ASSIGNMENT 2
FEATURE ENGINEERING STUDIO
PROFESSOR RYAN S.J.d. BAKER
DATA FAMILIARIZATION
"MUCKING AROUND"
DUE IN CLASS, MONDAY SEPTEMBER 23

The goal of this assignment is to get to know your data better!

Step 1: Get your data

- If you can't get your data, email me at least 48 hours before class, and I'll send you a practice data set

Step 2: Select your ground truth label and a small set of other variables

Step 3: For numerical variables

- What is its max, min, average, stdev?
- Create a histogram

Step 4: For categorical variables

- What is the distribution of each value

Step 5: Write a brief report for me

- And be prepared to discuss your report in class