

Response Summary:

1. Student Information *

First Name	Kyle
Last Name	Choi
Major	WEPD
Course (e.g. CGT 270-001)	CGT 270-003
Term (e.g. F2019)	S2022

2. Email Address *

(University Email Address is required.)

choi687@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	Crunchbase (Census)
Data source 2	Investment Network
Data source 3	Publicly Available Data

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

Structured data, All text, One source, Secondary data

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.)? *

Characters, Integers, Spreadsheet, Excel Pages

Evaluate

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

Company name, category list, market, country code, state/region, investor name, funding amount (usd), year founded, year funded.

Independent variable: Investment amount, funding amount raised, company field, company market

Dependent variable: Acquisitions made, investment amount

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

Target audience: investors, competing companies

Assumptions: All companies are legally filed, investors go through legal means of funding a business

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. *

Correlation between what field a company is in and how much money it has raised

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. *

Data is largely outdated, missing lots of bigger companies. The data is partially complete, with some companies missing key information such as investment and acquisition amount

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

The central focus of the data is the steady increase in money raised over time

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. *

The data is incredibly granulated, as there is both a summary and access to all the raw data used to compile the summary

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. *

The data includes companies founded from Q1 of 2007 up to Q4 of 2014. Topics that can be covered include investment amounts over time, venture vs seed funding, and private equity vs angel investors.
