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`.

```
incumbentsRegression1 <- lm(voteshare ~ difflog, data = incumbents)
  summary(incumbentsRegression1)
```

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

```
  ggplot(incumbents, aes(voteshare, difflog)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "lm")
```

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

```
  incumbents1 <- lm(voteshare ~ difflog, data = incumbents)
  incumbents1_resid <- resid(incumbents1)
```

4. Write the prediction equation.

```
incumbentsRegression1 <- lm(voteshare ~ difflog, data = incumbents)
  variableDifflog <- data.frame(difflog=c(0.5,0.67,0.46,0.6,1.0,0.73))
  predict(incumbentsRegression1, newdata = variableDifflog, interval =
  'confidence')
```

Y = voteshare, X = difflog

Y = 0.042X + 0.579

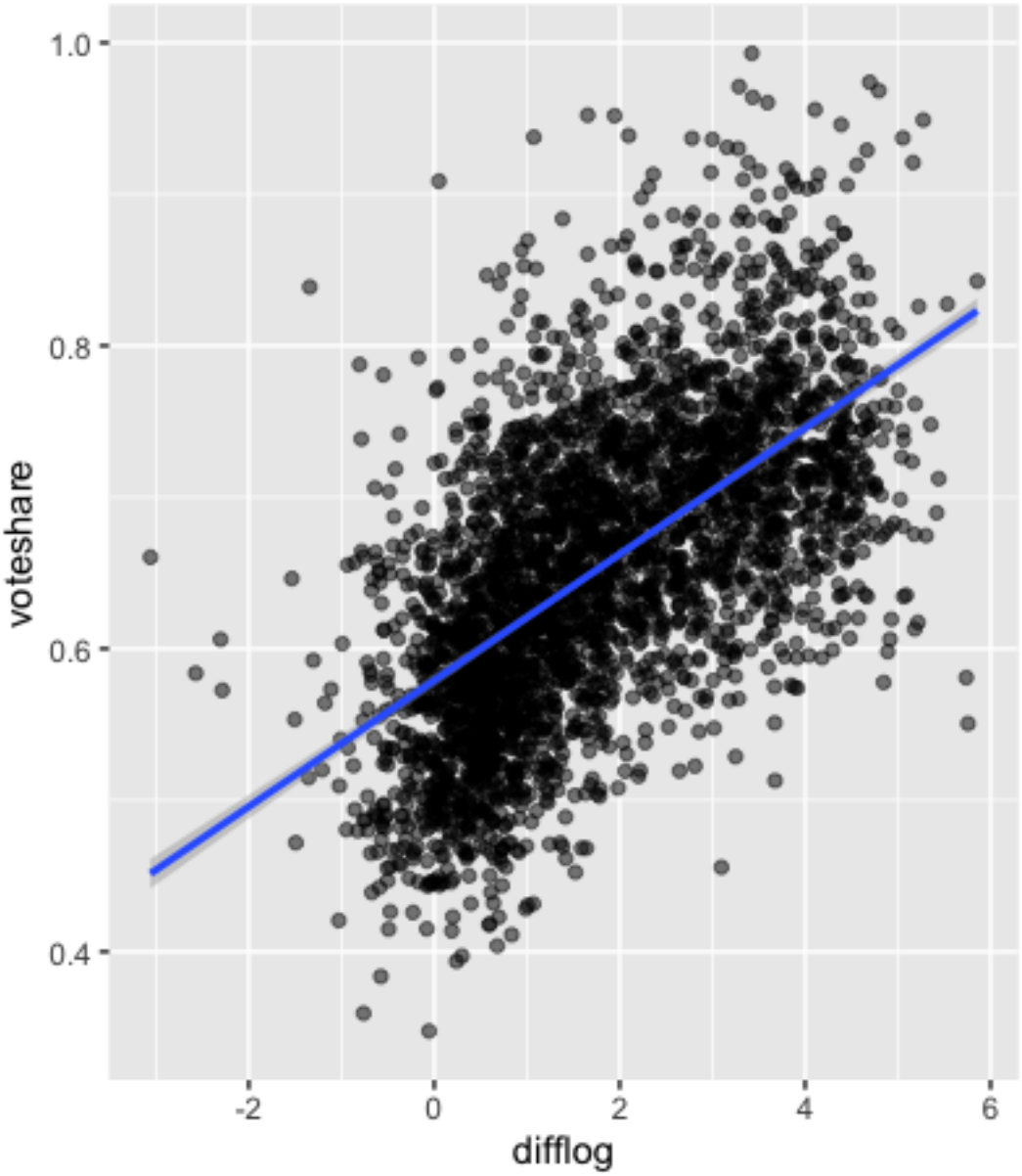

Figure 1: Scatterplot of voteshare and difflog

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`.

```
incumbents2 <- lm(presvote ~ difflog, data = incumbents)
summary(incumbents2)
```

