

County Level Turnouts and Graphs

Theodore Dounias

September 1, 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(),
    VOTER_ID = 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()))
```

Sidenote: Problems in VHist I did not initially find

I include this here as a reminder and description of the problem.

```
summary(as.factor(vhist$VOTING_METHOD))
```

##	Absentee Carry	Absentee Mail	Early Voting
##	106196	13258152	1224179
##	Early Voting - DRE	In Person	In Person - DRE
##	418364	187168	160042
##	Mail Ballot	Mail Ballot - DRE	Polling Place
##	8397030	2840	8746351
##	Vote Center	Vote Center - DRE	NA's
##	231354	181178	1269812

```
vhist_na <- vhist %>%
  filter(is.na(VOTING_METHOD))
```

```
summary(as.factor(vhist_na$COUNTY_NAME))
```

##	Adams	Alamosa	Archuleta	Boulder	Chaffee	Conejos
##	11	1	2	6	5	2
##	Delta	Douglas	Gilpin	Gunnison	Jackson	Jefferson
##	1	2	104	79	915	1264424
##	Kiowa	Kit Carson	La Plata	Lake	Las Animas	Logan
##	382	4	93	1458	18	7
##	Mineral	Moffat	Montezuma	Morgan	Otero	Ouray
##	10	5	48	4	4	3
##	Park	Phillips	Pitkin	Rio Blanco	Rio Grande	Saguache
##	1	114	2	4	1	13
##	San Miguel	Summit	Teller	Yuma		
##	2081	1	6	1		

```
summary(as.factor(vhist$COUNTY_NAME))
```

```
##      Adams      Alamosa      Arapahoe      Archuleta      Baca      Bent
##      2053966      96137      3709094      86775      36137      31366
##      Boulder      Broomfield      Chaffee      Cheyenne      Clear Creek      Conejos
##      2689444      326236      133980      23503      82810      68632
##      Costilla      Crowley      Custer      Delta      Denver      Dolores
##      38063      27871      45622      229703      4287463      18342
##      Douglas      Eagle      El Paso      Elbert      Fremont      Garfield
##      1940459      248526      3567938      192444      291829      313742
##      Gilpin      Grand      Gunnison      Hinsdale      Huerfano      Jackson
##      40622      110281      99259      10511      57919      15082
##      Jefferson      Kiowa      Kit Carson      La Plata      Lake      Larimer
##      4558186      17634      51812      346228      43923      2287320
##      Las Animas      Lincoln      Logan      Mesa      Mineral      Moffat
##      106573      44428      142804      1142624      9986      102418
##      Montezuma      Montrose      Morgan      Otero      Ouray      Park
##      156604      288434      143654      114665      49152      122798
##      Phillips      Pitkin      Prowers      Pueblo      Rio Blanco      Rio Grande
##      44681      130499      81201      1038130      51655      75302
##      Routt      Saguache      San Juan      San Miguel      Sedgwick      Summit
##      168492      40928      7771      63924      27954      170536
##      Teller      Washington      Weld      Yuma
##      150301      39491      1411792      77010
```

```
summary(as.factor(year(mdy(vhist_na$ELECTION_DATE))))
```

```
##      1988      1990      1991      1992      1993      1994      1995      1996      1997      1998
##      78288      104370      12150      136783      53392      97355      52841      155341      65252      142150
##      1999      2000      2001      2002      2003      2004      2005      2006      2007      2008
##      80369      211508      74738      446      118      1267      328      2147      938      31
```

```
summary(as.factor(vhist_na$ELECTION_TYPE))
```

```
## Coordinated      General      Primary      Recall      Special
##      327941      691002      236446      2272      12151
```

There appear to be *several* different counties here that have NA's in the Voting Method column. Most NAs are in Jefferson. Admittedly, that is the most populous of counties, but it is still about 1 in 40 voters! At first glance, the number of NAs rises exponentially with the totals of voters per county.

Every single NA is concentrated before 2013, when it could be reasonably supposed that at least local elections were conducted through VBM. The vast, VAST majority are before 2002.

The distribution of NAs across election types seems fairly consistent with the total amount of votes cast in each contest type. This is reasonable, since it suggests that counting method is independent of election type.

Data Wrangling for Graphs

First create a turnouts table with just raw number of votes per voting method. See file with tables for table on how each voting method was coded.

