

Final Project - Daniel Johnson

Daniel Johnson

December 12, 2018

World Bank Enterprise Survey Data Exploration

A Note on the Data

The World Bank Enterprise Survey is a survey conducted at several year intervals of firms throughout the world. It asks questions of owners relating to a huge variety of subjects ranging from the problems they've had with corruption, infrastructure, or the government to their sales and export numbers. As such it gives a wide swath of both numerical and character data that can be subdivided easily based on country, year, region, or a variety of other geographic or demographic details. The questionnaire from which the data came is included on Github and includes variable identification.

Research

My primary question in my analysis is what are some of the determinants of whether or not a firm experiences crime. I am especially interested in how this might vary across regions or countries. Overall, this project is meant to mimic another project that I'm doing using the Stata statistical analysis program. My overall goal is to see where my skill level with R surpasses that of Stata, and what things I'm able to do in one and not in the other.

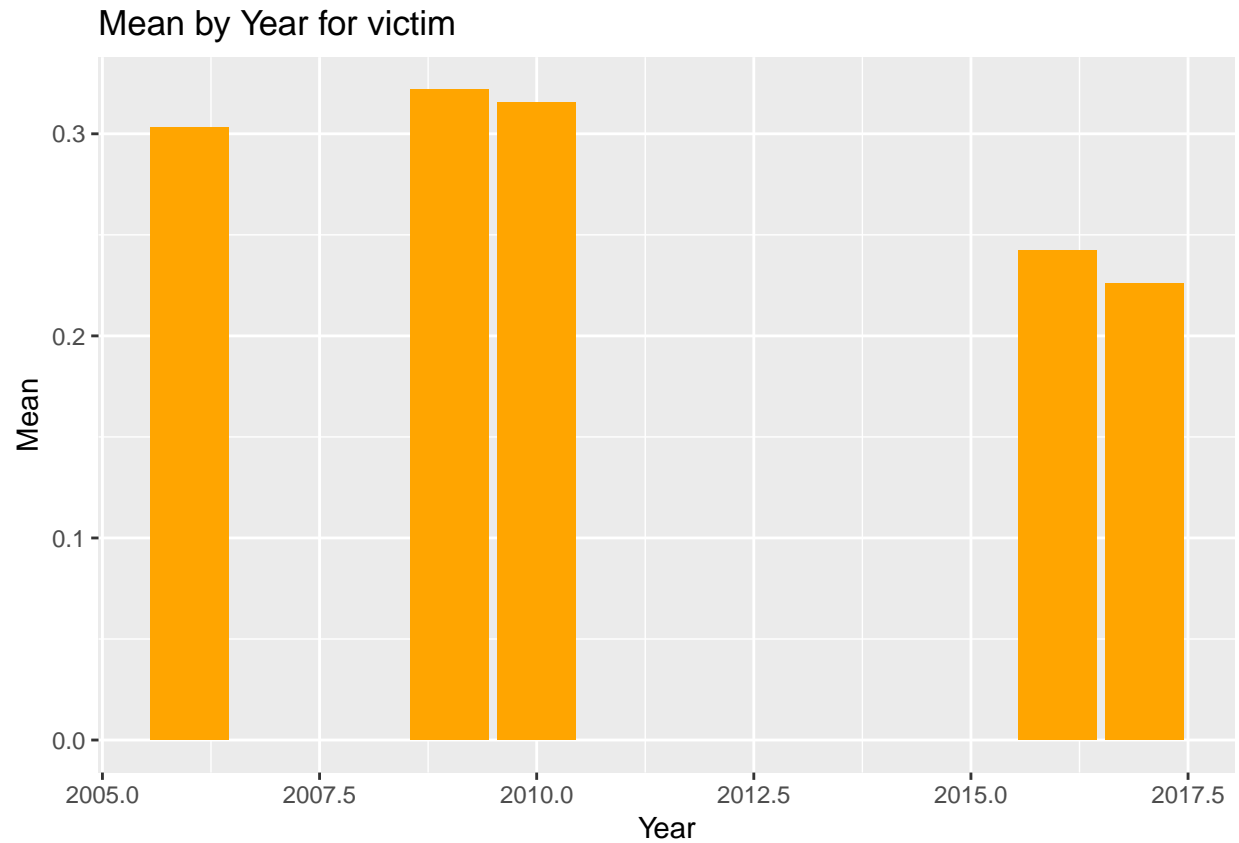
I'll start just by reading in the data

```
survey <- get_survey_df()
```

So to start, I'll look at how the values of certain variables have changed across a region over time

First, the percentage of firms that have been victims of crime in Latin America

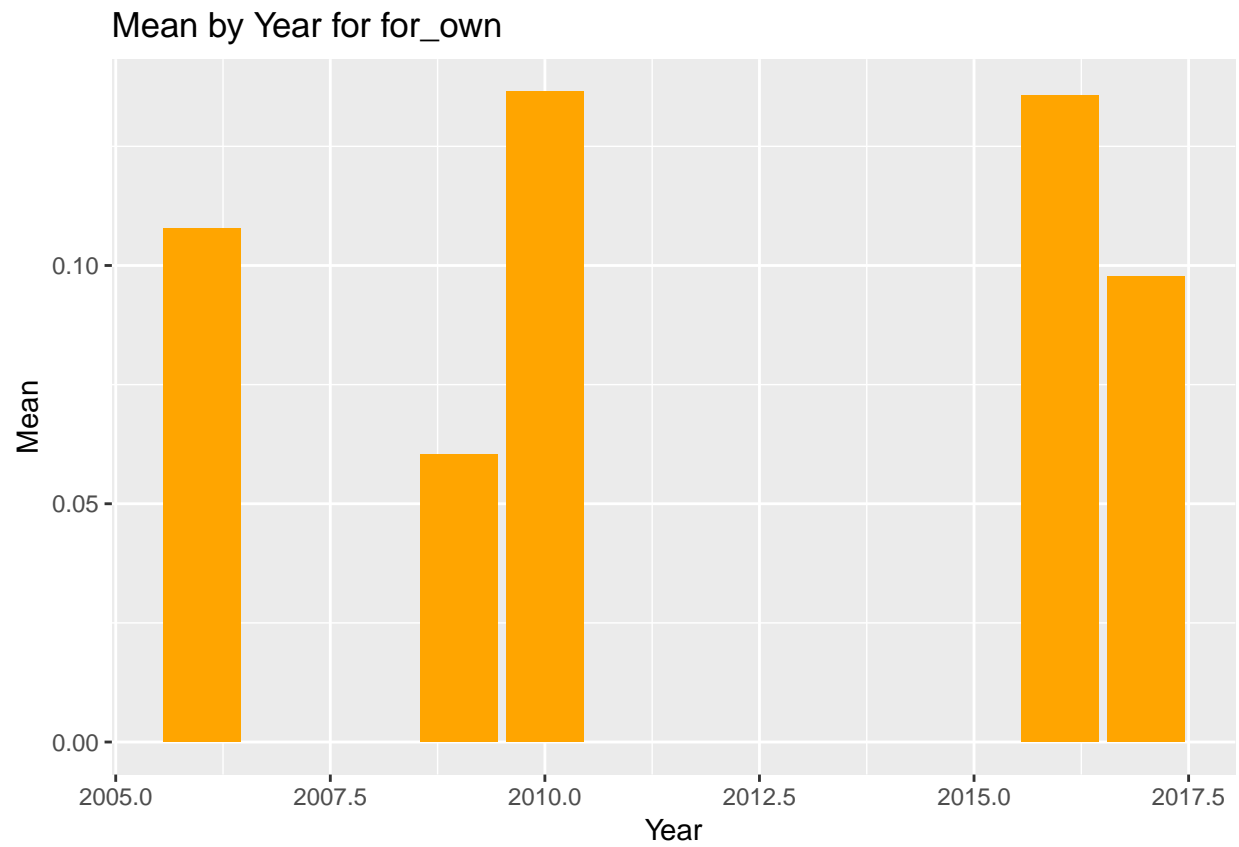
```
get_means_graph_region.Survey(survey, "LAC", "victim")
```



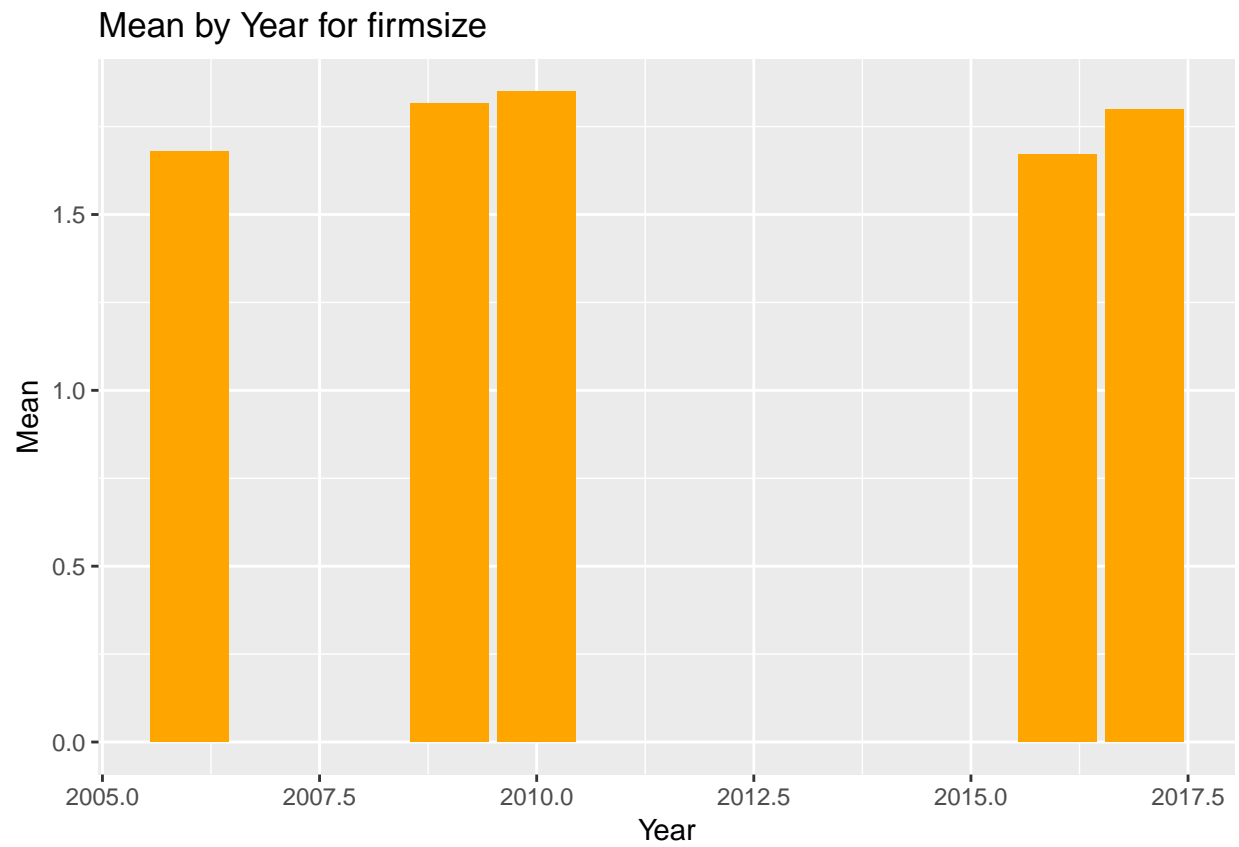
###The missing variables are simply a result of the fact that the survey is not conducted yearly

