R Statistical Analysis for "Dyadic Analysis of Fragile Middle Eastern States and Humanitarian Implications of Restrictive COVID-19 Policies"

By Daniel Habib, Naela Elmore, Seth Gulas, Nathan Ruhde, Daniel Mathew and Nicholas Parente

Preprocessing

```
library(lmtest)

data = read.csv('MLEG.csv')
data$Cases_Per_Capita = data$Cases / data$Population
data$Deaths_Per_Capita = data$Deaths / data$Population
```

Simple Tests

Despite their simplicity, these are really the tests we want to do. In their simplicity, they tease out the exact relationships we are interested in. All of these instances of ordinary least squares regression make the canonical assumptions that allow us to get these p-values, but we believe the general robustness of linear regression against deviations from our model supports this methodology. Later we will explore a non-parametric alternative.

Restrictive ~ FSI

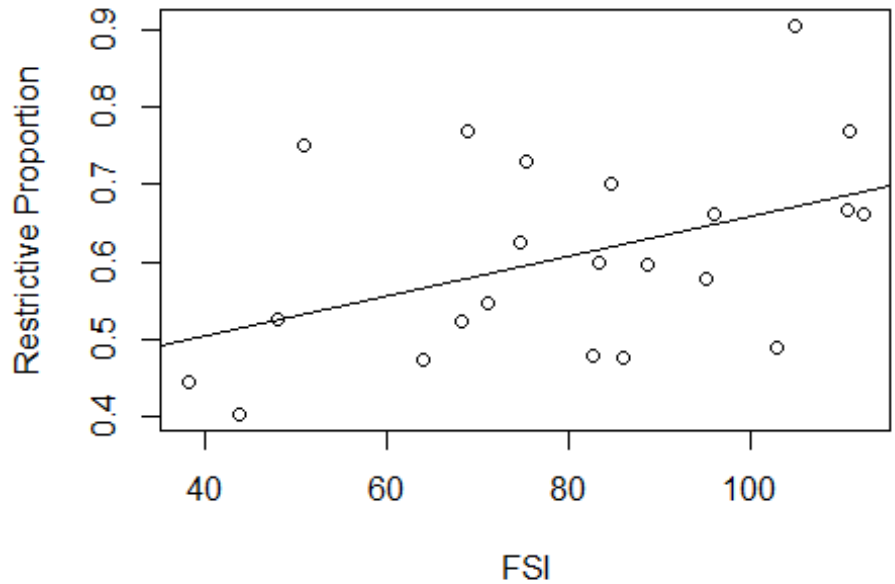
Simple ordinary least squares regression with the dependent variable being the proportion of COVID policies that were restrictive and predictor being the Fragility Index.

```
model_simple <- lm(data$Restrictive ~ data$FSI)
summary(model_simple)
```

```
##
## Call:
## lm(formula = data$Restrictive ~ data$FSI)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.17944 -0.06614 -0.02599  0.06797  0.23406
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.399335   0.096507   4.138  0.00051 ***
## data$FSI    0.002604   0.001164   2.237  0.03683 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1182 on 20 degrees of freedom
## Multiple R-squared:  0.2001, Adjusted R-squared:  0.1602
## F-statistic: 5.005 on 1 and 20 DF,  p-value: 0.03683
```

This is a 2-tailed test, so the p-value that our coefficient is > 0 is 0.018415, which is especially convincing. Now, we will explore multiple regression a bit later, though this paper is only trying to show that fragile Middle Eastern states tend to exhibit a greater proportion of restrictive policies, not claiming any sort of causation, so this evidence is convincing and there is no need to run multiple regression.

```
plot(data$FSI, data$Restrictive, xlab = 'FSI', ylab = 'Restrictive Proportion
')
abline(model_simple)
```



Quick tests of our model assumptions:

