Charlotte Is Creative Data Cleaning and Analysis

**Incentive Drawing-Bridget will do this**

1. Remove duplicate values
2. Assign random rank order
3. Select five lowest values

**Main Survey: Data Cleaning**

1. Exclusion/Inclusion
   1. Exclude participants who answer “No” to Q1 and/or Q2
   2. Exclude participants who list zip codes outside of North or South Carolina. If no zip code is listed, participant should be included if they meet other inclusion criteria. If a zip code appears to be mistyped but first three digits suggest that it is a NC or SC zip code (e.g. 282088), keep in survey.
      1. Let’s do two options for including/excluding zip codes:
         1. Including NC and SC zip codes (entries with no zip codes and zip codes indicating NC/SC are included)
         2. Include only MSA counties (list zip codes)- Bridget/Sydney to get list
2. Recoding
   1. Recode dichotomous variables to binary scale
   2. Recode all multiple response questions into single variable (e.g. 3\_1 to 3\_15)
   3. Recode race and ethnicity variables
      1. Start with race variable
         1. Include all single races with more than 5
            1. Single races categories with less than 5 can be classified as “Other race- single race”
         2. Include multiple race categories of more than 5
            1. Multiple race categories with less than 5 can be classified as “Other race- multiple races”
      2. Then add ethnicity
         1. Race/Ethnicity for each category
   4. Recode years of experience in creative industry (0-2,3-5, 6-10, above 10)
   5. Recode household income, combining variables Q48 and Q51 such that if Q49 is “Yes,” use value from Q48. If Q49 is “No,” use value from Q51
   6. Create a composite score for Qs 35 and 36 using the Abbreviated Survey scoring instructions at this [link.](https://files.consumerfinance.gov/f/201512_cfpb_financial-well-being-user-guide-scale.pdf) Please use the categories 18-64 and 65+ in place of the categories in the instructions. For individuals who did not list an age, provide the range indicated in the scoring instructions. Participants who “preferred not to respond” to any of the five questions needed for composite scoring should be excluded.

**Main Survey: Analysis**

1. Frequencies
   1. Create table of frequencies with data visualization for each survey question.
   2. For demographic variables, show frequencies for original and recoded variables
   3. NOTE: some questions were only asked to a subset of survey participants. Look at Qualtrics survey instrument to confirm which questions this applies to and make a note under frequency tables.
2. Crosstabs
   1. Use recoded versions of variables for crosstabs
   2. Create Crosstabs for all close-ended survey questions with the following demographic variables:
      1. Age, gender, born in Charlotte, years of experience, race/ethnicity, personal income, household income, Q4 (artistic discipline)
      2. Values less than 5 should not be reported
3. Share initial findings with Sydney and Bridget, discuss areas of interest that should be analyzed further

For discussion next week:

1. How do we want to recode race and ethnicity for analysis purposes? Combine, keep separate?
2. Do we want to do crosstabs for all variables and demographics, or just those of highest interest?