Now to compare to some potential determinants of crime: Foreign ownership of the company, size of the company, and if the firm is an exporter

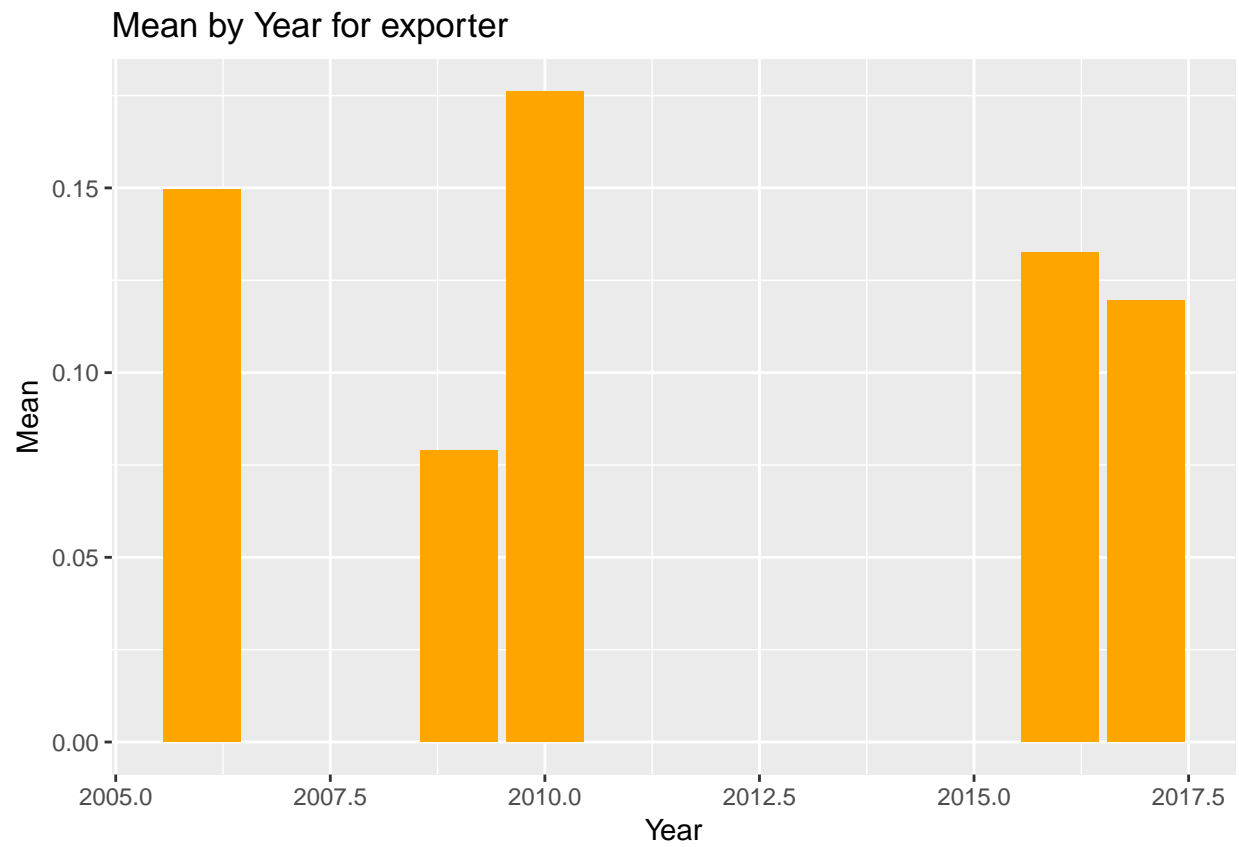
```
get_means_graph_region.Survey(survey, "LAC", "for_own")
```



```
get_means_graph_region.Survey(survey, "LAC", "firmsize")
```

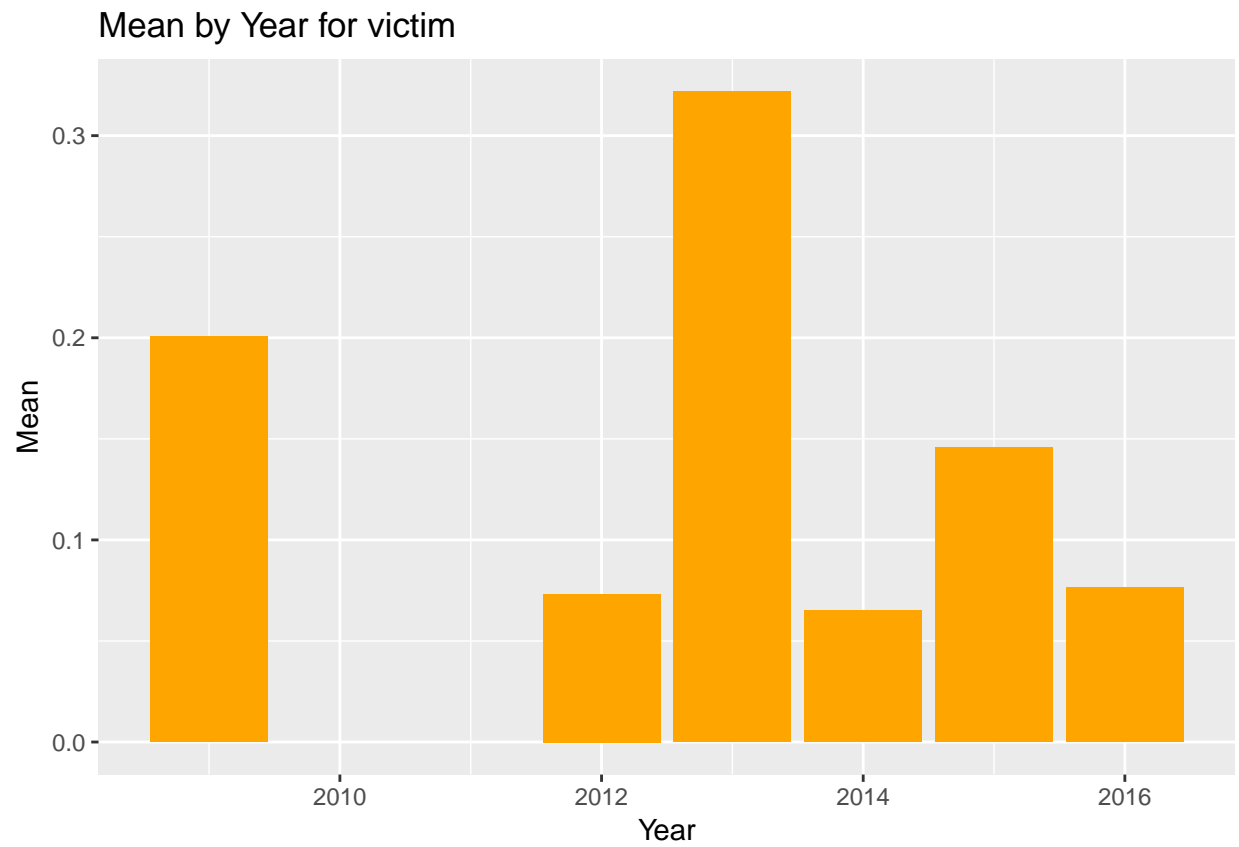


```
get_means_graph_region.Survey(survey, "LAC", "exporter")
```

