**CS109 Projects**

For your final assignment in this course you will work on a month-long data science project. The goal of the project is to go through the complete data science process to answer questions you have about some topic of your own choosing. You will acquire the data, design your visualizations, run statistical analysis, and communicate the results.

**Project Team**

You will work closely with other classmates in a **3-4 person project team**. You can come up with your own teams and use Piazza to find prospective team members. If you can’t find a partner we will team you up randomly. We recognize that individual schedules, different time zones, preferences, and other constraints might limit your ability to work in a team. If this the case, ask us for permission to work alone. In general, we do not anticipate that the grades for each group member will be different. However, we reserve the right to assign different grades to each group member based on peer assessments (see below).

**Project Milestones**

There are a few milestones for your final project. It is critical to note that **no extensions will be given** for any of the project due dates for any reason. Late days may not be used. Projects submitted after the final due date will not be graded. If you anticipate any issues (e.g., due to business travel) you need to send an email to the staff mailing list at least one week in advance.

Monday, November 11: Project proposals due

Week of November 18: Project review meeting with your TF

**Thursday, December 12: Projects due (including web site and 2 minute screencast)**

Tuesday, December 17: Project presentations and best project prizes

**Deliverables**

There are several deliverables for your project that will be graded individually to make up your final project score.

**Proposal**

You start your project by forming your groups and letting us know what topic you are interested in exploring by submitting this[project proposal form](https://docs.google.com/forms/d/1xNLa6dd0Yc4F7HriKYmReYda7PqmbYBmPjIgkUax4kA/viewform). **Each team will only need to submit one form.** Based on your proposals we will assign a TF to your team who will guide you through the rest of the project. You will schedule a project review meeting with your TF during regular lecture times of the week marked in the schedule. Make sure all of your team members are present at the meeting. Online students can schedule a Skype meeting with their TF.

**iPython Process Book**

An important part of your project is your iPython process book. Your process book details your steps in developing your solution, including how you collected the data, alternative solutions you tried, describing statistical methods you used, and the insights you got. **Equally important to your final results is how you got there!** Your process book is the place you describe and document the space of possibilities you explored at each step of your project. We strongly advise you to include many visualizations in your process book.

Your process book should include the following topics. Depending on your project type the amount of discussion you devote to each of them will vary:

* Overview and Motivation: Provide an overview of the project goals and the motivation for it. Consider that this will be read by people who did not see your project proposal.
* Related Work: Anything that inspired you, such as a paper, a web site, or something we discussed in class.
* Initial Questions: What questions are you trying to answer? How did these questions evolve over the course of the project? What new questions did you consider in the course of your analysis?
* Data: Source, scraping method, cleanup, etc.
* Exploratory Analysis: What visualizations did you use to look at your data in different ways? What are the different statistical methods you considered? Justify the decisions you made, and show any major changes to your ideas. How did you reach these conclusions?
* Final Analysis: What did you learn about the data? How did you answer the questions? How can you justify your answers?

As this will be your only chance to describe your project in detail make sure that your process book is a standalone document that fully describes your process and results.

**Code**

We expect you to write high-quality and readable Python code in your process book. You should strive for doing things the right way and think about aspects such as reusability, error handling, etc. We also expect you to document your code.

**Project Website**

You will create a public website for your project using [Google Sites](https://sites.google.com/) or Github Pages or any other web hosting service of your choice. The web site should effectively summarize the main results of your project and tell a story. Consider your audience (the site is public) and keep the level of discussion at the appropriate level. Your iPython process book and data should be linked to the web site as well, either using a zip file, [github](https://github.com/), [bitbucke](https://bitbucket.org/%E2%80%8E)t, or another code hosting site. Also embed your main visualizations and your screencast in your website.

**Project Screencast**

Each team will create a **two minute screencast with narration** showing a demo of your iPython process book and/or some slides. Information about how to prepare these screencasts can be found [here](https://docs.google.com/document/d/1alPLuBOW5YPoQDa57KZes1h72PoQDoDj21-UEKOHp1I/pub). Please make sure that the sound quality of your video is good - it may be worthwhile to invest in an external USB microphone. Upload the video to an online video-platform such as YouTube or Vimeo and embed it into your project web page. We will show the best videos in class.

We will strictly enforce the two minute time limit for the video, so please make sure you are not running longer. Use principles of good storytelling and presentations to get your key points across. Focus the majority of your screencast on your main contributions rather than on technical details. What do you feel is the best part of your project? What insights did you gain? What is the single most important thing you would like your audience to take away? Make sure it is upfront and center rather than at the end.

**Peer Assessment**

It is important to provide positive feedback to people who truly worked hard for the good of the team and to also make suggestions to those you perceived not to be working as effectively on team tasks. We ask you to provide an honest assessment of the contributions of the members of your team, including yourself. The feedback you provide should reflect your judgment of each team member’s:

* Preparation – were they prepared during team meetings?
* Contribution – did they contribute productively to the team discussion and work?
* Respect for others’ ideas – did they encourage others to contribute their ideas?
* Flexibility – were they flexible when disagreements occurred?

Your teammate’s assessment of your contributions and the accuracy of your self-assessment will be considered as part of your overall project score. Please fill out [this form](https://docs.google.com/a/g.harvard.edu/forms/d/1tf3SApcC5aqG2t6mjasDV82Oiq4hsdVg99oAQL4qZ7I/viewform) for the evaluation.

**Submission Instructions**

To submit your projects, create a folder named **lastname\_firstinitial\_project** and place your iPython notebook and your other files in this folder. Compress the folder (please use .zip compression) and submit it in the appropriate dropbox folder. *If we cannot access your work or links because these directions are not followed correctly, we will not grade your work.*You can submit once per team or submit once per team member.

**Grading**

80% of the grade will be based on your iPython process book. This includes the quality of your data analysis and Python code, the complexity and level of difficulty of your project, completeness and overall functionality of your analysis. The remaining 20% of your grade will be based on your web site and screencast and the quality of their storytelling aspects. Your individual project score will also be determined by your peer evaluations.