Homework 2: Corona and Democracy

Political Science Research Methods

The Institute of Political Science

Leiden University

Prof. Dr. Babak Rezaee Daryakenari

Alejandro Trenchi (s2593688)

<github.com/trenchi-a>

Spring 2020

**Data: V-Dem Project**

Corona=read.csv("https://raw.githubusercontent.com/trenchi-a/LU\_QA\_2020/master/Datasets/Corona\_mergedV-Dem.csv")  
names(Corona)

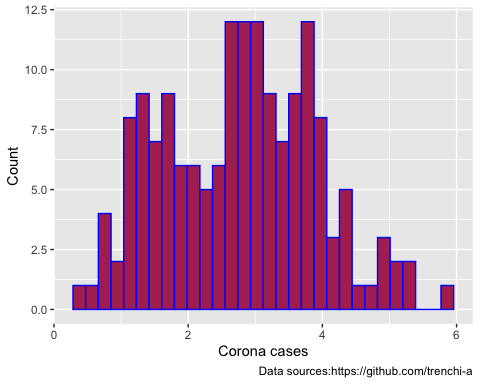
## [1] "country" "ISOcode" "cases" "deaths"   
## [5] "popdata2018" "popdata2018\_log" "cases\_log" "deaths\_log"   
## [9] "ccode" "X\_merge" "country\_name" "country\_id"   
## [13] "v2x\_polyarchy" "v2x\_libdem" "v2x\_partipdem"

**1a. Visualize the distribution of cases\_log, v2x\_polyarchy, v2x\_libdem, v2x\_partipdem.**

library(ggplot2)

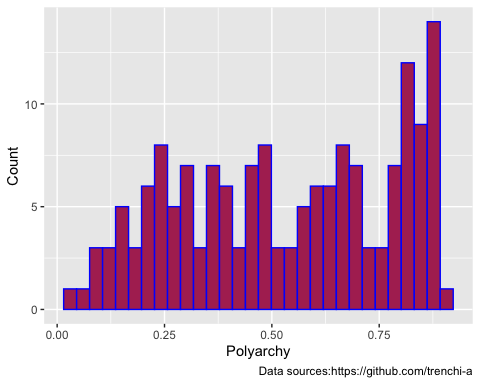
**Corona Cases**

ggplot(Corona,  
 aes(x = cases\_log)) +   
 geom\_histogram(col="blue", fill="maroon")+  
 labs(x = "Corona cases",  
 y = "Count",  
 caption="Data sources:https://github.com/trenchi-a")



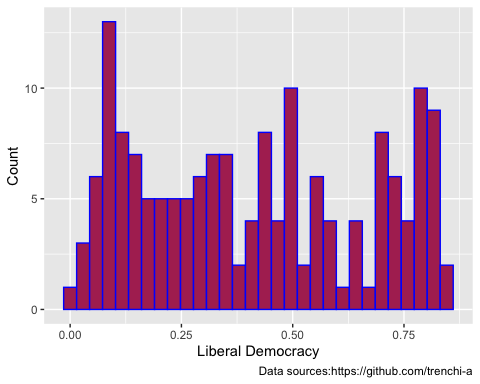
**Polyarchy**

ggplot(Corona,  
 aes(x= v2x\_polyarchy)) +  
 geom\_histogram(col='blue', fill='maroon')+  
 labs(x = "Polyarchy",  
 y = "Count",  
 caption="Data sources:https://github.com/trenchi-a")



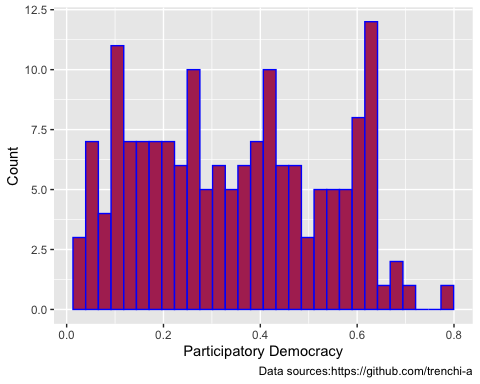
**Liberal Democracy**

ggplot(Corona,  
 aes(x = v2x\_libdem)) +  
 geom\_histogram(col="blue", fill='maroon')+  
 labs(x = "Liberal Democracy",  
 y = "Count",  
 caption="Data sources:https://github.com/trenchi-a")



