Problem Set 3

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.

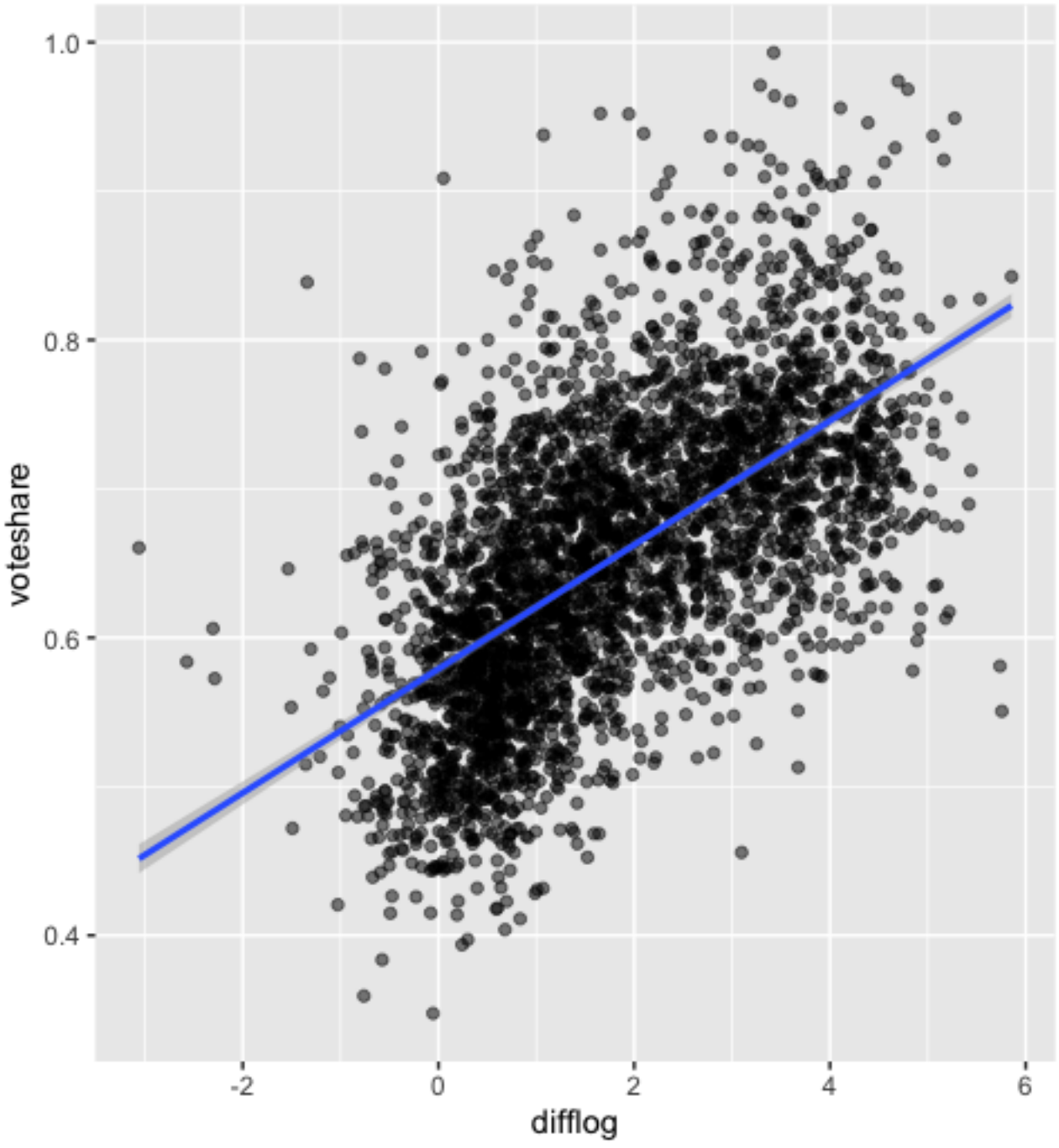
1. Run a regression where the outcome variable is `voteshare` and the explanatory variable is `difflog`.

```
#Load the data set incumbents
incumbents <- read_csv("https://raw.githubusercontent.com/jackoneillWHU/
    StatsI_Fall2021/main/datasets/incumbents_subset.csv")

#run regression with voteshare as outcome variable and difflog as
    explanatory variable
summary(lm( data = incumbents, voteshare ~ difflog))
lm1 <- lm( data = incumbents, voteshare ~ difflog)
```

