Project Proposal

iSchool Canvas Support

Start Assignment

Due Nov 5 by 11:59pm **Points** 30 **Submitting** a website url

Summary

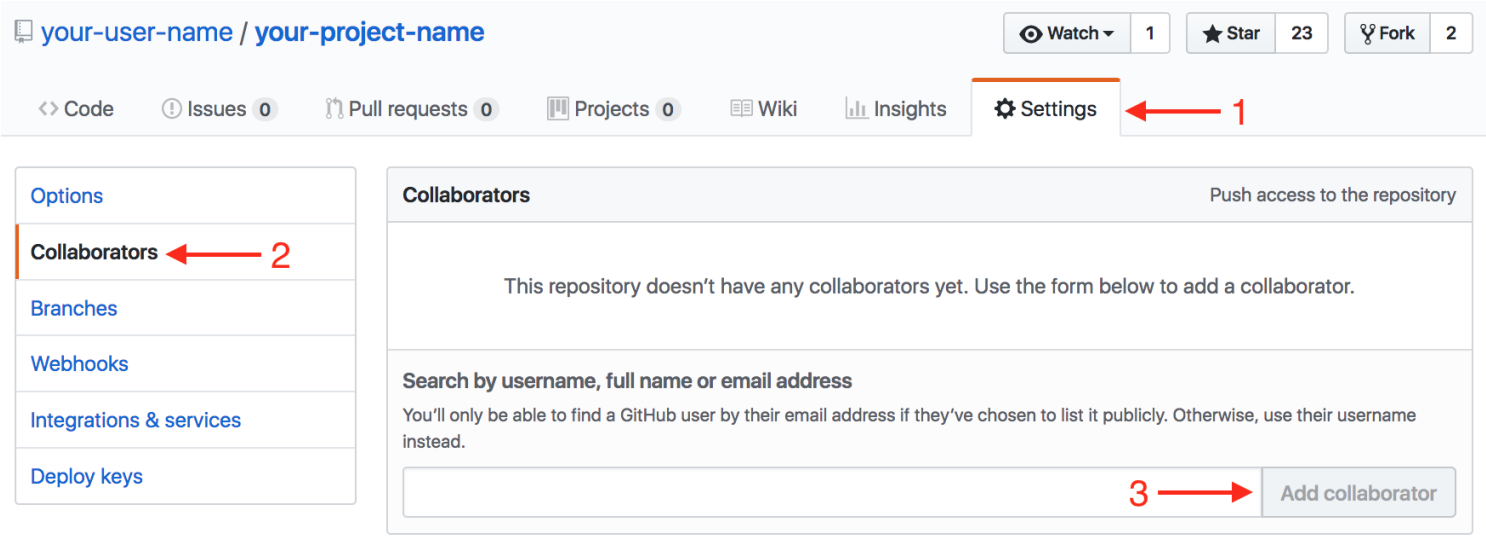
As described in the **Final Project Assignment**, this is your opportunity to work with your peers to explore a topic of your choice, and create an interactive web application for communicating your insights. As a team, you'll decide what domain you wish to explore, and how you want to present data about that topic. As the first step in the project, this proposal will allow you to:

- Describe your domain of interest
- Articulate pertinent questions in that domain
- Identify, download, and describe at least 3 datasets related to that domain

Project Set-up

Part of the purpose of this assignment is for you to create the repository in which you'll complete your final project. To do so, **only one person** will follow **this link** (https://classroom.github.com/a/AkNmzBY9) to create a repository ***for your group***. This will create a repository for your **group** to all work on.

Once **one person** has created a repository, they should be able to add collaborators to your team:



If you experience any difficulty with this, please contact your TA **as early as possible**.

Your final project description (see below) should be written in **well-structured and formatted Markdown** in the README.md file in this repository.

Domain of interest

As we've discussed, data science can expose underlying patterns in any domain that uses or collects data (which is nearly *any* domain). Anything from the **forced relocation of homeless individuals**

(https://www.theguardian.com/us-news/ng-interactive/2017/dec/20/bussed-out-america-moves-homeless-people-country-study) to how people gender representation (https://pudding.cool/2017/09/this-american-life/) in the media, data can expose interesting (and actionable) patterns. In this section, you'll identify a domain that you are interested in (e.g., music, education, dance, immigration -- any field of your interest) and answer the following questions in your README.md file:

- Why are you interested in this field/domain?
- What other examples of data driven project have you found related to this domain (share **at least 3**)?
- What data-driven questions do you hope to answer about this domain (share **at least 3**)?

We strongly suggest that you complete this section first, discussing what you might want to learn, then move forward with the data discovery process.

Finding Data

We are lucky enough to live in a time when there is lots of publicly available data made possible by governments, journalists, academics, and companies. In this section, you will **identify and download** at least 3 sources of data related to your domain of interest described above (into a folder you create called *data/*). You won't be required to use all of these sources, but it will give you practice discovering data. If your data is made available through a Web API, you don't need to download it. For each source of data, provide the following information:

- Where did you download the data (e.g., a web URL)?
- How was the data collected or generated? Make sure to explain *who* collected the data (not necessarily the same people that host the data), and who or what the data is about?
- How many observations (rows) are in your data?
- How many features (columns) are in the data?
- What questions (from above) can be answered using the data in this dataset?

Submission

For your submission, please **submit your website URL of the GitHub repository** you'll be using for you final project.