

CPSVote Package Test

Paul Gronke

3/18/2020

Load Data Files

For reasons we haven't tracked down, you can load all years of data *except* for 2018. First, I will download the 1994-2016 CPS and load them into memory (I already have the data downloaded).

Notes: 1. For the time being, we only read a random subset of all of the CPS data. We can turn that off if users request. 2. We are unable to read the 2018 data file as of the time of this writing (=the data are available and codes are ready, we simply have a file pattern match error).

```
# WARNING: THIS NEXT COMMAND DOWNLOADS A LOT OF DATA
#cpsvote::cps_download_data(year = seq(1994,2018,2))
cps <- cps_read(dir = "~/Google Drive/EVIC_Work/Code/gronke_github/CPSVote/cps_data", year = seq(1994, 2016, 2))

## No new data files downloaded
## Reading 12 data file(s)...
## 1994 file read
## 1996 file read
## 1998 file read
## 2000 file read
## 2002 file read
## 2004 file read
## 2006 file read
## 2008 file read
## 2010 file read
## 2012 file read
## 2014 file read
## 2016 file read

## Warning in cps_read(dir = "~/Google Drive/EVIC_Work/Code/gronke_github/CPSVote/
## cps_data", : The column names provided by the CPS do not refer to the same
## question across all years. Be cautious that you are joining columns which
## correspond across years.
```

Basic Data Descriptives

We will start by checking some basic data descriptives for voting

```
cps %>%
  tabyl(VRS_VOTE, YEAR)
```

```
## VRS_VOTE    94    96 1998 2000 2002 2004 2006 2008 2010 2012 2014
##      -9 3494 4087 3773 5965 6146 7065 6971 9389 11139 8740 10264
##      -3  489  670  776   805   998 1063 1082 1249 1392 1360 1695
##      -2 1698 1468 1974 1484 2310 1542 2113 1055 1858 1391 2261
##      -1 58184 51745 52110 52295 61075 60067 59160 58439 57954 57287 57975
##       1 47466 50025 39512 50010 47897 63052 47175 59975 44802 59597 42559
##       2 42125 28284 38041 24164 41287 23730 36754 20692 35017 23223 39488
##
## 2016
## 9979
## 1573
## 1471
## 58302
## 58533
## 22238
```

```
# Let's add nice percentages
```

```
cps %>%
  tabyl(VRS_VOTE, YEAR) %>%
  adorn_percentages("col") %>%
  adorn_pct_formatting() %>%
  adorn_ns()
```

```
## VRS_VOTE          94          96          1998          2000          2002
##      -9 2.3% (3494) 3.0% (4087) 2.8% (3773) 4.4% (5965) 3.8% (6146)
##      -3 0.3% (489) 0.5% (670) 0.6% (776) 0.6% (805) 0.6% (998)
##      -2 1.1% (1698) 1.1% (1468) 1.4% (1974) 1.1% (1484) 1.4% (2310)
##      -1 37.9% (58184) 38.0% (51745) 38.3% (52110) 38.8% (52295) 38.2% (61075)
##       1 30.9% (47466) 36.7% (50025) 29.0% (39512) 37.1% (50010) 30.0% (47897)
##       2 27.5% (42125) 20.8% (28284) 27.9% (38041) 17.9% (24164) 25.9% (41287)
##          2004          2006          2008          2010          2012
## 4.5% (7065) 4.5% (6971) 6.2% (9389) 7.3% (11139) 5.8% (8740)
## 0.7% (1063) 0.7% (1082) 0.8% (1249) 0.9% (1392) 0.9% (1360)
## 1.0% (1542) 1.4% (2113) 0.7% (1055) 1.2% (1858) 0.9% (1391)
## 38.4% (60067) 38.6% (59160) 38.8% (58439) 38.1% (57954) 37.8% (57287)
## 40.3% (63052) 30.8% (47175) 39.8% (59975) 29.4% (44802) 39.3% (59597)
## 15.2% (23730) 24.0% (36754) 13.7% (20692) 23.0% (35017) 15.3% (23223)
##          2014          2016
## 6.7% (10264) 6.6% (9979)
## 1.1% (1695) 1.0% (1573)
## 1.5% (2261) 1.0% (1471)
## 37.6% (57975) 38.3% (58302)
## 27.6% (42559) 38.5% (58533)
## 25.6% (39488) 14.6% (22238)
```

