This exam is inspired by the Longitudinal Study of American Youth which can be found at [https://www.icpsr.umich.edu/web/NADAC/studies/37336](https://www.icpsr.umich.edu/web/NADAC/studies/37336). The Longitudinal Study of American Youth (LSAY) is a project that was first funded by the National Science Foundation in 1985 and was designed to examine the development of: (1) student attitudes towards and achievement in science and mathematics and (2) student interest in and plans for a career in science, mathematics, or engineering, during middle school, high school, and the first four years post-high school. The relative influence parents, home, teachers, school, peers, media, and selected informal learning experiences had on these developmental patterns were considered as well. In the original cohort 2,829 tenth grade students throughout the United States were followed for a period of seven years, beginning in 1987 and ending four years after high-school.

1. (a) What is the population of interest in the Longitudinal Study of American Youth (LSAY)?
   (b) Approximately 100 of the original 2,829 tenth graders did not participate in the follow up interviews conducted six years after the study began. How might this lack of participation influence the study findings?

2. The Longitudinal Study of American Youth (LSAY) includes information about hundreds of variables. Below is a list of some of the types of questions collected in the LSAY. For each question identify the relevant variable as either qualitative or quantitative. For the qualitative variables state if the variable is nominal, ordinal or binary. For the quantitative variables state if the variable is discrete or continuous.
   (a) How many math courses have you completed during high school so far?
   (b) Do you have access to the internet at home? (Yes, No)
   (c) How would you rate your abilities in mathematics? (Excellent, Very Good, Good, Moderate, Weak, Poor)
   (d) How many adults in your household work full time jobs?
   (e) What is your favorite subject in school?
   (f) The quality of science and mathematics education in the United States is inadequate. Do you strongly agree, agree, disagree, or strongly disagree with this statement?
   (g) How long does it take you commute from home to school?
   (h) What was your score on your last math test?
3. For a class, your friend Damian was asked to create a data visualization based on the LSAY. Below is the visualization Damian created.

Damian needs your advice for improving this visualization before he submits the assignment.

(a) Identify problems with his visualization. Use the language of poor visualization practices discussed in class.
(b) Give three concrete suggestions for improvement. Be as specific as possible.
4. In class we discussed several common statistical misunderstandings including:

- Mistakes Interpreting Correlation
- Cherry Picking Data
- Survivorship Bias
- Sampling Bias
- Simpson’s Paradox
- The Prosecutor’s Fallacy
- Probable Coincidences

Each of the following is an example of a potential erroneous conclusion based on data from the LSAY. **Pick 3** of the conclusions below and identify which statistical misunderstanding is involved and provide a short explanation of the error in interpretation.

(a) Two of the respondents shared the same birthday and have mothers named Susan.

**Conclusion:** The students must have been lying when they reported their birthdays and mothers’ names! It seems impossibly unlikely that two students who both have moms named Susan with the exact same birthday end up in the sample; it must be a lie.

(b) The proportion of underclassmen (9th and 10th graders) who reported feeling adequately prepared for a 4-year university was larger than the proportion of underclassmen who reported feeling adequately prepared for a full time job. The same was true for upperclassmen (11th and 12th grade): The proportion of upperclassmen who reported feeling adequately prepared for a 4-year college was larger than the proportion of upperclassmen who reported feeling adequately prepared for a full time job.

**Conclusion:** Since the pattern holds for both the underclassmen and the upperclassmen, we can conclude that when considered together the proportion of high school students feeling adequately prepared for a 4-year college is larger than the proportion of high school students who reported feeling adequately prepared for a full time job.

(c) Researchers found no correlation between parental income and student happiness.

**Conclusion:** There is no relationship between parental income and student happiness.

(d) Nearly all of the respondents who participated in the follow up study graduated from high school.

**Conclusion:** Nearly all 10th graders will go on to graduate from high school.

(e) One student reported failing science class in the 9th grade. Over the summer he got a science tutor and earned an A in his 10th grade science class.

**Conclusion:** Tutoring will get you an A in science class, even if you previously failed a science class.

(f) Over 70% of the students who reported wanting careers in the sciences had taken a physics class in their high school.

**Conclusion:** If you enter a high physics class you can expect around 70% of the students in the class to want careers in the sciences.