UNIQUE ID: E7321008
ASWIN RAJ

R ASSESSMENT – CA2

 # 1)

#Importing the Covid data

```
confirmedraw <- read.csv("https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv")

str(confirmedraw)

View(confirmedraw)

deathsraw <- read.csv("https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_deaths_global.csv")

recoveredraw <- read.csv("https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_recovered_global.csv")

View(deathsraw)

View(recoveredraw)

library(tidyr)

library(dplyr)
```

 # Gathering and grouping the data by Country.Region & Date

```
confirmed <- confirmedraw %>% gather(key="date", value="confirmed", -c(Country.Region, Province.State, Lat, Long)) %>% group_by(Country.Region, date) %>% summarize(confirmed=sum(confirmed))

deaths <- deathsraw %>% gather(key="date", value="deaths", -c(Country.Region, Province.State, Lat, Long)) %>% group_by(Country.Region, date) %>% summarize(deaths=sum(deaths))

recovered <- recoveredraw %>% gather(key="date", value="recovered", -c(Country.Region, Province.State, Lat, Long)) %>% group_by(Country.Region, date) %>% summarize(recovered=sum(recovered))

summary(confirmed)

summary(deaths)

summary(recovered)
```

#Combining all three datasets

```
country <- full_join(confirmed, deaths) %>% full_join(recovered)

str(country) # checking character of date
```

#assigning a variable to date and converting the character to date

```
country$date <- country$date %>% sub("X", "", .) %>% as.Date("%m.%d.%y")

str(country) # checking if it is converted
```

 # creating a new variable for number of days

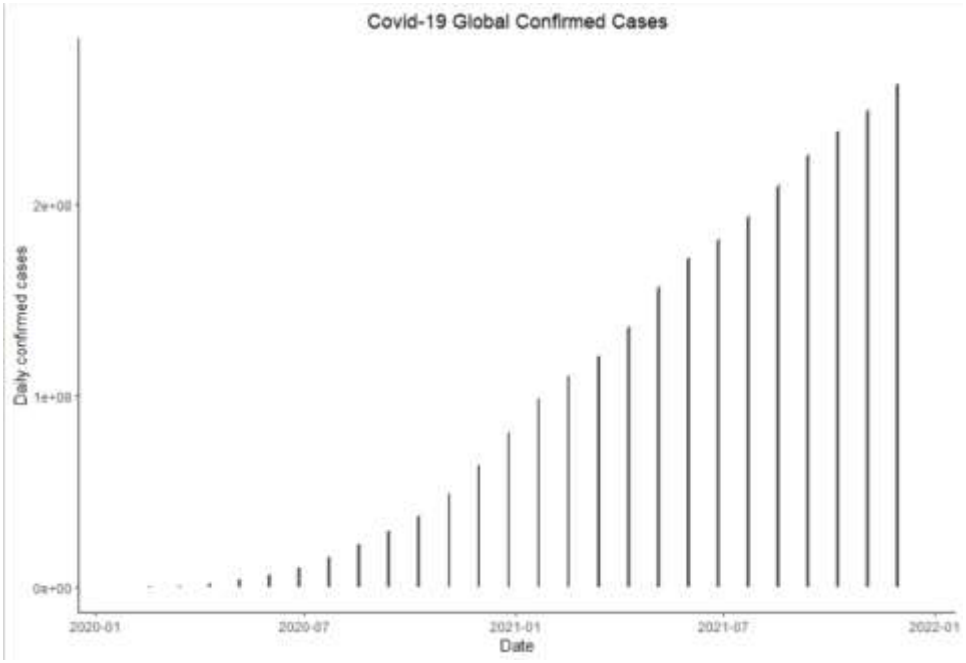
```
country <- country %>% group_by(Country.Region) %>% mutate(cumconfirmed=cumsum(confirmed), days = date - first(date) + 1)
```

#using the datasets assigned for each country and combining it to look at the world level of covid cases

```
world <- country %>% group_by(date) %>% summarize(confirmed=sum(confirmed),
cumconfirmed=sum(cumconfirmed), deaths=sum(deaths), recovered=sum(recovered)) %>% mutate(days = date -
first(date) + 1)
```

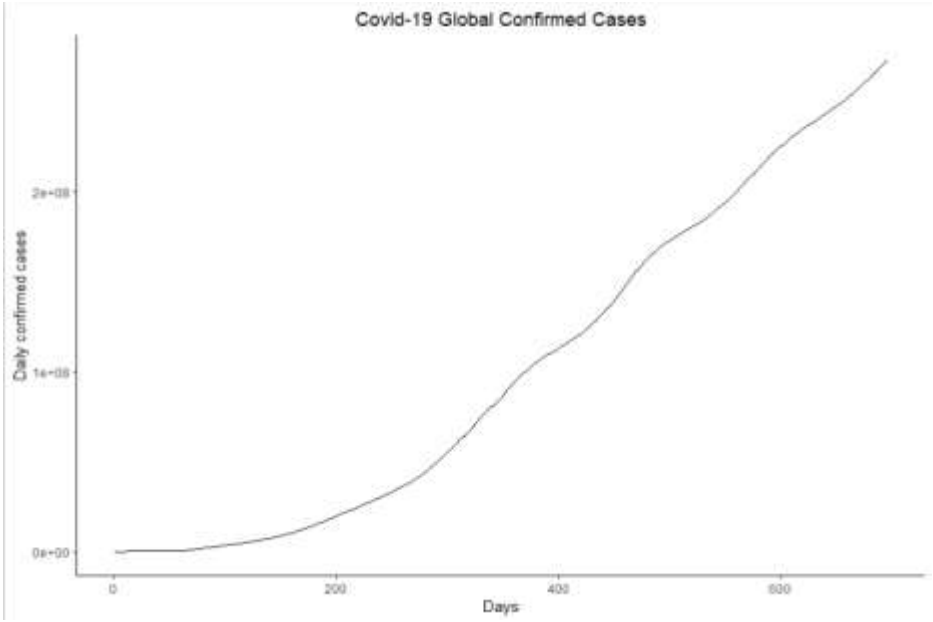
 # A) Barchart

```
library(ggplot2)
ggplot(world, aes(x=date, y=confirmed)) + geom_bar(stat="identity", width=0.1) +
 theme_classic() +
 labs(title = "Covid-19 Global Confirmed Cases", x= "Date", y= "Daily confirmed cases") +
 theme(plot.title = element_text(hjust = 0.5))
```



