Final Project Guidelines, or A Brief Walkthrough of Information Design Process

Prompt (from 170117-mit-4s02-syllabus-final.pdf)

1. What is the question? What's the story you are trying to tell (if it's narrative)? What are you helping people find? (if it's a tool for exploration)
2. Who is the audience? What is the context in which it will be used?
3. What would you use for source data?
4. How much data is it?

Data Collection

1. Data you already have
    - Personal data (e.g. fitbit steps, journal)
    - Research data (e.g. lab experiments)
    - Data received from clients, classes, etc.
2. Data you need to collect
    - Data obtainable from research (e.g. First of Her Kind, tampon tax)
    - Data obtainable from download or query (e.g. US census data, No Ceilings)
    - Data obtainable through purchase or specific request

- Tools: manual collection, google search, API, crowd-sourced materials

Data Cleaning

- File type
    - What is the file type of the data (e.g. csv, xls, json, txt)? Can you load it? Does it require conversion?
- Size
    - How big is the data? Is it easily loadable? Do you need to cut the data into multiple data frames or combine multiple data files into a single data frame?
- Format
    - How do you want to organize your data? For data frames, what do you want your rows and columns to be?
- Subset
    - What part of the data do you want to use? Drop parts that you don't want.
- Naming
    - How do you want to name your variables? Good names will save a great amount of time and confusion.
- Unit
    - Is the data in the unit that you want? Do you want to convert it? (e.g. fraction to percent, number of doctors per person to number of people per doctor, Fahrenheit to Celsius)
- Calculation
    - Are there parts of the data that you want to combine, separate or calculate with? (e.g. Total sales tax = state tax + local tax, separate population into age groups)

- Tools: manual organization, Excel, R, Python, etc.

Data Analysis

- Scope
  - How many observations are there? How big is the sample and the population? What to you know about the subject (e.g. gender, age, race)? What is the time span? What is the geographic span?
- Missing values
  - Are there missing values in the data? How are they encoded (e.g. NA, -9999, “”)? How much is missing? Can you guess why they are missing? Is there a need of data credibility re-evaluation? Do you want to drop observations with missing values, leave the data as it is, or impute the missing values?
- Summary statistics
  - range, percentile, mean, median, mode, variance, standard deviation
- Descriptive visuals
  - frequency table, bar chart, pie chart
  - histogram, scatterplot
  - correlation matrix
- Transformation
  - standardization, log-transformation

- Tools: Excel, R, Python, etc.

Data Representation (from 170117-mit-4s02-syllabus-final.pdf)

- What is the story you are trying to tell? What do you want your audience to uncover?

Refinement (from 170117-mit-4s02-syllabus-final.pdf)

1. What about the first time user? How will they understand and interact with your piece without you there to explain it to them?
2. Does your piece need additional context like sidebar/intro text to support the visualization?
3. How might interactivity help support your concept?
4. What is the context of use? Desktop, mobile, installation, print?
5. Consider your grid structure, color, typography, and animation. Keep it simple if you are struggling with formal design elements.