Econometrics-1
Data Assignment -1

Gunar Sindhwani
Roll Number - 2020199
Medha Hira
Roll Number -2021265
Kritika Gupta
Roll Number - 2021395
Sumeet Mehra
Roll Number - 2021428

1. The Environment Quality Measure chosen for this assignment is Ground Water level with the Quality Indicator as Ground Water Level.
2. First of all, State,District,YearCode and GroundWaterLevel are extracted from the main df and then NA values are filtered out. The Duplicacies are then removed and aggregated values of groundwater level are found for the entries where state district and yearcode are the same. Later the Unique District Names are found and corresponding District ID are found and then merged with the table and then in order to accommodate the i,t dependency, We updated the ID with ID = 10000*districtID + YEAR, this gave a unique ID to every entry. The Values are averaged in order to obtain a strict district year level dataset.
3. Merged with SDP on the basis of Year and State.
4. Merged with gini values on the basis of districtName
5. Summary Table is assigned below. Link
   The Summary Table contains
   - Summary of the Variables
   - Histogram Plots
   - Box Plots
   - Shape of Distribution Plots
   - Interpretations and Outliers.
6. Table Received after running the Regression on the model

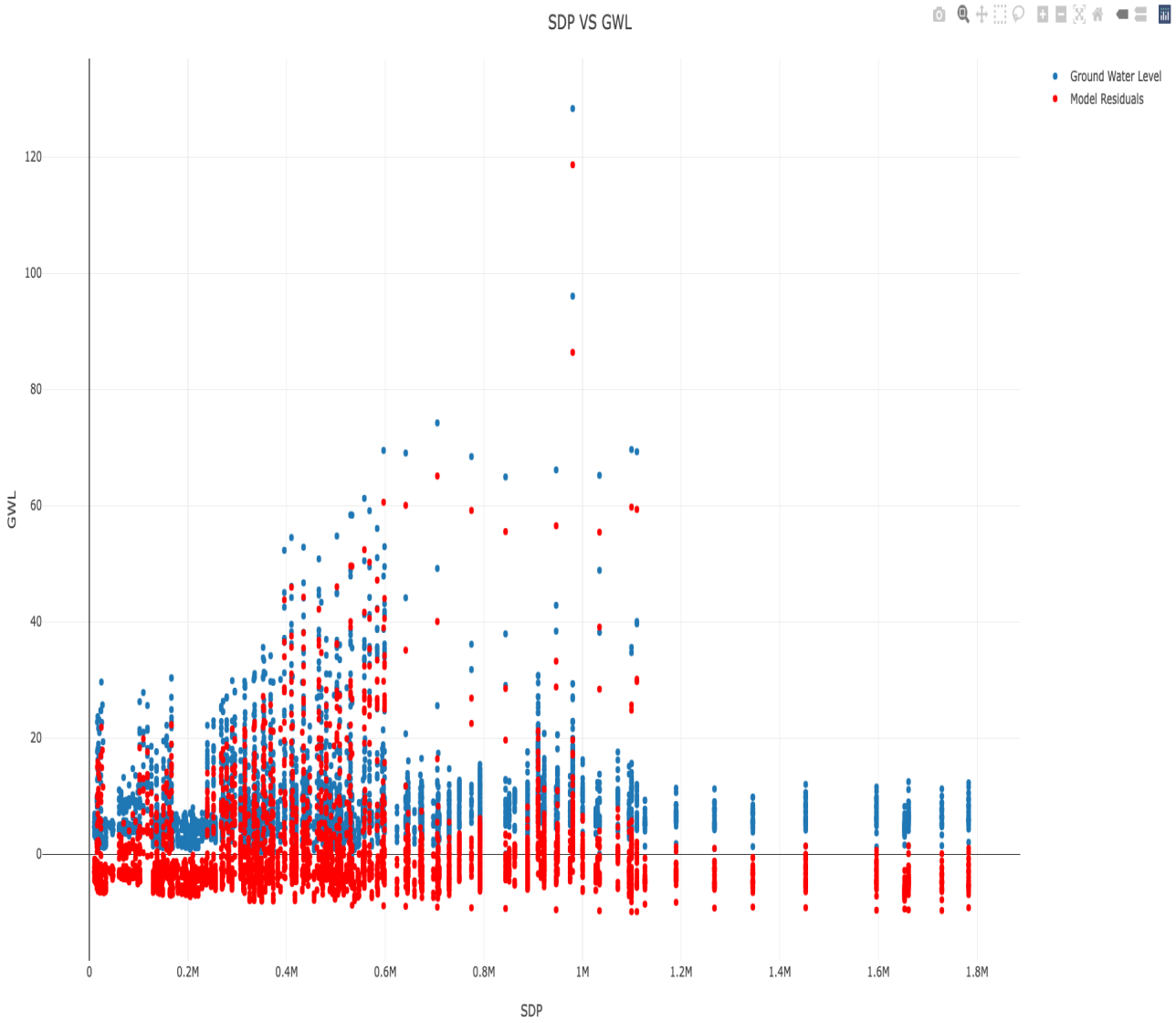
$$Environmental\ Quality\ Indicator\ (EQI)_{i,t} = \beta_0 + \beta_1 SDP_{i,t} + u_{i,t}$$

