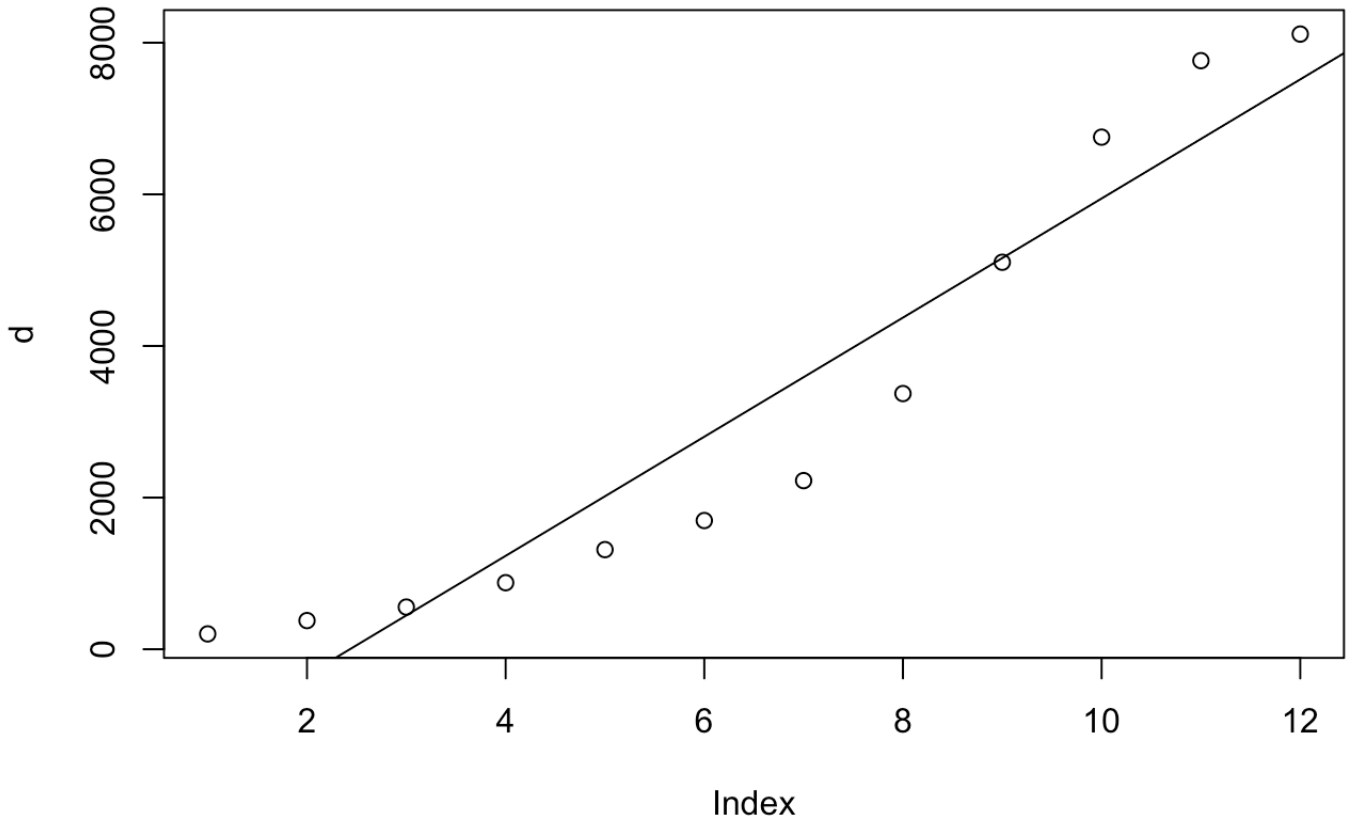
