Data Diagnostics

Theodore Dounias

9/23/2018

FULL VOTER HISTORY DIAGNOSTIC

```r
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()
## )
```

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

First I will go through all the variables. I will count unique values, count NAs, see if there is any pattern to the NAs, and clean up when necessary.

```r
#Number of NAs?
sum(is.na(vhist_full$VOTER_ID))
```

```
## [1] 8
```

```r
#These can probably be deleted
vhist_full <- vhist_full[!(is.na(vhist_full$VOTER_ID)), ]
#The new vhist is smaller by 8 observations!
```

```r
#Number of NAs?
sum(is.na(vhist_full$ELECTION_TYPE))
```

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

```r
#No NAs!

#Checking for "weird" values
summary(as.factor(vhist_full$ELECTION_TYPE))
```

```
##       Coordinated           General         Municipal Municipal Run-off
##           8502023          21822792            659511            150317
##           Primary            Recall            School           Special
##           5673882            162447             97376            515888
##  Special District
##             61301
```

```r
#All these types are straightforward, nothing to deal with!
```

```r
#Number of NAs?
sum(is.na(vhist_full$ELECTION_DESCRIPTION))
```

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

```
#No NAs!
```

```
#Number of NAs?
sum(is.na(vhist_full$ELECTION_DATE))
```

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

```
#No NAs!

#Checking for "weird" values
unique(vhist_full$ELECTION_DATE)
```

```
##   [1] "2006-11-07" "2004-11-02" "2000-11-07" "1998-11-03" "1996-11-05"
##   [6] "2016-06-28" "2014-06-24" "2012-06-26" "2010-08-10" "2008-08-12"
##  [11] "2004-08-10" "2002-08-13" "2015-11-03" "2013-11-05" "2009-11-03"
##  [16] "2007-11-06" "2005-11-01" "2001-11-06" "1999-11-02" "1997-11-04"
##  [21] "2016-11-08" "2014-11-04" "2012-11-06" "2010-11-02" "2008-11-04"
##  [26] "2002-11-05" "2006-08-08" "2000-03-10" "1998-08-11" "2011-11-01"
##  [31] "2003-11-04" "2000-08-08" "1994-11-08" "1992-11-03" "1996-08-13"
##  [36] "1995-11-07" "1993-11-02" "1988-11-08" "1996-03-05" "1992-08-11"
##  [41] "1994-08-09" "1995-06-13" "2012-04-03" "2012-08-14" "1990-11-06"
##  [46] "1989-12-05" "1989-11-07" "1990-08-14" "1996-11-01" "2014-05-06"
##  [51] "2013-09-10" "1988-08-09" "1991-11-05" "1990-03-20" "1998-09-10"
##  [56] "1992-03-03" "1991-10-29" "1989-05-16" "1989-02-07" "1992-03-17"
##  [61] "1988-12-06" "2002-11-26" "2002-08-30" "1992-11-01" "1994-08-01"
##  [66] "2011-12-20" "1991-03-19" "1994-11-01" "1991-11-01" "1995-11-01"
##  [71] "1996-08-01" "2016-05-03" "2010-05-04" "1995-09-01" "1995-01-01"
##  [76] "2015-04-07" "2013-04-02" "2012-05-08" "1990-07-10" "1991-01-01"
##  [81] "1988-05-17" "1997-11-14" "1999-05-04" "2009-08-18" "1986-11-04"
##  [86] "1986-08-12" "1989-08-01" "1997-01-03" "1995-06-20" "1994-04-05"
##  [91] "1998-02-23" "2001-03-06" "2015-05-05" "2011-06-07" "1900-01-01"
##  [96] "1984-11-06" "1991-05-07" "1989-05-02" "2000-09-12" "1996-12-19"
## [101] "1987-05-12" "1986-04-08" "1997-04-08" "2001-11-16" "2011-05-03"
## [106] "2003-04-01" "2015-09-01" "2003-04-08" "2001-04-03" "1999-04-06"
## [111] "1996-04-02" "2004-02-24" "1992-05-19" "2003-06-03" "2003-05-06"
## [116] "1996-09-01" "1992-09-01" "1992-08-01" "1993-11-01" "2015-06-02"
## [121] "2002-11-01" "2000-11-01" "1998-11-01" "1996-08-02" "1999-11-01"
## [126] "1989-03-07" "2005-05-03" "2001-05-08" "2005-07-19" "2005-04-05"
## [131] "1997-11-01" "2003-01-01" "2001-11-01" "2006-12-12" "2011-04-05"
## [136] "2009-04-07" "2007-04-03" "1998-04-07" "2004-04-06" "2002-04-02"
## [141] "2016-04-05" "2002-05-07" "2005-07-26" "2007-01-30" "2015-03-24"
## [146] "2015-02-10" "2010-04-06" "1997-02-04" "2008-04-01" "2012-01-31"
## [151] "2014-04-01" "2011-07-12" "2016-07-26" "1991-02-05" "2007-07-10"
## [156] "2010-03-02" "1992-07-14" "1992-06-02" "1991-07-09" "2005-12-13"
## [161] "2000-05-02" "1992-09-22" "2015-03-17" "2012-01-24" "1994-09-13"
## [166] "2002-06-04" "2015-01-27" "2009-04-28" "2007-05-01" "1998-12-01"
## [171] "2010-01-12" "1994-05-02" "1996-04-09" "2001-02-06" "1998-05-05"
## [176] "2009-02-17" "2005-03-08" "2009-12-15" "2005-10-04" "2006-04-04"
## [181] "2004-01-01" "2000-06-20" "1998-01-13" "2006-05-02" "2001-02-13"
## [186] "1991-05-21" "2006-06-27" "2014-04-08" "2007-06-05" "1999-12-07"
## [191] "2009-01-20" "2016-08-09" "2009-03-03" "1996-01-01" "1989-03-21"
## [196] "2004-11-16" "2000-04-04" "1989-01-17" "2003-12-09" "2001-06-19"
## [201] "1997-11-03" "2002-08-01" "2012-05-22" "1998-08-01" "2000-08-01"
## [206] "2000-08-02" "2001-10-16" "2010-12-14" "2010-12-07" "1999-08-03"
## [211] "2008-05-06" "2006-01-01" "2000-01-01" "1996-05-07" "1999-01-01"
## [216] "1999-08-02" "2007-11-07" "1998-08-02" "2003-07-29" "1997-01-01"
## [221] "1997-08-01" "1999-08-01" "1997-08-02" "2008-12-09" "1992-07-07"
## [226] "2009-08-25" "2013-04-23" "1997-05-06" "2005-04-19" "1991-11-06"
## [231] "1996-11-07" "2003-01-14" "2007-12-18" "2008-01-29" "2003-08-02"
## [236] "1900-01-02" "1997-03-05" "1997-08-13" "1997-04-16" "2004-06-08"
## [241] "2001-06-12" "1997-01-13" "1995-08-01" "1994-09-01" "1991-09-01"
```

```
#Nothing too weird, a few elections miscoded in the 1900s
```

```
#Number of NAs?
sum(is.na(vhist_full$VOTING_METHOD))
```

```
## [1] 1513468
```

