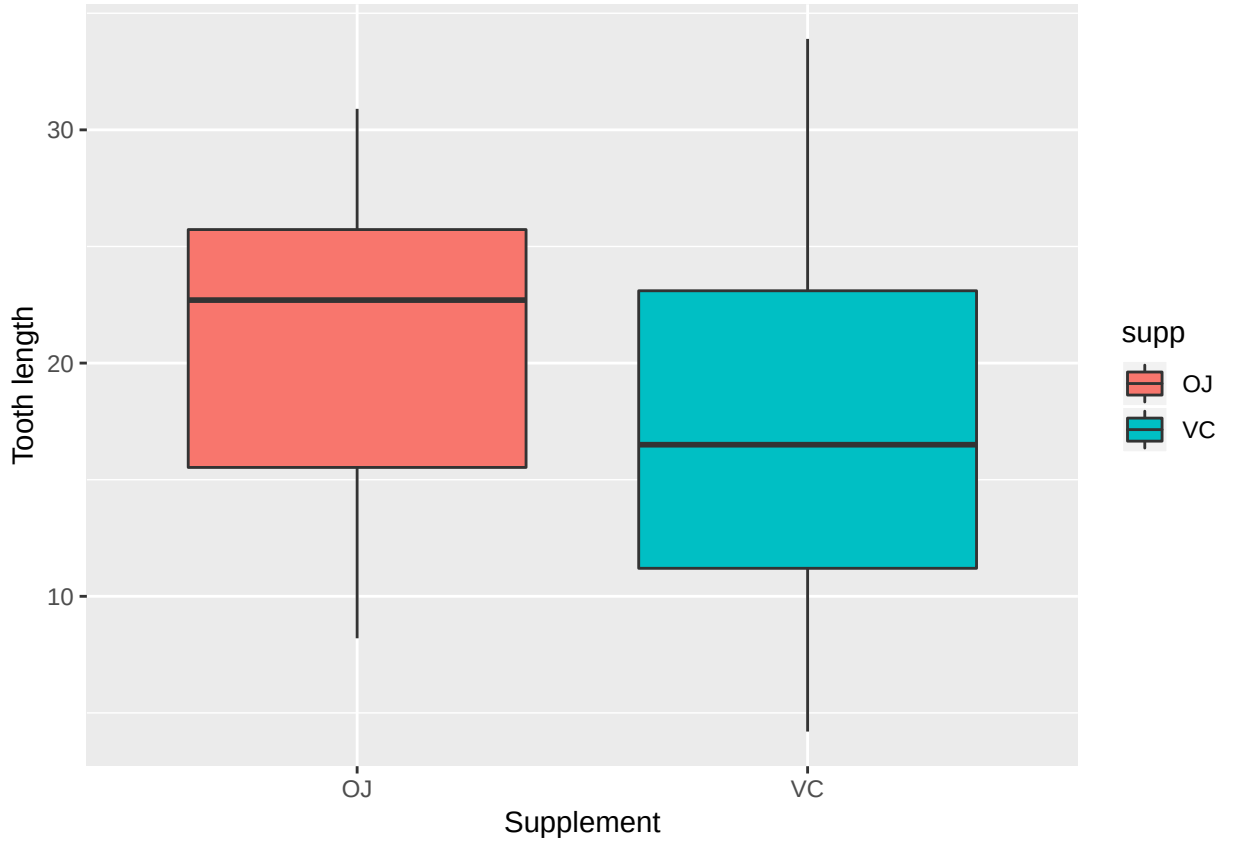
```
##     len supp dose
## 55 24.8   OJ    2
## 56 30.9   OJ    2
## 57 26.4   OJ    2
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

2. Provide a basic summary of the data.