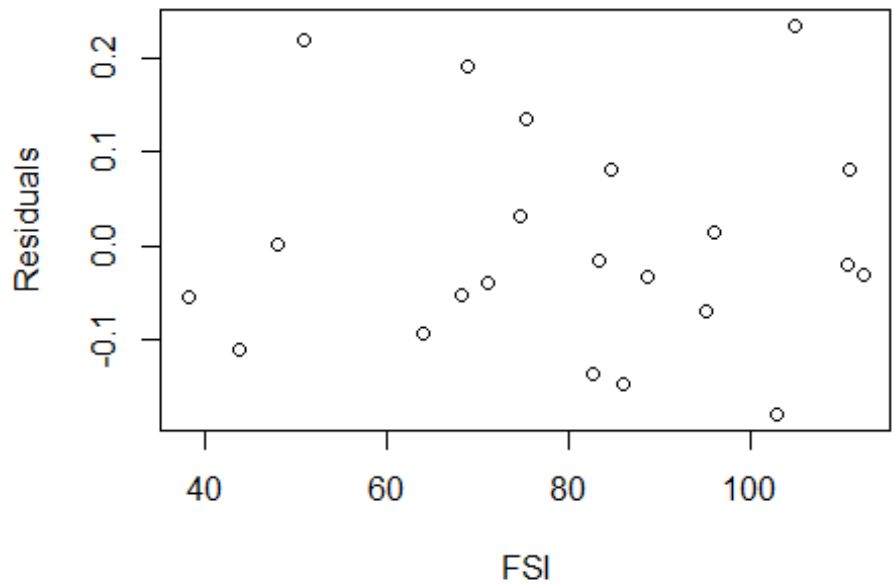
The Shapiro-Wilkes test returning a significant p-value would suggest there is good evidence to suggest our residuals aren't normally distributed, which would mean our model may be misspecified, but as discussed above, we believe OLS Regression is robust enough for our purposes and the catalog of evidence we present is extensive enough that this isn't overly concerning. Nonetheless, the p-value is not significant.

```
shapiro.test(resid(model_simple))

##
##  Shapiro-Wilk normality test
##
## data:  resid(model_simple)
## W = 0.94096, p-value = 0.2071



plot(data$FSI, resid(model_simple), xlab = 'FSI', ylab = 'Residuals')
```



Regarding outliers and high leverage points, we have so few data points here that visual inspection suggests no systematic pattern of concern.

We can also do a formal test for serialization of our residuals, though inspection of the residual plot likely more than suffices in this case:

```
dwtest(data$Restrictive ~ data$FSI)

##
##  Durbin-Watson test
##
## data:  data$Restrictive ~ data$FSI
## DW = 2.0589, p-value = 0.4904
## alternative hypothesis: true autocorrelation is greater than 0
```

Here we can see that at the very least we can't reject the idea that under our model we have serially uncorrelated errors.

Stringency ~ FSI

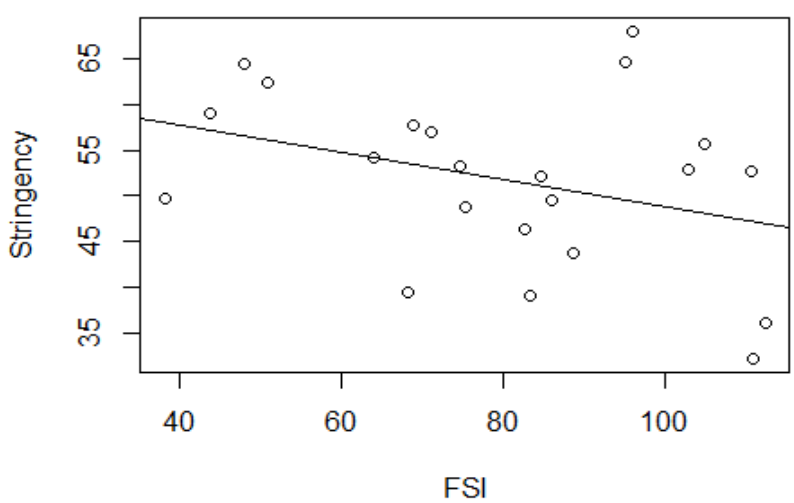
Simple ordinary least squares regression with dependent variable being the Stringency and predictor being the Fragility Index.

```
model_simple <- lm(data$Stringency ~ data$FSI)
summary(model_simple)
```

```
##
## Call:
## lm(formula = data$Stringency ~ data$FSI)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.0058  -6.3457   0.8458   5.2458  18.6537
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 63.81137    7.46101   8.553 4.09e-08 ***
## data$FSI    -0.15026    0.08997  -1.670     0.11
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.141 on 20 degrees of freedom
## Multiple R-squared:  0.1224, Adjusted R-squared:  0.0785
## F-statistic: 2.789 on 1 and 20 DF,  p-value: 0.1105
```

