Relationship Between Financial Data and Forest Coverage in 2018

https://github.com/emmadeangeli/Shapiro_DeAngeli_Culberson_ENV872_EDA_FinalProject.git

Ben Culberson, Emma DeAngeli, and Shana Shapiro

Contents

1 Rationale and Research Questions

We were initially interested in whether climate financing correlated to environmental health in various countries. However, we could not find enough data to lead to a robust analysis. Instead, we are looking into whether different financial indicators (in as many countries as we could find) lead specifically to forest coverage. In our project, we are using forest coverage as a proxy for environmental health. Developed countries may have the luxury of being able to preserve their forests. We were particularly interested in seeing if other financial indicators related to forest coverage. Our main research question was "do financial indicators correlate to forest coverage for different countries in 2018?" Therefore, our hypotheses are the following:

$$H_0 :$$

There are no financial indicators or combination of financial indicators that correlate to forest coverage.

$$H_a :$$

There is at least one financial indicator or combination of financial indicators that correlates to forest coverage.

2 Dataset Information

We pulled our data from three databases on the World Bank website: Global Financial Development, Global Economic Monitor, and Country Climate and Development Report. The first two have variables related to financial and economic data while the third has variables related to climate and development. After initial exploration into the most climate-vulnerable countries, we selected four variables to analyze in the year 2018:

Dataset	Description
Central Bank Assets to GDP (%)	This variable is equal to a given country's central bank assets divided by its GDP for the year 2018
Domestic Credit to Private Sector (% of GDP)	This variable is equal to the amount of private investment in a given countries debt divided by that countries GDP for the year 2018
Domestic Reserves (Million US$)	This variable is equal to a given country's monetary reserves held by its central bank in million USD for the year 2018
Dependent variable:	Share of surface occupied by forest (%)

3 Exploratory Analysis

3.1 Individual Linear Regressions with Each Financial Variable

```
forest.regression1 <- lm(data = finance_forests.df, Forest_Share_of_Surface ~ credit.to.
summary(forest.regression1)
```

```
##
## Call:
## lm(formula = Forest_Share_of_Surface ~ credit.to.private, data = finance_forests.df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.419 -19.682  -1.393  13.912  61.045
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       31.91587    4.05139   7.878 4.88e-12 ***
## credit.to.private  0.01868    0.05225   0.358    0.721
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.58 on 97 degrees of freedom
## Multiple R-squared:  0.001317,   Adjusted R-squared:  -0.008979
## F-statistic: 0.1279 on 1 and 97 DF,  p-value: 0.7214
```

```
forest.regression2 <- lm(data = finance_forests.df, Forest_Share_of_Surface ~ assets.to.
summary(forest.regression2)
```

```
##
## Call:
## lm(formula = Forest_Share_of_Surface ~ assets.to.gdp, data = finance_forests.df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.964 -16.681  -2.608  14.229  56.839
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    29.9110     2.4583  12.167   <2e-16 ***
## assets.to.gdp   0.5564     0.2197   2.532   0.0129 *
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.91 on 97 degrees of freedom
## Multiple R-squared:  0.062,  Adjusted R-squared:  0.05233
## F-statistic: 6.411 on 1 and 97 DF,  p-value: 0.01295
```

