**ANALYSIS FILES/STEPS:**

1. Data cleaning code (data-cleaning-code)
   1. Loaded individual datasets from each coder (measurement data only)
   2. Deleted one scan that was coded by two coders (Jared and Kayna, kept Kayna’s)
   3. Added “coder” column with name of each coder to each individual dataset
   4. Compiled all individual datasets into one dataset
   5. Renamed columns with abbreviations of measurements (Appendix A? of dissertation)
   6. Imported demographic dataset
   7. Recoded race/ethnicity with abbreviations in Table 3.3? of dissertation
   8. Added age group column
   9. Set and double checked variable classifications (numeric vs char vs factor)
   10. ***Created headscan\_full excel***
       1. ***All participants, all measurement values***
       2. ***Includes visual and statistic outliers, and NA values***
       3. ***All demographic categories are retained***
2. Demographic data exploration (demographic-data-exploration)
   1. Loaded headscan\_full
   2. Numeric age info:
      1. Table of sumstats
      2. Histogram of age frequency
   3. Race/ethnicity (all categories retained) info:
      1. Table of sumstats
      2. Boxplot of age by race/eth category
      3. Histogram of race/eth category frequencies
   4. Gender (all categories retained) info:
      1. Table of sumstats
      2. Boxplot of age by gender category
      3. Histogram of race/eth category frequencies
   5. Age group info:
      1. Table of sumstats
      2. Boxplot of age (num) by age group
      3. Histogram of age group frequencies
   6. Headscan sample demographics plot of age, sorted by race/eth, color by gender
   7. Couple of other plots (not that great)
3. Summary stats SA1 (summary-stats-SA1)
   1. Loaded headscan\_full
   2. Transformed all measurement data from centimeters to millimeters
   3. ***Created measureNAs excel***
      1. ***Sum of NA values per measurement***
      2. ***Proportion of NA values per measurement***
   4. Measurement NA info:
      1. Table of proportion and sum of na values per each measurement location (CREATE NEW TABLE IN ORDER OF NA)
      2. Histogram of missing value ordered by measurement location (one for count, one for prop)
   5. Overall, race/eth, gender, age sumstats for all measurements!
      1. Histogram of measurement frequency
      2. Table of overall measurement sumstats
      3. Table of all race/eth categories sumstats
      4. Boxplots (ordered and not) of measurement by all race/eth categories
      5. Table of all gender categories sumstats
      6. Boxplots (ordered and not) of measurement by all gender categories
      7. Table of all age group categories sumstats
      8. Boxplots (ordered and not) of measurement by age group categories
   6. Removed visual outliers and redeveloped all of above for following measurements:
      1. GoSub\_C
      2. NRB\_L
      3. SelDH\_C
      4. SnasM\_C
      5. TrGo\_C
      6. TrSman\_C
      7. TrSnas\_C
4. Removing visual outliers (removing-visual-outliers)
   1. For all 7 above measures:
      1. Looking at histogram with all data
      2. Checking max value
      3. Change values close to max (visual outliers) to NA
      4. Checking how many values were filtered out
      5. Histogram of filtered data
   2. ***Created headscan\_full1 excel***
      1. ***Same as headscan\_full, but 26 individual measurement values changed to NA based on visual outlier status***
   3. ***Created all\_vis\_out1 excel***
      1. ***2 columns:***
         1. ***ID: id number***
         2. ***Vis\_out: what was the measurement that was changed to NA in headscan\_full1 dataset***

**EVERYTHING BELOW HERE: NEEDS TO BE REKNIT!!!!**

1. Bivariate panel
   1. Used headscan\_full1 data, where visual outliers were individually changed to NA values (not entire row)
   2. If IDS for visual and stat outliers overlap, can keep this as-is
   3. One point that looks like an outlier (80mm Sellion to menton/face length) – try to label and compare to visual and stat outlier table (after step 7)
2. Removing statistical outliers
   1. NEXT STEPS: create dataset
3. compare these outliers and visual outliers (NEW FILE NEEDED)
   1. make table comparing IDs and measurements
   2. Consider: bivariate panel can have na values, PCA cannot