# Homework-1

All of these homework problems are from the *R for Data Science* book. The section numbers (e.g., "3.2.4 Exercises") refer to sections in this book. Although the questions are based on those in the book, some questions ask for additional details or analysis.

When solving these problems, you are allowed to use any method from the book or class, even if that method wasn't yet covered when the exercise was presented in the book.

Write answers that are as complete as possible. If a graph is helpful for formalizing the solution, provide the graph. If a table is helpful, provide a table. In the text part of the answer, outline the progression in your thinking as you perform the analysis.

Note that you should type your answers in RStudio, by typing into the file `Homework-1.sa.Rmd`. We will talk about R Markdown in class, and it is also briefly covered in the book, Chapter 27 (R Markdown).

As discussed in the syllabus, if you look up answers online, or get answers from your fellow students, or from Generative AI, these sources need to be cited.

Out of 115 points.

---

### 3.2.4 Exercises

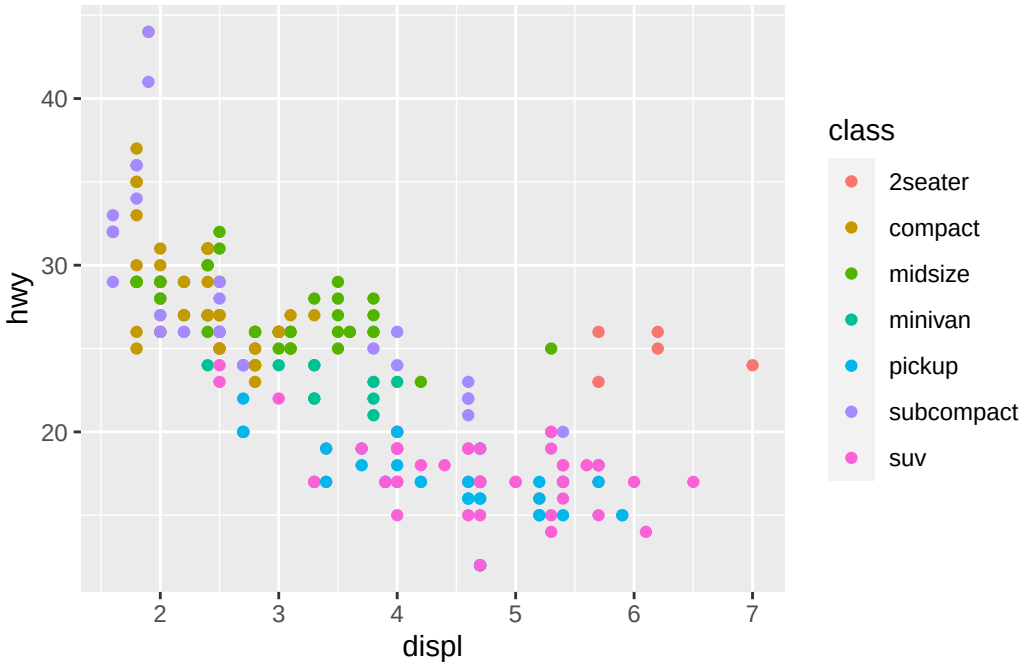
**(1) 3.2.4 Exercise 4 (10 pts)** For the `mpg` dataset, make a scatterplot of `hwy` vs `cyl`. What does this plot tell you about these two variables?

**(2) 3.2.4 Exercise 5 (10 pts)** What happens if you make a scatterplot of `class` vs. `drv`. Is this plot useful? Why or why not?

### 3.3.1 Exercises

**(3) 3.3.1 Exercise 2 (10 pts)** Which variables in `mpg` are categorical? Which are continuous? (Hint: type `?mpg` to read the documentation for the dataset). How can you see this information when you run `mpg`?

**(4) 3.3.1 Exercise 3 (10 pts)** Consider the scatterplot of displacement vs miles per gallon that we have been studying:




Map a continuous variable to `color`, `size`, and `shape`, and show the resulting graphs. How do these aesthetics behave differently for categorical vs. continuous variables?