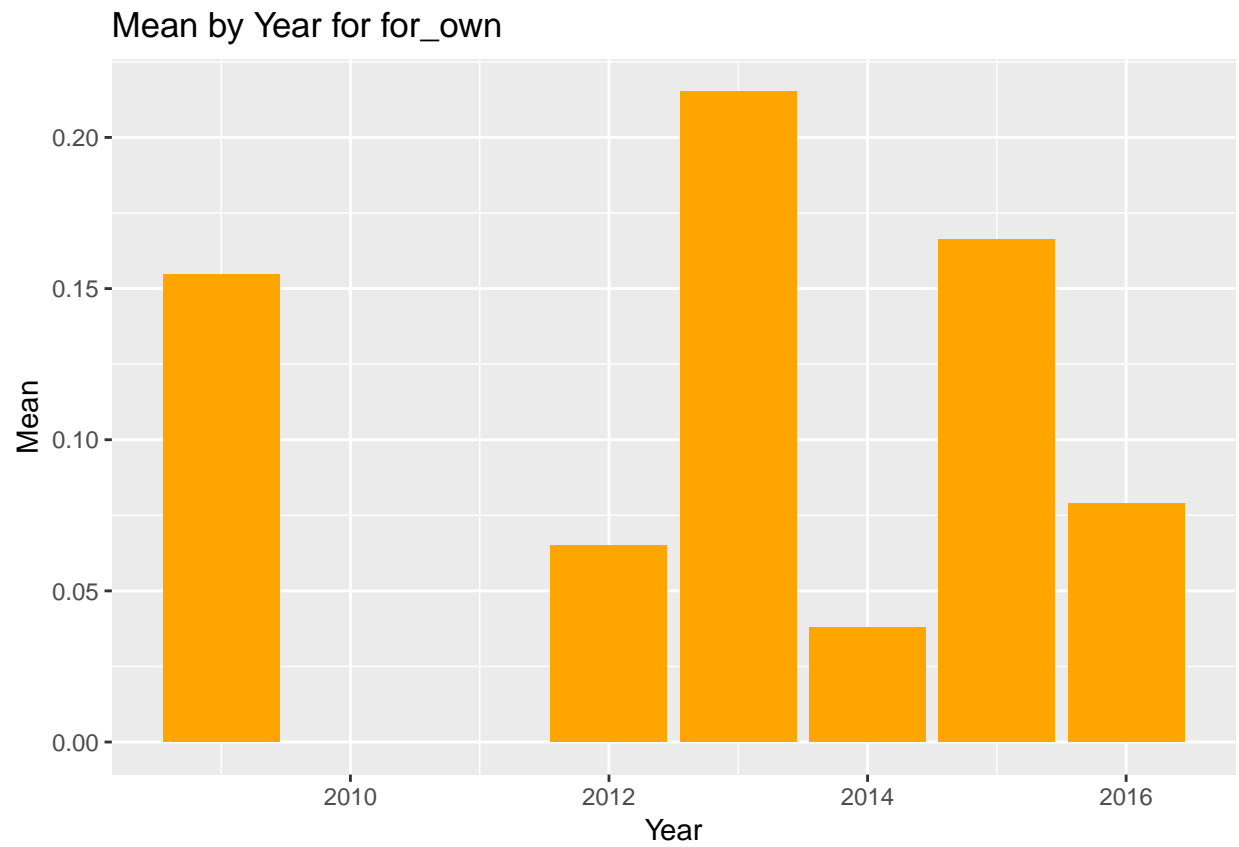


##Now the same sets of comparisons except for the East Asia Pacific region

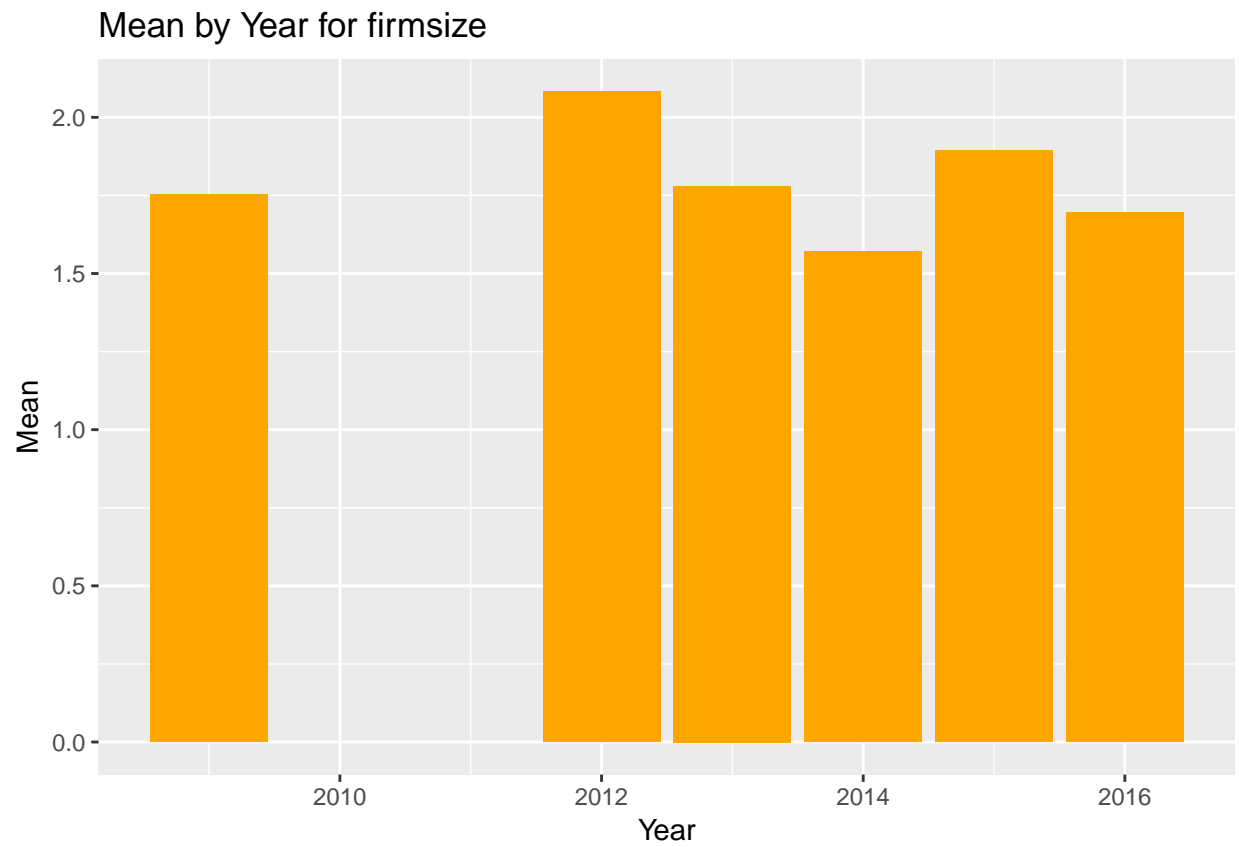
```
get_means_graph_region.Survey(survey, "EAP", "victim")
```



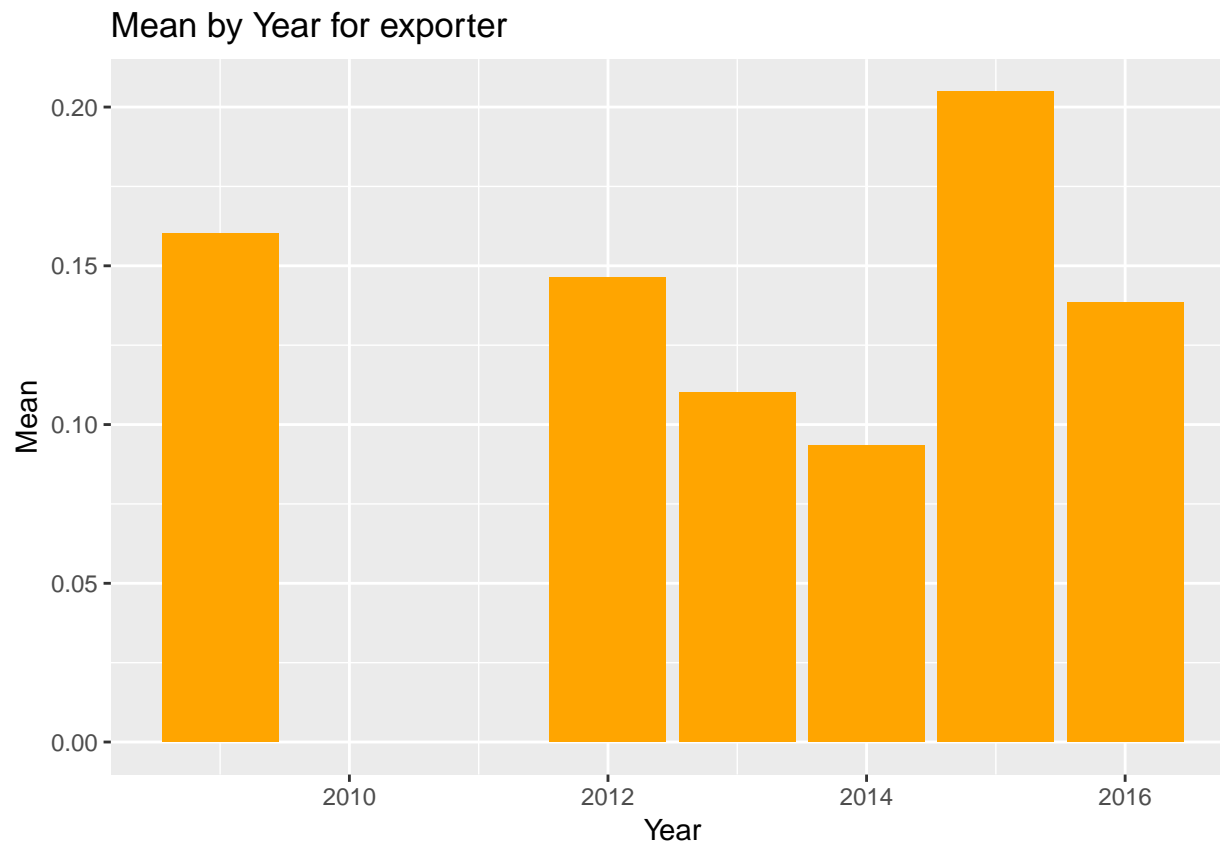
```
get_means_graph_region.Survey(survey, "EAP", "for_own")
```



```
get_means_graph_region.Survey(survey, "EAP", "firmsize")
```



```
get_means_graph_region.Survey(survey, "EAP", "exporter")
```

