Classwork\_Practise\_Chaytanya

Chaytanya

10/1/2021

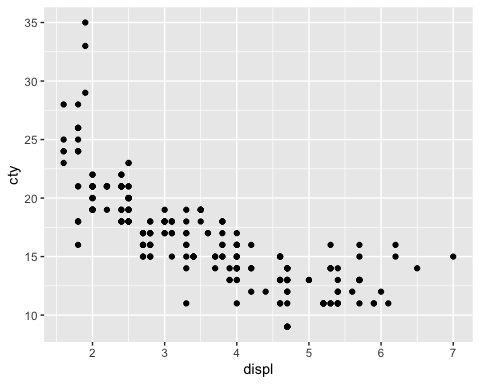
#PRACTISE TEST   
#GGPLOT PRACTISE   
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

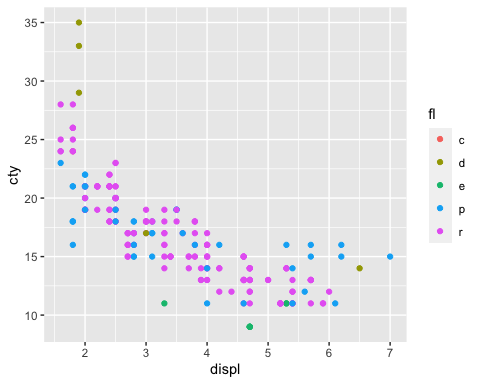
## ✓ ggplot2 3.3.5 ✓ purrr 0.3.4  
## ✓ tibble 3.1.4 ✓ dplyr 1.0.7  
## ✓ tidyr 1.1.3 ✓ stringr 1.4.0  
## ✓ readr 2.0.1 ✓ forcats 0.5.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

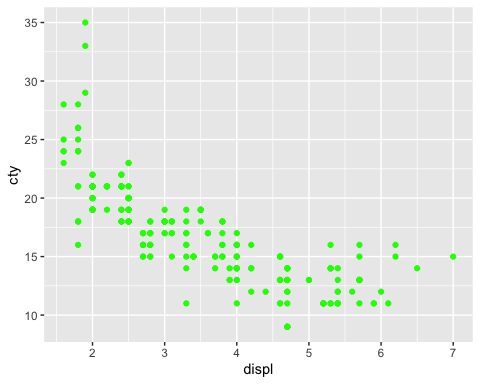
library(dplyr)  
library(ggplot2)  
#install.packages("nycflights13")  
library(nycflights13)  
#view(mpg)  
  
#Question1  
ggplot(data=mpg) +  
 geom\_point(mapping = aes(x = displ, y=cty))



#Question2  
  
ggplot(data=mpg) +  
 geom\_point(mapping = aes(x = displ, y=cty, color = fl))



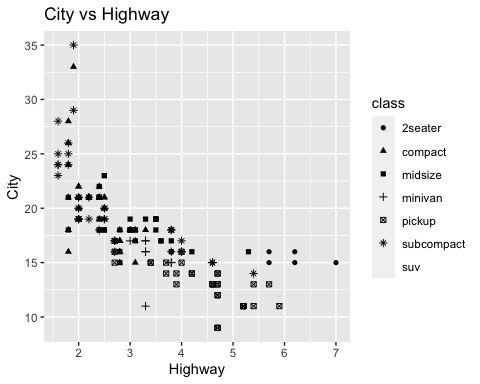
#Question3  
  
ggplot(data=mpg) +  
 geom\_point(mapping = aes(x = displ, y=cty),color ="green")



#Question4   
  
ggplot(data=mpg) +  
 geom\_point(mapping = aes(x = displ, y=cty,shape=class))+  
 ggtitle("City vs Highway")+  
 xlab("Highway")+  
 ylab("City")

