QTM 151

Week 8 – Final Project Lab

Umberto Mignozzetti

Final Project

- Today, we will talk about your final project.

- I have four types of projects that each of your groups can implement.

- I'll pitch each project, and at the end you have to select a project for your group.

Option 1 - National Health and Nutrition Examination Survey

NHANES

The National Health and Nutrition Examination Survey (NHANES) is a program designed to assess the health and nutritional status of adults and children in the US.

The survey examines a nationally representative sample of about 5,000 persons each year. It combines interviews and physical examinations.

The NHANES program began in the early 1960s and has been conducted as a series of surveys focusing on different population groups or health topics.

All questionnaires, datasets, and related documentation are publicly accessible at: https://wwwn.cdc.gov/nchs/nhanes/Default.aspx

NHANES

For your final project, as a first option, please use NHANES 2017-2018 data wave.
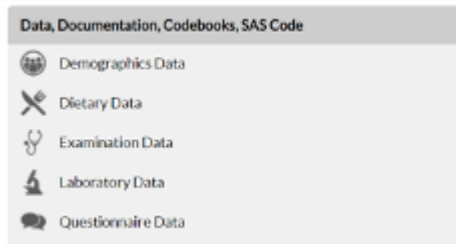
- The only reason for not doing this is if you want to analyze the data over time.

If you are interested in analyzing trends over the years, my second option suggestion, you can also look into other data collected in previous years.

- Note that each wave is cross-sectional: from year to year, it was NOT repeated measurements from the same individuals.

NHANES

- Choose 2-3 data files from the available list and merge them according to your analytical goals.

- The data files are in SAS format (XPT), which can be imported into R using the following commands:

```
library (Hmisc)
library (SASxport)
a ← read.xport ("file_name.xpt")
```

Option 2 -- Expedia

Expedia

This dataset is provided by Expedia.com, a travel website, which is used to book airline tickets, hotels, car rentals, and vacation packages.

The dataset file and video will be on Canvas.

Hint: this is a large dataset, please start small (one random sample of the dataset).

Option 3 -- COVID

COVID

COVID-19 is still disrupting our way of life.

There are several scholars collecting datasets on COVID around the world

Moreover, NYTimes has built great dataviz on the subject, all in R.

My suggestion is for you to explore the packages about the subject.

Option 4 -- World Data

World Data

The Quality of Government Institute collects data from all over the world, since 1946 (some variables even early!)

They organize a few datasets that merge together scholars' data, plus their own collected variables.

There are several types of datasets:

- Time-Series Cross-Section
- Cross-Section only data
- Surveys with experts
- And many others!

Rules

Final Project: Rules

To receive full marks you will have to apply skills you learned from this course to:

- import
- tidy, transform, wrangle
- and visualize the data

You also have to prepare 2-3 slides to present the results with your teammates. Five minutes presentations!

Final Project: Rules

Each project will be judge by:

- Insight and creativity of the project
- Data visualization of the project
- Clarity of the presentation

Important Deadlines:

- Nov 12: Submit your one-page project proposal on Canvas (20 pts)
- Nov 15: I will notify you my decision / talk to you if I feel the project is unrealistic.
- Dec 01: Final submission on Canvas (60 pts)
- Dec 03: Presentation (20 pts)

Questions?

Have a great weekend!