STAT 140: Homework 2

R Solutions

EXPLORING ggplot2

For the following exercises, you will be exploring a dataset collected on the game Pokemon Go. As noted in the dataset description on OpenIntro, "A key part of Pokémon Go is using evolutions to get stronger Pokemon, and a deeper understanding of evolutions is key to being the greatest Pokemon Go player of all time. This data set covers 75 Pokemon evolutions spread across four species. A wide set of variables are provided, allowing a deeper dive into what characteristics are important in predicting a Pokemon's final combat power (CP)."

```
library(ggplot2)
## source pokemon go data from OpenIntro website
source("http://www.openintro.org/stat/data/pokemon.R")
```

The book R for Data Science can be a great resource for you as you learn more about what we can do in R. While a lot of it is advanced, it does have great examples for data visualizations. Open the PDF of the book in your browser. For this exercise, you will use the table of contents: Click on **3. Data Visualization** and then on **3.8 Position Adjustments**. Using the pokemon data, and specifically the variables `attack_weak_type` and `attack_weak_type_new`, generate the same bar plots that were generated for the diamonds data set in the book example. You should make 7 plots in total (I have already included the R chunks for your convenience). Pay attention to the subtle differences in the code that let you change how these plots look.

- For these first two plots, choose either of the two variables to depict in your bar chart. What is the difference between these two plots? What piece of the code accomplished this change?

  Answer: In the first plot we specify colour=attack_weak_type; this gives us the color outline around the bars. In the second we specify fill=attack_weak_type; this gives us the filled color in the bars.

```
bc <- (ggplot(data = pokemon) + geom_bar(mapping = aes(x = attack_weak_type,
    colour = attack_weak_type)) + theme_bw() + xlab("Weak Attack Type"))
bc
```

```
bc <- (ggplot(data = pokemon) + geom_bar(mapping = aes(x = attack_weak_type,
    fill = attack_weak_type)) + theme_bw() + xlab("Weak Attack Type"))
bc
```

- For the rest of the plots, you will be using both variables. Identify the plots that we have studied in class. What are the differences among these five plots? What pieces of code accomplished these changes?

  Answer: Within aes() within geom_bar: color vs fill; alpha

```
bc <- (ggplot(data = pokemon) + geom_bar(mapping = aes(x = attack_weak_type,
    fill = attack_weak_type_new)) + theme_bw() + xlab("Weak Attack Type") +
    ylab("Count"))
bc
```