 # B) Lineplot

```
ggplot(world, aes(x=days, y=confirmed)) + geom_line() +
 theme_classic() +
 labs(title = "Covid-19 Global Confirmed Cases", x= "Days", y= "Daily confirmed cases") +
 theme(plot.title = element_text(hjust = 0.5))
```

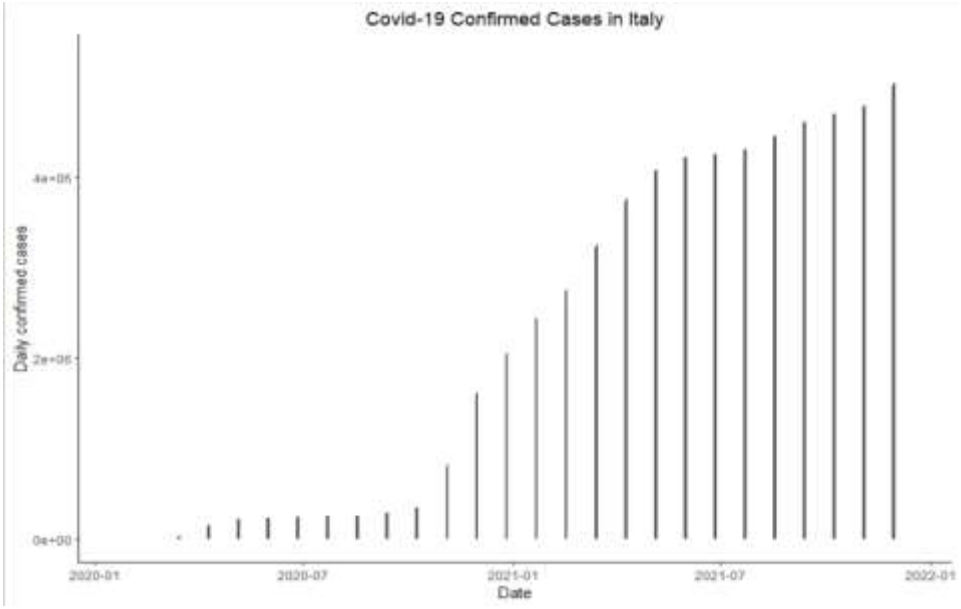


 # Confirmed Cases in Italy

```
italy <- country %>% filter(Country.Region=="Italy")
```

#confirmed cases in italy

```
ggplot(italy, aes(x=date, y=confirmed)) + geom_bar(stat="identity", width=0.1) +
  theme_classic() +
  labs(title = "Covid-19 Confirmed Cases in Italy", x= "Date", y= "Daily confirmed cases") +
  theme(plot.title = element_text(hjust = 0.5))
```

