* **Specific Aim 1:** Characterize face dimensions demographic groups of gender (Male, Female, Prefer not to say or Other), race (White, Black, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Prefer not to say or Other), ethnicity (LatinX), and age (18-29, 30-49, 50+).
  + Summary statistics (see examples in ANSUR II (Gordon et al., 2014))
  + Important for industry, gives measurements in cm terms
    - Mean
    - Standard deviation
    - Min
    - Max
    - Chosen percentiles (z-score)
* **Specific Aim 2:** Discover components that define measurement variability in dataset participants
  + Principal Component Analyses (will tell us what variables inform principal components that define the data, and provide visualization)
    - Between groups (overall, using all samples)
      * May require multiple iterations
        + Dropping NA values by participant (row)
        + Dropping NA values by measure (column)
        + Understanding if these NA values are important via correlation

ie if chin is correlated with another non-occluded measurement…

* + - Within groups
      * Gender (M, F, Prefer not to say or Other)
      * Race (White, Black, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Prefer not to say or Other)
      * Ethnicity (LatinX)
      * Age (18-29, 30-49, 50+)
    - Online tutorials of PCA in R;
      * give idea of outcome in proposal
        + plots: utilize color to show multiple iterations of same plots (PC1 an PC2)
      * how to interpret
    - COULD HELP INFORM MANOVA!! To avoid including too many measurements or too many demographic groups
      * Could use PC’s themselves in MANOVA
* **Specific Aim 3:** Quantify demographic sources of variation within the data
  + MANOVA analysis (two-way)
    - MANOVA tells us whether multiple levels of independent variables on their own or in combo have an effect on the dependent variables
      * To what degree do demographic variables have an effect on measurement variables?
        + And, do interactions of demographic variables have an effect on measurement variables?
        + Reference groups: Caucasian, male, young (based on LANL face panel)

Statistical power is higher when reference group n is high

(another justification for this reference group)

* + - May drop age from certain demographics from analysis (i.e., age) if above analyses do not indicate differences.
      * Including too many variables can create too many interactions.
      * Measurements that are highly correlated may create problems (collinearity)
    - MANOVA will significantly quantify sources of facial anthropometric variation in our sample
  + ASSUMPTIONS:
    - Outliers: remove or transform the data
    - Sample size: 20 observations for each group
    - Independence of observations: yes, each person was randomly sampled
    - Normality: all dependent variables must follow a multivariate normal distribution in each group
      * Linear combinations should also be normally distributed
    - Homogeneity of covariance matrices: population covariance matrices for each measurement in each group must be equal
      * Box’s test: if significant, assumption is violated
* **Specific Aim 4 AS NECCESARY:** Further quantify sources of variation in data using post-hoc analyses
  + Univariate ANOVAs (one-way and/or two-way, as needed)
    - Determine what individual measurement variables differ between demographic variables (one-way) or interactions between demographic variables (two-way).
    - Depending on Specific Aim 3 findings, the scale of these post-hoc analyses could be small or quite large
      * Will code for maximum number of post-hoc ANOVA analyses and remove those not needed