SI Proj Part 2

TridevDeka

24 June 2020

library(datasets) library(dplyr) library(ggplot2)

```
library(datasets)
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
```

```
library(ggplot2)
```

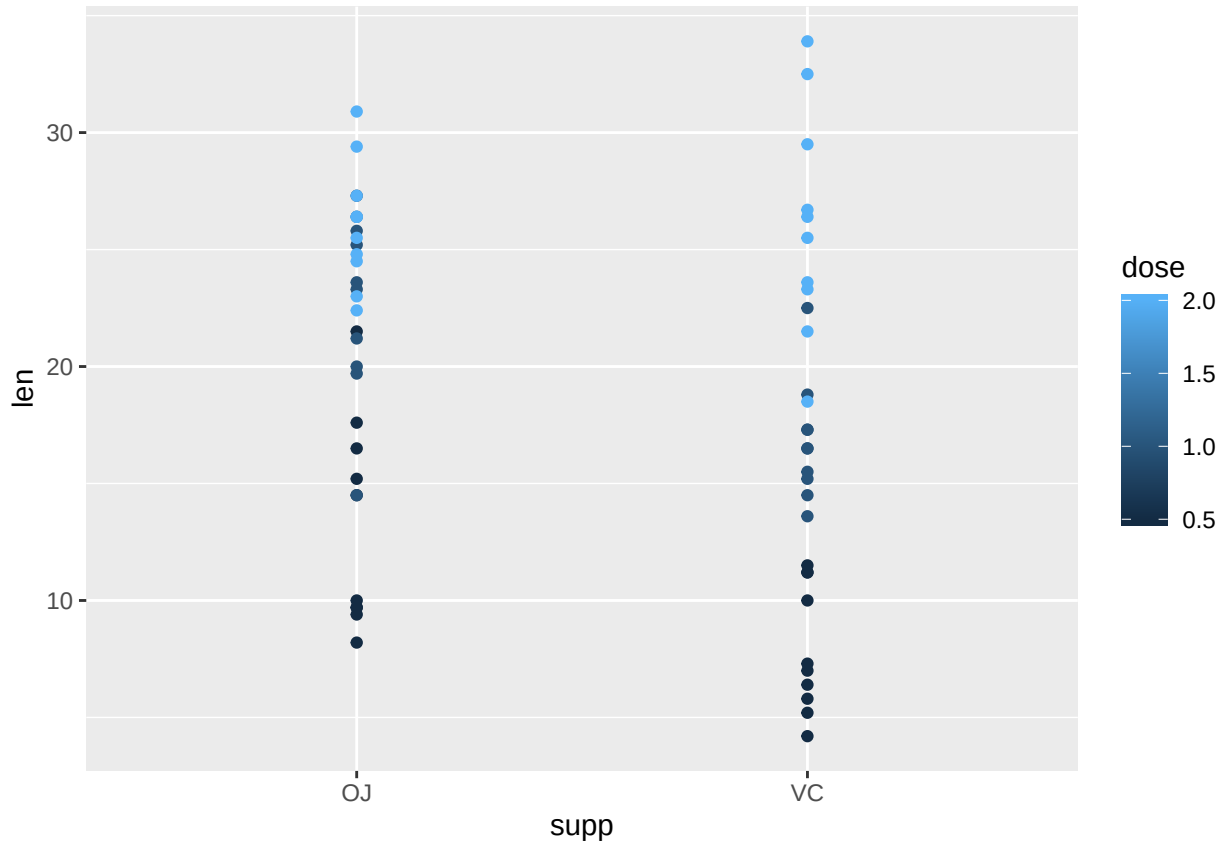
Load ToothGrowth Data

```
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 ...
```

Quick plot for discovery

```
qplot(supp,len, data = ToothGrowth, colour = dose)
```