The two-tailed p-value here is 0.11, and consequently the one-sided p-value that the coefficient here is <0 (so higher fragility yields lower Stringency) has a p-value of 0.055. While this may fail generally conservative tests (which we support), this p-value suggests that under the assumptions of our model (the canonical assumptions made when doing multiple regression, namely a linear relationship and i.i.d. normal errors), there is a probability of ~0.055 that we would observe this data if the true coefficient of FSI was ≥ 0.

```
plot(data$FSI, data$Stringency, xlab = 'FSI', ylab = 'Stringency')
abline(model_simple)
```



Simple test: Contact ~ FSI

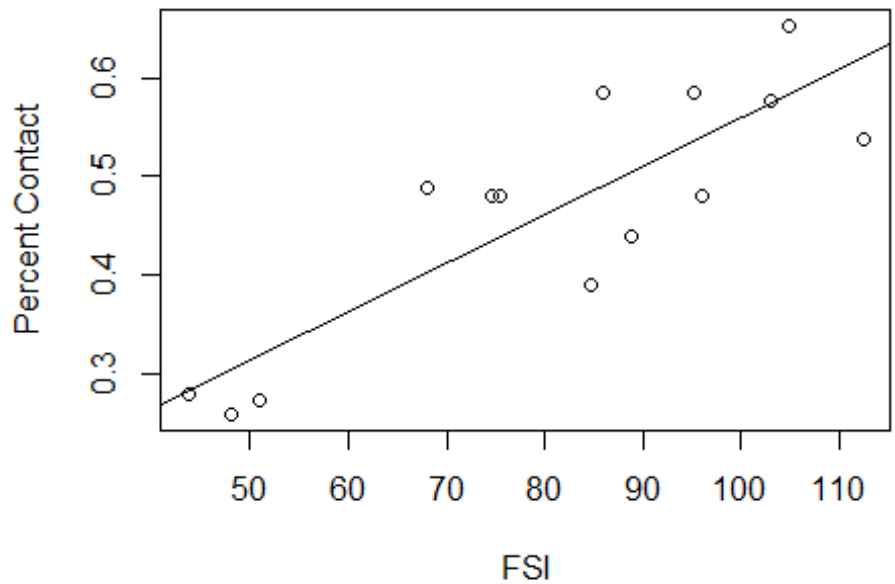
Simple ordinary least squares regression with the dependent variable being the Percent Contact and predictor being the Fragility Index.

```
data_contact = data[is.na(data['Contact']) == FALSE,]
model_simple <- lm(data_contact$Contact ~ data_contact$FSI)
summary(model_simple)

##
## Call:
## lm(formula = data_contact$Contact ~ data_contact$FSI)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.092604 -0.054404  0.000056  0.047577  0.094869
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.0697904  0.0712517   0.979    0.347
## data_contact$FSI 0.0048902  0.0008534   5.730 9.45e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06708 on 12 degrees of freedom
## Multiple R-squared:  0.7324, Adjusted R-squared:  0.7101
## F-statistic: 32.84 on 1 and 12 DF,  p-value: 9.45e-05
```