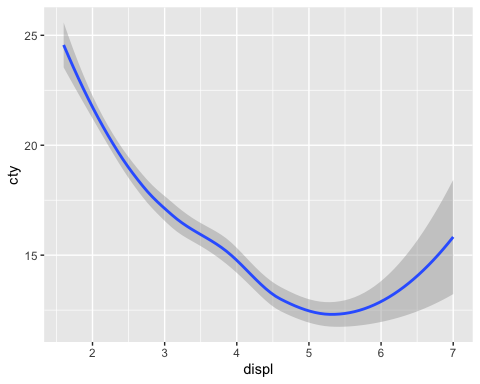
## Warning: The shape palette can deal with a maximum of 6 discrete values because  
## more than 6 becomes difficult to discriminate; you have 7. Consider  
## specifying shapes manually if you must have them.

## Warning: Removed 62 rows containing missing values (geom\_point).



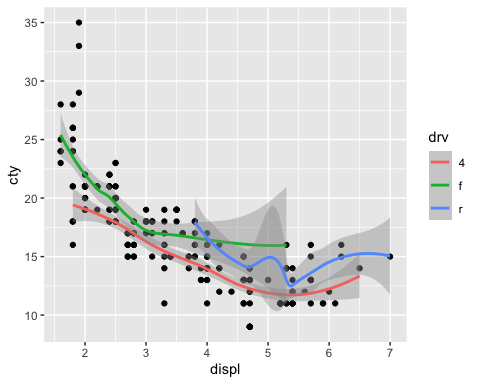
#Question5  
  
ggplot(data=mpg) +  
 geom\_smooth(mapping=aes(x=displ,y=cty))

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



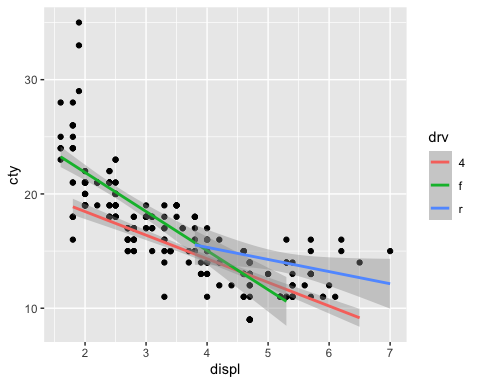
#Question6  
  
  
ggplot(data=mpg) +  
 geom\_point(mapping = aes(x=displ, y=cty)) +  
 geom\_smooth( mapping=aes(x=displ,y=cty,color=drv) )

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

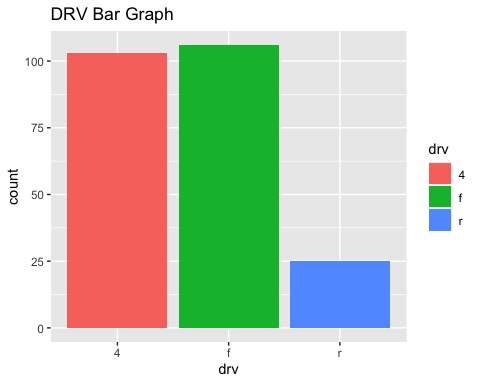


#Question7  
  
ggplot(data=mpg) +  
 geom\_point(mapping = aes(x=displ, y=cty)) +  
 geom\_smooth(method=lm, mapping=aes(x=displ,y=cty,color=drv) )

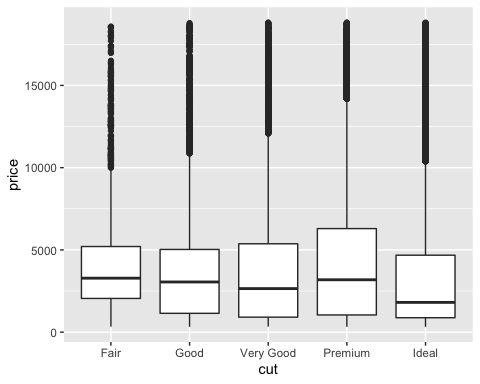
## `geom\_smooth()` using formula 'y ~ x'



