

Final Project

Devin Barkey

12/1/2019

Predicting the 2020 Presidential Election Using Education Data

One of the leading story lines following the 2016 presidential election and the 2018 midterms has been the increasing divide between non-college-educated voters and those with a bachelor's degree.¹

Given education played a critical role four years ago, could we use education to predict the 2020 presidential election? Below, we built a model based on the 2016 election and then predicted county-level margins for five swing states using 2018 American Community Survey data as a proxy for 2020 conditions.

```
#Import data provided
county <- read.csv("county.csv")
election2016 <- read.csv("election2016.csv")

#Import 2018 American Community Survey 1 Year Estimate Educational Attainment Data
acs2018 <- read.csv("acs2018_Education_by_County.csv")

#Trim and edit ACS data
acs2018 <- acs2018[seq(2,ncol(acs2018)), c("NAME", "S1501_C01_001E", "S1501_C01_006E",
                                           "S1501_C01_005E", "S1501_C01_015E")]
acs2018$S1501_C01_001E <- as.numeric(as.character(acs2018$S1501_C01_001E))
acs2018$S1501_C01_006E <- as.numeric(as.character(acs2018$S1501_C01_006E))
acs2018$S1501_C01_005E <- as.numeric(as.character(acs2018$S1501_C01_005E))
acs2018$S1501_C01_015E <- as.numeric(as.character(acs2018$S1501_C01_015E))
names(acs2018) <- c("county", "total_pop_18_to_24", "total_pop_25_over",
                   "total_bach_18_to_24", "total_bach_25_over")
acs2018$total_pop <- acs2018$total_pop_18_to_24 + acs2018$total_pop_25_over
acs2018$pct_bach_deg <- (acs2018$total_bach_18_to_24 +
                        acs2018$total_bach_25_over)/acs2018$total_pop * 100

#Merge election and county datasets by county ID
election2016$pct_margin <- (election2016$votes_dem_16 -
                            election2016$votes_gop_16)/election2016$votes_total_16
education2016 <- merge(county, election2016, by = "FIPS", sort = TRUE)
colnames(education2016)[3] <- "pct_bach_deg"
```

The Education-only Model

```
# Build model using 2016 election results and county-level education data from 2011-2015
#and predict 2020 margins
states <- c("FL", "MI", "NC", "PA", "WI")
state.names = c("Florida", "Michigan", "North Carolina", "Pennsylvania", "Wisconsin")
```