Here we get a very low p-value even for the 2-sided test, which tells us under the assumptions of our model, how likely we would be to see this data if the coefficient on FSI was ≤ 0, so we can certainly reject this idea in favor of the alternative that this coefficient is > 0. Namely larger FSI means higher contact.

```
plot(data_contact$FSI, data_contact$Contact, xlab = 'FSI', ylab = 'Percent Co
ntact')
abline(model_simple)
```



Multiple Regression

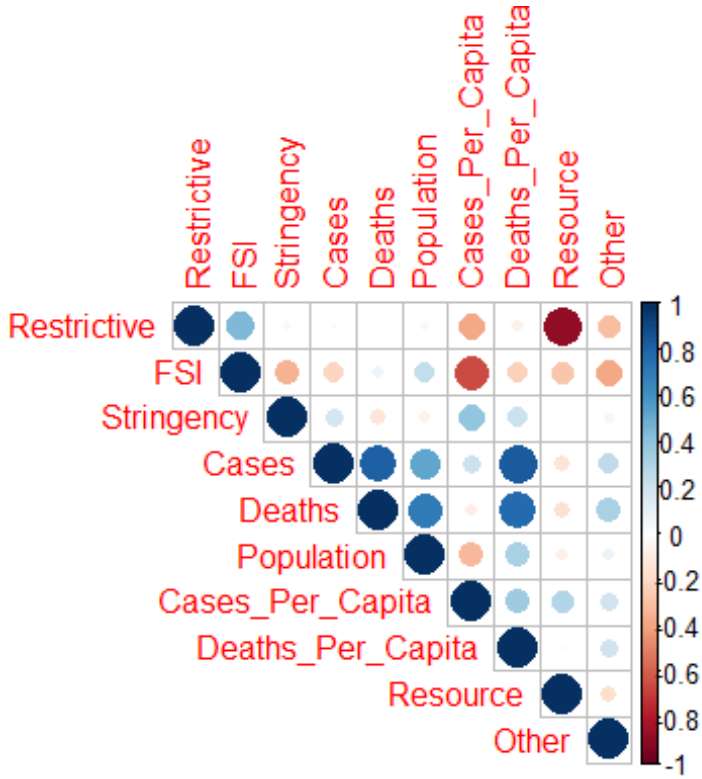
As stated above, we are not primarily interested in doing this. Although it "controls for other variables," this paper doesn't claim we have statistically significant evidence of *causation*. Further studies with more data could explore this claim. Before performing multiple regression, we should check for collinearity.

```
library(corrplot)

## corrplot 0.90 loaded

data <- data[c("Restrictive_Over_Total", "FSI", "Stringency", "Cases", "Death
s", "Population", "Cases_Per_Capita", "Deaths_Per_Capita", "Resource_Over_Tot
al", "Other_Over_Total")]
names(data)[names(data) == 'Restrictive_Over_Total'] <- 'Restrictive'
names(data)[names(data) == 'Resource_Over_Total'] <- 'Resource'
names(data)[names(data) == 'Other_Over_Total'] <- 'Other'
M <- cor(data)

corrplot(M, type="upper")
```



Here Restrictive is our dependent variable, so we search this (above) correlation matrix for collinearity among possible predictors (namely FSI). This suggests there is notable collinearity between FSI and cases_per_capita. Hence it would be unwise to perform multiple regression with both these predictors. (A more formal test of collinearity can be done by, for example, by looking at eigenvalues, but for our purposes, this simple "eye" test is likely more than sufficient).

Multiple Regression

This is a simple ordinary least-squares regression model with multiple predictors, and the dependent variable being the proportion of COVID policies that were restrictive.

```
model <- lm(data$Restrictive ~ data$FSI + data$Population + data$Cases + data
$Deaths + data$Deaths_Per_Capita)
summary(model)

##
## Call:
## lm(formula = data$Restrictive ~ data$FSI + data$Population +
##     data$Cases + data$Deaths + data$Deaths_Per_Capita)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.17614 -0.06812 -0.03040  0.05733  0.24864
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)             3.223e-01  1.409e-01   2.288   0.0361 *
## data$FSI                3.569e-03  1.466e-03   2.434   0.0270 *
## data$Population        -8.088e-10  1.671e-09  -0.484   0.6349
## data$Cases              8.242e-07  5.939e-07   1.388   0.1843
## data$Deaths            -8.686e-06  1.524e-05  -0.570   0.5767
## data$Deaths_Per_Capita -3.702e+02  9.789e+02  -0.378   0.7103
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1241 on 16 degrees of freedom
## Multiple R-squared:  0.2952, Adjusted R-squared:  0.07501
## F-statistic: 1.341 on 5 and 16 DF,  p-value: 0.2975
```

