Populations and Bias

Each of the following scenarios involves collecting data to learn about a population. State (a) what population is involved, and (b) why the sample is biased. To demonstrate that a sample is biased, you must argue that certain members of the population are more or less likely to be sampled than others. Note: there will usually be many valid answers for parts (a) and (b), but your answer to part (b) will depend on how you define the population in part (a).

1. You need a survey on household spending patterns. You take a random sample from the customer list of the local brokerage firm.

2. You want to learn about New York City residents' sentiments (positive or negative) towards their mayor, Bill de Blasio. You search for "de Blasio" on Twitter and read the first 100 relevant search results.

3. You need to know the opinions of Langone students with regard to some curriculum matters. You ask some of the people in your class.

4. You want to learn about the quality of the food at a local restaurant. You read the reviews for the restaurant on Yelp.com.

5. You want to estimate the rate of growth of stocks over the last fifty years. You take a random sample of the stocks listed today on either the New York Stock Exchange or the Nasdaq. Some of these stocks did not exist fifty years ago; you set these aside. For the other stocks, you identify their prices fifty years ago, and you use this to compute the growth rate.

Types of Data

6. The class survey asked each respondent to report the following information: gender; birth date; GMAT score; undergraduate major; time spent studying per week; interest level in the course; industry; job type; number of dinners out per month; number of pairs of shoes; cups of coffee consumed per week; and number of websites visited per day.

   (a) Which of the variables measured by the survey are categorical/qualitative?
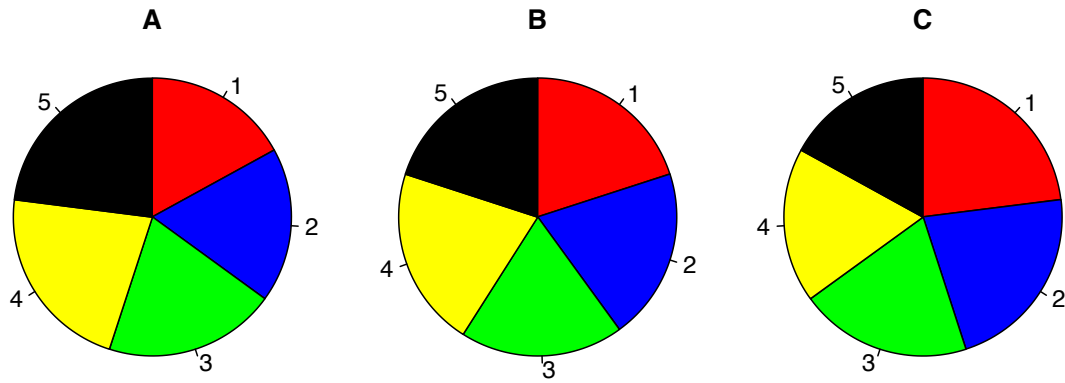
   (b) Which of the variables measured by the survey are numerical/quantitative?

7. What type of variable is the answer to the phone prompt "Enter '1' for English, '2' for Spanish."? Why?

8. Each Yelp restaurant includes a star rating (1–5). What type of variable is the star rating?

Describing Categorical (Qualitative) Data

9. Use the following pie charts to rank the categories (1–5) by size.

10. List two methods to describe the reported undergraduate majors of the class survey respondents.

11. Draw what you think the bar chart for the birth months of the survey respondents will look like.

Describing Numerical (Quantitative) Data

12. Draw what you think the histogram for "Websites Visited per Day" will look like.

13. Draw what you think the histogram for "Dinners per Month" will look like.

14. Draw what you think the histogram for "Interest in this Class" will look like.