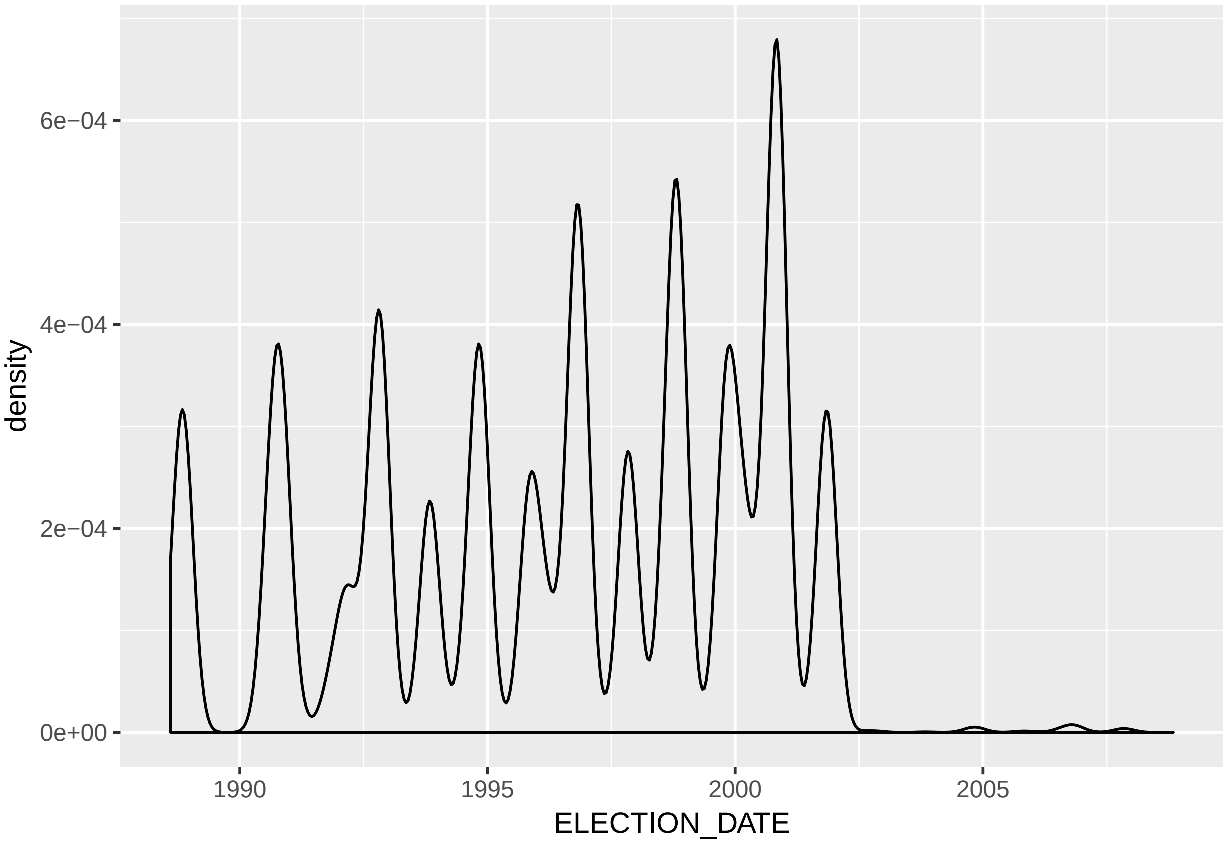
```
#A BUNCH of NAs! Let's check them out
missing_method <- vhist_full[(is.na(vhist_full$VOTING_METHOD)), ]

#Guess what county?
summary(as.factor(missing_method$COUNTY_NAME))
```

```
##      Adams    Alamosa  Archuleta    Boulder    Chaffee    Conejos 
##         12          1          4          9          6          2 
##      Delta    Douglas     Gilpin   Gunnison    Jackson  Jefferson 
##          1          2        127         97       1136    1507147 
##      Kiowa Kit Carson   La Plata       Lake Las Animas      Logan 
##        453          5        113       1684         19          9 
##    Mineral     Moffat  Montezuma     Morgan      Otero      Ouray 
##         14          5         70          6          4          3 
##       Park   Phillips     Pitkin Rio Blanco Rio Grande   Saguache 
##          1        138          2          4          1         14 
## San Miguel     Summit     Teller       Yuma 
##       2367          1         10          1
```

```
#Jefferson County!

#Date distribution
ggplot(missing_method, aes(x = ELECTION_DATE)) +
  geom_density()
```



```
#Note most if not all of the dates are before 2005. Therefore,
#if I only study elections after 2012, this should not be an issue.

#Checking for "weird" values
summary(as.factor(vhist_full$VOTING_METHOD))
```

```
##     Absentee Carry      Absentee Mail       Early Voting
##             128908           15270821            1449455
## Early Voting - DRE          In Person   In Person - DRE
##             471145             186888             156234
##        Mail Ballot  Mail Ballot - DRE      Polling Place
##            7624175               3190           10379166
##        Vote Center  Vote Center - DRE               NA's
##             261027             201060            1513468
```

```
#This is very familiar...so should be fine!
```

```
#Number of NAs?
sum(is.na(vhist_full$PARTY))
```

```
## [1] 725382
```

```
#A lot of NAs!
missing_party <- vhist_full[(is.na(vhist_full$PARTY)), ]

#Guess what county?
summary(as.factor(missing_party$COUNTY_NAME))
```

```
##       Adams     Alamosa     Arapahoe    Archuleta         Baca         Bent
##      124433         572       25076        4809         187         401
##     Boulder  Broomfield     Chaffee    Cheyenne Clear Creek     Conejos
##       26311       24092        8014         167         689        3566
##    Costilla     Crowley      Custer       Delta      Denver     Dolores
##         195         278         779        1524       48414          49
##     Douglas       Eagle     El Paso      Elbert     Fremont    Garfield
##       24526        3763       37207        1770        3959        2607
##      Gilpin       Grand    Gunnison    Hinsdale    Huerfano     Jackson
##        2692         842        1531          39         618         244
##   Jefferson       Kiowa  Kit Carson    La Plata        Lake     Larimer
##       25017         127         157        5419        2066      131793
##  Las Animas     Lincoln       Logan        Mesa     Mineral      Moffat
##        1003         164        1139       53662          48        3805
##   Montezuma    Montrose      Morgan       Otero       Ouray        Park
##        2052        2519        7568        5370         288        7336
##    Phillips      Pitkin     Prowers      Pueblo  Rio Blanco  Rio Grande
##         112        3424         777       17465         670         596
##       Routt    Saguache    San Juan  San Miguel    Sedgwick      Summit
##        4348         287         172         695          66       11098
##      Teller  Washington        Weld        Yuma
##        9322         147       76998         318
```

```
#Actually not Jefferson...seems evenly distributed

#Date distribution
unique(missing_party$ELECTION_DATE)
```

```
## [1] "2008-11-04" "2009-08-18" "2007-11-06" "2008-08-12" "2009-01-20"
## [6] "2009-03-03" "2008-05-06" "2008-04-01" "2008-12-09"
```

```
#All elections are between 2008-2009!
#I wonder what happened...

#Checking for "weird" values generally
summary(as.factor(vhist_full$PARTY))
```

```
##                    ACN AMERICAN CONSTITUTION                   DEM
##                   7601                  6902               1231304
##               DEMOCRAT            DEMOCRATIC                 GREEN
##                  20646               1713032                   427
##                    GRN     GUN OWNERS RIGHTS                   LBR
##                   9284                     6                 32485
##            LIBERTARIAN               NO DATA              PRO-LIFE
##                   3086              28985896                     2
##                 REFORM                   REP            REPUBLICAN
##                     45               1325306               2545506
##                    UAF          UNAFFILIATED                   UNI
##                 923487                114550                   590
##                   NA's
##                 725382
```

```
#Apart from double-coding (DEM/DEMOCRAT)
#Everything else seems fine
```