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

```
#create a scatterplot with regression line
ggplot(data = incumbents, aes(x = difflog, y = voteshare)) +
  geom_point(alpha = 0.5) + #add a scatterplot
  geom_smooth(method = "lm") #add a linear regression line
```



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

```
#save residuals as new object
residuals(lm1)
incumbents$residuals <- residuals(lm1)
```

4. Write the prediction equation.

   The mathematical formula of the linear regression can be written as y= a +b*x or y = b0 + b1*x, where: b0 and b1 are known as the regression beta coefficients or parameters: b0 is the intercept of the regression line; that is the predicted value when x = 0. b1 is the slope of the regression line. For our regression the prediction equation is: y = 0.579031 + 0.041666*x

Question 2

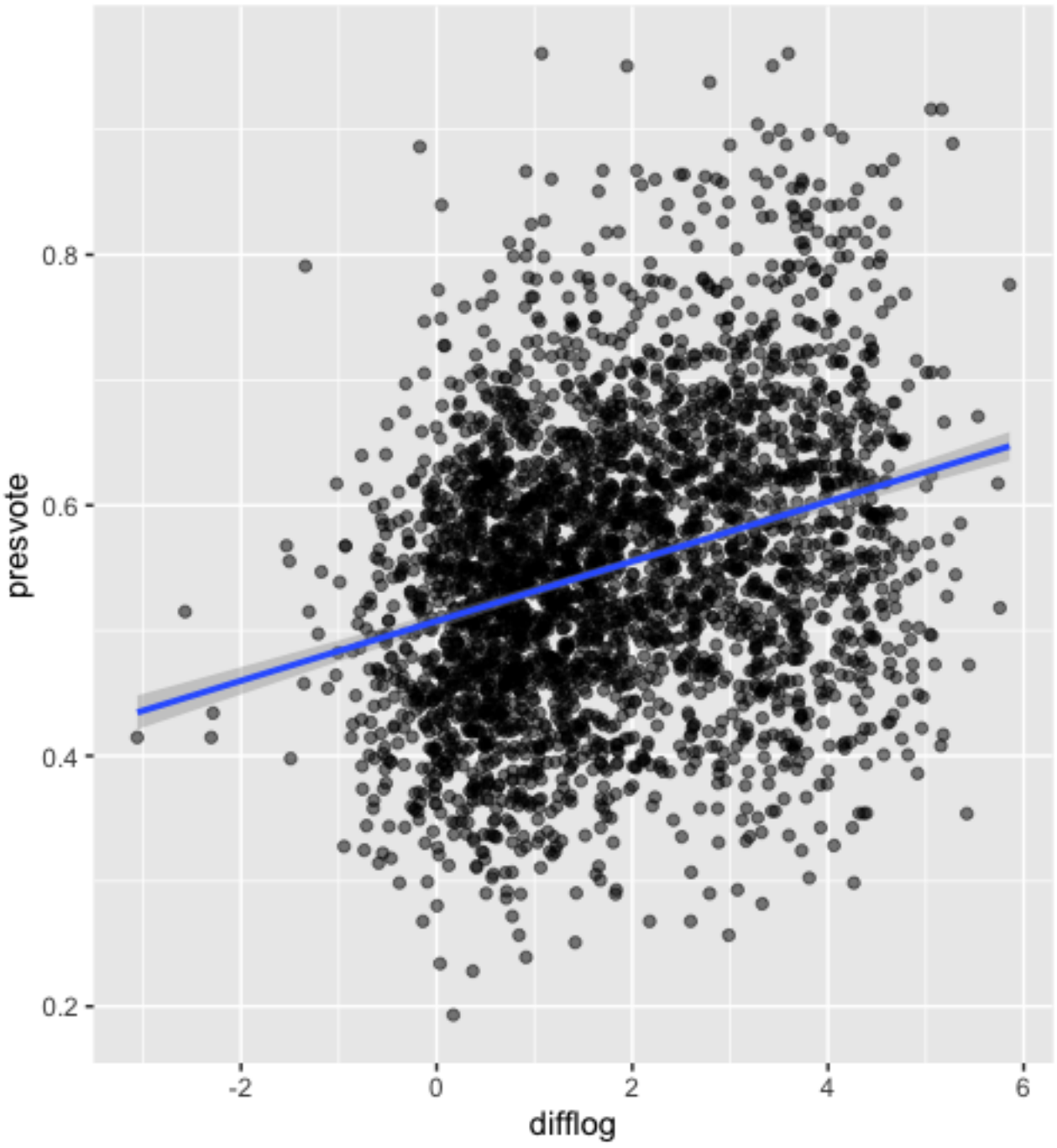
We are interested in knowing how the difference between incumbent and challenger's spending and the vote share of the presidential candidate of the incumbent's party are related.

