Grammar of Graphics

Each of the following questions should be answered using the `storms` data set built into `dplyr`. Provide the code to construct each of the following plots and make a sketch of the resulting plot. No need to enumerate all of the detail of the data in the plot but focus on the general shape and be sure to be clear what all axes and legends are (i.e. what the aesthetic mappings are).

1. Plot the distribution of the wind speed three ways: as a histogram, a box plot, and a density curve.

2. A bar chart showing the distribution of storm status found across all records in the data set.

3. A bar chart showing the distribution of storm status of when considering just one record for each named storm[1].

[1] There are several ways to reduce the data frame to contain only one row for each named storm. The simplest is probably to group the data frame, then slice it and take the first record

4. A scatterplot of the relationship between wind speed and pressure for every record in the data set.

5. A scatterplot of the relationship between average wind speed and average pressure for each storm.

6. A line plot of the relationship between wind speed and pressure for all of the storms in 2014. Each storm should be indicated with a different color[2].

7. Create a line plot that shows the average duration of storms over time. That is: the x-axis should be the year and the y-axis should be average storm duration.

[2]If you'd like to a slightly more sophisticated version of this plot, read about `geom_path()`. Instead of lines, this can plot paths with an arrowhead indicating the progression of that storm.

EDA

Your task is to understand the basic structure of a new data set from 2020 related to tuition costs at colleges and universities across the country. It is stored as a csv on github that can be read into your R session (or a Quarto document) using:

```
tuition_cost <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/
tidytuesday/main/data/2020/2020-03-10/tuition_cost.csv')
```

For each of the following, please answer the question in full sentences and also provide the code that you used. If you generated a plot, also sketch the plot.

8. What are the dimensions of the data set? What is the unit of observation? What is the nature of the variables (that they are, their units or levels, and their type)?

9. Is there any missing data? If so, where?

10. What is the distribution of the cost of in-state tuition? Answer with summary statistics, a plot, and a written description of its shape.

11. How does the distribution of in-state tuition differ between the three types of institutions? Answer with a plot and a written description of how the distributions compare.

12. Which are the three most expensive institutions? The three least expensive institutions?

13. Create one more visualization of at least two variables that you're curious about. Describe the structure that you see with a written description.