Conclusion here is that the only potentially worrying issue is some missing method values post 2012. Else the data look pretty solid.

VOTER REG FILE DIAGNOSTICS

This has the potential to become…repetitive. I will read in all variables, describe what each one is, and then exclusivelly keep the useful ones.

```r
reg12 <- read_csv("2012reg2.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   VOTER_ID = col_integer(),
##   HOUSE_NUM = col_integer(),
##   RESIDENTIAL_ZIP_CODE = col_integer(),
##   RESIDENTIAL_ZIP_PLUS = col_integer(),
##   PRECINCT_CODE = col_double(),
##   PRECINCT_NAME = col_double(),
##   `STATE SENATE` = col_integer(),
##   `STATE HOUSE` = col_integer()
## )
```

```
## See spec(...) for full column specifications.
```

```r
reg13 <- read_csv("2013reg2.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   VOTER_ID = col_integer(),
##   HOUSE_NUM = col_integer(),
##   RESIDENTIAL_ZIP_CODE = col_integer(),
##   RESIDENTIAL_ZIP_PLUS = col_integer(),
##   PRECINCT_NAME = col_double(),
##   CONGRESSIONAL = col_integer(),
##   `STATE HOUSE` = col_integer()
## )
## See spec(...) for full column specifications.
```

```r
reg14 <- read_csv("2014reg2.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   VOTER_ID = col_integer(),
##   HOUSE_NUM = col_integer(),
##   RESIDENTIAL_ZIP_CODE = col_integer(),
##   RESIDENTIAL_ZIP_PLUS = col_integer(),
##   PRECINCT_CODE = col_double(),
##   PRECINCT_NAME = col_double(),
##   `STATE SENATE` = col_integer(),
##   `STATE HOUSE` = col_integer()
## )
## See spec(...) for full column specifications.
```

```r
reg15 <- read_csv("2015reg2.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   VOTER_ID = col_integer(),
##   RESIDENTIAL_ZIP_CODE = col_integer(),
##   BIRTH_YEAR = col_integer(),
##   PRECINCT_CODE = col_double(),
##   PRECINCT_NAME = col_double(),
##   CONGRESSIONAL = col_integer(),
##   `STATE HOUSE` = col_integer(),
##   JUDICIAL = col_integer()
## )
## See spec(...) for full column specifications.
```

```
reg16 <- read_csv("2016reg2.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   VOTER_ID = col_integer(),
##   HOUSE_NUM = col_integer(),
##   RESIDENTIAL_ZIP_CODE = col_integer(),
##   PRECINCT_CODE = col_double(),
##   PRECINCT_NAME = col_double(),
##   `State Senate` = col_integer(),
##   `State House` = col_integer()
## )
## See spec(...) for full column specifications.
```

