1. A researcher studied the medical records of 400 randomly selected patients. For each patient, she recorded whether the person took aspirin every day and if the person had ever had a heart attack. Then she reported the percentage of heart attacks for the patients who took aspirin every day and for those who did not take aspirin every day. What type of study did this researcher conduct?

   (a) **An observational study**

   (b) An experiment

   (c) A census

   (d) None of the above
   Explain your choice of answer in a sentence or two: *The study is observational as no treatments were applied to the subjects.*

2. A second researcher studied 400 patients that visited a regional hospital in the previous year. He randomly assigned half of the patients to take aspirin every day and the other half to take a placebo everyday. Then after two years he reported the percentage patients having heart attacks for those who took aspirin every day and for those who did not take aspirin every day. What type of study did this researcher conduct?

   (a) An observational study

   (b) **An experiment**

   (c) A census

   (d) None of the above
   Explain your choice of answer in a sentence or two: *The study is an experiment, as the patients were assigned (at random) to a treatment (either aspirin or a placebo).*

3. Suppose the researchers of the studies in Q1 and Q2 both wished to explore the following research question: Does taking aspirin daily reduce the risk of having a heart attack? Which of the following states the study or studies that could address the research question?

   (a) Only the study in Q1.

   (b) **Only the study in Q2.**

(c) Both of the studies in Q1 and Q2.

(d) Neither of the studies in Q1 nor Q2.
Explain your choice of answer in a sentence or two: *Only the experiment described in Q2 could remove concerns about confounding variables. For example, perhaps people who choose to take aspirin everyday have other health issues (such as, for example, alcohol problems) associated with heart disease.*

4. You are interested in researching the following question: In Coquitlam, do restaurants with drive–through service have a higher frequency of health code violations, on average, than restaurants without a drive–through service? Which of the following study designs would best help address this question?

(a) Select restaurants in Coquitlam at random and assign them at random into two groups (with and without drive–through service). Use government records to determine the number of health code violations over the past 12 months.

(b) **Select a random sample of restaurants in Coquitlam with drive–through service and a random sample of restaurants in Coquitlam without drive–through service. Use government records to determine the number of healt code violations at each restaurant over the past 12 months.**

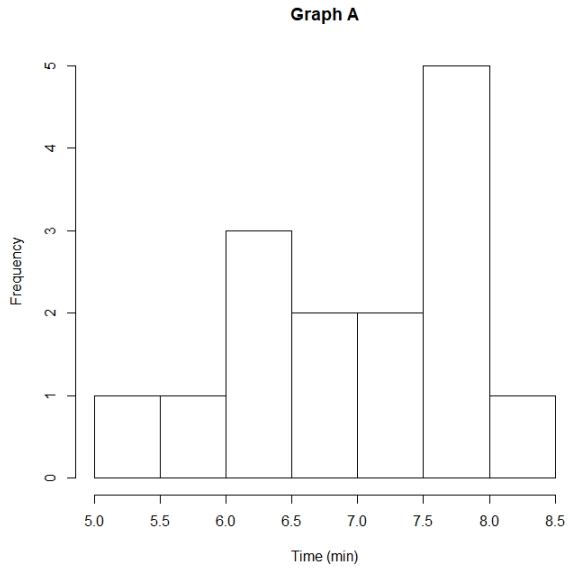
(c) Ask all Coquitlam restaurant owners to volunteer to complete a survey. Collect data on their responses to these questions: (1) Does your restaurant have drive–through service? (2) How many health code violations have you had over the past 12 months?