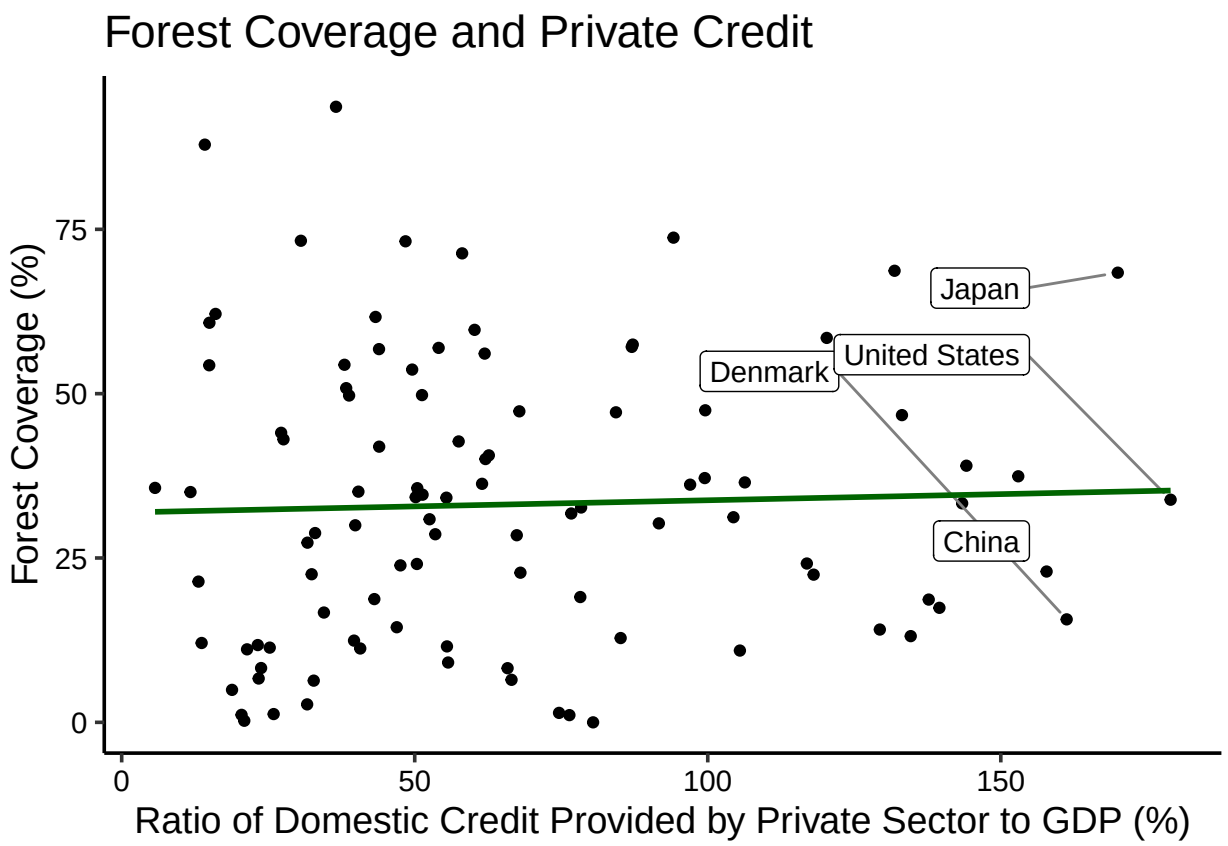
```
forest.regression3 <- lm(data = finance_forests.df, Forest_Share_of_Surface ~ Reserves)
summary(forest.regression3)
```

```
##
## Call:
## lm(formula = Forest_Share_of_Surface ~ Reserves, data = finance_forests.df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.189 -18.793  -0.428  14.346  60.609
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.303e+01  2.241e+00  14.742   <2e-16 ***
## Reserves    1.206e-06  6.431e-06   0.188    0.852
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.59 on 97 degrees of freedom
## Multiple R-squared:  0.0003626,  Adjusted R-squared:  -0.009943
## F-statistic: 0.03518 on 1 and 97 DF,  p-value: 0.8516
```