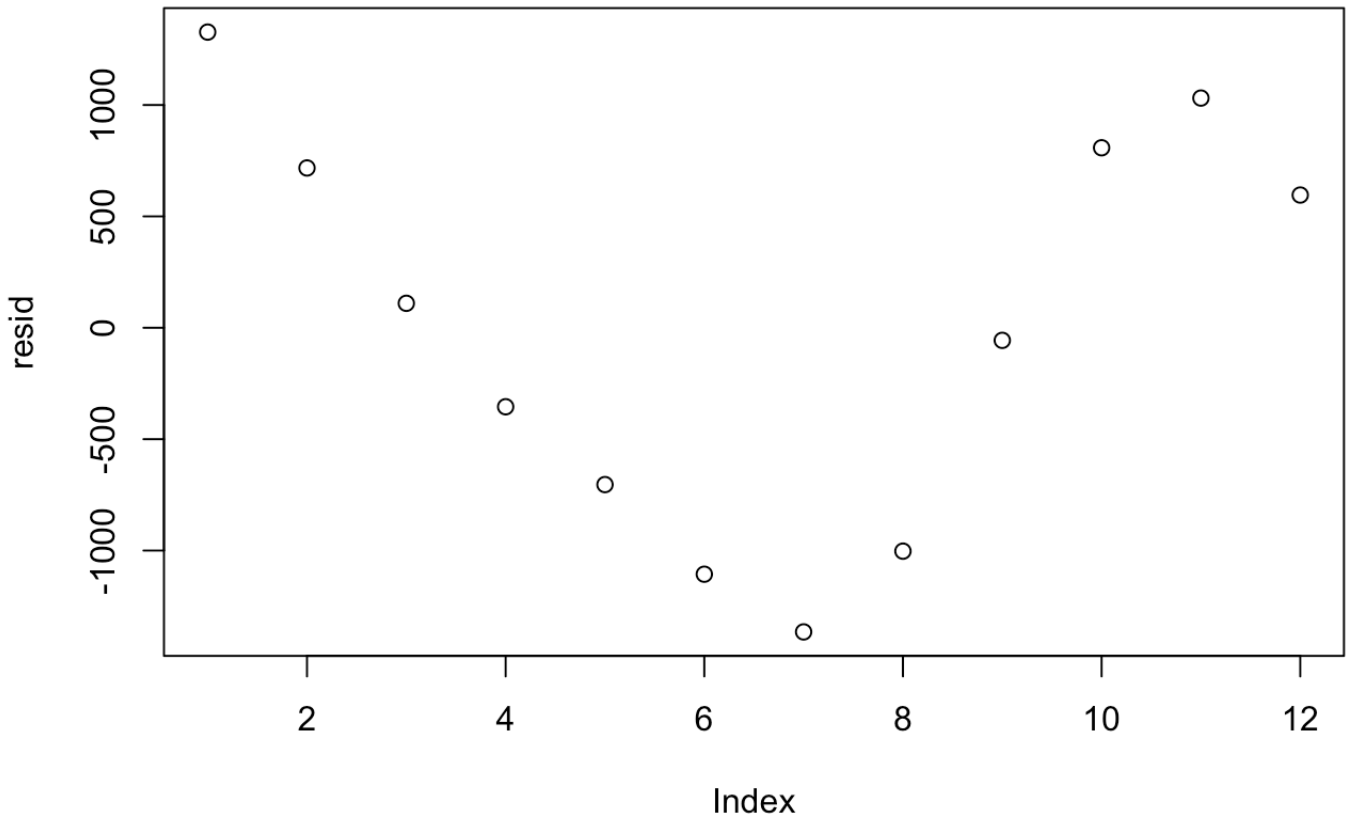
Covide Effect on Unemployment

