Abiola Gabriel Olofin

Professor Liew

DS 201

April 18, 2020

Project 2

For Part 1, we were asked to make a "box plot with error bars showing totals for all members of one category (e.g., mammals or birds)" and we had to make assumptions about errors in estimates. In order words, using one category(Mammals, Birds, Amphibian etc.), I was asked to create a box plot that showed the different members of an order of a category. I decided to use the category of Mammals. When looking at the graph, there are pros and cons. To start, one pro is that the graph tells me which family is more likely to be endangered based on where it is placed on the graph. This means that it is easier to identify which order of mammals are probably going endangered, which one are not, and which ones are unaccounted for. Another pro is that we are able to see where the outliers are in each order of mammal which can help us narrow down which mammal in particular is an outlier and why. However, there are some cons. This graph does not show the total population of mammals which can cause readers to question the credibility of the graph because we do not know whether I am actually showing the data in terms of population by park or overall.

For Part 2, we were asked to make a "stacked barchart showing totals for all members of one category and contributions from each park(they add up to the total)". In order words, using one category(Mammals, Birds, Amphibian etc.), I was asked to create a stacked bar graph that showed the different order of a category in each park. I decided to use the category of Mammals again. When looking at the graph, there are definitely some pros and cons. To start, one pro of this graph is that I was able to see how many total mammals are in each park. This allows us to

see the overall total number of mammals and which parks ,across the country, have the most mammals and which have the least. Another pro is that I was also able to see how many types of orders of mammals are in each park. This allows me to see which types of mammals are present in each park, where the highest and lowest concentration of those animals are and the total number of orders of mammals by parks. Another pro I noticed was that using order made the graph a lot easier to understand. I was debating on whether to use the mammals' family or order; however, after looking at the results of both, using the mammals' order provided a better representation of the data because not only did it allow us to see the total number of mammals by park and in general, it also allowed us to see the breakdown of mammals by park and see which parks might actually have issues with endangered animals when comparing each order of mammal by park. Finally, the last pro is that I was able to see every park even if they do not have a high population of mammals. This just makes it easier to identify which parks have which amount of mammals and mammals' scientific order. Even though there are pros to this graph and it was overall solid, there was one con that jumped. The major con we noticed was that even though we know that each park has mammals to some extent, the scale of the graph (Total), makes some parks look like they do not have mammals at first glance. This might cause misinterpretation of the graph seeing that some values are not able to be seen. Unless the park has a large number of mammals, it would be really difficult to compare the total number of mammals in each park and the total number of mammals' order in each park as well which is definitely a big con of this graph.

In Part 3, we are asked to "a histogram showing how much each category (you only need the 5 that we selected for Project 1) contributes to the dataset, i.e., how many entries in the dataset are from each category". In other words, this means we are creating a histogram that shows the breakdown of each category of animals and how much they contributed to the

data. There are many pros and cons. One pro is that we are able to see the total number of animals per category which gives us a good view of how many animals are in each category. However, one con is that if one category is too large, it makes the smaller categories non-existent which could cause misinterpretation of data.