1. Use your data to determine whether the mean or the median summarizes the data more meaningfully.

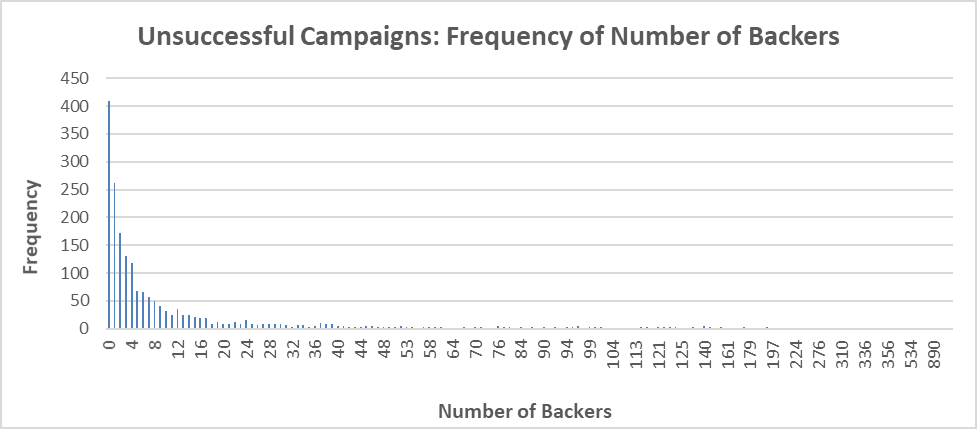
The dataset contains information about 4,000 projects. The current analysis divides the dataset by the result of the project, i.e. was it successful or unsuccessful. Then we compare the number of backers per project to the result of the project. As can be seen in Figure 2 and Figure 3, the dataset contains several outliers. For example, the campaigns that were successful range from having 1 backer to a project having over 25,000 backers. It a similar case for unsuccessful campaigns, which range from having no backers to certain projects having over 1500. Given the outliers and their magnitude, the median summarizes the data more meaningfully. This is because it presents the central number of the dataset, rather than taking the average.

Figure 2

Figure 1

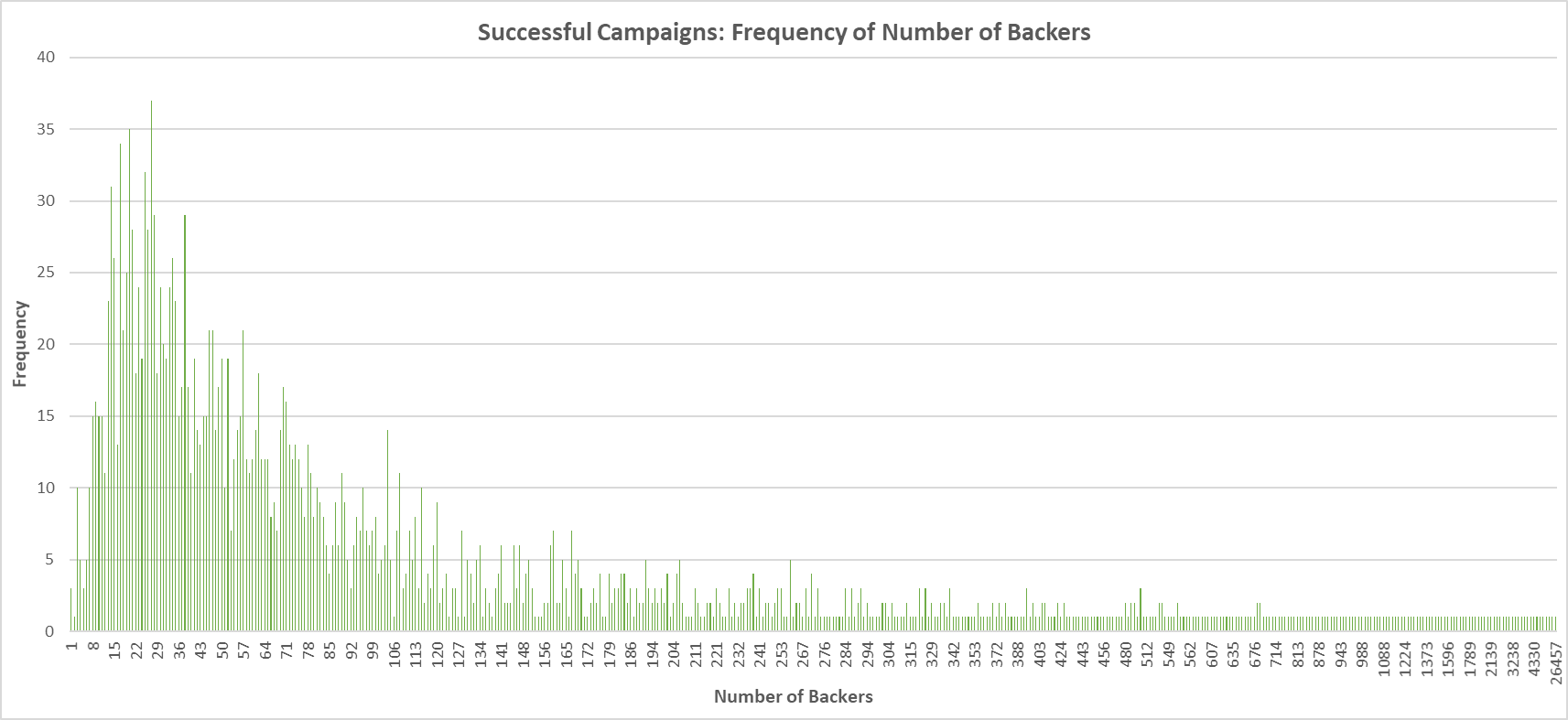
1. Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?

Figure 3

As seen in Figure 1, the standard deviation of successful and unsuccessful campaigns is 844 and 74, respectively. The standard deviation is a tool that is beneficial when trying to understand the spread of a dataset. Given these numbers, it is evident that there is more variability with successful campaigns. Furthermore, the 1.5 IQR rule states that anything outside of the interquartile range could be an outlier. Figure 4 displays what these numbers mean in this dataset. The numbers in Figure 4 when compared to the data represented in Figure 2 and Figure 3, clearly demonstrate that the variability is far greater within successful campaigns. These conclusions make sense because there are far more successful campaigns, as compared to unsuccessful ones. Of the successful campaigns, there are many that range from having 1 backer to some having over 25,000 backers.

Figure 4