* Remove missing data
* Plot histogram of age for relevant data (20k after NA removal) ~ age
* Plot histogram of ethnicity for relevant data (20k after NA removal) ~ age
* Test the hypothesis over race as well
* Separate the GSP files analysis for now
* Remove the diagnostics columns but keep them for later analysis
* If the minimal is 0 add the first non-zero distance to all observations

Experiment

1. Take a feature with relative high minimum
2. Apply log transformation – draw histogram
3. Remove 9/10 of the distance to zero
4. Repeat 2 for the moved distribution
5. Which value of *k* yields the most “normal” distribution

Mean FA – column 44 -91 include

Mean MD – column 92 – 139

Track data:

FA 140-166

MD 176 – 193

Volume 194 – 332

Compare 29 to 290

* KS test for normal and log-normal distributions
* Test for one distribution assumption – test the one distribution model vs the mixture model
* Repeat the analysis after brain size correction

Report

* Compute the sd of the log normal distribution (after log) – population, men, women:
  + *T-test* for mean similarity *Cohen’s d*
  + for single distribution vs mixture
  + Pure types vs Mixture
* Add proportion for each distribution – men and women
* In the cases of normal beats log-normal – what is the underline distribution (make sure that the log “fixes” the outliers)