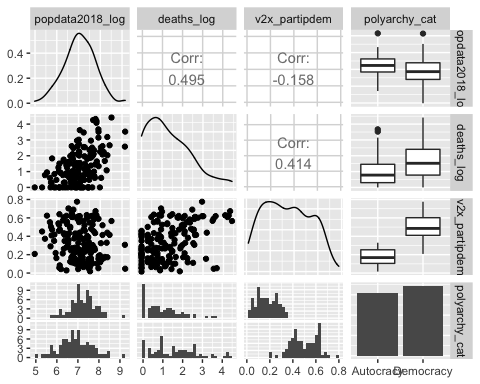
**Participatoy Democracy**

ggplot(Corona,  
 aes(x = v2x\_partipdem)) +  
 geom\_histogram(col="blue", fill="maroon")+  
 labs(x = "Participatory Democracy",  
 y = "Count",  
 caption="Data sources:https://github.com/trenchi-a")



**1b. Reproduce the below plot. (tips: use ggpairs from GGally package).**

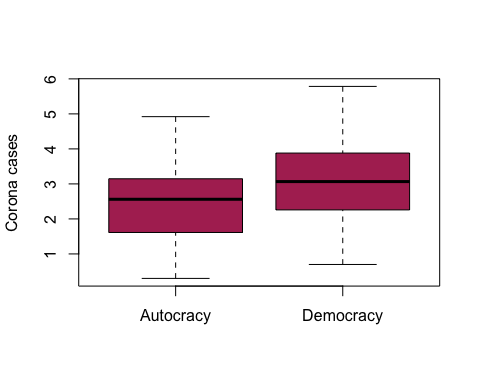
# If v2x\_polyarchy>.5: democracy(1); v2x\_polyarchy<.5: nodemocracy(0)  
Corona$polyarchy\_cat=ifelse(Corona$v2x\_polyarchy> .5, "Democracy", "Autocracy")  
Corona$polyarchy\_cat<-factor(Corona$polyarchy\_cat)  
  
library(GGally)  
  
ggpairs(Corona, columns = c("popdata2018\_log","deaths\_log","v2x\_partipdem","polyarchy\_cat"))



**1c. Define the categorical variables of democracy indices: larger than .5 democracy, otherwise autocracy. Plot the association of these variables with cases\_log.**

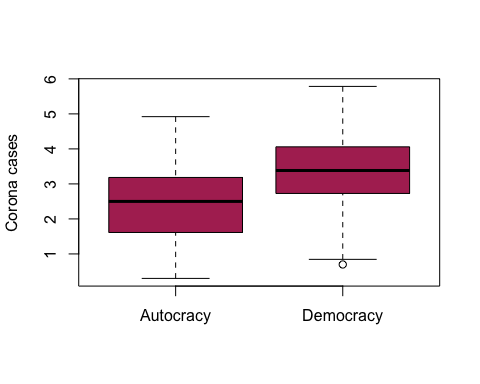
**Polyarchy**

# If v2x\_polyarchy>.5: democracy(1); v2x\_polyarchy<.5: nodemocracy(0)  
Corona$polyarchy\_cat=ifelse(Corona$v2x\_polyarchy> .5, "Democracy", "Autocracy")  
Corona$polyarchy\_cat<-factor(Corona$polyarchy\_cat)  
  
boxplot(Corona$cases\_log~Corona$polyarchy\_cat,  
 col='maroon',  
 xlab='', ylab='Corona cases')



