

National Footprint Accounts

Aira Domingo

Data Mining

INTRODUCTION

- Earth is facing an environmental crisis
 - Global temperature rising
 - Ocean getting warmer
 - Ice sheets shrinking
 - Sea levels rising
 - Many other extreme events





INTRODUCTION

- Humans contribution to the greenhouse effect is the leading cause to global warming
- Greenhouse effect: warming that results from gases radiating from Earth that are trapped in atmosphere
 - Water vapor
 - Methane
 - Nitrous oxide
 - Carbon dioxide
- Since Industrial Revolution, humans have increased atmospheric carbon dioxide concentration by 1/3



INTRODUCTION

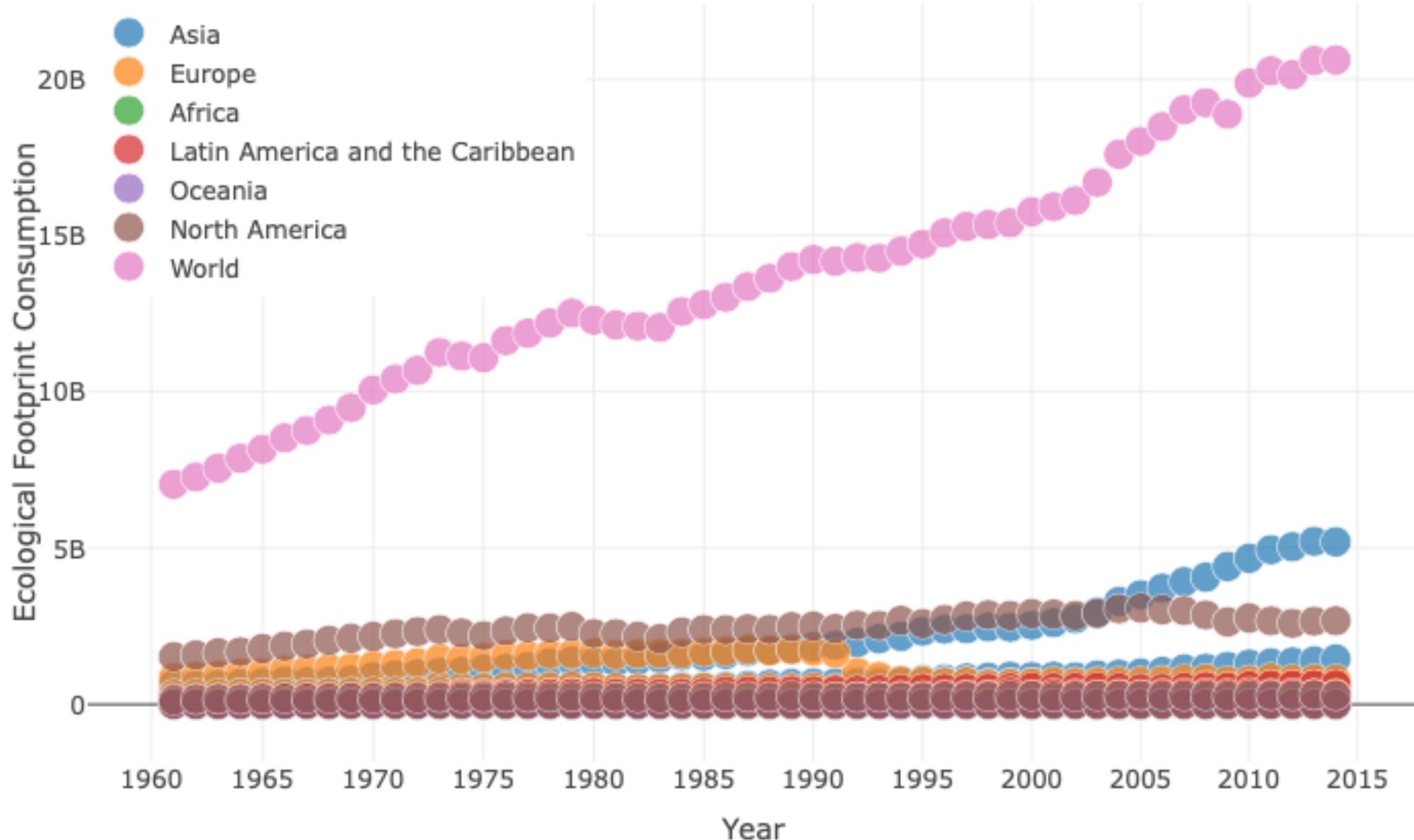
- In US, primary sources of greenhouse gases: transportation, electricity production, industry, commercial and residential, agriculture, land use forestry
- Earth is struggling to absorb carbon dioxide from humanity's increasing use of natural resources
- Earth in ecological deficit
 - Using equivalent of 1.7 Earths to provide our resources and absorb our waste

How is the amount of carbon
emission affected by a country's
land use?