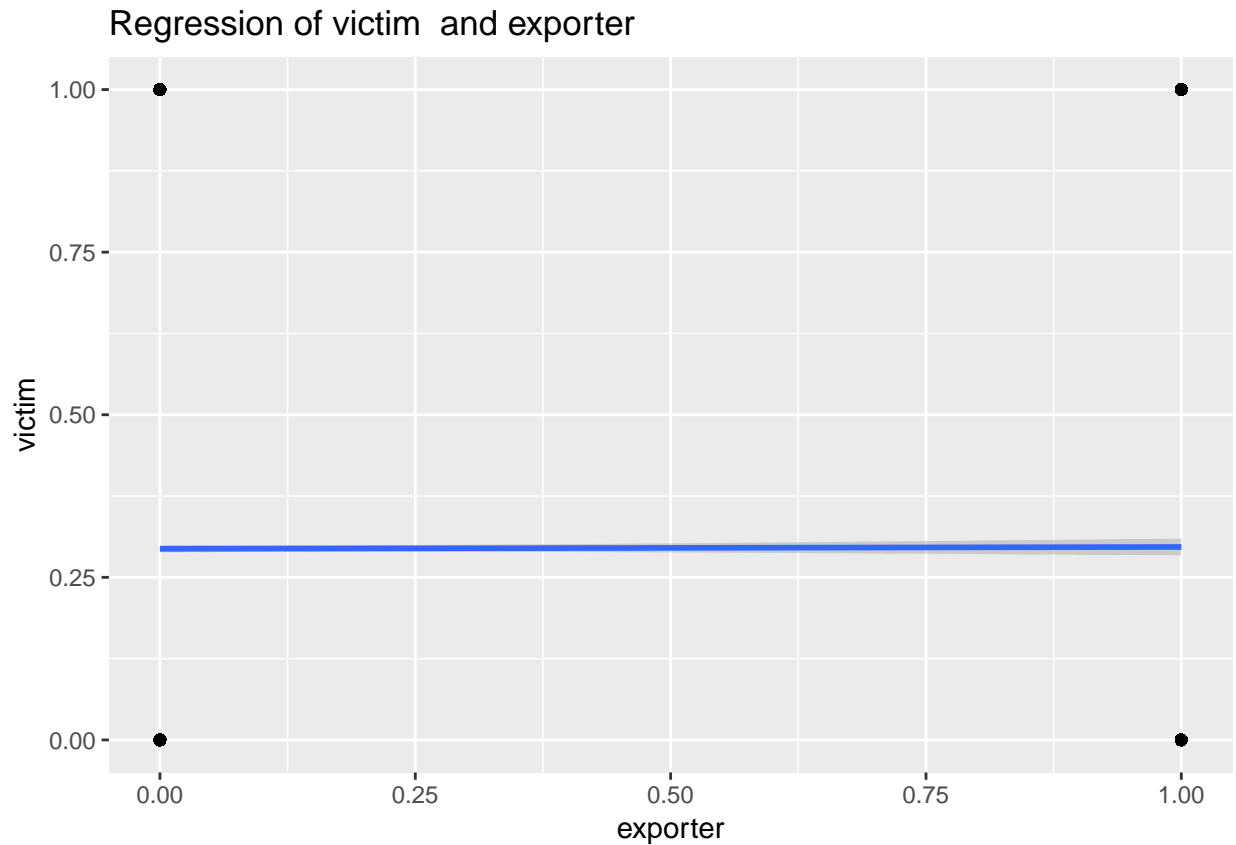
Interestingly, none of the variables seemed, at a glance, to have any obvious correlation in Latin America, but in the East Asia Pacific region the variable for foreign ownership appeared markedly similar to that of the victim. This, obviously, is just an eyeball test, so I'll now examine more objectively through some linear regressions.

First, victim regressed against exporter for Latin America

```
get_regression.Survey(survey, "region", "LAC", "victim", "exporter")
```

```
## [[1]]
##
## Call:
## lm(formula = y ~ x, data = df_reg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.2967 -0.2939 -0.2939  0.7033  0.7061
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.293881   0.002769 106.146  <2e-16 ***
## x            0.002831   0.007133   0.397   0.691
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4557 on 31900 degrees of freedom
```

```
## Multiple R-squared:  4.939e-06, Adjusted R-squared:  -2.641e-05
## F-statistic: 0.1576 on 1 and 31900 DF,  p-value: 0.6914
##
##
## [[2]]
```



Victim versus firmsize

```
get_regression.Survey(survey, "region", "LAC", "victim", "firmsize")
```

```
## [[1]]
##
## Call:
## lm(formula = y ~ x, data = df_reg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.3884 -0.3119 -0.2355  0.6116  0.8410
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.159023   0.006280   25.32  <2e-16 ***
## x             0.076454   0.003245   23.56  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.4519 on 31962 degrees of freedom
## Multiple R-squared:  0.01707,    Adjusted R-squared:  0.01704
## F-statistic: 554.9 on 1 and 31962 DF,  p-value: < 2.2e-16
##
##
## [[2]]
```

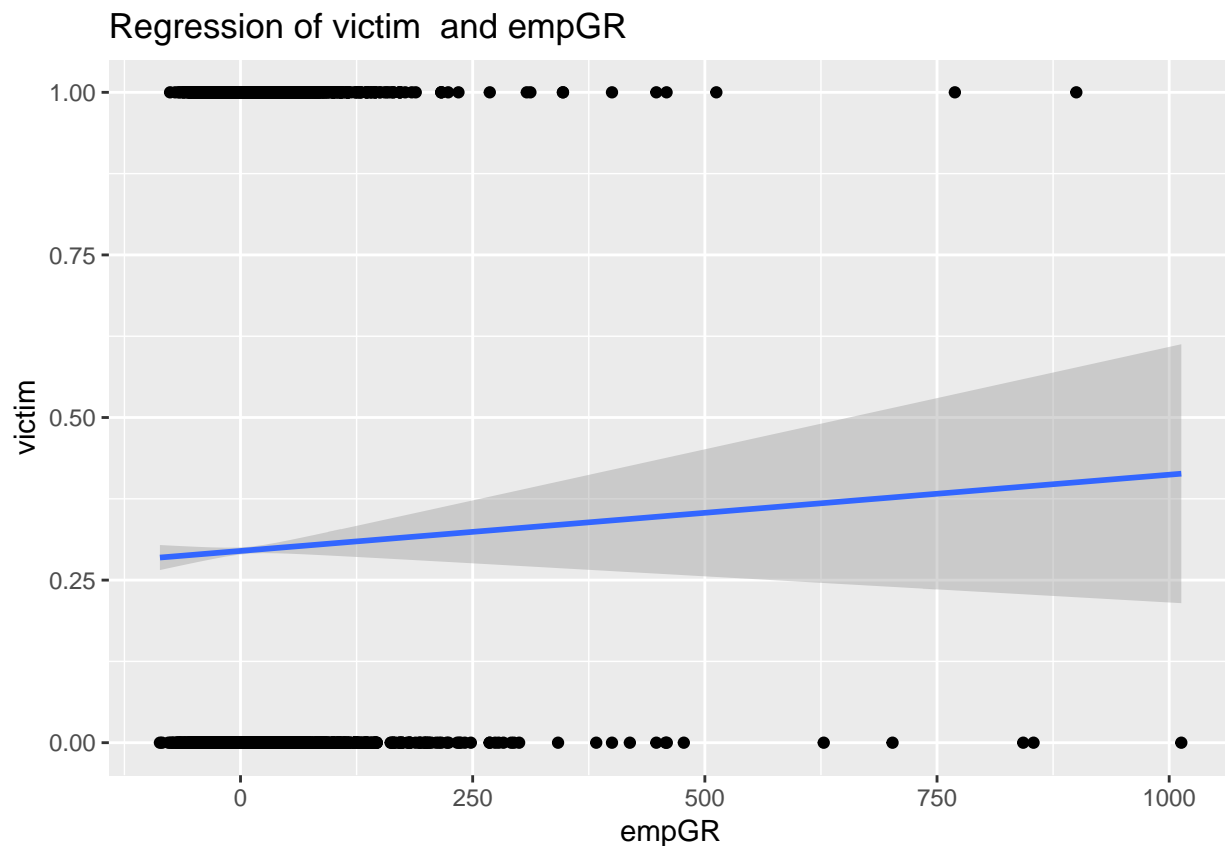


Victim versus employment growth

```
get_regression.Survey(survey, "region", "LAC", "victim", "empGR")
```

```
## [[1]]
##
## Call:
## lm(formula = y ~ x, data = df_reg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.4136 -0.2958 -0.2949  0.7030  0.7140
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.2948533  0.0027368 107.738  <2e-16 ***
## x            0.0001172  0.0001009   1.162    0.245
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.4563 on 29698 degrees of freedom
## Multiple R-squared:  4.544e-05, Adjusted R-squared:  1.177e-05
## F-statistic: 1.35 on 1 and 29698 DF, p-value: 0.2454
##
##
## [[2]]
```

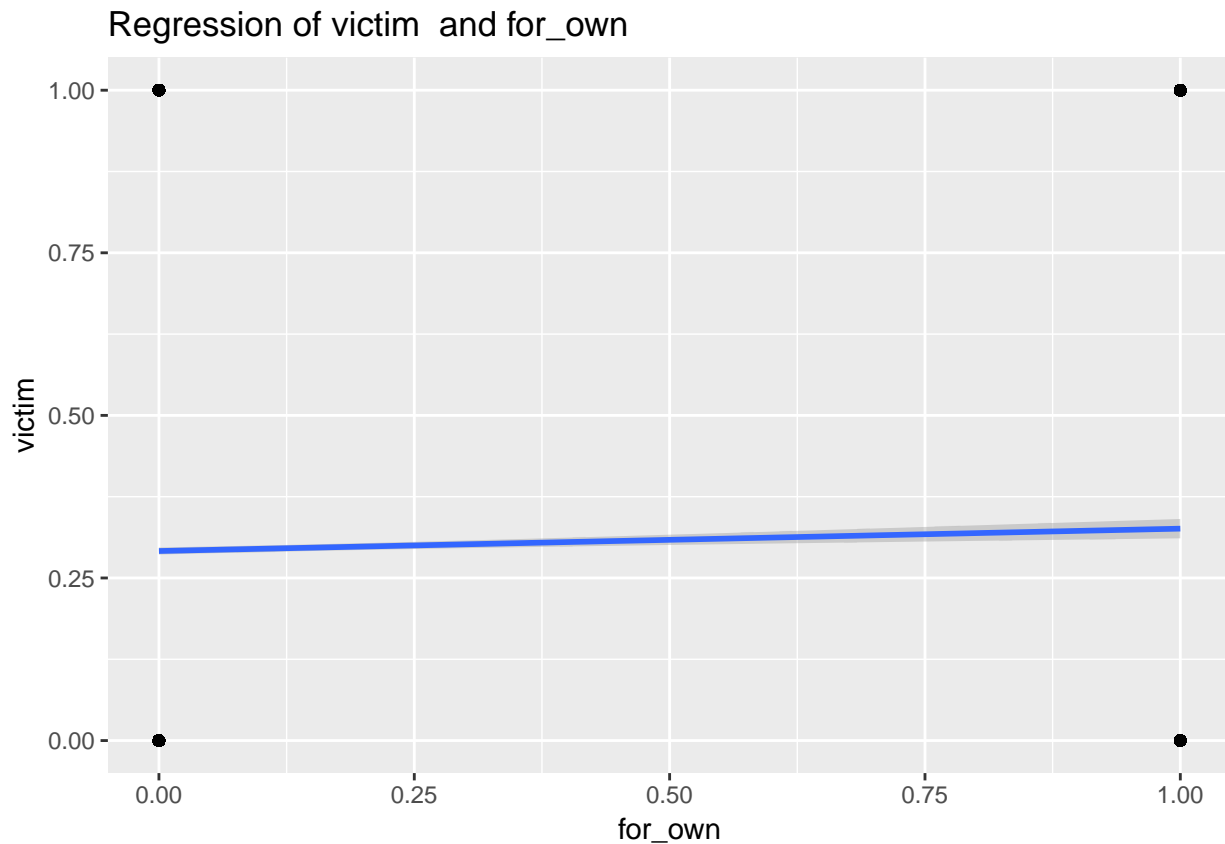


Victim versus foreign ownership

```
get_regression.Survey(survey, "region", "LAC", "victim", "for_own")
```

```
## [[1]]
##
## Call:
## lm(formula = y ~ x, data = df_reg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.3259 -0.2914 -0.2914  0.7086  0.7086
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.291453   0.002768 105.295 < 2e-16 ***
## x             0.034411   0.008076   4.261 2.04e-05 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4561 on 30770 degrees of freedom
## Multiple R-squared:  0.0005897, Adjusted R-squared:  0.0005572
## F-statistic: 18.16 on 1 and 30770 DF, p-value: 2.041e-05
##
##
## [[2]]
```



