

Response Summary:

Acquire Worksheet

Goal: Identify appropriate data sources, analyze the data, identify data types, variables, list assumptions about the data

Objectives: Students will identify and acquire data from appropriate data sources

Outcomes: Data for the current visualization challenge

1. Student Information *

First Name	Ayomikun Obayomi
Last Name	Obayomi
Course (e.g. CGT 270-001)	CGT 270-009
Term (e.g. F2019)	F2021

2. Email Address *

aobayomi@purdue.edu

3. Visualization Assignment *

- In-class Assignment

Generate

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

Data source 1	Yes, the data was publicly available. I used the site called Kaggle to find similar databases to the World Cup Squads of 2018 that show the stats of the players. The researcher got his inspiration from a website called, https://sofifa.com/ , (the search method used was a well-designed observation) which shows that the information he/she derived is accurate and valid as they provided us with basically everything the website had to offer
Data source 2	The second data source that I found that correlates to my initial dataset is also publicly available. The search method that the researcher used was well-designed research- even better than the first data source because the researcher doesn't limit himself to just one primary source but rather more than one primary source and the researcher had also done their research by narrowing down their checkboxes to a few questions that would guide his inspiration.
Data source 3	Data source 3 is publicly available. The researcher makes use of three reliable primary sources that are equipped with outcomes of a soccer players' ability to carry their country and team to football glory. However, I would say that there are some details missing from the data (as the researcher stated in the section for improving the dataset) but nevertheless, it is not the researcher's fault but from the main source.

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

All three data sources are structured because they were written in some form of a spreadsheet, have a ton of values and numbers for different purposes. My sources are a combination of data because it includes the date of births of these players, integers for how many caps (appearances) for their club and country, and characters for their names, names of clubs, countries, position, and group. It is secondary data because it is a summary, a sheet of information that organizes the players alphabetically, and also interprets everything about the players from age to country.

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 types from my data are dates, characters, integers, and strings. They are stored in a structured format. Access to the data is .xlsx. A spreadsheet holds the data together.

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:

Ages
date of birth
Caps
Goals
Country

The parameters:

Ages
Date of birth
Goals
Caps

The dependent variable is the position that the player plays that is affecting them not to score a certain amount of goals.

The independent variable is the number of goals that player has

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

The audience is anyone that interested in soccer or in particular shows a dataset of all information about all the stats about a player that participated in the World Cups(2018).

Generate

9. What real life behavior does the data reflect? Does it show patterns or activity, regularity or events, a timeline, population data, etc? Explain. *
The data reflects a timeline that happened in the year 2018 and also shows a regularity of events because the world cup happens every 4years

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. *
I believe that the source will be available in the future because it shows a past representation of the players that participated in the World Cup of 2018. The dataset is however not completed because it doesn't state which country was in a certain group (for those who might have forgotten), the winner of the group and which country won it. The quality of the data has not gone through any peer-review or editing because it is dataset that provides a visualization in the form of a spreadsheet. Yes, the data was in the format of xlsx.

12. What information is emphasized? What is the central focus of the data? Explain. *
The information that is emphasized is the generally who each player that participated in the World Cup of 2018 is, the country that they are from, the team that they play for and their date of birth.

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. *
Fortunately, the data is both summarized and I also have access to the raw data from the website <https://sofifa.com/>. The data has already been precategorized by the researcher into categories of ages and name (alphabetical order)

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 topics that can covered in the dataset can be the players' biographies from their countries squad. The time range/frame was the year 2018 because contemporarily some players have either retired from the national team or changed clubs. The data only acknowledges the participation of a player that played for their country in the year 2018.