	Ground.water.level		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	7.74	7.34 – 8.14	**<0.001**
VALUE	0.00	0.00 – 0.00	**<0.001**
Observations	4628		
R^2 / R^2 adjusted	0.009 / 0.008		

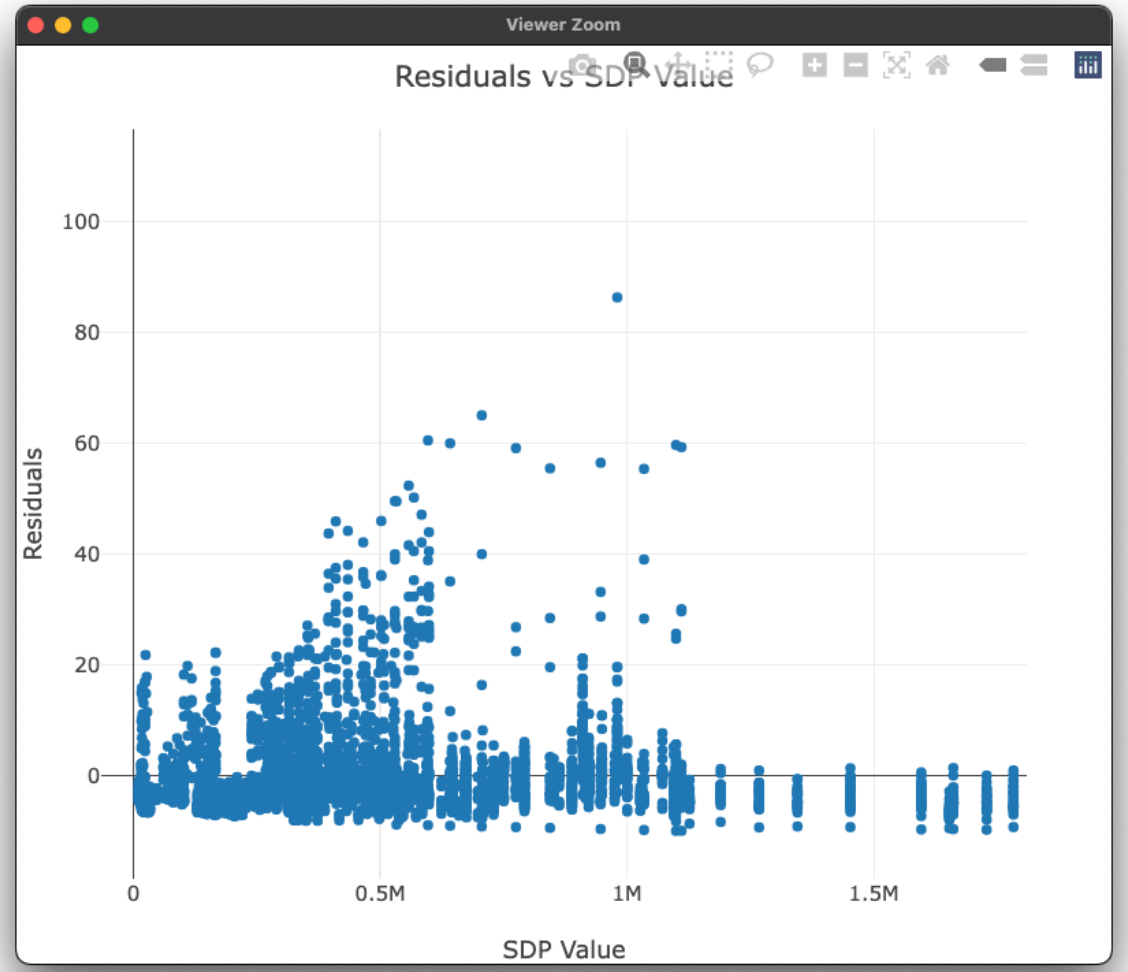
Dependent Variable: Ground.water.value ***Independent Variable***: VALUE (sdpValue) N=4628, R^2= 0.00868	
Coefficients	Estimate (Std. Error)
Intercept ($\hat{\beta}_0$)	7.740e+00 (2.044e-01)
Trend ($\hat{\beta}_1$)	1.976e-06 (3.105e-07)

