

Problem Set 3

Paula Montano/Applied Stats/Quant Methods 1

Due: November 12, 2021

Instructions

- Please show your work! You may lose points by simply writing in the answer. If the problem requires you to execute commands in R, please include the code you used to get your answers. Please also include the `.R` file that contains your code. If you are not sure if work needs to be shown for a particular problem, please ask.
- Your homework should be submitted electronically on GitHub in `.pdf` form.
- This problem set is due before class on Friday November 12, 2021. No late assignments will be accepted.
- Total available points for this homework is 80.

In this problem set, you will run several regressions and create an add variable plot (see the lecture slides) in R using the `incumbents_subset.csv` dataset. Include all of your code.

Question 1

We are interested in knowing how the difference in campaign spending between incumbent and challenger affects the incumbent's vote share.

1. Run a regression where the outcome variable is `voteshare` and the explanatory variable is `difflog`.

```

1 incumbents_subset <- read.csv("https://raw.githubusercontent.com/ASDS-TCD
  /StatsI_Fall2021/main/datasets/incumbents_subset.csv")
2 summary(incumbents_subset)

1 voteshare_diff <- lm(voteshare ~ difflog, data = incumbents_subset)
2 summary(lm(voteshare ~ difflog, data = incumbents_subset))

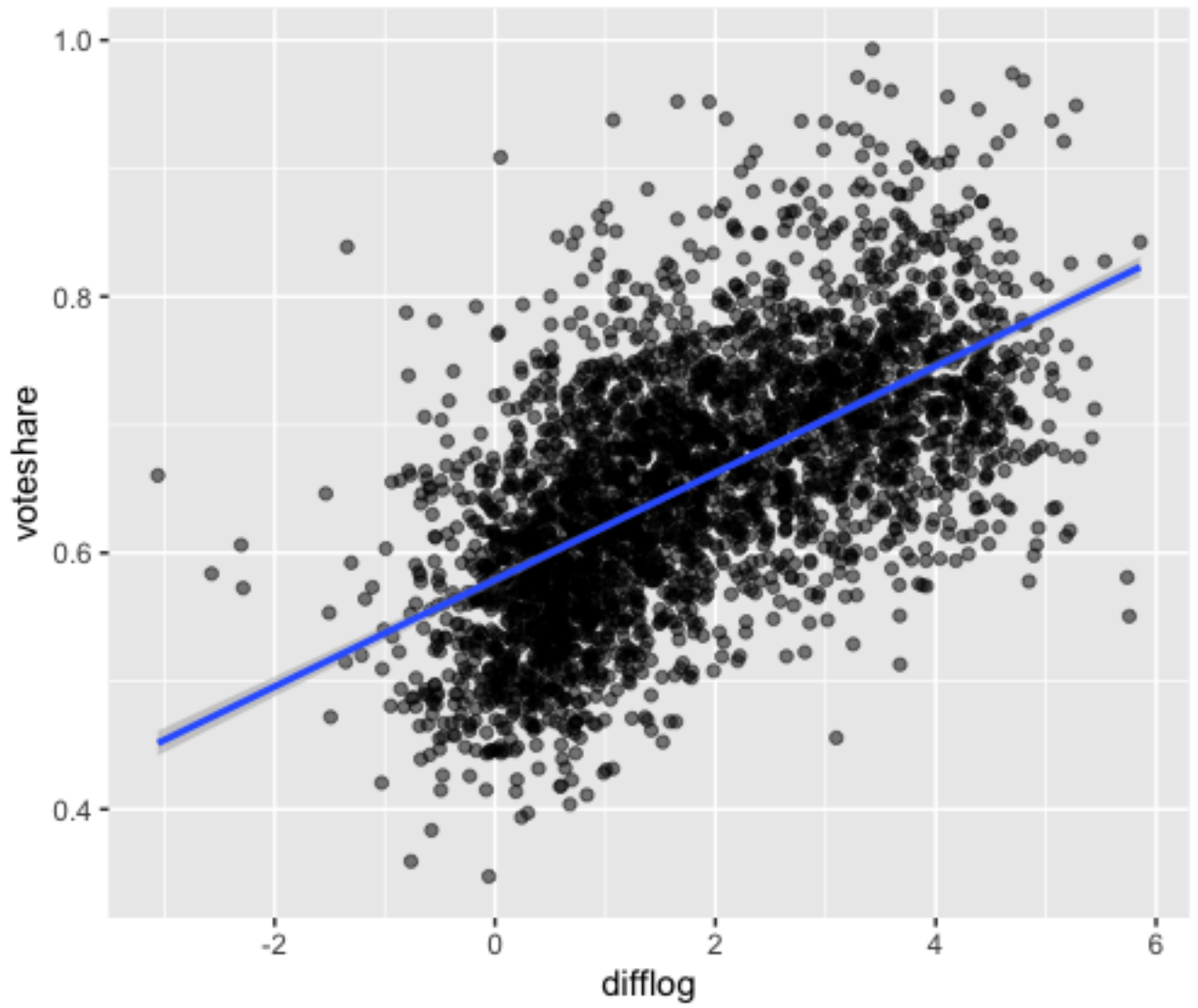
```

