Assignment: Statistical Inference Course Project

Part 2: ToothGrowth data analyze.

Overview

Now in the second portion of the class, we're going to analyze the ToothGrowth data in the R datasets package.

TodoList

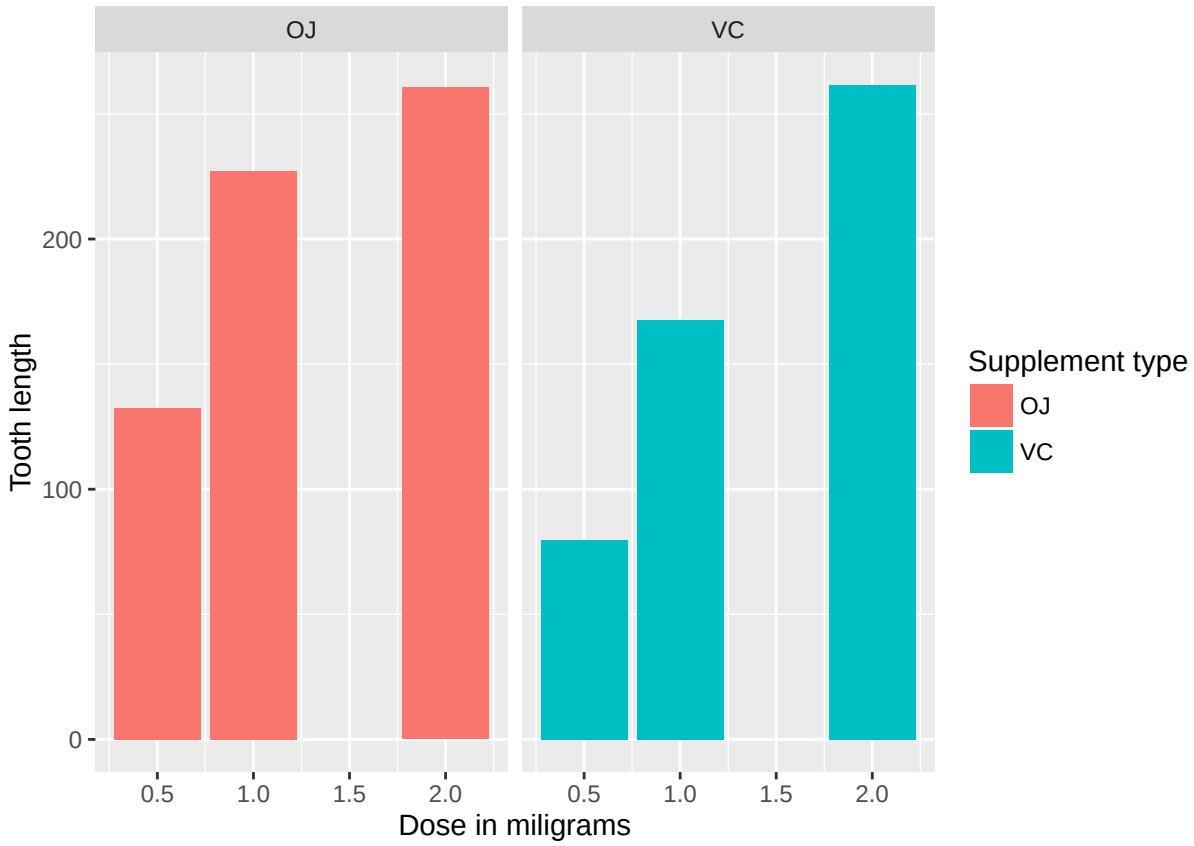
1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.
4. State your conclusions and the assumptions needed for your conclusions.

