Week 3 Workbook

Odelia

3/10/2021

Question 1: Why is this graph not printing any output?

This is because "geom_point" is not in the code. It is important for "geom_point" to be in the code as it tells R to create the scatterplot. Geom defines the layout of the ggplot layer while point defines the layout to be a scatterplot. If "geom_point" is missing from the code, R will not know that I want my data to be displayed as a plot and specifically, a scatterplot. For the plot to display successfully, after loading tidyverse, the code should be as such: ggplot(data = mtcars) + geom_point(mapping = aes(x = mpg, y = wt, colour=factor(cyl)))

Question 2. Using the mpg dataset, graph the relationship between city mileage and highway mileage by year manufacture.

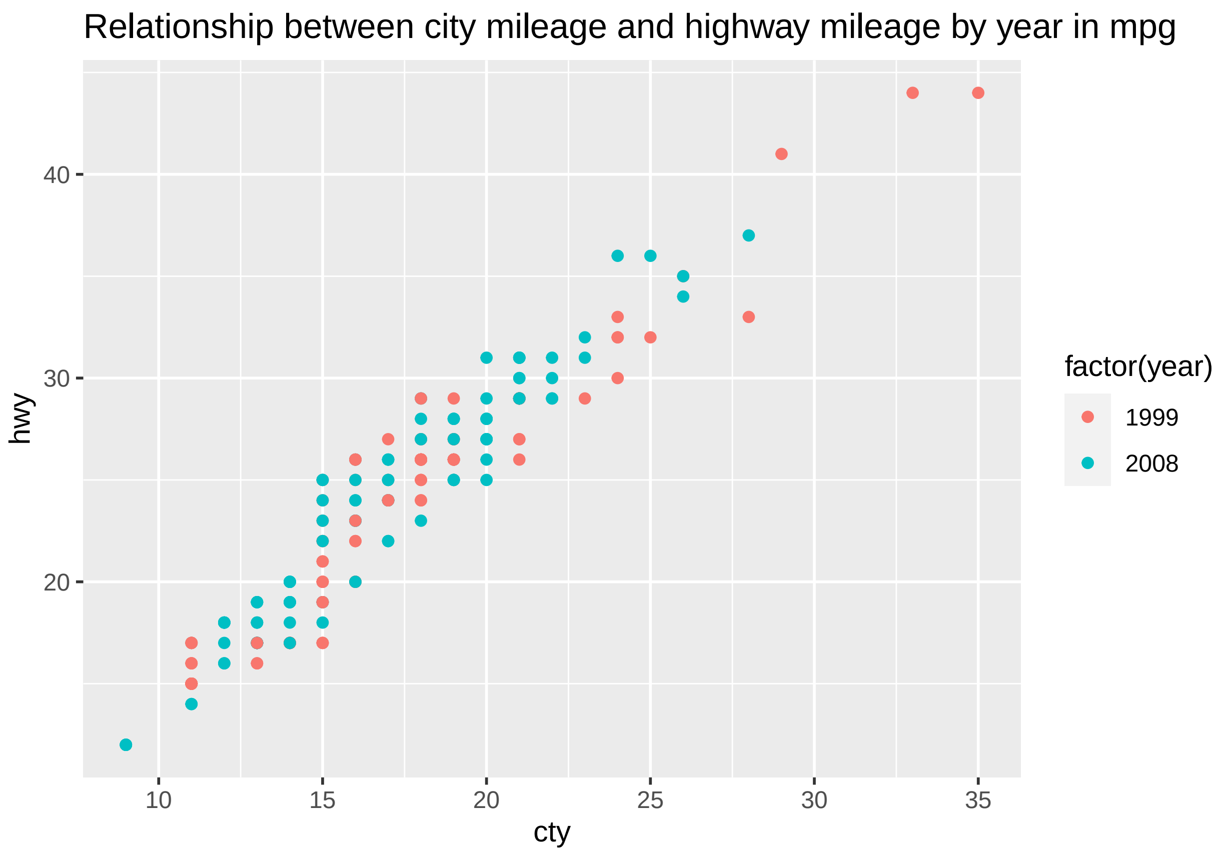
```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.0     v dplyr   1.0.4
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

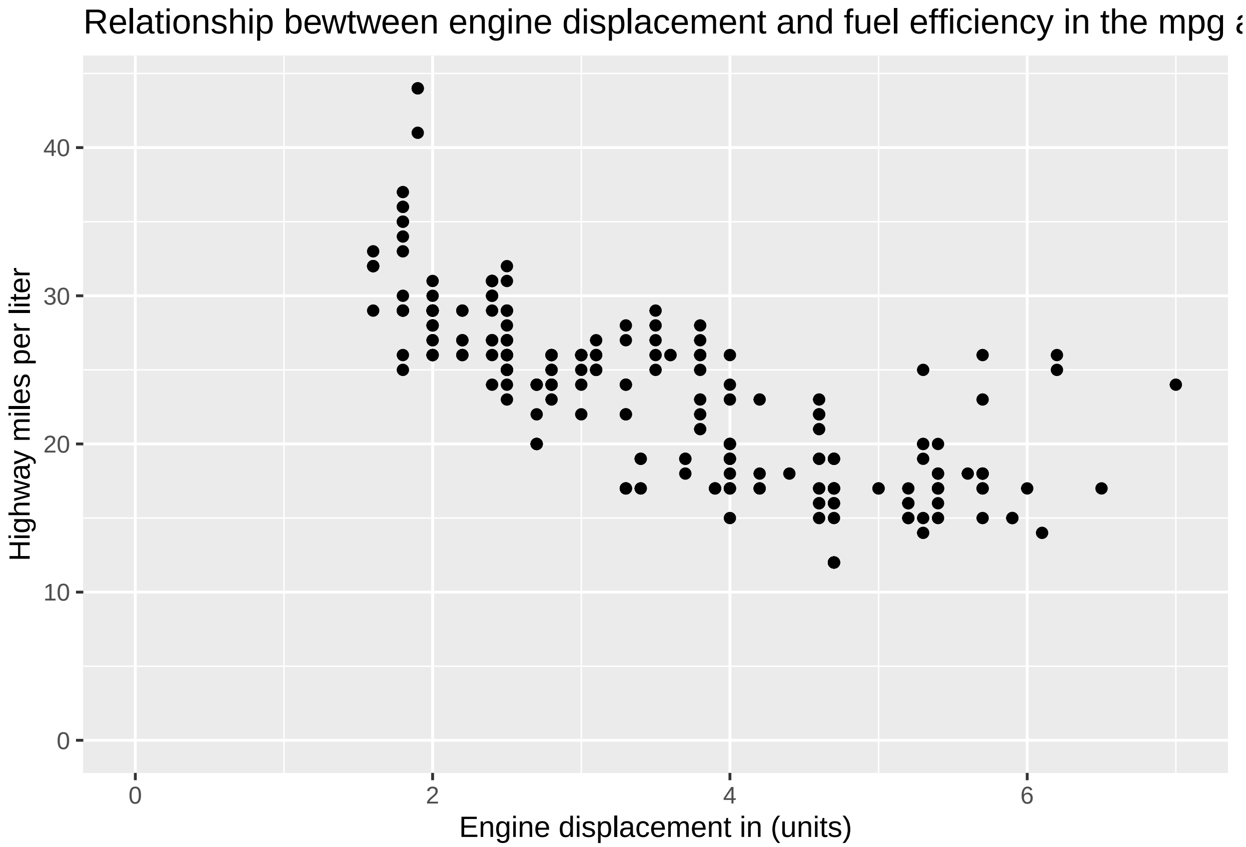
```r
ggplot(data = mpg) + geom_point(mapping = aes(x = cty, y = hwy, colour=factor(year))) + labs(title =  "
```



Question 3. Edit this graph so that the x axis and the y axis both start at 0

This can be done by adding "expand_limits(x = 0, y = 0)" to the code.

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  labs(title = "Relationship bewtween engine displacement and fuel efficiency in the mpg automobile dat
  xlab("Engine displacement in (units)") +
  ylab("Highway miles per liter") + expand_limits(x = 0, y = 0)
```



Question 4: what is one benefit and one limitation for this graph above (in which the x and y values start at 0?)

Limitation: Real world applications for internal combustion engines will always have a value greater than 0. Likewise for data that can never logically be 0, having both the x and y axes at 0 would unnecessarily make the graph's data points more compressed together and therefore harder to read than a truncated graph.

