

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

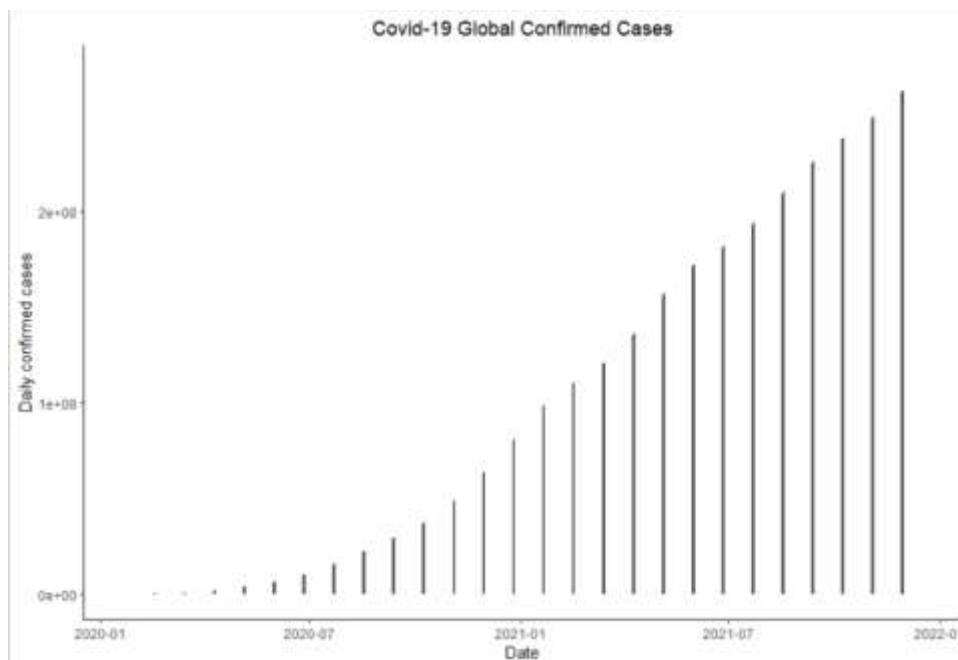
```
library(ggplot2)

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))
```



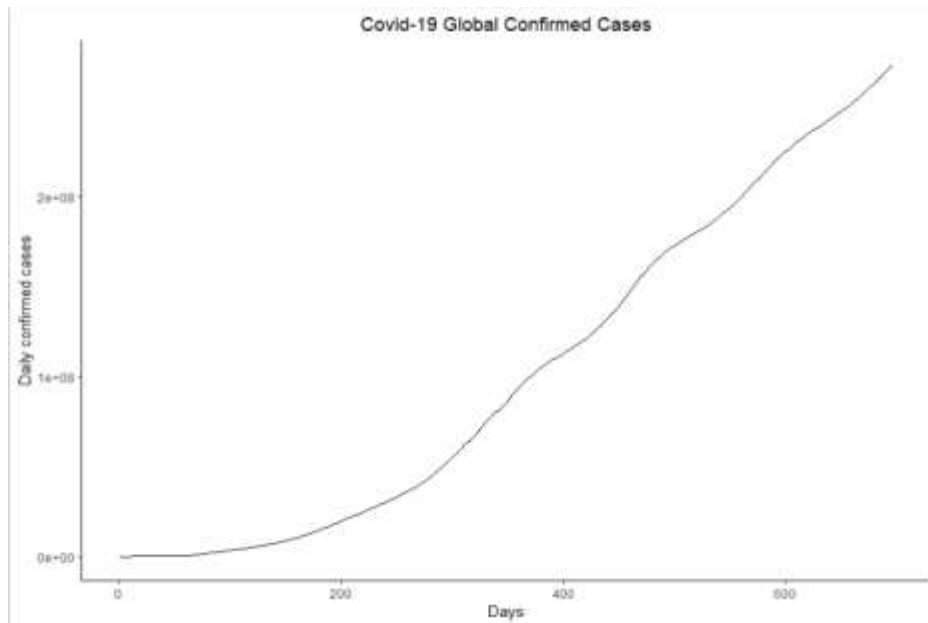
B) Lineplot

```
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))
```

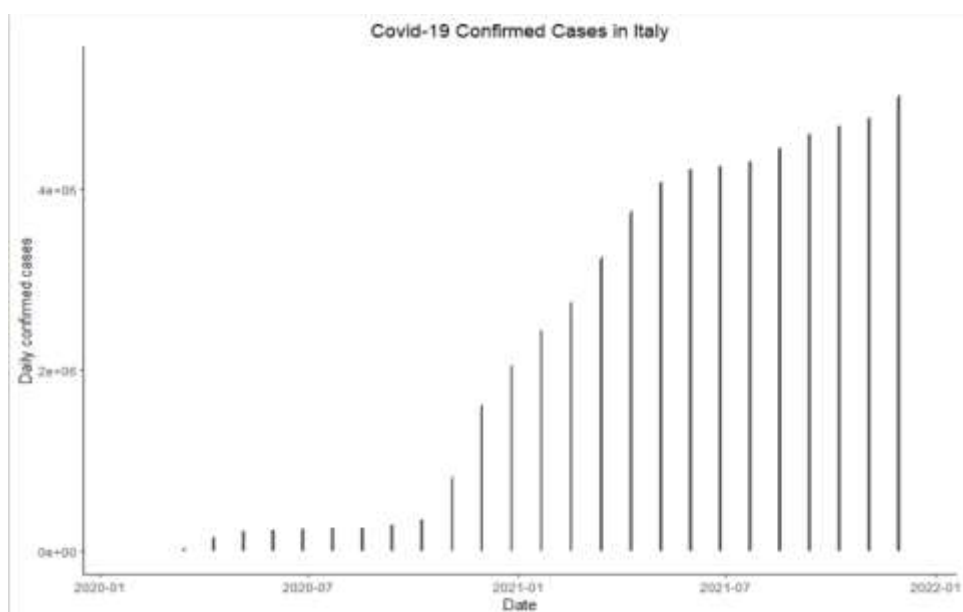


Confirmed Cases in Italy

```
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))
```