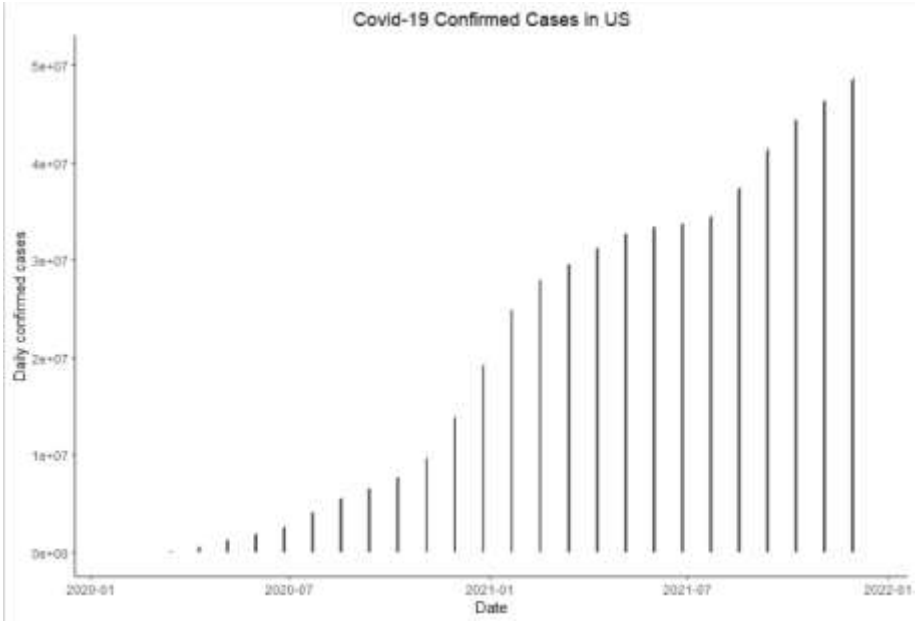


#Confirmed cases in USA

```
US <- country %>% filter(Country.Region=="US")
```

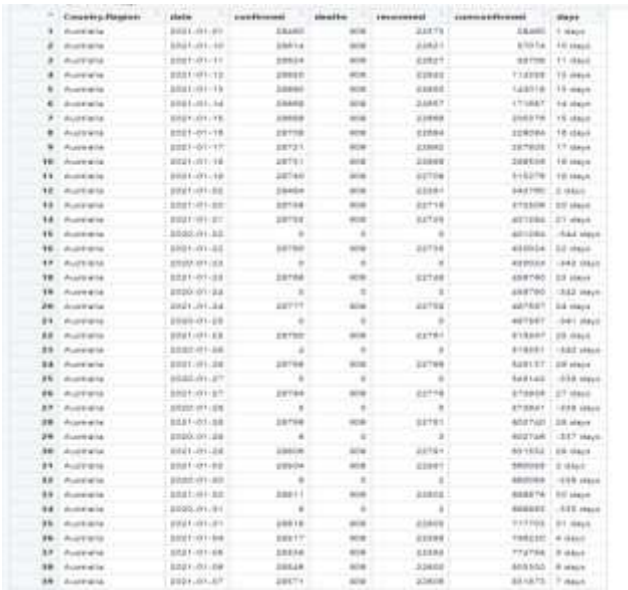
#confirmed cases in US

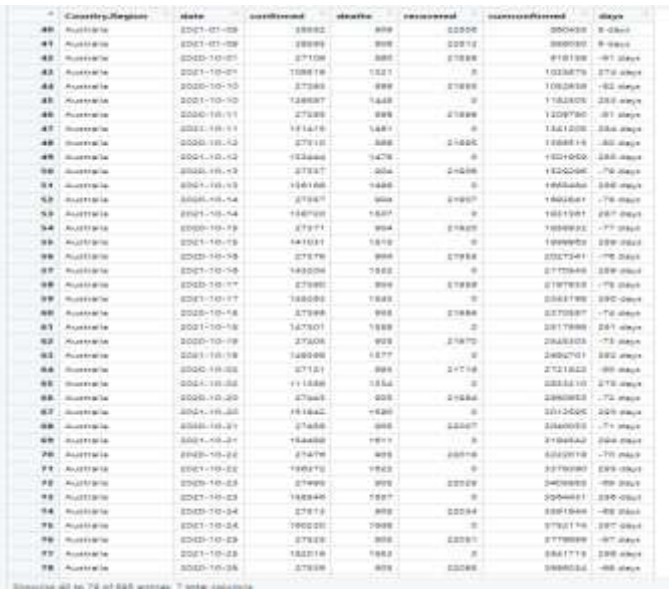
```
ggplot(US, aes(x=date, y=confirmed)) + geom_bar(stat="identity", width=0.1) +
  theme_classic() +
  labs(title = "Covid-19 Confirmed Cases in US", x= "Date", y= "Daily confirmed cases") +
  theme(plot.title = element_text(hjust = 0.5))
```



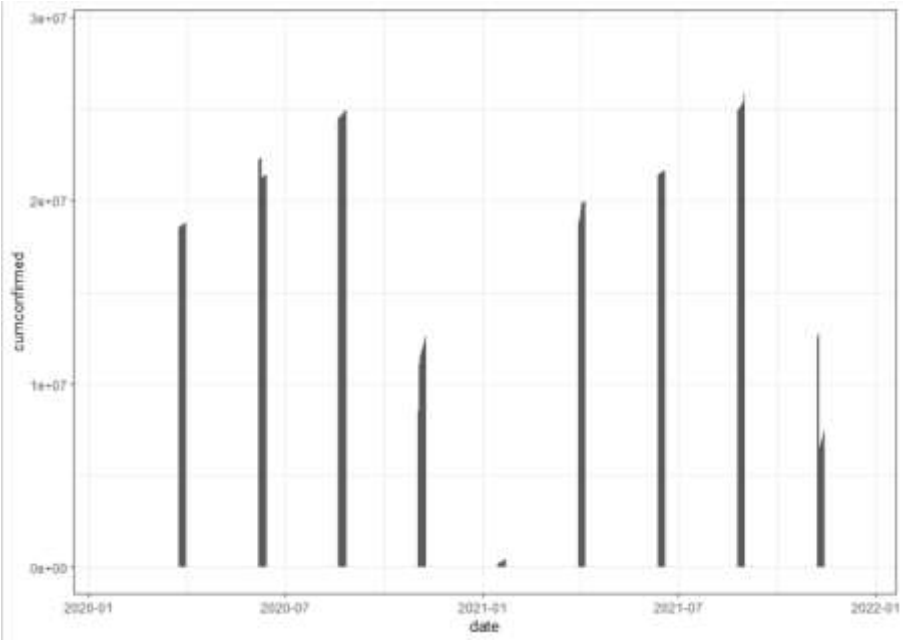
 # D)

```
afg<-country %>%
  filter(Country.Region=="Australia")
View(afg)
```



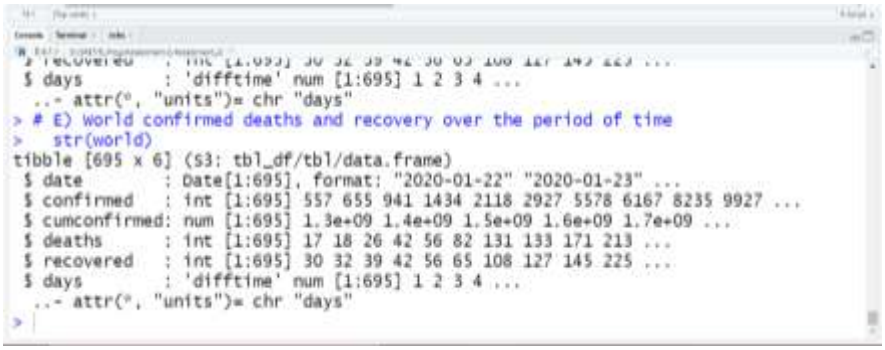


```
ggplot(afg,aes(x=date,y=cumconfirmed))+
  geom_bar(stat="identity",width=0.1)+
  theme_bw()
```

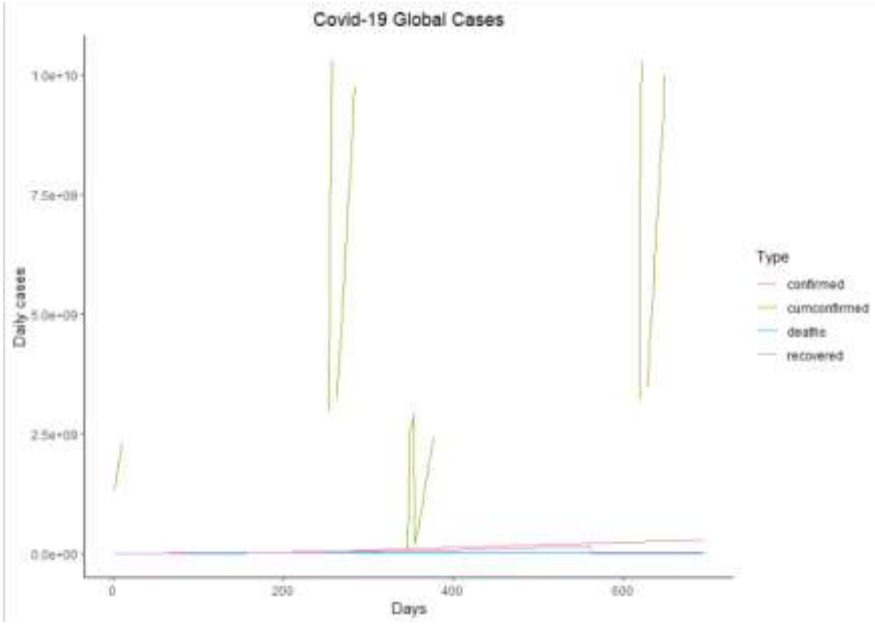


 # E) World confirmed deaths and recovery over the period of time

```
str(world)
```




```
world %>% gather("Type", "Cases", -c(date, days)) %>%
  ggplot(aes(x=days, y=Cases, colour=Type)) + geom_line() +
  theme_classic() +
  labs(title = "Covid-19 Global Cases", x= "Days", y= "Daily cases") +
  theme(plot.title = element_text(hjust = 0.5))
```



