Title Analysis of the ToothGrowth Data from R Datasets Package

Overview

This project explores the *ToothGrowth* data from the R *datasets* library by performing some preliminary analysis on the data and drawing some initial conclusions.

Exploratory Data Analysis

```
library(dplyr)      # Data manipulation (filter, mutate, group_by, etc.)
library(ggplot2)    # Plotting (qplot, ggplot etc.)
library(knitr)      # Dynamic Report Creation
library(datasets)   # R included sample data sets
data(ToothGrowth)
```

To setup the analysis we load the required libraries and the *ToothGrowth* data set. Per the R documentation for the *datasets* package [1], *ToothGrowth* is described as follows:

> The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

```
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 ...
```

```
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
```

```
table(ToothGrowth$supp)
```

```
##
## OJ VC
## 30 30
```

```
table(ToothGrowth$dose)
```

```
##
## 0.5   1   2
##  20  20  20
```

```
table(ToothGrowth %>% select(supp, dose))
```

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

```
## Transformation/Renaming for clarity

names(ToothGrowth) = c("Length", "Suppliment", "Dose")
ToothGrowth$Dose = factor(ToothGrowth$Dose)
ToothGrowth$Suppliment <- factor(
    ToothGrowth$Suppliment
    ,labels = c("Orange Juice", "Ascorbic Acid")
)

tgBySuppDose <- ToothGrowth %>%
    group_by(Suppliment, Dose) %>%
    summarize(Length = mean(Length)) %>%
    arrange(Suppliment, Dose, desc(Length))

kable(tgBySuppDose
      ,format = "markdown"
      ,caption = "ToothGrowth by Suppliment and Dose"
)
```

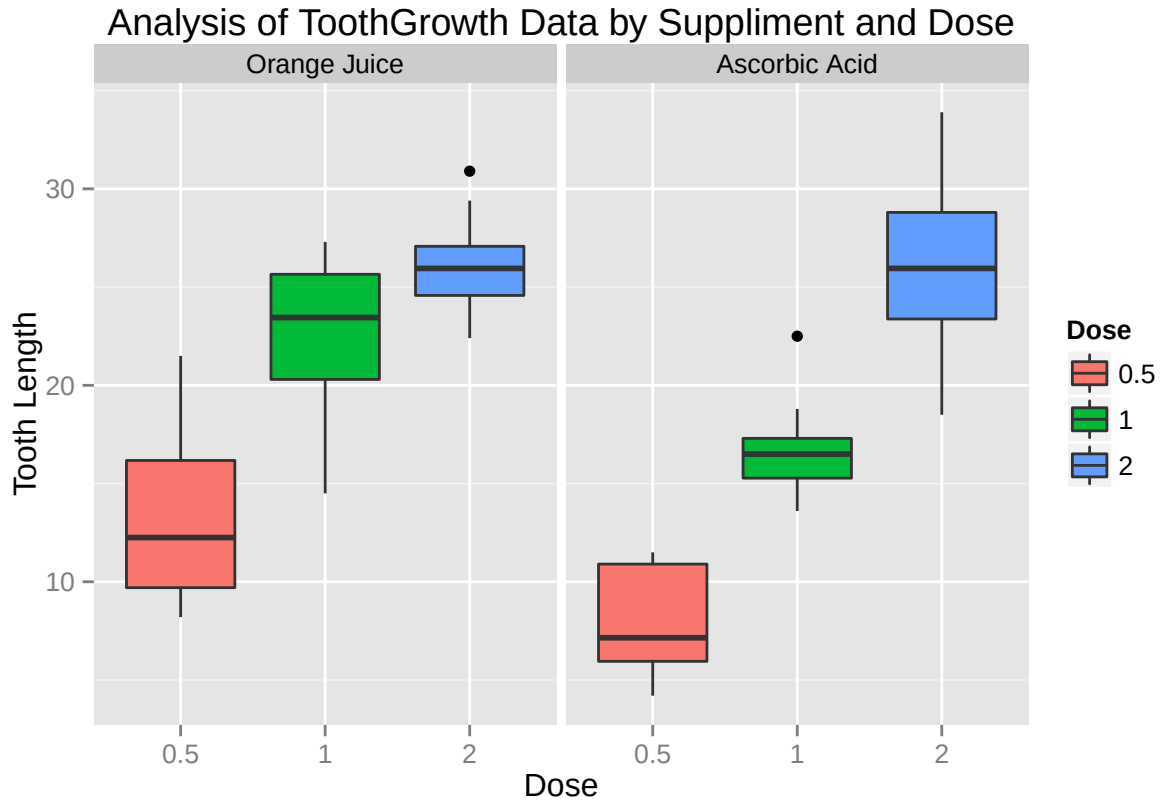
Suppliment	Dose	Length
Orange Juice	0.5	13.23
Orange Juice	1	22.70
Orange Juice	2	26.06
Ascorbic Acid	0.5	7.98
Ascorbic Acid	1	16.77
Ascorbic Acid	2	26.14

The exploratory analysis of the *ToothGrowth* data frame demonstrates that the data consist of 30 observations each of *supp* OJ (Orange Juice) and VC (Ascorbic Acid); the 30 observations of each *supp* consist of 10 observations each of the three *dose* levels (0.5, 1 and 2).

Data Summary

Confidence Intervals and Hypothesis Test

```
plot1 <- ggplot(
    ToothGrowth
    ,aes(
        x = Dose
        ,y = Length
        ,fill = Dose
    )
) + geom_boxplot() +
    facet_grid(.~Suppliment) +
    labs(
        title = "Analysis of ToothGrowth Data by Suppliment and Dose"
        ,y = "Tooth Length"
    )
print(plot1)
```



Conclusions

Reference

[1] https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html