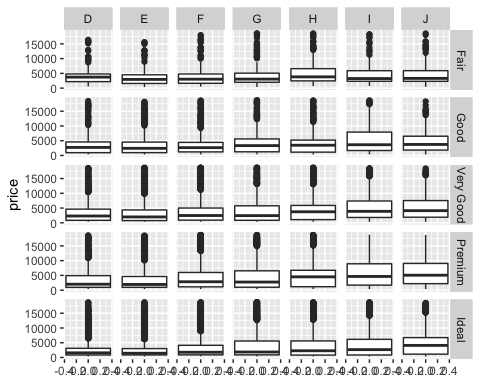
#Question8  
  
ggplot(data=mpg) +  
 geom\_bar(mapping = aes(x= drv,fill=drv))+  
 ggtitle("DRV Bar Graph")



#Question9 and Question10 - As discussed in the classs,Skipped.   
  
#Question11  
ggplot(data = diamonds) +  
 geom\_boxplot(mapping = aes(x=cut , y=price))

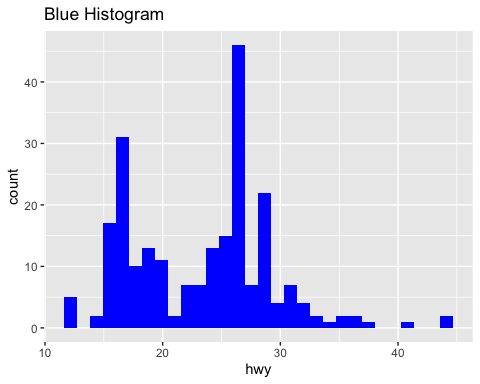


#Question12  
ggplot(data = diamonds) +  
 geom\_boxplot(mapping = aes(y=price))+  
 facet\_grid(cut~color)



#Question13  
  
ggplot(data = mpg) +  
 geom\_histogram(mapping = aes(x=hwy),fill="blue")+  
 ggtitle("Blue Histogram")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



#PART2  
 #Question1  
x1<-filter(mpg,manufacturer=="dodge",model=="durango 4wd") # (double =)  
head(x1)

## # A tibble: 6 × 11  
## manufacturer model displ year cyl trans drv cty hwy fl class  
## <chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>  
## 1 dodge durango 4wd 3.9 1999 6 auto… 4 13 17 r suv   
## 2 dodge durango 4wd 4.7 2008 8 auto… 4 13 17 r suv   
## 3 dodge durango 4wd 4.7 2008 8 auto… 4 9 12 e suv   
## 4 dodge durango 4wd 4.7 2008 8 auto… 4 13 17 r suv   
## 5 dodge durango 4wd 5.2 1999 8 auto… 4 11 16 r suv   
## 6 dodge durango 4wd 5.7 2008 8 auto… 4 13 18 r suv

#Question2  
  
x2<-filter(mpg, cty<10 & hwy<16) # (double =)  
head(x2)

## # A tibble: 5 × 11  
## manufacturer model displ year cyl trans drv cty hwy fl class  
## <chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>  
## 1 dodge dakota pi… 4.7 2008 8 auto(… 4 9 12 e pick…  
## 2 dodge durango 4… 4.7 2008 8 auto(… 4 9 12 e suv   
## 3 dodge ram 1500 … 4.7 2008 8 auto(… 4 9 12 e pick…  
## 4 dodge ram 1500 … 4.7 2008 8 manua… 4 9 12 e pick…  
## 5 jeep grand che… 4.7 2008 8 auto(… 4 9 12 e suv

#Question3  
  
x3<-mpg%>%  
arrange(desc(displ))  
head(x3)

