CSE3020 - Data Visualisation

Lab Mid Term

SANJIT KUMAR
18BCE0715
DR NALINI N
LAB - L13 + L14

Question/Task - No 4

Consider the mpg data set in the dplyr and plotly/ggplot2 package.
a. Which car(s) had the highest highway gas mileage? (For the purposes of this question, consider each observation a different car.)
b. Compute the mean city mileage for compact cars.
c. Compute the mean city mileage for each class of cars, and arrange in decreasing order.
d. Which cars have the smallest absolute difference between highway mileage and city mileage? (For the purposes of this question, consider each observation a different "car".)
e. Compute the mean highway mileage for each year, and arrange in decreasing order.
f. Show visualization for any one(above query).

Aim - To understand the central tendency of milage in the given mpg dataset

A)

 # Cars With Maximum Highway Milage

```
df[df$hwy == max(df$hwy),]
```

```
> df[df$hwy == max(df$hwy),]
# A tibble: 2 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 volkswagen   jetta        1.9  1999     4 manual(m5) f        33    44 d     compact
2 volkswagen   new beetle   1.9  1999     4 manual(m5) f        35    44 d     subcompact
```

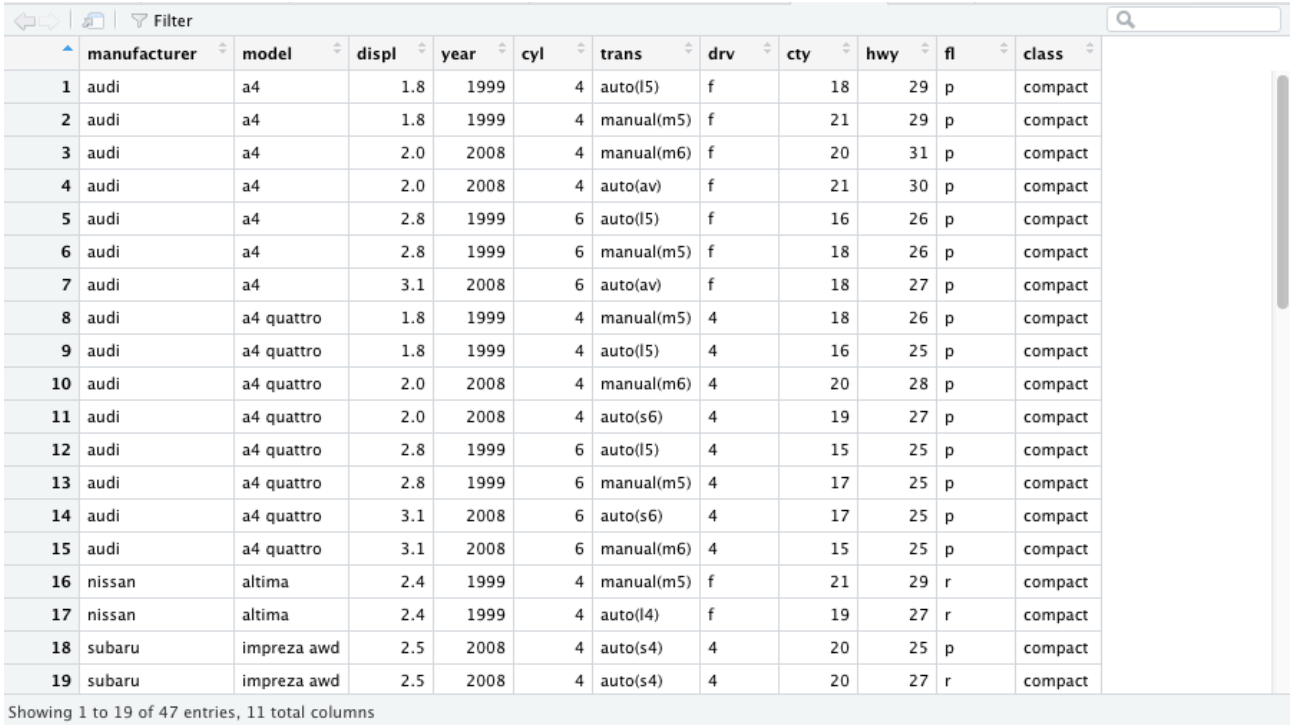
B)