```
#Moving on with code as if problems were not there
```

```
#Sorting out voting methods
```

```
vhist$VOTING_METHOD[vhist$VOTING_METHOD == "Early Voting - DRE" | vhist$VOTING_METHOD == "Early Voting"]
```

```

      vhist$VOTING_METHOD == "Vote Center" | vhist$VOTING_METHOD == "In Person - DRE" |
      vhist$VOTING_METHOD == "Polling Place" | vhist$VOTING_METHOD == "Vote Center - DRE" |
vhist$VOTING_METHOD[vhist$VOTING_METHOD == "Mail Ballot - DRE" | vhist$VOTING_METHOD == "Absentee Carry" |
      vhist$VOTING_METHOD == "Absentee Mail"] <- "Mail Ballot"

```

```
#Make date variable
```

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

```
#Create turnouts file with raw vote counts
```

```

turnouts <- vhist %>%
  mutate(count = 1) %>%
  group_by(COUNTY_NAME, ELECTION_DESCRIPTION, ELECTION_TYPE, ELECTION_DATE, VOTING_METHOD) %>%
  summarize(total_votes = sum(count)) %>%
  spread(key = VOTING_METHOD, value = total_votes) %>%
  filter(year(ELECTION_DATE) >= 1992)

```

```
#Rename so that no spaces exist
```

```
names(turnouts)[5:6] <- c("IN_PERSON", "VBM")
```

```
#Replace NA values with 0
```

```
turnouts$IN_PERSON[is.na(turnouts$IN_PERSON)] = 0
```

```
turnouts$VBM[is.na(turnouts$VBM)] = 0
```

```
head(turnouts)
```

```

## # A tibble: 6 x 7
## # Groups:   COUNTY_NAME, ELECTION_DESCRIPTION, ELECTION_TYPE,
## #   ELECTION_DATE [6]
##   COUNTY_NAME ELECTION_DESCR~ ELECTION_TYPE ELECTION_DATE IN_PERSON   VBM
##   <chr>        <chr>          <chr>        <date>         <dbl> <dbl>
## 1 Adams      1992-GENERAL      General    1992-11-03         0 33783
## 2 Adams      1992-PRIMARY      Primary    1992-03-03         0  7797
## 3 Adams      1992-PRIMARY      Primary    1992-08-11         0  8480
## 4 Adams      1993-SPECIAL      Special    1993-11-02         0 15853
## 5 Adams      1994-GENERAL      General    1994-11-08         0 27798
## 6 Adams      1994-PRIMARY      Primary    1994-08-09         0  3831
## # ... with 1 more variable: `<NA>` <dbl>

```

I then calculate the total amount of registered voters per county. I include active and non-active voters.

```
#Set DATE variables
```

```
vrf$REGISTRATION_DATE <- mdy(vrf$REGISTRATION_DATE)
```

```
vrf$EFFECTIVE_DATE <- mdy(vrf$EFFECTIVE_DATE)
```

```
#Filter inactive or corrupt voters
```

```

vrf <- vrf %>%
  filter(year(REGISTRATION_DATE) > 1910 && year(EFFECTIVE_DATE) > 1910)

```

```
#Create sums of registrants per year
```

```

col_county_registrants <- vrf %>%
  mutate(count = 1, YEAR_REGISTERED = year(REGISTRATION_DATE)) %>%
  group_by(COUNTY, YEAR_REGISTERED) %>%
  summarize(REGISTERED_IN_YEAR = sum(count))

```

```
#Create total registrants per year
```

```
col_county_registrants <- col_county_registrants %>%
  group_by(COUNTY) %>%
  mutate(REGISTRANTS = cumsum(REGISTERED_IN_YEAR)) %>%
  filter(YEAR_REGISTERED >= 1992)

head(col_county_registrants)

## # A tibble: 6 x 4
## # Groups:   COUNTY [1]
##   COUNTY YEAR_REGISTERED REGISTERED_IN_YEAR REGISTRANTS
##   <chr>      <dbl>          <dbl>          <dbl>
## 1 Adams      1992            3399            28394
## 2 Adams      1993             621            29015
## 3 Adams      1994            2714            31729
## 4 Adams      1995            3378            35107
## 5 Adams      1996            4358            39465
## 6 Adams      1997            3179            42644
```

Graph Creation Step

I will include four graphs: *A cumulative graph for the whole of Colorado* A county level graph including all counties *A county level graph with only the largest CO counties* A county level graph, with the largest counties coded a different color than the rest

Statewide Graph