Currently, I think I conceivably only need the following variables:

Voter ID County *Registration Date* Voter Status *Party* Gender *Birth Year* Precinct Code

The section that follow analyze these variables across the five datasets.

```
#Check for NAs
sum(is.na(reg16$VOTER_ID))
```

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

```
sum(is.na(reg15$VOTER_ID))
```

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

```
sum(is.na(reg14$VOTER_ID))
```

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

```
sum(is.na(reg13$VOTER_ID))
```

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

```
sum(is.na(reg12$VOTER_ID))
```

```
## [1] 4
```

```
#There are 5 total, 1 in 2013 and 4 in 2012.
#All in all nothing to write home about
```

```
sum(is.na(reg16$COUNTY))
```

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

```r
sum(is.na(reg15$COUNTY))
```

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

```r
sum(is.na(reg14$COUNTY))
```

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

```r
sum(is.na(reg13$COUNTY))
```

```
## [1] 4785
```

```r
sum(is.na(reg12$COUNTY))
```

```
## [1] 4232
```

```r
#A bit more concerning, some voters are missing county values! However,
#given the sheer number of total votes that I have at my disposal,
#8k is not a large number. I think I can disregard.
```

```r
sum(is.na(reg16$REGISTRATION_DATE))
```

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

```r
sum(is.na(reg15$REGISTRATION_DATE))
```

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

```r
sum(is.na(reg14$REGISTRATION_DATE))
```

```
## [1] 3
```

```r
sum(is.na(reg13$REGISTRATION_DATE))
```

```
## [1] 4789
```

```r
sum(is.na(reg12$REGISTRATION_DATE))
```

```
## [1] 4238
```

```r
#The numbers for 2012-13 are oddly familiar...
#Let's try pulling them out!
missing2013 <- reg13[(is.na(reg13$REGISTRATION_DATE)), ]
missing2012 <- reg12[(is.na(reg12$REGISTRATION_DATE)), ]

#NOTE THE FOLLOWING!
head(missing2012)[, 1:7]
```

```
## # A tibble: 6 x 7
##   VOTER_ID COUNTY LAST_NAME REGISTRATION_DA~ OLD_VOTER_ID HOUSE_NUM
##      <int> <chr>  <chr>     <chr>            <chr>            <int>
## 1  8207482 <NA>   MCCLAVE   <NA>             <NA>                NA
## 2   449852 <NA>   KELLY     <NA>             <NA>                NA
## 3  6305977 <NA>   ROBINSON  <NA>             <NA>                NA
## 4  1603225 <NA>   WINFREE   <NA>             <NA>                NA
## 5  8009912 <NA>   DOYLE     <NA>             <NA>                NA
## 6  6696418 <NA>   DINOWITZ  <NA>             <NA>                NA
## # ... with 1 more variable: HOUSE_SUFFIX <chr>
```

```r
head(missing2013)[, 1:7]
```

```
## # A tibble: 6 x 7
##   VOTER_ID COUNTY LAST_NAME REGISTRATION_DA~ OLD_VOTER_ID HOUSE_NUM
##      <int> <chr>  <chr>     <chr>            <chr>            <int>
## 1   6.34e6 <NA>   BARBER    <NA>             <NA>                NA
## 2   1.18e5 <NA>   EMANUELS~ <NA>             <NA>                NA
## 3   6.01e8 <NA>   FORZLEY   <NA>             <NA>                NA
## 4   4.13e6 <NA>   LAUTERMI~ <NA>             <NA>                NA
## 5   6.36e6 <NA>   BAUER     <NA>             <NA>                NA
## 6   4.19e6 <NA>   STARNES   <NA>             <NA>                NA
## # ... with 1 more variable: HOUSE_SUFFIX <chr>
```

```
#They have NAs for almost all exept ID!
#I think this may be due to privacy concerns.

#Let's run some more diagnostics here.
#This command should count how many date entries per year
#contain more characters than necessary
sum(na.omit(!(nchar(reg12$REGISTRATION_DATE) == 10)))
```

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

```
sum(na.omit(!(nchar(reg13$REGISTRATION_DATE) == 10)))
```

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

```
sum(na.omit(!(nchar(reg14$REGISTRATION_DATE) == 10)))
```

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

```
sum(na.omit(!(nchar(reg15$REGISTRATION_DATE) == 10)))
```

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

```
sum(na.omit(!(nchar(reg16$REGISTRATION_DATE) == 10)))
```