So we still find that we can reject the idea that FSI has a true coefficient ≤ 0. Now if we did include cases_per_capita, we would find this is no longer the case, but conclusions drawn from regression models suffering from collinearity are not very informative for us. Regardless, here's the output:

```
model <- lm(data$Restrictive ~ data$FSI + data$Population + data$Cases + data
$Deaths + data$Deaths_Per_Capita + data$Cases_Per_Capita)
summary(model)

##
## Call:
## lm(formula = data$Restrictive ~ data$FSI + data$Population +
##     data$Cases + data$Deaths + data$Deaths_Per_Capita + data$Cases_Per_Cap
ita)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.16858 -0.05429 -0.03158  0.04238  0.26176
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             4.009e-01  1.574e-01   2.548   0.0223 *
## data$FSI                2.699e-03  1.660e-03   1.626   0.1248
## data$Population        -9.120e-10  1.664e-09  -0.548   0.5916
## data$Cases              8.576e-07  5.911e-07   1.451   0.1674
## data$Deaths            -1.502e-05  1.622e-05  -0.926   0.3690
## data$Deaths_Per_Capita  1.635e+02  1.088e+03   0.150   0.8826
## data$Cases_Per_Capita  -4.364e+00  3.988e+00  -1.094   0.2910
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1233 on 15 degrees of freedom
## Multiple R-squared:  0.3474, Adjusted R-squared:  0.08631
## F-statistic: 1.331 on 6 and 15 DF,  p-value: 0.3036
```

Now, even in this case, our one-sided p-value is 0.0624, which is fairly small. Nonetheless, this output isn't necessarily useful to us, as there is collinearity. Other multiple regressions:

```
## Call:
## lm(formula = data$Stringency ~ data$FSI + data$Population + data$Cases +
##     data$Deaths + data$Deaths_Per_Capita)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.490  -4.667   1.035   4.501  16.862
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             4.531e+01  9.195e+00   4.928 0.000151 ***
## data$FSI               -2.850e-02  9.568e-02  -0.298 0.769616
## data$Population         1.644e-07  1.091e-07   1.507 0.151178
## data$Cases              3.942e-05  3.876e-05   1.017 0.324350
## data$Deaths            -2.945e-03  9.947e-04  -2.961 0.009197 **
## data$Deaths_Per_Capita  1.246e+05  6.389e+04   1.950 0.068862 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.099 on 16 degrees of freedom
## Multiple R-squared:  0.4488, Adjusted R-squared:  0.2766
## F-statistic: 2.606 on 5 and 16 DF,  p-value: 0.06586
```

```
## Call:
## lm(formula = data_contact$Contact ~ data_contact$FSI + data_contact$Popula
tion +
##     data_contact$Cases + data_contact$Deaths + data_contact$Deaths_Per_Cap
ita)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.089799 -0.015714 -0.001767  0.013837  0.071884
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    1.620e-01  1.068e-01   1.517   0.1678
## data_contact$FSI               3.863e-03  1.091e-03   3.542   0.0076 **
## data_contact$Population        1.397e-09  1.161e-09   1.204   0.2631
## data_contact$Cases             1.295e-07  8.033e-07   0.161   0.8760
## data_contact$Deaths           -3.860e-07  2.822e-05  -0.014   0.9894
## data_contact$Deaths_Per_Capita -8.630e+02  6.067e+02  -1.422   0.1927
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05086 on 8 degrees of freedom
## Multiple R-squared:  0.8974, Adjusted R-squared:  0.8333
## F-statistic:    14 on 5 and 8 DF,  p-value: 0.0008782
```

Here we lose significance for the stringency case but retain significance for percent contact.

