**Step 1 – part 1 (Wednesday Oct 6th)**

**As much as you can for Wednesday**

**1. Introduce your data:**

**What is the population of interest?**

**What kind of variables you have?**

**Specify the type of variables.**

**2. Data cleaning and manipulation:**

**Explain any data cleaning**

**Explain any data manipulation (changing structure, creating new variables,…)**

**(Final variables)**

**3. Research question**

**What is the question?**

**Why is this interesting?**

**Why logistic regression?**

**Step 1 – part 2: At least test on few variables for Wednesday**

**4. Selecting variables:**

**Test of significance: Binary response**

**Numerical Variables: Two sample T-test: (To compare the means between two populations)**

**t.test(group1, group2, var.equal=T)**

**t.test(group1, group2, var.equal=F)**

**Variance Test: var.test(group1, group2) Use the p-value to decide what kind of T-test you need to use.**

**Categorical Variables: Chi-square Test: (To determine if there is an association between explanatory & response)**

**count <- matrix(c(a,b,c,d), nrow=2)**

**chisq.test(count)**