3.2 Plots of share of surface occupied by foreast on all 3 financial variables

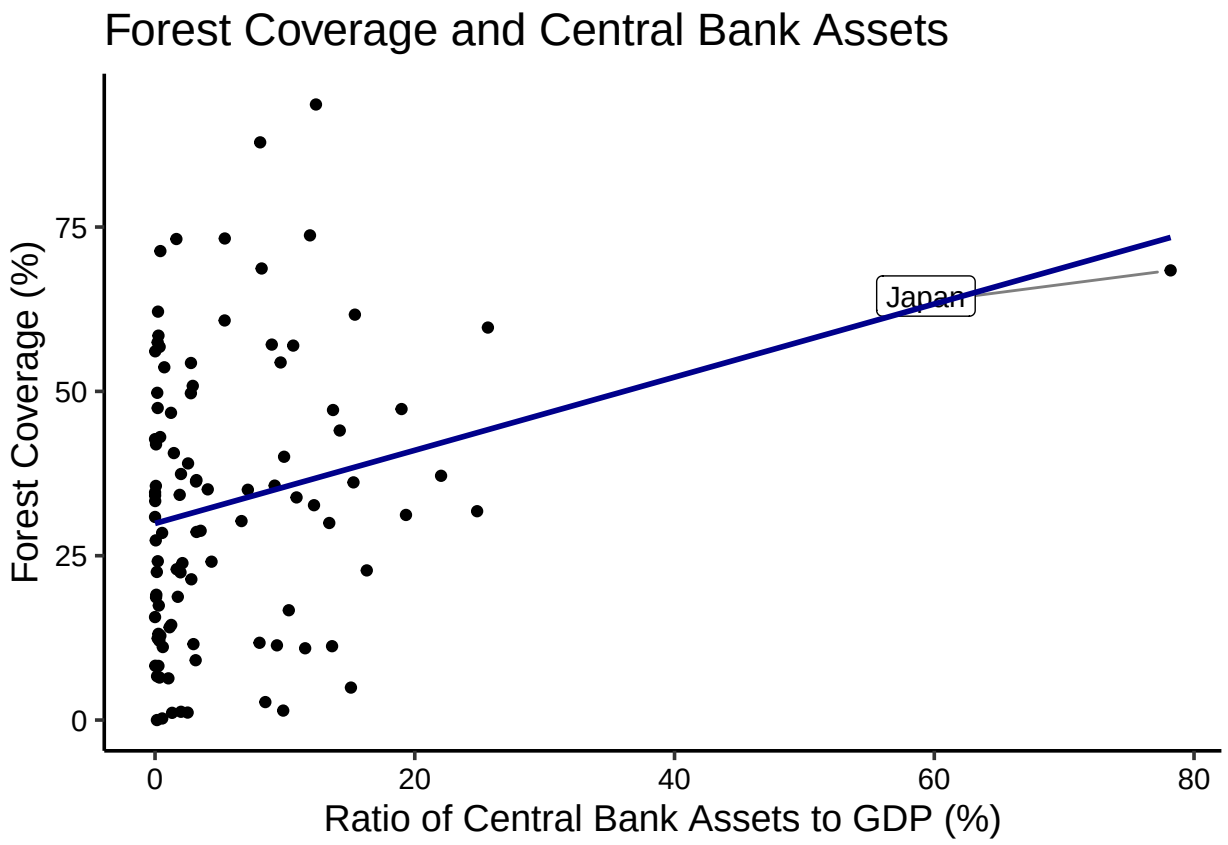
```
#Plot of Domestic Credit to Private Sector (% of GDP)
ggplot(finance_forests.df, aes(x = credit.to.private, y = Forest_Share_of_Surface)) +
  geom_point() +
  geom_label_repel(aes(label = Country),
                   box.padding   = 5,
                   point.padding = 0.5,