```
summary(ToothGrowth)
```

```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

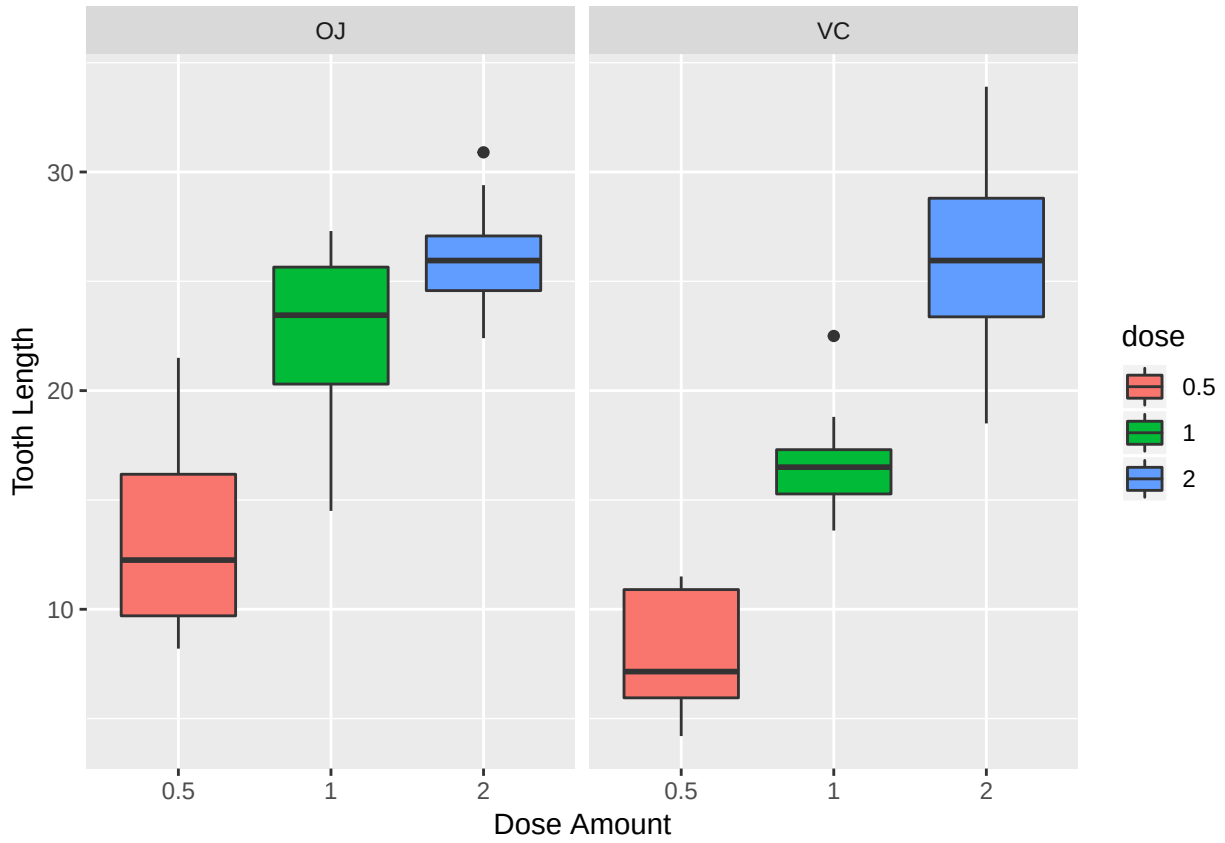
```r
ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp))+
        labs(x="Supplement", y= "Tooth length")
```



3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

```r
# Convert dose to a factor
ToothGrowth$dose<-as.factor(ToothGrowth$dose)

# Plot tooth length vs. the dose amount, broken out by supplement
ggplot(aes(x=dose, y=len), data=ToothGrowth) +
  geom_boxplot(aes(fill=dose))+
  facet_grid(~ supp)+
  labs(x = "Dose Amount", y = "Tooth Length")
```



The figure show there are different in tooth growth. To determine the difference is significante or not used t-test

```
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(1.0,0.5))
# T-test for vitamine c when dose 1 or 0.5 mg
t.test(len ~ dose, ToothGrowth_sub[ToothGrowth_sub$supp =='VC', ])$p.value
```

```
## [1] 6.811018e-07
```

```
# T-test for orange juice when dose 1 or 0.5 mg
t.test(len ~ dose, ToothGrowth_sub[ToothGrowth_sub$supp =='OJ', ])$p.value
```

```
## [1] 8.784919e-05
```