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

```
#All are exactly as they should be
```

```
sum(is.na(reg16$VOTER_STATUS))
```

```
## [1] 381
```

```
sum(is.na(reg15$VOTER_STATUS))
```

```
## [1] 47
```

```
sum(is.na(reg14$VOTER_STATUS))
```

```
## [1] 298
```

```
sum(is.na(reg13$VOTER_STATUS))
```

```
## [1] 5103
```

```
sum(is.na(reg12$VOTER_STATUS))
```

```
## [1] 4708
```

```
#Similarly as before, and including those hidden for privacy,
#There is a managable amount of NAs in these statuses.
```

```
#Moving on to counts
#2012
summary(as.factor(reg12$VOTER_STATUS))
```

```
##                   Active                 Inactive UNITED STATES OF AMERICA
##                  2607664                  1035140                     2303
##                   CANADA                AUSTRALIA           UNITED KINGDOM
##                      230                      121                      104
##              NEW ZEALAND                   MEXICO                   FRANCE
##                       56                       55                       45
##                    JAPAN                    CHINA                    SPAIN
##                       45                       32                       31
##                   ISRAEL                    ITALY GERMANY FEDERAL REPUBLIC
##                       30                       29                       23
##                   BRAZIL  BRITISH COLUMBIA CANADA                 THAILAND
##                       19                       19                       17
##               COSTA RICA                HONG KONG                  IRELAND
##                       16                       16                       15
##                ARGENTINA                    CHILE                    INDIA
##                       13                       13                       12
##              SWITZERLAND                   SWEDEN     UNITED ARAB EMIRATES
##                       12                       11                       11
##                  DENMARK                  ENGLAND              KOREA SOUTH
##                       10                       10                       10
##                     PERU      NOVA SCOTIA  CANADA                  BELGIUM
##                       10                        9                        8
##        KOREA REPUBLIC OF          ONTARIO  CANADA                 SCOTLAND
##                        8                        8                        8
##                   TAIWAN                  AUSTRIA                INDONESIA
##                        8                        7                        7
##                SINGAPORE                  ECUADOR                  MOROCCO
##                        7                        6                        6
##                   NORWAY                  VIETNAM          ALBERTA  CANADA
##                        6                        6                        5
##                    EGYPT                    KENYA                   KUWAIT
##                        5                        5                        5
##              PHILIPPINES                   RUSSIA             SAUDI ARABIA
##                        5                        5                        5
##                   BELIZE                 MALAYSIA              NETHERLANDS
##                        4                        4                        4
##                   PANAMA              PUERTO RICO           QUEBEC  CANADA
##                        4                        4                        4
##             SOUTH AFRICA                    WALES                       BE
##                        4                        4                        3
##                  BERMUDA                    GHANA                     GUAM
##                        3                        3                        3
##                   POLAND                 PORTUGAL                   TURKEY
##                        3                        3                        3
##                   ZAMBIA              AFGHANISTAN                 ANGUILLA
##                        3                        2                        2
##               ANTARCTICA                  BAHRAIN                 CAMEROON
##                        2                        2                        2
##                 COLOMBIA           CZECH REPUBLIC                  ESTONIA
##                        2                        2                        2
##                      FIJI            GREAT BRITAIN                   GREECE
##                         2                        2                        2
##                  HONDURAS                KAZAKSTAN                    MALTA
##                         2                        2                        2
##                   MOLDOVA                    NEPAL                NICARAGUA
##                         2                        2                        2
##                   NIGERIA                  ROMANIA                  SENEGAL
##                         2                        2                        2
##                  TANZANIA                    01002                    02139
##                         2                        1                        1
##                     09310                    09331                     1510
##                         1                        1                        1
##                      1888                     1890                    20052
##                         1                        1                        1
##                      2794                     3030                  (Other)
##                         1                        1                       41
##                      NA's
##                      4708
```

```
#2013
summary(as.factor(reg13$VOTER_STATUS))
```

```
##                       Active                     Inactive
##                      2801955                       752851
##     UNITED STATES OF AMERICA                       CANADA
##                         1582                          154
##                    AUSTRALIA               UNITED KINGDOM
##                           92                           73
##                  NEW ZEALAND                       MEXICO
##                           43                           33
##                       FRANCE                        JAPAN
##                           31                           28
##                        SPAIN                        CHINA
##                           23                           19
##      BRITISH COLUMBIA CANADA     GERMANY FEDERAL REPUBLIC
##                           18                           18
##                       ISRAEL                        ITALY
##                           18                           18
##                       BRAZIL                     THAILAND
##                           16                           16
##                   COSTA RICA                    HONG KONG
##                           12                           12
##                      IRELAND                  SWITZERLAND
##                           10                           10
##                      ENGLAND                        INDIA
##                            9                            9
##          NOVA SCOTIA  CANADA                      BELGIUM
##                            9                            8
##                       SWEDEN                    ARGENTINA
##                            8                            7
##                      DENMARK                  KOREA SOUTH
##                            7                            7
##              ONTARIO  CANADA                    INDONESIA
##                            7                            6
##            KOREA REPUBLIC OF              ALBERTA  CANADA
##                           6                           5
##                       CHILE                     MOROCCO
##                           5                           5
##                 PHILIPPINES                      RUSSIA
##                           5                           5
##                   SINGAPORE                      TAIWAN
##                           5                           5
##        UNITED ARAB EMIRATES                     VIETNAM
##                           5                           5
##                     AUSTRIA                        PERU
##                           4                           4
##                    SCOTLAND                SOUTH AFRICA
##                           4                           4
##                 AFGHANISTAN                      BELIZE
##                           3                           3
##                     ECUADOR                       EGYPT
##                           3                           3
##                        GUAM                 NETHERLANDS
##                           3                           3
##                 PUERTO