¹See <https://www.economist.com/graphic-detail/2019/11/11/poorly-educated-voters-hold-the-keys-to-the-white-house>;
<https://www.brookings.edu/blog/brown-center-chalkboard/2019/06/03/8-reasons-why-education-may-be-pivotal-in-the-2020-election-and-beyond>;
<https://fivethirtyeight.com/features/americas-electoral-map-is-changing/>

```

winner <- rep(NA, 5)
names(winner) <- states

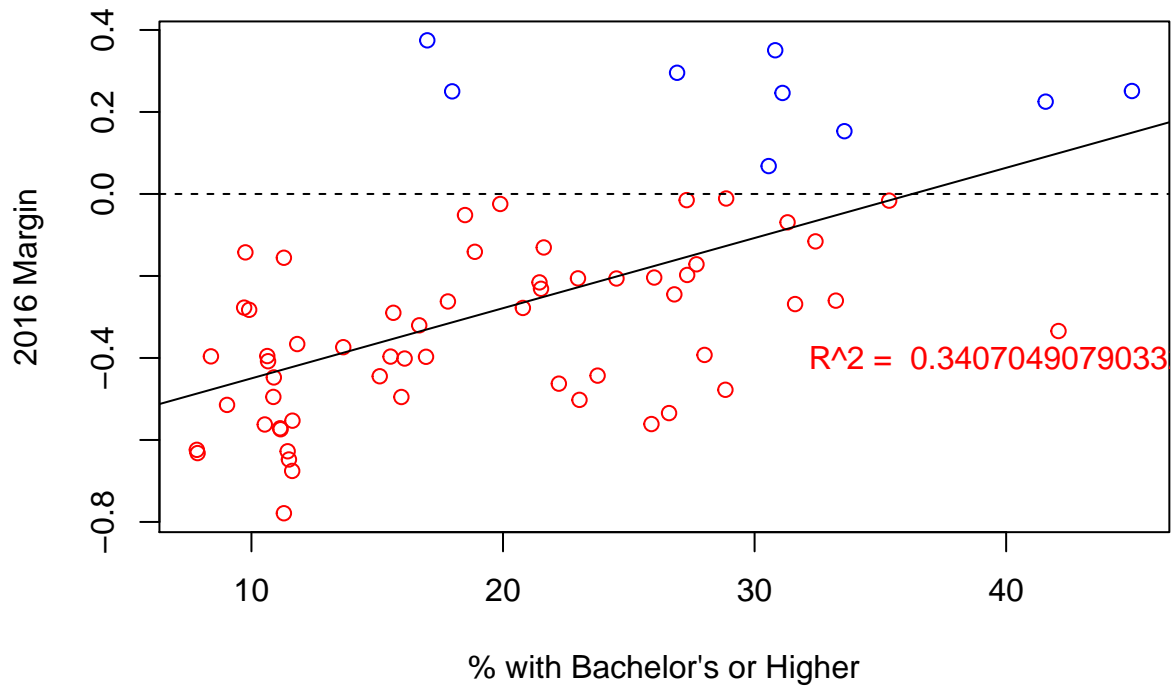
for (i in 1:5){
  #Build and plot model for state
  x = subset(education2016, subset = (education2016$state == states[i]))
  fit <- lm(pct_margin ~ pct_bach_deg, data = x)
  plot(x$pct_bach_deg, x$pct_margin, main = paste(states[i], " 2016 County Results"),
        xlab = "% with Bachelor's or Higher", ylab = "2016 Margin",
        col = ifelse(x$pct_margin > 0, "blue", "red"))