The Regression Model estimates the linear relationship between the independent variable VALUE (sdpValue) measured as the sdpValue and the dependent variable Ground.water.value. The Intercept $\hat{\beta}_0$ can be interpreted as the estimated average value of Ground Water Level Value when the VALUE or State Domestic Product Value (SDP Value) is Zero (0). $\hat{\beta}_1$ represents the estimated change in Ground Water level value per unit SDP Value (i.e when SDP value changes by 1).

7. Part - A : Trace 1 is Residuals vs Ground.water.level and Trace 2 is SDP Value vs Ground.water.level

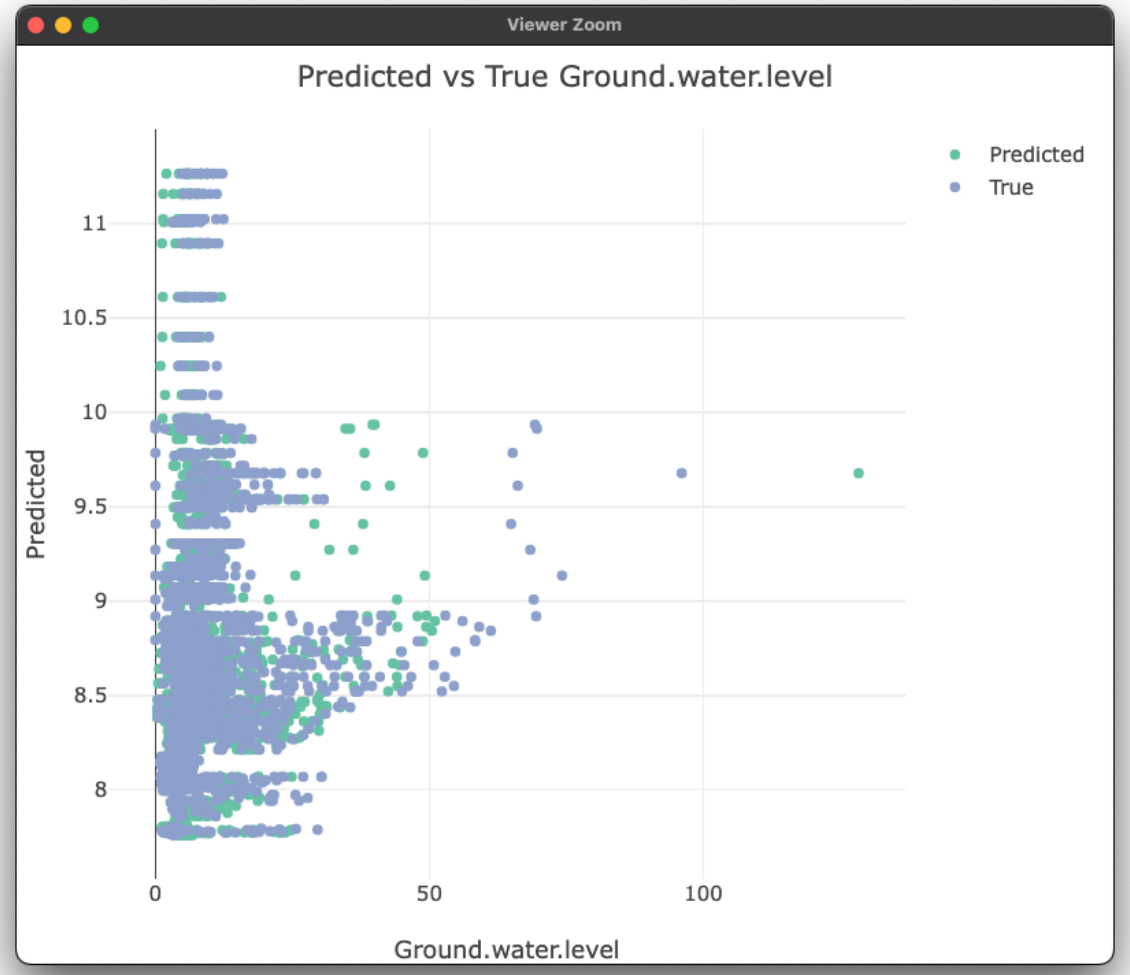


Part - B