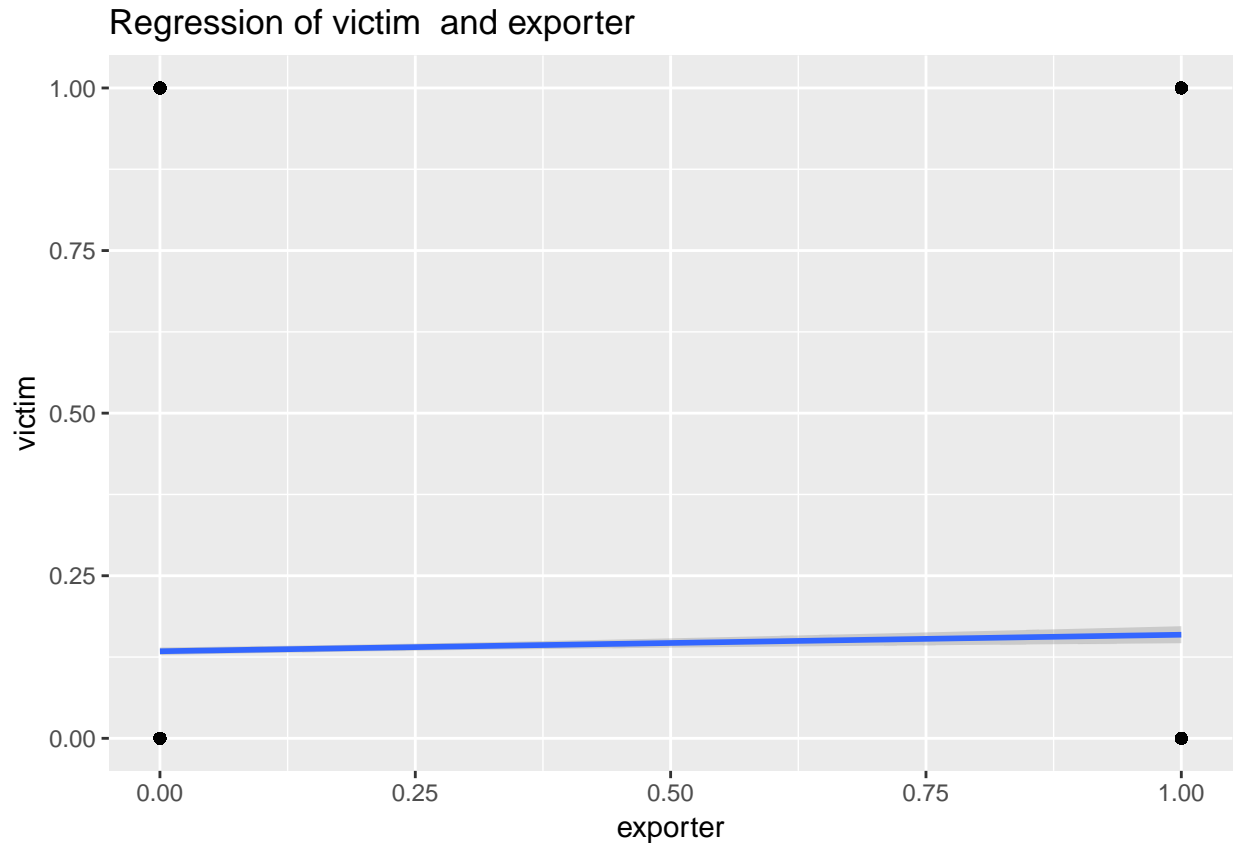
Only foreign ownership is statistically significant here

Repeating in East Asia and the Pacific

```
get_regression.Survey(survey, "region", "EAP", "victim", "exporter")
```

```
## [[1]]
##
## Call:
## lm(formula = y ~ x, data = df_reg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.1593 -0.1339 -0.1339 -0.1339  0.8661
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.133893   0.002946  45.455  < 2e-16 ***
```

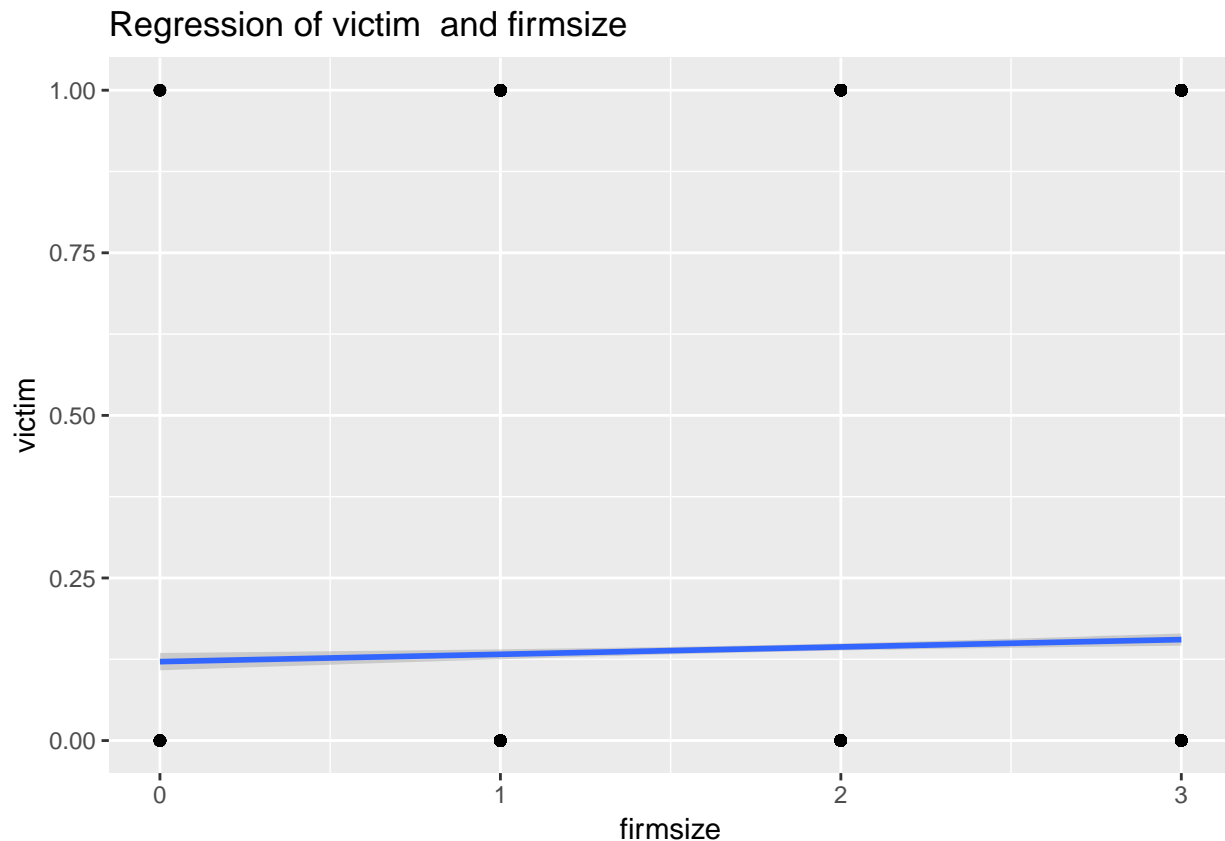
```
## x          0.025419  0.007290  3.487  0.00049 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3448 on 16377 degrees of freedom
## Multiple R-squared:  0.0007418, Adjusted R-squared:  0.0006808
## F-statistic: 12.16 on 1 and 16377 DF, p-value: 0.0004902
##
##
## [[2]]
```



```
get_regression.Survey(survey, "region", "EAP", "victim", "firmsize")
```

```
## [[1]]
##
## Call:
## lm(formula = y ~ x, data = df_reg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.1553 -0.1440 -0.1440 -0.1327  0.8786
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.121357   0.006788  17.878 < 2e-16 ***
## x            0.011305   0.003391   3.334 0.000858 ***
## ---
```