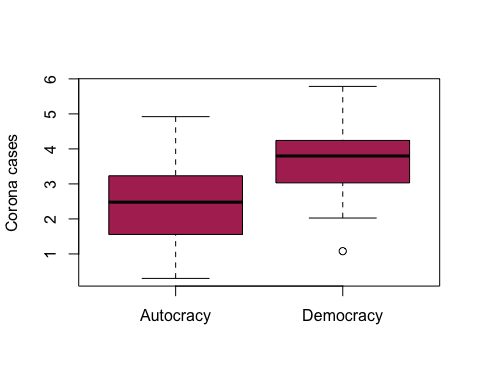
**Liberal Democracy**

# If v2x\_libdem>.5: democracy(1); v2x\_libdem<.5: nodemocracy(0)  
Corona$libdem\_cat=ifelse(Corona$v2x\_libdem> .5, "Democracy", "Autocracy")  
Corona$libdem\_cat<-factor(Corona$libdem\_cat)  
  
boxplot(Corona$cases\_log~Corona$libdem\_cat,  
 col='maroon',  
 xlab='', ylab='Corona cases')



**Participatory Democracy**

# If v2x\_partidem>.5: democracy(1); v2x\_partidem<.5: nodemocracy(0)  
Corona$partidem\_cat=ifelse(Corona$v2x\_partipdem> .5, "Democracy", "Autocracy")  
Corona$partidem\_cat<-factor(Corona$partidem\_cat)  
  
boxplot(Corona$cases\_log~Corona$partidem\_cat,  
 col='maroon',  
 xlab='', ylab='Corona cases')



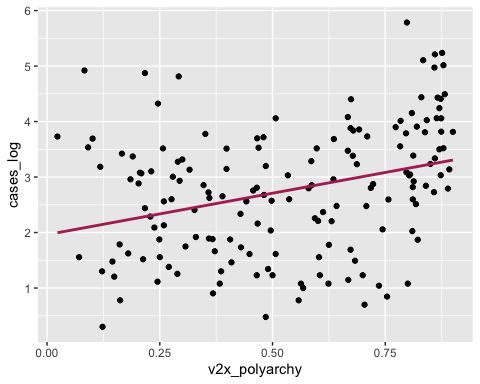
**d.Estimate below models and discuss their results:**

**1. cases\_log = 0 + 1 v2x\_polyarchy +**

OLS\_m1=lm(cases\_log~v2x\_polyarchy, data=Corona)  
summary(OLS\_m1)

##   
## Call:  
## lm(formula = cases\_log ~ v2x\_polyarchy, data = Corona)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.31471 -0.69695 -0.00291 0.79433 2.83502   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.9619 0.2034 9.646 < 2e-16 \*\*\*  
## v2x\_polyarchy 1.4940 0.3450 4.330 2.61e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.091 on 161 degrees of freedom  
## Multiple R-squared: 0.1043, Adjusted R-squared: 0.09876   
## F-statistic: 18.75 on 1 and 161 DF, p-value: 2.608e-05

ggplot(Corona, aes(x=v2x\_polyarchy, y=cases\_log)) +   
 geom\_point() + geom\_smooth(method=lm, color="maroon", se=FALSE)

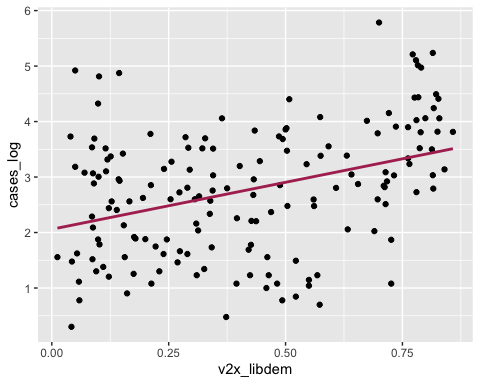


**2. cases\_log = 0 + 1 v2x\_libdem +**

OLS\_m2=lm(cases\_log~v2x\_libdem, data=Corona)  
summary(OLS\_m2)

##   
## Call:  
## lm(formula = cases\_log ~ v2x\_libdem, data = Corona)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.32986 -0.67076 0.06097 0.79488 2.77711   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.0592 0.1605 12.828 < 2e-16 \*\*\*  
## v2x\_libdem 1.6922 0.3302 5.125 8.44e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.069 on 161 degrees of freedom  
## Multiple R-squared: 0.1402, Adjusted R-squared: 0.1349   
## F-statistic: 26.26 on 1 and 161 DF, p-value: 8.441e-07

ggplot(Corona, aes(x=v2x\_libdem, y=cases\_log)) +   
 geom\_point() + geom\_smooth(method=lm, color="maroon", se=FALSE)