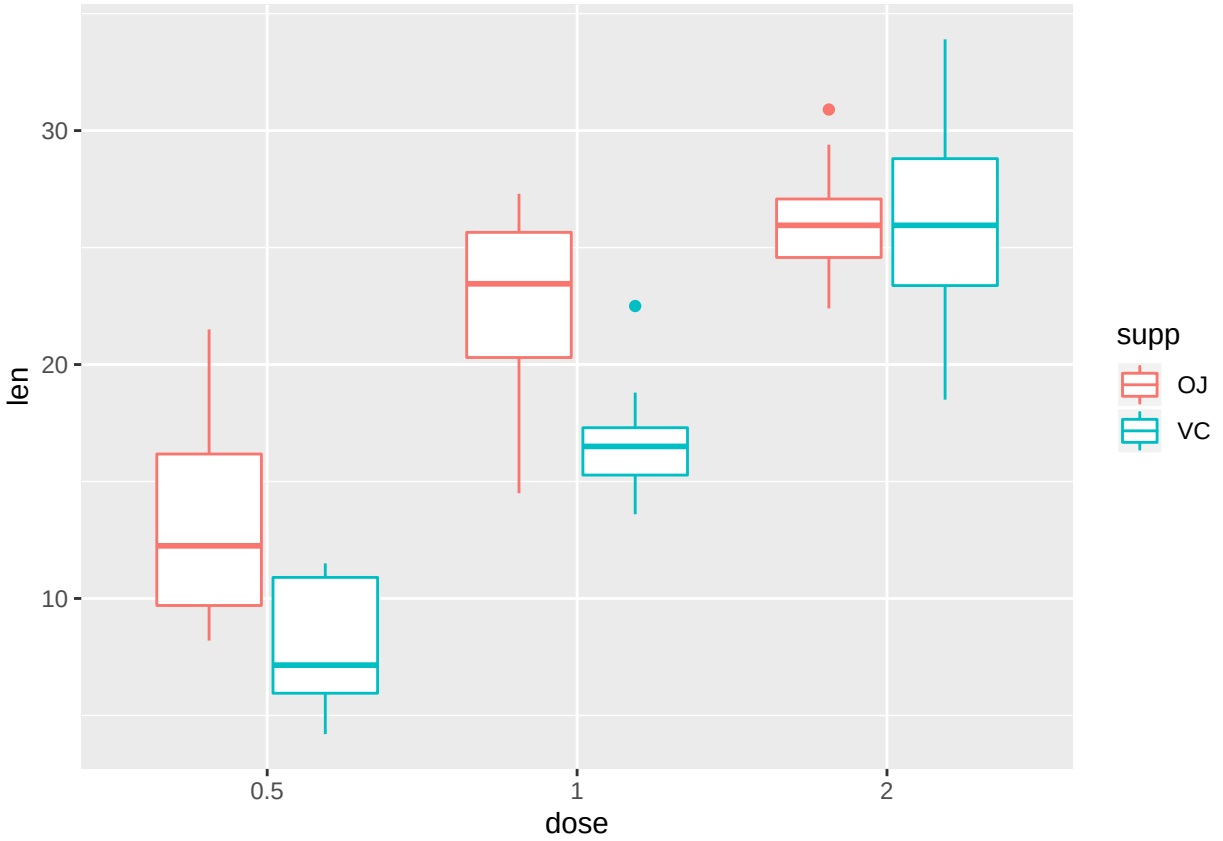
Gather plot by dose and see if there's any relevency

As dose is numeric, we have to convert it to factors so that we can group them

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

Boxplot for better view

```
g <- ggplot(ToothGrowth, aes(x=dose, y=len, colour=supp)) +
  geom_boxplot()
print(g)
```



Is there a difference between dosage and tooth growth?

compare dose between 0.5 & 1, 1 & 2

```
dose05 <- filter(ToothGrowth, dose == 0.5)
dose1 <- filter(ToothGrowth, dose == 1)
dose2 <- filter(ToothGrowth, dose == 2)

t.test(dose1$len, dose05$len, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  dose1$len and dose05$len
## t = 6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  6.753323      Inf
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

```
t.test(dose2$len, dose1$len, alternative = "greater")
```

```
##
```

```
##  Welch Two Sample t-test
##
## data:  dose2$len and dose1$len
## t = 4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  4.17387     Inf
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

Is OJ a more effective delivery method than VC?

```
OJdose05 <- filter(dose05, supp == "OJ")
VCdose05 <- filter(dose05, supp == "VC")
t.test(OJdose05$len, VCdose05$len, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  OJdose05$len and VCdose05$len
## t = 3.1697, df = 14.969, p-value = 0.003179
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  2.34604     Inf
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

```
OJdose1 <- filter(dose1, supp == "OJ")
VCdose1 <- filter(dose1, supp == "VC")
t.test(OJdose1$len, VCdose1$len, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  OJdose1$len and VCdose1$len
## t = 4.0328, df = 15.358, p-value = 0.0005192
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  3.356158      Inf
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

```
OJdose2 <- filter(dose2, supp == "OJ")
VCdose2 <- filter(dose2, supp == "VC")
t.test(OJdose2$len, VCdose2$len, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  OJdose2$len and VCdose2$len
## t = -0.046136, df = 14.04, p-value = 0.5181
## alternative hypothesis: true difference in means is greater than 0
```

```
## 95 percent confidence interval:
##  -3.1335     Inf
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```