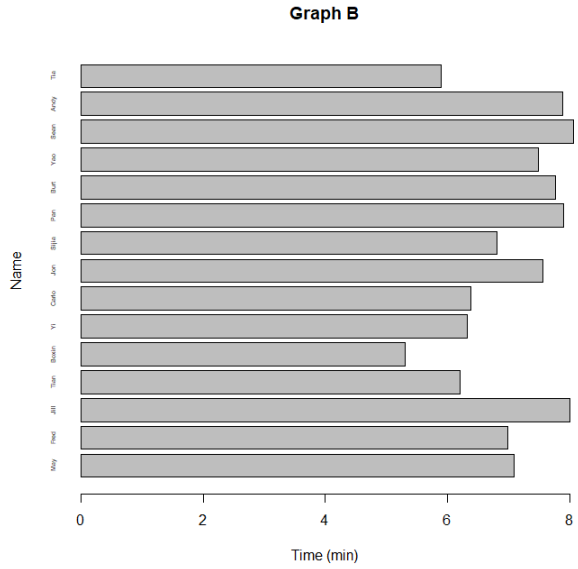
(d) Select Coquitlam restaurants at random and divide them into two groups at random. For one group, assign the restaurants to having a health code violation, while the other group you assign to having no health code violation. Compare the proportion of restaurants in each group having drive–through service.
Explain your choice of answer in a sentence or two, clarifying why you rejected an alternative choice: *In (b), two random samples are taken (and we could choose them to be proportionate to the respective population sizes) and an authoritative source is used to determine the number of health code violations for each. The*

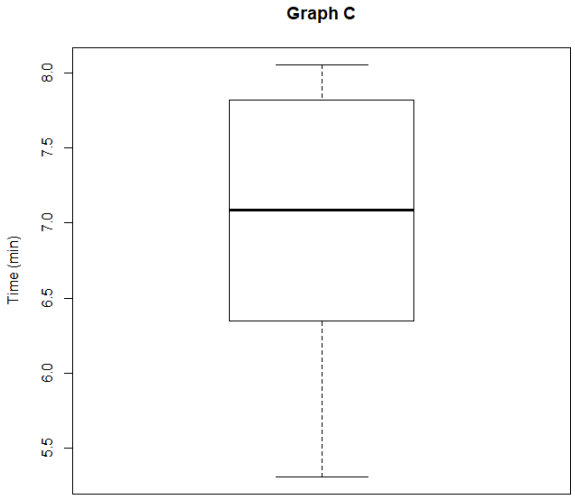
use of random samples will allow the results to be generalized to the populations of Coquitlam restaurants offering and not offering drive–through service. Option (a) implies we can assign restaurants to have a drive–through service or not; this is a characteristic of experiments and not observational studies. Option (c) is not correct because in general, a survey may be a reasonable choice for an observational study, though care must be taken to ensure that the respondents to the survey are representative of the population of interest. In this case, because generalization to all Coquitlam restaurants is desired, a volunteer survey will likely not result in an appropriate sample. As health code violations are undesirable, it is likely that respondents may not correctly report the number of violations they received, thus compromising the data in another way. Nonresponse would also be an issue, with owners of restaurants with health code violations being less likely to respond. Option (d) in nonsensical in that restaurants cannot be assigned to a treatment of health code violation status.

5. The athletics club at a high school record how long each of the members

take to run one mile. Below are three graphics of the data recorded:







Which of the above graphs most easily allows you to estimate the median running time?

(a) Graph A

(b) Graph B

(c) **Graph C**

(d) None of the graphs

(e) All of the graphs

Provide an estimate of the median time if you can, or otherwise explain why you cannot estimate the median: *The median appears to be about 7.1 minutes.*

6. For the data in the previous question, which graphic allows you to most easily see the shape of the distribution of running times?

(a) **Graph A**

(b) Graph B

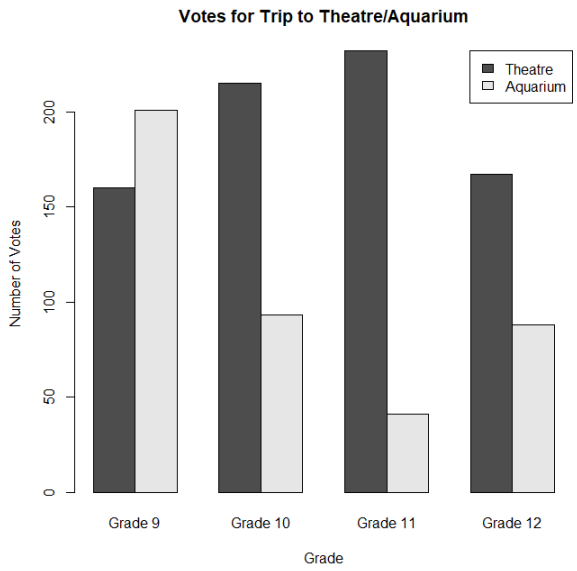
(c) Graph C

(d) None of the graphs

(e) All of the graphs

7. A high school is planning a field trip to the aquarium or to the theatre for students in grades 9 through 12. Students from each grade were encourage to vote for the trip they preferred. The bar graphs of votes

for the four grade levels are shown below.



