Milestone 3
Due Sunday, 11:59pm

There are **three** components to your submission:
1. R script that does the following:
   a. Reads in your data
   b. Cleans and describes your 10 variables (see template below).
2. Summary tables describing your ten variables (see template below)
3. 300 word writeup describing the patterns in your variables based on your summary table:
   a. Quantitative variables: describe the distribution, missing values
   b. Categorical variables: describe how you collapsed levels, describe the distribution, missing values
   c. Based on what you've found so far, what exploratory analysis would you like to do with this data?

For each categorical variable, you should copy the following template into your R script:

###name of variable (categorical)

#step 1: Convert to factor variable

#step 2: Show levels of the factor variable

#step 3: Reduce the number of levels (no more than 4-5 levels) by either removing levels when there is no valid answer given or combining levels

#step 4: Create a frequency table of the levels

For each quantitative variable, you should copy the following template into your R script:

#name of variable (quantitative)

##step 1: convert to numeric, date, or logical (whichever is the most appropriate). You can create additional variables if you have a date/time object (time of day, day of week, etc).

step 2: create a summary table of the variable (min, max, mean, median, n)

Template for other types of variables:
Geography (county, state) or names: report the number of levels instead of all the levels

Summary Table (Quantitative Variables)

variable name	N missing	mean	sd	min	p50	max
bride_age	245	30.5	6.12	20	29	90
groom_age	352	32.6	7.52	16	31	92

Summary Table (Categorical Variables)

Variable	Levels	Frequency
Race	White	1204
	Black	230
	Other	456
	NA	203
Sex	Male	546
	Female	630
	NA	0

Summary Table (geography, names)

Variable	nLevels
State	52
County	200
Instructor Name	4,500