

Reproducing Plots with R

```
install.packages("tidyverse")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)

library(dslabs)
library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

Using the pipe operator, write the code to compute the average, median and standard deviation of city and highway miles per gallon (mpg) by number of cylinders excluding cars with 5 cylinders. Round each statistic to 2 decimal places and sort the resulting table by city average mpg. Your output should be the following table:

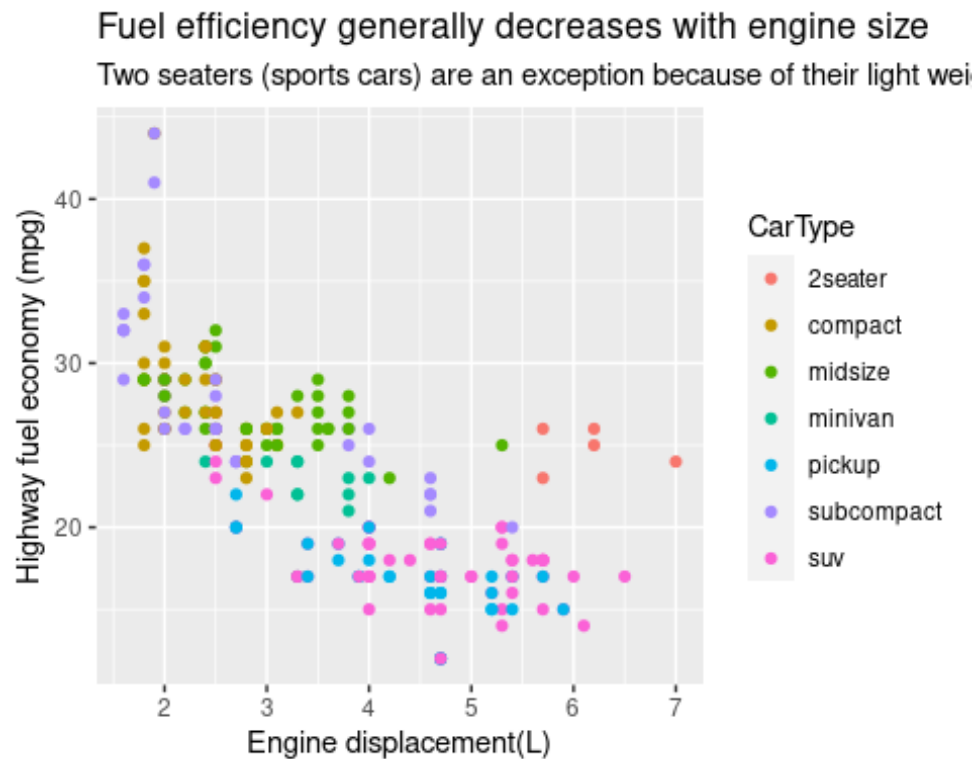
```
mpg %>%
  filter(cyl != 5) %>%
  group_by(cylinder=cyl) %>%
  summarize(city_avg= round(mean(cty),2),
            city_med= round(median(cty),2),
            city_sd = round(sd(cty),2),
            hwy_avg= round(mean(hwy),2),
            hwy_med= round(median(hwy),2),
            hwy_sd= round(sd(hwy),2)) %>%
  arrange(city_avg)