In which grade level did the responses vary the most?

(a) 9

(b) 10

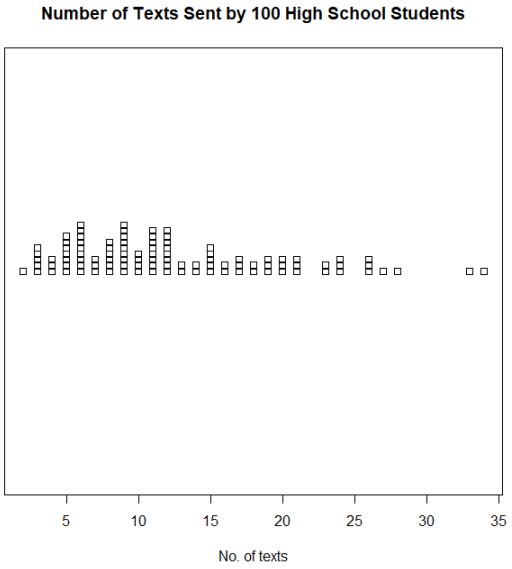
(c) 11

(d) 12

Explain your choice of answer in a sentence: *The greatest difference between the votes for theatre and aquarium are within the grade 11s. However, considered as numerical binary data (i.e., 0's and 1's), the data for grade 9s will have the highest variance.*

8. A sample of students at a high school was taken by using the first one hundred responses to an internet–based survey advertised by email. The students were asked how many text (or other internet–based) messages they had sent in the previous day. A dotplot of the counts is

shown below:

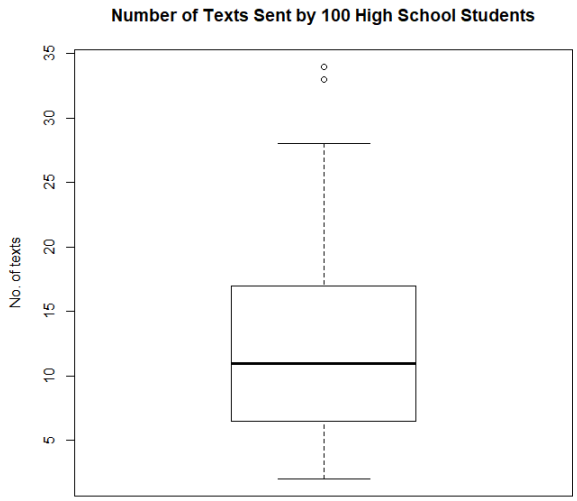


(a) What type of study is described above?
This is an observational study.

(b) What do you think is the target population for the study?
The target population is probably all students at the high school.

(c) Describe the data distribution you see above.
The distribution appears unimodal and right skewed.

(d) Below is a boxplot of the data.



Approximately what was the median number of texts (or other internet–based) messages sent by the students? Explain how you arrive at your answer.
The median appears to be about 11, as taken from the line in the box of the boxplot.

(e) Approximately what is the interquartile range for the counts? Explain how you arrive at your answer.
The lower quartile is about 7, the upper quartile about 17, judging by the limits on the box in the boxplot.

(f) Do there appear to be any outliers in the data? Explain your answer clearly.
Yes, the largest values appear to be outliers. These are above $Q_3 + 1.5IQR \approx 32$.

(g) Would you have any concerns about how the data were gathered with regards to the research question? Explain your answer clearly.
As invitation to participate was by email, students who regularly use email would be more likely to respond. This could bias the sample, as perhaps those responding would tend to differ in their use

of text and internet–based messaging compared to those students who do not use email as often.

(h) Suppose we wished to compare the distributions by sex (using data taken from the school records). What graphical method would you use to compare the data distributions?
We could use side–by–side boxplots (preferably) or back–to–back stemplots.