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

```
ggplot(incumbents, aes(difflog, presvote)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "lm")
```

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

```
incumbents2 <- lm(voteshare ~ difflog, data = incumbents)
incumbents2_resid <- resid(incumbents1)
```

4. Write the prediction equation.

```
variableDifflog <- data.frame(difflog=c(0.5,0.67,0.46,0.6,1.0,0.73))
predict(incumbents2, newdata = variableDifflog, interval = 'confidence')
```

   Y = presvote, X = difflog

   Y = 0.508 + 0.0238X

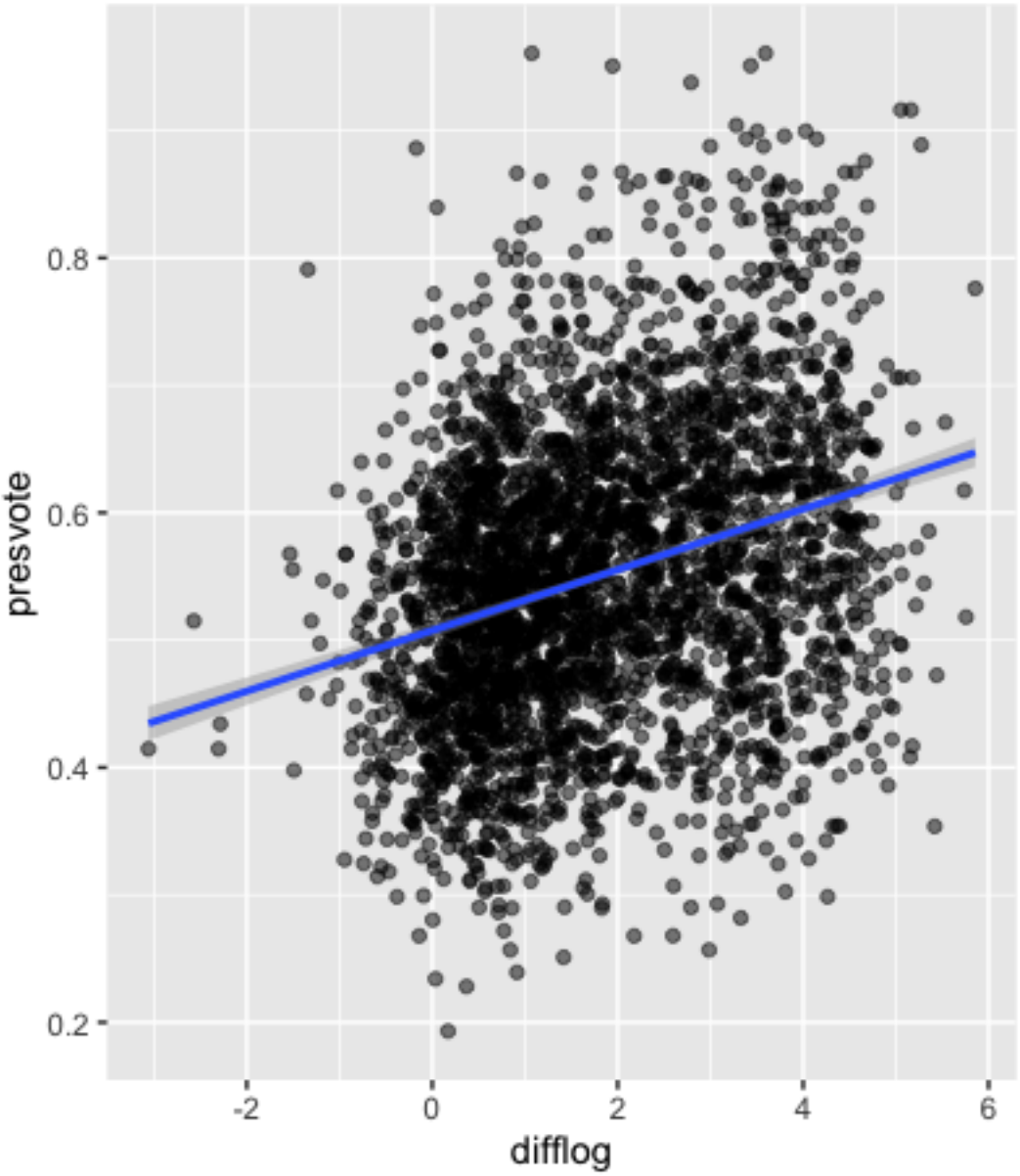

Figure 2: Scatterplot of presvote and difflog

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`.

```
incumbents3 <- lm(voteshare ~ presvote, data = incumbents)
summary(incumbents3)
```