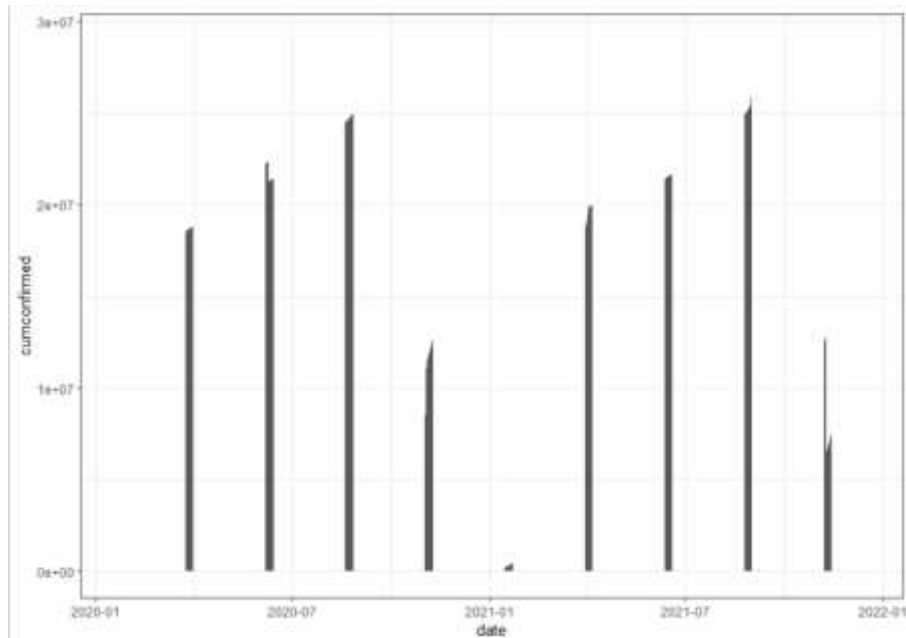
D)

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

#	Country.Region	date	confirmed	deaths	recovered	sumconfirmed	days
1	Australia	2021-01-01	25480	908	23875	25480	1 days
2	Australia	2021-01-02	25474	908	23873	50954	2 days
3	Australia	2021-01-03	25504	909	23927	76458	3 days
4	Australia	2021-01-04	25500	908	23932	101958	4 days
5	Australia	2021-01-05	25500	908	23932	127458	5 days
6	Australia	2021-01-06	25500	908	23932	152958	6 days
7	Australia	2021-01-07	25500	908	23932	178458	7 days
8	Australia	2021-01-08	25500	908	23932	203958	8 days
9	Australia	2021-01-09	25500	908	23932	229458	9 days
10	Australia	2021-01-10	25500	908	23932	254958	10 days
11	Australia	2021-01-11	25500	908	23932	280458	11 days
12	Australia	2021-01-12	25500	908	23932	305958	12 days
13	Australia	2021-01-13	25500	908	23932	331458	13 days
14	Australia	2021-01-14	25500	908	23932	356958	14 days
15	Australia	2021-01-15	25500	908	23932	382458	15 days
16	Australia	2021-01-16	25500	908	23932	407958	16 days
17	Australia	2021-01-17	25500	908	23932	433458	17 days
18	Australia	2021-01-18	25500	908	23932	458958	18 days
19	Australia	2021-01-19	25500	908	23932	484458	19 days
20	Australia	2021-01-20	25500	908	23932	509958	20 days
21	Australia	2021-01-21	25500	908	23932	535458	21 days
22	Australia	2021-01-22	25500	908	23932	560958	22 days
23	Australia	2021-01-23	25500	908	23932	586458	23 days
24	Australia	2021-01-24	25500	908	23932	611958	24 days
25	Australia	2021-01-25	25500	908	23932	637458	25 days
26	Australia	2021-01-26	25500	908	23932	662958	26 days
27	Australia	2021-01-27	25500	908	23932	688458	27 days
28	Australia	2021-01-28	25500	908	23932	713958	28 days
29	Australia	2021-01-29	25500	908	23932	739458	29 days
30	Australia	2021-01-30	25500	908	23932	764958	30 days
31	Australia	2021-01-31	25500	908	23932	790458	31 days
32	Australia	2021-02-01	25500	908	23932	815958	32 days
33	Australia	2021-02-02	25500	908	23932	841458	33 days
34	Australia	2021-02-03	25500	908	23932	866958	34 days
35	Australia	2021-02-04	25500	908	23932	892458	35 days
36	Australia	2021-02-05	25500	908	23932	917958	36 days
37	Australia	2021-02-06	25500	908	23932	943458	37 days
38	Australia	2021-02-07	25500	908	23932	968958	38 days
39	Australia	2021-02-08	25500	908	23932	994458	39 days
40	Australia	2021-02-09	25500	908	23932	1019958	40 days