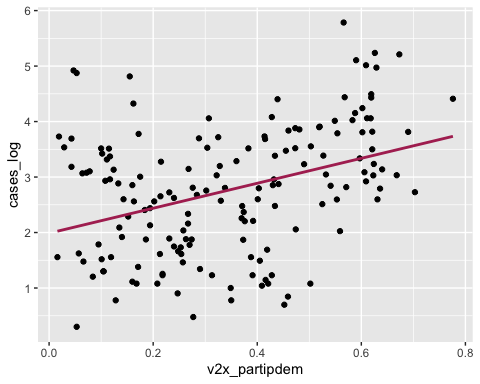


**3. cases\_log = 0 + 1 v2x\_partidem +**

OLS\_m3=lm(cases\_log~v2x\_partipdem, data=Corona)  
summary(OLS\_m3)

##   
## Call:  
## lm(formula = cases\_log ~ v2x\_partipdem, data = Corona)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.30577 -0.71433 0.00464 0.77544 2.82720   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.9880 0.1725 11.523 < 2e-16 \*\*\*  
## v2x\_partipdem 2.2494 0.4389 5.125 8.43e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.069 on 161 degrees of freedom  
## Multiple R-squared: 0.1403, Adjusted R-squared: 0.1349   
## F-statistic: 26.27 on 1 and 161 DF, p-value: 8.434e-07

ggplot(Corona, aes(x=v2x\_partipdem, y=cases\_log)) +   
 geom\_point() + geom\_smooth(method=lm, color="maroon", se=FALSE)



**4. cases\_log = + 1 v2x\_polyarchy + 2 v2x\_libdem + 3 v2x\_partidem +**

OLS\_m4=lm(cases\_log~v2x\_polyarchy + v2x\_libdem +v2x\_partipdem, data=Corona)  
summary(OLS\_m4)

##   
## Call:  
## lm(formula = cases\_log ~ v2x\_polyarchy + v2x\_libdem + v2x\_partipdem,   
## data = Corona)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.20338 -0.69834 -0.02883 0.76655 2.74390   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.6622 0.2562 10.392 < 2e-16 \*\*\*  
## v2x\_polyarchy -5.5996 1.6452 -3.404 0.000842 \*\*\*  
## v2x\_libdem 4.1004 1.4380 2.852 0.004928 \*\*   
## v2x\_partipdem 4.0546 1.6687 2.430 0.016219 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.036 on 159 degrees of freedom  
## Multiple R-squared: 0.2016, Adjusted R-squared: 0.1866   
## F-statistic: 13.39 on 3 and 159 DF, p-value: 7.792e-08

**Explanation**: The results show that there is a positive association between the number of cases of Corona and Democracy in the different types of Democracy indices (Polyarchy, Liberal Democracy and Participatory Democracy).

**Simple Linear Regression Model**

**cases\_log = 1.9619 + 1.4940 v2x\_polyarchy**

This means that if the index of v2x\_polyarchy increases by one unite in a certain country, the cases\_log increases by 1.4940.

**cases\_log = 2.0592 + 1.6922 v2x\_libdem**

This means that if the index of v2x\_libdem increases by one unite in a certain country, the cases\_log increases by 2.0592.

**cases\_log = 1.9880 + 2.2494 vx2\_partipdem**

This means that if the index of v2x\_partipdem increases by one unite in a certain country, the cases\_log increases by 1.9880.

**Multivariate Regression Model**

**cases\_log = 2.6622 + -5.5996 v2x\_polyarchy + 4.1004 v2x\_libdem + 4.0546 v2x\_partipdem**

This means that in this Multivariate Regression Model there is a possitive association between the number of cases of Corona and Liberal Democracy and Participatory Democracy variables. In addition, we found that for Polyarchy there is not significance and therefore it is possible to remove the variable from the model.