                   max.overlaps = 50,
                   segment.color = 'grey50')+
  geom_smooth(method = 'lm', se = FALSE, color = "darkgreen") +
  labs(x = "Ratio of Domestic Credit Provided by Private Sector to GDP (%)", y = "Forest
```

```
## `geom_smooth()` using formula 'y ~ x'

## Warning: ggrepel: 95 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



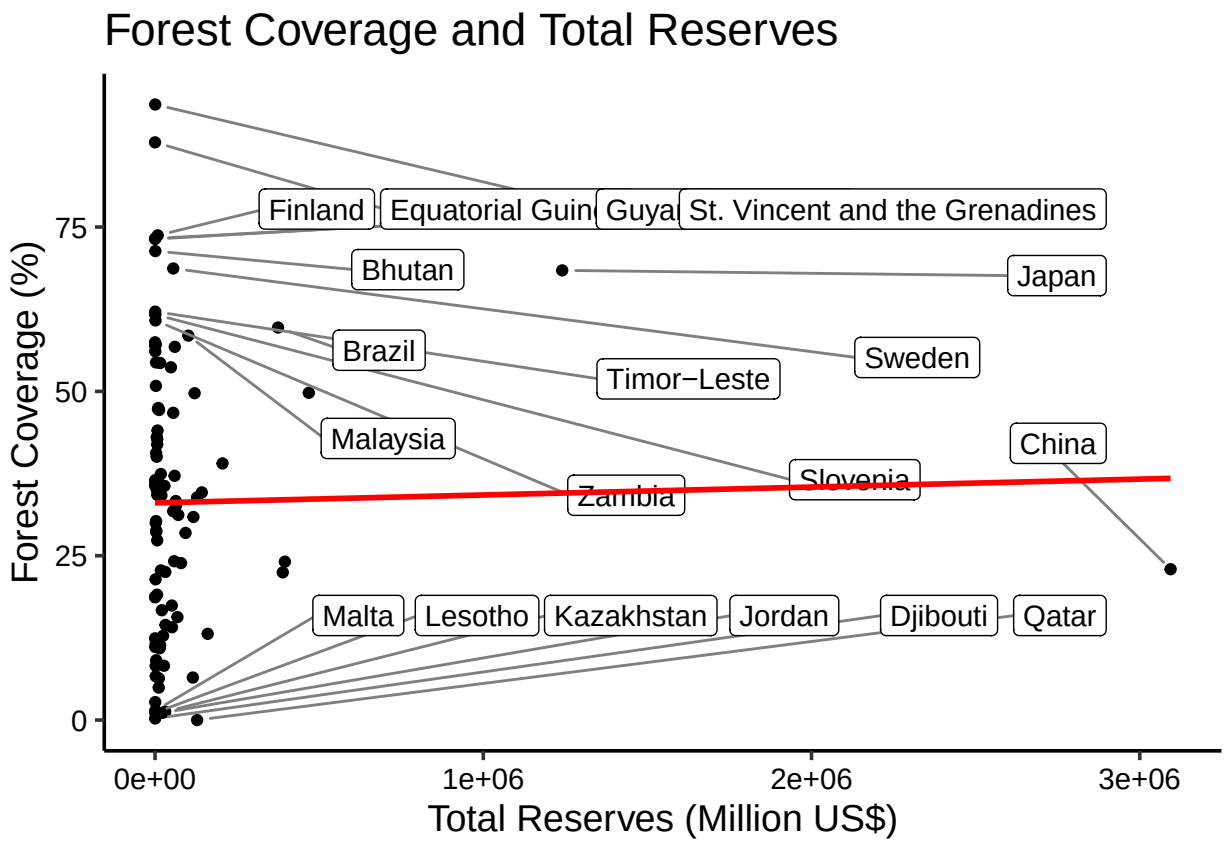

