

## Entry 1: Project Proposal

### (A) Find data source and formulate research questions

- (a) Data Source: <https://www.kaggle.com/socathie/kickstarter-project-statistics>
  - (i) most\_backed.csv - the 4000 most backed projects ever on Kickstarter (as of October 30, 2016)
- (b) Possible Research Questions - Brainstorm
  - (i) *What are the traits of the most funded projects?*
    - 1) What categories are most funded?
    - 2) What locations are most funded?
    - 3) Do certain words in the blurb affect the amount of funds?
  - (ii) *Where are the most popular areas for projects?*
    - 1) What categories are most popular in these areas?
  - (iii) *What is the average amount funded from each backer?*
    - 1) In each category?
    - 2) In each location (state)?
  - (iv) *How overfunded are the top projects?*
    - 1) In each category?
    - 2) In each location?

### (B) Project Plan

- (a) Tentative Title
  - (i) Baby Got Back(ers): Analysis of Kickstarter's Best Projects
- (b) Research questions and hypotheses
  - (i) *What are the traits of the most funded projects?*
    - 1) Which categories are most funded?
    - 2) Which locations are most funded?
    - 3) Do certain words in the blurb affect the amount of funds?

Hypothesis: If the Kickstarter project has high funding then it probably contains either a humorous blurb, is in the games/entertainment category, or is located in the California area.
  - (ii) *What is the average amount funded from each backer?*
    - 1) In each category?
    - 2) In each location (state)?

Hypothesis: If the Kickstarter Project has high funding, then the project backers pledged an average of \$50.
  - (iii) *How overfunded are the top projects?*
    - 1) In each category?
    - 2) In each location?

Hypothesis: If the Kickstarter Project has high funding, then the top projects are overfunded by 400%.

(c) Motivation

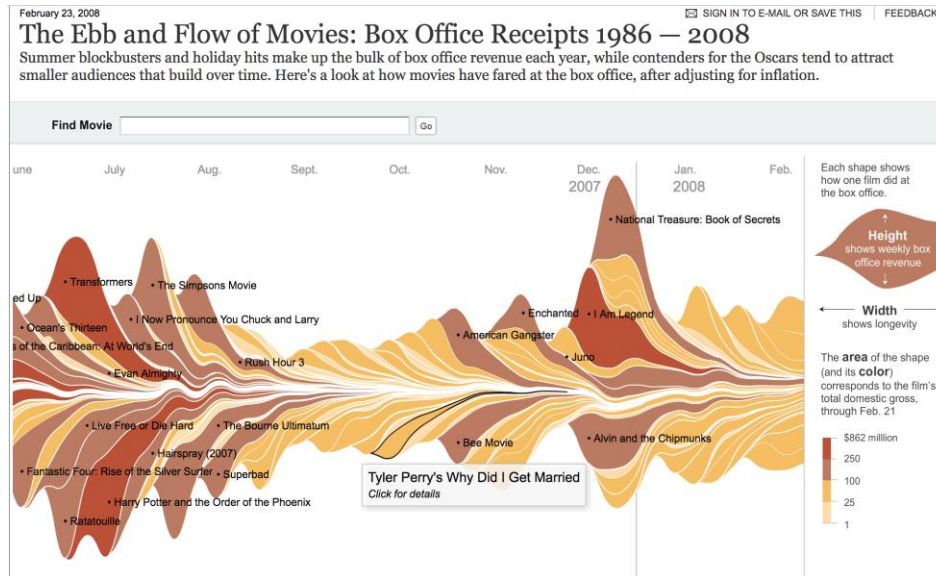
- (i) We feel that with the current climate on the Internet, ideas are being thrown around all over the place. It's harder for people to get their good ideas out to the public. We chose the Kickstarter data because we really like what Kickstarter is doing. Their mission is to help bring creative projects to life by helping struggling artists and other creators that just need a little help in getting their great idea rolling. We believe in what they're doing and are interested in learning more about this data, specifically the projects that have been the most successful in terms of raising funds.
- (ii) In terms of the motivation for our research questions, the dataset gave us access to information regarding the top 4,000 most backed projects ever on Kickstarter. The variables in the dataset include: amount pledged, a blurb about the product, the creator, product category, monetary goal for funding, number of backers, title of the project, and Kickstarter URL. Because of this information, we wanted to look at the various traits of the most funded projects, the average amount pledged by each backer, and how overfunded these most backed products could be. We came up with these initial questions while just discussing the data with one another. Knowing the traits of these successful projects could help others with good ideas get funded. The average amount each backer pledges gives good insight into what people are willing to spend on a potentially risky project. Finally, knowing how overfunded these projects are is interesting to see because it is most likely related to its popularity. That can spark more research as to why these products are so successful.

(d) Data sources and technical process

- (i) <https://www.kaggle.com/socathie/kickstarter-project-statistics>
- (ii) Most\_backed.csv - contains 4000 most funded projects on Kickstarter as of October 30, 2016 since the launch on April 20, 2009
- (iii) Technical process (transformations, data sources, merge etc.)
  - a) What data after this... Patents. Or we want to compare more platforms that are doing crowdsourcing?
  - 2) Convert all currencies to USD

