Assignment 1 MATH 208 (Question 1)

Prsa Yadollahi
260869949
MATH 208 - Assignment 1

a) i) ii)

```
data(ToothGrowth)
mode(head(ToothGrowth))
```

```
## [1] "list"
```

```
class(head(ToothGrowth))
```

```
## [1] "data.frame"
```

b)

```
nrow(head(ToothGrowth))
```

```
## [1] 6
```

```
ncol(head(ToothGrowth))
```

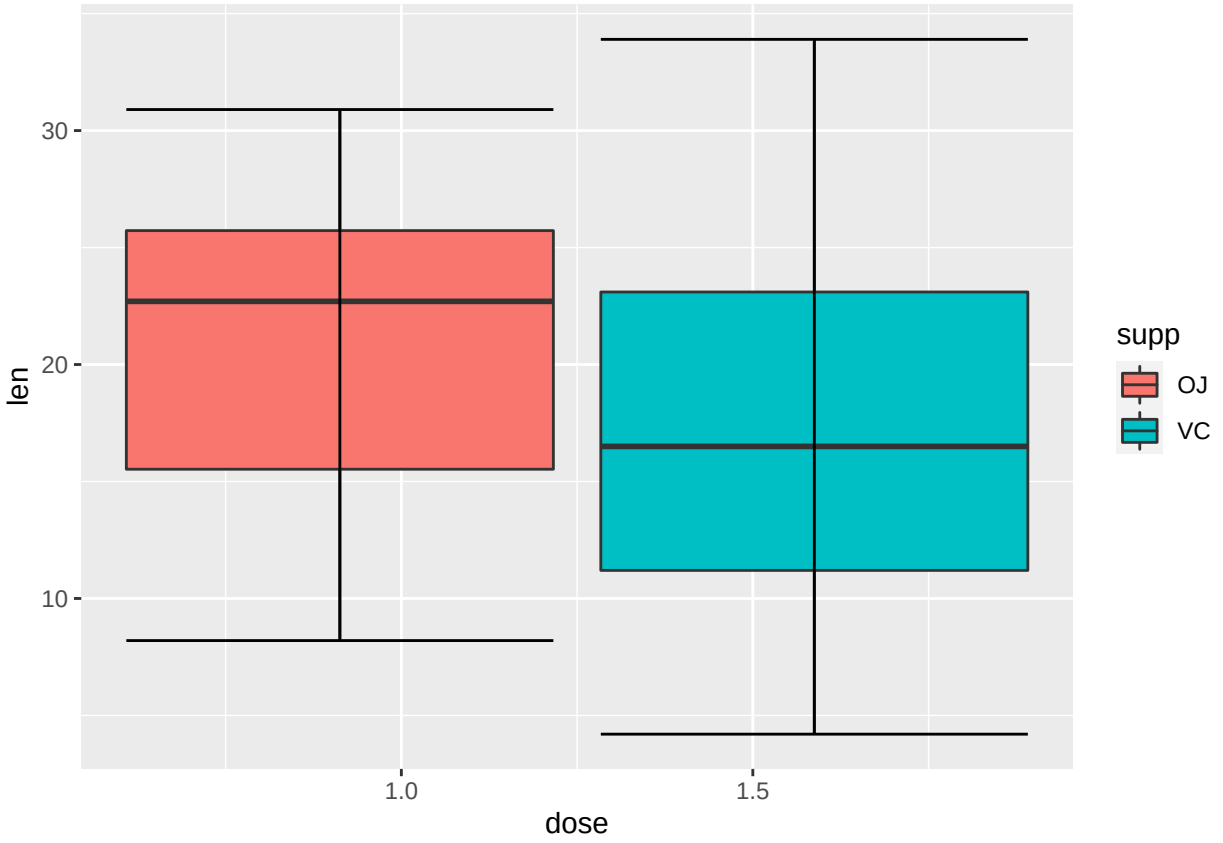
```
## [1] 3
```

```
head(ToothGrowth)
```

```
##    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
```