 # 2) A) mean, median, mode, trimmed mean

x<-c(100, 95, 95, 90, 85, 75, 65, 60, 55)

mean(x)

median(x)

library(modeest)

mode=mfv(x)

print(mode)

#trimmed mean

mean(x,trim=0.5)

#2 b)

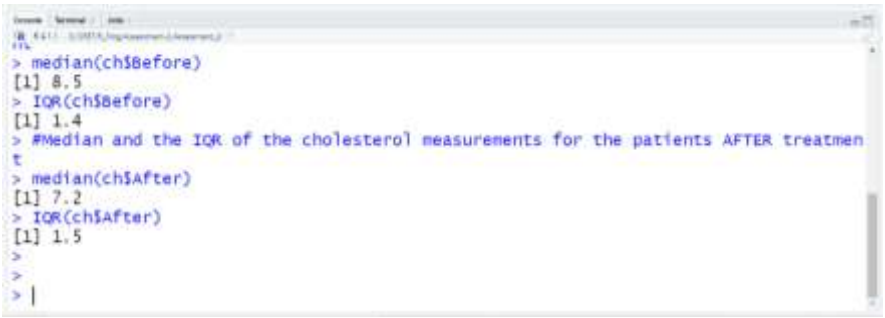
```
ch<-data.frame(Patient=c(1:15),
               Before=c(9.1,8.0,7.7,10.0,9.6,7.9,9.0,7.1,8.3,9.6,8.2,9.2,7.3,8.5,9.5),
               After=c(8.2,6.4,6.6,8.5,8.0,5.8,7.8,7.2,6.7,9.8,7.1,7.7,6.0,6.6,8.4))
View(ch)
```

#Median and the IQR of the cholesterol measurements for the patients BEFORE treatment

```
median(ch$Before)
IQR(ch$Before)
```

#Median and the IQR of the cholesterol measurements for the patients AFTER treatment

```
median(ch$After)
IQR(ch$After)
```



```
> median(ch$Before)
[1] 8.5
> IQR(ch$Before)
[1] 1.4
> #Median and the IQR of the cholesterol measurements for the patients AFTER treatmen
t
> median(ch$After)
[1] 7.2
> IQR(ch$After)
[1] 1.5
>
>
>
```

 # 2) C)

#installing packages "ggpubr" and "rstatix" for hypothesis testing

```
install.packages("ggpubr")
library(ggpubr)
install.packages("rstatix")
library(rstatix)
```

#installing package "MASS" for the dataset 'birthwt'

```
install.packages("MASS")
library(MASS)
birthwt
```

#one way anova to do hypothesis testing because we use it to compare dependent and independent

#H0 : smoking mother has no impact on weight of the baby during birth

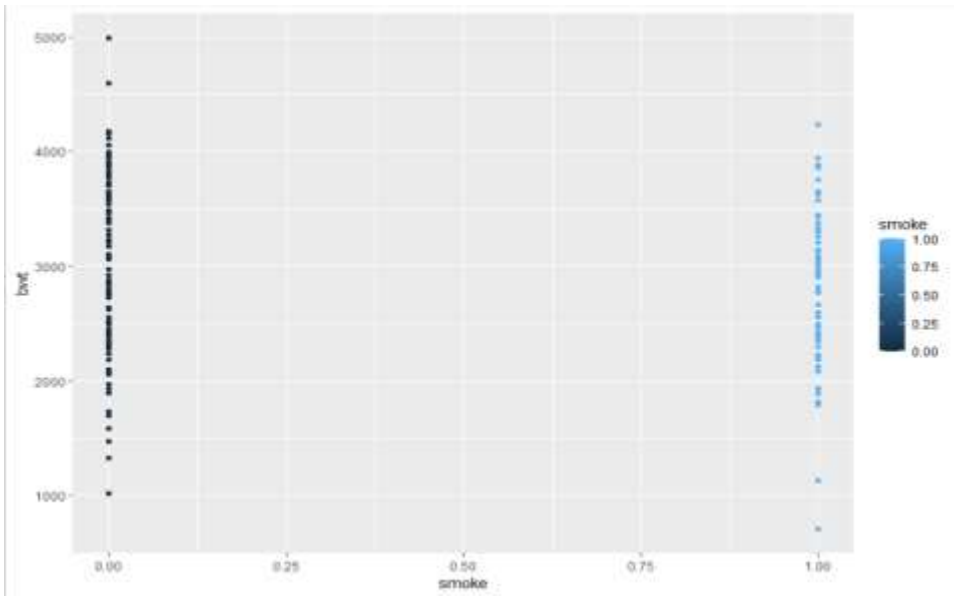
```
aov1=aov(birthwt$bwt~birthwt$smoke)
summary(aov1)
```

```
> #one way anova to do hypothesis testing because we use it to compare dependent and i
ndependent
> #H0 : smoking mother has no impact on weight of the baby during birth
> aov1=aov(birthwt$bwt~birthwt$smoke)
> summary(aov1)
                Df   Sum Sq Mean Sq F value  Pr(>F)
birthwt$smoke    1  3625946 3625946   7.038 0.00867 **
Residuals      187 96343710  515207
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

#null hypothesis is rejected because p-value is lesser than 0.05

#visualizing the plot

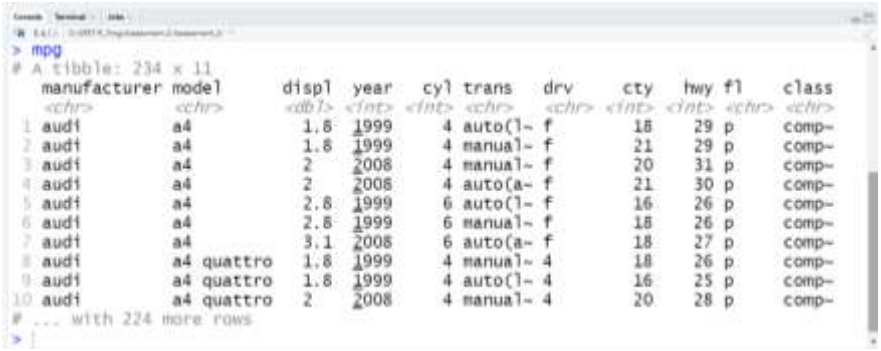
```
library(tidyverse)
ggplot(birthwt,aes(x=smoke,y=bwt,col=smoke))+
  geom_point()
```



#Hence smoking mothers have impact on weight of the baby during birth

 # 3) A)