2. Make a scatterplot of the two variables and add the regression line.

```

1 ggplot(incumbents_subset, aes(difflog, voteshare)) +
2   geom_point(alpha = 0.5) +
3   geom_smooth(method = "lm")

```

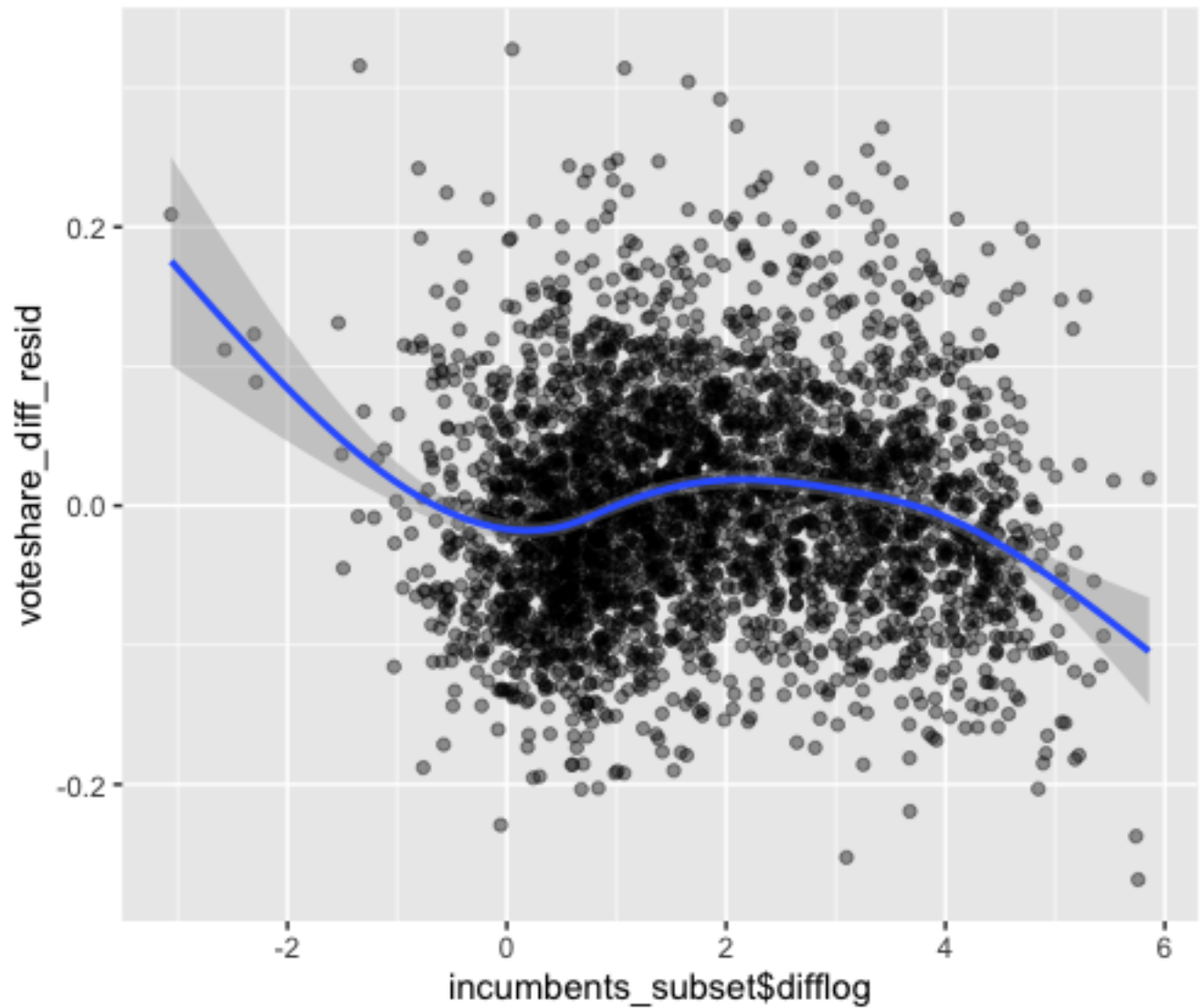


3. Save the residuals of the model in a separate object.

```

1 voteshare_diff_resid <- resid(voteshare_diff)
2
3 ## Plot residuals
4 ggplot(aes(incumbents_subset$difflog, voteshare_diff_resid), data = NULL)
5   +
6   geom_point(alpha = 0.4) +
7   geom_smooth()