#	Country.Region	date	confirmed	deaths	recovered	sumconfirmed	days
41	Australia	2021-02-10	25500	908	23932	1045458	41 days
42	Australia	2021-02-11	25500	908	23932	1070958	42 days
43	Australia	2021-02-12	25500	908	23932	1096458	43 days
44	Australia	2021-02-13	25500	908	23932	1121958	44 days
45	Australia	2021-02-14	25500	908	23932	1147458	45 days
46	Australia	2021-02-15	25500	908	23932	1172958	46 days
47	Australia	2021-02-16	25500	908	23932	1198458	47 days
48	Australia	2021-02-17	25500	908	23932	1223958	48 days
49	Australia	2021-02-18	25500	908	23932	1249458	49 days
50	Australia	2021-02-19	25500	908	23932	1274958	50 days
51	Australia	2021-02-20	25500	908	23932	1300458	51 days
52	Australia	2021-02-21	25500	908	23932	1325958	52 days
53	Australia	2021-02-22	25500	908	23932	1351458	53 days
54	Australia	2021-02-23	25500	908	23932	1376958	54 days
55	Australia	2021-02-24	25500	908	23932	1402458	55 days
56	Australia	2021-02-25	25500	908	23932	1427958	56 days
57	Australia	2021-02-26	25500	908	23932	1453458	57 days
58	Australia	2021-02-27	25500	908	23932	1478958	58 days
59	Australia	2021-02-28	25500	908	23932	1504458	59 days
60	Australia	2021-02-29	25500	908	23932	1529958	60 days
61	Australia	2021-03-01	25500	908	23932	1555458	61 days
62	Australia	2021-03-02	25500	908	23932	1580958	62 days
63	Australia	2021-03-03	25500	908	23932	1606458	63 days
64	Australia	2021-03-04	25500	908	23932	1631958	64 days
65	Australia	2021-03-05	25500	908	23932	1657458	65 days
66	Australia	2021-03-06	25500	908	23932	1682958	66 days
67	Australia	2021-03-07	25500	908	23932	1708458	67 days
68	Australia	2021-03-08	25500	908	23932	1733958	68 days
69	Australia	2021-03-09	25500	908	23932	1759458	69 days
70	Australia	2021-03-10	25500	908	23932	1784958	70 days
71	Australia	2021-03-11	25500	908	23932	1810458	71 days
72	Australia	2021-03-12	25500	908	23932	1835958	72 days
73	Australia	2021-03-13	25500	908	23932	1861458	73 days
74	Australia	2021-03-14	25500	908	23932	1886958	74 days
75	Australia	2021-03-15	25500	908	23932	1912458	75 days
76	Australia	2021-03-16	25500	908	23932	1937958	76 days
77	Australia	2021-03-17	25500	908	23932	1963458	77 days
78	Australia	2021-03-18	25500	908	23932	1988958	78 days
79	Australia	2021-03-19	25500	908	23932	2014458	79 days
80	Australia	2021-03-20	25500	908	23932	2039958	80 days

```
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

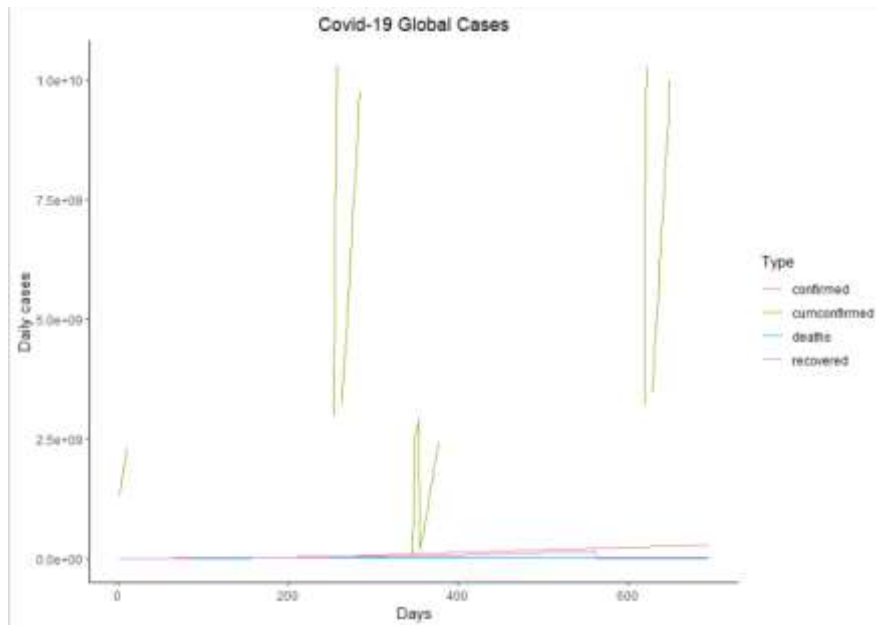
```
str(world)
```