```
##Graph of Colorado in general##

##### What are the off year general elections about?? ####

#Total Registered Colorado V0tesr by year
total_registrants <- col_county_registrants %>%
  ungroup() %>%
  group_by(YEAR_REGISTERED) %>%
  summarise(TOTAL_REGISTRANTS = sum(REGISTRANTS))

#Create statewide turnout dataset
turnouts_statewide <- turnouts %>%
  ungroup() %>%
  select(1, 3:6) %>%
  filter(ELECTION_TYPE %in% c("General", "Coordinated", "Primary")) %>%
  mutate(ELECTION_DATE = year(ELECTION_DATE)) %>%
  group_by(ELECTION_TYPE, ELECTION_DATE) %>%
  summarise(IN_PERSON = sum(IN_PERSON), VBM = sum(VBM), TOTAL_VOTES = sum(IN_PERSON, VBM))

#Remove even year coordinated
turnouts_statewide <- turnouts_statewide[!(turnouts_statewide$ELECTION_TYPE == "Coordinated" & as.integer(t

#Remove odd year Primaries
turnouts_statewide <- turnouts_statewide[!(turnouts_statewide$ELECTION_TYPE == "Primary" & as.integer(t

#Remove odd year Generals
```

```

turnouts_statewide <- turnouts_statewide[!(turnouts_statewide$ELECTION_TYPE == "General" & as.integer(t

#Join the data
names(turnouts_statewide)[2] <- "ELECTION_YEAR"
names(total_registrants)[1] <- "ELECTION_YEAR"

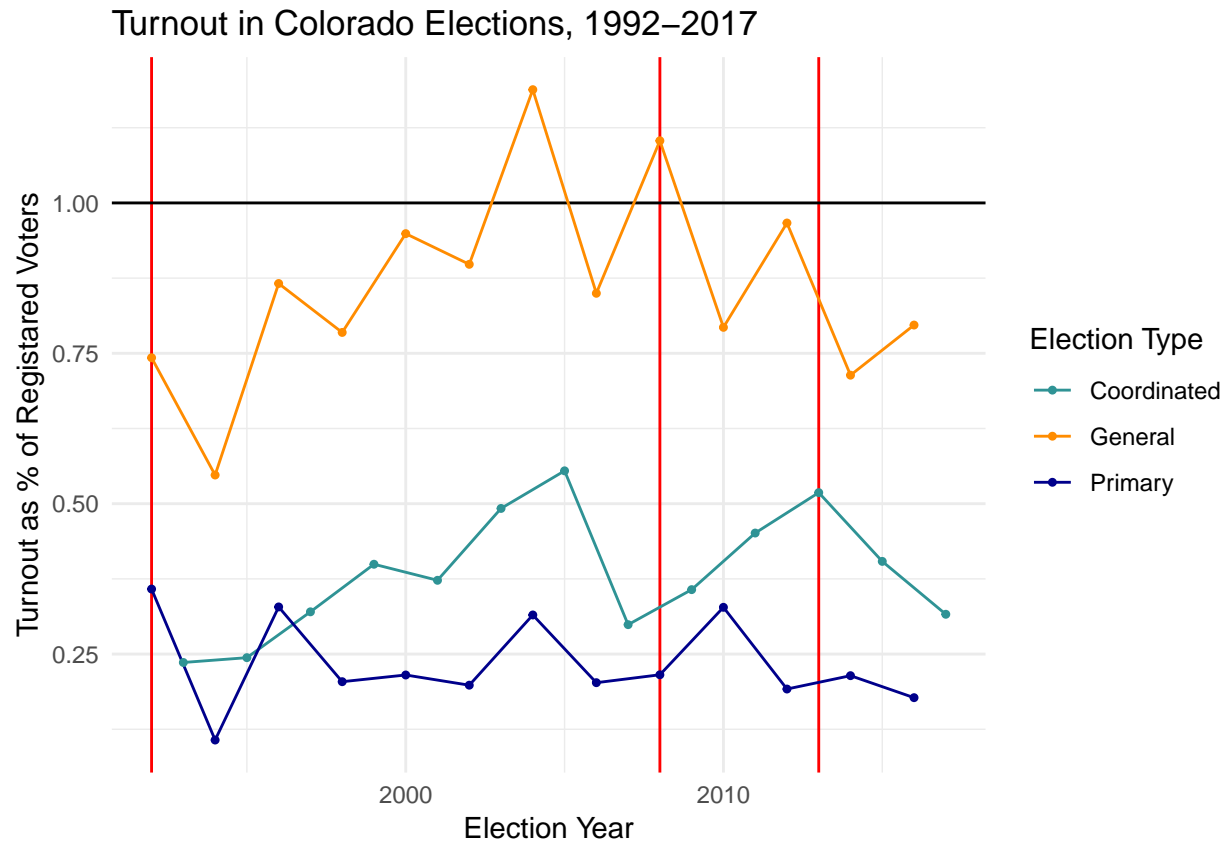
turnouts_statewide <- merge(turnouts_statewide, total_registrants, by = "ELECTION_YEAR")

