Turnout_calculations

Theodore Dounias

9/23/2018

TEST METHOD FOR TURNOUTS

```
county_reg_ref <- read_csv("County_Reg_Per_Year.csv")
```

```
## Parsed with column specification:
## cols(
##   X1 = col_integer(),
##   VOTER_ID = col_integer(),
##   COUNTY16 = col_character(),
##   COUNTY15 = col_character(),
##   COUNTY14 = col_character(),
##   COUNTY13 = col_character(),
##   COUNTY12 = col_character()
## )
```

```
vhist_full <- read_csv("full_voter_history.csv")
```

```
## Parsed with column specification:
## cols(
##   VOTER_ID = col_integer(),
##   ELECTION_TYPE = col_character(),
##   ELECTION_DATE = col_character(),
##   ELECTION_DESCRIPTION = col_character(),
##   VOTING_METHOD = col_character(),
##   PARTY = col_character(),
##   COUNTY_NAME = col_character()
## )
```

```
vhist_full$ELECTION_DATE <- mdy(vhist_full$ELECTION_DATE)
```

Returning now to El Paso, I will outline the method I will use to find turnout. I will then calculate turnout statistics for the 2014 midterms and check against reported figures. Let's check out this method on Monday and find ways to automate it for all counties.

I will do the following:

DENOMINATOR Find all IDs that were registered to vote in 2014. I will do this through the county/id file I just created. Match these IDs with those from the 2014 VRF. Filter out any with a registration date prior to 22 days before election day. *If calculating turnout over active voters, filter out inactive voters.* The length of the remaining dataset should be the denominator of turnout.

NUMERATOR Use the full history file and filter for election day and county. Match the resulting IDs with the registration file created above (inactive+active). They should all have a match. If any don't, investigate before disregarding. *The number of rows should be the numerator, and should be equal to ballots cast.

I will now apply this step by step to El Paso.

```
#Make base denominator
denom_ep_14 <- county_reg_ref %>%
  select(2, 5) %>%
```

```
  filter(COUNTY14 == "El Paso")


#Read in and wrangle registration file
reg14 <- read_csv("2014reg2.csv",
         col_types = cols_only(VOTER_ID = col_guess(), COUNTY = col_guess(),
                               VOTER_STATUS = col_guess(),
                               REGISTRATION_DATE = col_guess()))

reg14$REGISTRATION_DATE <- mdy(reg14$REGISTRATION_DATE)

#Check if all IDs are present in the reg file
sum(!(denom_ep_14$VOTER_ID %in% reg14$VOTER_ID))
```

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

```
#Matching IDs
denom_ep_14 <- left_join(denom_ep_14, reg14, by = "VOTER_ID")

#Filtering invalid dates
denom_ep_14 <- denom_ep_14 %>%
  filter(REGISTRATION_DATE <= as.Date("2014-11-04") - 22)

#I have turnout statistics from the SoS as percentage of
#Registered Voters, so no need for further filtering!

denominator <- nrow(denom_ep_14)
```

```
#Make file with votes
num_ep_14 <- vhist_full %>%
  filter(ELECTION_DATE == as.Date("2014-11-04") &
           COUNTY_NAME == "El Paso")

#Is any of them not in the reg file?
sum(!(num_ep_14$VOTER_ID %in% denom_ep_14$VOTER_ID))
```

```
## [1] 3039
```

```
#Investigate!
missing_voters <- num_ep_14[!(num_ep_14$VOTER_ID %in% denom_ep_14$VOTER_ID),]

#Similarities in PARTY registration...
summary(as.factor(missing_voters$PARTY))
```

```
## NO DATA
##    3039
```

```
#None appear to have data on their party registration!

#Do they exist in the 2014 reg file?
sum(!(missing_voters$VOTER_ID %in% reg14$VOTER_ID))
```

```
## [1] 41
```

```
#Most of them do!

#Do they exist in the cumulative county reg file?
```

```
sum(!(missing_voters$VOTER_ID %in% county_reg_ref$VOTER_ID))
```

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

```
#All of them do!

#Voting methods are different...
summary(as.factor(missing_voters$VOTING_METHOD))
```

```
##   Absentee Mail       In Person In Person - DRE     Mail Ballot
##             198             951             271            1619
```

```
#Check to see active/inactive status
missing_voters <- left_join(missing_voters, reg14, by = "VOTER_ID")

#Almost all active or NA (or weird stuff I will deal with)
summary(as.factor(missing_voters$VOTER_STATUS))
```

```
##                   Active                 Inactive UNITED STATES OF AMERICA
##                     2988                        1                        8
##                     NA's
##                       42
```

```
###HOWEVER
#Proceeding for now without their inclusion...
numerator <- sum(num_ep_14$VOTER_ID %in% denom_ep_14$VOTER_ID)

turnout <- numerator/denominator
```

It is therefore obvious that these numbers are quite close by the following table:

```
table_df <- data.frame(c("Calculated", "Reported"),
                       c(numerator, 231635), c(denominator, 424553),
                       c(signif(turnout, 3)*100, 54.56))

names(table_df) <- c("", "Ballots Cast", "Total Registered", "Turnout")

kable(table_df)
```

	Ballots Cast	Total Registered	Turnout
Calculated	228292	418931	54.50
Reported	231635	424553	54.56