```

#> tibble [1:695] (S3: tbl_df/tbl/data.frame)
#>   recovered : int [1:695] 30 32 39 42 56 65 108 127 145 225 ...
#>   $ date      : Date [1:695], format: "2020-01-22" "2020-01-23" ...
#>   $ confirmed : int [1:695] 557 655 941 1434 2118 2927 5578 6167 8235 9927 ...
#>   $ cumconfirmed: num [1:695] 1.3e+09 1.4e+09 1.5e+09 1.6e+09 1.7e+09 ...
#>   $ deaths    : int [1:695] 17 18 26 42 56 82 131 133 171 213 ...
#>   $ recovered  : int [1:695] 30 32 39 42 56 65 108 127 145 225 ...
#>   $ days       : 'difftime' num [1:695] 1 2 3 4 ...
#>   .. attr(,"units")= chr "days"

```

```
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)
```

```

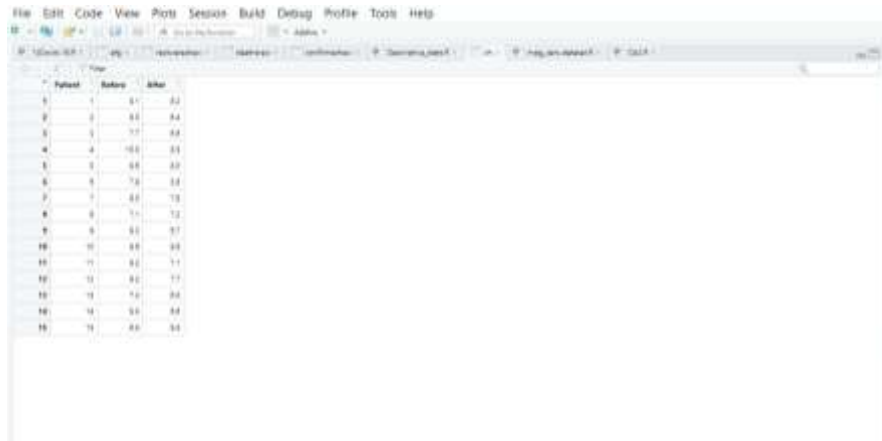
> mean(x)
[1] 80
> median(x)
[1] 85
> library(modeest)
> mode=mfv(x)
> print(mode)
[1] 95
> #trimmed mean
> mean(x,trim=0.5)
[1] 85
>

```

#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)
```



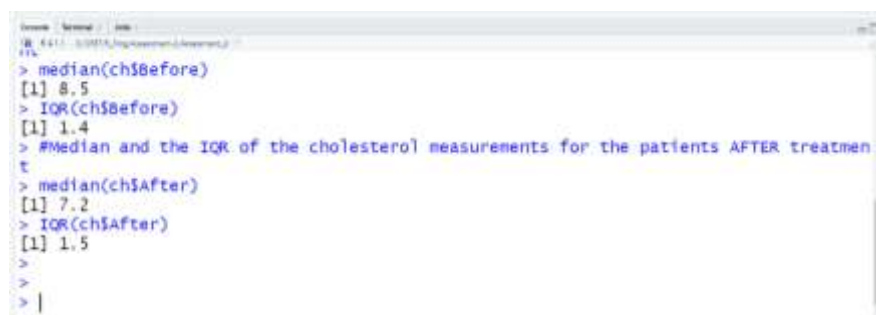
Patient	Before	After
1	9.1	8.2
2	8.0	6.4
3	7.7	6.6
4	10.0	8.5
5	9.6	8.0
6	7.9	5.8
7	9.0	7.8
8	7.1	7.2
9	8.3	6.7
10	9.6	9.8
11	8.2	7.1
12	9.2	7.7
13	7.3	6.0
14	8.5	6.6
15	9.5	8.4

#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 treatment
> 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
>

```

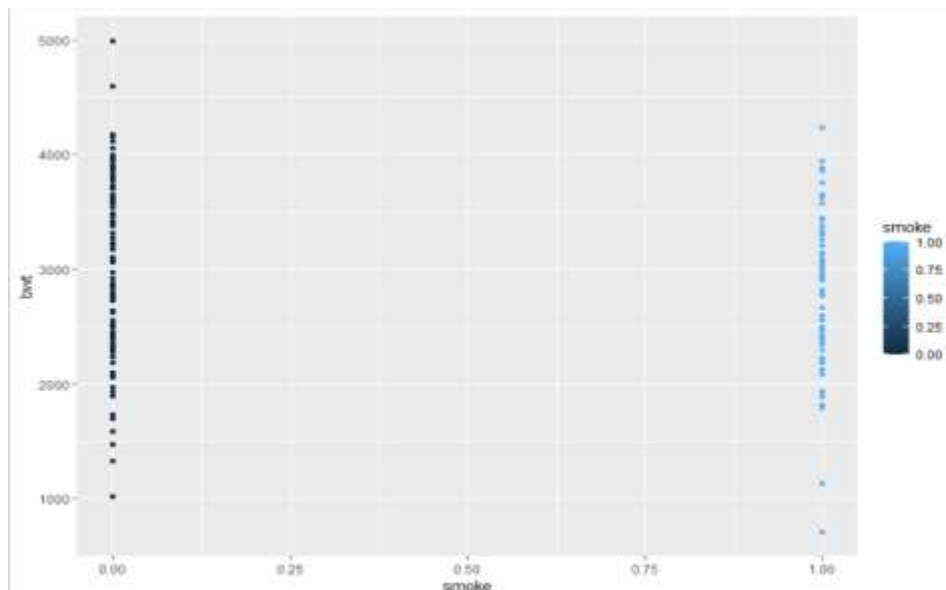
#null hypothesis is rejected because p-value is lesser than 0.05

#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)
```