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

```
#run regression with presvote as outcome variable and difflog as
    explanatory variable
summary(lm( data = incumbents, presvote ~ difflog))
lm2 <- lm( data = incumbents, presvote ~ difflog)
```

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

```
#run regression with presvote as outcome variable and difflog as
    explanatory variable
summary(lm( data = incumbents, presvote ~ difflog))
lm2 <- lm( data = incumbents, presvote ~ difflog)
```



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

```
#save residuals as new object
residuals(lm2)
incumbents$residuals2 <- residuals(lm2)
```

4. Write the prediction equation.

   For this regression the prediction equation is: y = 0.507583 + 0.023837*x

Question 3

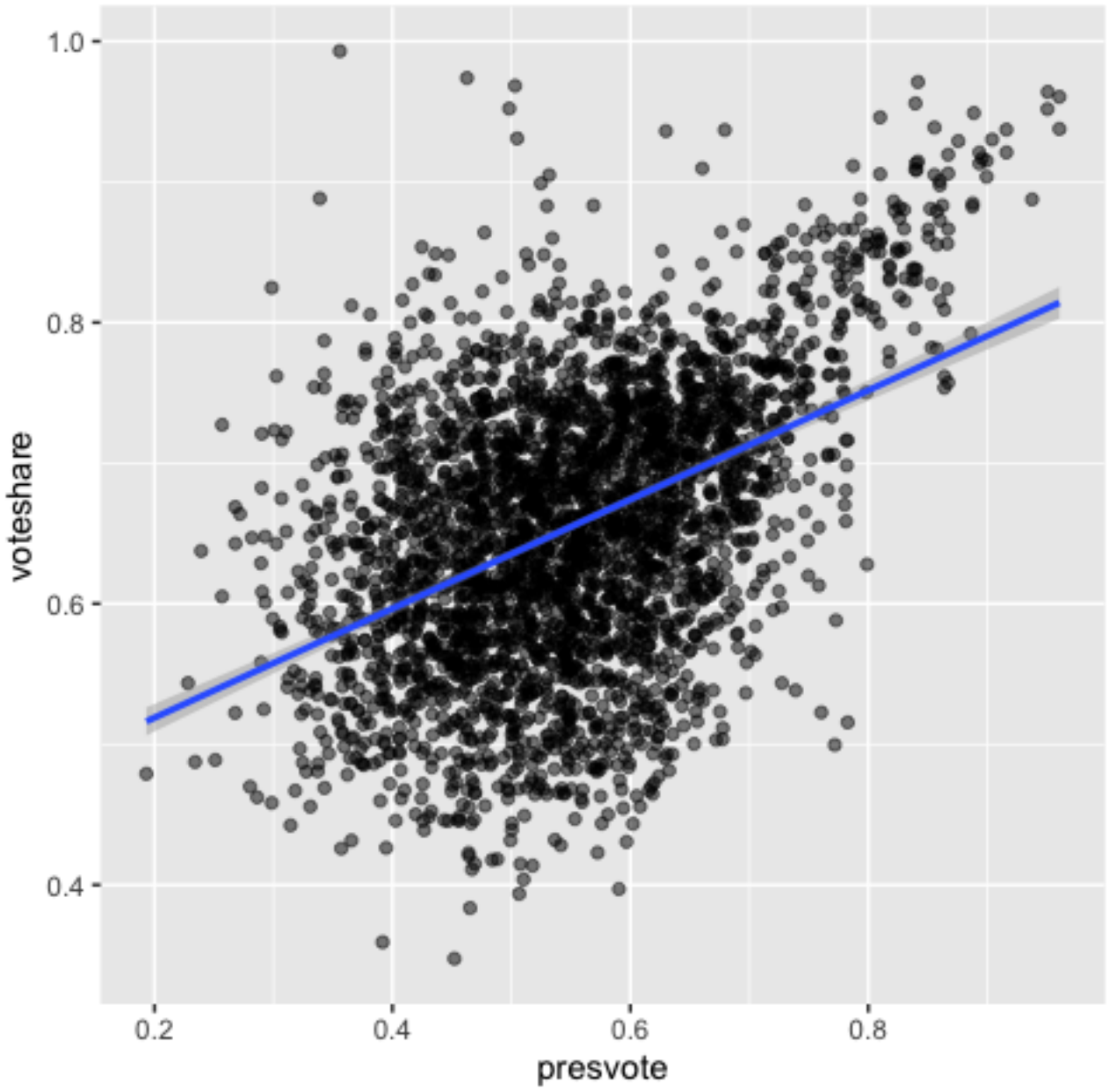
We are interested in knowing how the vote share of the presidential candidate of the incumbent's party is associated with the incumbent's electoral success.

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

```
#run regression with voteshare as outcome variable and presvote as
    explanatory variable
summary(lm( data = incumbents, voteshare ~ presvote))
lm3 <- lm( data = incumbents, voteshare ~ presvote)
```

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

```
#create a scatterplot with regression line
ggplot(data = incumbents, aes(x = presvote, y = voteshare)) +
  geom_point(alpha = 0.5) + #add a scatterplot
  geom_smooth(method = "lm") #add a linear regression line
```