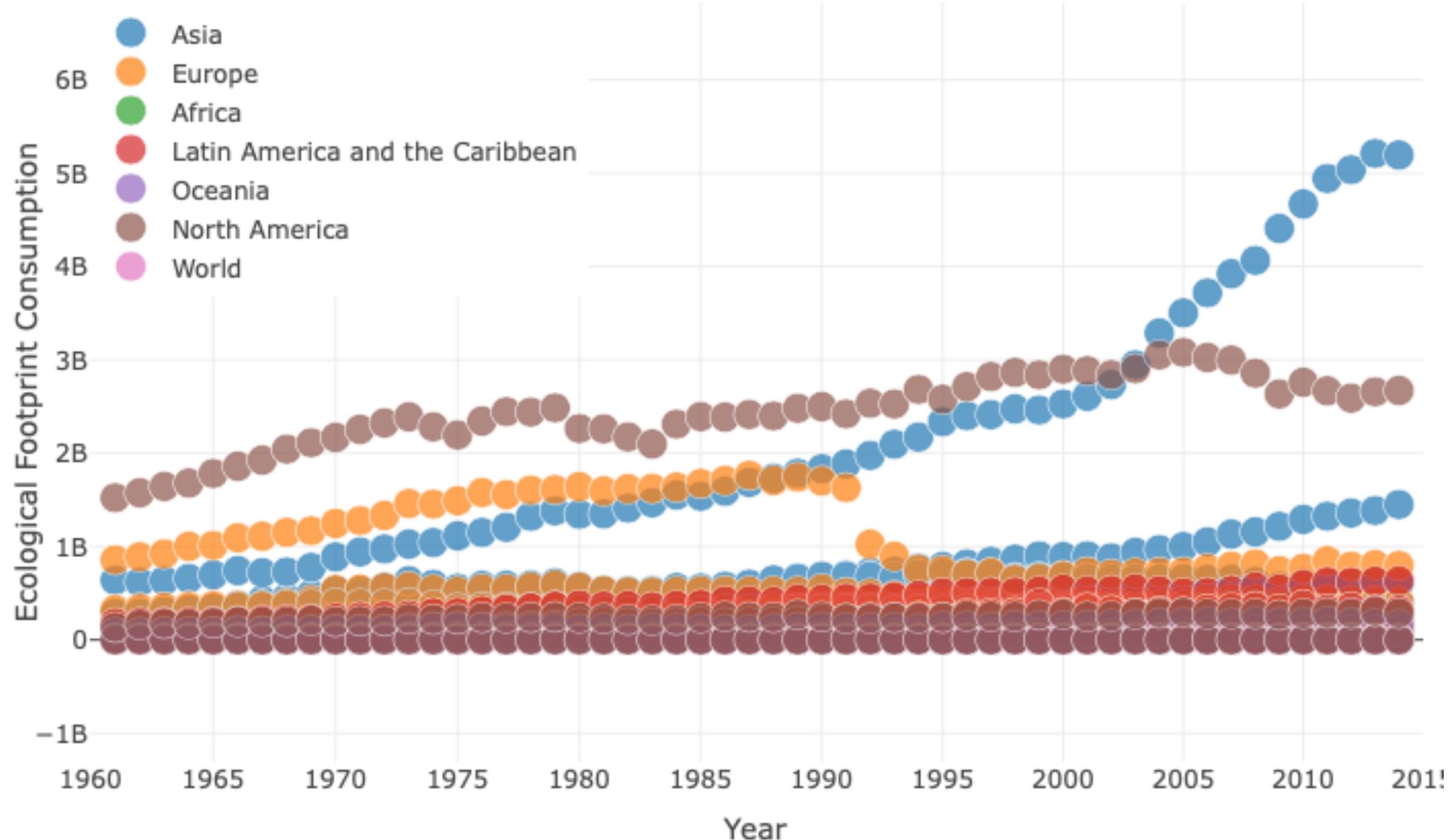
DESCRIPTION OF DATASET

- Global Footprint Network dataset from Kaggle
- 1,305,300 observations
- 15 features: country, ISO alpha-3 code, UN region, UN subregion, year, record, crop land, grazing land, forest land, fishing ground, built up land, carbon, total, Percapita GDP (2010 USD), population