A Non-Parametric Alternative

Here we perform a permutation test using the following function from the R library RVAideMemoire, "perm.cor.test." This function executes a permutation test based on the Pearson correlations. Documentation can be found online.

```
library(RVAideMemoire)

## *** Package RVAideMemoire v 0.9-80 ***

perm.cor.test(data$Restrictive, data$FSI)

##
##  Pearson's product-moment correlation - Permutation test
##
## data:  data$Restrictive and data$FSI
## 999 permutations
## t = 2.2371, p-value = 0.036
## alternative hypothesis: true correlation is not equal to 0


## data:  data$Stringency and data$FSI
## 999 permutations
## t = -1.67, p-value = 0.116
## alternative hypothesis: true correlation is not equal to 0

## data:  data_contact$Contact and data_contact$FSI
## 999 permutations
## t = 5.7304, p-value = 0.002
## alternative hypothesis: true correlation is not equal to 0
## sample estimates:
##       cor
## 0.8557841
```

These provide further strong evidence of our claim. The permutation test is nonparametric and makes no assumptions about the data being Gaussian in any way. The only assumption this test makes is the exchangeability of our data, which we think is reasonable here. In both the Restrictive Policies and Percent Contact tests, we see significant p-values that allow us to reject the idea of no correlation.

National Level Data

Country	Restrictive Policies / Total Policies	Resource-oriented Policies / Total Policies	Other Policies / Total	Resource-oriented / Restrictive Policies
Yemen	0.66	0.31	0.03	0.48
Syria	0.57	0.29	0.14	0.50
Afghanistan	0.53	0.42	0.05	0.79
Iraq	0.57	0.31	0.13	0.54
Egypt	0.42	0.49	0.09	1.16
Lebanon	0.72	0.23	0.05	0.31
Kuwait	0.73	0.22	0.05	0.30
Oman	0.42	0.43	0.16	1.03
Qatar	0.38	0.44	0.18	1.19
United Arab Emirates	0.58	0.36	0.06	0.63
Somalia	0.92	0.00	0.08	0.00
Sudan	0.89	0.09	0.02	0.10
Libya	0.54	0.44	0.03	0.81
Mauritania	0.66	0.34	0.00	0.52
Iran	0.56	0.26	0.18	0.46
Djibouti	0.58	0.38	0.04	0.64
Jordan	0.83	0.04	0.13	0.05
Algeria	0.63	0.32	0.05	0.51
Morocco	0.54	0.36	0.10	0.67
Saudi Arabia	0.71	0.25	0.04	0.36
Tunisia	0.53	0.41	0.05	0.77
Bahrain	0.62	0.34	0.04	0.55

Restrictive ~ FSI

```
model_simple <- lm(data$Restrictive ~ data$FSI)
summary(model_simple)

##
## Call:
## lm(formula = data$Restrictive ~ data$FSI)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.21031 -0.10725 -0.03804  0.08562  0.22686
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.427870   0.109409   3.911 0.000867 ***
## data$FSI    0.002362   0.001319   1.790 0.088577 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.134 on 20 degrees of freedom
## Multiple R-squared:  0.1381, Adjusted R-squared:  0.09501
## F-statistic: 3.205 on 1 and 20 DF,  p-value: 0.08858
```

When looking at only the national level data, we see reduced significance, but when we consider just the one-sided p-value the result is still significant at the 0.05 level.

Permutation Test: Restrictive ~ FSI

```
##  Pearson's product-moment correlation - Permutation test
##
## data:  data$Restrictive and data$FSI
## 999 permutations
## t = 1.7902, p-value = 0.088
## alternative hypothesis: true correlation is not equal to 0
## sample estimates:
##       cor
## 0.3716282
```

Here we see reduced significance, but results are significant at the 0.10 significance level.