RICO                       WALES
##                           3                           3
##                     BERMUDA              CZECH REPUBLIC
##                           2                           2
##                        FIJI               GREAT BRITAIN
##                           2                           2
##                       KENYA                    MALAYSIA
##                           2                           2
##                   NICARAGUA                      NORWAY
##                           2                           2
##                      PANAMA                      POLAND
##                           2                           2
##                    PORTUGAL              QUEBEC  CANADA
##                           2                           2
##               Returned Mail                SAUDI ARABIA
##                           2                           2
##                    TANZANIA                      TURKEY
##                           2                           2
##                      UGANDA                      ZAMBIA
##                           2                           2
##                       01002                       02139
##                           1                           1
##                       09310                        1510
##                           1                           1
##                        1888                        1890
##                           1                           1
##                        2794                        3030
##                           1                           1
##                       31401                       31905
##                           1                           1
##                        5000                       63103
##                           1                           1
##                        7103                       80230
##                           1                           1
##                        8803                        9110
##                           1                           1 
##                       93410                        9667 
##                           1                           1 
##                       97301                  AZERBAIJAN 
##                           1                           1 
##                     BAHRAIN                          BE 
##                           1                           1 
##      BRITISH VIRGIN ISLANDS                    BULGARIA 
##                           1                           1 
##                       BURMA CZECH AND SLOVAK FED REPUBL 
##                           1                           1 
##                     (Other)                        NA's 
##                          21                        5103
```

```
#2014
summary(as.factor(reg14$VOTER_STATUS))
```

```
##                    Active                 Inactive UNITED STATES OF AMERICA 
##                   2883194                   759033                     1300 
##                    CANADA                AUSTRALIA           UNITED KINGDOM 
##                       143                       98                       79 
##                    MEXICO              NEW ZEALAND                   FRANCE 
##                        41                       39                       30 
##                     JAPAN                    SPAIN                    ITALY 
##                        24                       22                       21 
##   BRITISH COLUMBIA CANADA                    CHINA                 THAILAND 
##                        18                       18                       18 
##                    ISRAEL                   BRAZIL               COSTA RICA 
##                        17                       15                       14 
##                 HONG KONG GERMANY FEDERAL REPUBLIC              SWITZERLAND 
##                        14                       12                       11 
##                   IRELAND      NOVA SCOTIA  CANADA                SINGAPORE 
##                        10                       10                       10 
##                   BELGIUM                ARGENTINA          ONTARIO  CANADA 
##                         9                        8                        8 
##                    SWEDEN                  DENMARK                  ENGLAND 
##                         8                        7                        7 
##                     CHILE                    INDIA              PHILIPPINES 
##                         6                        6                        6 
##              SOUTH AFRICA     UNITED ARAB EMIRATES                  VIETNAM 
##                         6                        6                        6 
##           ALBERTA  CANADA        KOREA REPUBLIC OF              KOREA SOUTH 
##                         5                        5                        5 
##                   MOROCCO                     PERU                  AUSTRIA 
##                         5                        5                        4 
##               NETHERLANDS                   RUSSIA                 SCOTLAND 
##                         4                        4                        4 
##                    TAIWAN                   BELIZE                    EGYPT 
##                         4                        3                        3 
##                   GEORGIA                    KENYA                 MALAYSIA 
##                         3                        3                        3 
##                    NORWAY              AFGHANISTAN                  BERMUDA 
##                         3                        2                        2 
##                   CROATIA           CZECH REPUBLIC                  ECUADOR 
##                         2                        2                        2
##                      FIJI            GREAT BRITAIN                GUATEMALA
##                         2                        2                        2
##                 INDONESIA                NICARAGUA                   POLAND
##                         2                        2                        2
##               PUERTO RICO           QUEBEC  CANADA            Returned Mail
##                         2                        2                        2
##                  TANZANIA                   TURKEY                   UGANDA
##                         2                        2                        2
##                   UKRAINE                    WALES                   ZAMBIA
##                         2                        2                        2
##                     01002                    02139                     0800
##                         1                        1                        1
##                      1888                     1890                    20052
##                         1                        1                        1
##                      3030                    31905                     5000
##                         1                        1                        1
##                      5902                    63103                     7103
##                         1                        1                        1
##                      7562                     8123                     8850
##                         1                        1                        1
##                      9110                     9667                    98433
##                         1                        1                        1
##                      9998                 ANGUILLA                  BAHRAIN
##                         1                        1                        1
##                BANGLADESH                  BOLIVIA                 BOTSWANA
##                         1                        1                        1
##                  BULGARIA                 CAMBODIA                  (Other)
##                         1                        1                       24
##                      NA's
##                       298
```

```
#2015
summary(as.factor(reg15$VOTER_STATUS))
```

```