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

```
ggplot(incumbents, aes(presvote, voteshare)) +
geom_point(alpha = 0.5) +
geom_smooth(method = "lm")
```

3. Write the prediction equation.

```
variablePresvote <- data.frame(presvote=c(0.5,0.37,0.46,0.32,0.41,0.52))
predict(incumbents3, newdata = variablePresvote, interval = 'confidence')
```

   Y = voteshare, X = presvote Y = 0.441 + 0.338X

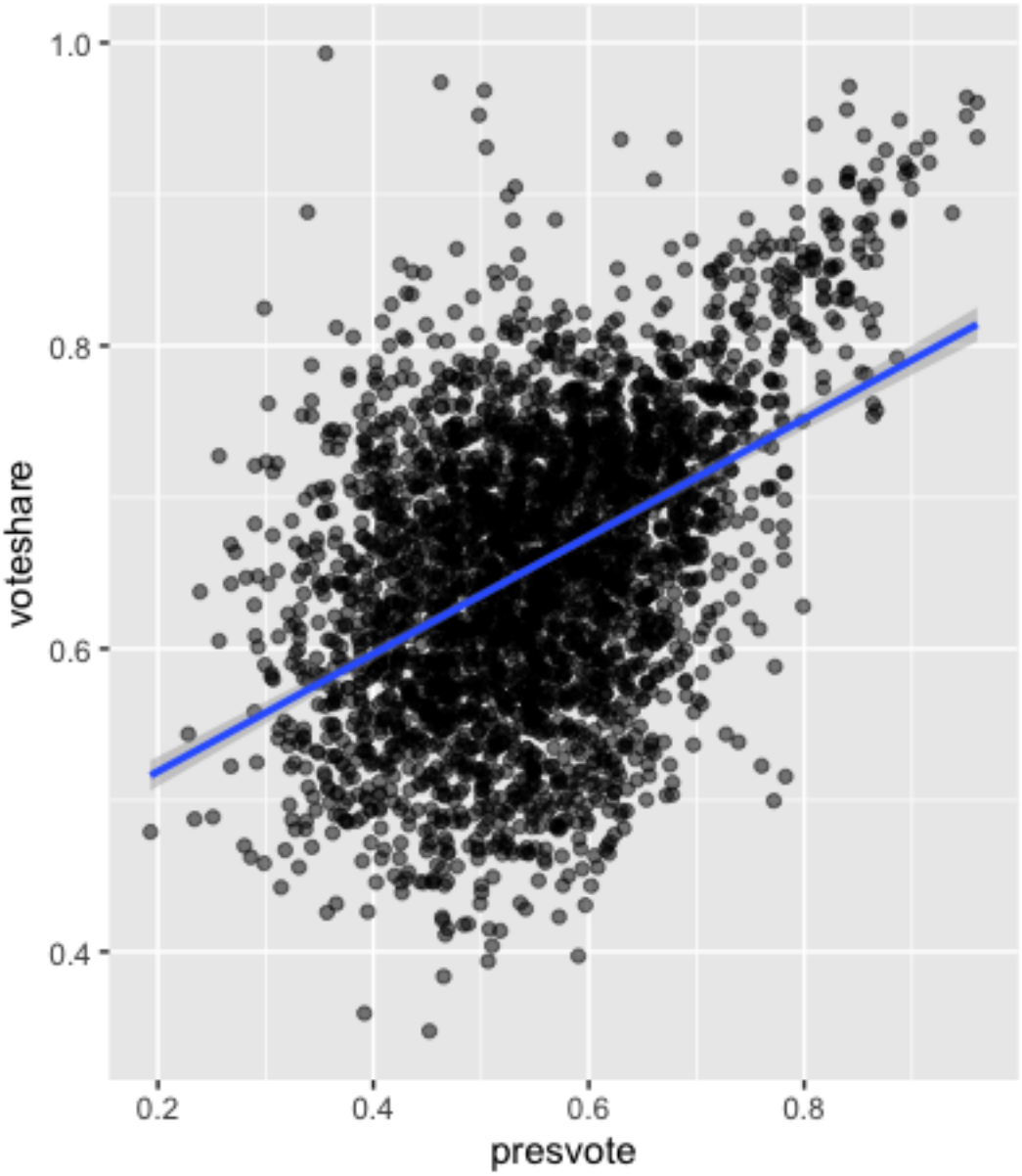

Figure 3: Scatterplot of voteshare and presvote

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.

this isn't right, I am not sure why !!

