**WTF am I doing**

**Bootstrapping**

* Bootstrap 2014 data to create fake 2015 data to look at
* I want to generate resident fecal N data and migratory fecal N data separately to hopefully maybe detect a difference
  + Just one statistic, split into 2 groups? Or treat as 2 separate statistics?
  + # desired replicates
    - (2res + 2mig) \* 12 sampling pds
* General bootstrapping notes
  + Nonparametric (not making assumptions about normality etc.)
  + Resamples dataset a given # of times
  + ~~boot(data= , statistic= , R= )~~
    - ~~data=df; statistic=FUNCTION YOU HAVE TO WRITE YOURSELF IN ADVANCE, R=#reps~~
      * ~~e.g. function to calculate means of fecal N, which I’ve called fn~~

~~the.means <- function(fn, i) {mean(fn[i])}~~

bootdata <- sample(1:nrow(data2014), SAMPLESIZEFILLTHISIN, replace = T)

newdata <- data2014[bootdata,]

this creates bootstrapped data that’s created by just randomly selecting entire rows from the 2014 data, replacing and reselecting, however many times I tell it to

**Prepping data for power analysis**

* I need to:
  + Split residents and migrants, then bootstrap each individually
  + Generate values for dates before June 18 – currently missing 3 sampling periods
  + Average resident and migrant fn values per sampling period
  + Compare averages – how?
    - Power.anova.test
      * Need variance for each group and between both groups
* Misc notes
  + This will be a preliminary first step just determining whether I have power to detect an overall difference in nutrition across the entire growing period – NOT temporally
    - If no difference can be detected, no need to see if I can detect temporal difference
* Steps to complete
  + Dplyr – subset by migratory status
    - Also by sampling period??
  + Bootstrap each set separately
    - NEED: 12 sampling periods, each with 2-3 values for residents and for migrants – call it 2 for simplicity’s sake
    - For residents and migrants separately:
      * 2 vals \* 12 sampling periods – 24 total numbers
        + Average sets of 2 to create final df of just 12 values
  + Average each set per sampling period
  + Average all sampling periods for each set
  + Compare averages
    - Create dataframe of the two means, concatenated
    - var(groupmeans) gives variance
    - I assume var(resmean) would give variance of resident means