Response Summary:

1. Student Information *

First Name	Saswat
Last Name	Mishra
Major	Web Programming
Course (e.g. CGT 270-001)	CGT 270-004
Term (e.g. F2019)	S2022

2. Email Address *
(University Email Address is required.)
mishra78@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	Wimbledon Champions
Data source 2	Not available
Data source 3	Not available

5. Data format: what format is the data in? Structured vs instructed? All text, a combination, multiple sources? Is it primary or secondary data? *
It is structured in rows and columns. It is a combination of text and numbers. It is Primary 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.)? *
There are a lot of text and numbers. They're stored in the spreadsheet. It is accessed through a csv. A spreadsheet holds the data

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 : Gender (String), Champion (String), Mins(Integer), Runner-up nationality(String), Champion Nationality(String), Runner-up(String), Score(String), Runner-up Seed(Integer), Champion Seed(Integer), Year(Integer). The independent variables are the Year and the Champion and Runner up variables. The dependent variable is the Score variable.

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

The assumptions are that all the data is entered correctly and no mistakes were accidentally entered. My audience are tennis enthusiasts who would like to see who the Wimbledon Champions are in previous years.

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 shows a timeline of events happening

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. *

There is some data missing in some of the columns. It is likely that this source won't be too useful for the future as it only shows Wimbledon Champions till 2016

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

The central focus of the data is who is the Champion players in certain years.

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. *

Data is already separated into categories that aren't filtered.

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 is from 1877-2016. The data doesn't go beyond this time range. Furthermore if only Male champions is to be looked at, the time range is from 1877-1955 while for Female champions the time range is from 1955-2016