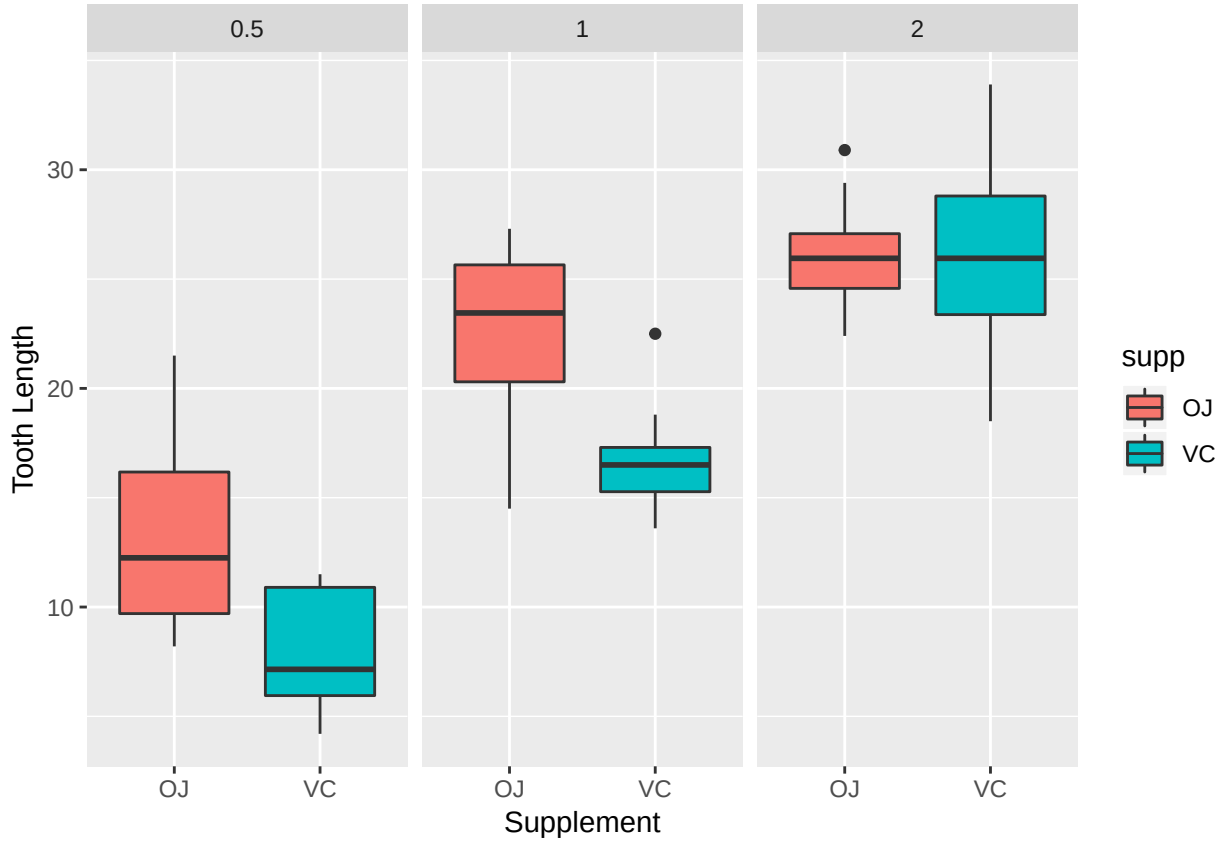
```
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(2.0, 1.0))
# T-test for vitamine c when dose 2 or 1 mg
t.test(len ~ dose, ToothGrowth_sub[ToothGrowth_sub$supp =='VC', ])$p.value
```

```
## [1] 9.155603e-05
```

```
# T-test for orange juice when dose 2 or 1 mg
t.test(len ~ dose, ToothGrowth_sub[ToothGrowth_sub$supp =='OJ', ])$p.value
```

```
## [1] 0.03919514
```

```
# Plot tooth length vs. supplement  method broken out by the dose
ggplot(aes(x=supp, y=len), data=ToothGrowth) +
  geom_boxplot(aes(fill=supp))+
  facet_grid(~ dose) +
  labs(x = "Supplement", y = "Tooth Length")
```



The figure show there are different in tooth growth. To determine the difference is significante or not used t-test

```
# T-test for supplement when dose 0.5 mg
t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == .5, ])$p.value
```

```
## [1] 0.006358607
```

```
# T-test for supplement when dose 1 mg
t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == 1, ])$p.value
```

```
## [1] 0.001038376
```

```
# T-test for supplement when dose 2 mg
t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == 2, ])$p.value
```

```
## [1] 0.9638516
```

4. State your conclusions and the assumptions needed for your conclusions.

Because the alternative hypothesis H_a is cost and may be dangerous therefore set alpha to 0.01 $\alpha = 0.01$. According to p-value of T-test the orange juice (OJ) more effective on tooth growth than vitamine c (VC) when dose is 0.5 or 1 mg but simlar effective in vitamine c when dose 2 mg