

Response Summary:

1. Student Information *

First Name	David
Last Name	Luo
Major	Data Science
Course (e.g. CGT 270-001)	CGT 270-LEC
Term (e.g. F2019)	S2022

2. Email Address *

(University Email Address is required.)

luo354@purdue.edu

3. Visualization Assignment *

- Homework Assignment

Generate

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

Data source 1	Labor Force Statistics from the Current Population Survey From the US Bureau of Labor Statistics.
Data source 2	Only data source required
Data source 3	Only data source required

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

This is a structured data set containing text and numerical values

This is 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.)? *

This alphanumeric data held in an excel spreadsheet.

Evaluate

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

Occupation: Text data

Total employed: Integer

*All fields below are given in percent of total employed

Women: Integer

White: Integer

Black or African American: Integer

Asian: Integer

Hispanic or Latino: Integer

*Percentages may not add to 100% due to data not being presented for all races

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

Assumptions:

The "Men" category is not given, I will assume that category is 100 - "Women"

Gender categories are not broken down by race. I will assume that percentages within each race are similar.

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. *

Employment data of the US in for 2021.

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. *

Weaknesses in the data source include the fact that it was a survey, and response bias may or may not affect the data.

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

Demographics for each occupation in America.

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 summarized, I do not have access to the raw survey data.

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 scope of this data is United States in 2021