Recoding Vote Turnout

We know that CPS has an unusual method for coding turnout. Let's compare how the CPS codes turnout and the method recommended by Hur and Achen.

```
cps %>%
  filter(YEAR == 2016) %>%
```

```
cps_label() %>%          # Convert columns with factor labels
cps_recode_vote() %>%    # Create two new vote turnout variables to correspond to CPS and Hur-Achen c
tabyl(VRS_VOTE, cps_turnout)
```

```
##      VRS_VOTE  YES    NO    NA_
##      YES 58533      0      0
##      NO      0 22238      0
##      DON'T KNOW      0 1471      0
##      REFUSED      0 1573      0
##      NO RESPONSE      0 9979      0
##      <NA>      0      0 58302
```

```
cps %>%
  filter(YEAR == 2016) %>%
  cps_label() %>%          # Convert columns with factor labels
  cps_recode_vote() %>%    # Create two new vote turnout variables to correspond to CPS and Hur-Achen c
  tabyl(VRS_VOTE, achenhur_turnout)
```

```
##      VRS_VOTE  YES    NO    NA_
##      YES 58533      0      0
##      NO      0 22238      0
##      DON'T KNOW      0      0 1471
##      REFUSED      0      0 1573
##      NO RESPONSE      0      0 9979
##      <NA>      0      0 58302
```

```
cps %>%
  filter(YEAR == 2016) %>%
  cps_label() %>%          # Convert columns with factor labels
  cps_recode_vote() %>%    # Create two new vote turnout variables to correspond to CPS and Hur-Achen c
  tabyl(cps_turnout, achenhur_turnout)
```

```
##      cps_turnout  YES    NO    NA_
##      YES 58533      0      0
##      NO      0 22238 13023
##      <NA>      0      0 58302
```

Voting Mode By Year With Appropriate Weights

The CPS requires that we use proper survey weights. Below, I demonstrate how to use the `srvyr` package to identify the survey design, weight data properly, and produce turnout by mode of balloting.

There are some glitches below, most notably, Voting mode appears to be available only from 2002 forward.

```
cps_recoded <- cps %>%
  cps_label() %>%          # Convert columns with factor labels
  cps_recode_vote() %>%    # Create two new vote turnout variables to correspond to CPS and Hur-Achen c
  mutate(
    census_region = case_when(
      STATE %in% c("ME", "NH", "VT", "MA", "CT", "RI",
                  "NY", "PA", "NJ") ~ "Northeast",
      STATE %in% c("ME", "DE", "WV", "DC", "VA", "NC", "SC", "GA", "FL",
                  "KY", "TN", "MS", "AL",
                  "OK", "AR", "LA", "TX") ~ "South",
      STATE %in% c("WI", "MI", "IL", "IN", "OH",
                  "ND", "MN", "SD", "IA", "NE", "MO", "KS") ~ "Midwest",
```

```

    STATE %in% c("MT", "ID", "WY", "NV", "UT", "CO", "AZ", "NM",
                "WA", "OR", "CA", "AK", "HI") ~ "West"),
  census_division = case_when(
    STATE %in% c("ME", "NH", "VT", "MA", "CT", "RI") ~ "New England",
    STATE %in% c("NY", "PA", "NJ") ~ "Middle Atlantic",
    STATE %in% c("ME", "DE", "WV", "DC", "VA", "NC", "SC", "GA", "FL") ~ "South Atlantic",
    STATE %in% c("KY", "TN", "MS", "AL") ~ "East South Central",
    STATE %in% c("OK", "AR", "LA", "TX") ~ "West South Central",
    STATE %in% c("WI", "MI", "IL", "IN", "OH") ~ "East North Central",
    STATE %in% c("ND", "MN", "SD", "IA", "NE", "MO", "KS") ~ "West North Central",
    STATE %in% c("MT", "ID", "WY", "NV", "UT", "CO", "AZ", "NM") ~ "Mountain",
    STATE %in% c("WA", "OR", "CA", "AK", "HI") ~ "Pacific"),
  vote_mode = case_when(
    VRS_VOTE_MAIL == "IN PERSON" & VRS_VOTE_EDAY == "ON ELECTION DAY" ~ "Election Day",
    VRS_VOTE_MAIL == "IN PERSON" & VRS_VOTE_EDAY == "BEFORE ELECTION DAY" ~ "Early In Person",
    VRS_VOTE_MAIL == "BY MAIL" ~ "Vote By Mail"
  )
)

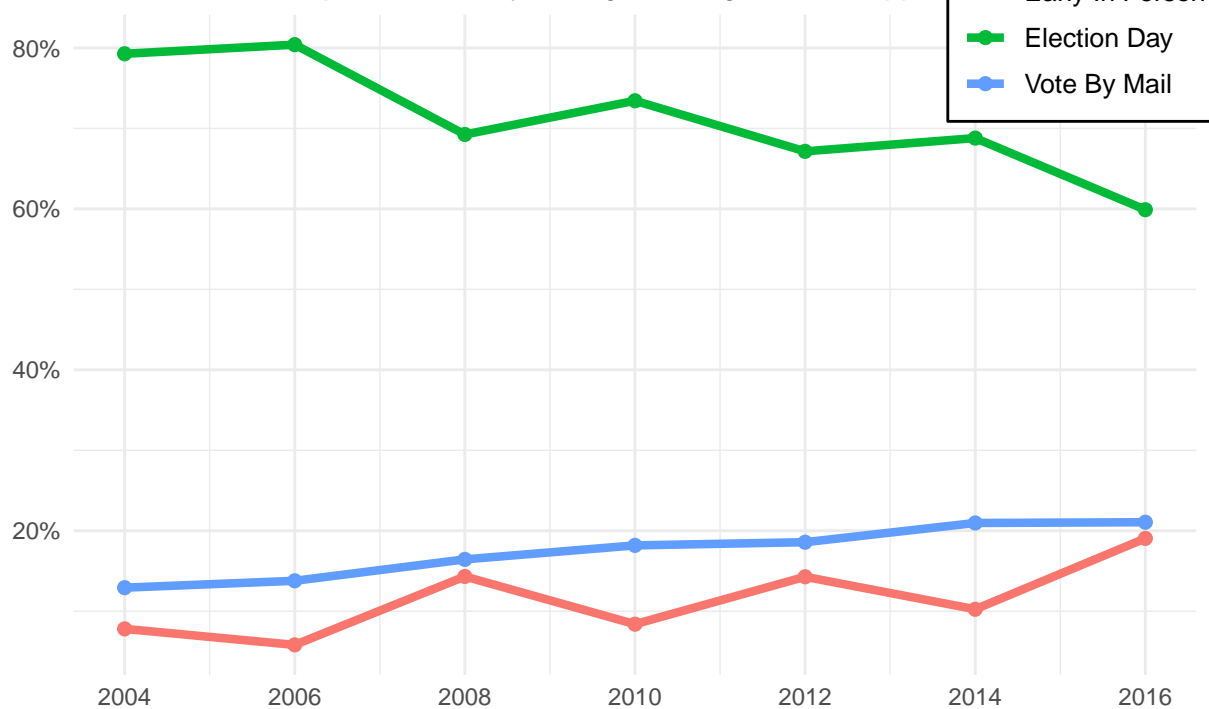
#
# Set up the survey design. Weights must be used for the CPS
#
cps_weight <- as_survey_design(cps_recoded, weights = WEIGHT)

cps_weight %>%
  filter(YEAR > 1994 & !is.na(vote_mode)) %>%
  group_by(YEAR, vote_mode) %>%
  summarize(value = survey_mean(na.rm = TRUE)) %>%
  ggplot(aes(x = YEAR, y = value, col = vote_mode, group = vote_mode)) +
  geom_line(size = 1.5) +
  geom_point(aes(x = YEAR, y = value, color = vote_mode), size = 2) +
  scale_x_continuous(breaks = seq(1996, 2016, by = 2)) +
  scale_y_continuous(labels = scales::percent) +
  labs(title = "The Growth of Early Voting, 1996 - 2016",
       subtitle = "Source: Current Population Survey, Voting and Registration Supplement",
       color = "Mode of Voting",
       y = "",
       x = "") +
  theme_minimal() +
  theme(plot.title = element_text(size = 20, family = "Times", face = "bold.italic", colour = "red"),
        legend.position = c(.9, 1),
        legend.background = element_rect(),
        legend.title = element_text(size = 12, face = "bold"),
        legend.text = element_text(size = 10))