**5. There are two solutions to this problem in regression analysis. First, you can directly add population as one of the control variables:**

**cases\_log = 0 + 1 v2x\_polyarchy + 2 popdata2018 +**

OLS\_m5=lm(cases\_log~v2x\_polyarchy + popdata2018, data=Corona)  
summary(OLS\_m5)

##   
## Call:  
## lm(formula = cases\_log ~ v2x\_polyarchy + popdata2018, data = Corona)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.2351 -0.6518 0.0457 0.8279 2.5522   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.769e+00 2.005e-01 8.823 1.90e-15 \*\*\*  
## v2x\_polyarchy 1.654e+00 3.326e-01 4.972 1.70e-06 \*\*\*  
## popdata2018 2.371e-09 5.233e-10 4.531 1.15e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.032 on 159 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.2024, Adjusted R-squared: 0.1924   
## F-statistic: 20.18 on 2 and 159 DF, p-value: 1.554e-08

**Another solution is to normalize the number of cases by dividing the number Corona cases by population size, the cases per population, and use it as dependent variable.**

**norm\_case = 0 + 1 v2x\_polyarchy + 2 popdata2018 +**

Corona$norm\_case=Corona$cases\_log/Corona$popdata2018\_log  
  
OLS\_m6=lm(norm\_case~v2x\_polyarchy, data=Corona)  
summary(OLS\_m6)

##   
## Call:  
## lm(formula = norm\_case ~ v2x\_polyarchy, data = Corona)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.30767 -0.10060 0.01564 0.10220 0.30345   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.26082 0.02589 10.075 < 2e-16 \*\*\*  
## v2x\_polyarchy 0.23846 0.04378 5.447 1.9e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1369 on 160 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.1564, Adjusted R-squared: 0.1511   
## F-statistic: 29.67 on 1 and 160 DF, p-value: 1.901e-07

**Compare the results**:

In order to better undestand the effect of democracy on the reported cases of Corona, it is necessary to add the population size.

It could be added as a control variable.

**cases\_log = 0 + 1 v2x\_polyarchy + 2 popdata2018 +**

or as a dependent variable by normalizing the number of cases by dividing the number Corona cases by population size.

**norm\_case = 0 + 1 v2x\_polyarchy + 2 popdata2018 +**

The first solution provides the total number of the population size, while the second solution gives the total number of cases per capita.

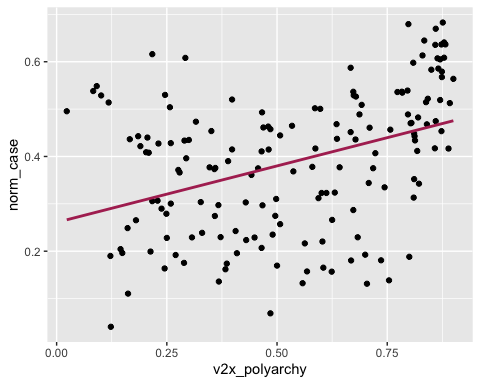
**Repeat your analysis using popdata2018 (normalized version of the equation)**

**norm\_case = 0 + 1 v2x\_polyarchy +**

OLS\_m7=lm(norm\_case~v2x\_polyarchy, data=Corona)  
summary(OLS\_m7)

##   
## Call:  
## lm(formula = norm\_case ~ v2x\_polyarchy, data = Corona)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.30767 -0.10060 0.01564 0.10220 0.30345   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.26082 0.02589 10.075 < 2e-16 \*\*\*  
## v2x\_polyarchy 0.23846 0.04378 5.447 1.9e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1369 on 160 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.1564, Adjusted R-squared: 0.1511   
## F-statistic: 29.67 on 1 and 160 DF, p-value: 1.901e-07

ggplot(Corona, aes(x=v2x\_polyarchy, y=norm\_case)) +   
 geom\_point() + geom\_smooth(method=lm, color="maroon", se=FALSE)