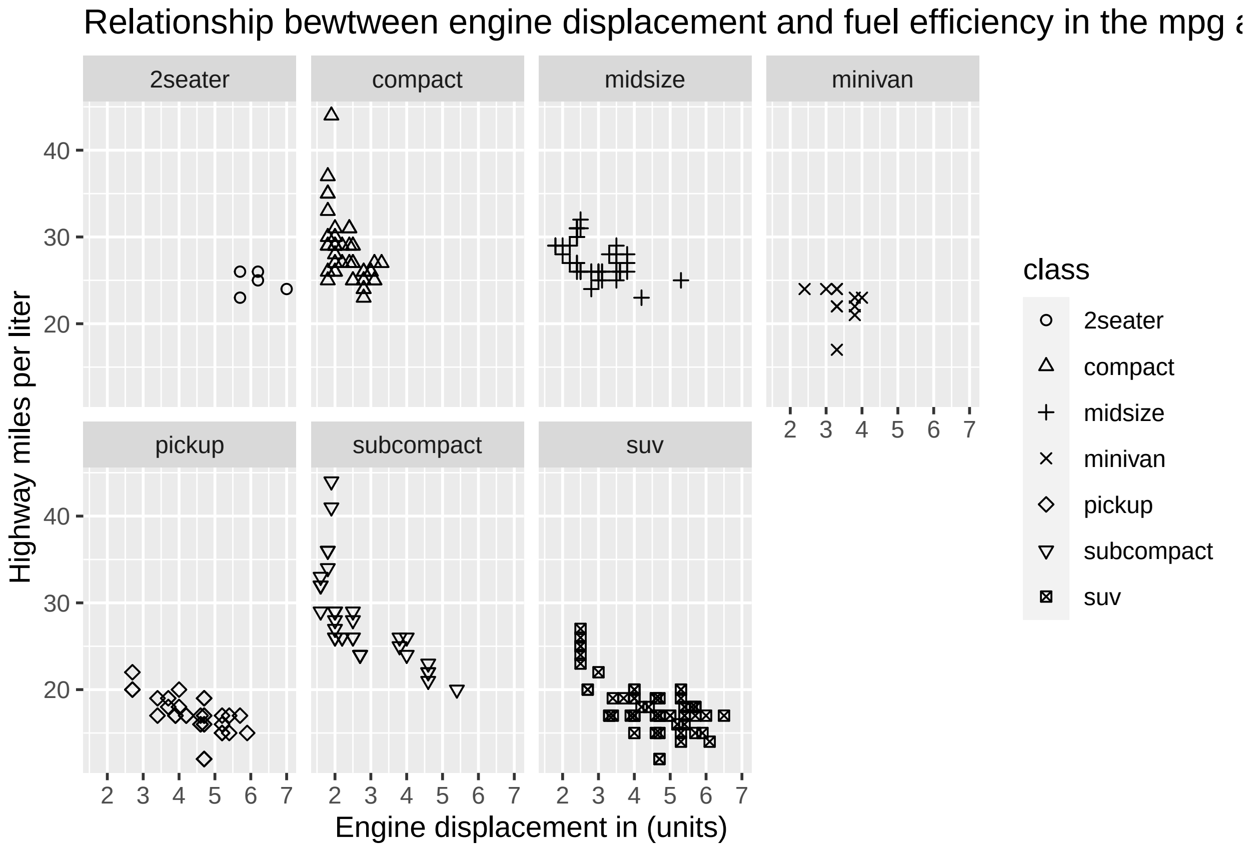
Benefit: On the other hand, truncated diagrams will distort the underlying numbers visually. A diagram that starts at 0 minimises the likelihood of people visually overestimating differences in data points.

Question 5. Which of these two graphs do you prefer and why?

I prefer the graph with the data points categorised by colours. Categorising data points by colours makes it easier and more comfortable for the brain to differentiate data points than if they were categorised by shapes. Especially in a dataset with a large number of data points or data points with underlying numbers that are really close or stack on top of one another, having the data points differentiated by colours would be easier to discern one data point from another, and also notice any trends in the data. It is also easier to associate the categories with their relative colours as colours attract the human eye and stimulate the brain more so than shapes. It is also empirically evident in research studies that colours have the tendency to capture better attention level, and thus, better memory than shapes.