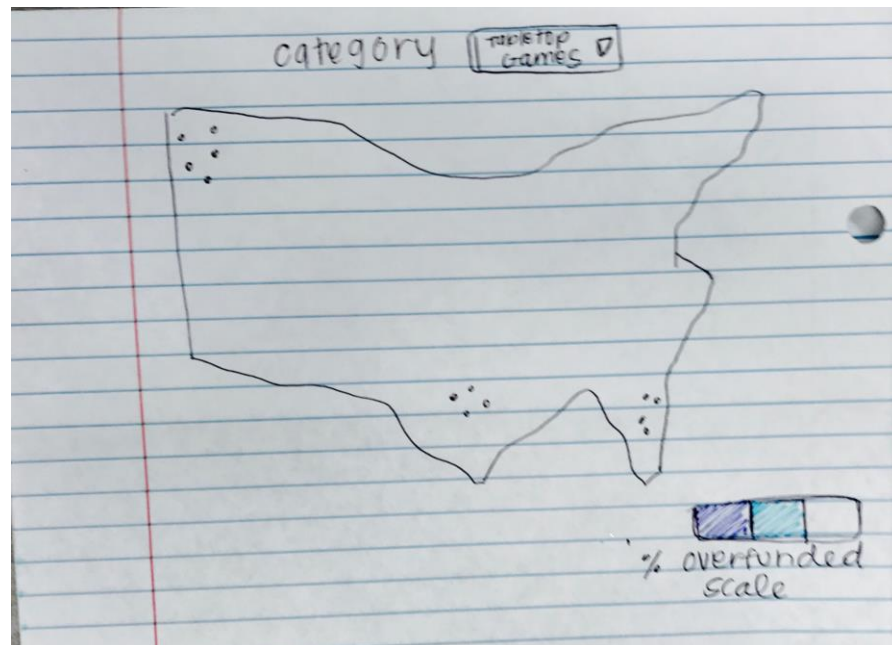
(e) Potential Visualizations (appropriate and why?)

- (i) Possible Visual 1
  - 1) Link: [http://www.nytimes.com/interactive/2008/02/23/movies/20080223\\_REVENUE\\_GRAPHIC.html](http://www.nytimes.com/interactive/2008/02/23/movies/20080223_REVENUE_GRAPHIC.html)



- 2) This interactive visual is a loose example of something that we envision for our data set. We think that it is an appropriate graph because it allows the viewer to select a Kickstarter project and learn more about the project in a pop-up. Visually this graph encodes the data, similar to the way we could potentially encode our data.

(ii) Possible Visual 2



- 1) This interactive visual could be used to help answer the geographic influences of overfunding a project. One of our research questions was interested in seeing if there was a relationship to the category and the location of where the Kickstarter was founded. (i.e. Would more films be started in Los Angeles?) Also we think this could answer the question whether

geographic location has any influence with the percent of the overfunding that the project receives attributing to the success of the campaign.

## Entry 2: Project Part 1

### [Revised Project Proposal](#)

In this stage of our project, we made further efforts in cleaning our data and merging multiple sources. We spent time using Tableau to develop visualizations that would be useful for a potential investor to make an investment decision. The following tasks were performed on our data and dashboard.

#### **In Excel:**

- We pulled data from World Bank about country GDP Ranking.
- We converted the goal and amount pledged to USD using the appropriate exchange rates found online. We created a column with the country based on the currency and then we manually ensured that the country matched the location column because some currencies were used in other countries.
- We also calculated percent overfunded for each project.
- Ensured that all the titles were easy to code when moving to R.
- Added “kickstarter.com” to prefix the URL column.

#### **In R:**

- We used the XML package in R to scrape data from the URL. We then enlisted help regarding the number of comments/images etc. from a webpage in R. We were able to save the html source code as a value in R. From that, we wanted to know if there was a way to do a CTRL+F type search for the number value that appears a bit after a string of text. We have the number that appears after “data-comments-count” however, we just want the number.
- Looked at number of unique values for Category and Country.
- Did a word count of the title of the project and the project blurb.
- Used “readLines” in a for loop in order to read the lines of HTML for each of the 4,000 project URL’s. This process was used to find both the number of comments and then updates on each of these project’s Kickstarter profiles.
  - In the loop we also put information about the pattern of the line of code surrounding the numbers of comments and updates we wanted, and then used functions like grep, getexpr, and mapply in order to extract the exact numbers we needed from the HTML coding.
  - One for loop was used for comments, one was used for updates.
- Exported data to csv/excel.

### **In Tableau:**

- In tableau, corrected all variable types and only made relevant columns visible.
- Split LOCATION into City and State/Country columns
- We made graphs for:
  - AvgAmtPledbyGDP- Amount Pledged by GDP Ranked 1-195
  - NumRecords- Total Projects by Category
  - WordCounts- Title/Blurb Wordcount
  - AmtPledgedbyLoc- Amount Pledged by City
  - Updates by Category
  - Comments by Category
- Changed color scheme so each graph on the dashboard was consistent.

## **Entry 3: Project Part 3**

At this stage in our project, we took the feedback received on the adjustments from Entry 2: Project Part 2 and applied it to our dashboard. We also published our dashboard and storyboard to Tableau and published this process book.

### **In Tableau**

- We fixed the “Amount Pledged by GDP” to be displayed horizontally.
  - We moved the GDP Rank and Country Name to the y-axis to make it easier to read, regardless of how many countries are listed.
- Created a StoryBoard in Tableau to assist the user in understanding the components of the Dashboard. This helps the user create insights that facilitate their decision making.
- Adjust Dashboard to “generic desktop” (1366x768) to ensure that it will be aesthetically pleasing on any computer screen.
- Published Dashboard on Tableau.
  - <https://public.tableau.com/profile/publish/BabyGotBackers/BabyGotBackersDashboard#!/publish-confirm>
  - Under the metadata, you can view the individual visualizations, the dashboard (“Baby Got Back(ers) Dashboard”) and the storyboard (“Baby Got Back(ers)”).

### **Final Task**

- Finalized Process Book
- Published on Github