Part 0 - Load the ToothGrowth data and perform some basic exploratory data analyses

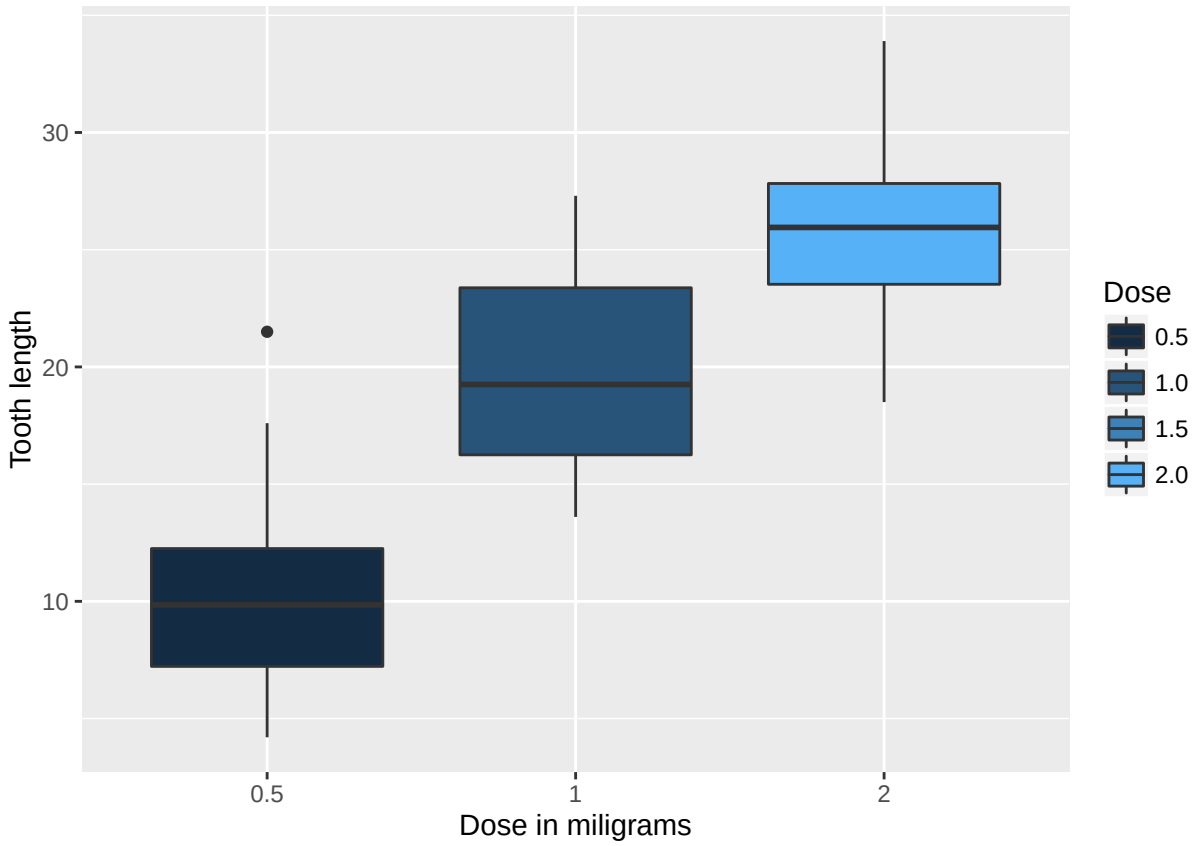
```
set.seed(42)
library(ggplot2)
library(datasets)
data <- ToothGrowth
```