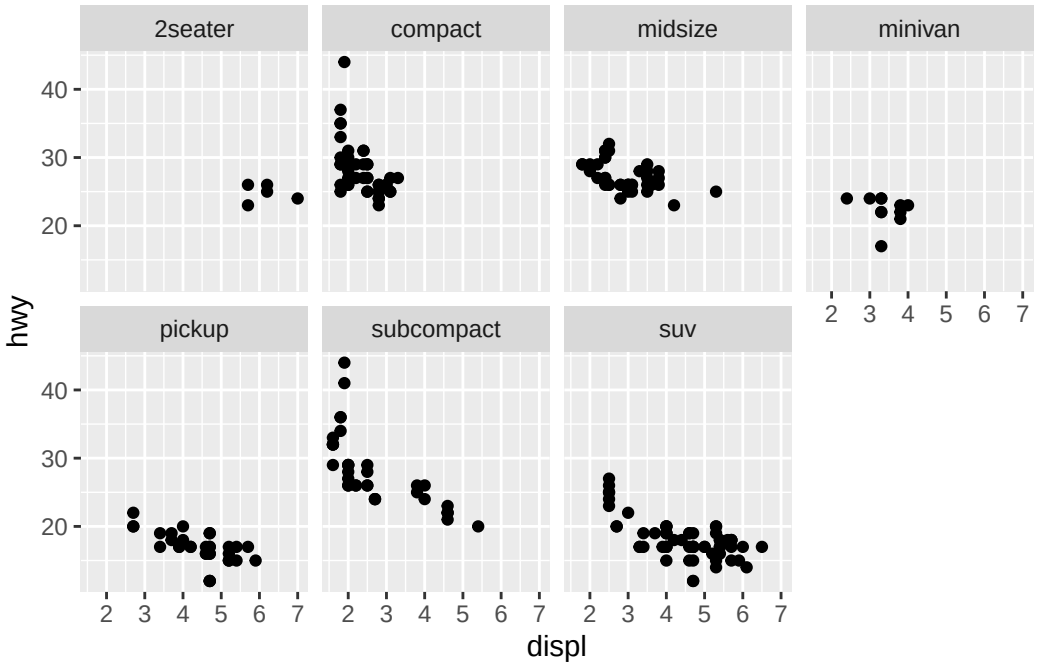
**(5) 3.3.1 Exercise 6 (10 pts)** What happens if you map an aesthetic to something other than a variable name, like `aes(color = displ < 4)`? Note that you will also need to specify $x$ and $y$.

### 3.5.1 Exercises

**(6) 3.5.1 Exercise 1 (5 pts)** What happens if you facet on a continuous variable? Show an example. Discuss when this might be useful, and when it is likely not useful.