Scatter Plot of Residuals vs SDP Value

Part - C

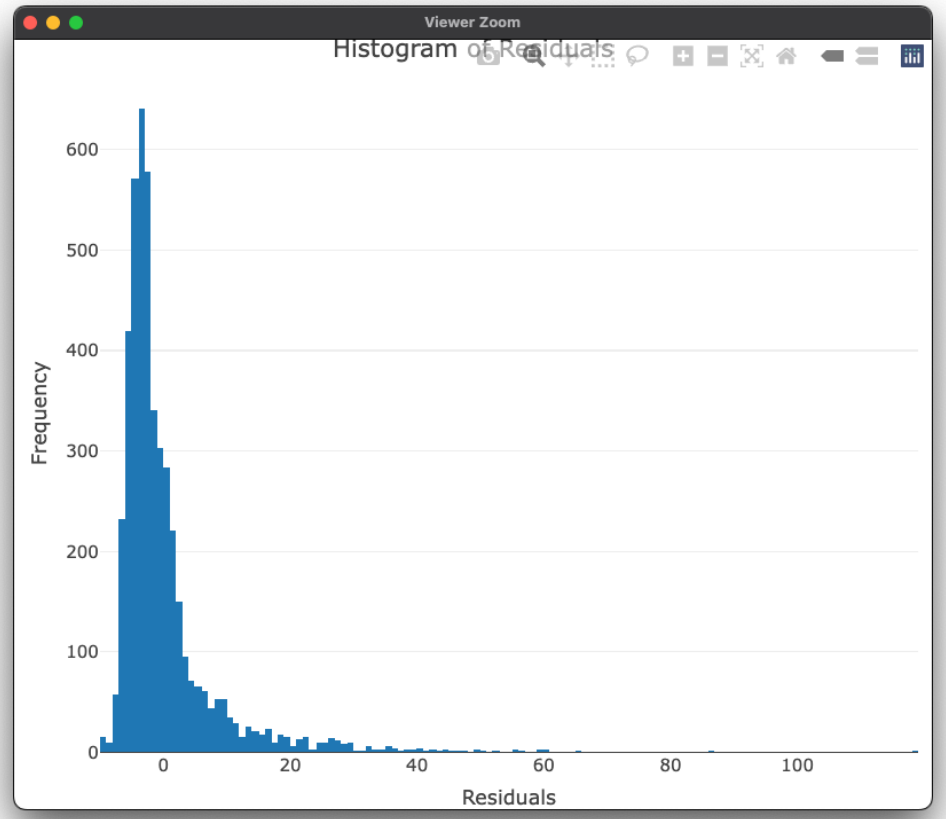


Predicted Values of EQI vs Real Values of EQI

Relation between Plots

1) The R^2 value is low which suggests that the model does not accurately fit the dataset
2) The Residuals are mostly negative values which could be seen in Part A and Part B which would mean that Predicted Values are greater than observed values, which could be seen in Part C. The results of Part A and Part B together prove the observations in Part C.
3) Graph 1 shows the distribution of residuals providing an overall view of the residuals and how they vary with different values.