Some usefull plots

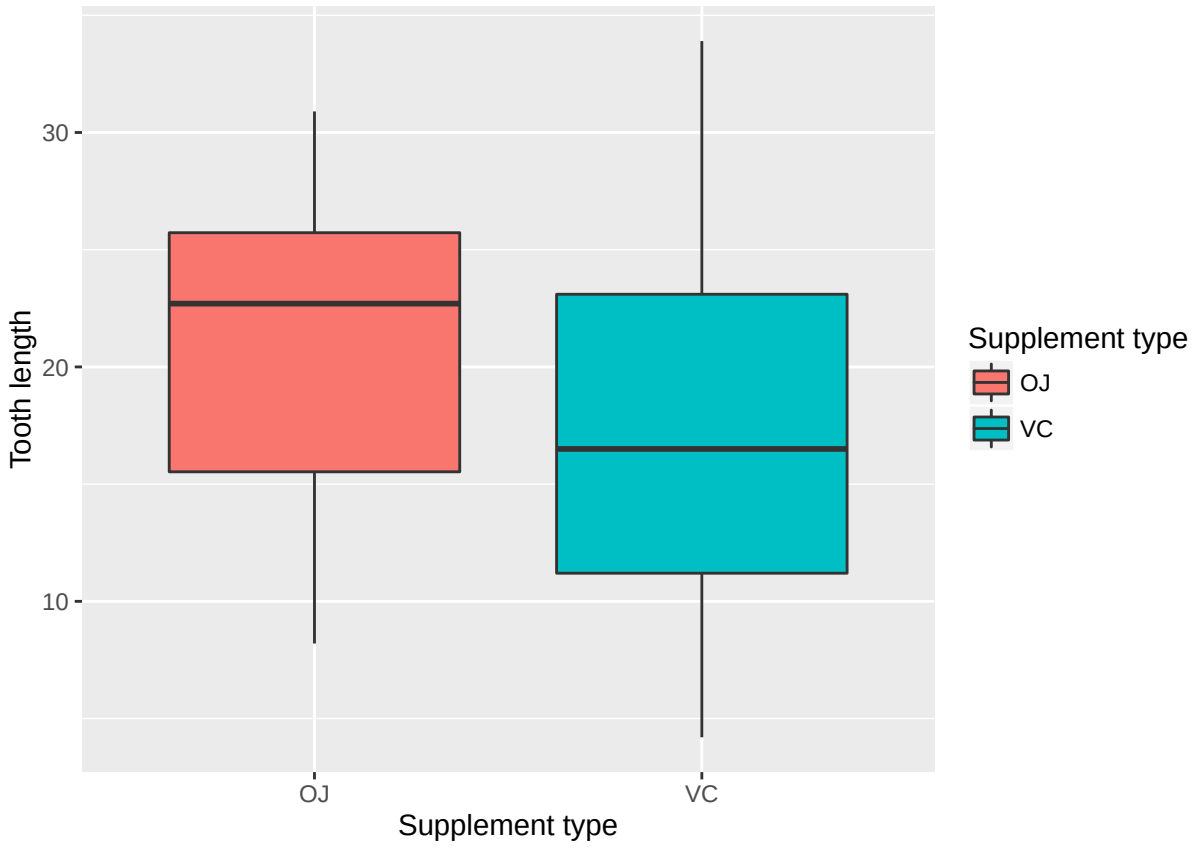
```
ggplot(data = data, aes(x = dose, y = len, fill = supp))  +
  geom_bar(stat = "identity") +
  facet_grid(. ~ supp) +
  xlab("Dose in miligrams") +
  ylab("Tooth length") +
  guides(fill = guide_legend(title = "Supplement type"))
```




```
ggplot(data = data, aes(x = as.factor(dose), y = len, fill = dose)) +
  geom_boxplot(aes(fill = dose)) +
  xlab("Dose in miligrams") +
  ylab("Tooth length") +
  guides(fill = guide_legend(title = "Dose"))
```




```
ggplot(data = data, aes(x = supp, y = len, fill = dose))  +
  geom_boxplot(aes(fill = supp)) +
  xlab("Supplement type") +
  ylab("Tooth length") +
  guides(fill = guide_legend(title = "Supplement type"))
```



Part 1 - Provide a basic summary of the data

```
summary(data)
```

```
##       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
```

```
str(data)
```

```
## '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 ...
```

```
table(data$dose, data$supp)
```

```
##
##       OJ VC
##   0.5 10 10
##   1   10 10
##   2   10 10
```

Part 2 - Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

Since the sample size is small, let's use T distribution.

```
t.test(len ~ supp, data = data)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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
```

Null hypothesis can not be rejected as confindence intervals contain zero and p-value is bigger than the 5% significance level.

Let's also compare the differences between the the different dosis level, since bigger dosis may yield contradicting evidence.

```
t.test(data$len, data$dose)
```

```
##
##  Welch Two Sample t-test
##
## data:  data$len and data$dose
## t = 17.81, df = 59.798, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  15.66453 19.62881
## sample estimates:
## mean of x mean of y
## 18.813333  1.166667
```

Comparing the difference between the two supplements yields convincing evidence to reject the null hypothesis, since the p-value aproximates to 0, and is thus substancially smaller in comparison to the significance level.

Part 3 - State your conclusions and the assumptions needed for your conclusions.

1. There is no convincing evidence that there is a difference between the two type of supplements based on the existing datasets and T statistics.
2. There is convincing evidence that there is a difference between the dosis level, and the growth.