```
#Plot of Central Bank Assets to GDP
ggplot(finance_forests.df, aes(x = assets.to.gdp, y = Forest_Share_of_Surface, label=Cou
  geom_point() +
  geom_label_repel(aes(label = Country),
                   box.padding   = 5,
                   point.padding = 0.5,
                   max.overlaps = 10,
                   segment.color = 'grey50')+
  geom_smooth(method = 'lm', se = FALSE, color = "darkblue") +
  labs(x = "Ratio of Central Bank Assets to GDP (%)", y = "Forest Coverage (%)", title =
```

```
## `geom_smooth()` using formula 'y ~ x'

## Warning: ggrepel: 98 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



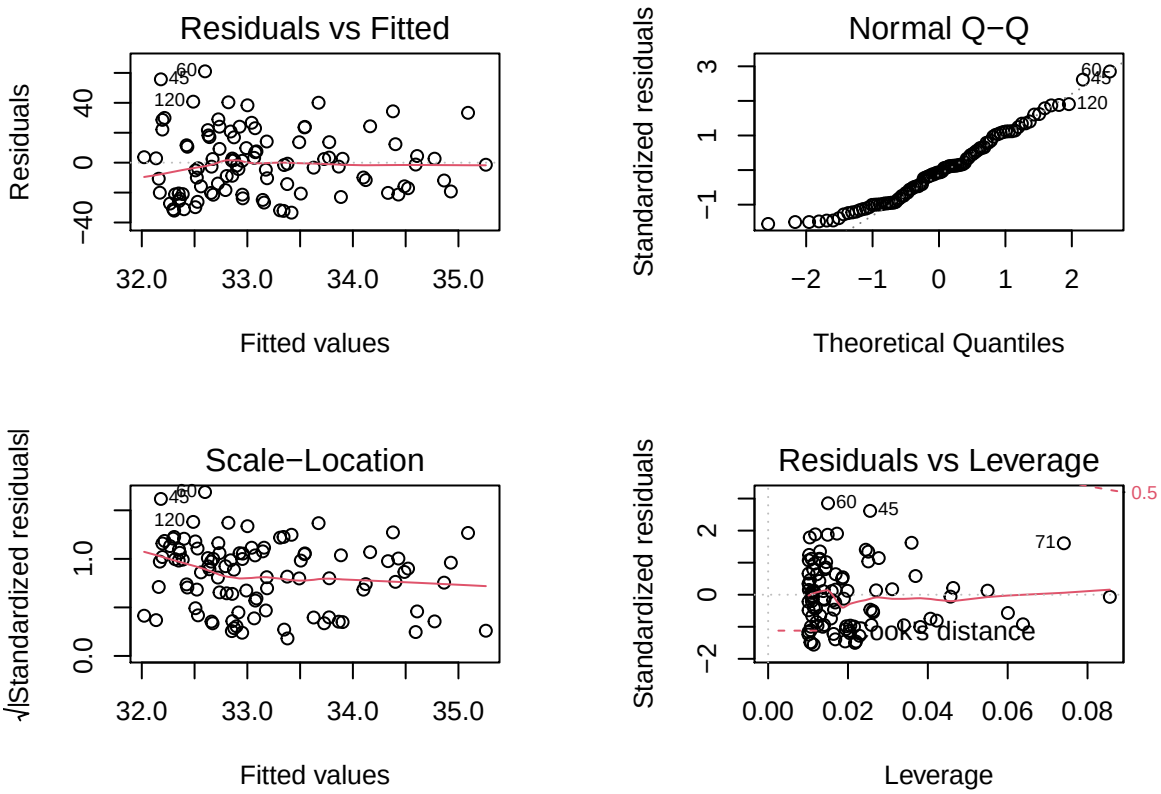

```
#Plot of Domestic Reserves (Million US$)

ggplot(finance_forests.df, aes(x = Reserves , y = Forest_Share_of_Surface, label=Country
  geom_point() +
  geom_label_repel(aes(label = Country),
                   box.padding   = 3,
                   point.padding = 0.5,
                   max.overlaps = 100,
                   segment.color = 'grey50')+
  geom_smooth(method = 'lm', se = FALSE, color = "red") +
  labs(x= 'Total Reserves (Million US$)', y = "Forest Coverage (%)", title = "Forest Cov
```

```
## `geom_smooth()` using formula 'y ~ x'

## Warning: ggrepel: 79 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



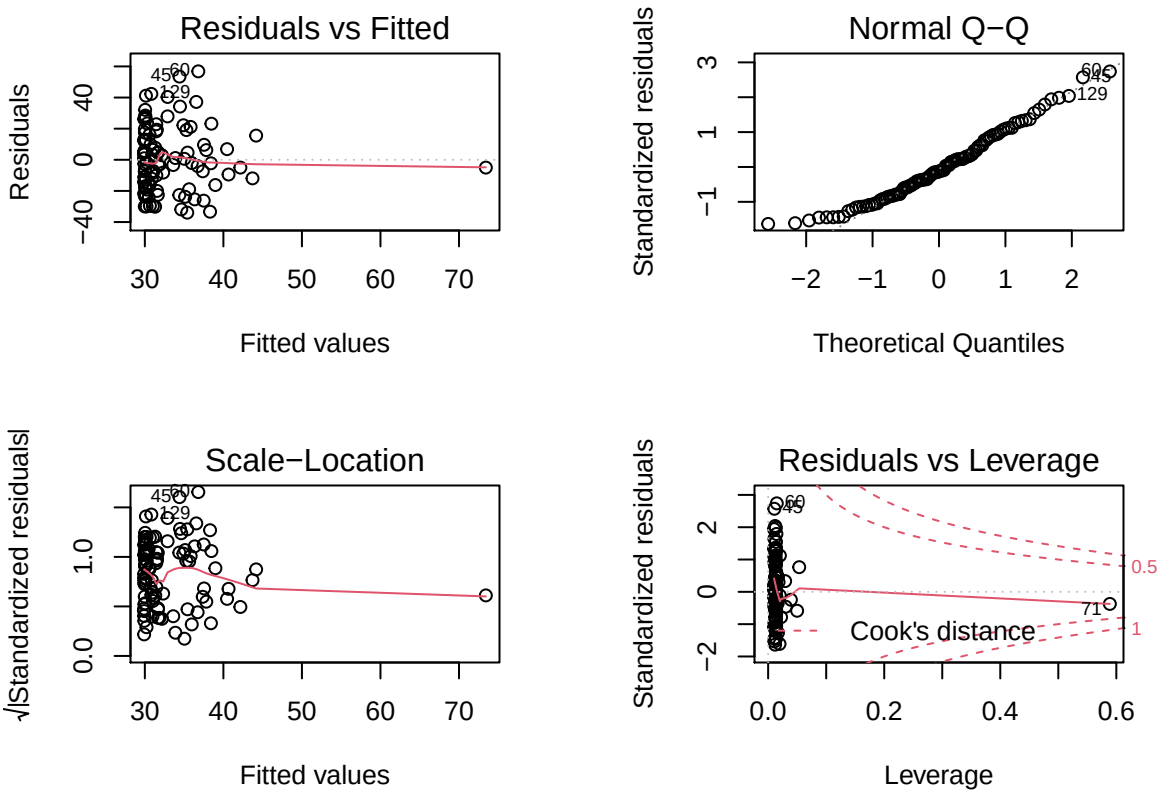
3.3 Residuals and Errors

