load("acs2017\_ny\_data.RData")

#glimpse(acs2017\_ny) try this later

acs2017\_ny[1:10,1:7]

attach(acs2017\_ny)

summary(acs2017\_ny)

print(NN\_obs <- length(AGE))

summary(AGE[female == 1])

summary(AGE[!female])

sd(AGE[female == 1])

mean(AGE[!female])

sd(AGE[!female])

hist(AGE[(AGE > 90)])

mean(AGE[ (female == 1) & (AGE<90) ])

#Top coding would push the avg age up by grouping more people

#into the top coded age thus moving up the average instead of

#people 92 and older

str(as.numeric(PUMA))

mean(PUMA)

PUMA <- as.factor(PUMA)

female <- as.factor(female)

print(levels(female))

levels(female) <- c("male","female")

educ\_indx <- factor((educ\_nohs + 2\*educ\_hs + 3\*educ\_somecoll + 4\*educ\_college + 5\*educ\_advdeg), levels=c(1,2,3,4,5),labels = c("No HS","HS","SmColl","Bach","Adv"))

install.packages("tidyverse")

install.packages("plyr")

library(tidyverse)

levels\_n <- read.csv("PUMA\_levels.csv")

levels\_orig <- levels(PUMA)

levels\_new <- join(data.frame(levels\_orig),data.frame(levels\_n))

levels(PUMA) <- levels\_new$New\_Level

summary(female)

summary(PUMA)

summary(educ\_indx)

ddply(acs2017\_ny, .(PUMA), summarize, mean = round(mean(AGE), 2), sd = round(sd(AGE), 2), n\_obsv = length(PUMA))

#Errors- no acs2017\_ny, found? or ddply which also happens below in line 37

dat\_use1 <- subset(acs2017\_ny,((INCWAGE > 0) & in\_NYC))

ddply(dat\_use1, .(PUMA), summarize, inc90 = quantile(INCWAGE,probs = 0.9), inc10 = quantile(INCWAGE,probs = 0.1), n\_obs = length(INCWAGE))

table(educ\_indx,female)

xtabs(~educ\_indx + female)

prop.table(table(educ\_indx,female))

#test using table with PUMA

prop.table(table(educ\_indx, PUMA))

#it appears by PUMA code a breakdown of all education levels and the percntage of

#of im guessing NYC population as a whole? Maybe? It would be easier to understand

#if rounded to the first two significant digits.

#This data includes not just whether a person has a college degree but also what field was the

#degree in: Economics or Psychology, for instance. Look over the codebook about DEGFIELD and DEGFIELDD

#(that second D means more detail) to see the codes. Maybe look at 10th and 90th percentiles by degree field?

###IF I OPEN THE DATA TABLE IN THE ENVIRONMENT HOW CAN I SEE THE DEGREES LISTED?

view(acs2017\_ny)

#Nevermind lol, i just called the data from the envronment. It has alot of information, rent too.

mean(educ\_nohs[(AGE >= 25)&(AGE <= 55)])

mean(educ\_hs[(AGE >= 25)&(AGE <= 55)])

mean(educ\_somecoll[(AGE >= 25)&(AGE <= 55)])

mean(educ\_college[(AGE >= 25)&(AGE <= 55)])

mean(educ\_advdeg[(AGE >= 25)&(AGE <= 55)])

# alternatively

restrict1 <- as.logical((AGE >= 25)&(AGE <= 55))

dat\_age\_primeage <- subset(acs2017\_ny, restrict1)

detach()

attach(dat\_age\_primeage)

mean(educ\_nohs)

mean(educ\_hs)

mean(educ\_college)

mean(educ\_advdeg)

detach()