Leoul Berhanu, Frederick Huang, Rhys Kittleson, Sakshum Kulshrestha, Hoa Minh Le

Introduction: Our project used the statistical tools we learned this semester to see how unemployment rates were affected by COVID 19. Our first experiment was conducted via a linear regression on unemployment rates. Our second experiment was a Hypothesis test (T-test) that compared unemployment rates in those with High School Degrees Versus Bachelor's Degrees.

Here you will find the construction of our linear regression. We found values of Sxy, Sxx, b hat and alpha from our data set and then graphed the points along with a line of best fit. Along with the regression, we constructed a plot of the residuals.





The fact that the residual plot has such a clear pattern is a bit concerning, because it implies that the data is not well suited to a linear relationship. We can further analyze our regression with a hypothesis test on the slope parameter.

```
## [1] 608.2154
```

```
## [1] 963.2393
```

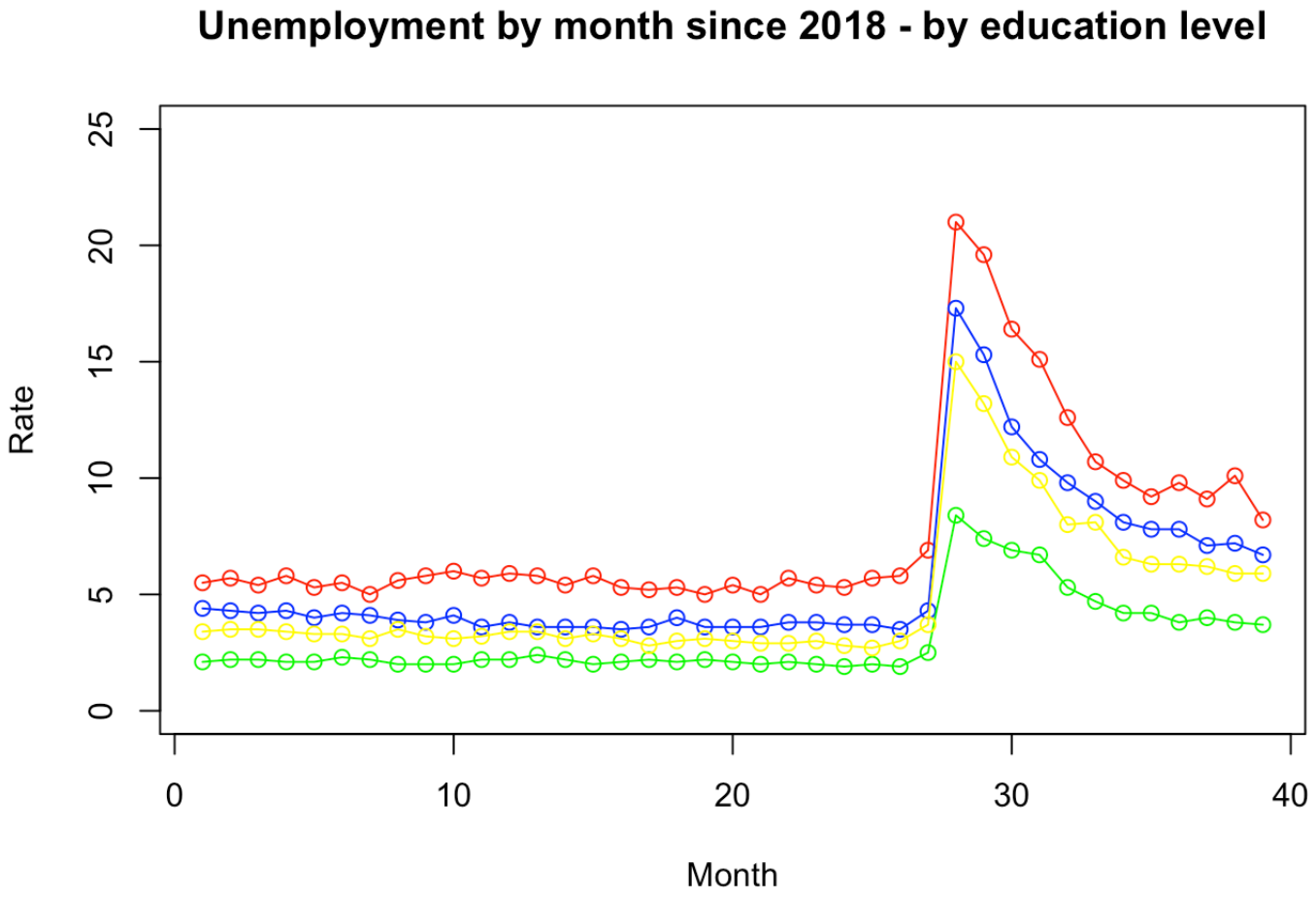
This 95 percent confidence interval for the B parameter is bounded at 608.21 to 963.24. Also, our estimated b_hat is included in this range, which is a promising result. Next, we can check how well the change in x reflects in the change in y in our linear regression

```
## [1] 0.897192
```

The r squared value is around 0.9, which is a relatively good r squared value, meaning that the linear regression fit this data decently. However, it is clear that the regression wouldn't be very good at predicting future values.

We now transition to our second test, which is a hypothesis test (T-test) that sees how unemployment rates

