Data Visualization Project Proposal

Figma Visualization Planning Process book

Team Members:
Stella Wong, Darci Kovacs, Chiamaka Azodo

Title: Fruit and Vegetables in the United States

Primary goal: To map out what fruit and vegetables are available seasonally in the United States, and how the price and yield has changed over time.

Key Questions:

- What fruits and vegetables are in season each month of the year?
- What fruits and vegetables are available in different regions/states of the US?
- How has crop yields in the U.S. changed over time?
- How have fruit and vegetable prices changed over time?

Procedure/Scope:

1. Import the data from NASS
2. Get NASS data into dataframe
3. (Order unclear)
    a. Create US map and show fruit and vegetable seasonality chart by state
    b. Create chart visualization of historic fruit and vegetable yield
    c. Create chart visualization of historic fruit and vegetable prices

Data:
National Agriculture Statistics Service data of agriculture in the United States.
Historical yield of major crops data of Pangea publisher for earth and environment
Fruit and vegetable prices - USDA Ag Data commons
Configurations for the annual data
Configuration for monthly data
Configuration for State data 2022-2011
Configuration for State data 2010-2002
FAO - Worldwide crop production data
FAO - Worldwide import/export data - partner level
FAO - Worldwide import/export data - crop level
FAO - Code book

Seasonal acres harvested
Seasonal monthly data (filtered fruits and veg) (sold at market) (only in the last complete year 2022) adding fresh market + processing (filter for every option that has price received – parity? Maybe unnecessary). Fruits and veg removed grouped fruits and vegetables (endive & escarole), and overall categories (total, mixed, other). If the fruit/veg exists for that month of that year, it goes on the calendar.

Reddit data scraping for the r/fruit subreddit to get qualitative data for text analysis

- The intent is to hopeful filter by time frame and to show the fruits that are being spoken about; and show that discussions can be influenced by seasonality?

Nice to have(s) if time allows:
Data of agriculture worldwide
D3 visualization

Visualizations:

- Map of fruits and vegetables available in different states in different seasons.
    - If time, ability to put in your zipcode or county and see what fruits and vegetables are available locally.
- Calendar of when fruits and vegetables are available during the year.
- Line chart of historical yield of fruits and vegetables.
- Visualizations of price fluctuations over time.

FAO Fruit and Vegetables list

- Tomatoes
- Bananas
- Watermelons
- Apples
- Grapes
- Oranges
- Mango, guava, mongosteen
- Plantains
- Tangerine, mandarin, clementine
- Pineapple
- Cantelope/Melon
- Peaches/necarine
- Pears
- Lemon and lime
- Papaya
- Plum and sloes
- Grapefruit and pomelos
- Dates
- Strawberries
- Avocadoes
- Onion/shallots
- Cucumber
- Cabbage
- Eggplant
- Carrot/turnip
- Chili and pepper
- Lettuce
- Spinach
- asparagus
- Cauliflower and Broccoli
- maize(corn)
- Potatoes
- Sweet potatoes
- Green garlic

Templates for Shiny App tabs

- https://shiny.rstudio.com/gallery/ncaa-swim-team-finder.html
- https://shiny.rstudio.com/gallery/ncaa-swim-team-finder.html

Meeting Notes

02-27-2023 Kick Off

- Individually research the dataset and come back together with an idea/suggestions on what we each want to do and then we decide.
- 03/06/2023 at 7pm follow-up meeting

04-16-2023:

- Made a figma for the project
- Downloaded data and shared to collective github
- Started working on code

Data:

- List of publicly available datasets (from Thesis courseworks)
    - Github list of publicly available datasets

Project Description on Syllabus/Lecture: (Section to be deleted before submission)

Project proposal instructions

Final group project (30%): A final group project (3-4 students) presented in the form of a website and accompanying project book. You will analyze data of your own choosing and report the results using:

1. static images based on ggplot2,
2. maps using geospatial data,
3. visualizations of text analyses, and/or
4. network visualizations AND
5. prepare a hosted, interactive display of some of your visualizations. There will be in-class presentations of the final projects (if class size allows).