**(7) 3.5.1 Exercise 4 (10 pts)** Consider the first faceted plot in this section:

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  facet_wrap(facets = vars(class), nrow = 2)
```

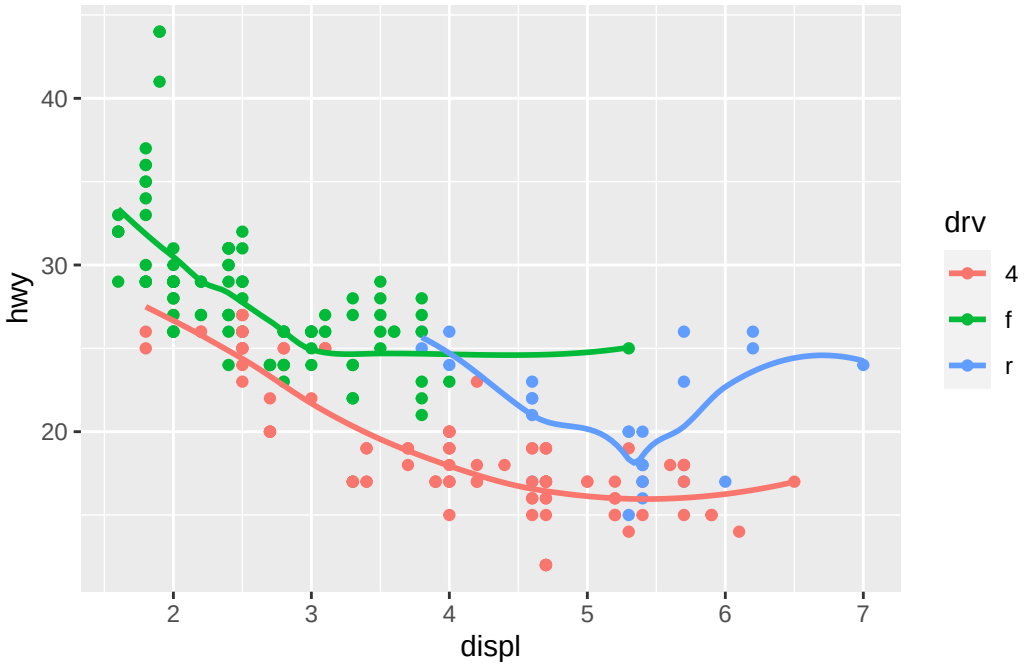



Create a comparable graph that uses the color aesthetic. What are the advantages to using faceting instead of the color aesthetic? What are the disadvantages? How might the balance change if you had a larger dataset?

### 3.6.1 Exercises

**(8) 3.6.1 Exercises 2 (10 pts)** Run this code in your head and predict what the output will look like. Then, run the code in R and check your predictions. Discuss your findings. Were you surprised?

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



**(9) 3.6.1 Exercises 4 (5 pts)** What does the `se` argument to `geom_smooth()` do?

**(10) 3.6.1 Exercises 5 (5 pts)** Will these two graphs look different? Why or why not?

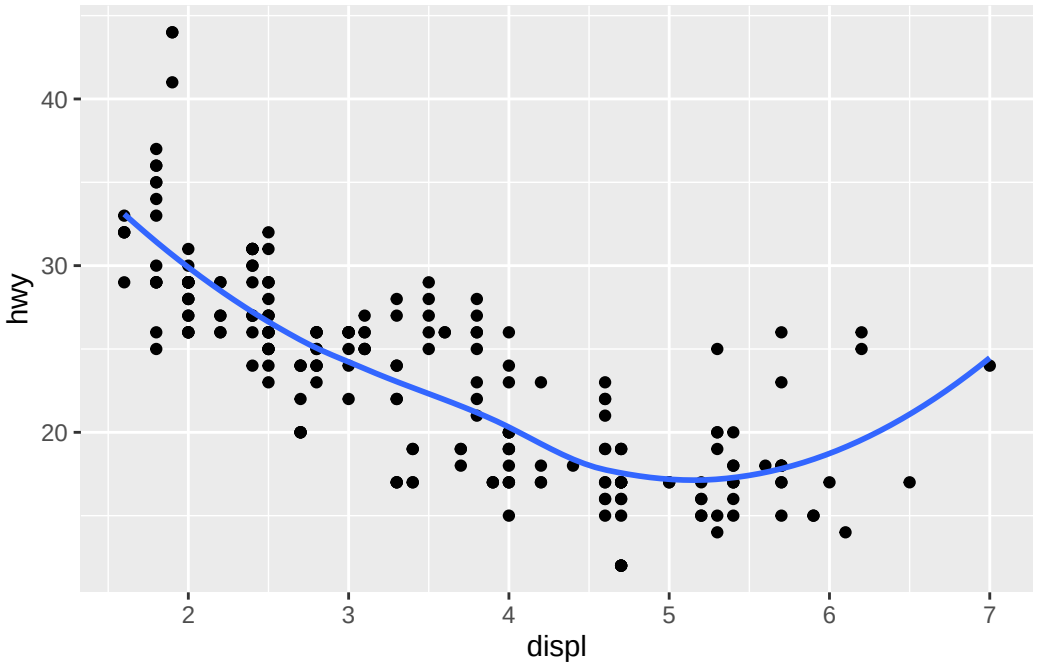
```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +
  geom_point() +
  geom_smooth()