```



4. Write the prediction equation.

```

1 ## 4. Prediction equation
2 ## Y= B0 (intercept: 0.579031) + B1 (difflog: 0.041666) * X
3 ## Y = voteshare (outcome variable)
4 ## X = the specific value on difflog (explanatory variable)
5 ## B1 = difflog coefficient

```

Question 2

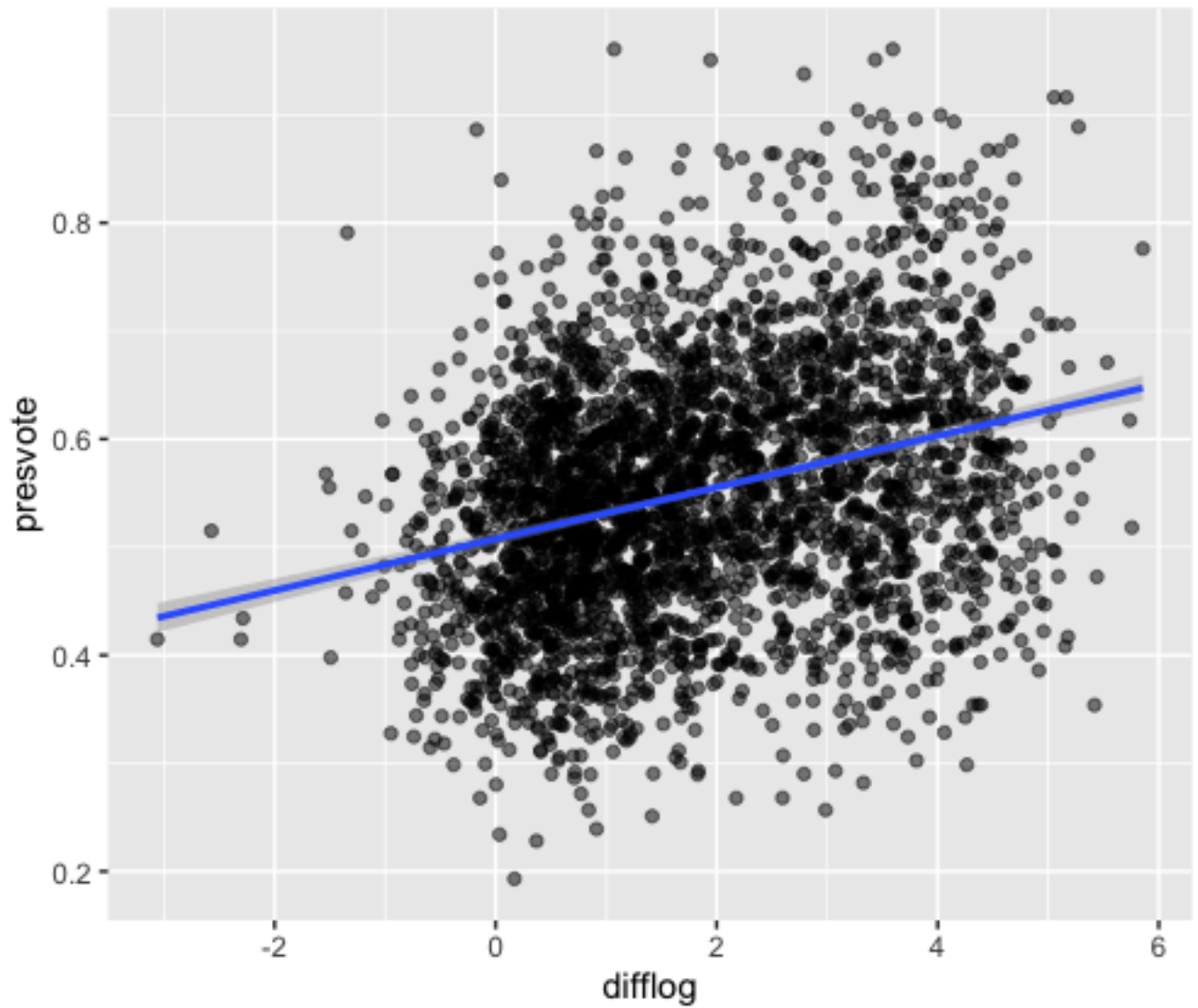
We are interested in knowing how the difference between incumbent and challenger's spending and the vote share of the presidential candidate of the incumbent's party are related.

