

bootstrap.R

March 3, 2019

```
# load data
unique_cols <- read.csv('unique_cols.csv', stringsAsFactors = FALSE,
                        colClasses = c('NULL', 'double', 'character'))

# subset conditions
cond1 <- !is.na(unique_cols$total_obligation) # remove NA
cond2 <- unique_cols$total_obligation >= 1 # Remove -, 0's
col_sub <- unique_cols[cond1 & cond2,]

parent_id_freq <- table(col_sub$parent_recipient_unique_id, useNA = "always")
parent_id_subset <- names(parent_id_freq[parent_id_freq > 99])
cond3 <- col_sub$parent_recipient_unique_id %in% parent_id_subset # subset to 100+
data_sub <- col_sub[cond3,]

# keep only the first digit from each row in total_obligation
data_sub$total_obligation <- as.integer(substr(data_sub$total_obligation, start = 1, stop = 1))

# save to csv
write.csv(data_sub, 'bootstrap.csv')
```