

4.4 Exercises

Basic Concepts

1. Describe the purpose of data subsetting.
2. Describe a data set where data subsetting should be implemented. What are the disadvantages of not subsetting the data?
3. What is Simpson's paradox?

Exercises

4. Suppose you are a craft beer lover taking a trip to Denver on business and you want to be sure to stop at one of the local breweries while you are there. Using the Beers and Breweries data set from the companion website, subset the data to only show beers brewed in Denver, Colorado, and answer the following questions.
 - a. What level of measurement do each of the variables represent?
 - b. What variables other than *City* could be used to subset the data?
 - c. How many craft breweries are in Denver?
 - d. Which craft beer has the highest *Alcohol by Volume (ABV)* of the beers brewed in Denver? Give the name of the beer and the brewery.
 - e. For the Renegade Brewing Company, how many different IPA styles of beer do they make? What are they?
 - f. What is the mean and standard deviation of the *ABV* values for the craft beers made by the Wynkoop Brewing Company?
 - g. Compute the coefficient of variation of the *ABV* values for both the Renegade and Wynkoop breweries. Which brewery has more consistent *ABV* values?

5. Suppose you are looking for a house in Mount Pleasant, SC, which is near Charleston, and you have limited your search to three subdivisions: Park West, Dunes West, and Carolina Park. Using the Mount Pleasant Real Estate Data from the companion website, answer the following questions.
 - a. What level of measurement do each of the variables represent?
 - b. Which variables could be used to subset the data?
 - c. How could you subset the data using quantitative variables such as *List Price* and *Acreage*?
 - d. How many different house styles are represented in these three subdivisions? What are the styles?
 - e. How many of the houses were built in the years 2015 to 2017? Which subdivision has the most homes built during this timeframe?
 - f. What is the average price of homes built in the years 2015 to 2017 in Carolina Park? Round your answer to the nearest whole dollar.
 - g. For all homes built in the years 2015 to 2017 in the three subdivisions, what is the minimum and maximum priced homes and in which subdivision are they?
 - h. What is the price per square foot of the two homes in part g.?
 - i. What variables do you think may contribute to the high price of the house with the maximum price?

Technology

To learn how to perform calculations on a filtered data set, please visit stat.hawkeslearning.com and navigate to **Discovering Statistics and Data, Fourth Edition > Technology Instructions > Data Manipulation > Subset Calculations.**

Data

stat.hawkeslearning.com
Discovering Statistics and Data, Fourth Edition > Data Sets > Beers and Breweries



Data

stat.hawkeslearning.com
Discovering Statistics and Data, Fourth Edition > Data Sets > Mount Pleasant Real Estate Data