3. Write the prediction equation.

   For this regression the prediction equation is: y = 0.441330 + 0.388018*x

Question 4

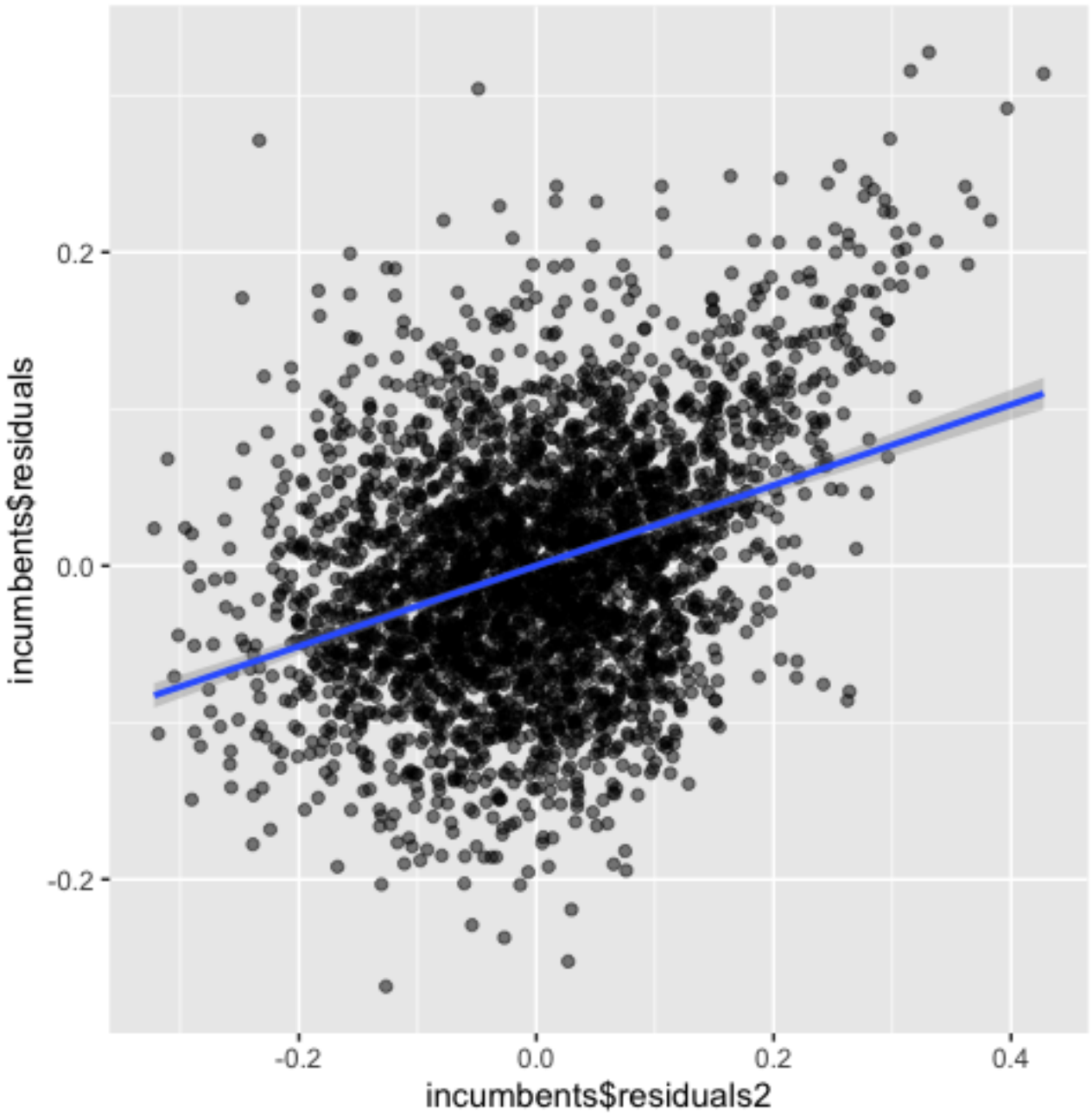
The residuals from part (a) tell us how much of the variation in `voteshare` is *not* explained by the difference in spending between incumbent and challenger. The residuals in part (b) tell us how much of the variation in `presvote` is *not* explained by the difference in spending between incumbent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

```
#run regression with incumbents$residuals as outcome variable and
    incumbents$residuals2 as the explanatory variable
summary(lm(data=incumbents, incumbents$residuals ~incumbents$residuals2))
```

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

```
#create a scatterplot with regression line
ggplot(data = incumbents, aes(x = incumbents$residuals2, y = incumbents$
    residuals)) +
  geom_point(alpha = 0.5) + #add a scatterplot
  geom_smooth(method = "lm") #add a linear regression line
```



3. Write the prediction equation.

   For this regression the prediction equation is: y = - 0.000000000000000005207 + 0.2568770127000978287 24*x

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`.

```
#run regression with voteshare as outcome variable and difflog and
    presvote as the explanatory variables
summary(lm(data = incumbents, voteshare ~ difflog + presvote))
```

2. Write the prediction equation.

   For this regression the prediction equation is: y = - 0.4486442 + 0.0355431*`difflog` + 0.2568770 *`presvote`

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