ggplot() +
  geom_point(data = mpg, mapping = aes(x = displ, y = hwy)) +
  geom_smooth(data = mpg, mapping = aes(x = displ, y = hwy))
```

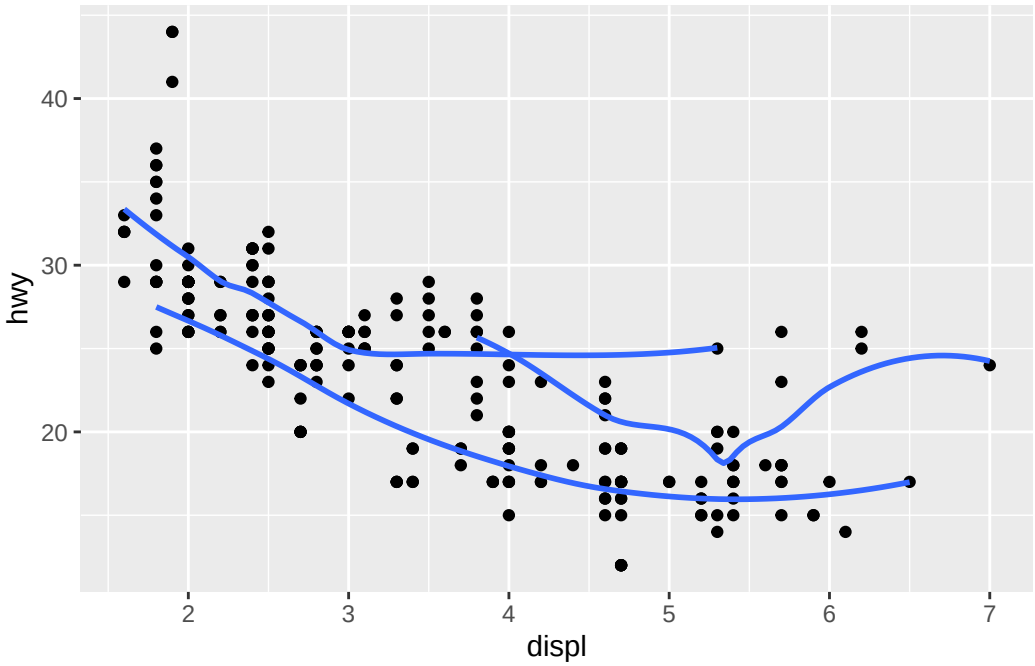
**3.6.1 Exercises 6** Recreate the R code needed to