

SOCI 385 - Fall 2023

Problem Set 1

This assignment is due via Canvas by 11:59 PM on Thursday, October 12, 2023.

Show all your code, order factor variable levels as appropriate, format your tables using Kable, write your responses in complete sentences, and label axes and plots. Submit your rendered pdf and the Quarto (.qmd) file for your notebook. Remember to include your name in the header. And remember that problem sets must be completed individually. No collaboration with other students is allowed.

This problem set requires two data files from a 2022 wave of the Pew Research Center's "American Trends Panel." Questions 1-4 use the `ps1.csv` file on Canvas. Question 5 uses the `ps1_long.csv` file on Canvas.

The `ps1.csv` file should have 11 variables and 5356 observations. Missing values are labeled NA. The variables are:

- `qkey`: a unique identifier for each respondent;
- `gender`: self-identification of gender, categorized as Woman, Man, or Other;
- `use_facebook`: binary variable noting whether the respondent reports using Facebook (1) or not using Facebook (0);
- `use_twitter`: binary variable noting whether the respondent reports using Twitter (1) or not using Twitter (0);
- `news_social_media`: categorical variable capturing how often the respondent reports using a social media site to get news headlines or information about events and issues that involve more than just their friends or family, with values of "Never", "Rarely", "Sometimes", or "Often";
- `age_cat`: respondent's age, categorized as "18-29", "30-49", "50-64", or "65+";
- `educ_cat`: respondent's education, categorized as "HS Grad or Less", "Some College", or "College Grad+";
- `racethn`: respondent's self-reported race or ethnicity, categorized as "Asian", "Black", "Hispanic", "Other", or "White";
- `news_trust`: categorical variable capturing how much, if at all, the respondent reports trusting the information they get from social media sites such as Facebook, Twitter, or YouTube, with values of "Not at all", "Not too much", "Some", or "A lot";
- `ideology`: categorical variable capturing respondent's self-reported political views, with values of "Very Conservative", "Conservative", "Moderate", "Liberal", or "Very Liberal";

- **misinformation:** which of these statements the respondent chose as coming closer to their own view, with 1 equal to “The U.S. government should take steps to restrict false information online, even if it limits people from freely publishing or accessing information” and 2 equal to “People’s freedom to publish and access information should be protected, even if it means false information can also be published.”

The `ps1_long.csv` file has multiple rows for each respondent. Each row corresponds to a different social media platform identified in the `site` column. The `use` variable takes the value of 0 if the respondent reports not using the site and takes the value of 1 if the respondent reports using the site. The `news` variable takes the value of 0 if the respondent reports not regularly getting news or headlines beyond information about family and friends from the site and takes the value of 1 if the respondent reports regularly getting such news or headlines from the site.

Questions

1. In a few sentences, describe any gender differences in who uses Facebook vs who uses Twitter. Consider the proportion self-identifying with each gender who report using the sites as well as the proportion of each site’s users self-identifying with each gender.
2. In a paragraph, describe the responses to the `news_social_media` question. What strikes you as sociologically interesting about the responses? What would you want to know more about? Your paragraph should include:
 - The overall proportion of respondents reporting each answer
 - The proportion of respondents in each age category reporting each answer
 - The proportion of respondents in each education category reporting each answer
3. In a few sentences, describe racial/ethnic differences in responses to the `news_trust` question. What is an additional variable (not in the dataset for this assignment) you would want to have in order to explain the relationship between these variables? Why might this additional variable be important? How would you phrase a survey question and possible responses to gather data about this additional variable?
4. Collapse the political ideology responses into three categories: any liberal, moderate, any conservative. In a few sentences, compare differences across these three new categories in responses to the misinformation question.
5. Use the `ps1_long.csv` file to create two figures. In the first figure, show the proportion of respondents using each social media site. In the second figure, show the proportion of users of each site who regularly get news from it. In a few sentences, describe the similarities and differences across these two figures. What do these figures suggest to you about social media as a source of news and information?