

STAT 380 Final Project

Updated: April 2, 2019.

Goal

Use your tools from STAT 380 and your skills as a data scientist for social good.

Requirements (Still developing)

- You must work with a partner
- You must use GitHub to manage your project & collaboration
- Project must conform to STAT 380 Style Guide
- Project must showcase a wide range of skills from STAT 380
- The project must join data from two different sources (i.e., not two data sets from the same source)
- You and your partner will present your key findings to the class during the scheduled final exam period (May 1 @ 2:30pm - 4:20pm in 211 Keller Bldg)

Data Sources

There are tons of data sources out there. It's best to think of a **cause first** and then search for data (it's probably out there). To help illustrate some of the breadth of data sources that could lend themselves to analyses for "social good" here's a list that I compiled in approximately 10 minutes.

- Large open data initiatives
 - Philly: <https://www.opendataphilly.org/>
 - NYC: <https://opendata.cityofnewyork.us/>
 - US Government: <https://www.data.gov/>
 - UK Government: <https://data.gov.uk/>
 - European Union: <http://data.europa.eu/euodp/en/home>
 - India: <https://data.gov.in/>
 - **Everywhere**: <https://www.opendatasoft.com/a-comprehensive-list-of-all-open-data-portals-around-the-world/>
- Independent organizations:
 - Microfinance: <https://build.kiva.org/>
 - Gapminder: <https://www.gapminder.org/>
- Official government functions, e.g., do an internet search for "US Bureau of" and look at suggestions:
 - Bureau of economic analysis
 - Bureau of labor statistics
 - Bureau of prisons
 - Bureau of justice
 - Bureau of land management
 - Bureau of census
 - World Bank
 - World Health Organization

Q & A

Q: What counts as “social good”

It can be defined pretty broadly, but part of the assignment will include a chance for you to explain how you feel your work contributes to some kind of social good that personally matters to you and/or your partner. For example, you could track down some data to analyze a topic in the popular media or among your community.

Q: What if we can’t solve it?

The purpose of the project is to try and use your skills to address a problem for the benefit of others. Clearly, these are often very difficult problems, so it’s not required that you succeed in “solving” the problem. Just that you use your skills to contribute a thoughtful effort to address it.

Q: Can I do the project by myself?

No. Every student needs to work with a partner. Each team will present their results during the final exam period scheduled for our class, and we will not have time for all teams to present if people work individually.

Q: Do we have to use GitHub?

Yes.

Q: Where’s the GitHub Classroom Repo?

No Repo will be provided through GitHub Classroom. You need to create a public (i.e. not private) GitHub Repo for your work. You will also need to invite your partner to the Repo to collaborate. **It needs to be public so I can see it to grade your work**, but I also want you to have this access so you can use the project in a work portfolio that you can share with others when you apply for jobs, etc.

Q: Do we have to use the STAT 380 Style Guide?

Yes.

Q: How many GitHub commits are required?

Each partner must make at least 5 commits to the project, and **each partner** must be responsible for a significant proportion of the final content (whether code or narrative or both).

Q: What if I have a final exam conflict?

If you have a University approved final exam conflict, you need to talk with Dr. Beckman immediately to discuss arrangements. The process for declaring final exam conflicts is managed by the Registrar’s office https://www.registrar.psu.edu/exams/exam__overload.cfm.

Q: Do I have to do everything we've done in STAT 380?

Not **everything**. Since this is a final project for STAT 380, you should try to show as much depth and breadth of STAT 380 content as possible. In some cases, this will result in doing an extra task or two that is loosely related to your research question just to showcase that you know how to use some of the additional tools learned in STAT 380.