## # A tibble: 6 × 11  
## manufacturer model displ year cyl trans drv cty hwy fl class  
## <chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>  
## 1 chevrolet corvette 7 2008 8 manual… r 15 24 p 2sea…  
## 2 chevrolet k1500 ta… 6.5 1999 8 auto(l… 4 14 17 d suv   
## 3 chevrolet corvette 6.2 2008 8 manual… r 16 26 p 2sea…  
## 4 chevrolet corvette 6.2 2008 8 auto(s… r 15 25 p 2sea…  
## 5 jeep grand ch… 6.1 2008 8 auto(l… 4 11 14 p suv   
## 6 chevrolet c1500 su… 6 2008 8 auto(l… r 12 17 r suv

#Question4  
  
x4<-mpg%>%  
arrange((cty))%>%  
print(n=30)

## # A tibble: 234 × 11  
## manufacturer model displ year cyl trans drv cty hwy fl class  
## <chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>  
## 1 dodge dakota p… 4.7 2008 8 auto(… 4 9 12 e pick…  
## 2 dodge durango … 4.7 2008 8 auto(… 4 9 12 e suv   
## 3 dodge ram 1500… 4.7 2008 8 auto(… 4 9 12 e pick…  
## 4 dodge ram 1500… 4.7 2008 8 manua… 4 9 12 e pick…  
## 5 jeep grand ch… 4.7 2008 8 auto(… 4 9 12 e suv   
## 6 chevrolet c1500 su… 5.3 2008 8 auto(… r 11 15 e suv   
## 7 chevrolet k1500 ta… 5.3 2008 8 auto(… 4 11 14 e suv   
## 8 chevrolet k1500 ta… 5.7 1999 8 auto(… 4 11 15 r suv   
## 9 dodge caravan … 3.3 2008 6 auto(… f 11 17 e mini…  
## 10 dodge dakota p… 5.2 1999 8 manua… 4 11 17 r pick…  
## 11 dodge dakota p… 5.2 1999 8 auto(… 4 11 15 r pick…  
## 12 dodge durango … 5.2 1999 8 auto(… 4 11 16 r suv   
## 13 dodge durango … 5.9 1999 8 auto(… 4 11 15 r suv   
## 14 dodge ram 1500… 5.2 1999 8 auto(… 4 11 15 r pick…  
## 15 dodge ram 1500… 5.2 1999 8 manua… 4 11 16 r pick…  
## 16 dodge ram 1500… 5.9 1999 8 auto(… 4 11 15 r pick…  
## 17 ford expediti… 4.6 1999 8 auto(… r 11 17 r suv   
## 18 ford expediti… 5.4 1999 8 auto(… r 11 17 r suv   
## 19 ford f150 pic… 5.4 1999 8 auto(… 4 11 15 r pick…  
## 20 jeep grand ch… 6.1 2008 8 auto(… 4 11 14 p suv   
## 21 land rover range ro… 4 1999 8 auto(… 4 11 15 p suv   
## 22 land rover range ro… 4.6 1999 8 auto(… 4 11 15 p suv   
## 23 lincoln navigato… 5.4 1999 8 auto(… r 11 17 r suv   
## 24 lincoln navigato… 5.4 1999 8 auto(… r 11 16 p suv   
## 25 toyota land cru… 4.7 1999 8 auto(… 4 11 15 r suv   
## 26 chevrolet c1500 su… 6 2008 8 auto(… r 12 17 r suv   
## 27 dodge ram 1500… 4.7 2008 8 manua… 4 12 16 r pick…  
## 28 dodge ram 1500… 4.7 2008 8 manua… 4 12 16 r pick…  
## 29 ford expediti… 5.4 2008 8 auto(… r 12 18 r suv   
## 30 land rover range ro… 4.2 2008 8 auto(… 4 12 18 r suv   
## # … with 204 more rows

#"dakota pickup 4wd , durango 4wd , ram 1500 pickup 4wd (auto), ram 1500 pickup 4wd(manual), grand cherokee 4wd has the lowest city milege   
  
#Question5  
  
 mpg%>%  
filter(manufacturer=="ford",year==1999,cty<=16 & hwy<16)

## # A tibble: 1 × 11  
## manufacturer model displ year cyl trans drv cty hwy fl class  
## <chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>  
## 1 ford f150 pick… 5.4 1999 8 auto(… 4 11 15 r pick…