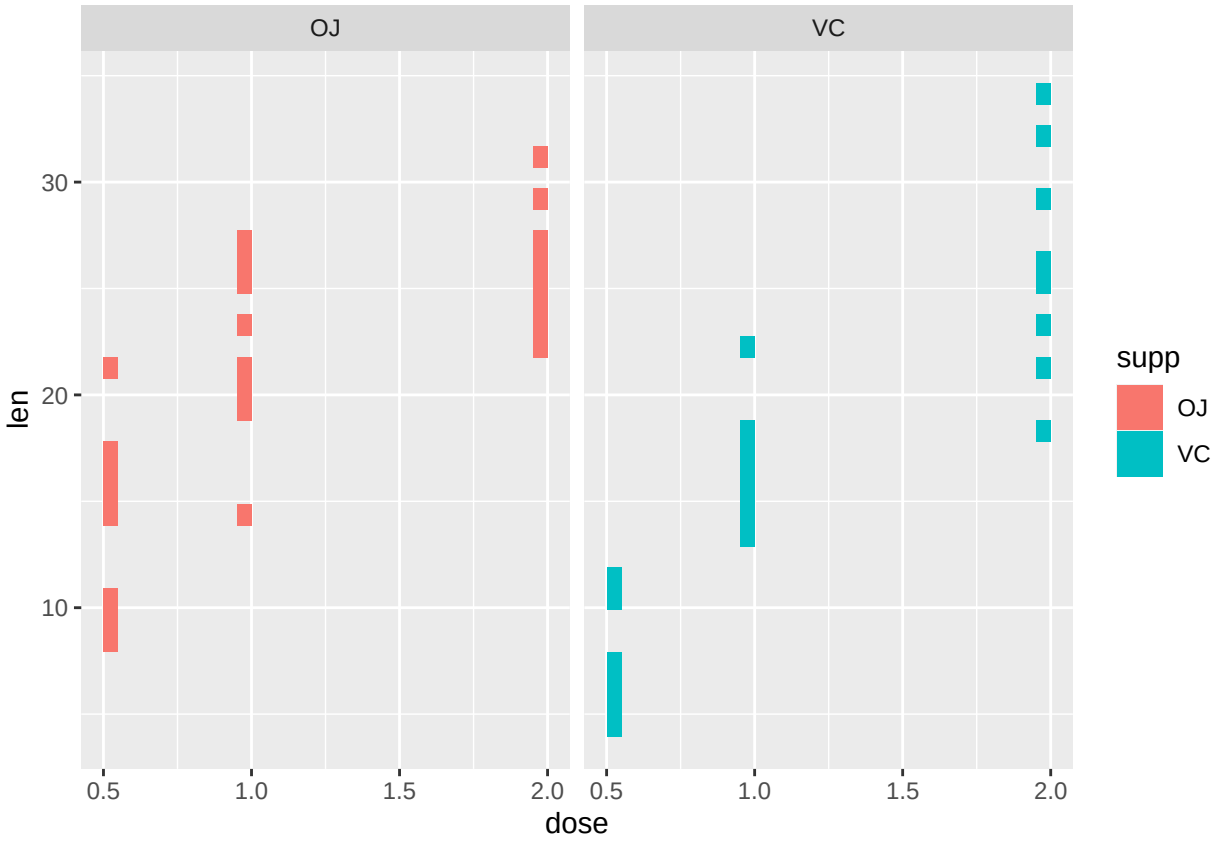
c) #Boxplot

```
ggplot(ToothGrowth, aes(x=dose, y=len, fill=supp, group=supp)) +
  geom_boxplot() +stat_boxplot(geom="errorbar",width=0.5)
```