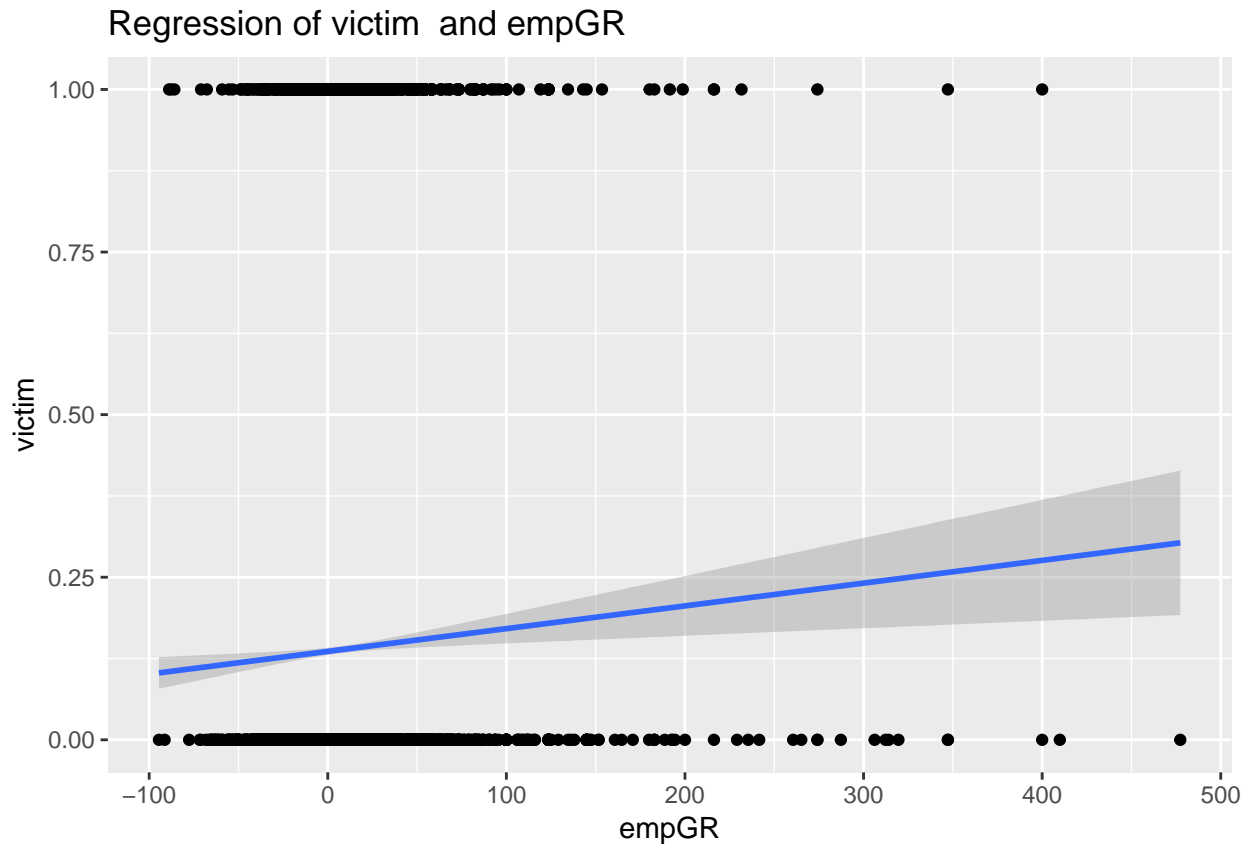
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3491 on 16605 degrees of freedom
## Multiple R-squared:  0.000669,    Adjusted R-squared:  0.0006088
## F-statistic: 11.12 on 1 and 16605 DF,  p-value: 0.0008576
##
##
## [[2]]
```



```
get_regression.Survey(survey, "region", "EAP", "victim", "empGR")
```

```
## [[1]]
##
## Call:
## lm(formula = y ~ x, data = df_reg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.3029 -0.1393 -0.1360 -0.1331  0.8951
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.1359623  0.0028684  47.400  <2e-16 ***
## x            0.0003498  0.0001201   2.912   0.0036 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.345 on 15525 degrees of freedom
## Multiple R-squared:  0.0005458, Adjusted R-squared:  0.0004814
## F-statistic: 8.478 on 1 and 15525 DF, p-value: 0.0036
##
##
## [[2]]
```

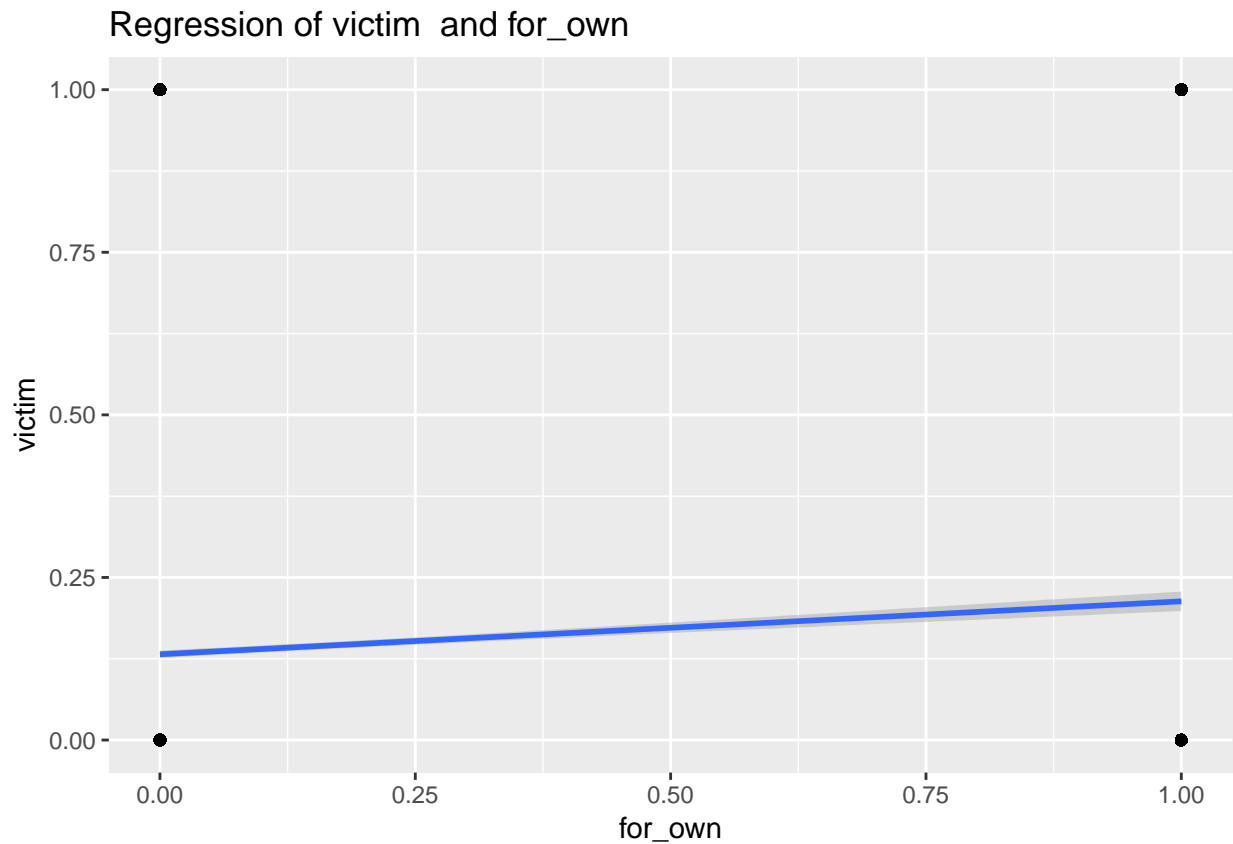


```
get_regression.Survey(survey, "region", "EAP", "victim", "for_own")
```

```
## [[1]]
##
## Call:
## lm(formula = y ~ x, data = df_reg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.2133 -0.1319 -0.1319 -0.1319  0.8681
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.131894   0.002896  45.536  <2e-16 ***
## x             0.081433   0.008155   9.985  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3482 on 16535 degrees of freedom
## Multiple R-squared:  0.005994, Adjusted R-squared:  0.005934
```



```
## F-statistic: 99.7 on 1 and 16535 DF, p-value: < 2.2e-16
##
##
## [[2]]
```



###Here, all are significant

Because it was significant in both, I'm going to focus in on foreign ownership

A multivariable regression to control for other potential explanations of the differences in whether or not a firm is a victim. Using foreign ownership, employment growth, exporter status, firm size, presence of female ownership, and country

Start with Latin America

```
get_lm_multi.Survey(survey, "region", "LAC", "victim", "for_own", "empGR", "exporter", "firm size", "fem
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6, data = df_reg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6167 -0.3268 -0.2411  0.5757  1.0270
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    7.689e-02  3.767e-02   2.041 0.041264 *
```

```
## x1 -1.014e-02 8.999e-03 -1.127 0.259807
## x2 2.362e-05 1.071e-04 0.220 0.825508
## x3 -5.038e-02 7.850e-03 -6.417 1.41e-10 ***
## x4 8.447e-02 3.720e-03 22.708 < 2e-16 ***
## x5 1.374e-02 5.556e-03 2.474 0.013379 *
## x6Argentina 8.934e-02 3.831e-02 2.332 0.019723 *
## x6Bahamas 2.724e-01 5.361e-02 5.082 3.77e-07 ***
## x6Barbados -7.725e-02 5.311e-02 -1.455 0.145781
## x6Belize -1.048e-01 5.242e-02 -1.999 0.045584 *
## x6Bolivia 7.977e-02 3.958e-02 2.015 0.043866 *
## x6Brazil 8.416e-02 3.977e-02 2.116 0.034332 *
## x6Chile 1.323e-01 3.874e-02 3.414 0.000641 ***
## x6Colombia 7.922e-03 3.835e-02 0.207 0.836332
## x6Costarica 1.663e-01 4.262e-02 3.902 9.54e-05 ***
## x6Dominica -1.520e-01 5.263e-02 -2.889 0.003868 **
## x6DominicanRepublic -1.790e-02 4.149e-02 -0.431 0.666118
## x6Ecuador 8.654e-02 3.944e-02 2.195 0.028207 *
## x6ElSalvador 1.528e-01 3.898e-02 3.920 8.88e-05 ***
## x6Grenada 1.372e-01 5.351e-02 2.565 0.010333 *
## x6Guatemala 1.658e-01 3.928e-02 4.221 2.44e-05 ***
## x6Guyana 2.389e-01 5.226e-02 4.572 4.85e-06 ***
## x6Honduras 1.127e-01 3.986e-02 2.828 0.004690 **
## x6Jamaica -2.479e-02 4.578e-02 -0.541 0.588234
## x6Mexico 4.711e-03 3.835e-02 0.123 0.902241
## x6Nicaragua 3.090e-02 3.976e-02 0.777 0.436981
## x6Panama -2.627e-02 4.036e-02 -0.651 0.515135
## x6Paraguay 8.112e-02 3.956e-02 2.051 0.040306 *
## x6Peru 2.184e-02 3.843e-02 0.568 0.569846
## x6StKittsandNevis 2.677e-01 5.393e-02 4.964 6.96e-07 ***
## x6StLucia -1.359e-01 5.218e-02 -2.604 0.009232 **
## x6StVincentandGrenadines 9.365e-02 5.451e-02 1.718 0.085788 .
## x6Suriname -5.859e-02 5.200e-02 -1.127 0.259909
## x6TrinidadandTobago 8.260e-02 4.468e-02 1.849 0.064501 .
## x6Uruguay 1.765e-01 3.930e-02 4.492 7.10e-06 ***
## x6Venezuela 1.610e-01 4.670e-02 3.447 0.000568 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4463 on 28138 degrees of freedom
## Multiple R-squared: 0.04367, Adjusted R-squared: 0.04248
## F-statistic: 36.71 on 35 and 28138 DF, p-value: < 2.2e-16
```

In this regression, foreign ownership (x1) is no longer significant

Trying now with EAP

