STAT 287 Final Project Check-In 2

Julia Zimmerman, Philip Nguyen

October 30, 2020

1

1.1

Major activities/milestones planned for this week (in the timeline that you had going into the start of the week).

1. Organizing and cleaning the information I've extracted – 1 week (deliverable: vectors, matrices, or other refined structure)
2. *Check progress to see if project idea should be abandoned; last chance*

1.2

Major activities/milestones accomplished this week (note: this may or may not actually be what was stated in 1a – that's okay, as long as you're making progress and learning!).
Progress compared to original timeline:

1. Organizing and cleaning the information we've extracted – 1 week (deliverable: vectors, matrices, or other refined structure) - done
2. *Check progress to see if project idea should be abandoned; last chance* – done (no reason to think it should)

Progress as tracked on GitHub via closed issues:
We have both been working on most of the items, but since the assignment instructions say to make it clear who has done what, we listed the main person who resolved the issues below.

1. Final work verifying our foundational data - that we had a sane definition of trope, a good list of tropes, our scripts were doing what we thought they were, our results matched what we'd expect if we looked at the website manually, etc. Not glamorous but important to convince ourselves we have a reasonably solid base for the rest of the project.
   Verifying script behaviour (Julia)
   Organizing the data we already have in a way that is practical (Phil)
   Verifying data matches what we would get going over the website manually (Phil) Bug-fix for script behaviour (Julia)
   Figuring out which tropes to include: what lives where (Julia), what is the "Main" category (Phil)
2. We made our first networks! (Both - we each made visualizations of different parts of the network.)

Please note that although Julia tends to be more verbose in workstyle on GitHub with issues and comments, Phil has contributed at least as much to resolving the issues.

2

2.1

Open challenges and questions (including what – if anything – are the challenges that Ethan or I can help provide feedback or pointers on?)
We're not currently stuck on anything, but this week we really need to nail down and prioritize our research questions. We also need to explore our visualization options. The network of tropes is larger than we had anticipated (although we didn't really have a good size estimate going into this), with about 27000 tropes and potentially around 3 million edges. So some of our initial exploratory data visualization has been less than readable - and Gephi may not be the right tool, since it has crashed several times when I've been working on a graph (which sometimes hasn't saved!). Phil has been trying out Python for the visualization. I do like writing the networks in gml file format, since it is easy to automate based on the data structures we made in week 1. We are going to meet with Jane Adams next week for advice on visualizations.

2.2

Major changes to research plan (if any, based on what you've learned or accomplished thus far, and the unexpected challenges you've faced this week) None so far (fingers' crossed).

3 Revised anticipated timeline

1. Setting up Github, etc., to begin work – 2 hours (deliverable: being able to push and pull from Github)
2. Downloading or otherwise preparing a data set for use – 2 hours (deliverable: having a reasonable looking file)
3. Parsing the data to find structural information – this could be arbitrarily hard, so generous time should be allotted – 1 week (deliverable: data frames, vectors, or CSV)
4. *Check progress to see if project idea should be abandoned*
5. Organizing and cleaning the information I've extracted – 1 week (deliverable: vectors, matrices, or other refined structure)
6. *Check progress to see if project idea should be abandoned; last chance*
7. Exploratory data analysis and visualization – 1 week (deliverable: interpretable images)
8. Writing up results – 3 days (deliverable: report in at least outline form)
9. Creating "birth certificate" summary of project [Eli20] – 1 day (deliverable: neat summary)
10. Audit for transparency, ethics, etc. – 2 days (deliverable: action items or approval)
11. Realize something actually makes no sense or wasn't doing what you thought – 1 week (deliverable: a fixed version of whatever needs fixing)
12. Refining visualizations – 3 days
13. Refining write-up – 3 days (deliverable: a polished report)
14. Making presentation – 2 days (deliverable: a presentation)
15. Donating money to (1) a social justice cause and (2) an environmental cause[1] – 1 hour (deliverable: receipts)

Total estimate: $\approx$ 6 weeks and 1 half day $\approx$ 6 weeks (I got the 6 weeks total from looking at the course schedule on Blackboard).

[1] We have the resources and opportunity to complete this project due to luck and privileges granted by society, some of which come at the cost of other people and the environment. Therefore some remuneration is within the scope of this project.