Final Project

Project Proposal Due: Nov 28, 2022

Design Review and Feedback: Dec 2, 2022

Final Code and Video Due: Wednesday, Dec 7, 2022, latest

There is an opportunity to show your work at a visualization workshop on Dec 5, 2022.

Final project themes

Your final project should be one of the following themes:

1. **Data analysis/explainer projects** involve analyzing a data set that has not previously been analyzed extensively and producing an interactive visual explainer of the most interesting findings you uncover in the data set (in the style of the pudding.cool for example). For data analysis/explainer projects, the final implementation deliverable is an interactive article containing at least **three substantially different interactive charts with accompanying text written in the form of an article** and a short video demo (up to 2 min long) explaining the work.

Here are some examples:

Newspaper Navigator

The Newspaper Navigator is a Library of Congress dataset that extracts the visual content of historic newspaper pages. They apply crowdsourcing and machine learning techniques to identify photographs, illustrations, maps, comics, cartoons, headlines, and advertisements. Design an interactive explainer to let people explore different aspects of this data set.

Data source: https://news-navigator.labs.loc.gov/

Why is my flight delayed? Investigating Flight Delays & Cancellations

Flight delays are estimated to have cost air travelers billions of dollars. FAA/Nextor estimated the annual costs of delays (direct cost to airlines and passengers, lost demand, and indirect costs) in 2017 to be $26.6 billion. With an external dataset, you can try to uncover the correlation between flight delays and factors like weather. Using the geological information, you can also identify and visualize the geological pattern of flight delays. Investigate the common causes or potential patterns of the delays and present the insights you find with visualization.

Data source: https://www.kaggle.com/usdot/flight-delays#flights.csv

Other data sources:

As noted in Assignment 2, there are other datasets:

- World Bank Indicators, 1960–2017. The World Bank has tracked global human developed by indicators such as climate change, economy, education, environment, gender equality, health, and science and technology since 1960. The linked repository contains indicators that have been formatted to facilitate use with Tableau and other data visualization tools. However, you're also welcome to browse and use the original data by indicator or by country. Click on an indicator category or country to download the CSV file.
- Chicago Crimes, 2001–present (click Export to download a CSV file). This dataset reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2001 to present, minus the most recent seven days. Data is extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system.
- Daily Weather in the U.S., 2017. This dataset contains daily U.S. weather measurements in 2017, provided by the NOAA Daily Global Historical Climatology Network. This data has been transformed: some weather stations with only sparse measurements have been filtered out.
- Social mobility in the U.S. Raj Chetty's group at Harvard studies the factors that contribute to (or hinder) upward mobility in the United States (i.e., will our children earn more than we will). Their work has been extensively featured in The New York Times. This page lists data from all of their papers, broken down by geographic level or by topic. We recommend downloading data in the CSV/Excel format and encourage you to consider joining multiple datasets from the same paper (under the same heading on the page) for a sufficiently rich exploratory process.
- The Yelp Open Dataset provides information about businesses, user reviews, and more from Yelp's database. The data is split into separate files (*business*, *check-ins*, *photos*, *review*, *tip*, and *user*), and is available in either JSON or SQL format. You might use this to investigate the distributions of scores on Yelp, look at how many reviews users typically leave, or look for regional trends about restaurants. Note that this is a large, structured dataset and you don't need to look at all of the data to answer interesting questions. In order to download the data you will need to enter your email and agree to Yelp's Dataset License.
- Kaggle Datasets are part of Kaggle, an online community platform for data scientists and machine learning enthusiasts. Kaggle allows users to collaborate with other users, find and publish datasets, use GPU-integrated notebooks, and compete with other data scientists to solve data science challenges.

2. **Improve Tableau dashboards downloaded from Tableau Public.** For improving Tableau dashboard projects, the final implementation deliverable is a Tableau workbook file and a short video demo (up to 2 min long) explaining the work. **Refer to assignment 3 for dashboard examples.**

Projects will be carried out by a team of up to 3 people.

The first step is for you to identify the type of project (data analysis/explainer or Tableau dashboard improvement) and the specific topic you will work on. We would like you to write these up as a project proposal which is due on Monday, Nov 28, 2022. Each group will then be responsible for presenting the project in draft form on Dec 2, 2022. This design review will give you a chance to obtain feedback on the work and prepare for the final deliverables which are due on Dec 7, 2022, latest.

Project Proposal Due: Wednesday, Nov 28, 2022

As a first step, you should create a project proposal that includes the names of the members of your group and a short (about 2 paragraphs) description of the data analysis/explainer or Tableau dashboard improvement project you plan to work on. You should submit the proposal as a PDF with your team members' names to Bharath and me.

Design Review and Feedback Due: Friday, Dec 2, 2022

I will review your project to provide feedback on the project and help you prepare for the final presentation. It is fine if your project is not yet in a fully "complete" state, but by this point you should have made substantive progress, including working (if still rough) prototypes of your main visualizations and interactions. You should prepare a short presentation of your work focusing on a demo. More information on the timing of these presentations will be posted as we get closer to the deadline.

Final Project Code and Video Due: Wednesday, Dec 7, 2022

The final deliverables include:

- **Code:** an implementation of your project.
- **Video:** a 2-minute video demo (with voiceover) explaining your project. If it is a data analysis/explainer project it should demo the interactions in the visualizations and describe your main findings. If it is an improved Tableau workbook, you will have a video demonstrating the improvements.

Submission: You should submit your final deliverables to Bharath and me (access to a running executable and a zip file of the code, or a link to a GitHub repository as well as the video).