

Response Summary:

1. Student Information *

First Name	Jack
Last Name	Kinzig
Major	Data Visualization
Course (e.g. CGT 270-001)	CGT 270
Term (e.g. F2019)	S2022

2. Email Address *

(University Email Address is required.)

kinzig@purdue.edu

3. Visualization Assignment *

- Lab Assignment

Generate

4. Identify appropriate data sources: is the data publicly available? What search methods were used? *

Data source 1	Federal Election Commission
Data source 2	Candidates' monthly finance reports
Data source 3	Candidates' fundraising reports

5. Data format: what format is the data in? Structured vs instructed? All text, a combination, multiple sources? Is it primary or secondary data? *

The data is structured, and is all text. It is primary data from the Federal Election Commission.

6. Data types: what types of data are in the data? How are they stored? What is the access to the data (API, JSON, txt, csv, etc.)? What structure holds the data (data base, spreadsheet, etc.)? *

Locations, amounts, dates, names. The data is stored in an Excel document in a .csv format.

Evaluate

7. Variables: list the data variables? What are the parameters? Give them names. What are the dependent variables and independent variables? *

Which candidate made the expenditure, amount spent, what the money was spent on, when it was spent, where it was spent, who received the payment. The only identifiable dependent variable is the recipient of the payment, as larger expenditures will often be used differently than smaller ones.

8. Audience & Assumptions: list any assumptions you have about the data. Who is your audience? *

I don't see any assumptions to be made about the data. The audience is anyone interested in politics or finance.

Generate

9. What real life behavior does the data reflect? Does it show patterns of activity, regularity of events, a timeline, population data, etc? Explain. *

It represents a timeline of a competition

11. What are the weaknesses of the data source? Is it likely that the source will be available in the future? Is the data complete? What is the quality of the data? Is it specific to your needs for the current project? Is the data in the format you need? Are there missing data? Explain. *

The only weakness I can see in the data is the sheer amount of it - it's daunting to look at.

12. What information is emphasized? What is the central focus of the data? Explain. *

The main emphasis is on the size and purpose of expenditures, as well as who made said expenditures.

13. At what level of granularity is the data provided? Is the data summarized, or do you have access to the raw data? Is the data categorized or is the data in a format that allows you to create your own categories, etc. Explain. *

It is raw data, and is broken down into categories. The excel document uses columns for each variable.

14. What is the scope of the data? What topics can be covered using the data? Is there a time range/frame? Is the data for a specific area/discipline/demographic etc.? Explain. *

Topics that can be covered include: the rise and fall of political campaigns, how much each candidate spent, on what and where the most money spent. The time range is mid-late 2015 to late 2016.