##                     Active               AFGHANISTAN
##                    2822516                         1
##                  ARGENTINA                 AUSTRALIA
##                          1                        22
##                    BELGIUM                   BERMUDA
##                          2                         1
##                     BRAZIL                    CANADA
##                          3                        78
##                      CHILE                     CHINA
##                          1                         7
##                 COSTA RICA                  ETHIOPIA
##                          2                         1
##             Failed to Vote                    FRANCE
##                         21                         2
##   GERMANY FEDERAL REPUBLIC             GREAT BRITAIN
##                          5                         2
##                       GUAM                 HONG KONG
##                          1                         3
##                   Inactive                     INDIA
##                     686039                         1
##                  INDONESIA                   IRELAND
##                         1                         2
##                    ISRAEL                     JAPAN
##                         9                         5
##         KOREA REPUBLIC OF               KOREA SOUTH
##                         1                         1
##                    MEXICO MICRONESIA  FEDERATED STS
##                        18                         1
##                      NCOA               NETHERLANDS
##                         1                         1
##               NEW ZEALAND                    NORWAY
##                         9                         1
##       NOVA SCOTIA  CANADA                      PERU
##                         1                         2
##               PHILIPPINES                    POLAND
##                         4                         2
##              REPUBLIC OF"                    RUSSIA
##                         3                         1
##                  SCOTLAND                     SPAIN
##                         1                         3
##               SWITZERLAND                  THAILAND
##                         1                         1
##      Undeliverable Ballot            UNITED KINGDOM
##                         2                        15
##  UNITED STATES OF AMERICA                   VIETNAM
##                        13                         1
##                      NA's
##                        47
```

```
#2016
summary(as.factor(reg16$VOTER_STATUS))
```

```
##                       Active                    Inactive
##                      3283797                      550320
##     UNITED STATES OF AMERICA                      CANADA
##                         1243                         187
##               UNITED KINGDOM                   AUSTRALIA
##                          100                          97
##                       MEXICO                       SPAIN
##                           45                          32
##                        JAPAN                 NEW ZEALAND
##                           31                          30
##                       FRANCE    GERMANY FEDERAL REPUBLIC
##                           28                          28
##                       ISRAEL                       ITALY
##                           18                          16
##                        CHINA                   HONG KONG
##                           12                          12
##                        INDIA                     IRELAND
##                           12                          12
##                  KOREA SOUTH                 NETHERLANDS
##                           12                          11
##                       BRAZIL                    THAILAND
##                            9                           9
##      BRITISH COLUMBIA CANADA                     DENMARK
##                            8                           8
##                       SWEDEN                       CHILE
##                            7                           6
##                   COSTA RICA                   SINGAPORE
##                            6                           6
##                       TURKEY                     ENGLAND
##                            6                           5
##            KOREA REPUBLIC OF                      NORWAY
##                            5                           5
##                       PANAMA                        PERU
##                            5                           5
##                  PHILIPPINES                SOUTH AFRICA
##                            5                           5
##                       TAIWAN                   ARGENTINA
##                            5                           4
##                      BELGIUM                    COLOMBIA
##                            4                           4
##                      ECUADOR               GREAT BRITAIN
##                            4                           4
##                      GRENADA                    PORTUGAL
##                            4                           4
##                        EGYPT                      GREECE
##                            3                           3
##                      HUNGARY                      POLAND
##                            3                           3
##                     SCOTLAND                 SWITZERLAND
##                            3                           3
##                      UKRAINE                     VIETNAM
##                            3                           3
##                      AUSTRIA                     BELARUS
##                            2                           2
##                      BERMUDA                        GUAM
##                            2                           2
##                    GUATEMALA                       KENYA
##                            2                           2
##                   LUXEMBOURG                      MALAWI
##                            2                           2
##                      MOROCCO                       NEPAL
##                            2                           2
##                     PAKISTAN                       QATAR
##                            2                           2
##                Returned Mail                    TANZANIA
##                            2                           2
##                       UGANDA                       09305
##                            2                           1
##                        20057                       28310
##                            1                           1
##                         3010                        4470
##                            1                           1
##                        49546                        6100
##                            1                           1
##                         6903                        7001
##                            1                           1
##                         7817                       80401
##                            1                           1
##                       80631                        8428
##                           1                           1
##                       96251                     BAHRAIN
##                           1                           1
##                      BELIZE                       BENIN
##                           1                           1
##                BURKINA FASO                    CAMBODIA
##                           1                           1
## CZECH AND SLOVAK FED REPUBL                 EL SALVADOR
##                           1                           1
##                    ETHIOPIA                     GEORGIA
##                           1                           1
##                      GUINEA                    HONDURAS
##                           1                           1
##                   INDONESIA                        IRAQ
##                           1                           1
##                     LEBANON                   LITHUANIA
##                           1                           1
##                    MALAYSIA                   NICARAGUA
##                           1                           1
##                     (Other)                        NA's
##                           8                         381
```

```r
#Voter status here clearly includes absentee voters
#and the country they live in. In fact, filtering
#one of the years we get:
country_status <- reg12[!(reg12$VOTER_STATUS %in% c("Active", "Inactive")), ]

#See the problem!
head(country_status)[,20:25]
```

```
## # A tibble: 6 x 6
##   VOTER_STATUS         PARTY GENDER BIRTH_YEAR PRECINCT_CODE PRECINCT_NAME
##   <chr>                <chr> <chr>  <chr>              <dbl>         <dbl>
## 1 <NA>                 <NA>  <NA>   <NA>                  NA            NA
## 2 UNITED STATES OF AM~ <NA>  UAF    Male                1993    2052619003
## 3 UNITED STATES OF AM~ <NA>  REP    Female              1994    2162530070
## 4 <NA>                 Acti~ <NA>   DEM                   NA          1987
## 5 ITALY                <NA>  DEM    Male                1961    1310816843
## 6 BELGIUM              <NA>  UAF    Female              1980    1320616637
```

```r
#These voters' records have been a bit mishandled.
#The variables seem "shifted" to one side
#However, they again are very few per year,
#So I think I can ignore them.
```

```r
sum(is.na(reg16$PARTY))
```

```
## [1] 1870
```

```r
sum(is.na(reg15$PARTY))
```