## # A tibble: 3 × 7
##   cylinder city_avg city_med city_sd hwy_avg hwy_med hwy_sd
##   <int>     <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1      8      12.6     13     1.81    17.6     17     3.26
## 2      6      16.2     16     1.77    22.8     24     3.69
## 3      4      21.0     21     3.5     28.8     29     4.52
```

#2) Using the pipe operator, write the code to re-produce the following scatter plot.

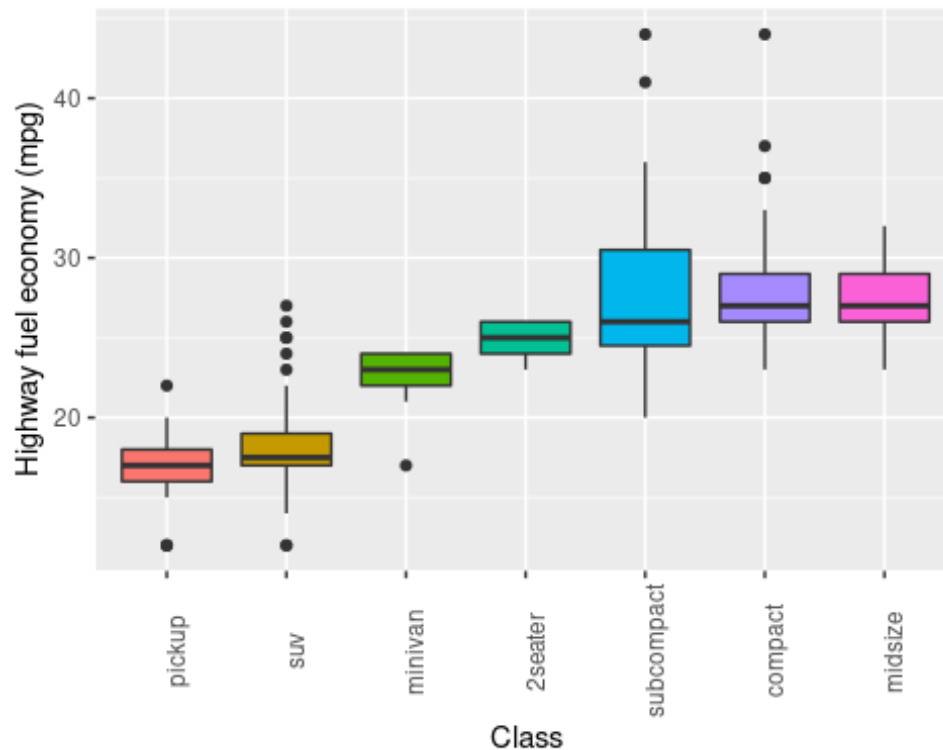
```
mpg %>%
  ggplot(aes(x=displ, y= hwy))+
```

```
geom_point(aes(col=class))+
  labs(title = "Fuel efficiency generally decreases with engine size")+
  labs(subtitle = "Two seaters (sports cars) are an exception because of
their light weight")+
  labs(x="Engine displacement(L)", y="Highway fuel economy (mpg)", col=
"CarType") #col here changes label title
```



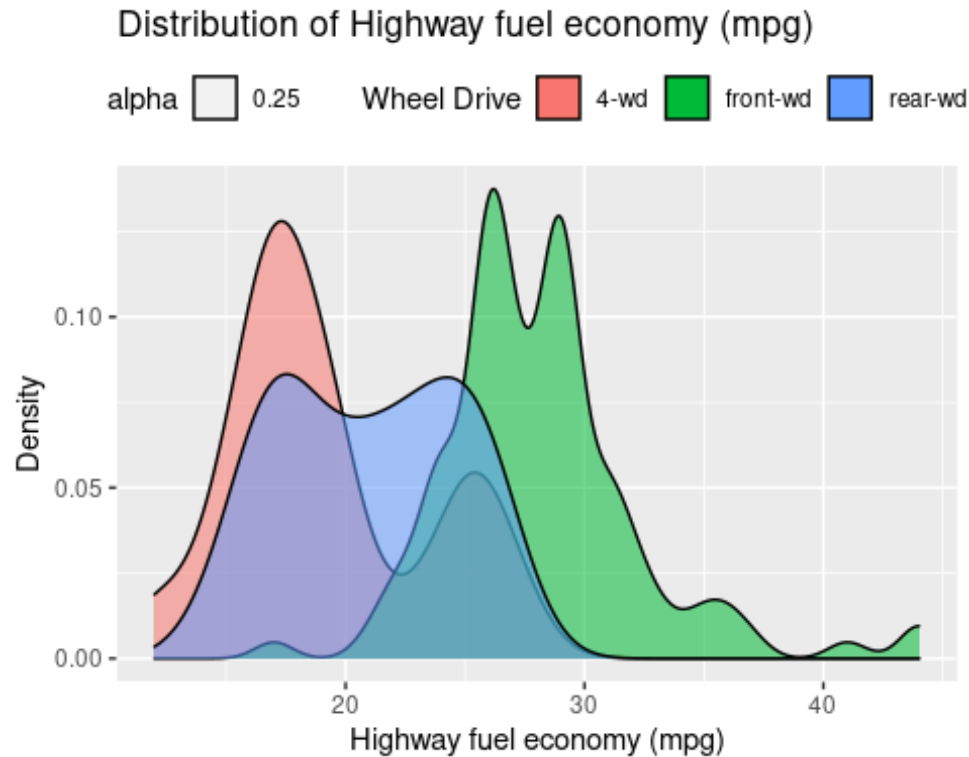
#3) Using the pipe operator, write the code to re-produce the following plot.