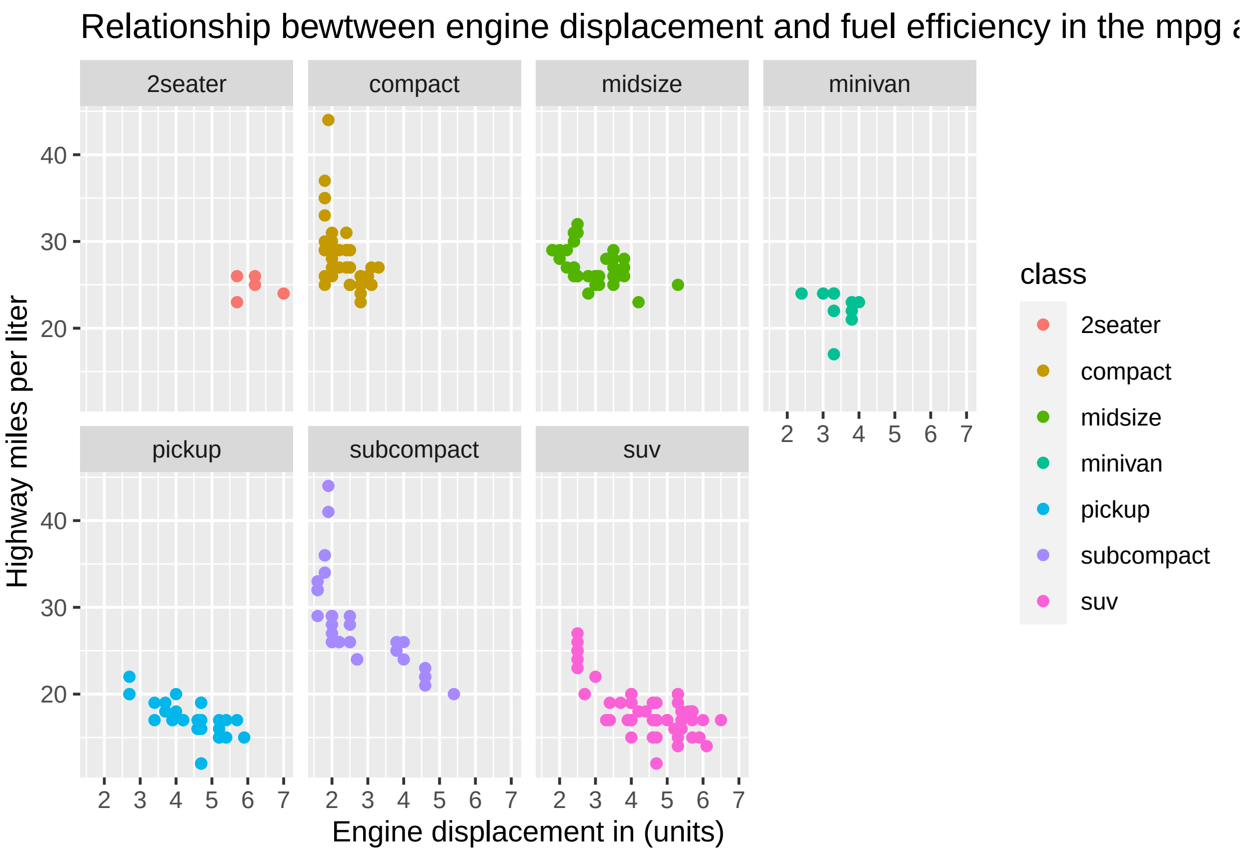
Question 6. add a facet to this graph for the "class" variable

```
g2 <-ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, shape =  class )) + scale_shape_manual(values = c(1:7)) +
  labs(title = "Relationship bewtween engine displacement and fuel efficiency in the mpg automobile data
  xlab("Engine displacement in (units)") +
  ylab("Highway miles per liter")
g2 + facet_wrap(~ class, nrow = 2)
```



nb: There are no data points seen under “suv” as there is no 7th shape for it. I have taken the liberty to work the code again with colours instead. Please don’t deduct my marks for this! :’)

```
g1 <-ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, colour =  class )) +
  labs(title = "Relationship bewtween engine displacement and fuel efficiency in the mpg automobile dat
  xlab("Engine displacement in (units)") +
  ylab("Highway miles per liter")
g1 + facet_wrap(~ class, nrow = 2)
```



Question 7. which graph is more informative and why?