

Final Project Proposal Form

Name:

Gabriel Krotkov

Partner's name (if applicable):

1. A 3- or 4-sentence description of your topic:

In the FIRST Robotics Competition, the most prestigious award is the "Impact Award" - an award for the team that best changes culture to advance STEM education and inspiration in their community. Part of the criteria for winning the Impact Award is a text submission summarizing the team's community impact over the past 5 years - and the winning submissions for every Regional and every District Championship event are publicly available. I'm interested in understanding the dynamics of Impact Award submissions over time: from copycat effects to shifting trends, there are a number of interesting dynamics to investigate.

2. A list of your research question(s):

- 1) How has the focus of Impact Award winning teams changed over the years?
- 2) Do District Championship winning teams cluster apart from Regional winning teams?
- 3) Are there strong geographical effects? (Do teams coming from certain teams tend to prioritize similar efforts?)
- 4) Of this corpus, each year 1 or 2 teams are selected to join the "Hall of Fame" - what traits separate those teams?
- 5) Are there observable "copycat" effects? (When a team enters the Hall of Fame, do subsequent submissions shift to mirror their submission?)

3. A short summary of any working hunches you have regarding your questions:

- 1) I expect there are big "current events" changes (teams responding to COVID, e.g.) but also a few macro trends - anecdotally I notice that later years tend to have a lot of focus on advocacy and engagement with government.
- 2) Truly don't have any idea about this one. Very curious!
- 3) I could imagine a few reasons this could be. For one, perhaps a team just keeps winning in a certain area with a similar year to year message. But that would be a less interesting result than something along the lines of "American Mountain West teams tend to focus on special needs education while Turkish teams focus on driving down the costs of robotics"
- 4) Also stumped on this one.

4. A description of your proposed corpus (or corpora) and a statement of its strengths and limitations in addressing your research question(s):

A set of more than 600 text responses (a few short answer and a length-limited essay) averaging around 3000 words each. I converted each response from pdf to txt for easy code readability.

Strengths:

- Completeness. Only 3 essays are missing between 2016 and 2023.
- Relevance. There is no off-topic text in the corpus
- Since the submission field only takes a simple set of characters each text is well-formatted (no weird punctuation marks.)

Limitations:

- There is some amount of text in the pdf files that isn't really content, like question prompts. This will require some data cleaning.
- I do not have access to the non-winning submissions for baseline setting.

5. A list of 2 or 3 secondary sources that will inform your methodology and a very brief discussion of their relevance to your project:

Biber - "The Multi-Dimensional Approach to Linguistic Analyses of Genre Variation: An Overview of Methodology and Findings"; for the MDA approach to understanding the influence of different dimensions of text features

Greies Et Al. "N-grams and the clustering of registers" - this paper could help parse out the registers in these corpora