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

```
incumbents4 <- lm(incumbents1_resid ~ incumbents2_resid)
summary(incumbents4)
```

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

```
ggplot(incumbents, aes(incumbents2_resid, incumbents1_resid)) +
geom_point(alpha = 0.5) +
geom_smooth(method = "lm")
```

3. Write the prediction equation.

```
y = incumbents1_resid, x = incumbents2_resid
y = −3.490e^33 + 1.000e^00x
```

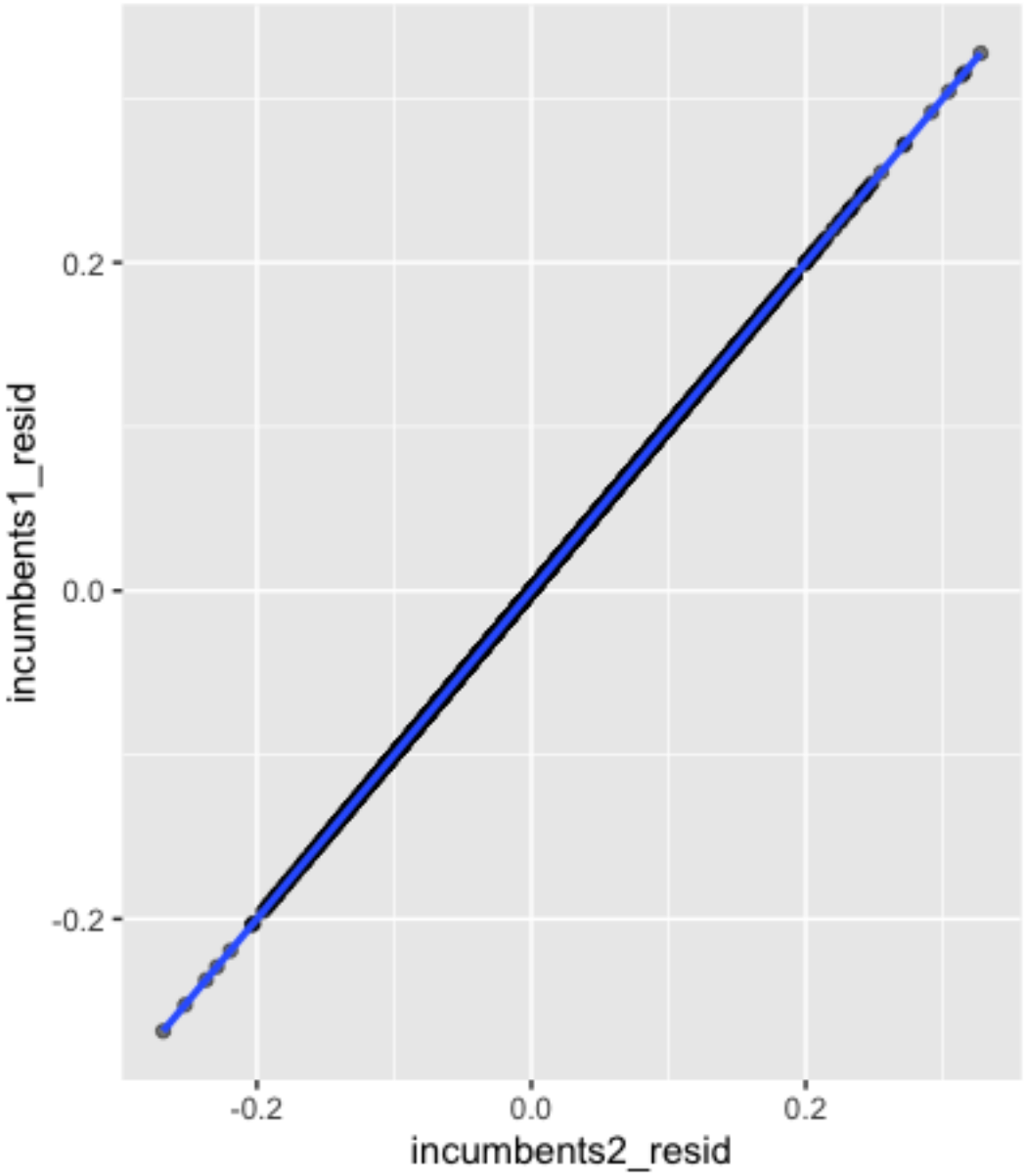

Figure 4: Scatterplot of residuals

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`.

```
incumbents5 <- lm(voteshare ~ difflog + presvote, data = incumbents)
summary(incumbents5)
```

2. Write the prediction equation.

```
y = vote share, x0 = difflog, x1 = presvote
y = 0.449 + 0.0355x0 + 0.257x1
```

3. What is it in this output that is identical to the output in Question 4? Why do you think this is the case? From the other girls, the residuals should be the same.

   The residuals tell us how much of the variation in `voteshare` is *not* explained by the difference in spending between incumbent and challenger and how much of the variation in `presvote` is *not* explained by the difference in spending between incumbent and challenger in the district. Are the residuals the same because in Q5 we are just running a multivariate regression analysis on the three variables we were previously investigating in single linear regressions? That is my best guess.