Ecological Footprint Total Consumption from 1960-2014



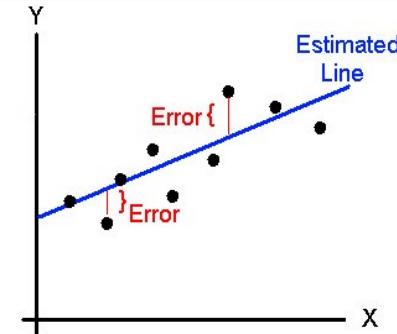
Ecological Footprint Total Consumption from 1960-2014



LINEAR REGRESSION

$$\hat{Y}_i = b_0 + b_1 X_i$$

Estimated (or predicted) Y value for observation i
Estimate of the regression intercept
Estimate of the regression slope
Value of X for observation i



- Target variable (Y): Carbon
- Predictors(X): crop land, grazing land, forest land, fishing ground, built-up land



DATA PREPROCESSING

- Instance Selection: subset observations with 'record' == 'EFConsTotGHA'
 - footprintm = datasetm[datasetm.record == 'EFConsTotGHA']
- Dropped countries with no data before 2014
- Dropped observations from 'World'
- Filled the remaining empty cells with 0 for imputation



DATA PREPROCESSING

- Split data into X and Y
- Train test split 70/30
 - `x_train, x_test, y_train, y_test = train_test_split(X, Y, test_size=.30, random_state=50)`
- Impute missing values with mean
 - `imputer = SimpleImputer(missing_values=0, strategy='mean')`
- Scale Min-Max with Max = 100
 - `scaler = MinMaxScaler(feature_range=(0, 100))`



FITTING TO MODEL

- `model = LinearRegression(normalize=True)`
`model.fit(scaled_x_train, scaled_y_train)`
- `carbon_y_pred = model.predict(scaled_x_test)`

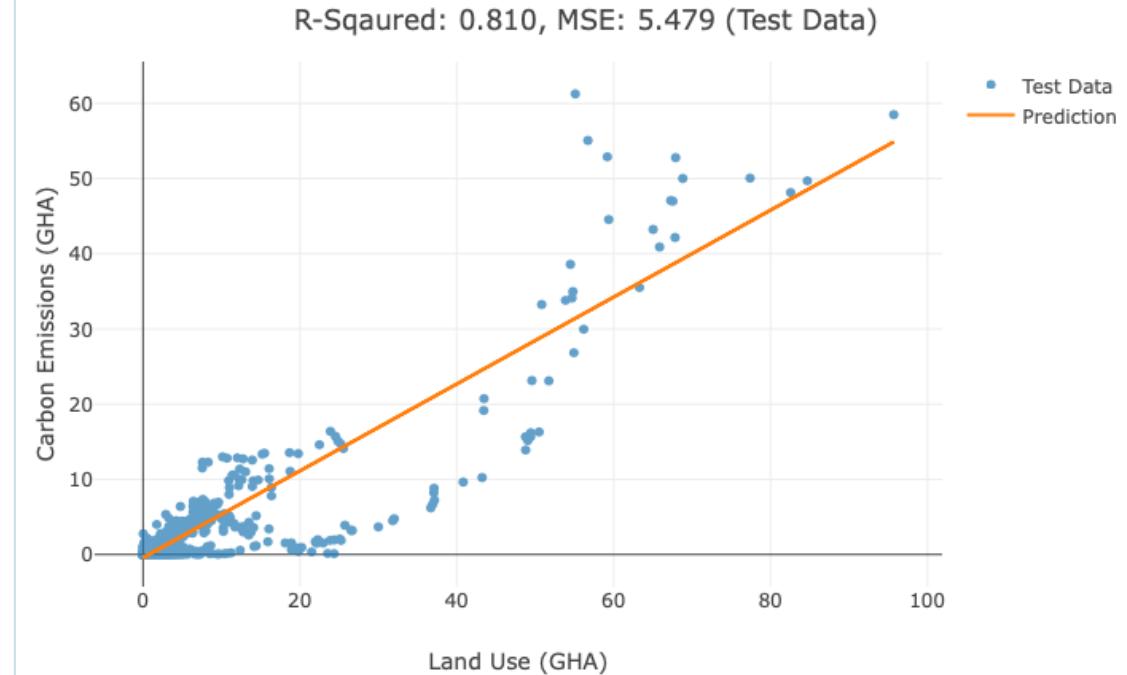
PERFORMANCE OF MODELS

- Predictor: Crop Land
 - R² = 0.59=, error= 11.722
- Predictor: Grazing Land
 - R² = 0.469, error= 15.308
- Predictor: Forest Land
 - R² = 0.81, error= 5.48
- Predictor: Fishing Grounds
 - R² = 0.438, error= 16.191
- Predictor: Built up Land
 - R² = 0.354, error= 18.620

