Bradley Curthoys

UWA Data Analytics Boot Camp

Challenge #1

1. **Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?**
2. Regardless of country, the success rate of crowdfunding campaigns is between 50-60%, with an average of 54.64%. This leads us to conclude that based on this data, crowdfunding campaigns are successful approximately half the time.
3. There is a high degree of variability in the data set as indicated by large variance and standard deviation values. The distributions of the data sets are also heavily skewed, which leads us to conclude that the data is not a good sample of the population.
4. The three most prolific types of crowdfunding campaigns were shown to be Film & Video, Music, and Theatre. We can use this fact to make an inference that crowdfunding campaigns are more popular among those groups who participate in the Performing Arts.
5. **What are some limitations of this dataset?**

There is a methodology limitation in that no information about where this dataset came from or how it was collected is provided, so the validity or reliability of the data could be called into question.

Also, on the face of it, it seems that the data is from mostly white, economically affluent countries. It has the potential to be perceived as a cultural bias, which would be a research process limitation.

1. **What are some other possible tables and/or graphs that we could create, and what additional value would they provide?**

We could create a box-and-whisker plot, which would help us to visually identify potential outliers in the data. We could also utilize a distribution plot, in order to visually represent how the data is skewed.

1. **Use your data to determine whether the mean or the median better summarises the data.**

The median is a better summary of the data because the mean is being heavily influenced by outliers and is therefore a misleading measure of central tendency in this case.

1. **Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?**

According to the data, there is more variability with the successful campaigns. This does not make sense, as variability it meant to have an inverse relationship with sample size. Thus, the larger the sample size the smaller the variability. That is not what we see here in this data.