#Make some tweaks
turnouts_statewide <- turnouts_statewide %>%
  mutate(TURNOUT = TOTAL_VOTES/TOTAL_REGISTRANTS)

turnouts_statewide$ELECTION_TYPE <- as.factor(turnouts_statewide$ELECTION_TYPE)

#GRAPH!!
ggplot(turnouts_statewide, aes(x = ELECTION_YEAR, y = TURNOUT, col = ELECTION_TYPE)) +
  scale_color_manual(values = c("#2F9395", "Dark Orange", "dark blue")) +
  geom_vline(xintercept = 1992, col = "red") +
  geom_vline(xintercept = 2008, col = "red") +
  geom_vline(xintercept = 2013, col = "red") +
  geom_hline(yintercept = 1) +
  geom_line() +
  geom_point(size = .9) +
  labs(title = "Turnout in Colorado Elections, 1992-2017", x = "Election Year", y = "Turnout as % of Reg
  theme_minimal()

```



County Level Graphs

```
#Create County-level turnout file
turnouts_by_county <- turnouts %>%
  ungroup() %>%
  select(1, 3:6) %>%
  filter(ELECTION_TYPE %in% c("General", "Coordinated", "Primary")) %>%
  mutate(ELECTION_DATE = year(ELECTION_DATE))

#Rename in preparation for merger
names(turnouts_by_county)[1:3] <- c("COUNTY", "ELECTION_TYPE", "ELECTION_YEAR")
names(col_county_registrants)[2] <- "ELECTION_YEAR"

#Merge
turnouts_by_county <- merge(turnouts_by_county, col_county_registrants, by = c("COUNTY", "ELECTION_YEAR"))

#Tidy data for graphing
turnouts_by_county <- turnouts_by_county %>%
  ungroup() %>%
  mutate(TURNOUT = (VBM + IN_PERSON)/REGISTRANTS) %>%
  select(-6)

#Remove even year coordinated
turnouts_by_county <- turnouts_by_county[!(turnouts_by_county$ELECTION_TYPE == "Coordinated" & as.integer(ELECTION_YEAR) % 2 == 0), ]

#Remove odd year Primaries
turnouts_by_county <- turnouts_by_county[!(turnouts_by_county$ELECTION_TYPE == "Primary" & as.integer(ELECTION_YEAR) % 2 == 1), ]
```

```

turnouts_by_county <- turnouts_by_county[!(turnouts_by_county$ELECTION_TYPE == "Primary" & as.integer(t

#Remove odd year Generals
turnouts_by_county <- turnouts_by_county[!(turnouts_by_county$ELECTION_TYPE == "General" & as.integer(t

