WS #1 - DS problems
DS 002R, Jo Hardin

Monday, August 25, 2025

Your Name: _____

Names of people you worked with: _____

Which piece of the material for DS 002R seems most daunting? Which piece seems most fun?

Task:

1. What problem or task would you like to investigate using data?
2. What would be hard about executing the project?
3. What are the potential ethical frameworks to consider?
4. How would you define success for the project?

Solution:

There is no real *solution* to this task. But some things to consider include:

1. Ideas: so many great ideas! The beauty of data science is that it touches every discipline.
2. Difficulties: how will you store the data? how will you clean the data? how will you analyze the data?
3. Ethics: who is benefiting from your study? Who is being left out?
4. Success: oftentimes success is measured through some kind of metric (e.g., accuracy). But we need to be careful with our definition of success because it doesn't always form to the problem at hand. For example, consider an email filter whose success mechanism is accuracy. But let's also say that you don't get much SPAM email, maybe 2%. So if your filter decides that **everything** is not SPAM email (i.e., it doesn't even use a model), then it will have a 98% accuracty rate.