

Arapahoe and Denver County Diagnostic

Theodore Dounias
September 18, 2018

Reading in the Data

```
#Read in voter history data
vhist <- read_csv("data/2017_CO/VHist_2017/CO_2017_VHist_full.csv",
  col_types = cols(PARTY = col_skip(), X1 = col_skip()))

#Read in vrf data
vrf <- read_csv("data/2017_CO/VRF_2017/CO_2017_VRF_full.csv",
  col_types = cols_only(VOTER_ID = col_guess(), COUNTY = col_guess(),
    REGISTRATION_DATE = col_guess(),
    EFFECTIVE_DATE = col_guess(),
    VOTER_STATUS = col_guess(),
    PARTY = col_guess()))

#Are all voter ids from the history data in the registration data?
sum(!(unique(vhist$VOTER_ID) %in% unique(vrf$VOTER_ID)))
```

```
## [1] 0
```

```
#Yes
```

Data Wrangling, using mostly or only filter

```
vhist_den_arap <- vhist %>%
  filter(COUNTY_NAME %in% c("Denver", "Arapahoe"))

vrf_den_arap <- vrf %>%
  filter(COUNTY %in% c("Denver", "Arapahoe"))

#Are all voter ids in both of them? This outputs the NUMBER of IDs missing
sum(!(unique(vhist_den_arap$VOTER_ID) %in% vrf_den_arap$VOTER_ID))
```

```
## [1] 159581
```

```
#...no. So let's check out the problems
not_common_voter_ids <- vhist_den_arap[!(vhist_den_arap$VOTER_ID %in% vrf_den_arap$VOTER_ID), ]

#Let's merge with the general voter registration file.
#I will do a left merge, to preserve observations only from the
#dataset of "not common" voters

not_common_voter_ids <- left_join(not_common_voter_ids, vrf, by = "VOTER_ID")

names(not_common_voter_ids)[6:7] <- c("VOTED_IN", "CURRENTLY_REGISTERED_IN")

#See if you can spot the problem...
head(not_common_voter_ids[, c(1, 2, 6, 7)])
```

```
## # A tibble: 6 x 4
##   VOTER_ID ELECTION_TYPE VOTED_IN CURRENTLY_REGISTERED_IN
##   <int> <chr>          <chr>          <chr>
## 1    72891 General      Denver      Douglas
## 2    73359 General      Denver      Douglas
## 3    73410 General      Arapahoe    El Paso
## 4    73410 General      Arapahoe    El Paso
## 5    73410 General      Arapahoe    El Paso
## 6    73410 General      Arapahoe    El Paso
```

Plainly stated, the problem is that voters who moved from one county to the next are not counted in the denominator of the turnout statistic (Registrants in a given county), but *are* counted in the numerator (total votes cast in an election). I'm not sure what the patterns of people moving between counties actually is, but it seems plausible that their doing so is causing this calculation to be a bit out of whack.