```
get_lm_multi.Survey(survey, "region", "EAP", "victim", "for_own", "empGR", "exporter", "firmsize", "fem
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6, data = df_reg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```

## -0.74280 -0.18106 -0.07009 -0.02980 1.00837
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.1787296  0.0187310   9.542 < 2e-16 ***
## x1             -0.0086969  0.0090465  -0.961 0.336391
## x2              0.0001783  0.0001167   1.528 0.126586
## x3             -0.0023491  0.0080022  -0.294 0.769098
## x4              0.0216048  0.0037039   5.833 5.55e-09 ***
## x5              0.0047005  0.0055203   0.852 0.394503
## x6China        -0.1762158  0.0188513  -9.348 < 2e-16 ***
## x6Fiji          0.0886425  0.0342106   2.591 0.009577 **
## x6Indonesia    -0.1738018  0.0187381  -9.275 < 2e-16 ***
## x6LaoPDR       -0.0288848  0.0205455  -1.406 0.159776
## x6Malaysia     -0.0314804  0.0211000  -1.492 0.135729
## x6Micronesia   0.5147489  0.0784284   6.563 5.44e-11 ***
## x6Mongolia      0.0081127  0.0215539   0.376 0.706632
## x6Myanmar      -0.1333388  0.0201192  -6.627 3.53e-11 ***
## x6PapuaNewGuinea 0.3983140  0.0447687   8.897 < 2e-16 ***
## x6Philippines  0.0036713  0.0189612   0.194 0.846476
## x6Samoa         0.3307284  0.0378454   8.739 < 2e-16 ***
## x6Solomon Islands 0.1274050  0.0346670   3.675 0.000239 ***
## x6Thailand      -0.2070418  0.0207909  -9.958 < 2e-16 ***
## x6Timor-Leste  -0.0317565  0.0277667  -1.144 0.252770
## x6Tonga         0.0098174  0.0326472   0.301 0.763639
## x6Vanuatu       0.3964732  0.1039393   3.814 0.000137 ***
## x6Vietnam      -0.0769662  0.0192017  -4.008 6.15e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3236 on 15016 degrees of freedom
## Multiple R-squared:  0.08179,    Adjusted R-squared:  0.08044
## F-statistic: 60.79 on 22 and 15016 DF,  p-value: < 2.2e-16

```

Again, foreign ownership is no longer significant. However, firm size was significant across both regions. It seems that with each increase in size level, a firm is about 8.4 percentage points more likely to be the victim of a crime in Latin America, and about 2.2 percentage points more likely in East Asia and the Pacific.

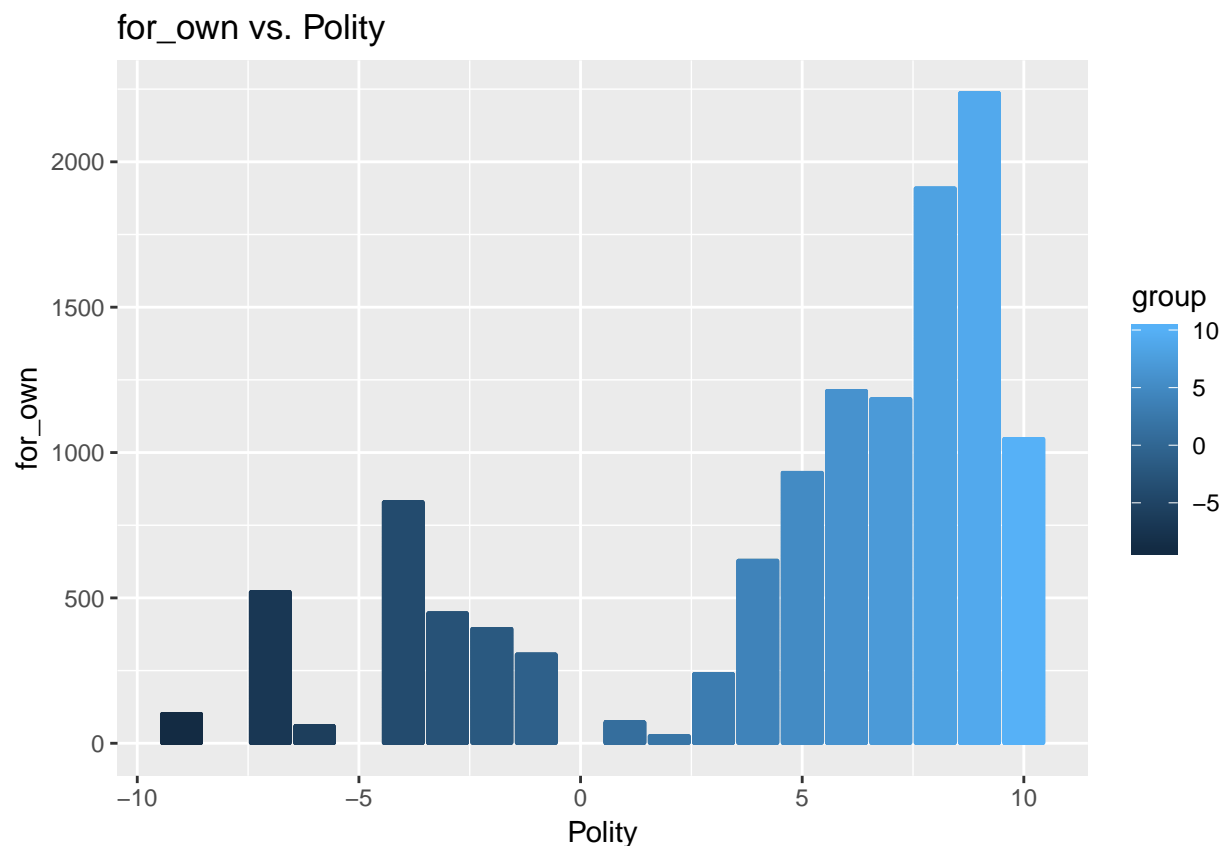
Exploring use of Polity Scores

Polity scores are a ranking on a scale of -10 to 10 of how democratic (closer to ten) or autocratic (closer to -10) a country is. I found these scores through the polity data set and then added them to the World Bank survey for this part of my analysis.

Although the above analysis suggested that foreign ownership is not a significant variable when controlling for other potential explanatory variables, I am interested in returning to it here. In essence, I want to know if countries that are more democratic or more autocratic are, in somewhat loose terms, friendlier to foreign business. I would hypothesize that highly democratic countries are relatively friendly from a sense of global spirit or camaraderie. At the same time, I would expect highly authoritarian countries to be relatively friendly as well, because the government will be careful about protecting any foreign businesses that they actually allow into the country. Thus, I would expect graphing crime against foreign owners against polity scores to have an inverse U shape

Here, I graph polity scores against foreign ownership

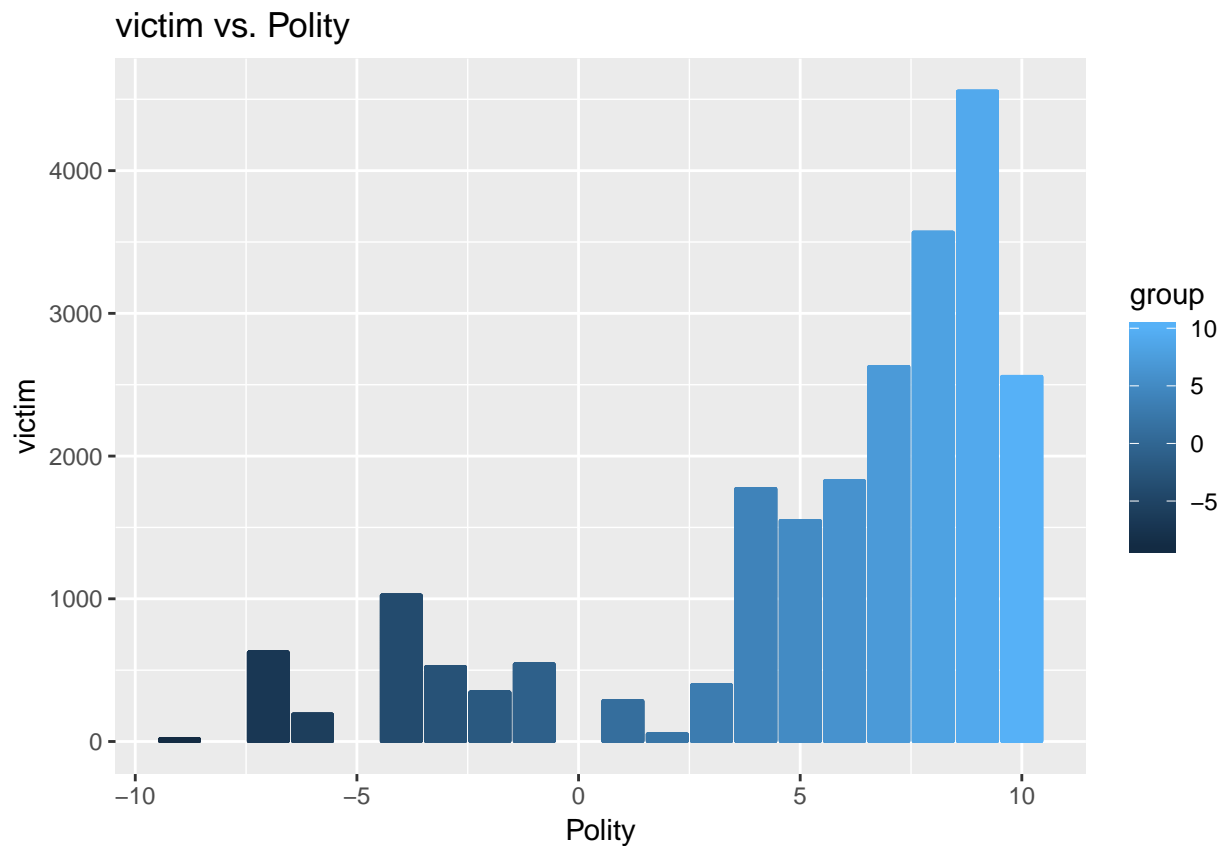
```
graph_two_variables.Survey(survey, "for_own", "Polity", "Polity")
```



This at least offers some support for the idea that autocracies accept foreign firms sparingly, but do incentivize them to some degree, in keeping with economic theory and my hypothesis

Now polity scores against crime

