Practice Set 1

Dhyey Mavani

Due by 10pm ET on Friday

Practice Set Information

During the week, you will get further practice with the material by working through the Practice Set, a set of problems designed to give you practice beyond the examples produced in the text.

You may work through these problems with peers, but all work must be completed by you (see the Honor Code in the syllabus) and you must indicate who you worked with below.

Even then, the best approach here is to try the problems on your own before discussing them with peers, and then write your final solutions yourself.

GitHub Workflow

1. Before editing this file, verify you are working on the copy saved in *your* repo for the course (check the filepath and the project name in the top right corner).
2. Before editing this file, make an initial commit of the file to your repo to add your copy of the problem set.
3. Change your name at the top of the file and get started!
4. You should *save, knit, and commit* the .Rmd file each time you've finished a question, if not more often. You should also *push* your commits back onto GitHub occasionally (you can do this after each commit).
5. When you think you are done with the assignment, save the pdf as "*Name_thisfilename_date*.pdf" before committing and pushing (this is generally good practice but also helps me in those times where I need to download all student homework files).

Gradescope Upload

For each question (e.g., 3.1), allocate all pages associated with the specific question. If your work for a question runs onto a page that you did not select, you may not get credit for the work. If you do not allocate *any* pages when you upload your pdf, you may get a zero for the assignment.

You can resubmit your work as many times as you want before the deadline, so you should not wait until the last minute to submit some version of your work. Unexpected delays/crises that occur on the day the assignment is due do not warrant extensions (please submit whatever you have done to receive partial credit).

Practicing Academic Integrity

If you worked with others or used resources outside of provided course material (notes, textbook, etc) to complete this assignment, please acknowledge them below using a bulleted list.

I acknowledge the following individuals with whom I worked on this assignment:

Name(s) and corresponding problem(s) N/A *

I used the following sources to help complete this assignment:

Source(s) and corresponding problem(s) N/A *

Problem 1 **MDSR Exercise 2.5 (modified)** Consider the data graphic for Career Paths at Williams College. Focus on the graphic under the "Major-Career" tab.

1.1 What story does the data graphic tell? What is the main message that you take away from it?

The data graphic indicates the major-wise career trajectories of the alums of Williams College. This graphic mainly helps in visualizing major-wise possible career trajectories. We can say that almost every career path is possible from any of the majors because we can see a combination of almost all vivid shades (signifying different majors) in each of the different grey career arcs (possible careers). This creates a mesh network as we can observe in the Compilation graphic. This tells us that there is open curriculum and alums from Williams are having overall development of skills. So, we can say that it's a liberal arts school.

1.2 Can the data graphic be described in terms of the taxonomy presented in this chapter? If so, list the visual cues, coordinate system, and scale(s). If not, describe the feature of this data graphic that lies outside of that taxonomy.

The scale and coordinate system is very unclear in the graphic. Still, we can say that the length of each of te sectional arcs of majors and careers represent the relative number of alums associated with that field. Moreover, there are markings of different majors and careers so they count in the legend and labels.

But, vivid colors are used in the graphic to represent different majors, so that is a visual cue. The usage of thickness of lines linking the careers and majors together is a new visual cue.

1.3 Critique and/or praise the visualization choices made by the designer. Do they work? Are they misleading? Thought-provoking? Brilliant? Are there things that you would have done differently? Justify your response.

Critique:

1. The scale and coordinate system is very unclear in the graphic. Hence, it may be difficult to visualize correct scales and measures in terms of quantity.
2. The labels indicating majors and careers are too small to read.
3. The graph is more leaning towards the qualitative side, and hence does not give that of an exact idea of the relationship to an average reader.

Praise:

1. The usage of different shades of colors in combination is perfect and sticks out to eyes.
2. I really liked the idea of making the middle part of connecting bands thin, so that readers could easily walk their way through the majors to careers section without getting entangled in colors.
3. The spacing between each of the 30 arcs serves a brilliant purpose of keeping the mixing at minimum and ensuring that the vision in clear even in the case of compilation figure.

Things that I would have done differently:

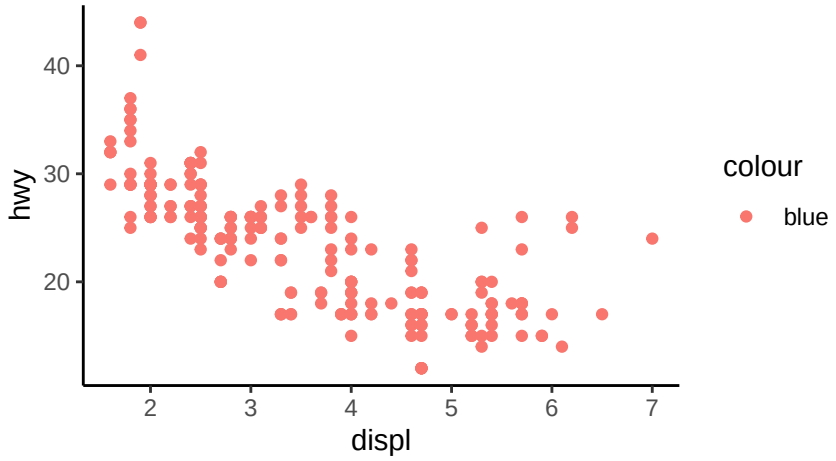
1. Introduce some type of scale in terms of length of arc or angle preferably.
2. Increase the font size of labels signifying different major and career names, and locate them to the respective colored arcs by the arrows of same color in order to accommodate the large sixe of text while maintaining the clarity of label correspondence.
3. Add some quantitative data to the graph either by incorporating into graphic or by building legends.