1. Run a regression where the outcome variable is `presvote` and the explanatory variable is `difflog`.

```
1 presvote_diff <- lm(presvote ~ difflog, data = incumbents_subset)
2 summary(lm(presvote ~ difflog, data = incumbents_subset))
```

2. Make a scatterplot of the two variables and add the regression line.

```
1 ggplot(incumbents_subset, aes(difflog, presvote)) +
2   geom_point(alpha = 0.5) +
3   geom_smooth(method = "lm")
```



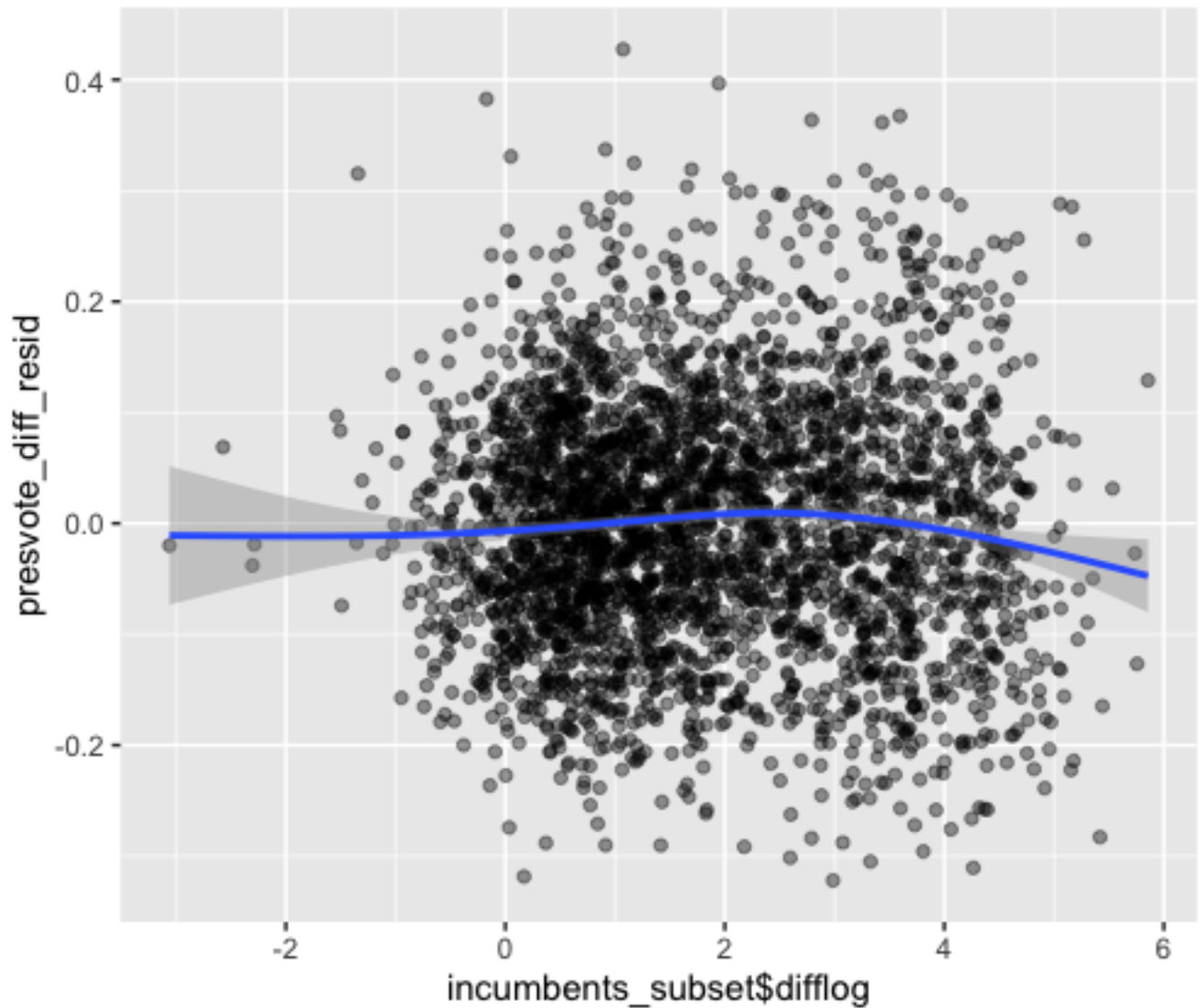
3. Save the residuals of the model in a separate object.

```
1 presvote_diff_resid <- resid(presvote_diff)
2
```

```

3 ## Plot residuals
4 ggplot(aes(incumbents_subset$difflog, presvote_diff_resid), data = NULL)
  +
5   geom_point(alpha = 0.4) +
6   geom_smooth()

