Early Senate Polls

Chris Park

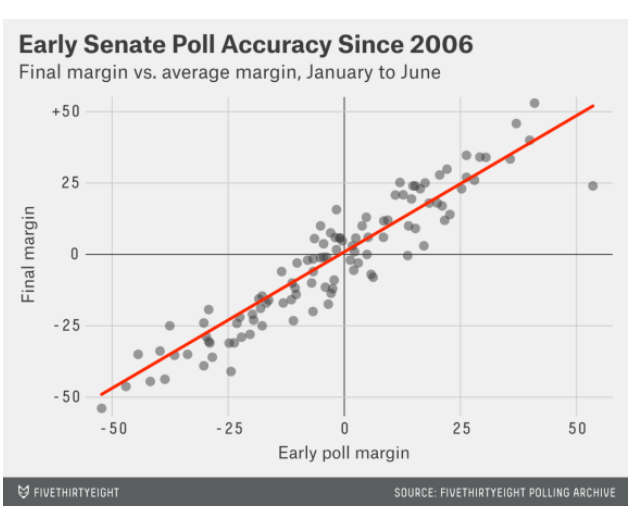
10/21/2020

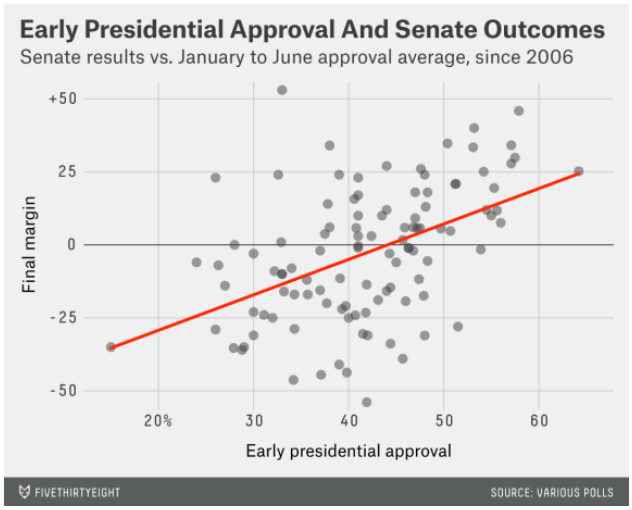
Article Information:

Article Title: Early Senate Polls Have Plenty to Tell Us About November

Click for link: *ARTICLE HYPERLINK*

Here are the two visual graphics that I will be replicating:





Harry Enton, the author of this article, did an analysis by collecting the average of all the polls for 107 races since 2006. He also collected the president's approval rating in the first half of the year for comparison.

Visual Graphic 1: Early Senate Poll Accuracy Since 2006

In this first visual graphic, Enton was proving how early polls do a pretty good job of predicting the final vote margin. Through this, he presents the average error between the early polls. Enton later explains that the president's party's candidate won 83% of the time he or she led in the early poll average and lost 88 percent of the time when he or she trailed.

First, I isolated the poll_average column and election_result column. The poll_average is the "Early poll margin" and the election_result is the "Final poll margin." I also arranged the poll_average column from smallest to largest for my own visual purposes (the x-axis goes from least to greatest). There was minimal data wrangling involved.

