**Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?**

1. **Backers:** Most backers from the data provided come from the USA

This can be represented with a simple bar graph, as below:

However, below are a few examples of limitations for this insight:

* It is not globally representative; we can only see data of backers from 7 countries.
* There are only 1000 campaigns listed in this dataset, which may not accurately represent the entire Crowdfunding portfolio.
* The data doesn’t consider respective currency value, making direct comparison difficult. Exchange rates may vary over time, compounding this issue.

1. **Funding:** Corresponding to the chart above, the majority of total funds pledged was in USD, however on average, Canadians pledged the most per backer, indicating that campaigns might do better in Canada. By comparison, Denmark backers pledged the least on average.

As a side note, when looking at the funding goal, amount pledged and the campaign outcome, it is clear that perhaps the funding expectation was too high for campaigns that ultimately failed to reach their goals.

1. **Categories:** The parent categories of Film & Video, Music, and Theatre had the largest number of backers compared to other categories.

However, upon further analysis, despite having a large number of backers, these categories still failed to meet their funding goals in some countries.

You can see this in China’s data, and Italy’s data, against the top-funded parent categories (namely Film & Video, Music and/or Theatre) in the charts below:

**What are some limitations of this dataset?**

* Geographically ambiguous (country does not indicate project location or backer location)
* No data on demographics of backers (could provide further insight)
* Sample size is too small for fair analysis
* Currency (as mentioned above)
* Campaign details lacking (eg: incentives to encourage funding and therefore higher chance of success)
* Number of countries is limited (geographical representation)
* Parent and/or sub-categories may be too broad/vague compared to blurb descriptions

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

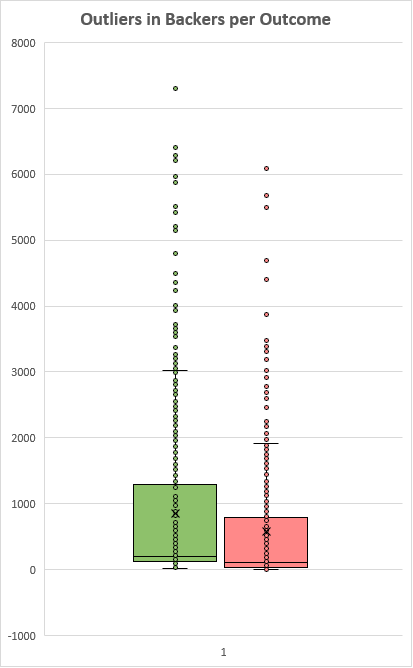
* We could also use a pie graph to visualise the first insight
* Goal vs pledged scatter plot to visualise successful and failed projects (with trend line)
* Box and whisker plot to identify outliers, giving insight into potential skew in data
* Pie graph to visualise the percentage of funding per category, to show which category was most preferred by backers

**BONUS**

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

The median better summarises the data because there is a large number of outliers present. Having outliers skews the mean, and therefore the median is a better measure.

The below chart shows this, with green representing successful campaigns and red representing failed campaigns.



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

I need help answering this question – I am confused by the large numbers!