


Practice Assignment: Data Collection and Organization

Questions 1–6: Sometimes there are confrontations on planes because passengers disagree about flying etiquette.¹ *FiveThirtyEight* conducted a survey to ask air passengers which habits they consider to be rude.² The spreadsheet below shows part of the resulting dataset. (It is not necessary to download the full dataset to complete the questions below.)

	RespondentID	How often do you travel by plane?	How tall are you?	Is it rude to recline your seat on a plane?
1	3434278696	Once a year or less	6'3"	Yes, somewhat rude
2	3434275578	Once a year or less	5'8"	No, not rude at all
3	3434268208	Once a year or less	5'11"	No, not rude at all
4	3434250245	Once a month or less	5'7"	No, not rude at all
5	3434245875	Once a year or less	5'9"	No, not rude at all
6	3434235351	Once a month or less	6'2"	Yes, somewhat rude
7	3434218031	Once a year or less	6'0"	No, not rude at all
8	3434213681	Once a year or less	6'0"	No, not rude at all
9	3434172894	Once a year or less	5'6"	Yes, very rude
10	3434165659	Once a month or less	6'0"	No, not rude at all

- 1) Each row in the dataset represents one ____.
- observational unit
 - variable

Answer: a

- 2) Each column in the dataset represents one ____.
- observational unit
 - variable

Answer: b

¹ Hickey, W. (2014, September 6). *41 percent of fliers think you're rude if you recline your seat.* FiveThirtyEight. <https://fivethirtyeight.com/features/airplane-etiquette-recline-seat/>

² *Flying-etiquette.csv [Data file].* (2014, September 5). FiveThirtyEight. <https://github.com/fivethirtyeight/data/blob/master/flying-etiquette-survey/flying-etiquette.csv>

3) What are the observational units in this dataset?

- a) Airlines
- b) Air passengers
- c) Opinions about flying etiquette
- d) Flights

Answer: b

4) Which of the following characteristics are being treated as categorical in the dataset? There may be more than one correct answer.

- a) Frequency of airplane travel
- b) Height
- c) Opinion about whether reclining your seat is rude

Answer: a, c

5) Is Respondent ID a quantitative variable? Explain why or why not.

Answers will vary.

Sample answer: No; even though Respondent ID takes a numerical value, it wouldn't be used in arithmetic. For example, the average Respondent ID doesn't really make sense.

6) Why do you think the dataset is organized using Respondent IDs instead of the respondents' names?

Answers will vary.

Sample answer: The researchers probably assured the respondents that their personal information would be kept private. In fact, they may not have collected that information in the first place.

Questions 7 and 8: An *R* package called *nycflights13* provides information about all flights departing from New York City in the year 2013.³ The dataset has been used to explain which flights were delayed, and it includes information like scheduled departure and arrival times, actual departure and arrival times, airline carrier, amount of time spent in the air, and distance between airports.

³ Wickham, H. (2021, April 4). *nycflights13: Flights that departed NYC in 2013*. R package version 1.0.2. <https://CRAN.R-project.org/package=nycflights13>

Note: *R* is a language and environment for statistical computing and graphics. *R packages* contain *R* code, data, and documentation about data that can be used by users of *R*. You can learn more about *R* at <https://www.r-project.org/about.html>.

7) What are the observational units in the *nycflights13* dataset?

- a) Airlines
- b) Air passengers
- c) Opinions about flying etiquette
- d) Flights

Answer: d

8) Consider two ways of measuring arrival delay.

Part A: You could measure arrival delay by recording the amount of time between actual arrival time and scheduled arrival time (with positive numbers representing delays and negative numbers representing early arrivals). Is this variable categorical or quantitative? If the variable is quantitative, classify it as discrete or continuous.

Answer: Quantitative - Continuous

Part B: You could measure arrival delay by recording whether the actual arrival time is later than the scheduled arrival time (yes or no). Is this variable categorical or quantitative? If the variable is quantitative, classify it as discrete or continuous.

Answer: Categorical