#Question6  
   
 mpg%>%  
 filter(manufacturer=="ford",model=="mustang")%>%  
 mutate(difference= hwy-cty)%>%  
 select(manufacturer,model,cty,hwy,difference)

## # A tibble: 9 × 5  
## manufacturer model cty hwy difference  
## <chr> <chr> <int> <int> <int>  
## 1 ford mustang 18 26 8  
## 2 ford mustang 18 25 7  
## 3 ford mustang 17 26 9  
## 4 ford mustang 16 24 8  
## 5 ford mustang 15 21 6  
## 6 ford mustang 15 22 7  
## 7 ford mustang 15 23 8  
## 8 ford mustang 15 22 7  
## 9 ford mustang 14 20 6

# Question7  
 flights

## # A tibble: 336,776 × 19  
## year month day dep\_time sched\_dep\_time dep\_delay arr\_time sched\_arr\_time  
## <int> <int> <int> <int> <int> <dbl> <int> <int>  
## 1 2013 1 1 517 515 2 830 819  
## 2 2013 1 1 533 529 4 850 830  
## 3 2013 1 1 542 540 2 923 850  
## 4 2013 1 1 544 545 -1 1004 1022  
## 5 2013 1 1 554 600 -6 812 837  
## 6 2013 1 1 554 558 -4 740 728  
## 7 2013 1 1 555 600 -5 913 854  
## 8 2013 1 1 557 600 -3 709 723  
## 9 2013 1 1 557 600 -3 838 846  
## 10 2013 1 1 558 600 -2 753 745  
## # … with 336,766 more rows, and 11 more variables: arr\_delay <dbl>,  
## # carrier <chr>, flight <int>, tailnum <chr>, origin <chr>, dest <chr>,  
## # air\_time <dbl>, distance <dbl>, hour <dbl>, minute <dbl>, time\_hour <dttm>

?flights  
 view(flights)  
   
 flights%>%  
 filter(carrier=="AA",month==3 ,day==17 , year==2013)%>%  
 select(carrier,year,month,day,arr\_delay)%>%  
 arrange(desc(arr\_delay))

## # A tibble: 88 × 5  
## carrier year month day arr\_delay  
## <chr> <int> <int> <int> <dbl>  
## 1 AA 2013 3 17 67  
## 2 AA 2013 3 17 39  
## 3 AA 2013 3 17 39  
## 4 AA 2013 3 17 36  
## 5 AA 2013 3 17 33  
## 6 AA 2013 3 17 22  
## 7 AA 2013 3 17 22  
## 8 AA 2013 3 17 21  
## 9 AA 2013 3 17 19  
## 10 AA 2013 3 17 19  
## # … with 78 more rows

#Question8  
   
 mpg%>%  
 filter(year==1999)%>%  
 group\_by(manufacturer,year==1999) %>%  
 summarise(meanmilage=mean(cty))%>%  
 arrange(desc(meanmilage))

## `summarise()` has grouped output by 'manufacturer'. You can override using the `.groups` argument.

## # A tibble: 15 × 3  
## # Groups: manufacturer [15]  
## manufacturer `year == 1999` meanmilage  
## <chr> <lgl> <dbl>  
## 1 honda TRUE 24.8  
## 2 volkswagen TRUE 21.2  
## 3 subaru TRUE 19   
## 4 hyundai TRUE 18.3  
## 5 toyota TRUE 18.2  
## 6 nissan TRUE 17.7  
## 7 audi TRUE 17.1  
## 8 pontiac TRUE 17   
## 9 chevrolet TRUE 15.1  
## 10 jeep TRUE 14.5  
## 11 ford TRUE 13.9  
## 12 mercury TRUE 13.5  
## 13 dodge TRUE 13.4  
## 14 land rover TRUE 11   
## 15 lincoln TRUE 11

#Honda has got the best average gas milege in 1999