```



4. Write the prediction equation.

```

1 ## Prediction equation
2 ## Y (presvote) = B0 (intercept: 0.507583) + B1 (difflog coefficient:
   0.023837) * x
3 ## B0 = (intercept: 0.507583)
4 ## Y = presvote outcome variable, value that we want to predict
5 ## B1 = difflog coefficient (0.023837) * x

```

Question 3

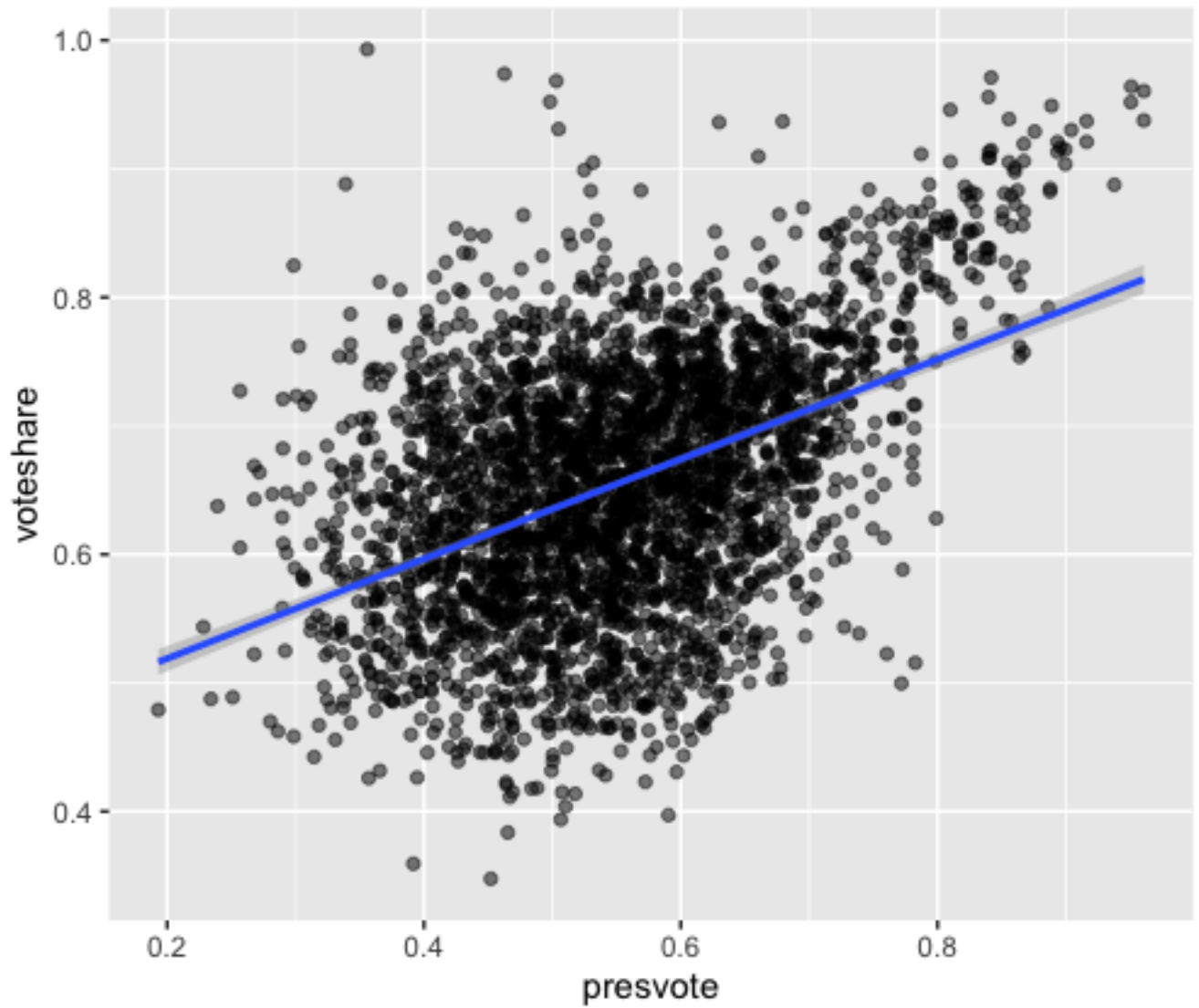
We are interested in knowing how the vote share of the presidential candidate of the incumbent's party is associated with the incumbent's electoral success.