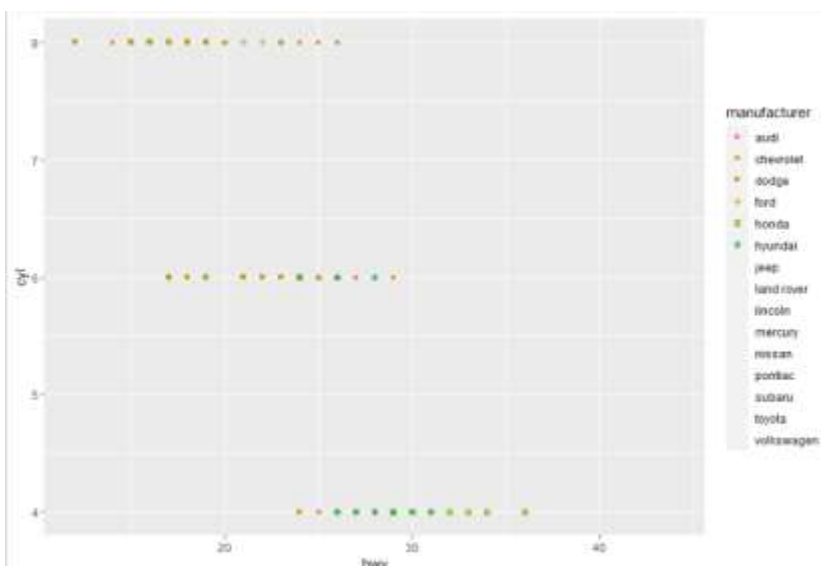
```
> data=mpg
> #Displaying the information about the dataset
> str(data)
tibble [234 x 11] (s3: tbl_df/tbl/data.frame)
 $ manufacturer: chr [1:234] "audi" "audi" "audi" "audi" ...
 $ model       : chr [1:234] "a4" "a4" "a4" "a4" ...
 $ displ      : num [1:234] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
 $ 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

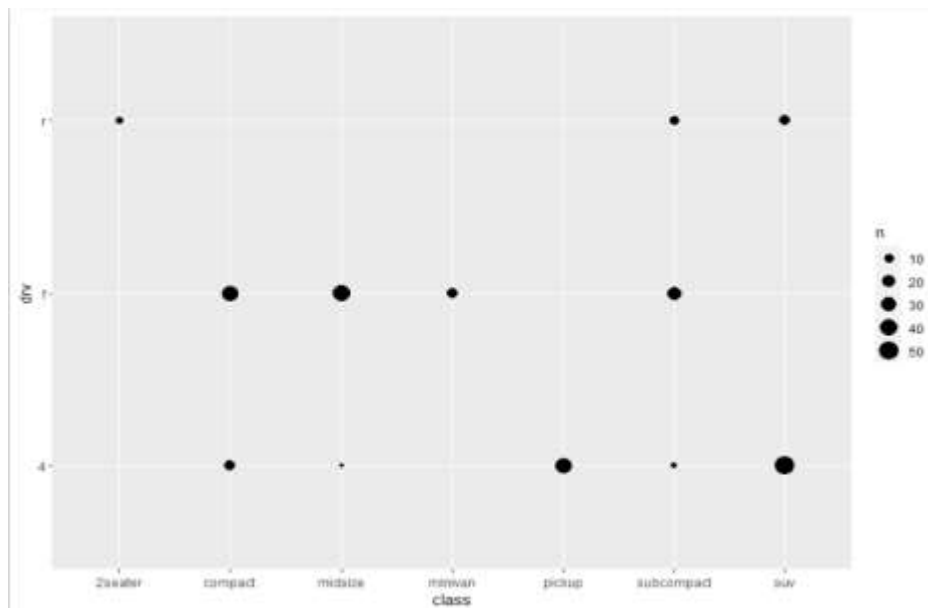
```
ggplot(mpg,aes(x=hwy,y=cyl,col=manufacturer,shape=manufacturer))+
```