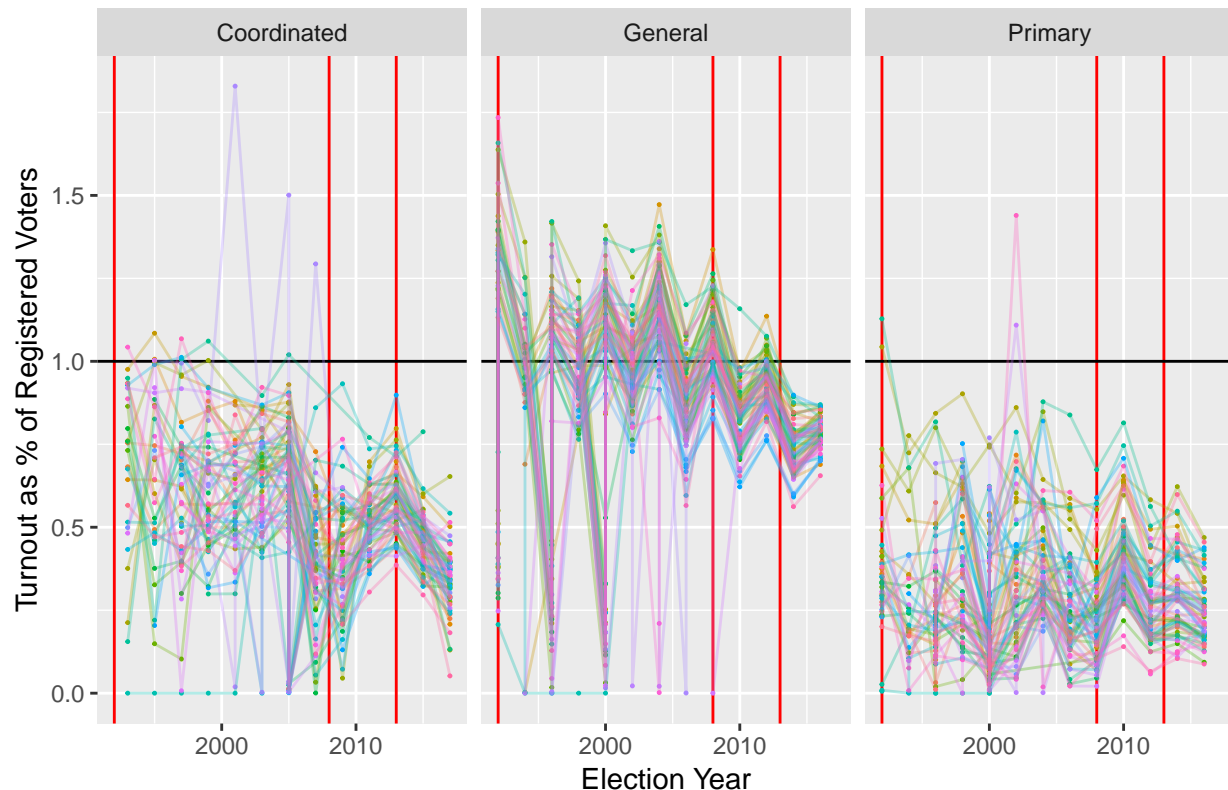
#Make County a Factor
turnouts_by_county$COUNTY <- as.factor(turnouts_by_county$COUNTY)