1. Run a regression where the outcome variable is **voteshare** and the explanatory variable is **presvote**.

```
1 voteshare_presv <- lm(voteshare ~ presvote, data = incumbents_subset)
2 summary(lm(voteshare ~ presvote, data = incumbents_subset))
```

2. Make a scatterplot of the two variables and add the regression line.

```
1 ggplot(incumbents_subset, aes(presvote, voteshare)) +
2   geom_point(alpha = 0.5) +
3   geom_smooth(method = "lm")
```

3. Write the prediction equation.

```
1 ## 3. Prediction equation
```

```
2 ## Y (voteshare) = B0 (intercept: 0.441330) + B1 (presvote coefficient:  
   0.388018) * X  
3 ## Y = voteshare is outcome variable , value that we want to predict  
4 ## B0 = 0.441330intercept  
5 ## X = the specific value on presvote (explanatory variable)  
6 ## B1 = presvote coefficient (0.388018)
```

Question 4

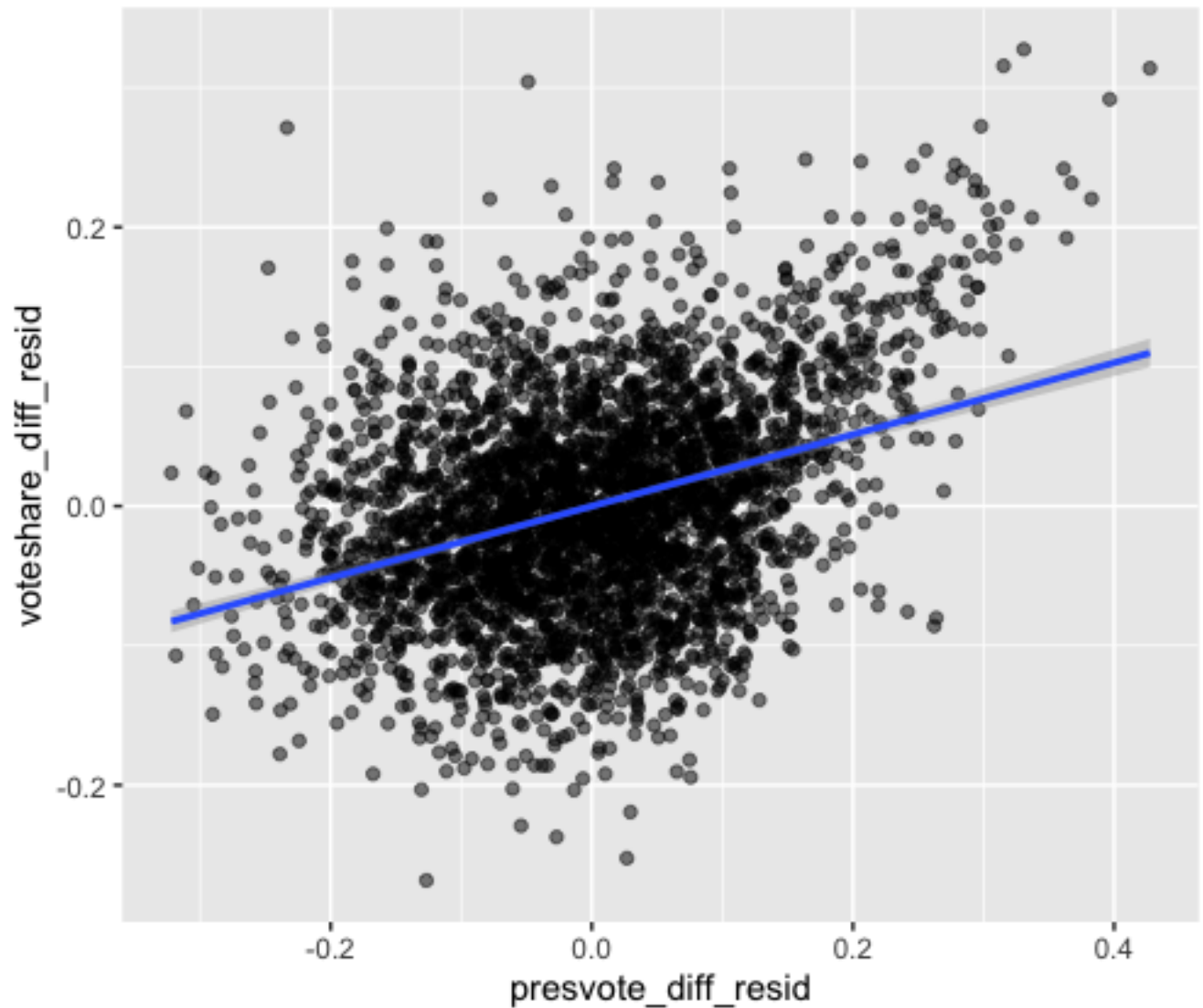
The residuals from part (a) tell us how much of the variation in **voteshare** is *not* explained by the difference in spending between incumbent and challenger. The residuals in part (b) tell us how much of the variation in **presvote** is *not* explained by the difference in spending between incumbent and challenger in the district.

