Name:

**Lab 1 (Chapters 1-4 Review): Mueller & Dweck**

Using the same study that we looked at in Classwork 5, complete this lab assignment.

Here is the code to get the data:

**PRAISESTUDY <- read.csv("http://bit.ly/muellerdweck\_study1", header=TRUE)**

As a reminder: Mueller and Dweck did a study with kids. First, they gave the kids a set of 10 easy puzzle questions (of the type commonly used on IQ tests). That set will be called **PS1** (easy problems). Then the kids received one of three types of **FEEDBACK** (Control, Intelligence Praise, or Effort Praise). Then they gave the kids a set of 10 very difficult puzzle questions (called **PS2**). Lastly, they gave them a set of easy questions again at the end (called **PS3**).

1. What is the observational unit in this data frame? Could we put the average school’s PS3 score into this data frame? Why or why not?
2. Were there the same number of kids in each of these three conditions? Write the code you would use and answer the question here.
3. Let’s say we want to create a new variable called **PSDIFF** in this **PRAISESTUDY** data frame. We want to figure out how many more problems a kid solved in **PS3** than in **PS1** and save that number as a variable called **PSDIFF**. Write code to create such a variable in R.
4. Create a visualization to look at the variation in **PSDIFF** in R. Write the code here and sketch the visualization here. What are some things you notice about the distribution (think: shape, center, spread, and weird things).
   1. There were only 11 possible answers (since there were 10 problems and it’s possible that a kid got them all wrong). Can you modify your code so that there are only 11 bins in this histogram?
   2. Most people wouldn’t know what **PSDIFF** means. Add code to put a better label on the x-axis.
   3. Add the code to also add a density plot to this histogram. (Note: To add a density plot, you may need to plot density on the y-axis instead of count.)
5. Let’s say we wanted to write a word equation to explain the variation in **PSDIFF**. Maybe **FEEDBACK** is an important explanatory variable. What would the word equation look like?
6. What percentage of kids improved over the course of the study (answered more questions right on **PS3** than **PS1**)? Write R code that would give you the answer to this question as well as the percentage.
7. If you look at the histogram for **PSDIFF**, you’ll see there is just one kid who got 6 more correct on the third set of problems than the first set. What condition was that kid in?
8. How would you create a plot to explore whether **FEEDBACK** explains some of the variation in **PSDIFF**? Write the code here.
9. Explain what you would look for in your plot that would show that **FEEDBACK** does indeed explain some of the variation in **PSDIFF**.
10. Mueller & Dweck also asked kids to rate how much they enjoyed doing these puzzles. A researcher has an intuition that how much the kid enjoys these puzzles (**ENJOY**) is associated with **PSDIFF**. Which visualization will help use explore this idea? Write the code to produce such a visualization below.
11. Although there are a number of visualizations to choose from, a faceted histogram is not a great way to explore whether **ENJOY** explains some of the variation in **PSDIFF**. Explain why this is.
12. Write a word equation to represent the idea that enjoyment of puzzles (**ENJOY**) might predict **PSDIFF**.
13. Explain what you would look for in your plot that would show that **ENJOY** does indeed explain some of the variation in **PSDIFF**.
14. Is that what you see in the plot when you actually make it? What do you think about this hypothesis? (This question is not graded correct/incorrect. I just want you to think it through and make an argument.)
15. Children were also asked what kind of problems they liked to do. They were asked to choose from four types of problems: (a) *problems that aren't too hard, so I don't get many wrong*; (b) *problems that I'll learn a lot from, even if I won't look so smart*; (c) *problems that are pretty easy, so I'll do well*; (d) *problems that I'm pretty good at, so I can show that I'm smart*. The children’s choices are recorded in the variable called **GOAL** with the letters *a, b, c, d*. Recode this variable so that instead of *a, b, c, d* the values are *not hard, learning, easy, good at.* Write the code for that here.
16. Three of the goals (a, c, and d) are what researchers call “performance goal” and only b is a “learning goal.” Researchers categorized whether students exhibited a performance goal or learning goal in the variable **GOALCODE**. Performance goal is given a value of 0 and learning goal has a value of 1. Label the variable **GOALCODE** with the labels *performance* and *learning* instead of 0 and 1. Write the R code here.
17. Notice that 15 and 16 are similar tasks but used different functions (**recode** versus **factor**). Why do we use different functions in these two situations? (What’s different about the situations?)

1. What percentage of kids in the Intelligence condition chose learning goals? What percentage of kids in the Effort condition chose learning goals?
2. Now that you have explored the PRAISESTUDY data, write a research question that you are wondering about (maybe even one that could be answered with this data set!).
3. Using the PRAISESTUDY data, write a quiz question from chapter 1.
4. Using the PRAISESTUDY data, write a quiz question from chapter 2.
5. Using the PRAISESTUDY data, write a quiz question from chapter 3.
6. Using the PRAISESTUDY data, write a quiz question from chapter 4.