```
library(tidyverse)
mpg
?mpg
```

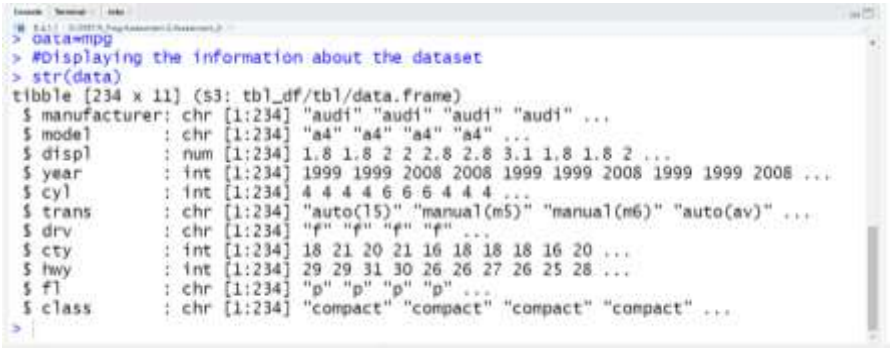


```
> mpg
# A tibble: 234 x 11
   manufacturer model      displ  year   cyl trans   drv     cty   hwy fl    class
   <chr>        <chr>      <dbl> <int> <int> <chr>   <chr> <int> <int> <chr> <chr>
 1 audi         a4           1.8  1999     4 auto(l~ f        18    29 p     comp~
 2 audi         a4           1.8  1999     4 manual~ f        21    29 p     comp~
 3 audi         a4           2    2008     4 manual~ f        20    31 p     comp~
 4 audi         a4           2    2008     4 auto(a~ f        21    30 p     comp~
 5 audi         a4           2.8  1999     6 auto(l~ f        16    26 p     comp~
 6 audi         a4           2.8  1999     6 manual~ f        18    26 p     comp~
 7 audi         a4           3.1  2008     6 auto(a~ f        18    27 p     comp~
 8 audi         a4 quattro   1.8  1999     4 manual~ 4        18    26 p     comp~
 9 audi         a4 quattro   1.8  1999     4 auto(l~ 4        16    25 p     comp~
10 audi         a4 quattro   2    2008     4 manual~ 4        20    28 p     comp~
# ... with 224 more rows
>
```

data=mpg

#Displaying the information about the dataset

str(data)

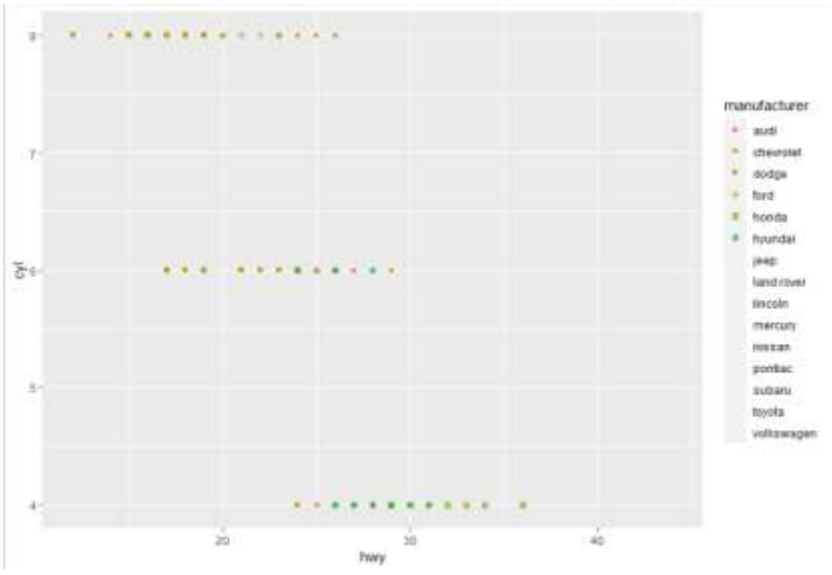


```
> data=mpg
> #Displaying the information about the dataset
> str(data)
tibble [234 x 11] (S3: tbl_df/tbl/data.frame)
 $ manufacturer: chr [1:234] "audi" "audi" "audi" "audi" ...
 $ model       : chr [1:234] "a4" "a4" "a4" "a4" ...
 $ displ       : num [1:234] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
 $ year        : int [1:234] 1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...
 $ cyl         : int [1:234] 4 4 4 4 6 6 6 4 4 4 ...
 $ trans       : chr [1:234] "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...
 $ drv         : chr [1:234] "f" "f" "f" "f" ...
 $ cty         : int [1:234] 18 21 20 21 16 18 18 18 16 20 ...
 $ hwy         : int [1:234] 29 29 31 30 26 26 27 26 25 28 ...
 $ fl          : chr [1:234] "p" "p" "p" "p" ...
 $ class       : chr [1:234] "compact" "compact" "compact" "compact" ...
>
```

 # 3) B)

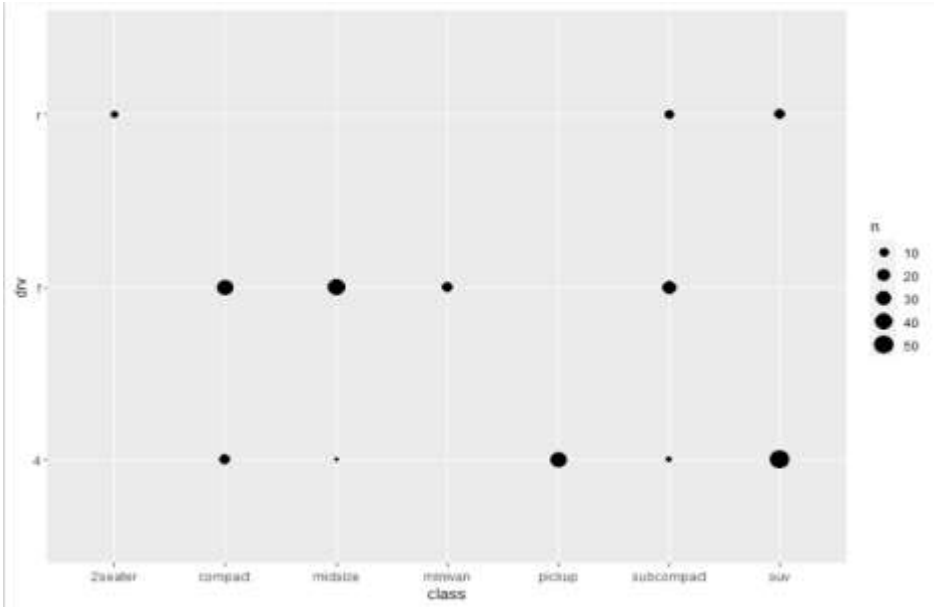
#scatterplot