```
geom_point()
```



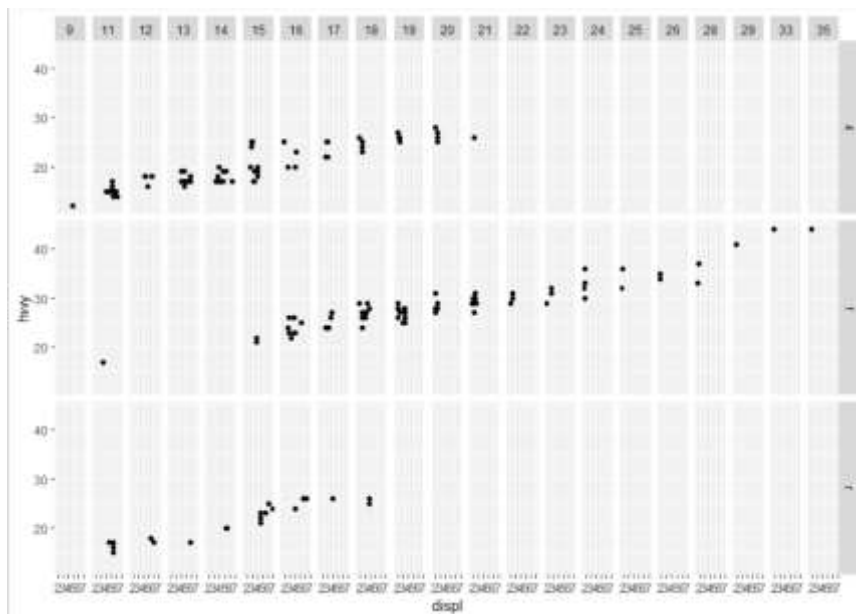
3) C)

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

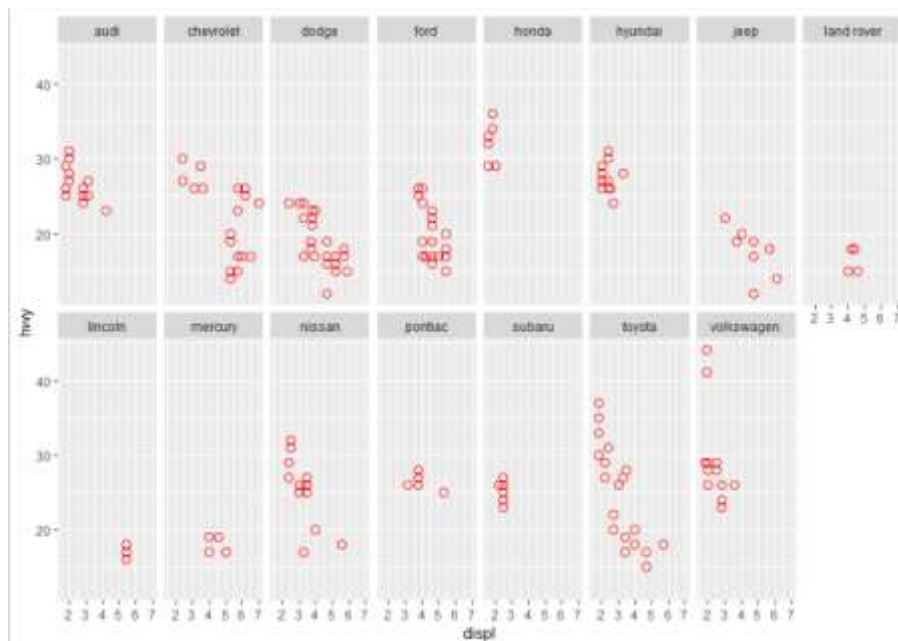


3) D)

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

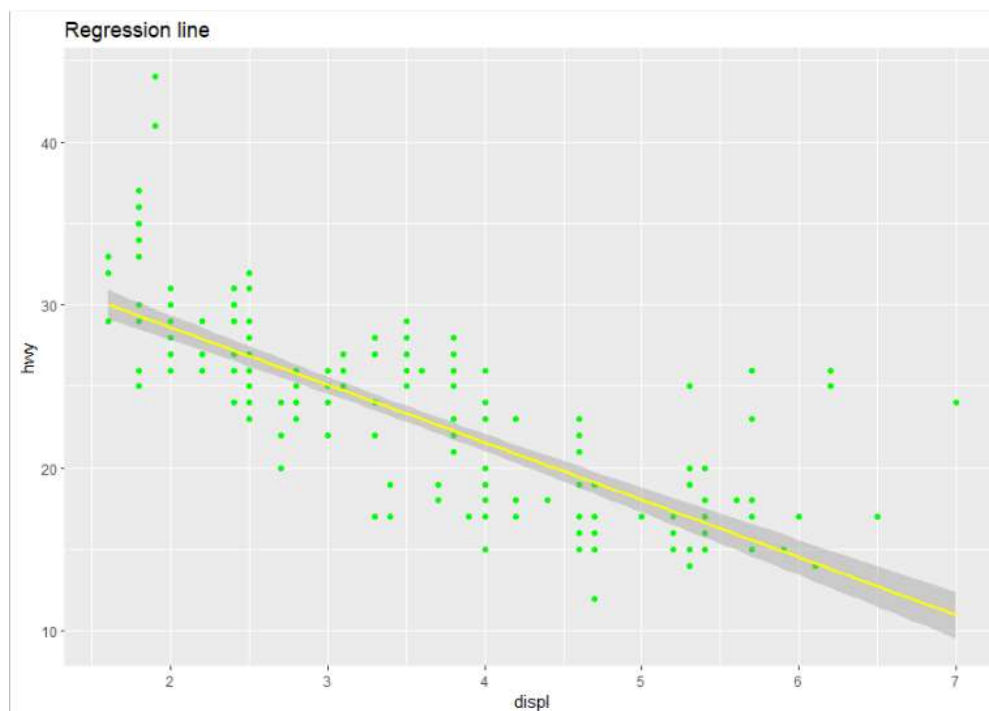


