Problem 1: 25 points
The Meaning of Life

In 2021, Pew Research Center conducted a survey of 2,596 adults in the United States (report here). They asked "What aspects of your life do you currently find meaningful, fulfilling, or satisfying?" They did not report raw data, but below is a generated dataset based on reported results.

Topic	Frequency
Family	716
Friends	300
Material Well-being	265
Career	262
Challenges	285
Spirituality	250
Society	210
Health	172
Hobbies	136

(a)
Compute the Relative Frequencies for each response category.

(b)
Construct a bar graph of the Relative Frequencies.

(c)
Interpret the data in a paragraph (2 or more sentences).

Problem 2: 25 points
Board Game Weights

The file `bgg.csv` on Canvas contains a database of every board game on the popular site "Board Game Geek." These data can be used to answer the following questions.

(a)
The column `averageweight` gives the average user assessment of the "weight" (i.e., complexity) of each game. Games with a value of 0 have not been rated and should not be included in any of the following analysis.

Create a table of summary statistics for the average weight of games. This table should include the Minimum, 1st Quartile, Median, Mean, 3rd Quartile, Maximum, Sample Variance, and Sample Standard Deviation (note the use of the word 'Sample' even though this is arguably a census).

(b)
Create Box Plots for the average weight of games by whether it is ranked as a Family Game or not. If `Family Game Rank` is blank (coded as `NaN`), then it is not ranked as a Family Game. I recommend adding a new column using the function `np.isnan`.

Are family games more or less complex? What can you say about the relative complexity of family games and non-family games?

(c)
Create a scatterplot of weight to average rating (`average`). What can you say about this relationship?

Problem 3: 50 points
Baseball Hall-of-Famers

Baseball Hall-of-Famers (HoF) played during different eras of baseball. One common classification of eras is '19th Century' (up to the 1900 season), 'Dead Ball' (1901–1919), 'Lively Ball' (1920–1941), 'Integration' (1942–1960), 'Expansion' (1961–1976), 'Free Agency' (1977–1993), and 'Long Ball' (after 1993). For this exercise, define the era of a player based on the mid-point of their career (rounding up if necessary).

Using the file `hofbatting.csv`, containing non-pitching HoFs as of 2013, classify each player according to their era to answer the following questions.

(a)
Create a Bar Graph for the number of HoFs from each era as of 2013. Interpret the data.

You will need to take multiple steps to solve this problem.

- Import the CSV.
- Create a column of data to define the mid-career.
- Find a way to count the number of HoFs in each era by the mid-career column.
- Create the graphs.
- Provide a written interpretation of the data.

(b)
Create a histogram showing the distribution of non-pitching HoFs' Mid-Career year.

(c)
There are two major dimensions to hitting: the ability to get on base (measured by the on-base percentage `OBP`) and the ability to advance runners already on base (measured by the slugging percentage `SLG`). Create a scatterplot of OBP vs. SLG. Are there any outliers? If so, identify them by name. Is there a relationship between OBP and SLG?

(d)
Consider a combined metric for hitting, the On-base Plus Slugging (`OPS`) statistic, which is the sum of OBP and SLG. Normalize this data (i.e., calculate the z-scores), then create a scatterplot with OPS on the y-axis and Mid-Career Year on the x-axis. Identify any outliers by name. Do you notice any patterns in the scatterplot of the data? What can you say (if anything) about the cause of any pattern?

(e)
Create a Box Plot for the Home-Run Rate (HRR), defined as home-runs per at-bat (HR/AB), of HoFs during each era (i.e., you should have 7 box-plots). Also calculate descriptive statistics of HRR including Min, Q1, Median, Q3, Max, Mean, Range, and Sample Standard Deviation for each era. Provide a table of these values from the Expansion era only (to limit time spent copying and pasting from Python).