```
ggplot(mpg,aes(x=hwy,y=cyl,col=manufacturer,shape=manufacturer))+
geom_point()
```



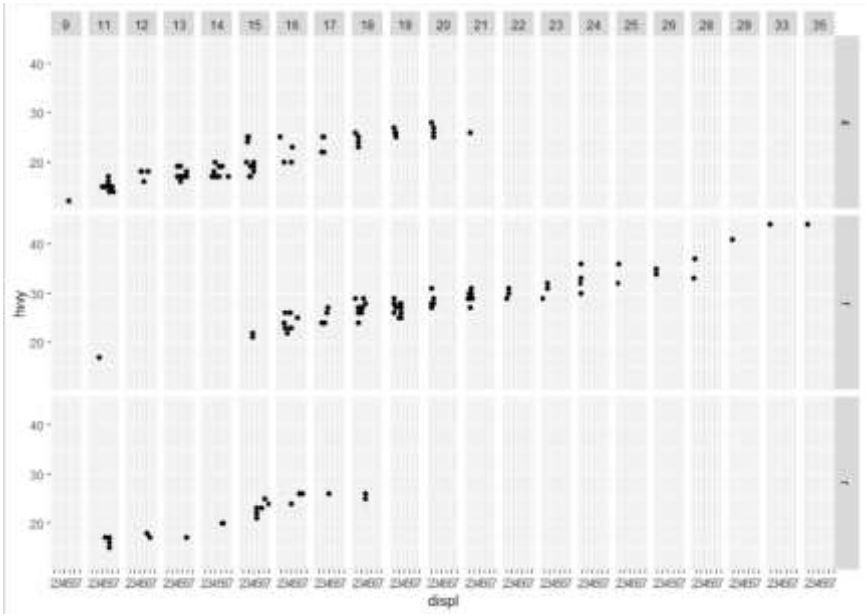
 # 3) C)

```
ggplot(mpg, aes(x = class, y = drv)) +
  geom_count()
```

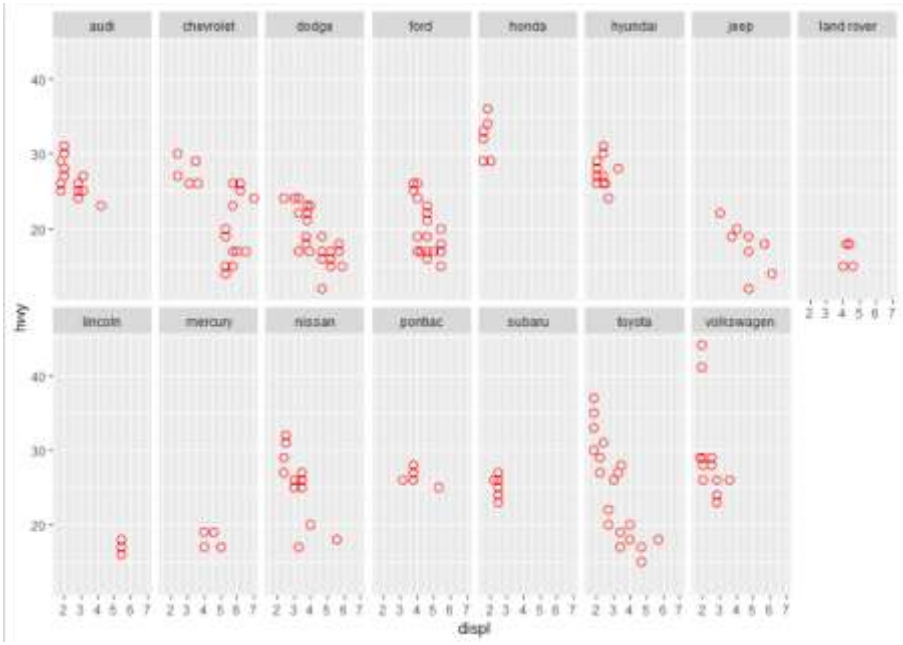


 # 3) D)

```
ggplot(mpg, aes(x = displ, y = hwy)) +
  geom_point()+
  facet_grid(drv~cty)
```

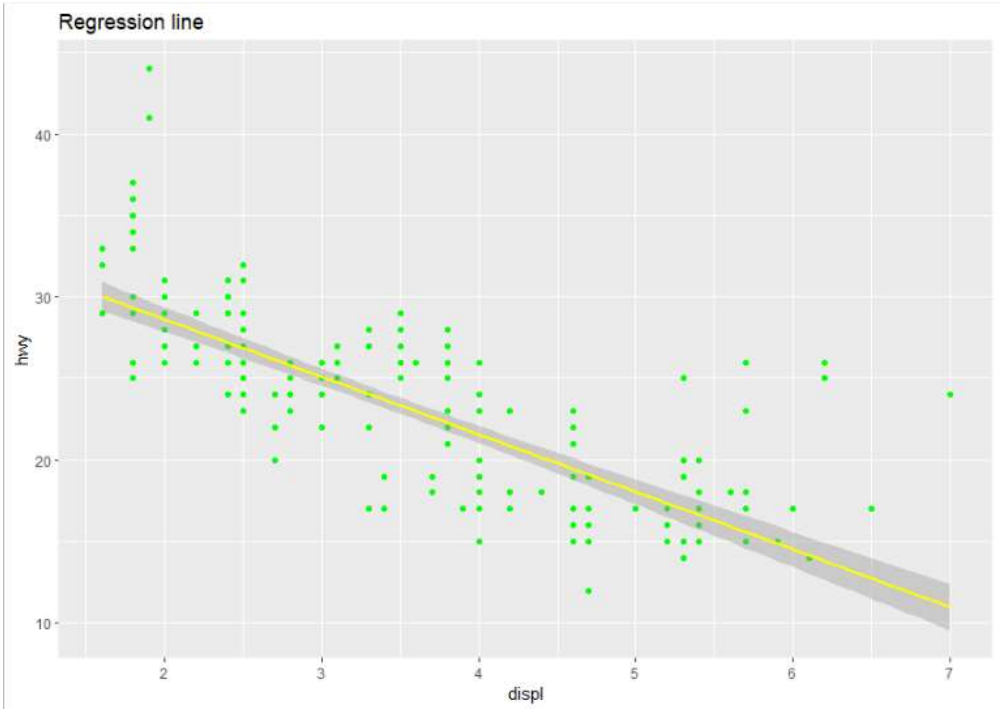