```
ggplot(mpg) +
  geom_point(mapping=aes(x=displ,y=hwy),colour="red",size=3,shape=21)+
  facet_wrap(~manufacturer,nrow=2)
```



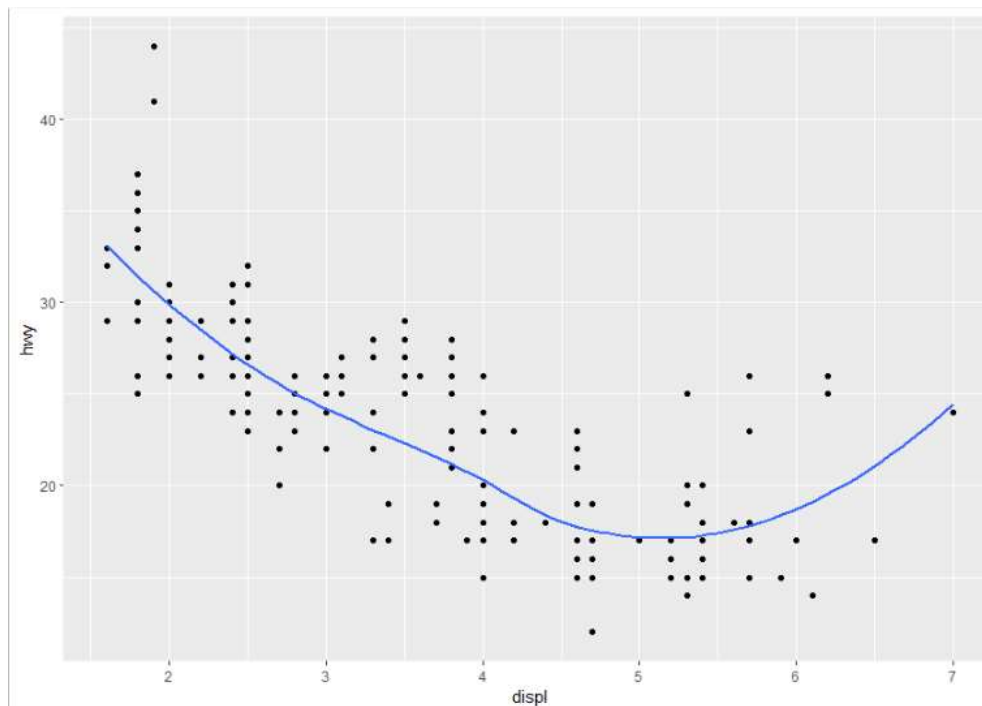
3) E) Regression line