Problem 2 **Spot the Error** Explain why the following command does not color the data points blue, then write down (in a new code chunk) the command that will turn the points blue. Use the help file for the dataset to additionally update the graphic with informative axis labels and a title.

Assigning the color within `aes()` provides a level based mapping of color to the x and y coordinates.

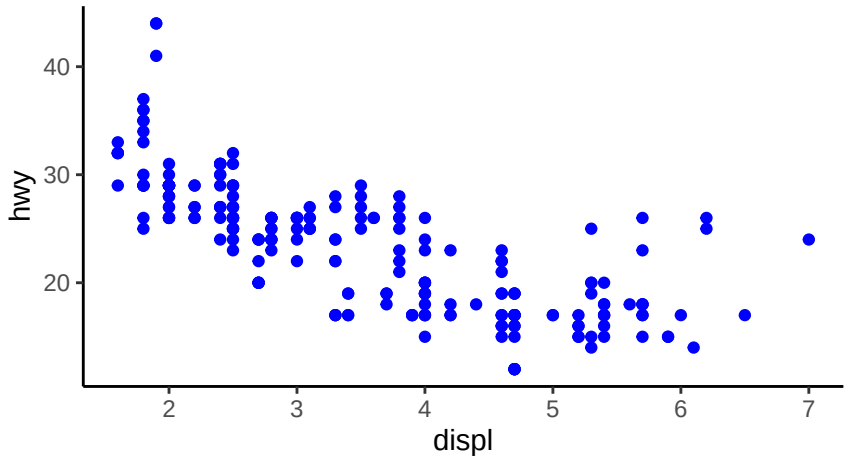
To make all the points blue, we should set the color outside `aes()`, but inside the `geom_point()`. This does not require mapping as it just passes the property `"blue"` to all the data points.

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



The corrected code is as follows...

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

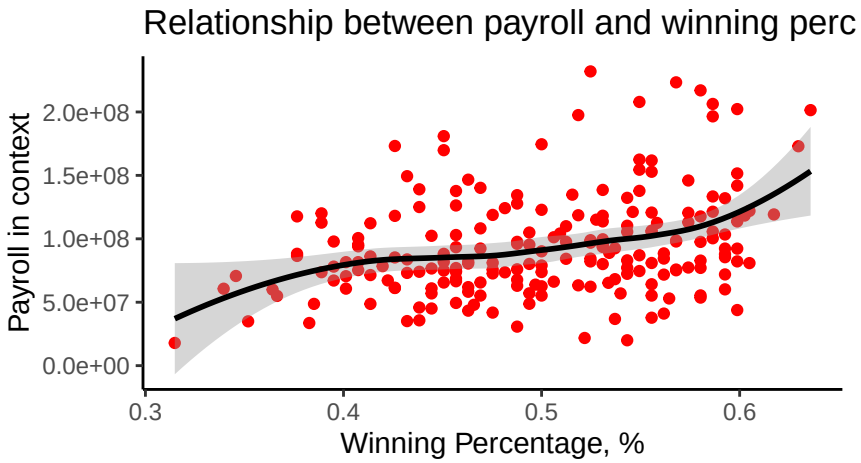


Problem 3 **MDSR Exercise 3.6 (modified)** Use the `MLB_teams` data in the **mdsr** package to create an informative data graphic that illustrates the relationship between winning percentage and payroll in context. What story does your graph tell?

The graph tells us that the curve mapping payroll and winning percentages together is approximately of a Sigmoid type. Moreover, the big picture is that the payroll increase leads to an increase in winning percentage. But, due to the Sigmoid nature of the curve, the rates of increase in the winning percentage may vary upon the payroll in context amount. As we can see as the payroll gets to the lower or higher side, the winning percentage shows more change compared to the case when the payroll in context is in the middle range.

```
library(mdsr)
MLBPlot <- ggplot(data = MLB_teams,aes(x = WPct,y = payroll)) +
  geom_point(color = "red") +
  labs(x = "Winning Percentage, %", y = "Payroll in context",
       title = "Relationship between payroll and winning percentages") +
  geom_smooth(color = "black")

MLBPlot
```



Problem 4 **MDSR Exercise 3.10 (modified)** Using data from the **nasaweather** package, use the `geom_path()` function to plot the path of each tropical storm in the `storms` data table (use variables `lat` (y-axis!) and `long` (x-axis!)). Use color to distinguish the storms from one another, and use facetting to plot each `year` in its own panel. Remove the legend of storm names/colors by adding `scale_color_discrete(guide = "none")`.

```
library(mdsr)
library(nasaweather)

storms <- storms

Tropical_storms <- storms %>%
  filter(type == "Tropical Storm")

ggplot(Tropical_storms, aes(x = lat, y = long)) +
  geom_path(show.legend = FALSE, aes(color = name)) +
  facet_wrap(~ year, ncol = 2) +
  labs(y = "Longitude", x = "Latitude", title = "Path of
       Tropical Storms") +
  scale_color_discrete(guide = "none")
```