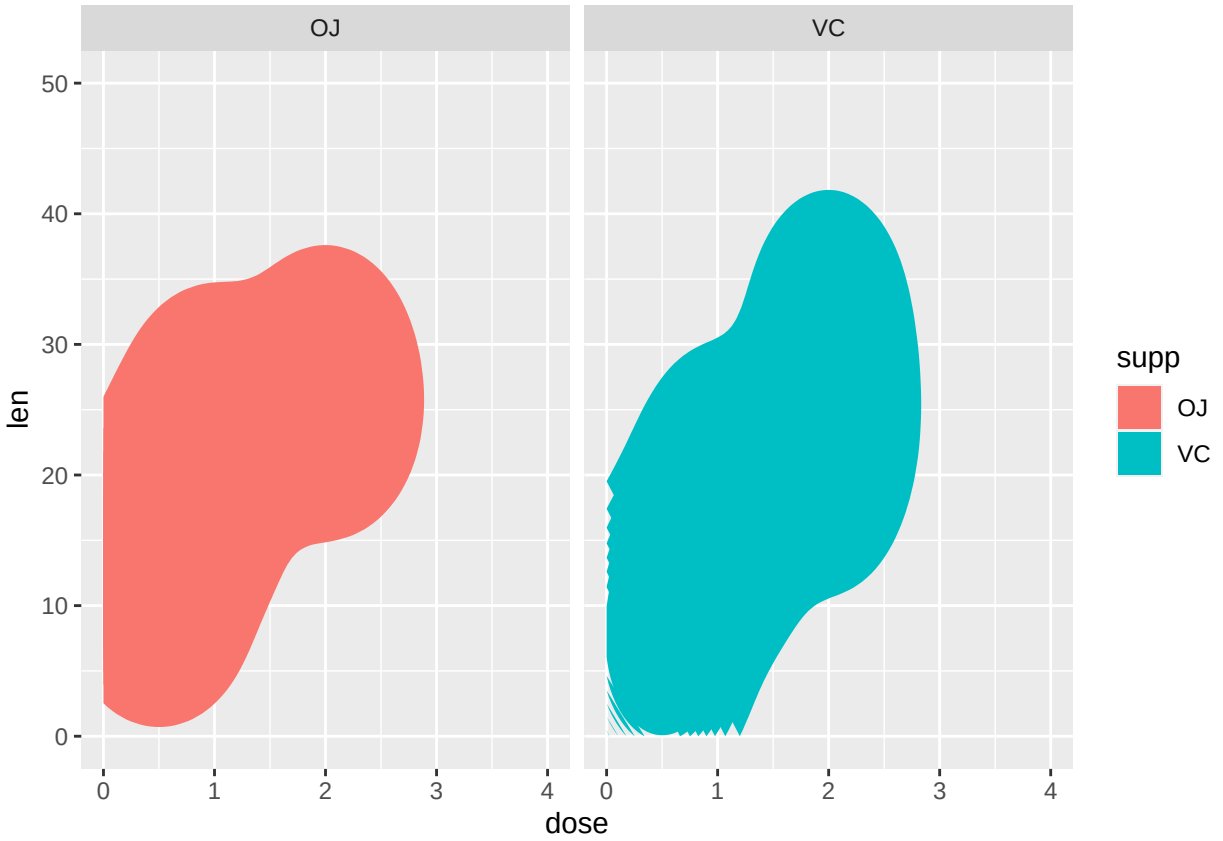
Histogram

```
ggplot(ToothGrowth, aes(x=dose, y=len, fill=supp, group=supp)) +
  geom_bin2d(bins=30) +
  facet_wrap(~supp)
```



#Density plots

```
ggplot(ToothGrowth, aes(x=dose,y=len,fill=supp,group=supp)) +
  stat_density_2d(bins=30, aes(fill = supp),
                  geom="polygon") +
  ylim(c(0,50)) +
  xlim(c(0,4)) +
  facet_wrap(~supp)
```

