Problem 1. A study compared percent of internet users and life expectancy at birth in 208 countries and found a positive association. What is a possible confounding variable that might explain this relationship?

Problem 2. Here is some summary data about the success/failure of space launches by different types of launching agencies.

	1957–1999		2000–2018	
	Failure	Success	Failure	Success
Private	13	295	10	562
State	281	3751	33	711
Startup	-	-	5	65

Describe a data frame that one might use to create the above summary table. How many variables are involved? What type is each variable? What is the total number of observations in the data frame? How many rows and columns?

Problem 3. You would like to determine if your classmates prefer the taste of regular Coke or Diet Coke. You have $50 to spend. Here are some potentially relevant costs:

- A 1 L bottle of Coke or Diet Coke costs $2.
- You must pay each participant in your study at least $5.
- You must pay anyone helping you administer the study at least minimum wage ($12.56/hour in CO).

Outline a design for this study. Can you collect data for all of your classmates? If not, what sampling strategy will you use? How will you best eliminate biases?

Problem 4. In 1936, Landon sought the Republican presidential nomination opposing the re-election of FDR. The Literary Digest polled about 10 million Americans and got responses from about 2.4 million. The magazine had surveyed its own readers, registered automobile owners, and registered telephone users. The poll showed that Landon would likely be the overwhelming winner and FDR would get only 43% of the votes.

Election result: FDR won, with 62% of the votes.

The magazine was completely discredited because of the poll and was soon discontinued. What might have gone wrong?