```
par(mfrow = c(2,2), mar=c(4,4,4,4))
plot(forest.regression1)
```



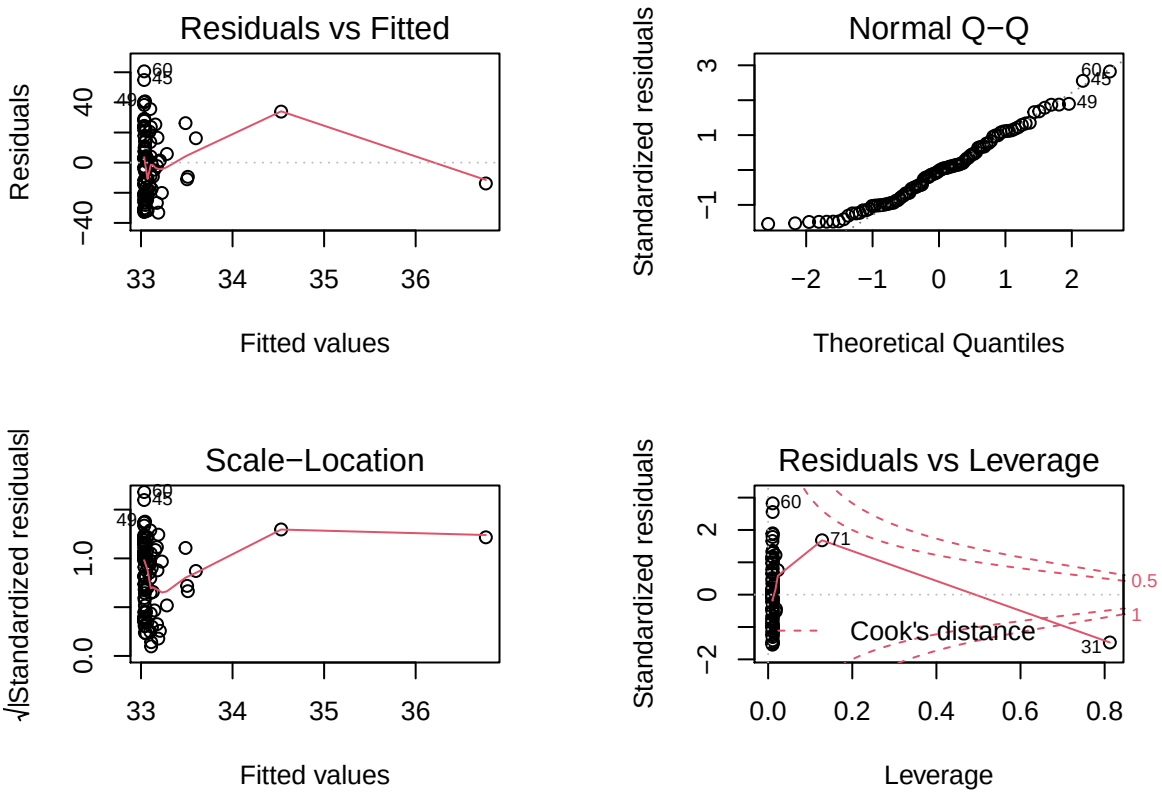

```
par(mfrow = c(1,1))
```

```
par(mfrow = c(2,2), mar=c(4,4,4,4))
plot(forest.regression2)
```




```
par(mfrow = c(1,1))
```

```
par(mfrow = c(2,2), mar=c(4,4,4,4))
plot(forest.regression3)
```




```
par(mfrow = c(1,1))
```

3.4 Creating a Multivariate Linear Model

```
Forestfinance.AIC <- lm(data = finance_forests.df, Forest_Share_of_Surface ~ credit.to.p
 step(Forestfinance.AIC)
```

```
## Start:  AIC=607.79
## Forest_Share_of_Surface ~ credit.to.private + assets.to.gdp +
##     Reserves
##
##                      Df Sum of Sq   RSS    AIC
## - credit.to.private   1      0.40 42345 605.79
## - Reserves            1     69.15 42413 605.95
## <none>                            42344 607.79
## - assets.to.gdp       1   2815.90 45160 612.16
##
## Step:  AIC=605.79
```

```
## Forest_Share_of_Surface ~ assets.to.gdp + Reserves
##
##                 Df Sum of Sq   RSS    AIC
## - Reserves       1     73.52 42418 603.96
## <none>                       42345 605.79
## - assets.to.gdp  1   2860.66 45205 610.26
##
## Step:  AIC=603.96
## Forest_Share_of_Surface ~ assets.to.gdp
##
##                 Df Sum of Sq   RSS    AIC
## <none>                       42418 603.96
## - assets.to.gdp  1    2803.5 45222 608.30

