```r
source("lib/utils.R")

# load the election results
results <- read.csv("data/results.csv", colClasses=c(rep("character", 3),
rep("numeric", 3)))
results$OTHER <- results$TOTAL - (results$CLINTON + results$TRUMP)
results$TRUMP_PERCENT   <- 100 * results$TRUMP    / results$TOTAL
results$CLINTON_PERCENT <- 100 * results$CLINTON  / results$TOTAL
results$TRUMP_MARGIN    <- results$TRUMP_PERCENT - results$CLINTON_PERCENT
results$CLINTON_MARGIN  <- results$CLINTON_PERCENT - results$TRUMP_PERCENT
results$MARGIN          <- abs(results$CLINTON_MARGIN)

state_abbrs <- results$ABBR

f <- function(abbr) {
  print(paste("Reading", abbr))
  read_state(abbr, 2015)
}

names <- do.call(rbind, lapply(state_abbrs, f))

# If you're on the right OS, you can run this command line to make sure we got every
name
system("grep -o ',2015,' data/states/*.TXT | wc -l")

names$HANDLE <- paste(names$NAME, "_", names$GENDER, sep="")

# tally state counts--number of states a name appears in--for each name+gender, and
join them
name_counts <- as.data.frame(table(names$HANDLE))
colnames(name_counts) <- c("HANDLE", "STATE_COUNT")

names <- merge(names, name_counts, by="HANDLE")

# sort by state rank, with tie going to state with more of that name
sorted <- names[order(names$GENDER, names$HANDLE, names$RANK, -names$VALUE),]

# reduce to those in 10+ states
top_10_roster <- unique(names$HANDLE[names$STATE_COUNT >= 10])
top_10_names <- subset(sorted, sorted$STATE_COUNT >= 10)

# for each name, rank which states are highest now that we've sorted them
top_10_names$RANK_N <- 0
for (handle in top_10_roster) {
  top_10_names$RANK_N[top_10_names$HANDLE == handle] <-
seq(1,NROW(top_10_names$RANK[top_10_names$HANDLE == handle]))
}

# and filter down to top ten
filtered <- subset(top_10_names, top_10_names$RANK_N <= 10)

# join names with political results
filtered <- merge(filtered, results, by="ABBR")
# and re-sort because merges always mess that up
filtered <- filtered[order(filtered$GENDER, filtered$HANDLE, filtered$RANK_N),]

```

```r
# Thx, http://stackoverflow.com/questions/3443687/formatting-decimal-places-in-r
specify_decimal <- function(x, k) format(round(x, k), nsmall=k)

# reduce long decimals for csv files
cleaned <- filtered
cleaned$TRUMP_PERCENT   <- specify_decimal(filtered$TRUMP_PERCENT, 2)
cleaned$CLINTON_PERCENT <- specify_decimal(filtered$CLINTON_PERCENT, 2)
cleaned$TRUMP_MARGIN    <- specify_decimal(filtered$TRUMP_MARGIN, 2)
cleaned$CLINTON_MARGIN  <- specify_decimal(filtered$CLINTON_MARGIN, 2)
cleaned$MARGIN          <- specify_decimal(filtered$MARGIN, 2)

# eyeball this
for (i in 1:100) {
  for (c in 15:19) {
    print(paste(paste(filtered[i,c], collapse=" "), paste(cleaned[i,c], collapse=" "),
sep=" -- "))
  }
}

# let's put these values in the baby box
write.csv(cleaned, "csv/all_names.csv", row.names=FALSE)

f <- function(handle) {
  print(handle)
  write.csv(subset(cleaned, cleaned$HANDLE==handle), paste("csv/names/", handle,
".csv", sep=""), row.names=FALSE)
}

# write csvs for every name
lapply(top_10_roster, f)

# total votes for each name among it's top-ten states
names_trump <- aggregate(TRUMP ~ HANDLE, FUN=sum, data=filtered)
names_clinton <- aggregate(CLINTON ~ HANDLE, FUN=sum, data=filtered)
names_other <- aggregate(OTHER ~ HANDLE, FUN=sum, data=filtered)
names_total <- aggregate(TOTAL ~ HANDLE, FUN=sum, data=filtered)

# add to roster
roster <- merge(names_trump, names_clinton, by="HANDLE")
roster <- merge(roster, names_other, by="HANDLE")
roster <- merge(roster, names_total, by="HANDLE")

roster$TRUMP_PERCENT   <- 100 * roster$TRUMP    / roster$TOTAL
roster$CLINTON_PERCENT <- 100 * roster$CLINTON  / roster$TOTAL
roster$SPLIT          <- roster$TRUMP_PERCENT - roster$CLINTON_PERCENT

# join with some of the original info that didn't survive aggregation
info <- top_10_names[,c("HANDLE", "NAME", "GENDER", "STATE_COUNT")]
info <- info[!duplicated(info), ]

roster <- merge(roster, info, by="HANDLE")

# add nat'l data
national_names <- read_national(2015)
national_names$HANDLE <- paste(national_names$NAME, "_", national_names$GENDER, sep="")

# tally state counts
```

```r
roster <- merge(roster, national_names[,c("HANDLE", "VALUE", "RANK")], by="HANDLE" )
roster$WINNER <- ""
roster$WINNER[roster$TRUMP < roster$CLINTON] <- "D"
roster$WINNER[roster$TRUMP > roster$CLINTON] <- "R"

roster$D_COUNT <- 0
roster$R_COUNT <- 0

for (i in 1:NROW(roster)) {
  handle <- roster[i,]$HANDLE
  print(handle)
  top_10 <- subset(filtered, filtered$HANDLE == handle)
  roster[i,]$D_COUNT <- NROW(subset(top_10, top_10$CLINTON > top_10$TRUMP))
  roster[i,]$R_COUNT <- NROW(subset(top_10, top_10$CLINTON < top_10$TRUMP))
}

# we're done!
write.csv(roster[,c(9:16,1:8)], "csv/roster.csv", row.names=FALSE)
```