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

```r
sum(is.na(reg14$PARTY))
```

```
## [1] 1062
```

```r
sum(is.na(reg13$PARTY))
```

```
## [1] 6704
```

```r
sum(is.na(reg12$PARTY))
```

```
## [1] 6740
```

```r
#Similarly as before, and including those hidden for privacy,
#There is a managable amount of NAs in these statuses.

#Moving on to counts
#2012
summary(as.factor(reg12$PARTY))
```

```
##                       ACN                    Active
##                      6996                       262
##                       AEL                 AUSTRALIA
##                      3319                         2
##                    BRAZIL   BRITISH COLUMBIA CANADA
##                         3                         1
##                    CANADA                     CHILE
##                         6                         1
##                     CHINA                Conversion
##                         5                        47
##                COSTA RICA                       DEM
##                         1                   1150752
##                     EGYPT                   ENGLAND
##                         1                         1
##                  ETHIOPIA            Failed to Vote
##                         1                       558
##                    FRANCE  GERMANY FEDERAL REPUBLIC
##                         1                         1
##                       GRN                 GUATEMALA
##                     10200                         1
##                  HONDURAS                 HONG KONG
##                         1                         3
##                  Inactive                     INDIA
##                       125                         4
##                   IRELAND                     JAPAN
##                         2                         8
##  KOREA DEMOCRATIC PEOPLES         KOREA REPUBLIC OF
##                         1                         1
##               KOREA SOUTH                       LBR
##                         4                     24529
##                    MEXICO MICRONESIA  FEDERATED STS
##                         5                         1
##                      NCOA               NEW ZEALAND
##                        36                         1
##                     QATAR                       REP
##                         1                   1155877
##             Returned Mail                     SPAIN
##                       263                         3
##                    TAIWAN                   TUNISIA
##                         3                         1
##                       UAF                        UC
##                   1291100                         3
##      Undeliverable Ballot            UNITED KINGDOM
##                       224                         3
##  UNITED STATES OF AMERICA                      NA's
##                        10                      6740
```

```r
#2013
summary(as.factor(reg13$PARTY))
```

```
##                       ACN                    Active
##                      7282                       225
##                 AUSTRALIA                    BRAZIL
##                         2                         3
##   BRITISH COLUMBIA CANADA                    CANADA
##                         1                         5
##                     CHILE                     CHINA
##                         1                         4
##                Conversion                COSTA RICA
##                         8                         2
##                       DEM                     EGYPT
##                   1106935                         1
##                   ENGLAND                    FRANCE
##                         2                         1
##  GERMANY FEDERAL REPUBLIC                       GRN
##                         1                      9916
##                  HONDURAS                 HONG KONG
##                         1                         2
##                  Inactive                     JAPAN
##                        33                         6
##         KOREA REPUBLIC OF               KOREA SOUTH
##                         1                         4
##                       LBR                    MEXICO
##                     26126                         4
## MICRONESIA  FEDERATED STS                      NCOA
##                         1                       103
##               NEW ZEALAND                       REP
##                         1                   1120226
##             Returned Mail                     SPAIN
##                        99                         2
##                    TAIWAN                       UAF
##                         3                   1284303
##                        UC      Undeliverable Ballot
##                         1                       365
##            UNITED KINGDOM  UNITED STATES OF AMERICA
##                         3                         9
##                      NA's
##                      6704
```

```r
#2014
summary(as.factor(reg14$PARTY))
```

```
##                       ACN                    Active
##                      9130                       221
##                 AUSTRALIA                    BRAZIL
##                         1                         2
##   BRITISH COLUMBIA CANADA                    CANADA
##                         1                         5
##                     CHILE                     CHINA
##                         3                         4
##                COSTA RICA                       DEM
##                         1                   1117727
##                     EGYPT                   ENGLAND
##                         1                         1
##            Failed to Vote                    FRANCE
##                       884                         2
##  GERMANY FEDERAL REPUBLIC                       GRN
##                         1                     11685
##                  HONDURAS                 HONG KONG
##                         1                         1
##                  Inactive                 INDONESIA
##                        22                         1
##                     JAPAN        KOREA REPUBLIC OF
##                         2                         1
##               KOREA SOUTH                       LBR
##                         4                     32084
##                  MALAYSIA                    MEXICO
##                         1                         3
## MICRONESIA  FEDERATED STS                      NCOA
##                         1                        37
##               NEW ZEALAND                       REP
##                         1                   1140302
##             Returned Mail                     SPAIN
##                        45                         3
##                    TAIWAN                  THAILAND
##                         3                         1
##                       UAF                   UKRAINE
##                   1331121                         1
##      Undeliverable Ballot                       UNI
##                       196                       156
##            UNITED KINGDOM  UNITED STATES OF AMERICA
##                         1                         6
##                      NA's
##                      1062
```

```
#2015
summary(as.factor(reg15$PARTY))
```

```
## integer(0)
```

```
#THERE IS NO PARTY VARIABLE IN 2015!

#2016
summary(as.factor(reg16$PARTY))
```

```
##                      ACN                   Active                AUSTRALIA
##                    11933                      286                        1
##                   BRAZIL                   CANADA                    CHINA
##                        4                        3                       11
##               COSTA RICA                      DEM                  ECUADOR
##                        2                  1199642                        2
##                  ENGLAND           Failed to Vote GERMANY FEDERAL REPUBLIC
##                        1                      304                        1
##                      GRN                HONG KONG                 Inactive
##                    14094                        1                        2
##                    INDIA                INDONESIA                  IRELAND
##                        2                        1                        2
##                    ITALY                    JAPAN        KOREA REPUBLIC OF
##                        3                        1                        4
##              KOREA SOUTH                      LBR                   MEXICO
##                        1                    44761                        3
##                     NCOA              NEW ZEALAND                      REP
##                        4                        1                  1176745
##            Returned Mail                  ROMANIA                 SCOTLAND
##                        4                        2                        1
##             SOUTH AFRICA               TAJIKISTAN                 THAILAND
##                        1                        1                        4
##                      UAF                   UGANDA                  UKRAINE
##                  1385949                        1                        1
##     Undeliverable Ballot                      UNI     UNITED ARAB EMIRATES
##                       18                      965                        2
##           UNITED KINGDOM UNITED STATES OF AMERICA                VENEZUELA
##                        5                        7                        1
##                  VIETNAM                 ZIMBABWE                     NA's
##                        2                        1                     1870
```

```
#Same issue as before; some files are mismatched,
#so values have "bled" into eachother.
#Need to be careful when doing any analysis that includes Party
```

```
sum(is.na(reg16$GENDER))
```

```
## [1] 247
```

```
sum(is.na(reg15$GENDER))
```

```
## [1] 7
```

```
sum(is.na(reg14$GENDER))
```

```
## [1] 135
```

```
sum(is.na(reg13$GENDER))
```

```
## [1] 5048
```

```
sum(is.na(reg12$GENDER))
```

```
## [1] 4625
```

```
#Similarly as before, and including those hidden for privacy,
#There is a managable amount of NAs in these statuses.

#I am not going to do counts. The results will be the same.
```

The issue that arises is taht nothing seems completelly clean past the first few variables, due to the issue of variable “displacement”. I will need to be very careful when using any of them for any purpose, and make sure to tidy first.