#GRAPH

##Just the first I could pump out
ggplot(turnouts_by_county, aes(x = ELECTION_YEAR, y = TURNOUT, col = COUNTY)) +
  facet_wrap(facets = "ELECTION_TYPE") +
  geom_vline(xintercept = 1992, col = "red") +
  geom_vline(xintercept = 2008, col = "red") +
  geom_vline(xintercept = 2013, col = "red") +
  geom_hline(yintercept = 1) +
  geom_point(size = .2) +
  geom_line(alpha = .3) +
  labs(title = "Turnout in Colorado Elections by County, 1992-2017", x = "Election Year", y = "Turnout a
  guides(col=FALSE)

```

Turnout in Colorado Elections by County, 1992–2017



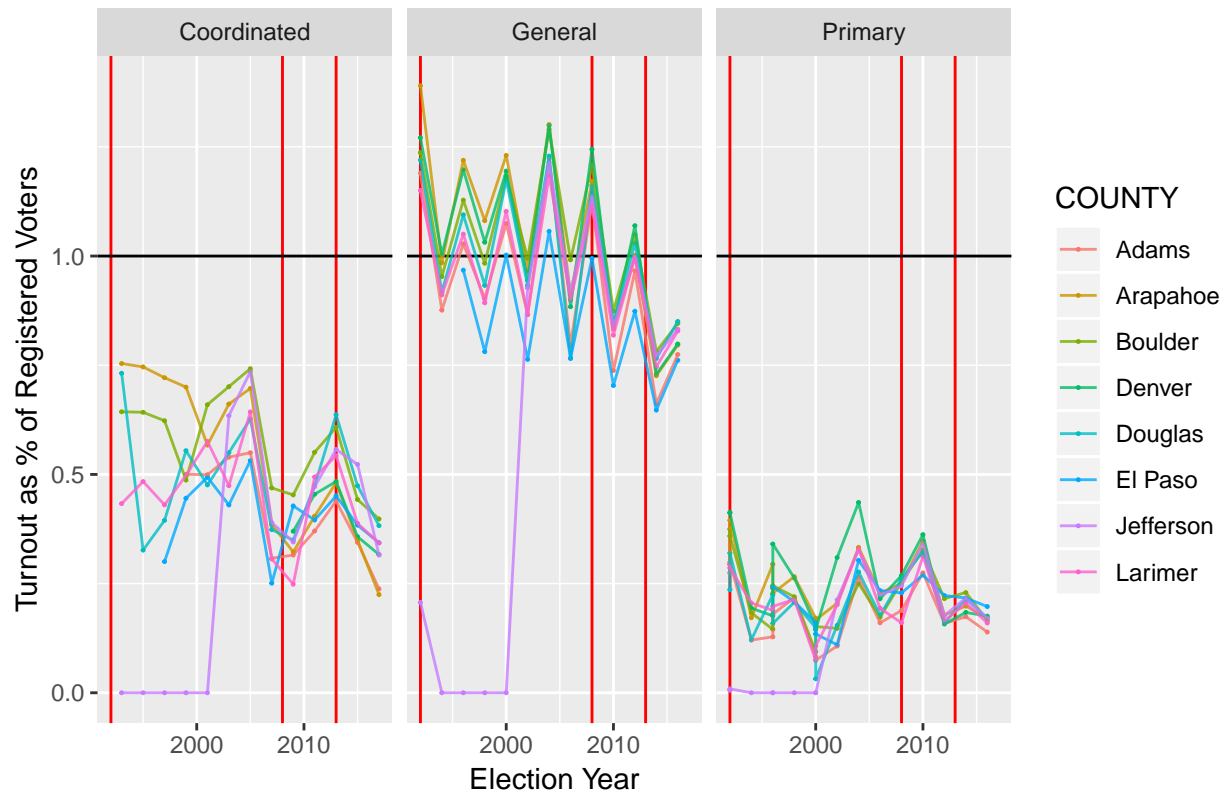
```

##A graph with only the largest 8 counties
big_seven_turnouts <- turnouts_by_county %>%
  filter(COUNTY %in% c("Jefferson", "El Paso", "Denver", "Arapahoe", "Adams", "Larimer", "Boulder", "Do

```

```
ggplot(big_seven_turnouts, aes(x = ELECTION_YEAR, y = TURNOUT, col = COUNTY)) +
  facet_wrap(facets = "ELECTION_TYPE") +
  geom_vline(xintercept = 1992, col = "red") +
  geom_vline(xintercept = 2008, col = "red") +
  geom_vline(xintercept = 2013, col = "red") +
  geom_hline(yintercept = 1) +
  geom_point(size = .2) +
  geom_line(alpha = .8) +
  labs(title = "Turnout in Colorado Elections for the Largest 8 Counties, 1992-2017", x = "Election Year")
```

Turnout in Colorado Elections for the Largest 8 Counties, 1992–2017



#Note what happens to Jefferson!

```
##Now the same graph, but with the rest of the counties included in grayscale
grayscale_turnouts <- turnouts_by_county %>%
  mutate(BIG_SEVEN <- ifelse(COUNTY %in% c("Jefferson", "El Paso", "Denver", "Arapahoe", "Adams", "La...

names(grayscale_turnouts)[8] <- "BIG_EIGHT"

ggplot(grayscale_turnouts, aes(x = ELECTION_YEAR, y = TURNOUT, col = BIG_EIGHT)) +
  scale_color_manual(values = c("dark orange", "grey")) +
  facet_wrap(facets = "ELECTION_TYPE") +
  geom_vline(xintercept = 1992, col = "red") +
  geom_vline(xintercept = 2008, col = "red") +
  geom_vline(xintercept = 2013, col = "red") +
  geom_hline(yintercept = 1) +
  geom_point(aes(shape = BIG_EIGHT)) +
```



```
labs(title = "Turnout in Colorado Elections on a County Level, 1992-2017", x = "Election Year", y =
```