```
ggplot(mpg) +
  geom_point(mapping=aes(x=displ,y=hwy),colour="red",size=3,shape=21)+
  facet_wrap(~manufacturer,nrow=2)
```



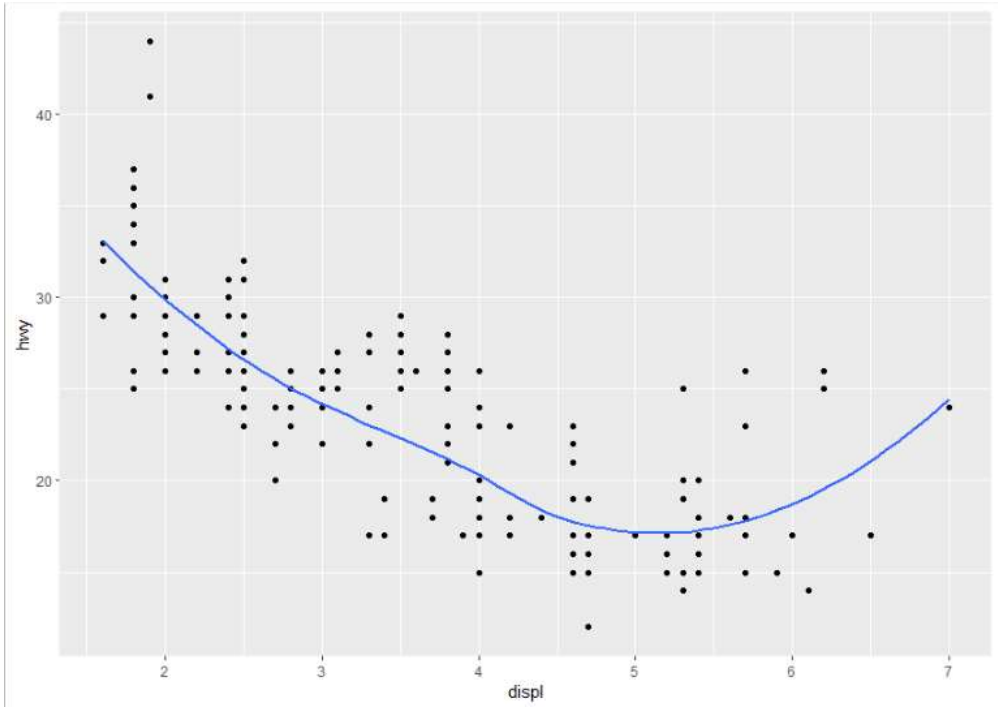
 # 3) E) Regression line

```
ggplot(data = mpg, aes(x = displ, y = hwy)) +
  geom_point(color="green") +
  geom_smooth(method="lm",se=TRUE,color="yellow") +
  labs(title="Regression line",y="hwy", x = "displ")
```



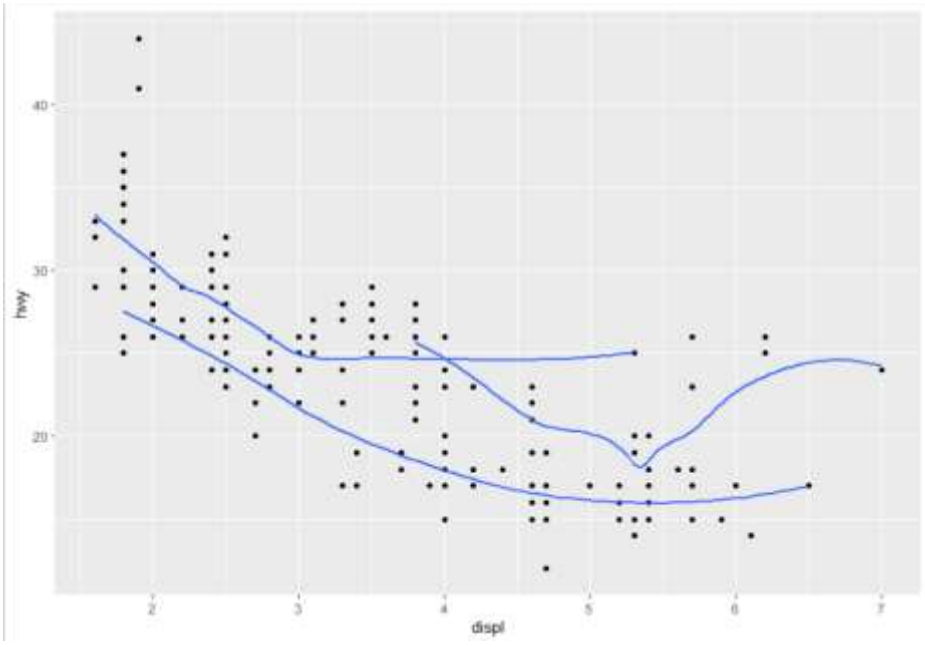
 # 3) F1)