```

The Growth of Early Voting, 1996 – 2016

Source: Current Population Survey, Voting and Registration Supplement



Graph 2: Rate of Early In Person Voting By Year and By Region

```
cps_weight %>%
  filter(YEAR > 1994) %>%
  mutate(earlyinperson = case_when(
    vote_mode == "Early In Person" ~ 1,
    vote_mode == "Election Day" |
    vote_mode == "Vote By Mail" ~ 0)
  ) %>%
  group_by(YEAR, census_region) %>%
  summarize(value = survey_mean(earlyinperson, na.rm = TRUE)) %>%
  ggplot(aes(x = YEAR, y = value, col = census_region, group = census_region)) +
  geom_line(size = 1.5) +
  geom_point(aes(x = YEAR, y = value, color = census_region), size = 2) +
  scale_x_continuous(breaks = seq(1996, 2016, by = 2)) +
  scale_y_continuous(labels = scales::percent) +
  theme_minimal() +
  labs(title = "Regional Use of Early In Person Voting, 1996 - 2016",
    subtitle = "Source: Current Population Survey, Voting and Registration Supplement",
    color = "Region") +
  ylab("") + xlab("") +
  theme(plot.title = element_text(size = 20, family = "Times", face = "bold.italic", colour = "red"),
    legend.position = c(.15,.8), legend.background = element_rect(),
    legend.title = element_text(size = 12, face = "bold"),
    legend.text = element_text(size = 10))
```

Warning: Removed 11 row(s) containing missing values (geom_path).

Warning: Removed 11 rows containing missing values (geom_point).

Regional Use of Early In Person Voting, 1996 – 2016

Source: Current Population Survey, Voting and Registration Supplement

