

Final Project*

PSTAT 231

Villaseñor-Derbez J.C. / 8749749

Contents

1. What makes voter behavior prediction (and thus election forecasting) a hard problem?
2. What was unique to Nate Silver's approach in 2012 that allowed him to achieve good predictions?
3. What went wrong in 2016? What do you think should be done to make future predictions better?
4. Remove summary rows from election.raw data:
 - Federal-level summary into a election_federal.
 - State-level summary into a election_state.
 - Only county-level data is to be in election.

```
election_raw <- read.csv(here("data", "election", "election.csv")) %>%
  as_tibble()

census_meta <- read.csv(here("data", "census", "metadata.csv"), sep = ";") %>%
  as_tibble()

census <- read.csv(here("data", "census", "census.csv")) %>%
  as_tibble() %>%
  mutate(CensusTract = as.factor(CensusTract))

election_federal <- election_raw %>%
  filter(fips == "US")

election_state <- election_raw %>%
  filter(state != "US", is.na(county))

election <- election_raw %>%
  filter(!is.na(county))
```

5. How many named presidential candidates were there in the 2016 election? Draw a bar chart of all votes received by each candidate

There were 31 explicitly mentioned presidential candidates, plus a category of **None of these acandidates**. Figure 1 shows the votes (on a \log_{10} -scale) that each candidate received.

```
election_federal %>%
  group_by(candidate) %>%
  summarize(votes = sum(votes, na.rm = T)) %>%
  ungroup() %>%
  mutate(candidate = fct_reorder(.f = candidate, .x = votes)) %>%
  ggplot(aes(x = candidate, y = votes)) +
  geom_col() +
  coord_flip() +
```

*Code available on GitHub at: https://github.com/jcvdav/PSTAT231/tree/master/final_project

```
scale_y_continuous(trans = "log10") +  
labs(x = "Candidate", y = "Votes (log-10 Scale)")
```

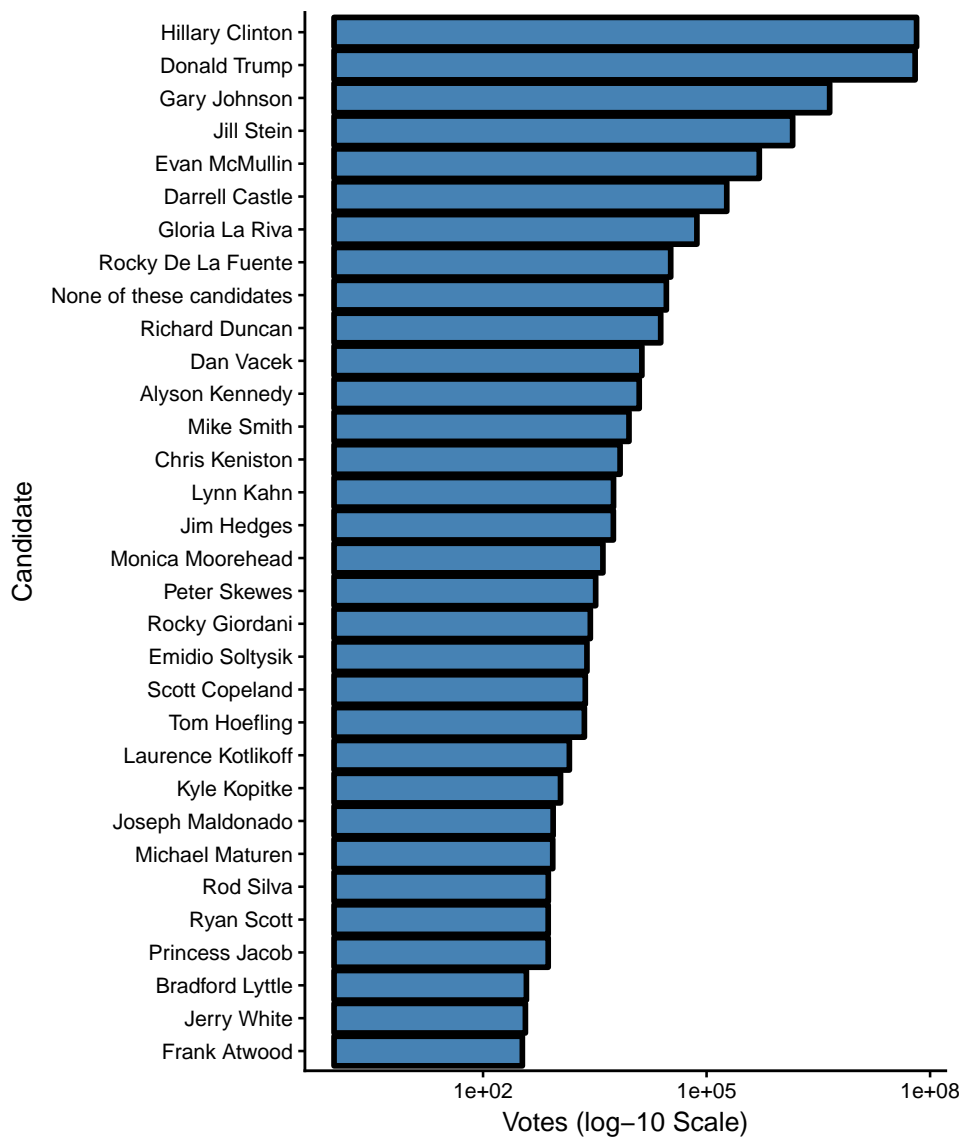


Figure 1: Number of votes that each presidential candidate received in the 2018 Presidential Elections.