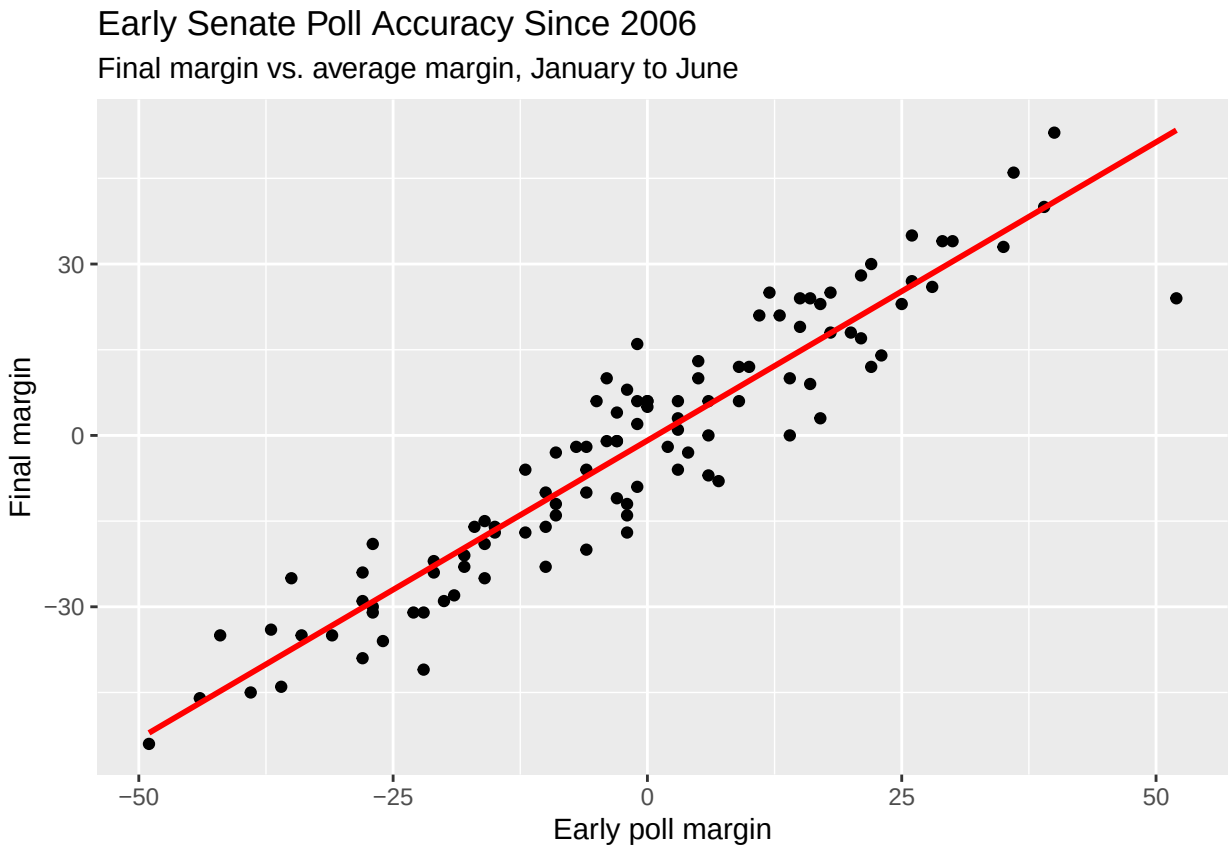
```
senate_accuracy = senate_polls %>%
  select(poll_average, election_result) %>%
  arrange(desc(-poll_average))
senate_accuracy
```

```
## # A tibble: 107 x 2
##    poll_average election_result
##           <int>           <int>
##  1          -49             -54
##  2          -44             -46
##  3          -42             -35
##  4          -39             -45
##  5          -37             -34
##  6          -36             -44
##  7          -35             -25
##  8          -34             -35
##  9          -31             -35
## 10          -28             -39
## # ... with 97 more rows
```

Next, I made a scatter plot with a line of best fit by making my x-axis the poll_average (later renamed to Early poll margin) and my y-axis the election_result (later renamed to Final margin).

```
senate_accuracy_plot = senate_accuracy %>%
  ggplot(aes(x = poll_average, y = election_result)) +
  geom_point() +
  geom_smooth(colour = "red", method = "lm", se = FALSE) +
  xlab("Early poll margin") +
  ylab("Final margin") +
  labs(
    title = "Early Senate Poll Accuracy Since 2006",
    subtitle = "Final margin vs. average margin, January to June"
  )
senate_accuracy_plot
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Visual Graphic 2: Early Presidential Approval And Senate Outcomes

In this second graphic, Enton was shows how a president's approval rating isn't as strongly tied to the ultimate result. The data is much more scattered in this graphic compared to the first graphic. Through this, he later explains how the 58 races in the sample where the president's statewide approval rating was less than 43 percent, the president's party candidate lost. However, when a president's approval rating was greater than 43%, the president's candidate lost 35 percent of races.

First, I isolated the presidential_approval column and election_result column. The presidential_approval is the "Early presidential approval rating" and the election_result is the "Final poll margin." I also arranged the presidential_approval column from smallest to largest for my own visual purposes (the x-axis goes from least to greatest). There was minimal data wrangling involved.