1. Run a regression where the outcome variable is the residuals from Question 1 and the explanatory variable is the residuals from Question 2.

```
1 voteshareresid_presvresid <- lm(voteshare_diff_resid ~ presvote_diff_resid, data = incumbents_subset)
2 summary(lm(voteshare_diff_resid ~ presvote_diff_resid, data = incumbents_subset))
```

2. Make a scatterplot of the two residuals and add the regression line.

```
1 ggplot(incumbents_subset, aes(presvote_diff_resid, voteshare_diff_resid))
  +
2   geom_point(alpha = 0.5) +
3   geom_smooth(method = "lm")
```



3. Write the prediction equation.

```

1 ## 3. Prediction equation
2 ## Y (voteshare_diff_resid) = B0 (intercept: -0.00000000000000000486) + B1
   (presvote_diff_resid coefficient: 0.25687701270009788423) * X
3 ## Y = voteshare_diff_resid (outcome variable), value that we want to
   predict
4 ## X = is the specific value on the explanatory variable
5 ## B1 = presvote_diff_resid coefficient

```

Question 5

What if the incumbent's vote share is affected by both the president's popularity and the difference in spending between incumbent and challenger?

1. Run a regression where the outcome variable is the incumbent's `voteshare` and the explanatory variables are `difflog` and `presvote`.

```
1 voteshare ~ difflog + presvote <- lm(voteshare ~ difflog + presvote, data =  
  incumbents_subset)  
2 summary(lm(voteshare ~ difflog + presvote, data = incumbents_subset))
```

2. Write the prediction equation.

```
1 ## 2. Prediction equation  
2 ## UY (voteshare) = B0 (intercept: 0.4486442) + B1 (difflog coefficient  
  :0.035543) * X1 + B2 (presvote coefficient:0.2568770) * X2  
3 ## UY = voteshare (outcome variable), value that we want to predict  
4 ## B1 = difflog coefficient:0.035543 * X1  
5 ## B2 = presvote coefficient:0.256877 * X2
```

3. What is it in this output that is identical to the output in Question 4? Why do you think this is the case?

The output that is identical in question 4 is the residual standard error coefficient (0.07339). In question 4 the variation in voteshare is not explained by the difference

in spending between incumbent and challenger, while in question 5 the voteshare variation is explained by the difference between the incumbent and challenger.