ToothGrowth Data Analysis

Elliot Smith

Overview

In this project, I will be performing some analyses on the ToothGrowth dataset. Using techniques learned in the course, I will draw some conclusions on this dataset.

Load the Date and Basic Analyses

First I will begin by loading the ggplot2 package:

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.1.2
```

The following code will load the data for us. The basic analysis will shed some light into why I chose the dose variable to be a factor:

```
data(ToothGrowth)
data <- ToothGrowth
data$dose <- as.factor(data$dose)
```

Now for a basic analysis:

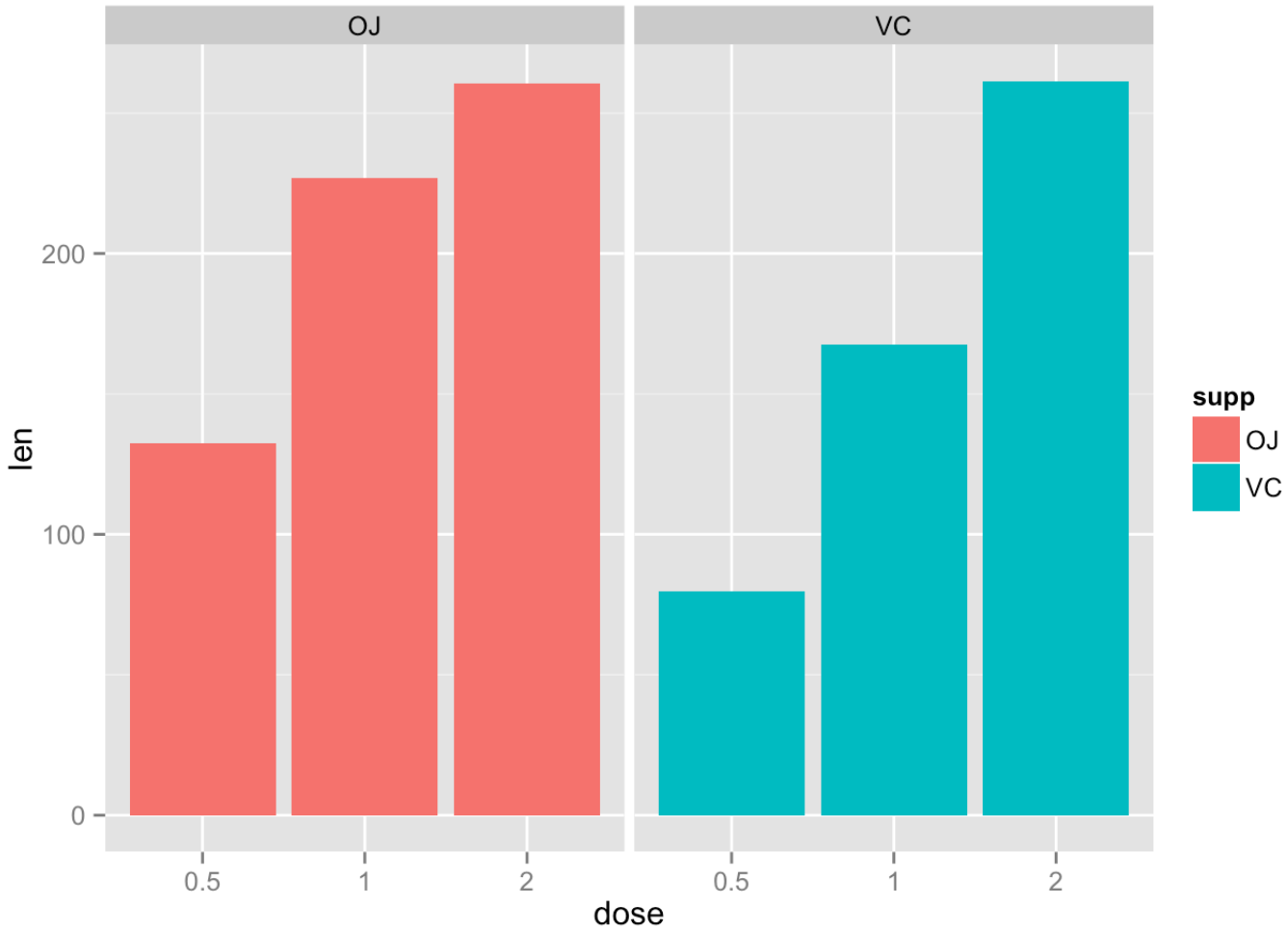
```
str(data)
```

```
## '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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

```
summary(data)
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

```
ggplot(data=data, aes(x=dose, y=len, fill=supp)) +
geom_bar(stat="identity",) +
facet_grid(. ~ supp)
```



Some things of notes from our basic analysis: * The dose variable is best set as a factor of 3 levels: 0.5, 1, 2 * The max value of len is 33.9, while the minimum value is 4.2 * len has a Mean of 18.81 and a Median of 19.25, very close indeed! * There exists a positive correlation between dose levels and tooth length

Advanced Analysis

First Analysis

```
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
```

Summary:

Second Analysis

```
t.test(len ~ supp, data=data[data$dose==0.5, 1:3])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```
t.test(len ~ supp, data=data[data$dose==1, 1:3])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```
t.test(len ~ supp, data=data[data$dose==2, 1:3])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.0461, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

Summary:

Conclusions and Assumptions

Conclusions

- Based on the dataset provided, it is safe to say that increasing the dose will result in an increase in the tooth length (len)
- However, the type of supplement used has a much smaller impact than the dose of the supplement

Assumptions

- The sample is representative of the population
- The variances of the two populations are different