For the final project, your data should:

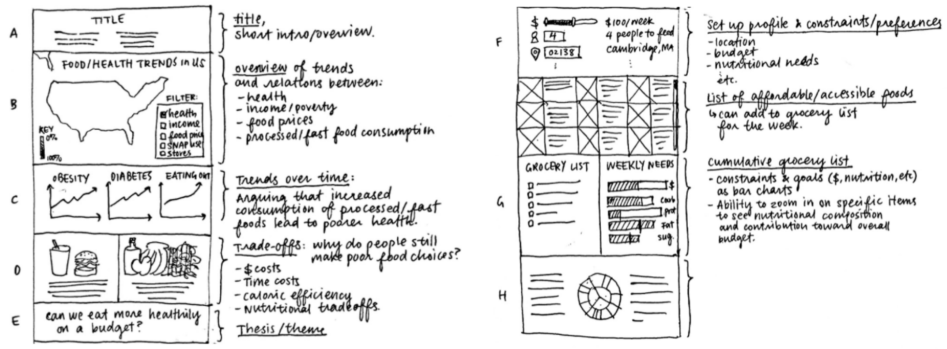
1. be a larger dataset to allow analysis in different forms
2. contain at best one (or more) of the following elements
    - some geographic identifiers (cities, addresses. lat-long etc.) to allow mapping
    - some text to allow simple text analyses
    - contain a network matrix (or item relationships) to plot visual networks
3. Preference is for data that you either put together yourself from existing datasets, and/or using some API to collect the data you need from an online source (e.g. Twitter).

Final Project Output:

- final output should be something that you could (if you choose to) use as a portfolio for job applications etc. to show your skill for that, it needs to be visually appealing, informative, and (partially) interactive.
- Example for successful project: Language and the 2016 US Presidential Election

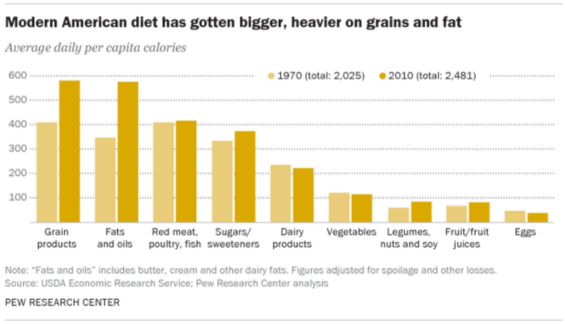
Final Project - Process Book

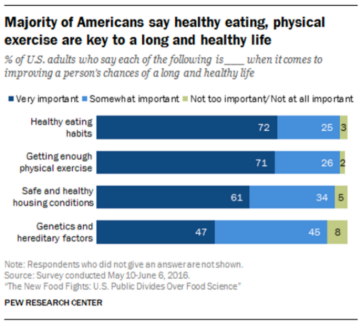
Part of the project is a **process book** in which you note (as you go along) how your thinking developed, perhaps including some plots that you later discarded, some (handwritten) sketches of how you imagine the final design etc.

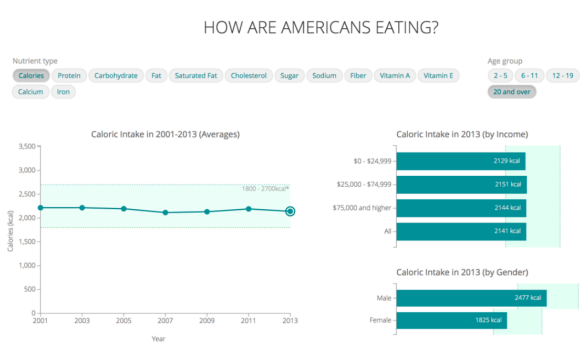


Final Project Hypothetical Example:

- **Title:** Unhealthy habits - Bad diets are the root of the obesity epidemic
- **Abstract:** obesity epidemic in the U.S. ... bad diets are to blame ... explore how have diets changed over time ... how are diets related to socio-economic factors etc.
- **Data**:
    - adult and changed obesity rates in the U.S.(CDC) and the World (WHO) (pair with geographic info for mapping)
    - Pew Research Center's surveys on food and nutrition attitudes and how diets have changed
    - Content of Recipes changed over time (text analysis of recipes, extracting measures)
    - Spread of Obesity in a Large Social Network