generate each of these graphs:

**(11) (6.1, 5 pts)** .

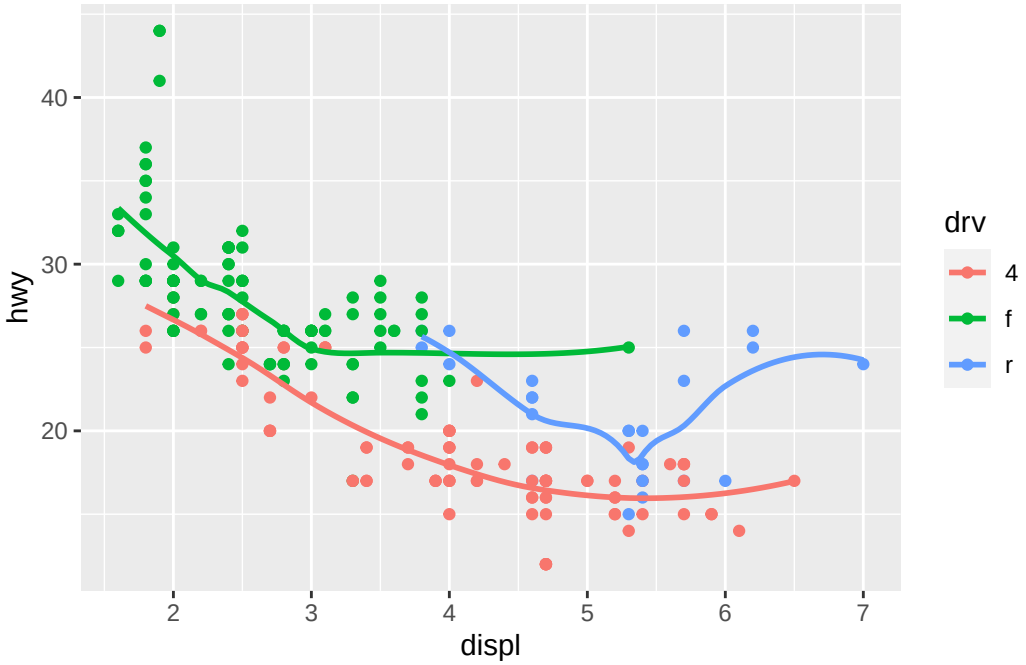
```
# Show or hide all graphs below
knitr::opts_chunk$set(echo=FALSE)
```



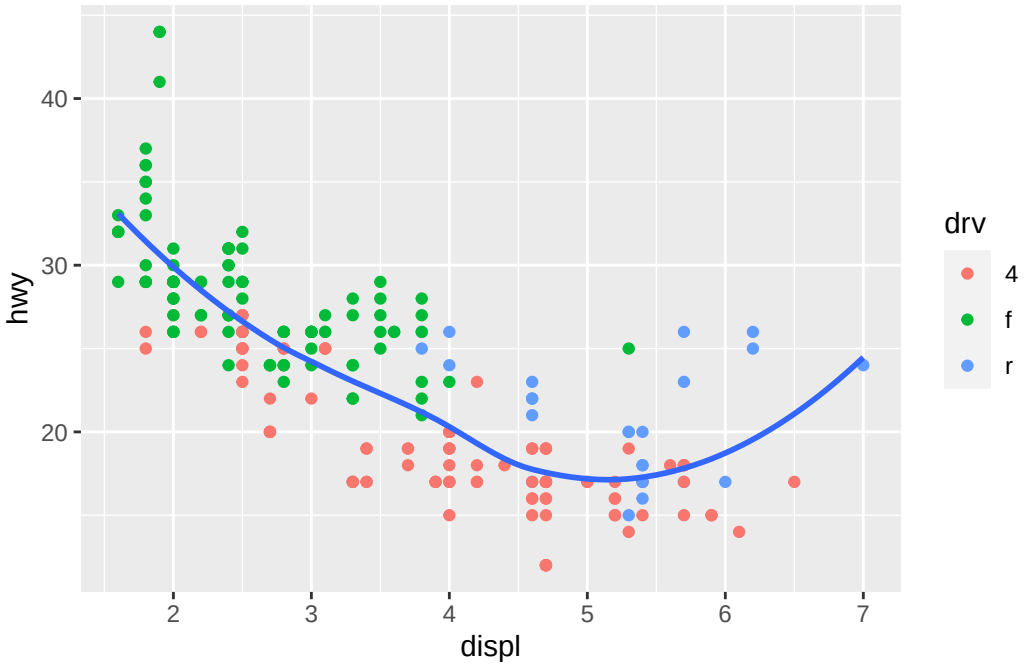

