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
```

Wrangling and turnout stats

I first select the correct counties:

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

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

Check Voter IDs and Troubleshoot

```
#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...
kable(head(not_common_voter_ids[, c(1, 2, 6, 7)]))
```

VOTER_ID	ELECTION_TYPE	VOTED_IN	CURRENTLY_REGISTERED_IN
72891	General	Denver	Douglas
73359	General	Denver	Douglas
73410	General	Arapahoe	El Paso
73410	General	Arapahoe	El Paso
73410	General	Arapahoe	El Paso
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 enumerator (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.

This makes it necessary to find a way to count in the voters that moved counties. Therefore, I will use two seperate methods to do so, and then check my results. The methods are:

- *Method 1*: When I create a count of voters, I subtract any voter that is not in the current reg file. This would approximate the effect of subtracting one from the numerator and one from the denominator. This assumes that the voter that moved was never present in the first county to begin with.Turnout equasion is:

$$turnout = \frac{Count\ of\ Voters\ in\ Election\ -\ Voters\ that\ Moved\ County}{Sum\ of\ Registrants\ Shown\ in\ VRF\ for\ Election}\ \%$$

- *Method 2*: If any of the voters has an ID that does not show up as a registrant in that county, I add them to the numerator as normal, but add +1 to the denominator. This assumes that the voter was also registered for a time period I know nothing about. Turnout equasion is:

$$turnout = \frac{Count\ of\ Voters\ in\ Election}{Sum\ of\ Registrants\ Shown\ in\ VRF\ for\ Election\ +\ Any\ Voters\ who\ Voted\ but\ Moved}\ \%$$

In either case it is necessary to wrangle the data into the following datasets:

- All voters that voted in election $ A_{d,C} $, where d is the date and C is the county.
- Registrants in county C on date d.
- Voters in the Vote History dataset but not in the Voter Registration dataset.

Data Wrangling

I will start with the voter history file. I will start by looking at two elections across Denver and Arapahoe counties: presidential for 2012 and presidential for 2008.

```
#Format variables: date and method
vhist_den_arap$ELECTION_DATE <- mdy(vhist_den_arap$ELECTION_DATE)
```

```
vhist_den_arap$VOTING_METHOD <- as.factor(vhist_den_arap$VOTING_METHOD)

#Wrangle, see County Lvl Turnouts and Graphs
vhist_den_arap_08_12 <- vhist_den_arap %>%
  #Make voting method vector correct
  mutate(VOTING_METHOD = fct_collapse(VOTING_METHOD,
    VBM = c("Absentee Carry", "Absentee Mail", "Mail Ballot",
            "Mail Ballot - DRE"),
    In_Person = c("Early Voting", "Early Voting - DRE",
                  "In Person", "In Person - DRE", "Polling Place",
                  "Vote Center", "Vote Center - DRE"))) %>%
  #Chose Elections
  filter(ELECTION_DATE %in% c(as.Date("2012-11-06"), as.Date("2008-11-04")))
```

Next up, the voter registration file.

```
#Recode dates
vrf_den_arap$REGISTRATION_DATE <- mdy(vrf_den_arap$REGISTRATION_DATE)

#Output dataset
vrf_den_arap_08 <- vrf_den_arap %>%
  filter(REGISTRATION_DATE <= as.Date("2008-11-04") -22)

vrf_den_arap_12 <- vrf_den_arap %>%
  filter(REGISTRATION_DATE <= as.Date("2012-11-06") - 22)
```

With these files I can output som initial turnout statistics along with actual reported turnout, on all voters–Active or Inactive–was:

```
#Turnouts calculated without application of methods
denver_08_simple <- nrow(filter(vhist_den_arap_08_12,
                      year(ELECTION_DATE) == "2008" &
                  COUNTY_NAME == "Denver"))/nrow(filter(vrf_den_arap_08,
                                                        COUNTY == "Denver" &
                                                            VOTER_STATUS == "Active"))

denver_12_simple <- nrow(filter(vhist_den_arap_08_12,
                      year(ELECTION_DATE) == "2012" &
                  COUNTY_NAME == "Denver"))/nrow(filter(vrf_den_arap_12, COUNTY == "Denver"))

arap_08_simple <- nrow(filter(vhist_den_arap_08_12,
                      year(ELECTION_DATE) == "2008" &
                  COUNTY_NAME == "Arapahoe"))/nrow(filter(vrf_den_arap_08,
                                                          COUNTY == "Denver" &
                                                            VOTER_STATUS == "Active"))

arap_12_simple <- nrow(filter(vhist_den_arap_08_12,
                      year(ELECTION_DATE) == "2012" &
                  COUNTY_NAME == "Arapahoe"))/nrow(filter(vrf_den_arap_12, COUNTY == "Denver"))

#Source: CO Secretary of State
actual_turnouts <- c( .8849, .9332, .6348, .7491)
simple_calc_turnouts <- c(denver_08_simple, arap_08_simple,
                          denver_12_simple, arap_12_simple)
```

```
turnouts <- data.frame(c("Denver 08", "Arapahoe 08", "Denver 12", "Arapahoe 12"),
           actual_turnouts, simple_calc_turnouts)

names(turnouts) <- c("", "Reported", "Simple Calculation")

kable(turnouts)
```

	Reported	Simple Calculation
Denver 08	0.8849	1.368301
Arapahoe 08	0.9332	1.385794
Denver 12	0.6348	1.077840
Arapahoe 12	0.7491	1.028967

A quick note: I could not find turnout for 2008 over all registered voters, so I further filtered for Active voters. This is probably due to the legislative change that occured in 2013, changing the designation of inactive voters.