```
presidential_senate = senate_polls %>%
  select(presidential_approval, election_result) %>%
  arrange(desc(-presidential_approval))
presidential_senate
```

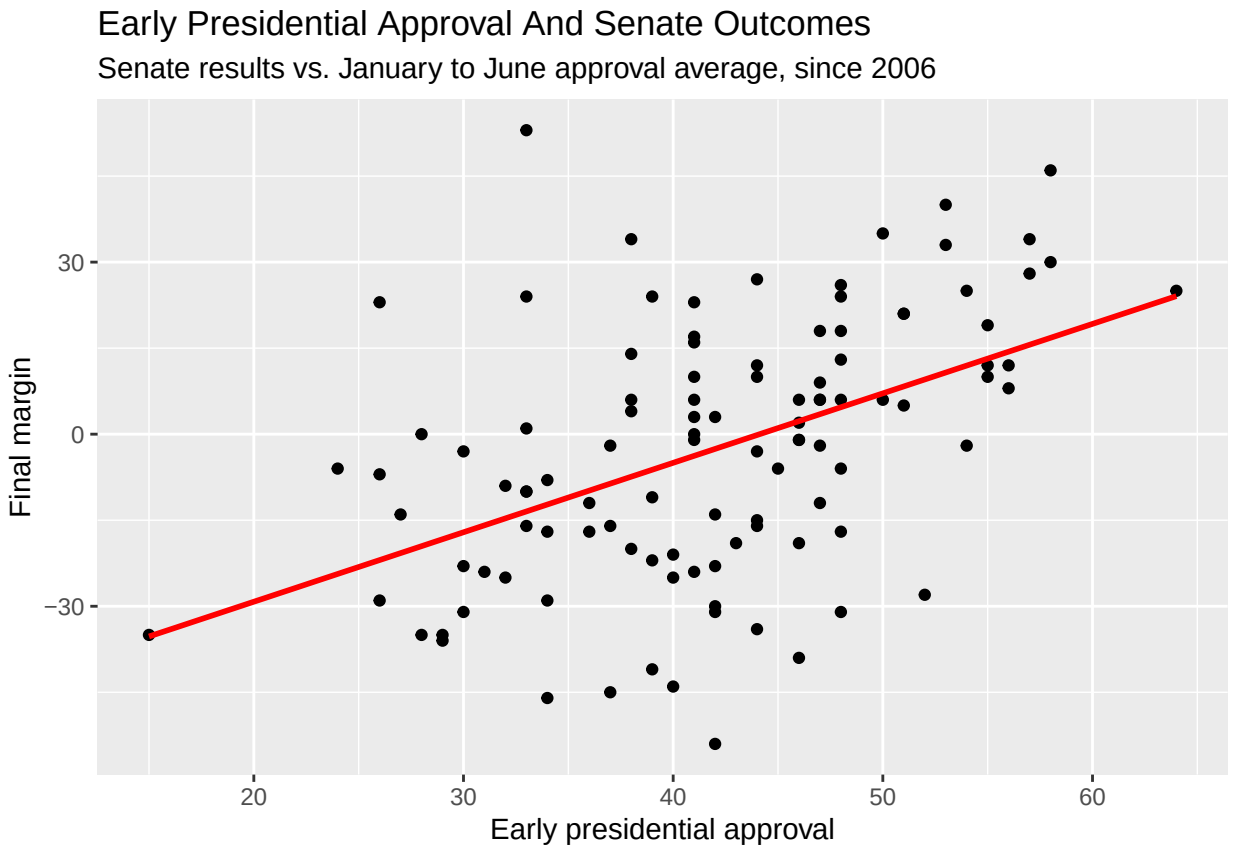
```
## # A tibble: 107 x 2
```

```
##    presidential_approval election_result
##                    <int>           <int>
##  1                    15             -35
##  2                    24              -6
##  3                    26              -7
##  4                    26             -29
##  5                    26              23
##  6                    27             -14
##  7                    28               0
##  8                    28             -35
##  9                    29             -35
## 10                    29             -36
## # ... with 97 more rows
```

Next, I made a scatter plot with a line of best fit by making my x-axis the presidential_approval (later renamed to Early presidential approval rating) and my y-axis the election_result (later renamed to Final margin).

```
presidential_senate_plot = presidential_senate %>%
  ggplot(aes(x = presidential_approval, y = election_result)) +
  geom_point() +
  geom_smooth(colour = "red", method = "lm", se = FALSE) +
  xlab("Early presidential approval") +
  ylab("Final margin") +
  labs(
    title = "Early Presidential Approval And Senate Outcomes",
    subtitle = "Senate results vs. January to June approval average, since 2006"
  )
presidential_senate_plot
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Ultimately, a president's approval rating is far from a perfect predictor of how Sente races will turn out. Polls are better predictors than presidential approval. However, implementing data from both will result in the best prediction according to Enten.