  abline(fit)
  abline(h = 0, lty = "dashed")
  rtext <- paste("R^2 = ", as.character(summary(fit)$r.squared))
  text(40, -.4, rtext, col = "red")

  #Plot predicted margins
  y = subset(acs2018, subset = (grepl(state.names[i], acs2018$county)))
  y$pred_marg <- predict(fit, newdata = y)
  plot(y$pct_bach_deg, y$pred_marg, main = paste(states[i], " Predicted 2020 County Results"),
        xlab = "% with Bachelor's or Higher",
        ylab = "Predicted Margin", col = ifelse(y$pred_marg > 0, "blue", "red"))
  abline(h = 0, lty = "dashed")

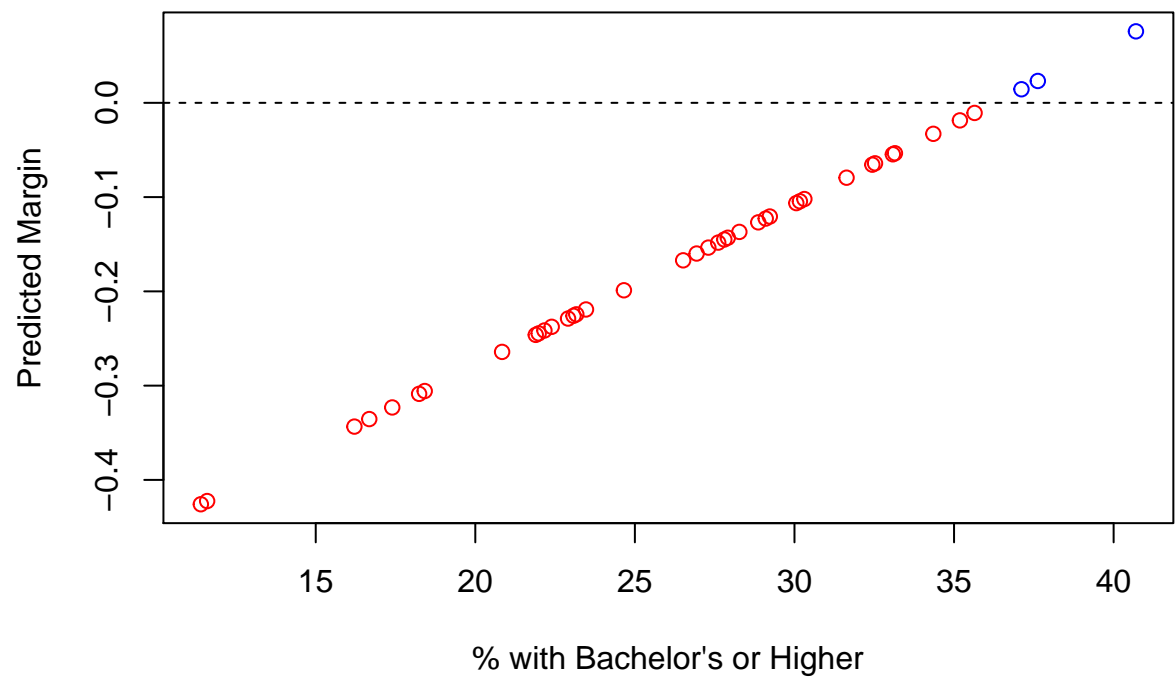
  #Weighting counties by total adult population from ACS data,
  #determine winner of each state
  winner[i] <- ifelse(sum(y$total_pop * y$pred_marg) > 0, "D", "R")
}