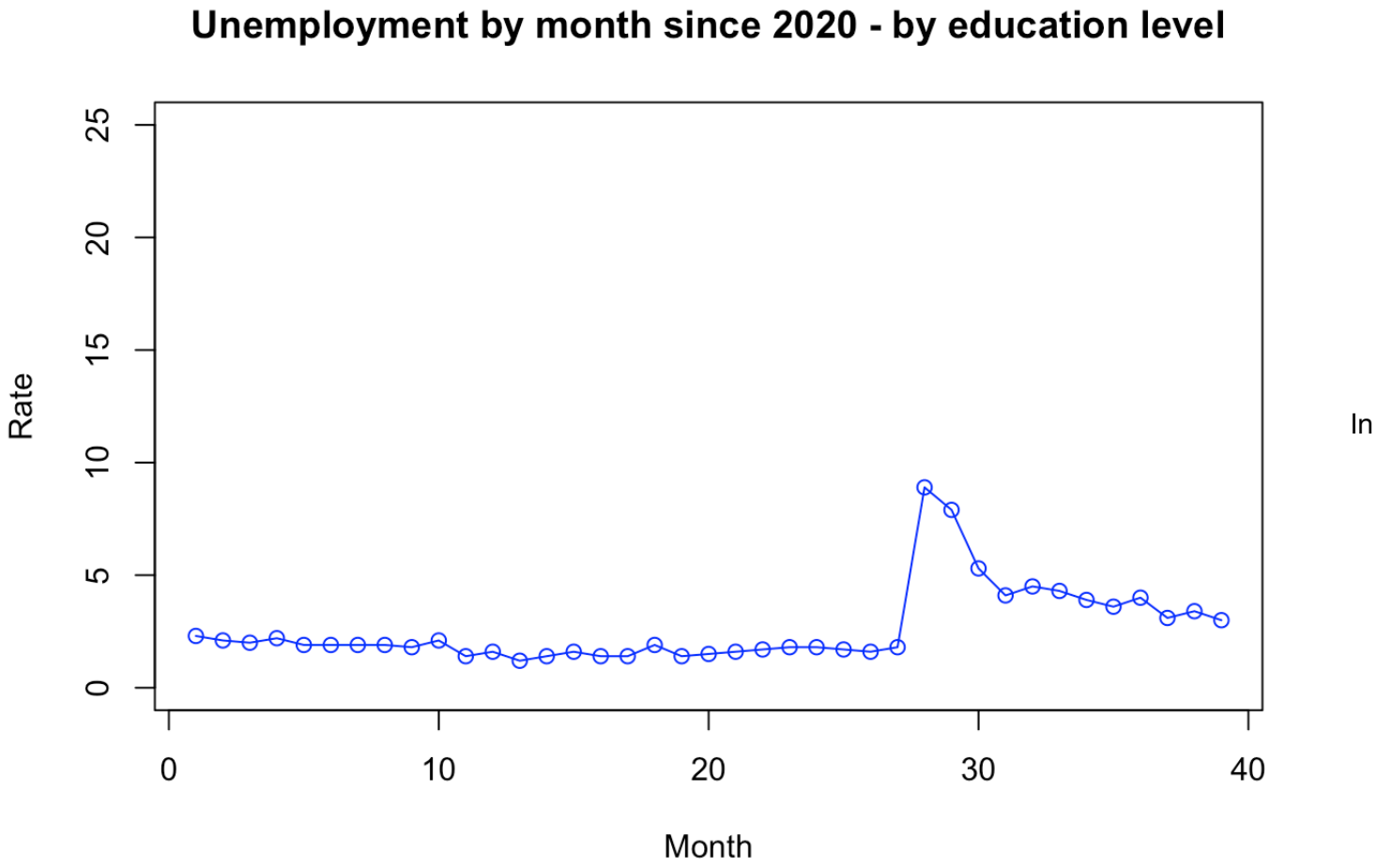
are represented by education level. Our Null Hypothesis was covid did not affect different groups differently, so the difference in the gap should be 0. In addition, there are graphs to represent the data.



Here is a simple plot of unemployment by month since 2016.

```
## [1] 0.2813188
```

With a P value of 0.28, we fail to reject the null hypothesis that unemployment is not affected by education level. This failure indicates that there is not significant evidence to support the claim that unemployment levels are different among different education groups.



In

conclusion, our linear regression estimated the b value very nicely, and did a great job at aligning y to x values with an R^2 of 0.9, but do to the poor residual plot it is safe to say that the model wouldn't do a good job at predicting future values. This makes sense because the spike in unemployment is very hard to explain without context, which ruins the linearity of it. In regards to our hypothesis test, the findings show that different education groups recieved the same treatment in terms of unemployment. This finding may be seen as odd since high school jobs are a lot less secure than jobs that require a bachelors, but due to the large scale of people who have a bachelors but work an average job, the findings do make some sense.

For future work, this project opened a lot of doors as to what we could study in the future. One thing our group was thinking of looking at was comparing unmployment rates during other pandemics, and seeing what results that would yield. If we had more tools in our arsenal, we would have tried to see if unemployment would have fit an exponential model, since the rise was so quick and extreme.