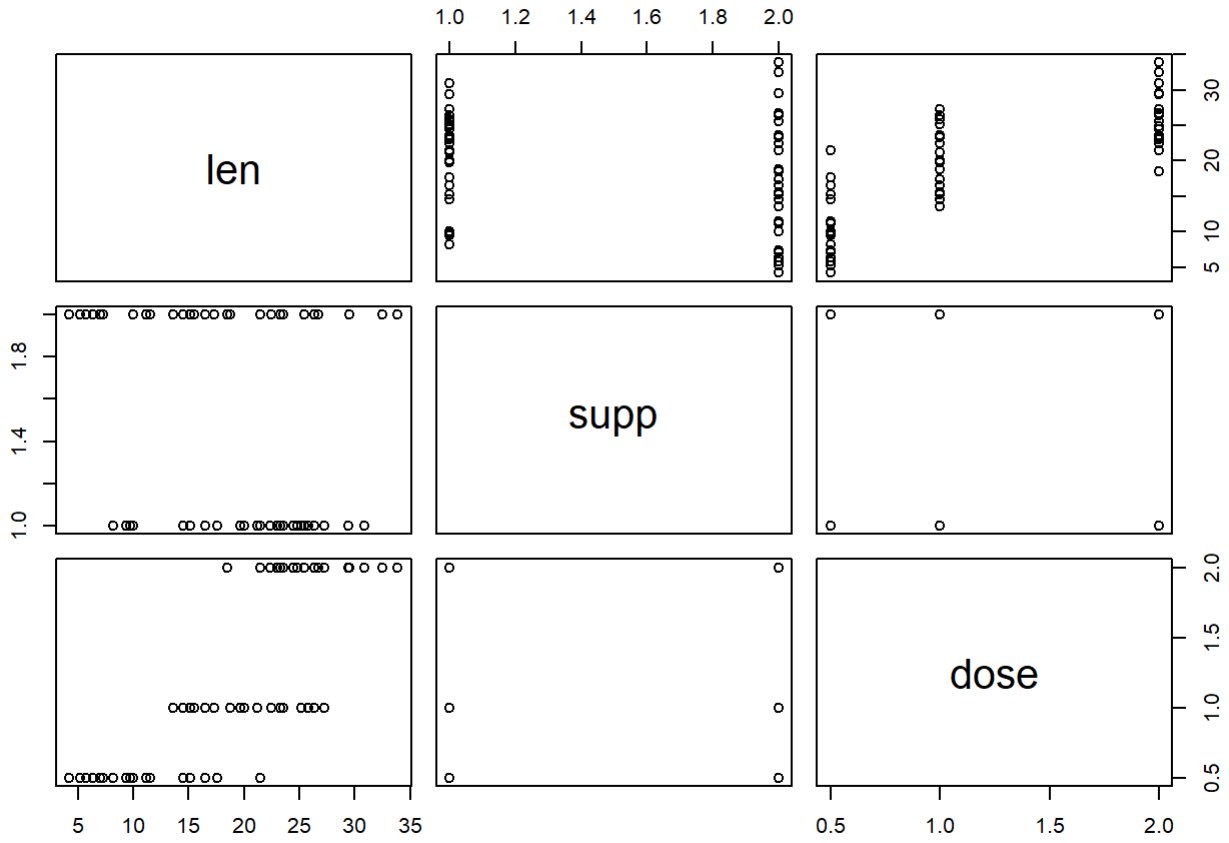
R Notebook

Overview: in the second portion of the project, I am going to analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses

```
library(ggplot2)
## Loading data
data("ToothGrowth")
plot(ToothGrowth)
```