```
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")
```



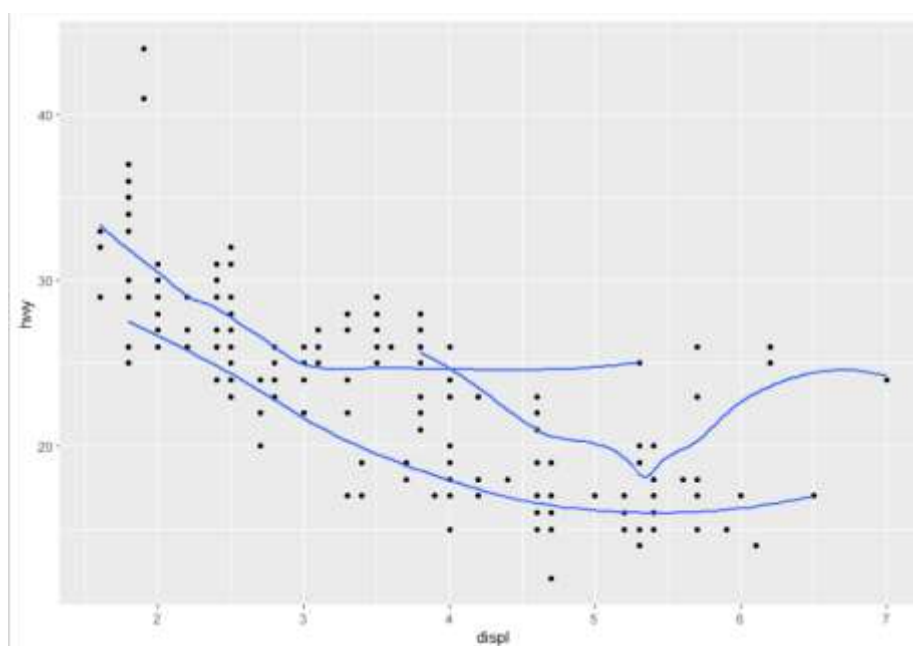
3) F1)

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



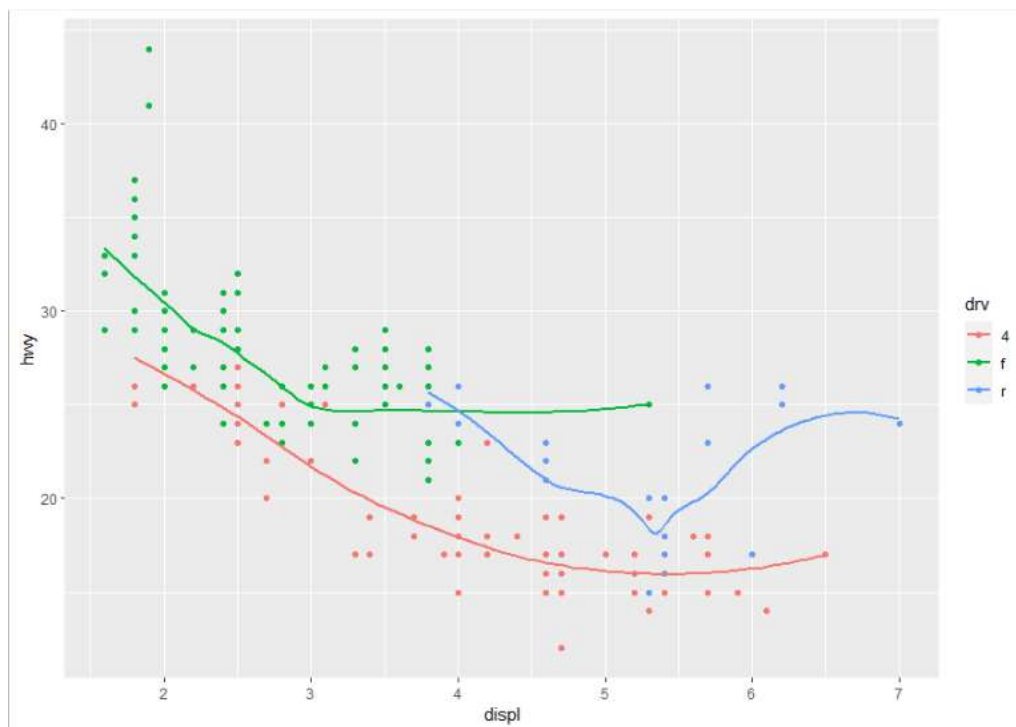
3) F2)

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



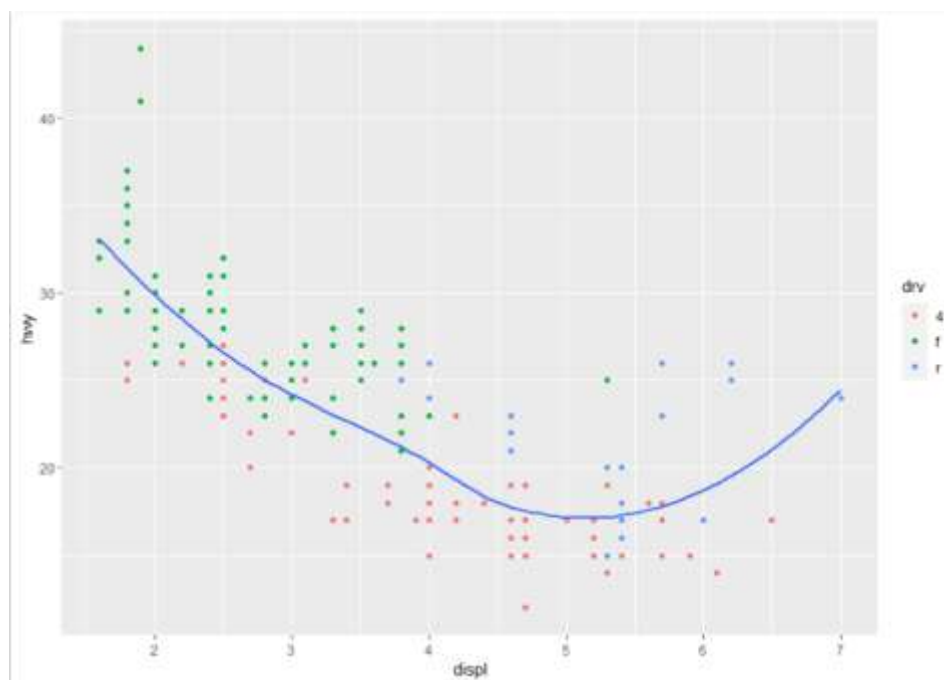
3) F3)

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



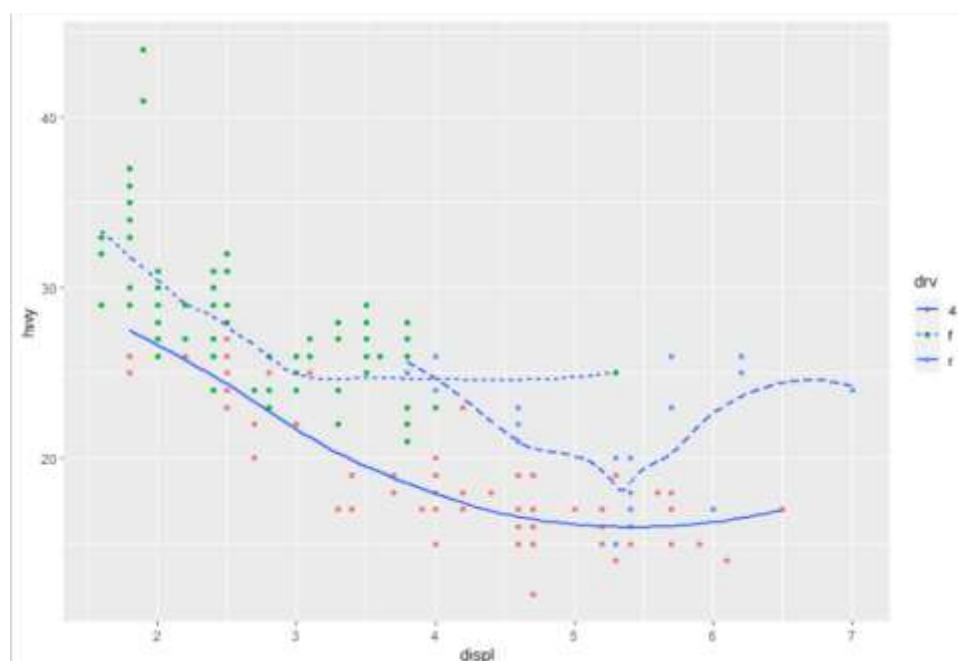
3) F4)

```
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))
```

