

Homework 3

Due 10/9

1 Description

Download the survey data below, and write a report at least three pages in length about things you found interesting. Many of my students in previous semesters answered questions like “does gender affect answer to question 3”; if you are going to look at such a thing, think about it in terms of “which questions does gender affect, if any.”

Your project must include at least one hypothesis test, as well as at least one visualization. The visualization must look good—do not submit something that just looks like a default matplotlib graph that has colors added. Write this report as if the head of your department is reading it (they are mathematically literate but expect something polished).

As discussed in class, have an introduction that summarizes your report and includes interesting findings, a background section explaining the dataset. This section should also include any significant data cleaning issues and how you handled them. Following this a section on findings, and a conclusion. You will also need to submit a PDF of the code used to analyze this dataset.

There will also be a notebook where you will need to answer specific questions to test the correctness of your analysis; this will be posted at a later date.

2 Data

The data is here. It consists of a survey from this semester, and two surveys of different sections last semester. Each table is slightly different. One includes a priming question to experiment with the effects of (priming. The most recent one, 2024, has many of the genders reversed in the questions themselves.

This data is also extremely messy—I have made no attempts to clean it for you. You will need to fix missing values, merge columns, and figure out how to handle any other mistakes you may find.