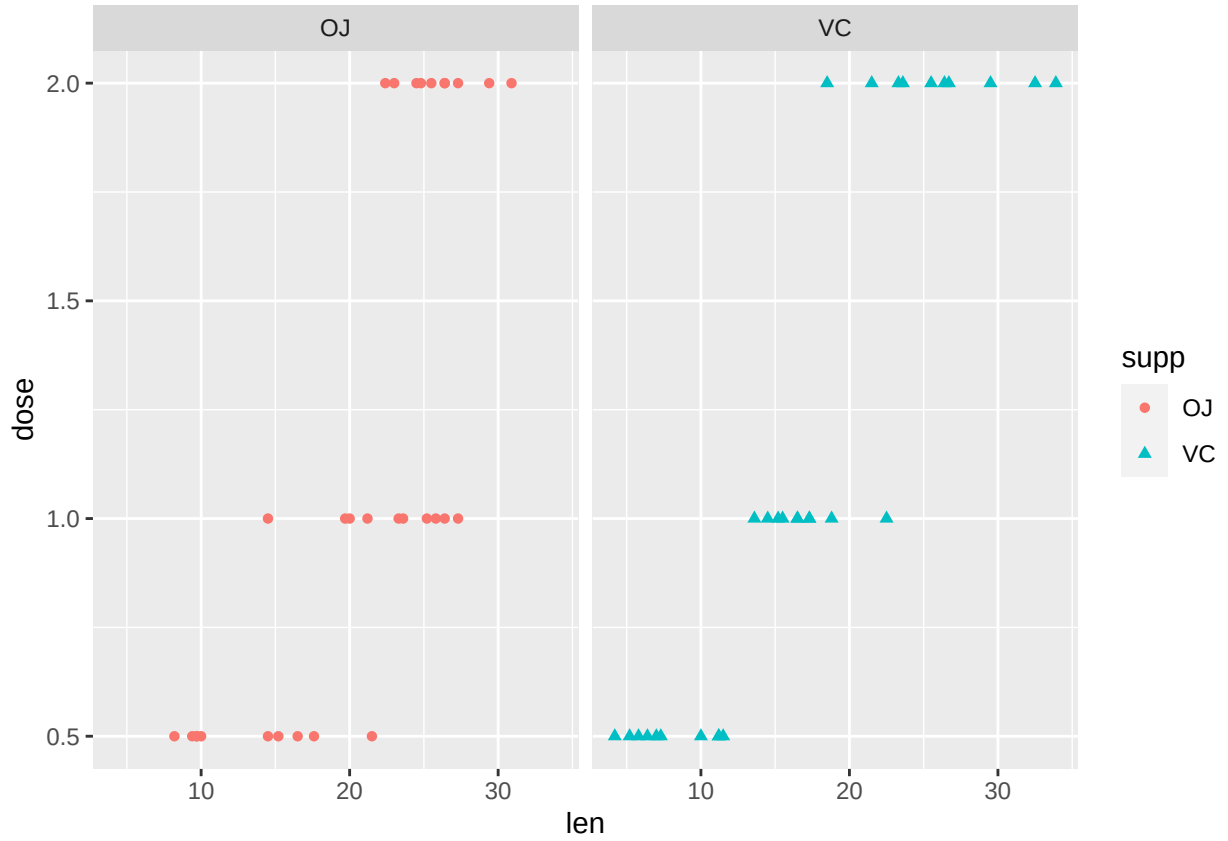


From the plots drawn above, we can deduct that there is no supplement that is associated with greater lengths. This can be observed throughout the plots since they all show a large variation of lengths. Take, for example, the boxplot that shows a similiar distribution of lengths but not a very large difference in length or the density plot where both supplements have similar distributions.

d) The density plot is more effective for assessing whether there is a difference in distribution of lengths between the OJ and VC. This is mainly because we have a larger view of the distribution making it easier to see how it is shaped. The boxplot shows a similar distribution of supplements but with less points.

e)

```
ggplot(ToothGrowth, aes(x=len, y=dose, shape = supp,
                        group=supp, col = supp)) +
  geom_point() + facet_wrap(~supp)
```



No, the association between length and dose does not depend on the type of supplement. Take lengths 0.5, 1.0 and 2.0 that have similar distributions of lengths for doses.

f)

g) len

```
ToothGrowth %>% group_by(supp) %>%  summarise(Mean = mean(len),
                          Median = median(len),
                          SD=sd(len))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 2 x 4
##   supp   Mean Median    SD
##   <fct> <dbl>  <dbl> <dbl>
## 1 OJ     20.7   22.7  6.61
## 2 VC     17.0   16.5  8.27
```

ii) dose

```
ToothGrowth %>% group_by(supp) %>%  summarise(Mean = mean(dose),
                          Median = median(dose),
                          SD=sd(dose))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 2 x 4
##   supp   Mean Median    SD
##   <fct> <dbl>  <dbl> <dbl>
## 1 OJ     1.17      1 0.634
## 2 VC     1.17      1 0.634
```