```
ggplot(mpg, aes(x = displ, y = hwy)) +
  geom_point() +
  geom_smooth(se = FALSE)
```



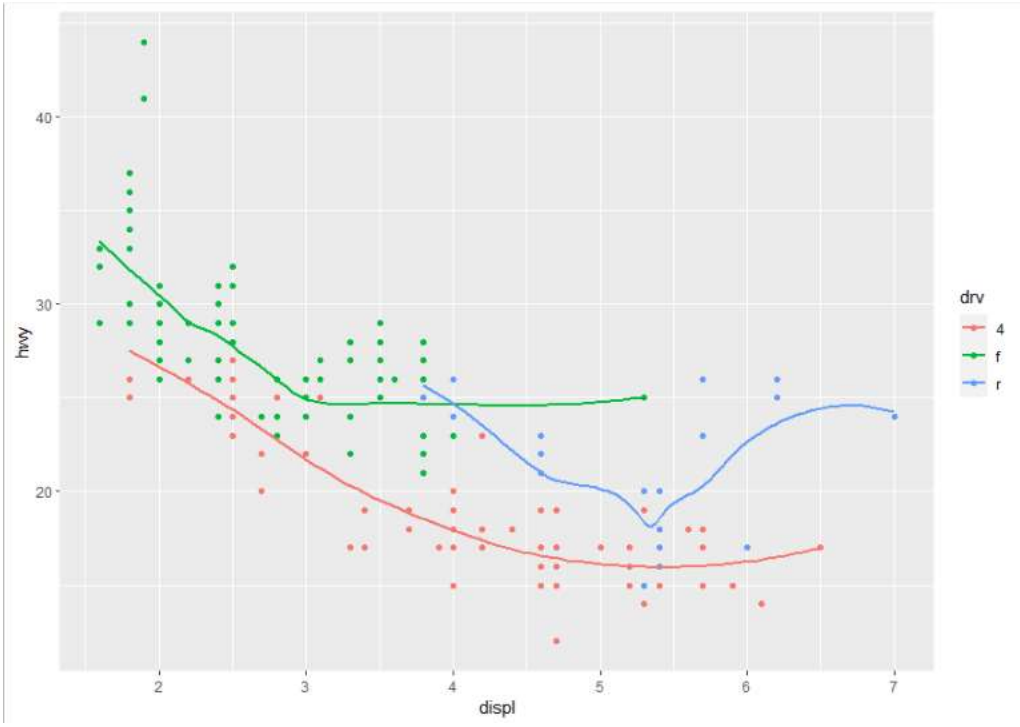
 # 3) F2)

```
ggplot(mpg, aes(x = displ, y = hwy)) +
  geom_point() +
  geom_smooth(mapping = aes(group = drv), se = FALSE)
```



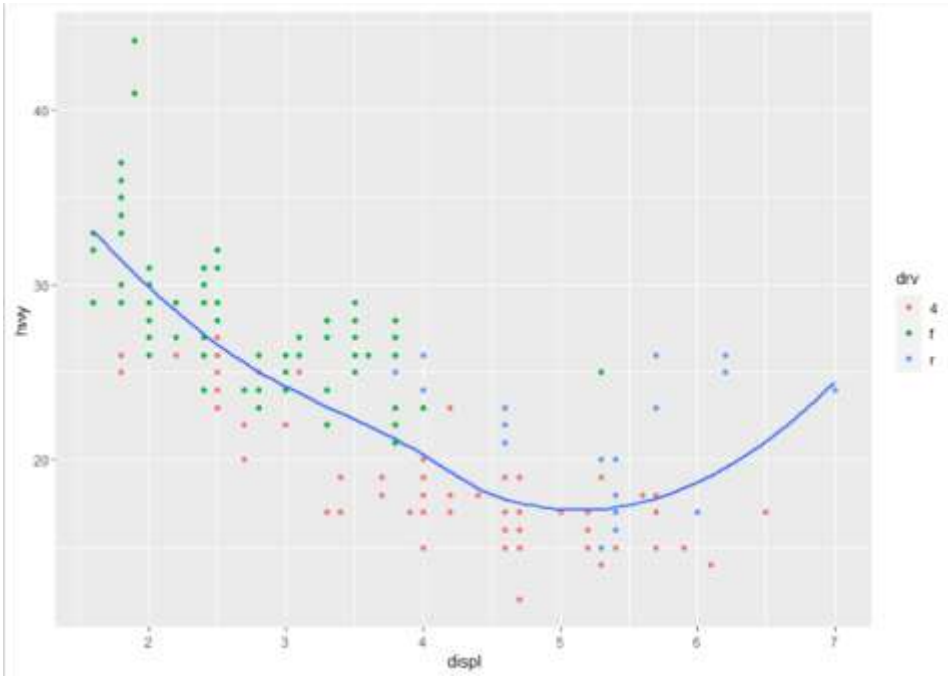
 # 3) F3)

```
ggplot(mpg, aes(x = displ, y = hwy, colour = drv)) +
  geom_point() +
  geom_smooth(se = FALSE)
```



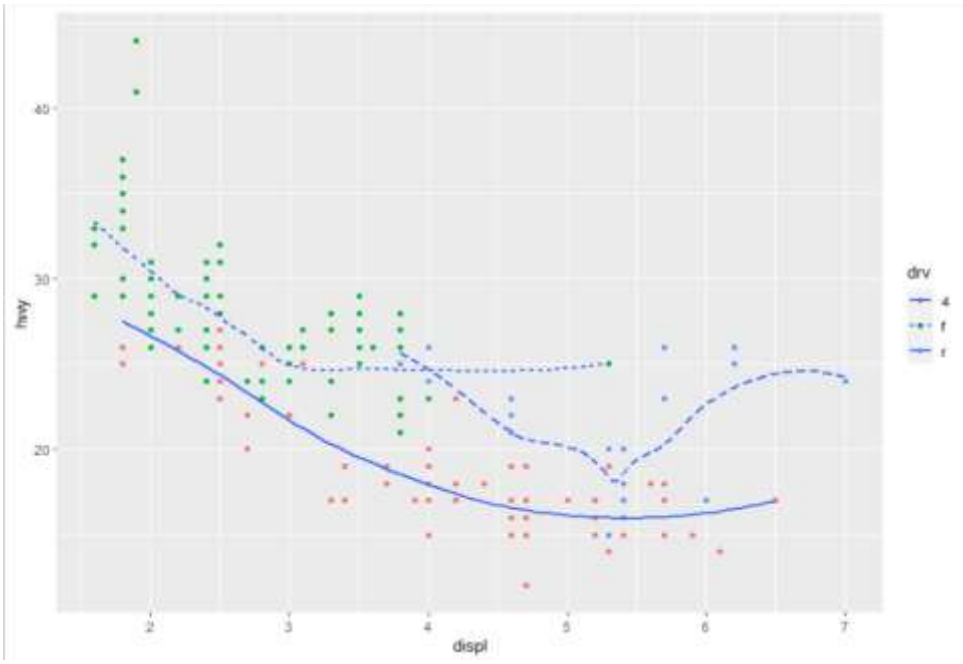
 # 3) F4)

```
ggplot(mpg, aes(x = displ, y = hwy)) +
  geom_point(aes(colour=drv)) +
  geom_smooth(se = FALSE)
```



 # 3) F5)

```
ggplot(mpg, aes(x = displ, y = hwy)) +
  geom_point(aes(colour=drv)) +
  geom_smooth(aes(linetype=drv),se = FALSE)
```



 # 3) F6)

```
ggplot(mpg, aes(x = displ, y = hwy)) +
  geom_point(colour='white',size=3) +
  geom_point(aes(colour=drv))
```