```
mpg %>%
  mutate(class=reorder(class, hwy, FUN=median)) %>%
  ggplot(aes(x=class, y=hwy, fill=class))+
  geom_boxplot()+
  labs(x='Class', y='Highway fuel economy (mpg)')+
  theme(legend.position = 'none', axis.text.x = element_text(angle=90))
```



#4) Using the pipe operator, write the code to re-produce the following plot. Hint: use the theme # of the Economist Magazine from the ggthemes package.

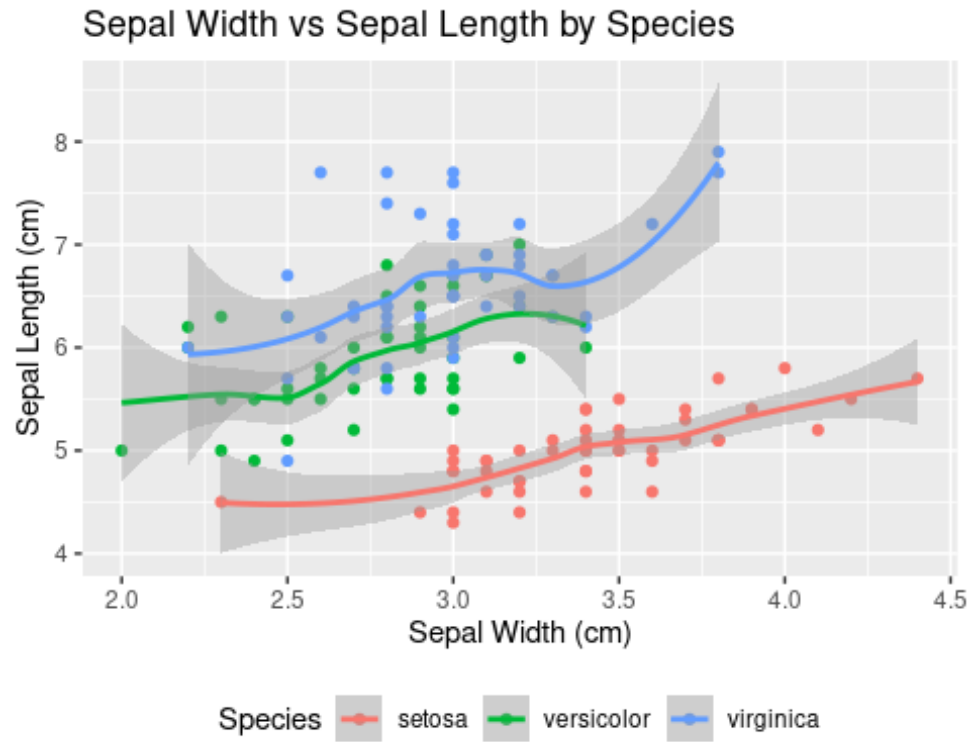
```
mpg %>%
  ggplot(aes(x=hwy, fill=drv))+
  geom_density(aes(group =, fill=, alpha= .25))+
  scale_fill_discrete(name='Wheel Drive', labels = c('4-wd', 'front-wd',
'rear-wd')) +
  labs(x='Highway fuel economy (mpg)', y='Density',
title='Distribution of Highway fuel economy (mpg)')+
  theme(legend.position = "top")
```



#1) Using the pipe operator, write the R code to re-produce the following scatter plot. What does the shaded region around each curve represent?

```
iris %>%
  ggplot(aes(x=Sepal.Width, y=Sepal.Length, col= Species))+
  geom_point()+
  geom_smooth()+
  labs(title="Sepal Width vs Sepal Length by Species", x= "Sepal Width (cm)",
y= "Sepal Length (cm)")+
  theme(legend.position = "bottom")

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



#The shaded region represents the confidence interval.