Toothgrowth Report By Charlie Chen

Overview

This is a report on the Toothgrowth dataset in R. We will examine which combination of supp and dosage is most beneficial to toothgrowth.

Report

```
data("ToothGrowth") ## Loading the ToothGrowth dataset
names(ToothGrowth) ## Examine the column names
```

```
## [1] "len"  "supp" "dose"
```

```
## Examine the class of each column
class(ToothGrowth[,1])
```

```
## [1] "numeric"
```

```
class(ToothGrowth[,2])
```

```
## [1] "factor"
```

```
class(ToothGrowth[,3])
```

```
## [1] "numeric"
```

```
head(ToothGrowth) ## Get a peek at the dataset
```

```
##    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
```

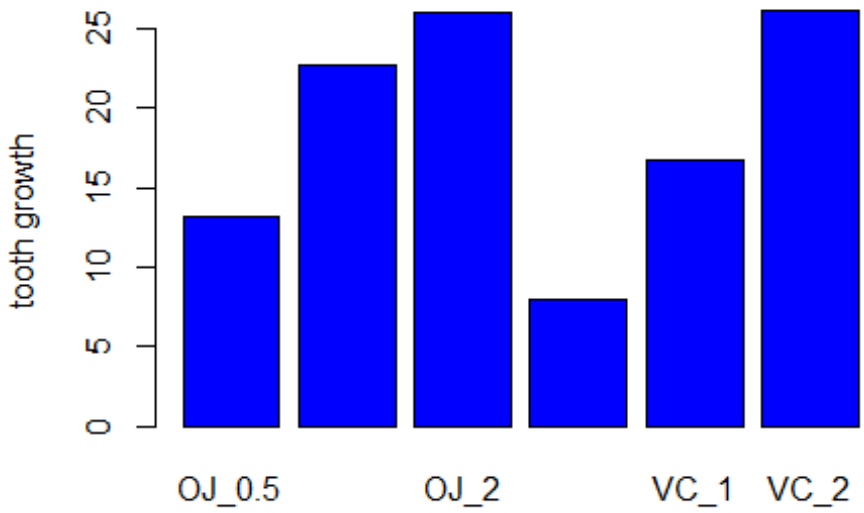
```
summary(ToothGrowth) ## Summary of the data
```

```
##       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
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.5.3
```

```
##
## 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

grouped <- ToothGrowth %>%
  group_by(supp, dose) %>%  ## Group the data by supp and dose
  mutate(supp_dose = gsub(' ','_', paste(supp, dose))) ## Add a column that c
ombines column supp and dose

head(grouped)

## # A tibble: 6 x 4
## # Groups:   supp, dose [1]
##     len supp   dose supp_dose
##   <dbl> <fct> <dbl> <chr>
## 1   4.2 VC      0.5 VC_0.5
## 2  11.5 VC      0.5 VC_0.5
## 3   7.3 VC      0.5 VC_0.5
## 4   5.8 VC      0.5 VC_0.5
## 5   6.4 VC      0.5 VC_0.5
## 6  10   VC      0.5 VC_0.5

grouped_summary <- ToothGrowth %>%
  group_by(supp, dose) %>%  ## Group the data by supp and dose
  summarize_all(mean) %>%  ## Get the mean of each group
  mutate(supp_dose = gsub(' ','_', paste(supp, dose))) ## Add a column that c
ombines column supp and dose

grouped_summary

## # A tibble: 6 x 4
## # Groups:   supp [2]
##   supp   dose   len supp_dose
##   <fct> <dbl> <dbl> <chr>
## 1 OJ      0.5 13.2  OJ_0.5
## 2 OJ      1   22.7  OJ_1
## 3 OJ      2   26.1  OJ_2
## 4 VC      0.5  7.98 VC_0.5
## 5 VC      1   16.8  VC_1
## 6 VC      2   26.1  VC_2

barplot(grouped_summary$len, ylab = 'tooth growth', names.arg = grouped_summa
ry$supp_dose, col = 'blue') ## barplot the tooth growth vs supp + dose used
```




```
## From the graph, we can see pretty clearly that using dosage of 2 is the wa
y to go, regardless if you use OJ or VC, since OJ_2 and VC_2 have the same am
ount of toothgrowth.

## We are, however, interested in if OJ_2 outperforms OJ_1 by a statistically
 significant margin
## So we perform a t test, assuming unequal variance and alpha = 0.05
t.test(grouped$len[grouped$supp_dose == 'OJ_2'], grouped$len[grouped$supp_dos
e == 'OJ_1'], alternative = "greater", var.equal = FALSE)$p.value

## [1] 0.01959757
```

Conclusion

We can see from the graph and the t-test that using a dosage of 2 is the way and only way to go if you are looking for maximum tooth growth. OJ_2 and VC_2 exhibit the same amount of toothgrowth while OJ_2 outperforms OJ_1 by a statistically significant (assuming alpha = 0.05 and unequal variance for the two sets of data) margin.