```
graph_two_variables.Survey(survey, "victim", "Polity", "Polity")
```



Roughly speaking, crime also reflects this prediction. These are both simple analyses though. The best way to test this will be through a regression

For Latin America

```
get_lm_multi.Survey(survey, "region", "LAC", "victim", "for_own", "Polity", "ForxPol", "firmsize", "emp
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6, data = df_reg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.4754 -0.3189 -0.2374  0.6000  0.8279
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.265e-01  1.476e-02  8.566  < 2e-16 ***
## x1           7.460e-02  3.789e-02  1.969   0.0490 *
## x2           3.001e-03  1.667e-03  1.800   0.0719 .
```

```
## x3          -1.021e-02  4.688e-03  -2.177   0.0295 *
## x4           8.423e-02  4.001e-03  21.051  < 2e-16 ***
## x5          -6.387e-06  1.068e-04  -0.060   0.9523
## x6          -4.115e-02  8.610e-03  -4.779  1.77e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4515 on 24280 degrees of freedom
## Multiple R-squared:  0.01901,    Adjusted R-squared:  0.01876
## F-statistic:  78.4 on 6 and 24280 DF,  p-value: < 2.2e-16
```

Here, the value for Polity is not significant, but foreign ownership and the interaction term between foreign ownership is significant, offering some support for my hypothesis, and, with a negative coefficient, mainly supporting the idea that if a firm is foreign owned, it is less likely to be a victim of a crime in a more democratic country

Now East Asia and the Pacific

```
get_lm_multi.Survey(survey, "region", "EAP", "victim", "for_own", "Polity", "ForxPol", "firmsize", "empGR", "emp")

##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6, data = df_reg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.31939 -0.14589 -0.11765 -0.08897  0.94561
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0854056  0.0079785  10.704  < 2e-16 ***
## x1           0.0592527  0.0105376   5.623  1.92e-08 ***
## x2           0.0036929  0.0004361   8.467  < 2e-16 ***
## x3          -0.0043195  0.0013294  -3.249  0.001161 **
## x4           0.0147065  0.0040188   3.659  0.000254 ***
## x5           0.0003358  0.0001255   2.676  0.007459 **
## x6           0.0001774  0.0086187   0.021  0.983579
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3315 on 12664 degrees of freedom
## Multiple R-squared:  0.01004,    Adjusted R-squared:  0.009566
## F-statistic:  21.4 on 6 and 12664 DF,  p-value: < 2.2e-16
```

Here, all three are significant, and the coefficient on the interaction term continues to be negative, again suggesting that foreign firms are less likely to be the victim of crimes in more democratic countries. Thus far, however, we have limited analysis to just two regions of particular interest. I'll now look at these regression results in every region

```
get_every_reg.Survey(survey, "region", "victim", "for_own", "Polity", "ForxPol", "firmsize", "empGR", "emp")

## $LAC
##
## Call:
```

```

## lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6, data = df_reg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.4754 -0.3189 -0.2374  0.6000  0.8279
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.265e-01  1.476e-02   8.566 < 2e-16 ***
## x1           7.460e-02  3.789e-02   1.969  0.0490 *
## x2           3.001e-03  1.667e-03   1.800  0.0719 .
## x3          -1.021e-02  4.688e-03  -2.177  0.0295 *
## x4           8.423e-02  4.001e-03  21.051 < 2e-16 ***
## x5          -6.387e-06  1.068e-04  -0.060  0.9523
## x6          -4.115e-02  8.610e-03  -4.779 1.77e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4515 on 24280 degrees of freedom
## Multiple R-squared:  0.01901,    Adjusted R-squared:  0.01876
## F-statistic: 78.4 on 6 and 24280 DF,  p-value: < 2.2e-16
##
##
## $MNA
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6, data = df_reg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.12842 -0.08764 -0.07779 -0.06306  0.95496
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0550919  0.0078474   7.020 2.41e-12 ***
## x1           0.0027257  0.0122589   0.222 0.824055
## x2           0.0023555  0.0007221   3.262 0.001112 **
## x3          -0.0044867  0.0025333  -1.771 0.076586 .
## x4           0.0160618  0.0044151   3.638 0.000277 ***
## x5          -0.0002460  0.0001667  -1.475 0.140194
## x6          -0.0101724  0.0087851  -1.158 0.246936
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2707 on 7348 degrees of freedom
## Multiple R-squared:  0.003146,    Adjusted R-squared:  0.002332
## F-statistic: 3.866 on 6 and 7348 DF,  p-value: 0.0007437
##
##
## $EAP
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6, data = df_reg)
##

```

```

## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.31939 -0.14589 -0.11765 -0.08897  0.94561
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0854056  0.0079785  10.704 < 2e-16 ***
## x1           0.0592527  0.0105376   5.623 1.92e-08 ***
## x2           0.0036929  0.0004361   8.467 < 2e-16 ***
## x3          -0.0043195  0.0013294  -3.249 0.001161 **
## x4           0.0147065  0.0040188   3.659 0.000254 ***
## x5           0.0003358  0.0001255   2.676 0.007459 **
## x6           0.0001774  0.0086187   0.021 0.983579
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3315 on 12664 degrees of freedom
## Multiple R-squared:  0.01004,    Adjusted R-squared:  0.009566
## F-statistic: 21.4 on 6 and 12664 DF,  p-value: < 2.2e-16
##
##
## $ECA
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6, data = df_reg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.3922 -0.2036 -0.1474 -0.0675  1.0309
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.887e-02  6.499e-03   5.981 2.26e-09 ***
## x1          -8.481e-03  1.209e-02  -0.702   0.483
## x2           9.297e-03  4.431e-04  20.982 < 2e-16 ***
## x3           2.255e-04  1.471e-03   0.153   0.878
## x4           6.681e-02  3.415e-03  19.564 < 2e-16 ***
## x5           7.809e-05  7.833e-05   0.997   0.319
## x6          -7.326e-02  6.966e-03 -10.516 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3774 on 23221 degrees of freedom
## Multiple R-squared:  0.03468,    Adjusted R-squared:  0.03443
## F-statistic: 139 on 6 and 23221 DF,  p-value: < 2.2e-16
##
##
## $AFR
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6, data = df_reg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max

```



```
## -0.4560 -0.2408 -0.2106 -0.1557  0.8624
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.109e-01  7.733e-03  14.346 < 2e-16 ***
## x1           1.531e-02  1.179e-02   1.299  0.19400
## x2           5.278e-03  7.575e-04   6.968 3.32e-12 ***
## x3           5.329e-03  1.994e-03   2.673  0.00752 **
## x4           6.798e-02  4.560e-03  14.908 < 2e-16 ***
## x5          -1.392e-04  8.191e-05  -1.700  0.08919 .
## x6           1.998e-02  1.023e-02   1.953  0.05089 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.423 on 21465 degrees of freedom
## Multiple R-squared:  0.019, Adjusted R-squared:  0.01873
## F-statistic: 69.29 on 6 and 21465 DF, p-value: < 2.2e-16
##
##
## $SAR
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6, data = df_reg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.15640 -0.07517 -0.05657 -0.05628  0.97042
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.318e-01  7.049e-03  18.696 < 2e-16 ***
## x1          -2.551e-02  2.919e-02  -0.874  0.3823
## x2          -8.344e-03  6.522e-04 -12.793 < 2e-16 ***
## x3           8.013e-03  4.533e-03   1.768  0.0771 .
## x4          -2.105e-04  2.918e-03  -0.072  0.9425
## x5           5.806e-05  1.045e-04   0.555  0.5786
## x6          -2.602e-02  6.562e-03  -3.965 7.38e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.256 on 14936 degrees of freedom
## Multiple R-squared:  0.01196, Adjusted R-squared:  0.01157
## F-statistic: 30.14 on 6 and 14936 DF, p-value: < 2.2e-16
```

Results here are much less robust. Polity scores are consistently significant, but foreign ownership and the interaction term are much less so. Overall, the results here don't seem to offer particularly strong support for my hypothesis. Finally, however, I'll conduct a regression on the entire survey, this time using region only as a control.

```
get_overall_reg(survey, "victim", "for_own", "Polity", "ForxPol", "firmsize", "empGR", "exporter", "reg

##
## Call:
## lm(formula = Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7, data = df_reg)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.39134 -0.23242 -0.14618 -0.04329  1.03838
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.440e-01  3.742e-03  38.472 < 2e-16 ***
## X1           2.674e-02  5.602e-03   4.772 1.82e-06 ***
## X2           5.030e-03  2.656e-04  18.941 < 2e-16 ***
## X3          -3.671e-04  7.721e-04  -0.475  0.635
## X4           4.979e-02  1.670e-03  29.826 < 2e-16 ***
## X5           3.091e-05  4.170e-05   0.741  0.458
## X6          -3.067e-02  3.562e-03  -8.610 < 2e-16 ***
## X7EAP        -1.180e-01  4.391e-03 -26.862 < 2e-16 ***
## X7ECA        -6.762e-02  3.662e-03 -18.467 < 2e-16 ***
## X7LAC         2.325e-02  3.758e-03   6.187 6.14e-10 ***
## X7MNA        -1.447e-01  5.343e-03 -27.083 < 2e-16 ***
## X7SAR        -1.957e-01  4.238e-03 -46.182 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3815 on 103944 degrees of freedom
## Multiple R-squared:  0.05416,    Adjusted R-squared:  0.05406
## F-statistic: 541.1 on 11 and 103944 DF,  p-value: < 2.2e-16
```

Overall, then, the results do not seem to back up my hypothesis concerning the interaction between polity scores, foreign ownership, and crime. While we can see my hypothesized results in certain regions, they don't hold in many others, and, globally, seem not to hold for the totality of firms in the survey.

As far as my alternate goal, I was able to achieve the same regressions and graphs using R as I was with Stata, but I found that particularly when it came to cleaning data sets and merging in polity scores, R offered me far more extensive capabilities than Stata did. Additionally, graphing processes through ggplot are both easier and more robust than in Stata.