```

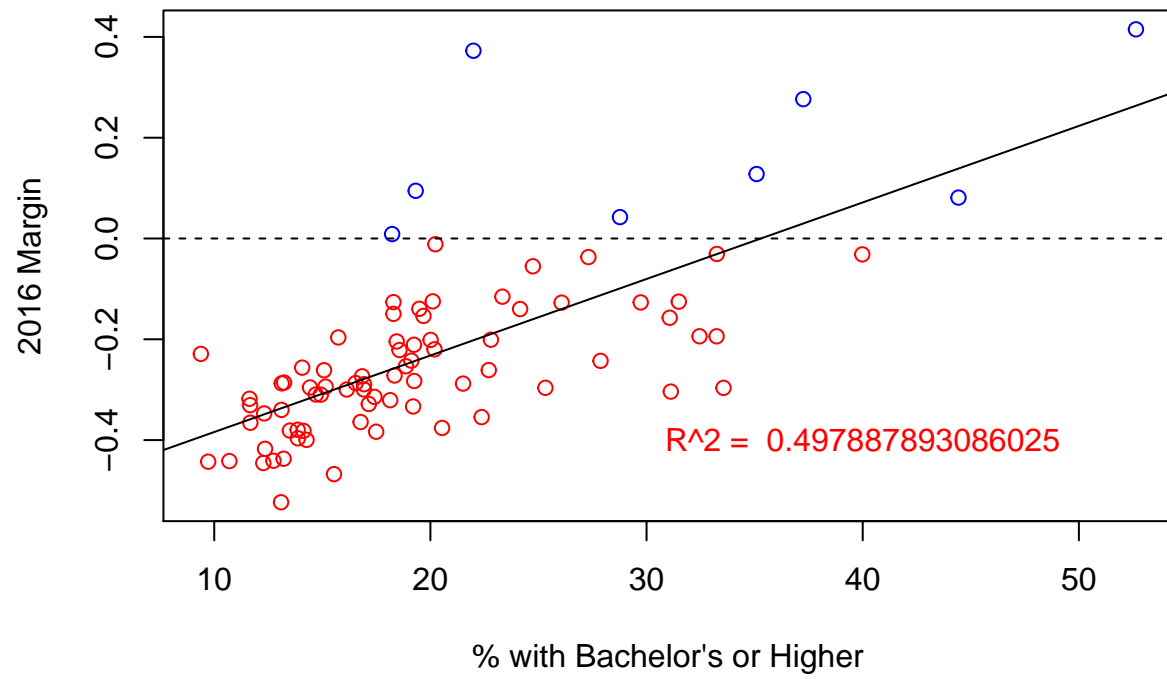
FL 2016 County Results



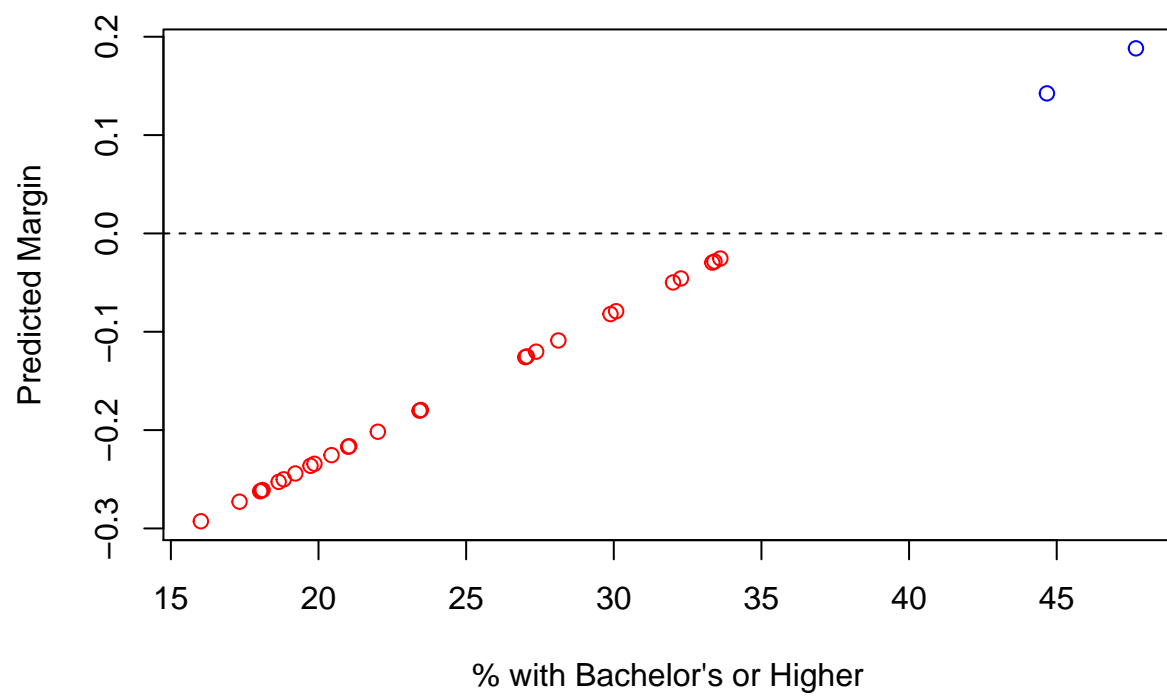
FL Predicted 2020 County Results



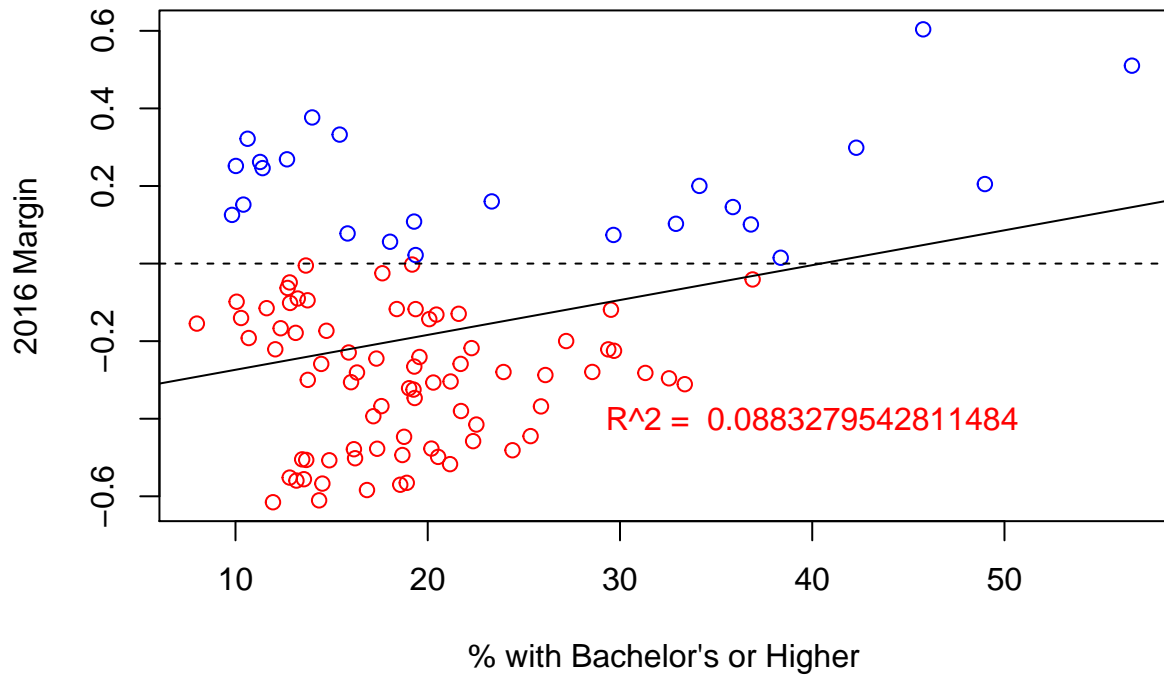
MI 2016 County Results



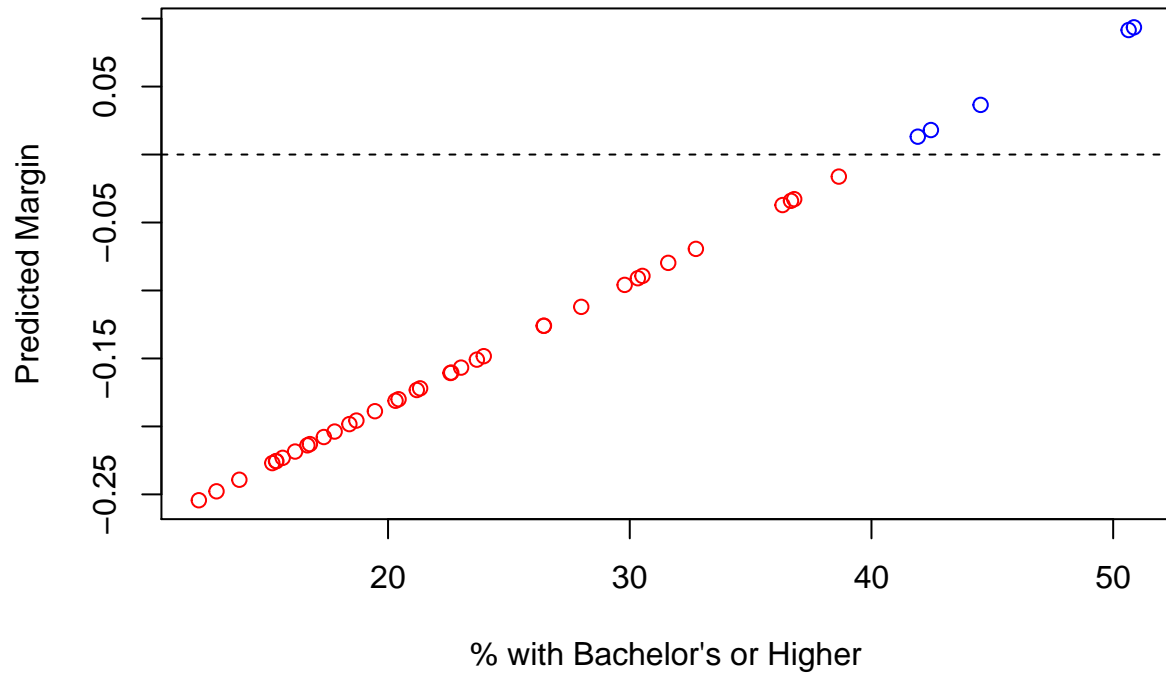
MI Predicted 2020 County Results



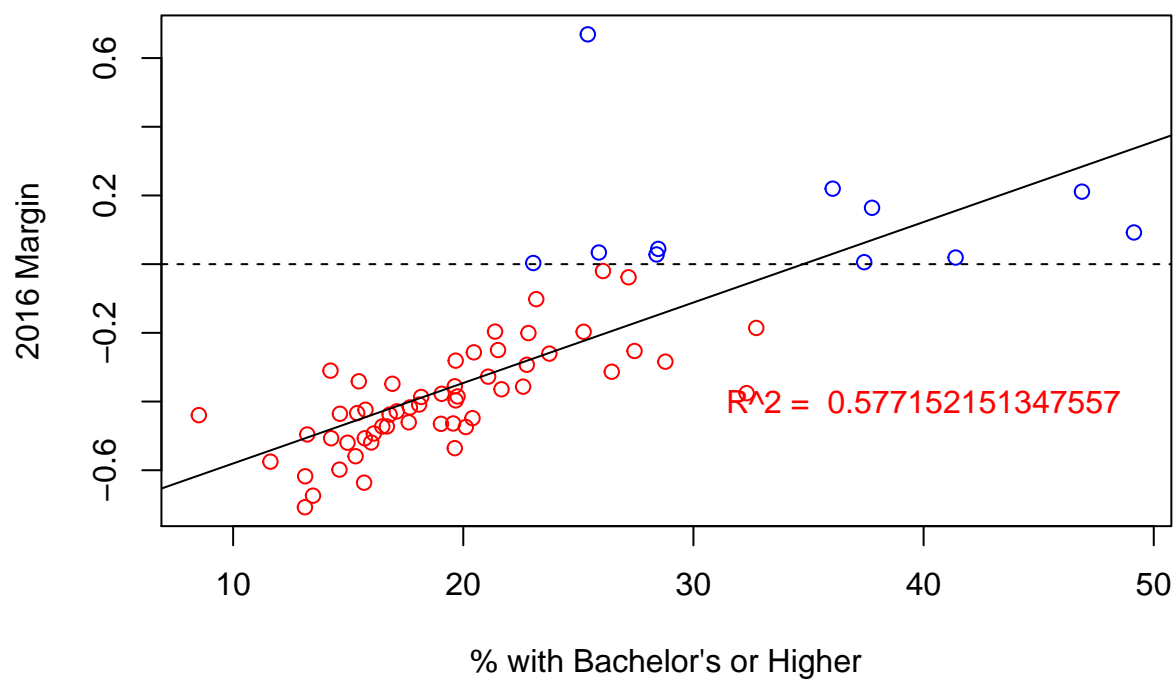
NC 2016 County Results



