Working With Open Data – Project Progress Report 1 – April 8, 2014

|  |  |
| --- | --- |
| Team: | Soteria |
| Members: | Evie Phan, Shreyas, Ashley Desouza, Victor Starostenko |

Progress since April 2 Project Proposal Submission:

* Explored the data
* Picked the variables we want to look at
* Solved the issue of how we will tackle the Boolean variables we have in the data set by deciding to look at similarities. We worked out Jaccard Similarity for our datasubset ( ie. n(AᴖB)/n(AᴗB) )
* Looked at different front end visualizations to display the information

Ongoing work this week:

* Cleaning data
* Data Exploration/Analysis
* Recoding data in IPython data frame
* Wireframe front end visualization
* Brainstorming visualization techniques

Major obstacles:

* Figuring how to use the data to arrive at an actual conclusion that can be useful for the open source community
* How to visualize the data to make it understand to a person who doesn't have domain expertise
* Breadth of data