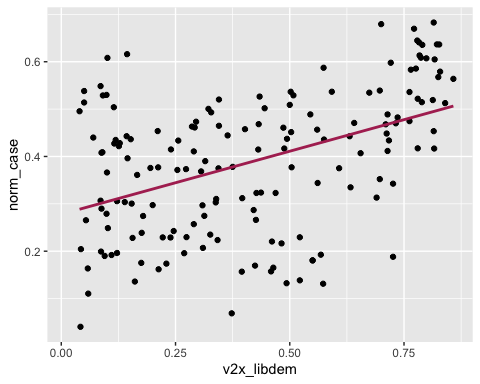


**norm\_case = 0 + 1 v2x\_libdem +**

OLS\_m8=lm(norm\_case~v2x\_libdem, data=Corona)  
summary(OLS\_m8)

##   
## Call:  
## lm(formula = norm\_case ~ v2x\_libdem, data = Corona)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.308474 -0.096410 0.008992 0.106282 0.303292   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.27794 0.02017 13.777 < 2e-16 \*\*\*  
## v2x\_libdem 0.26631 0.04137 6.437 1.35e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1329 on 160 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.2057, Adjusted R-squared: 0.2007   
## F-statistic: 41.44 on 1 and 160 DF, p-value: 1.353e-09

ggplot(Corona, aes(x=v2x\_libdem, y=norm\_case)) +   
 geom\_point() + geom\_smooth(method=lm, color="maroon", se=FALSE)

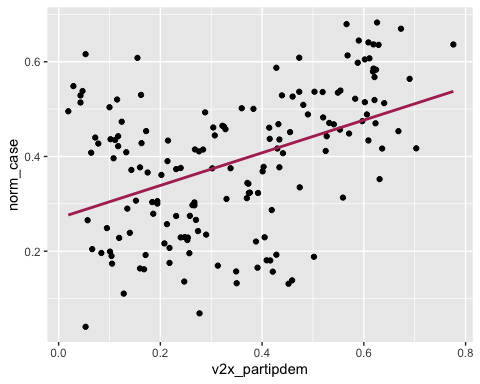


**norm\_case = 0 + 1 v2x\_partidem +**

OLS\_m9=lm(norm\_case~v2x\_partipdem, data=Corona)  
summary(OLS\_m9)

##   
## Call:  
## lm(formula = norm\_case ~ v2x\_partipdem, data = Corona)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.29657 -0.08773 0.01070 0.09810 0.32790   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.26983 0.02186 12.34 < 2e-16 \*\*\*  
## v2x\_partipdem 0.34489 0.05545 6.22 4.16e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1338 on 160 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.1947, Adjusted R-squared: 0.1897   
## F-statistic: 38.69 on 1 and 160 DF, p-value: 4.16e-09

ggplot(Corona, aes(x=v2x\_partipdem, y=norm\_case)) +   
 geom\_point() + geom\_smooth(method=lm, color="maroon", se=FALSE)



**norm\_case = + 1 v2x\_polyarchy + 2 v2x\_libdem + 3 v2x\_partidem +**

OLS\_m10=lm(norm\_case~v2x\_polyarchy + v2x\_libdem +v2x\_partipdem, data=Corona)  
summary(OLS\_m10)

##   
## Call:  
## lm(formula = norm\_case ~ v2x\_polyarchy + v2x\_libdem + v2x\_partipdem,   
## data = Corona)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.29976 -0.09599 0.00163 0.10199 0.32649   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.36168 0.03227 11.207 < 2e-16 \*\*\*  
## v2x\_polyarchy -0.73468 0.20501 -3.584 0.000451 \*\*\*  
## v2x\_libdem 0.63823 0.17882 3.569 0.000475 \*\*\*  
## v2x\_partipdem 0.45178 0.20692 2.183 0.030485 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1285 on 158 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.2662, Adjusted R-squared: 0.2522   
## F-statistic: 19.1 on 3 and 158 DF, p-value: 1.275e-10

**Do the results change? How?**:

After repeating the analysis using popdata2018 in its normalized equation (the number of cases by dividing the number Corona cases by population size), we can state that that there is a similar but a bit stonger correlation between the variables and therefore we can confirm that in order to the population size should definitely be included when analyzing the effect of democracy on the reported Corona cases.