##
## Call:
## lm(formula = Forest_Share_of_Surface ~ assets.to.gdp, data = finance_forests.df)
##
## Coefficients:
##   (Intercept)  assets.to.gdp
##       29.9110         0.5564
```

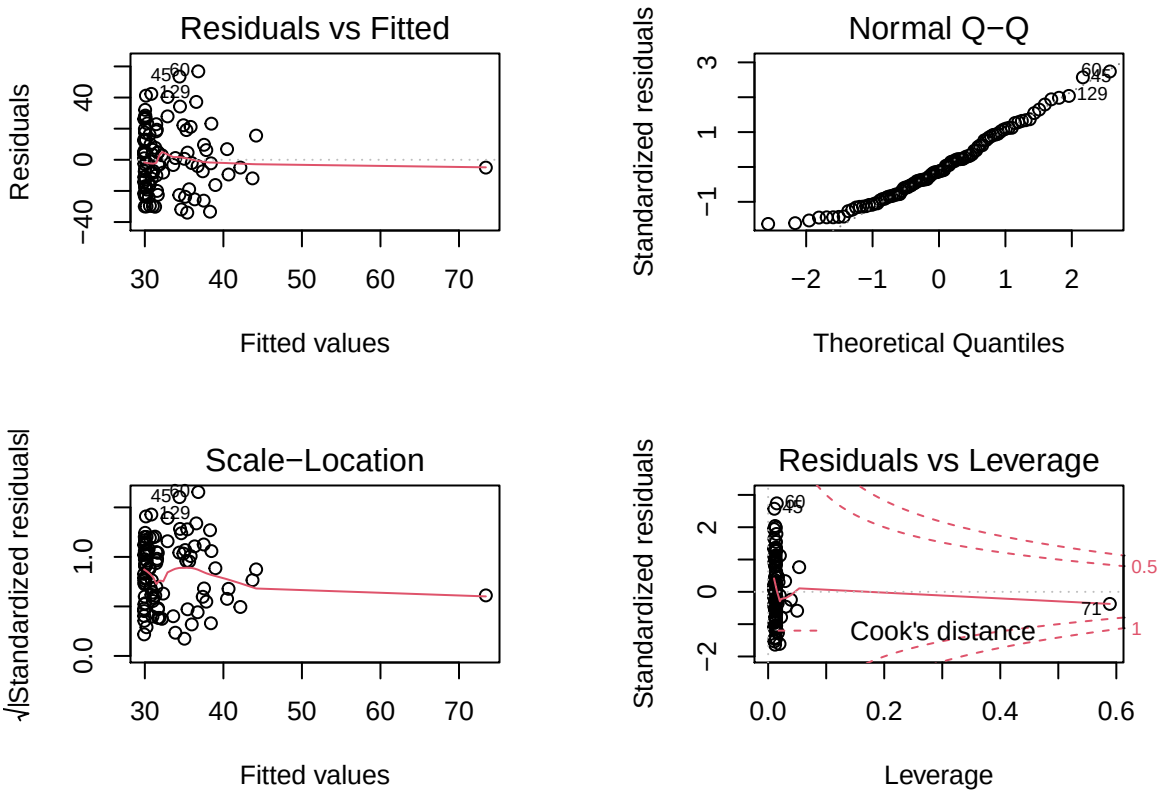
```
summary(Forestfinance.AIC)
```

```
##
## Call:
## lm(formula = Forest_Share_of_Surface ~ credit.to.private + assets.to.gdp +
##     Reserves, data = finance_forests.df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -34.296 -16.685  -2.649  14.391  56.515
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        2.992e+01  4.103e+00   7.292 9.07e-11 ***
## credit.to.private  1.639e-03  5.499e-02   0.030   0.9763
## assets.to.gdp      5.773e-01  2.297e-01   2.513   0.0136 *
## Reserves          -2.693e-06  6.837e-06  -0.394   0.6945
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.11 on 95 degrees of freedom
## Multiple R-squared:  0.06363,    Adjusted R-squared:  0.03406
## F-statistic: 2.152 on 3 and 95 DF,  p-value: 0.09883
```

```r
# the best result is to eliminate credit.to.private and Reserves, so "forest.regressio

Forestfinance.MR <- lm(data = finance_forests.df, Forest_Share_of_Surface ~ assets.to.gd
summary(Forestfinance.MR)
```

```
##
## Call:
## lm(formula = Forest_Share_of_Surface ~ assets.to.gdp, data = finance_forests.df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.964 -16.681  -2.608  14.229  56.839
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    29.9110     2.4583  12.167   <2e-16 ***
## assets.to.gdp   0.5564     0.2197   2.532   0.0129 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.91 on 97 degrees of freedom
## Multiple R-squared:  0.062,  Adjusted R-squared:  0.05233
## F-statistic: 6.411 on 1 and 97 DF,  p-value: 0.01295
```

