

Response Summary:

1. Student Information *

First Name	Rae
Last Name	Fu
Major	Animation and Visual Effects
Course (e.g. CGT 270-001)	CGT 270-004
Term (e.g. F2019)	S2022

2. Email Address *

(University Email Address is required.)

fu345@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	https://www.bls.gov/cps/cpsaat11.htm The data is publicly available. It could be found by searching it on google.
Data source 2	N/A
Data source 3	N/A

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. It is a combination of text, integers, and decimals. It is primary data since it is taken directly from a survey.

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.)? *

The data is categorical qualitative and discrete measures. They are stored as text and numeric. The access to the data is .xlsx. The structure that holds the data is a spreadsheet.

Evaluate

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

The data variables are occupation, total employed, women, white, African American, Asian, Hispanic or Latino. The parameters are a combination of African American, Asian and Hispanic or Latino as Minorities. Occupation is the independent variable, and the others are the dependent variables.

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

Assumptions are the same number of participants who took the survey fall into each of the variables except for occupation and total employed and that the data is accurate. The audience is people curious about the spread of occupations African Americans have now.

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. *

The data reflects population data, because the data shows what occupation people have, which is a part of people's lives.

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 source is likely to be available in the future because the data is given by the US Bureau of Labor Statistics. There are some parts of the data that are missing. The quality is good, since there are a few nulls. It is not specific to only my needs for the assignment but it satisfies my need. The data is editable, which will allow reformatting to filter data. There are missing data, typically for that other than total employed.

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

The central focus of the data is the breakdown of total employed by occupation and race.

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 raw and is in a format that allows the creation of new categories, since it is in an editable spreadsheet.

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 can cover topics involving occupation, race, sex, and Hispanic or Latino ethnicity. The time range is the year 2020. The data is for demographics in occupations.
