­Name:

**Classwork 6**

At least until the quiz (later today), we’ll use the same data set that we looked at in the lab assignment.

Here is the code to get the data:

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

**PRAISESTUDY$PSDIFF <- PRAISESTUDY$PS3 - PRAISESTUDY$PS1**

1. Let’s say we want to create boxplots to look at whether the variation in **PSDIFF** is explained by **FEEDBACK**. Write code to do this.
2. Explain what you would look for in your boxplots that would show that **FEEDBACK** does indeed explain some of the variation in **PSDIFF**.
3. If we shuffled **PSDIFF** in the boxplots, what would most of those shuffled boxplots look like?
4. Usually someone’s past performance is a pretty good indicator of their future performance. Do you think the kids who did pretty well on **PS1** would also do well on **PS3**? How would we write this as a word equation?
5. What visualization should we make? Write the code here.
6. If you count the dots, there won’t be 123 observations (even though we have data from 123 kids). Why? How can we improve our visualization?
7. Describe the general pattern you see in the scatterplot. Is that pattern true for every kid?
8. If we shuffled one of the variables (either PS1 or PS3) – how would that change the scatterplot?
9. What do you think is the likelihood of getting a pattern like the real one we got: High? Medium? Low? Why?