```r
par(mfrow = c(2,2), mar=c(4,4,4,4))
plot(Forestfinance.MR)
```




```
par(mfrow = c(1,1))
```

```
Forestfinance.anova <- aov(data = finance_forests.df, Forest_Share_of_Surface ~ assets.t
summary(Forestfinance.anova)
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## assets.to.gdp  1   2804  2803.5   6.411 0.0129 *
## Residuals     97  42418   437.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4 Analysis

After running 3 individual linear regressions with each of the financial variables we selected, only the regression using the Central Bank Assets to GDP variable had statistically significant p-value at 0.01295. The adjusted R-squared for regression case was 0.052, meaning that the linear model only explains roughly 5% of the variation in our dependent variable, Share of Surface Occupied by Forest (%). Despite this low adjusted R-squared, the output from this regression tells us that a 1% increase in the Assets to GDP Ratio of a given country is correlated with a 0.56% increase in the share of that country's surface covered by forest. The other two linear regressions had no statistically significant explanatory variables and had adjusted R-squared values that were close to zero so we choose not to interpret their coefficients.

After running these linear regressions, we plotted the Share of Surface Occupied by Forest on each explanatory financial variable in 3 separate scatterplots. On each scatterplot, we also plotted the fitted linear regression. In all three plots, it becomes clear the low adjusted R-squared values from our regressions were appropriate. The fitted linear regressions do not appear to explain much of the variation in our dependent variable. In the case of the Forest Coverage and Private Credit and Forest Coverage and Total Reserves plots, the linear fit does an exceptionally poor job of relating the explanatory variables to the Share of Surface Occupied by Forest dependent variable. Only for the Forest Coverage and Central Bank Assets plot does the fitted linear regression seem to make sense and even in this case, there is stil a considerable amount of the data that the fitted regression does not explain.

From the residual plots of these 3 individual linear regressions, we see that the second and third regressions (for the Total Reserves explanatory variable and the Assets to GDP explanatory variable, respectively) had residuals concentrated highly on the left hand side of the plots, and had fitted lines that did little to minimize the magnitude of these residuals. These two observations indicate that our models are not well fitted, and the corresponding adjusted-R squared values align with that indication. The first regression on the hand (for the Domestic Credit to Private Sector explanatory variable), seemed to indicate a better fitting model. In this case, there is less drastic asymmetry in the residuals and the residuals are much closer to zero than the other two regressions. However, the output for this single variable regression still shows that the model does not do a good job of fitting the data.

When we put all three explanatory financial variables together in a multivariate regression, we use the Akaike's Information Criterion (AIC) to select which variables are useful. After just 3 steps, the AIC tells us that the only needed explanatory financial variable is the Central Bank Assets to GDP ratio. In other words, the Total Reserves explanatory variable and the Domestic Credit to Private Sector explanatory variable are not terribly useful in modeling the Share Forest Coverage, possibly because they co-vary with other explanatory variables. The full multivariate linear model with all 3 financial variables still has an F-statistic that is somewhat statistically significant at the 0.1 level, but the regression with only the Central Bank Assets to GDP ratio has an F-statistic significant at the 0.05 level (equal to the p-value of the Central Bank Assets to GDP variable). Our integration of our final model then, is equivalent to the interpretation of single variable regression for Central Bank Assets to GDP

(they are the same model). A 1% increase in the Assets to GDP Ratio of a given country is correlated with a 0.56% increase in the share of that country's surface covered by forest. Any change in our other explanatory variables has no statistically significant correlation with the Share of Forest Coverage according to the models we run.

4.1 Question: Do financial indicators correlate to forest coverage for different countries in 2018?

As stated in our analysis, at least the Central Bank Assets to GDP has a statistically significant correlation with Share of Forest Coverage.

5 Summary and Conclusions

In summary, we reject the null hypothesis at the 0.05 level that there are no financial indicators or combination of financial indicators that correlate to forest coverage. Through our multiple linear regression, we find that the Central Bank Assets to GDP does have a statistically significant correlation with Share of Forest Coverage ($p = 0.0129$).

With our result, we can answer that yes, financial indicators correlate with various countries' forest coverage in 2018. While there is a statistically significant correlation, there are limitations to our analysis and we cannot reliably conclude where the causation of the correlation originates. Our analysis does not consider the exhaustive list of variables contributing to a country's forest cover, but it does provide some insight into the statistical relationship between financial indicators and environmental variables. Further research including additional financial indicators or environmental responses may be warranted.