**(12) (6.2, 5 pts)** .



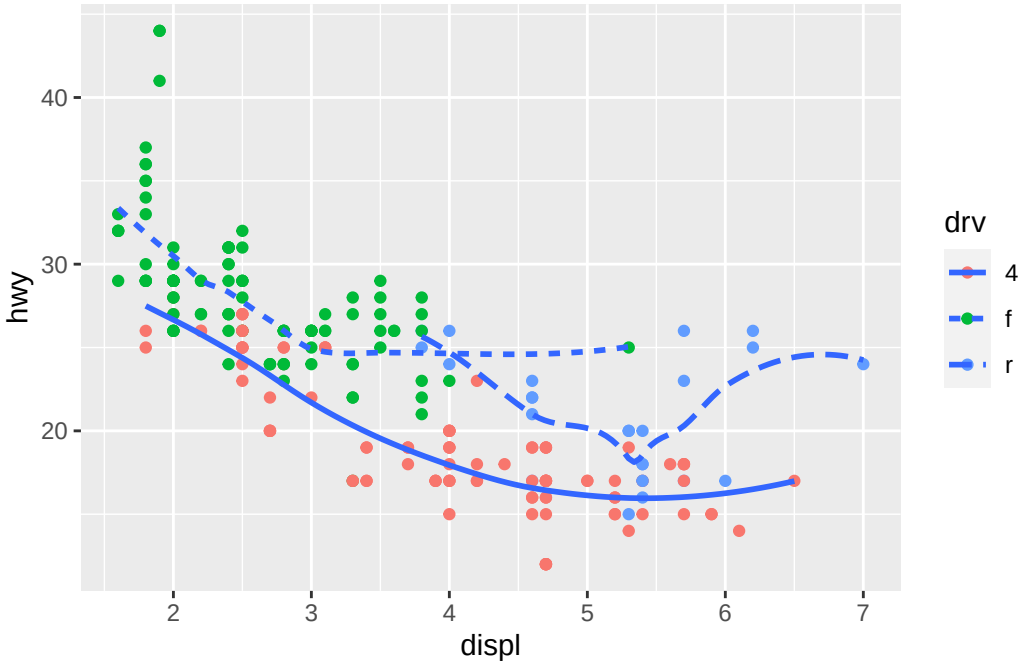

**(13) (6.3, 5 pts)** .



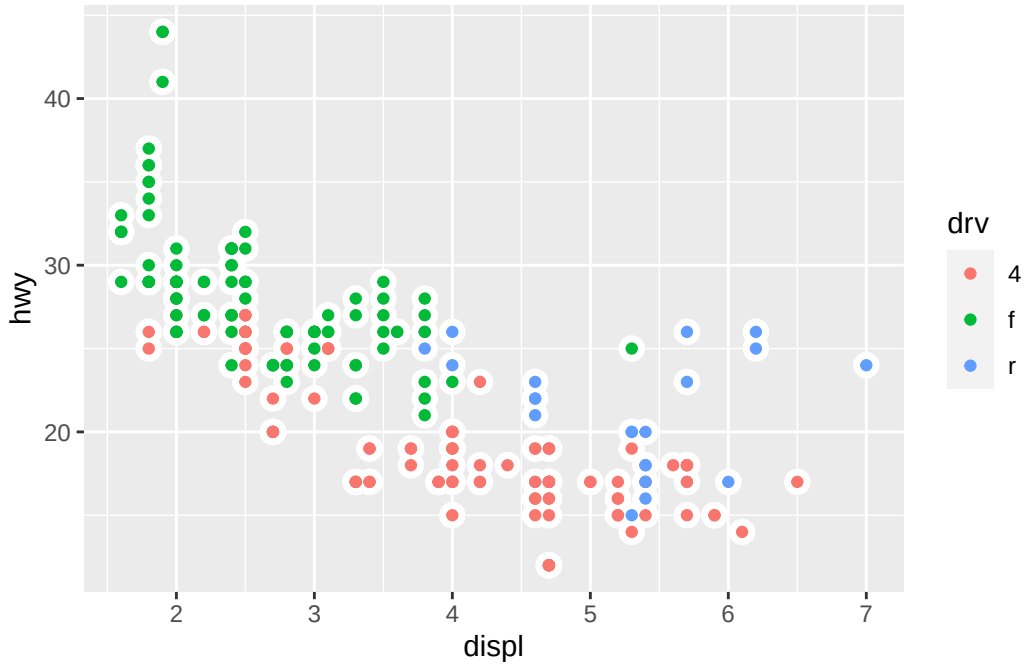

**(14) (6.4, 5 pts)** .




**(15) (6.5, 5 pts)** .




**(16) (6.6, 5 pts)** .


hwy
40
30
20
2
3
4
5
6
7
displ
drv
4
f
r