Show **all** of your work on this assignment and answer each question fully in the given context. Each individual part is worth 5 points and partial credit is awarded for close answers. Regardless of the total number of points on this homework, it will have the same weight as all other homeworks in terms of its impact on course grade. So, if a specific homework has 50 points and you lose 5, then your grade on that homework will be 45/50=90%. If there are 10 points on a homework and you lose 5, then your grade on that homework will be 5/10=50%. The average of those two homeworks would be $(90 + 50)/2 = 70$.
Please staple your assignment!

- **Chapter 1, Exercise 1 (page 23)**

- **Chapter 1, Exercise 9 (page 24)**

- **The Problems with Observational Studies**

  In class this week I mentioned that in an observational study, it is not possible to make conclusions about causation - meaning that even if we notice that every time "Event A" occurs that "Event B" will occur soon after, we are not able to say that Event A was the cause of Event B, or that if we stopped Event A from occuring this would stop Event B from occuring.

  Consider the following anecdote:

  > A marine biologist studying interactions between humans and non-elasmobranch ocean animals (meaning, "not sharks") has been keeping record of attacks by these animals on humans. As the amount of data built up over the course of several years, the biologist began to notice a relationship in the data: in almost every region, during periods where the average temperature was increasing there was an increase in the number of attacks reported in that region. Further, looking at regions where the average temperature was decreasing it was clear that there was a decrease in the number of attacks. Confident that the records of temperature and the records of attack were accurate, the scientist came to an incredible conclusion: animal attacks were driving increases in temperature.

  The scientist is probably wrong here, but hopefully this illustrates a point: just because two things can be observed to increase together or decrease together, that does not mean that we can say that one causes the other. In this case, it could be that rising temperatures cause people to spend time near the ocean, which increases the number of interactions between people and marine animals, which makes attack more frequent. In other words, I could use the same data (increaseing attacks related to increasing temperature) and make up the exact opposite cause-and-effect relationship. Still, even if my explanation seems more believable to you, it still doesn't have any support in the data - we don't have any information about the number of people visiting the ocean or interacting with animals. If the "temperature causes attacks" seems more believable, it's just because the things I assume (rising temperatures send people to the ocean) seem more believable to you. That's not rigorous science!

  Another problem with observational studies is that while we may record important data, we will also ignore important data. Consider the following:

  > I perform an observational study in which I observe the number of students crossing campus with an open umbrella and the number of students running across campus. My data indicates that when the number of students using an umbrella is low, the number

of students running is also low. I also observe that when the number of students using umbrellas is high, the number of students running is also high. I conclude that students using umbrellas are causing other students to run.

Hopefully, you can recognize that I may have ignored an important piece of information that might be the true cause of both the number of students using umbrellas and the number of students running: whether or not it was raining at the time. We call missing information that may be the real cause of our observations **Lurking Variables**.

Suggest a possible lurking variable in each case that might explain why the cause and effect relationship suggested could be wrong.

note: this question is fairly open-ended - as long as you can explain how the lurking variable you suggest could be driving both of the other variables you will get full credit. For example, in the umbrella/running example, a satisfactory response would be "The presence of rain may be a lurking variable. When it rains, people use umbrellas to stay dry and run between shelter if they don't have an umbrella. When it is not raining, neither umbrellas or running between classes are needed. Because of this, we would expect to see the number of people using umbrellas and the number of students running between classes increasing and decreasing at the same time, but another factor may be the actual cause.

a.) Researchers collecting data on outdoor temperature and the number of cases of common illnesses notice that when the temperature is lower more people get sick and when the temperature is higher fewer people get sick. They suggest that cold temperatures are causing these illnesses.

b. Researchers collecting data on work history notice that people are more likely to pass away in the five years after retirement than they are to pass away during any five year window of their careers. They conclude that retirement causes the person to pass away.

c. Over five years, the Coast Guard gathered monthly data for the number of swimsuit sales and the number of shark attacks. They noticed that when the number of swimsuits sold in a month was higher, the number of shark attacks in that month was also higher. They concluded that swimsuit sales were causing shark attacks.

- **Hockey game attendance.**

Caroline performs the following study to see if outside temperature has an effect on attendance at her college's hockey games. For each hockey game at her college, Caroline records the outside temperature and the attendance. Here are her results:

Day of Week		Temperature, deg. F	Attendance
Friday	12/14	35	840
Wednesday	12/19	20	560
Tuesday	1/ 8	−5	340
Friday	1/11	23	775
Wednesday	1/23	14	680
Saturday	2/ 2	30	950
Friday	2/ 8	28	950

1. Is this an experiment or observational study?
2. What type of variable is attendance?

Caroline analyzes her results and finds that outside temperature and attendance have a strong positive correlation (i.e., as one increases, the other also increases). She concludes that higher game day temperatures causes higher attendance at their college's hockey games.

1. Did she come to a proper conclusion for this study? Why or why not?
2. Look at the day of the week of the hockey games. What type of variable is this?
3. Rewrite the data table, adding a new column "School Night" (using the values "no" if the game is on a Friday or Saturday, and "yes" if the game is on any other day). How does Attendance relate to School Night?
4. For what type of studies do you have to worry about possible lurking variables affecting the results?

- **Washer stretching.**

  George works for a company that manufactures rubber washers. He randomly selects 1000 washers off the assembly line throughout two weeks for a study on the durability of these washers under stretching. To make sure that the washers are fit to be used in the real world, George must test the washers. Holding heat constant, George subjects each washer to one of various methods of stretching. The washers are randomly assigned to be stretched under one of five different forces (low, medium-low, medium, medium-high, and high). After each test, George classifies a washer as either defective or non-defective.

  1. Is this an experiment or observational study?
  2. What type of variable is heat?
  3. What type of variable is the amount of stretching?
  4. What type of variable is response to the stretching method?
  5. The 100 selected washers constitutes the sample. What is the population?
  6. George analyzes the results and finds that the defect rate increases with the amount of stretching. Can George conclude that the amount of stretching causes a change in the defect rate of the washers? Why or why not?