Hi classmate,

I keep going on my individual project.

recap: <https://my.uclaextension.edu/courses/39499/discussion_topics/689063?module_item_id=2237101> (please see my presentation.

Actually, I used them for my individual project, but I am late delivering the presentation, SO, I put my slides as attachments, if you are interested in my project,  do not be shy and please email me.

From the above website I put in, I am intrigued by why Five states have overturned in 2020 in the United States.

As a result, in this discussion, I will put the State of Arizona as an example, and use another database to start today's analysis.

Database: ANES, it's a questionnaire during and post-survey in the 2020 USA election.

In this table, I will use several variables

V201575  <- ANES$V201575 # where are you grew up?  
V201152  <- ANES$V201152 # How would you rate Donald Trump?  
V202110x <- ANES$V202110x # summary : 2020 president vote   
V201127 <- ANES$V201127 # Approve or disapprove of Donald trump handling Job   
V201130 <- ANES$V201130 # Approve or disapprove of Donald trump handling Economy   
V201133 <- ANES$V201133 # Approve or disapprove of Donald trump handling Foreign relations   
V201136 <- ANES$V201136 # Approve or disapprove of Donald trump handling Health care   
V201139 <- ANES$V201139 # Approve or disapprove of Donald trump handling Immigration   
V201142 <- ANES$V201142 # Approve or disapprove of Donald trump handling COVID-19

Then, in order to easily check, I re-name them.

V201575  <- GrewUpState ( Arizona = 4)  
V201152  <- RateForTrump (0~100)  
V202110x <- VoteForIn2020 (Biden = 1, Trump = 2)  
V201127 <- JobHandling (Approve = 1, Disapprove = 2)  
V201130 <- EconomyHandling (Approve = 1, Disapprove = 2)  
V201133 <- ForeignRelationHandling (Approve = 1, Disapprove = 2)  
V201136 <- HealthcareHandling (Approve = 1, Disapprove = 2)  
V201139 <- ImmigrationHandling (Approve = 1, Disapprove = 2)  
V201142 <- COVID19Handling (Approve = 1, Disapprove = 2)

Hence, I start my linear regression journey.

First of all, I use correlation to thoroughly check all the variables.

You can see, ggscatmat can hardly tell me the whole story ( **because all variables are binary data except Rate for Trump and Rate for Biden variable**)

A picture containing table

Description automatically generated

Except that, the corrplot is not good enough to show the correlation because of the same above reason.

Chart

Description automatically generated

As a result, I had to do a linear regression model (**model 1**) directly.

I possess all the other variables with the variable RateForTrump to check which attitude affects the Rate for Trump the most.

Only four variables have a comparatively acceptable P-value ( **JobHandling, EconomyHandling, ForeignRelationHandling, ImmigrationHandling**). However, two of them's P-value are >0.05

Table

Description automatically generatedI use the acceptable four variables to calculate the regression line (model 2) again:

it looks better.

A picture containing text

Description automatically generated

Besides, I try another regression line (model 3).

Text

Description automatically generated

Finally, I calculate the **logistic regression using the model 2 variables**:

Text

Description automatically generated

Hence, I want to calculate if a person all approves Trump's policy or all disapproves Trump's policy, what it will happen?Table

Description automatically generated

it looks normal because if you all appreciate Trump's policy, you will definitely vote for Trump, and vice versa.