8. Histogram of the Residuals for the Regression model for Part-6.
   The Inference can be taken that the values on the right hand side of the 0-Line and on the left hand side of the 0-Line are almost the same which may seem that they visually cut each other to represent 0 sum of the Residuals.



9.

   This is later tested on the system as well.

```
> sum(residuals)
[1] -6.090684e-12
```

   The Sum of the Residuals comes out to be -6.09 * 10^(-12) which is a very low value and which tends to zero.
   Probable reasons could be processing errors and the summation due to the floating point error. Thus the residual is approximately 0 only.

10. Table Received after running the Regression on the model

$$EQI_{i,t} = \alpha_0 + \alpha_1 SDP_{i,t} + \alpha_2 SDP_{i.t}^2 + \alpha_3 SDP_{i.t}^3 + \alpha_4 GINI_i + \gamma_{i,t}$$

	Ground.water.level		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.25	-1.52 – 1.01	0.694
VALUE	0.00	0.00 – 0.00	**<0.001**
VALUE^2	-0.00	-0.00 – -0.00	**<0.001**
VALUE^3	0.00	0.00 – 0.00	**0.002**
ginivalue	16.87	12.92 – 20.82	**<0.001**
Observations	4628		
R^2 / R^2 adjusted	0.061 / 0.060		

Dependent Variable: Ground.water.value ***Independent Variables***: VALUE,VALUE^2,VALUE^3,ginivalue N=4628, R^2=0.0.06124	
Coefficients	Estimate (Std. Error)
Intercept ($\hat{\alpha}_0$)	-2.538e-01 (6.458e-01)
($\hat{\alpha}_1$)	1.892e-05(2.177e-06)
($\hat{\alpha}_2$)	-1.707e-11(3.176e-12)
($\hat{\alpha}_3$)	3.898e-18 (1.260e-18)
($\hat{\alpha}_4$)	1.687e+01(2.014e+00)

The Regression Model estimates the relationship between the independent variable VALUE (sdpValue), value^2,value^3, ginivalue measured as the sdpValue and the dependent variable Ground.water.value.

$(\hat{\alpha}_0)$: This is the intercept term, which represents the value of the dependent variable (EQI) when all of the predictor variables (VALUE, VALUE^2, VALUE^3, and ginivalue) are equal to zero. In other words, a represents the y-intercept of the regression line. This is basically the Ground Water Level Value when SDP and Ginivalue are null and zero.

$(\hat{\alpha}_1)$:This is the coefficient for the first predictor variable VALUE), which represents the linear relationship between VALUE and ginivalue. It indicates the change in Ground.water.value for a one-unit increase in VALUE, holding all other variables constant.**

$(\hat{\alpha}_2)$: This is the coefficient for the second predictor variable (VALUE^2), which represents the quadratic relationship between VALUE and Ground.water.value. It indicates the change in Ground.water.value for a one-unit increase in VALUE^2, holding all other variables constant. **

$(\hat{\alpha}_3)$: This is the coefficient for the third predictor variable (VALUE^3), which represents the cubic relationship between VALUE and Ground.water.value . It indicates the change in Ground.water.value for a one-unit increase in VALUE^3, holding all other variables constant.**

$(\hat{\alpha}_4)$: This is the coefficient for the fourth predictor variable (ginivalue),which represents the linear relationship between ginivalue and Ground.water.value. It indicates the change in Ground.water.value for a one unit increase in ginivalue, holding all other variables constant.

** - In practice, it is often difficult to hold some variables constant while manipulating others.It is mathematically not possible to perform an increase in x^2 or x^3 without increasing x or vice versa, but the interpretations of individual coefficients can be only found out when this is performed.
The Variables are not independent of each other,