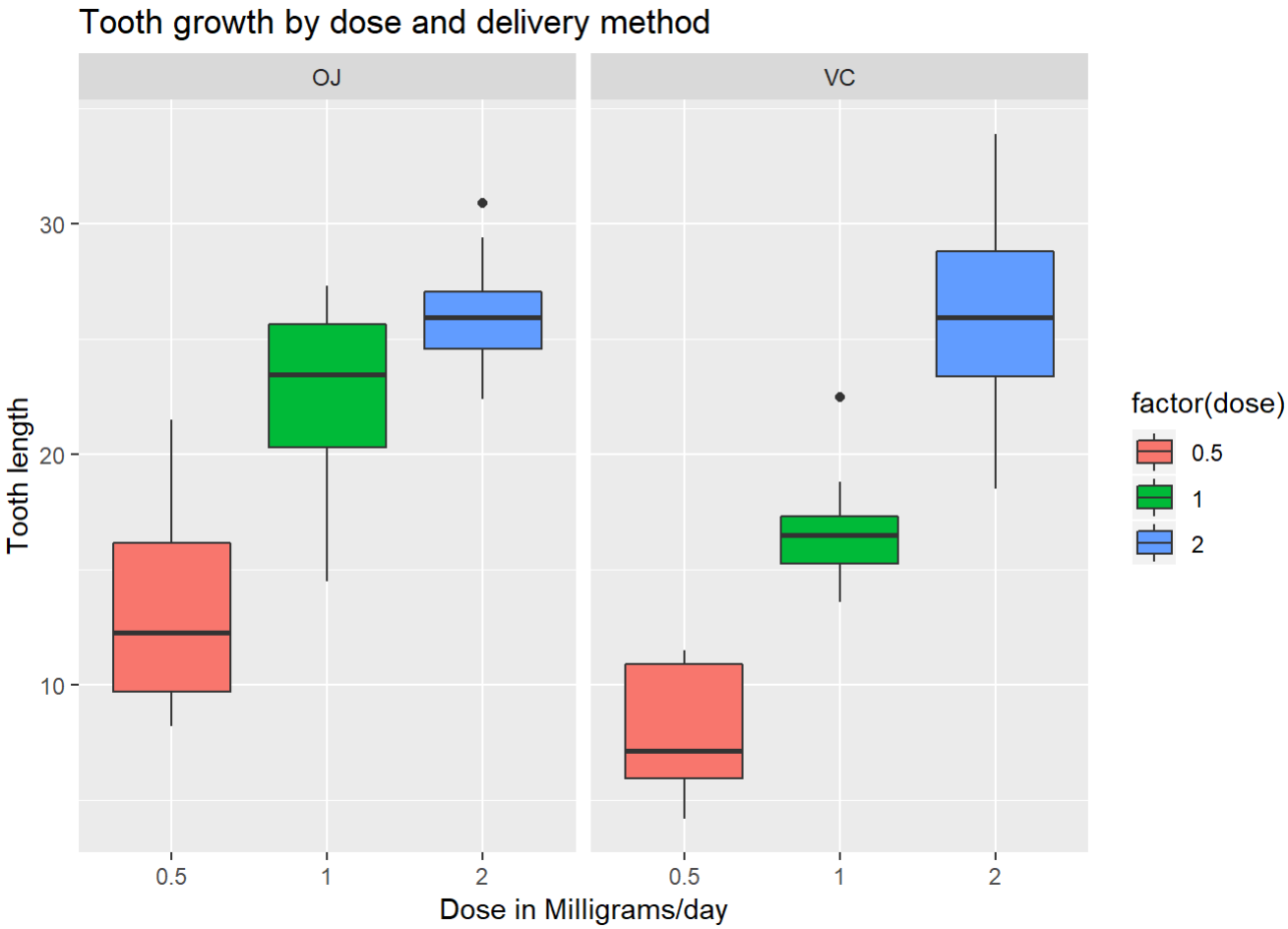


According to the plots, we can see the pairwise relationship among three variables, len,supp and dose.

```
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 ...
```

```
### Creating Boxplot of the data
ggplot(ToothGrowth, aes(factor(dose), len)) + geom_boxplot(aes(fill = factor(dose))) + face
t_grid(.~supp) + labs(title = "Tooth growth by dose and delivery method", x= "Dose in Milli
grams/day", y= "Tooth length")
```




```
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
```

2. According to the summary, we can understand the distribution of the data.

```
t.05.oj <- ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 0.5,]
t.05.vc <- ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == 0.5, ]
t.05 <- t.test(t.05.oj$len, t.05.vc$len, var.equal = F)
t.05
```

```
##
##  Welch Two Sample t-test
##
## data:  t.05.oj$len and t.05.vc$len
## 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 of x mean of y
##     13.23      7.98
```

```
t.1.oj <- ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 1,]
t.1.vc <- ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == 1, ]
t.1 <- t.test(t.1.oj$len, t.1.vc$len, var.equal = F)
t.1
```

```
##
##  Welch Two Sample t-test
##
## data:  t.1.oj$len and t.1.vc$len
## 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 of x mean of y
##     22.70     16.77
```

```
t.2.oj <- ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 2,]
t.2.vc <- ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == 2, ]
t.2 <- t.test(t.2.oj$len, t.2.vc$len, var.equal = F)
t.2
```

```
##
##  Welch Two Sample t-test
##
## data:  t.2.oj$len and t.2.vc$len
## t = -0.046136, 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 of x mean of y
##     26.06     26.14
```

```
data.frame("p-value" = c(t.05$p.value, t.1$p.value, t.2$p.value), row.names = c("Dose 0.5",
"Dose 1","Dose 2"))
```

	p.value <dbl>
Dose 0.5	0.006358607
Dose 1	0.001038376
Dose 2	0.963851589

3 rows

T.tests were performed for testing tooth growth by supp(OJ,VC) and dose(0.5, 1 and 2) By comparing the p-values, At 95% signifance level (95% confidence interval) we are rejecting the null hypothesis of equal means by delivery methods for dose 0.5 and 1 but we don`t reject the null hypothesis for dose 2