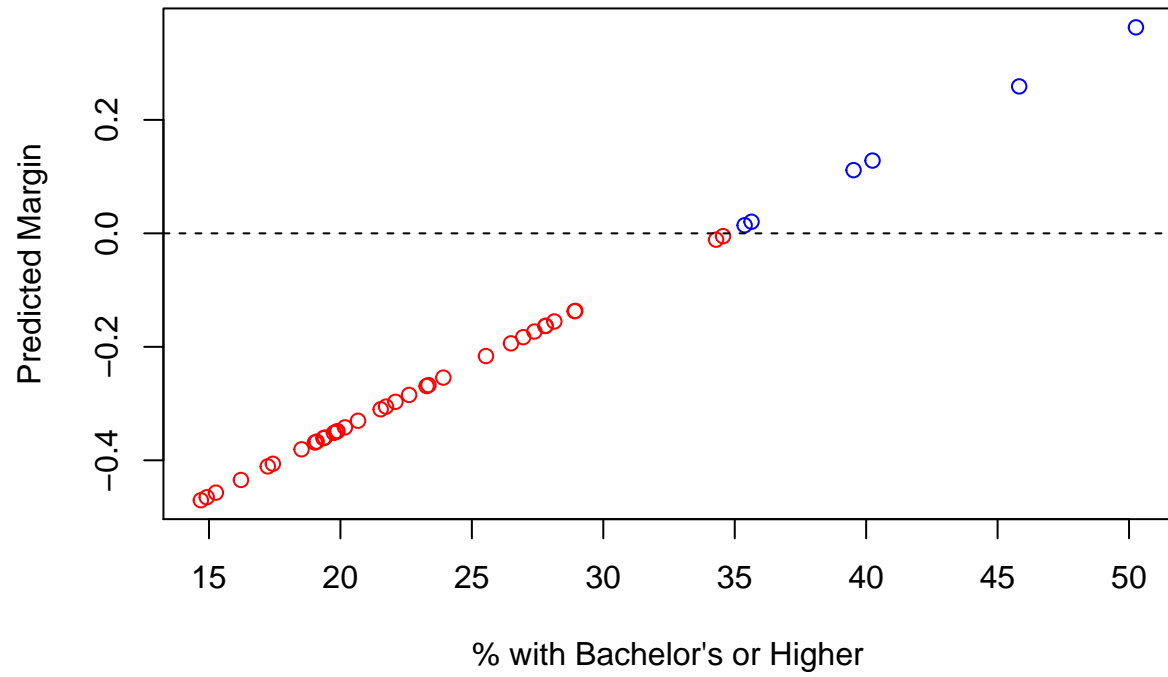
NC Predicted 2020 County Results



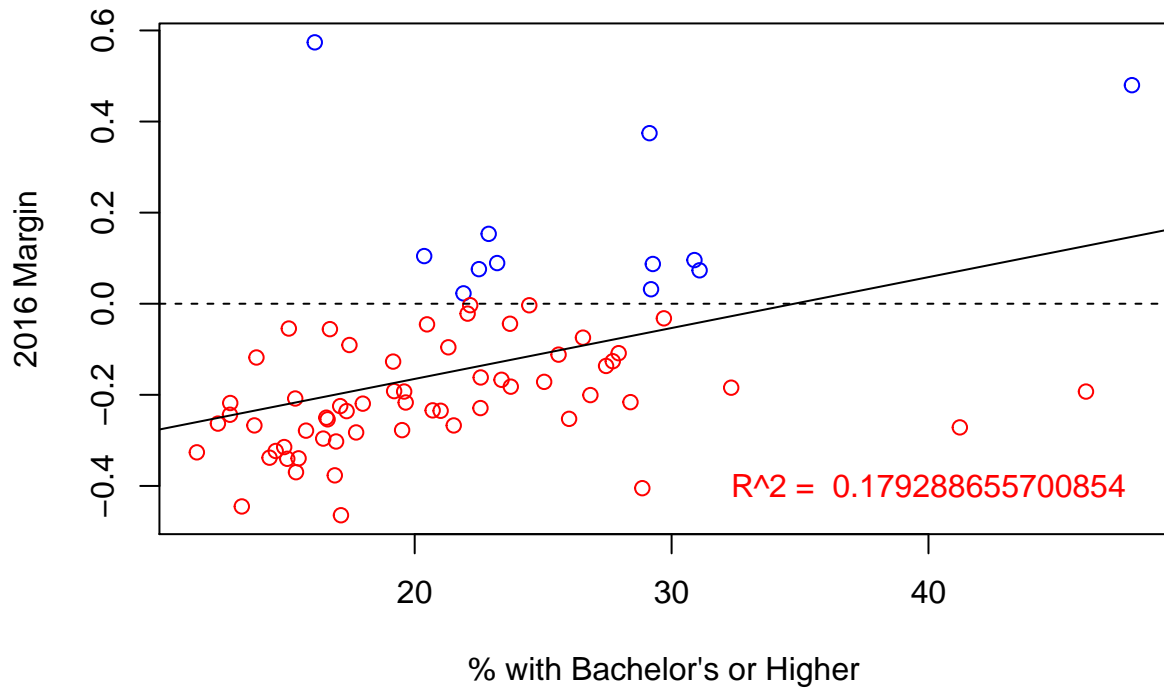
PA 2016 County Results



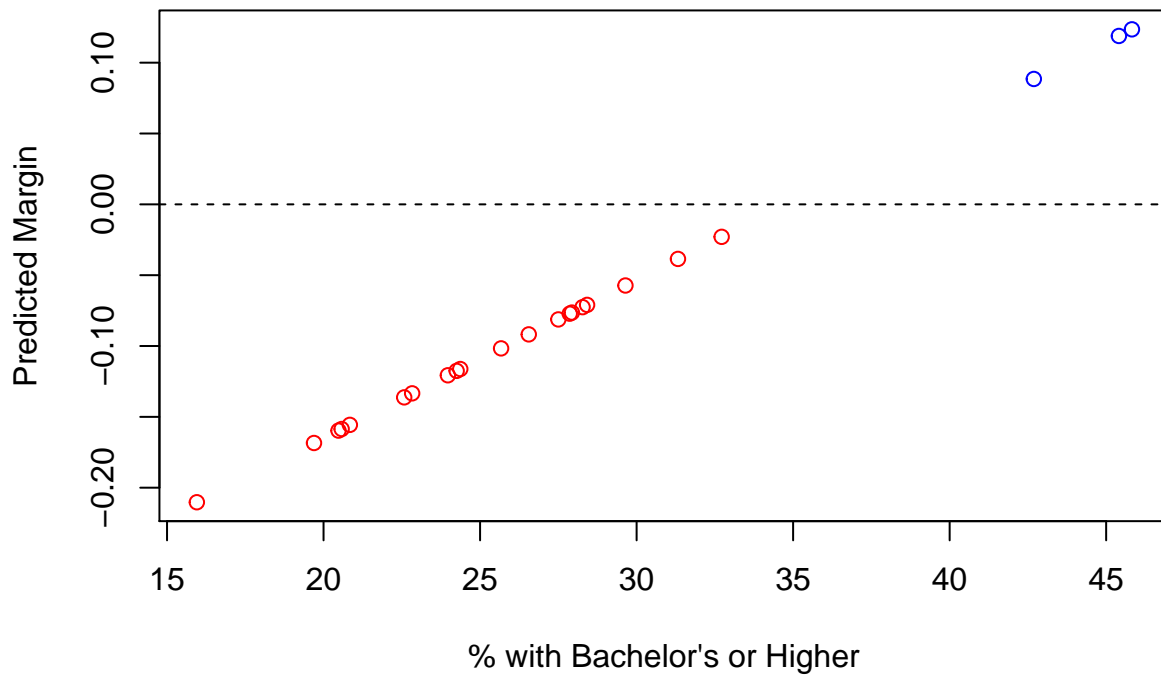
PA Predicted 2020 County Results



WI 2016 County Results



WI Predicted 2020 County Results



```
winner
```

```
## FL MI NC PA WI  
## "R" "R" "R" "R" "R"
```

Observations

We made the following observations from the results above:

If Education Is the Only Predictor...

Assuming a county's education levels is the sole predictor of its vote margin, 2020 appears bleak for Democrats. A Trump win in these states would yield 90 electoral votes and almost certainly lead to the President's reelection.

Varying Predictive Value

The R^2 values for the states modeled differed significantly, ranging from North Carolina's meager 0.088 to Pennsylvania's not-meaningless 0.577. We conjectured that this discrepancy might be due to other demographic factors that affect the vote margin in each state. For instance, the composition of the Democratic coalition in North Carolina or Wisconsin will be quite different from that of Pennsylvania.

More Questions than Answers

Because education did not prove to be a significant predictor in the majority of swing states tested, how would this factor interact with race and gender? How would this information affect the Democratic nomination in which “electability” is playing an outsized role?