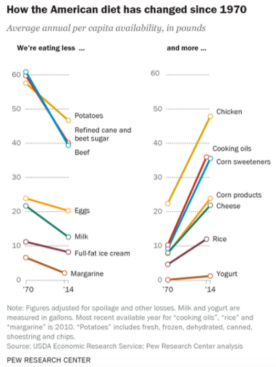
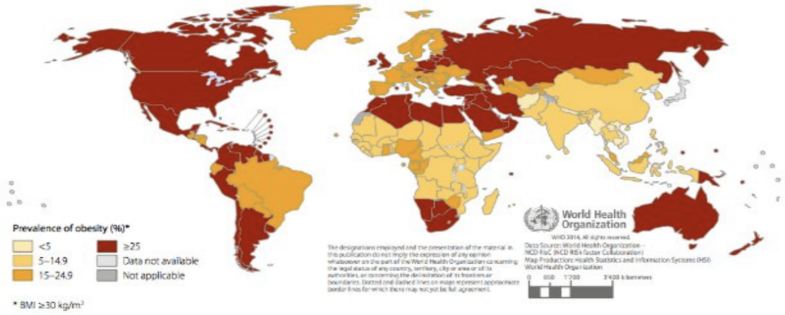
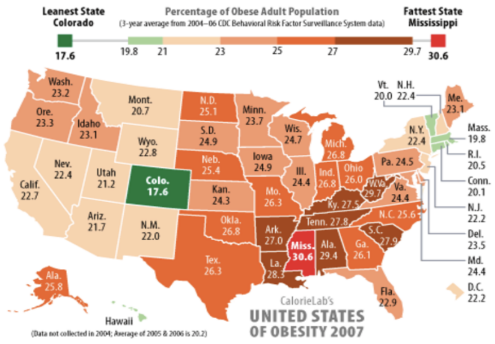


Fig. 7.2 Age-standardized prevalence of obesity in women aged 18 years and over (BMI ≥30 kg/m²), 2014





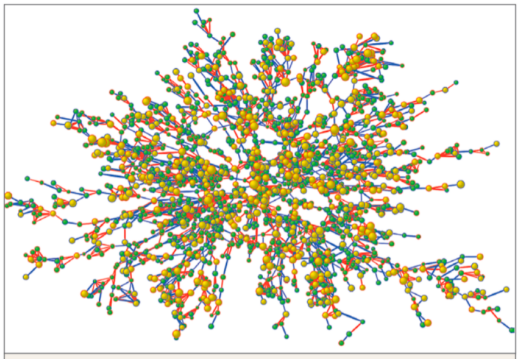
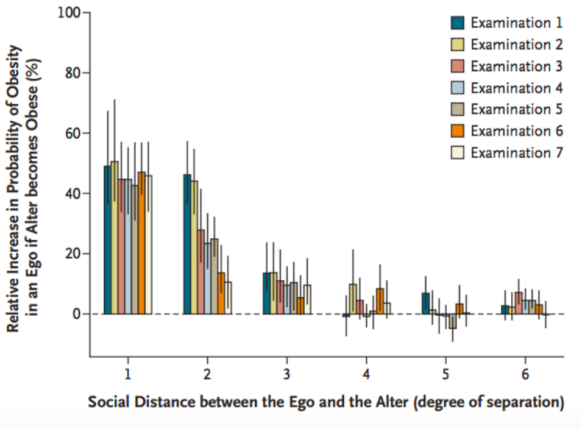


Figure 1. Largest Connected Subcomponent of the Social Network in the Framingham Heart Study in the Year 2000.
Each circle (node) represents one person in the data set. There are 2200 persons in this subcomponent of the social network. Circles with red borders denote women, and circles with blue borders denote men. The size of each circle is proportional to the person's body-mass index. The interior color of the circles indicates the person's obesity status: yellow denotes an obese person (body-mass index, ≥30) and green denotes a nonobese person. The colors of the ties between the nodes indicate the relationship between them: purple denotes a friendship or marital tie and orange denotes a familial tie.



●