 # Mean City Milage for Compact Cars

```
df2 = filter(df,df$class=="compact")

mean_city_milage = mean(df2$cty)

mean_city_milage
```



Filter

	manufacturer	model	displ	year	cyl	trans	drv	cty	hwy	fl	class
1	audi	a4	1.8	1999	4	auto(l5)	f	18	29	p	compact
2	audi	a4	1.8	1999	4	manual(m5)	f	21	29	p	compact
3	audi	a4	2.0	2008	4	manual(m6)	f	20	31	p	compact
4	audi	a4	2.0	2008	4	auto(av)	f	21	30	p	compact
5	audi	a4	2.8	1999	6	auto(l5)	f	16	26	p	compact
6	audi	a4	2.8	1999	6	manual(m5)	f	18	26	p	compact
7	audi	a4	3.1	2008	6	auto(av)	f	18	27	p	compact
8	audi	a4 quattro	1.8	1999	4	manual(m5)	4	18	26	p	compact
9	audi	a4 quattro	1.8	1999	4	auto(l5)	4	16	25	p	compact
10	audi	a4 quattro	2.0	2008	4	manual(m6)	4	20	28	p	compact
11	audi	a4 quattro	2.0	2008	4	auto(s6)	4	19	27	p	compact
12	audi	a4 quattro	2.8	1999	6	auto(l5)	4	15	25	p	compact
13	audi	a4 quattro	2.8	1999	6	manual(m5)	4	17	25	p	compact
14	audi	a4 quattro	3.1	2008	6	auto(s6)	4	17	25	p	compact
15	audi	a4 quattro	3.1	2008	6	manual(m6)	4	15	25	p	compact
16	nissan	altima	2.4	1999	4	manual(m5)	f	21	29	r	compact
17	nissan	altima	2.4	1999	4	auto(l4)	f	19	27	r	compact
18	subaru	impreza awd	2.5	2008	4	auto(s4)	4	20	25	p	compact
19	subaru	impreza awd	2.5	2008	4	auto(s4)	4	20	27	r	compact

Showing 1 to 19 of 47 entries, 11 total columns

```
> mean_city_milage = mean(df2$cty)
> mean_city_milage
[1] 20.12766
```

C)

 # Group by Class and Find Mean in desc

```
mean_classes = group_by(df, class) %>% summarize(m = mean(cty)) %>%
arrange(desc(m))

mean_classes
```

```
> # Group by Class and Find Mean
> mean_classes = group_by(df, class) %>% summarize(m = mean(cty)) %>% arrange(desc(m))
> mean_classes
# A tibble: 7 x 2
  class          m
  <chr>      <dbl>
1 subcompact  20.4
2 compact     20.1
3 midsize     18.8
4 minivan     15.8
5 2seater     15.4
6 suv         13.5
7 pickup      13
>
```

D)

 # cars with smallest absolute difference between highway and city milage

df$absdiff = abs(df$hwy - df$cty)

df[df$absdiff == max(df$absdiff),]

```
> df[df$absdiff == max(df$absdiff),]
# A tibble: 2 x 12
  manufacturer model      displ  year   cyl trans    drv     cty   hwy fl    class      absdiff
  <chr>        <chr>      <dbl> <int> <int> <chr>    <chr> <int> <int> <chr> <chr>        <int>
1 honda        civic        1.8  2008     4 auto(l5) f        24    36 c     subcompact      12
2 volkswagen   new beetle   1.9  1999     4 auto(l4) f        29    41 d     subcompact      12
>
```

E)

#mean highway milage of each year in desc

mean_year_hwy = group_by(df, year) %>% summarize(m = mean(hwy)) %>% arrange(desc(m))

mean_year_hwy

```
> #mean highway milage of each year
> mean_year_hwy = group_by(df, year) %>% summarize(m = mean(hwy))  %>% arrange(desc(m))
> mean_year_hwy
# A tibble: 2 x 2
   year     m
  <int> <dbl>
1  2008  23.5
2  1999  23.4
>
```

F)

#visualisation of mean milage by class

```
p<-ggplot(mean_classes, aes(x=class, y=m)) + geom_bar(stat='identity')
p
```