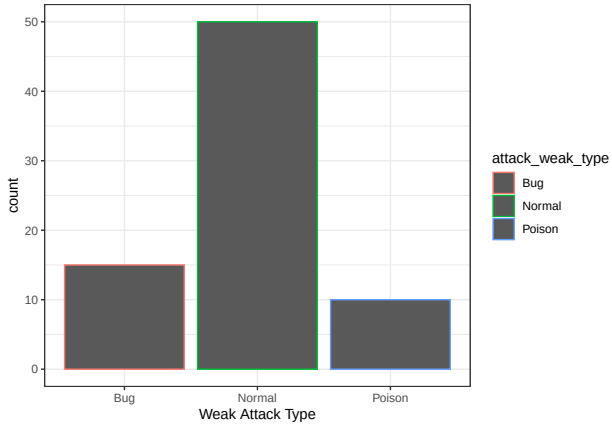

Figure 1: R Exercise Plot 1

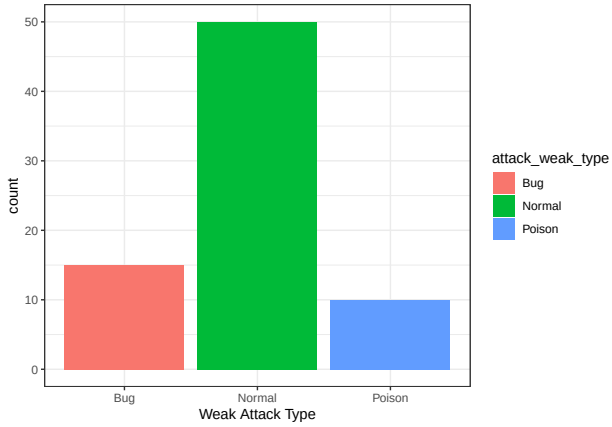

Figure 2: R Exercise Plot 2

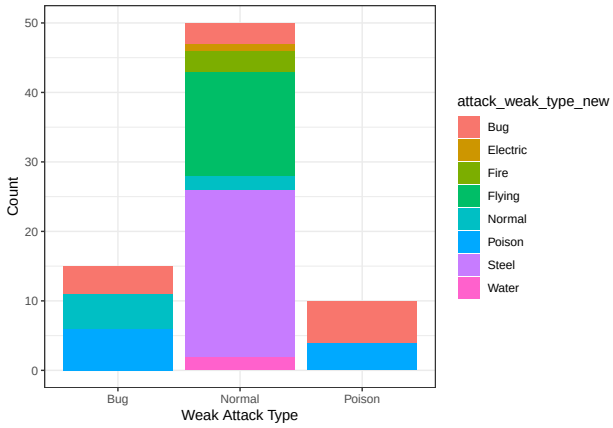

Figure 3: R Exercise Plot 3

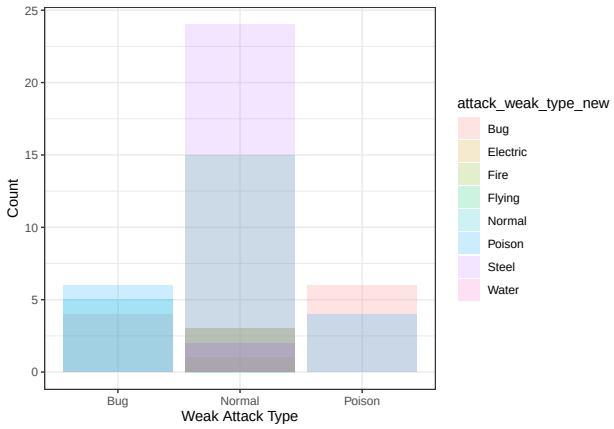

Figure 4: R Exercise Plot 4

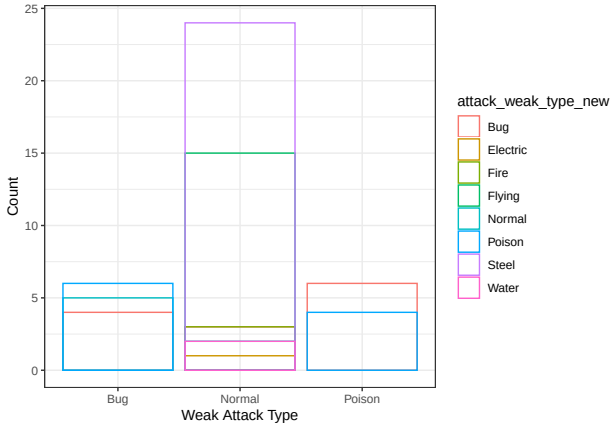

Figure 5: R Exercise Plot 5

```
bc <- (ggplot(data = pokemon) + geom_bar(mapping = aes(x = attack_weak_type,
    fill = attack_weak_type_new), alpha = 1/5, position = "identity") + theme_bw() +
    xlab("Weak Attack Type") + ylab("Count"))
bc
```

```
bc <- (ggplot(data = pokemon) + geom_bar(mapping = aes(x = attack_weak_type,
    color = attack_weak_type_new), fill = NA, position = "identity") + theme_bw() +
    xlab("Weak Attack Type") + ylab("Count"))
bc
```

- Based on the last three plots you generated, what is the default argument for position?

  Answer: Default is 'identity' - wasn't specified in Plot 3, but the position was the same for Plot 4 and Plot 5 where we specified Position='identity'.

```
bc <- (ggplot(data = pokemon) + geom_bar(mapping = aes(x = attack_weak_type,
    fill = attack_weak_type_new), position = "fill") + theme_bw() + xlab("Weak Attack Type") +
    ylab("Count"))
bc
```

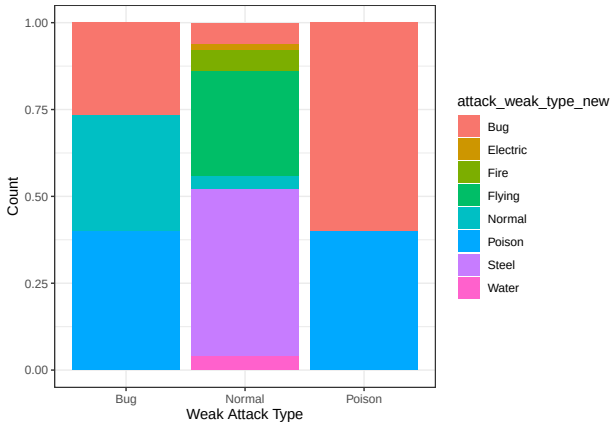

Figure 6: R Exercise Plot 6

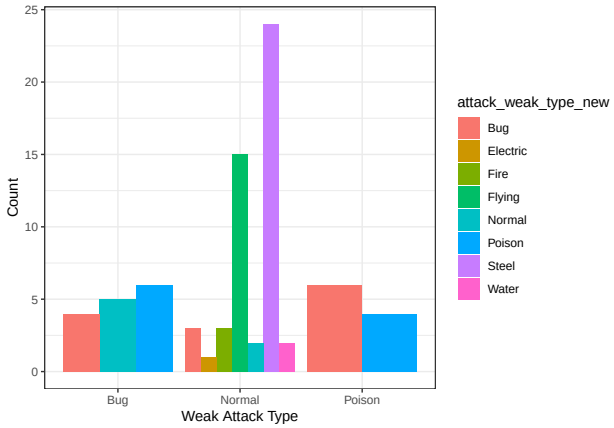

Figure 7: R Exercise Plot 7

```
bc <- (ggplot(data = pokemon) + geom_bar(mapping = aes(x = attack_weak_type,
    fill = attack_weak_type_new), position = "dodge") + theme_bw() + xlab("Weak Attack Type") +
    ylab("Count"))
bc
```