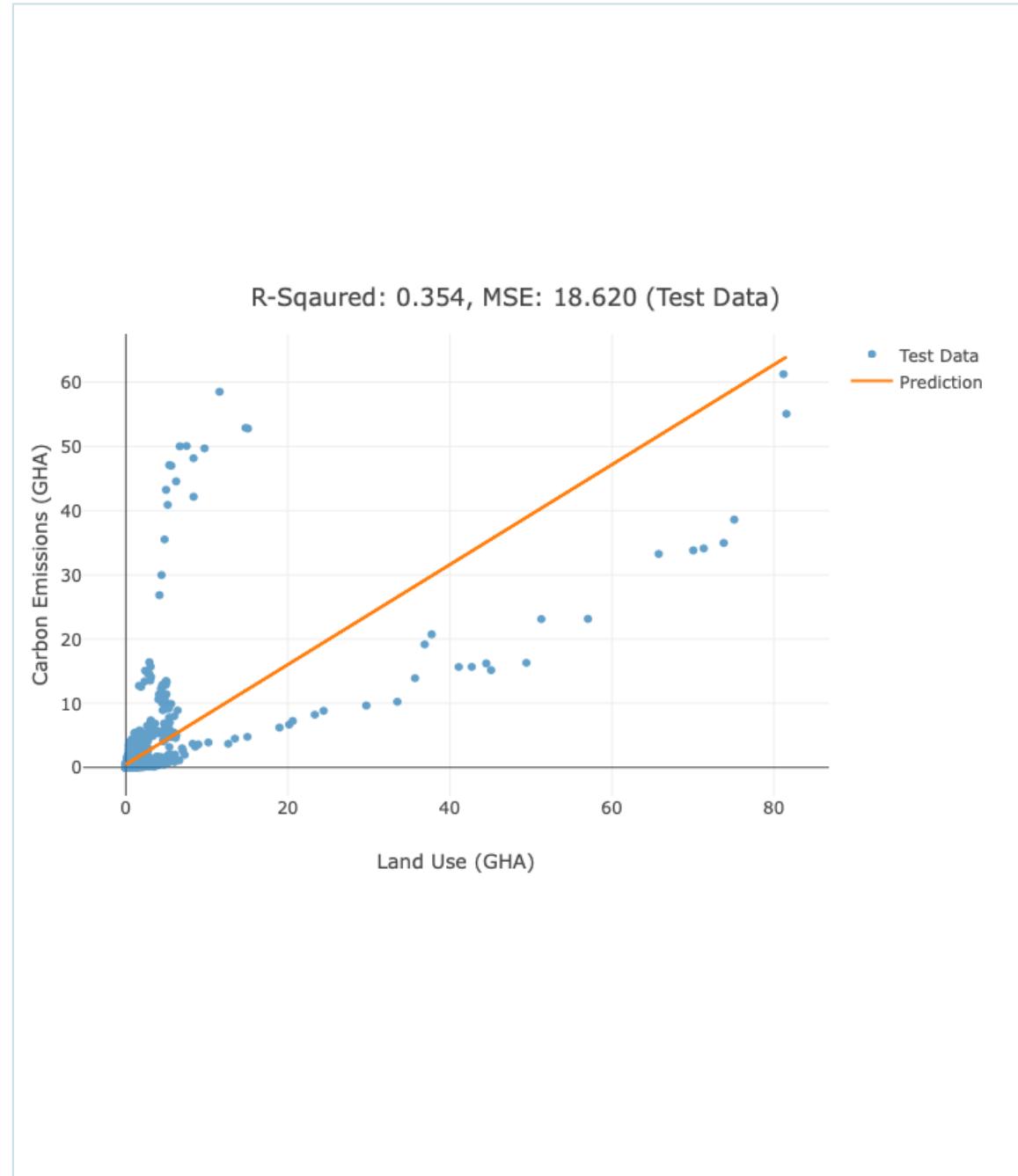
RESULTS

- Best Model : Predictor: **Forest Land**
 - $R^2 = 0.81$
 - error= 5.48
 - $Y = 0.57807228x + (-0.440467)$
- Worst Model: Predictor: **Built Up Land**
 - $R^2 = 0.35427300573754483$
 - error= 18.62
 - $Y = 0.77867656x + 0.451244$

- Predictor: **Forest Land**
 - $R^2 = 0.815$
 - $MSE = 5.48$
 - $Y = 0.57807228x + (-0.440467)$



- Predictor: Built-Up Land
 - R² = 0.354
 - MSE= 18.62
 - Y = 0.7786x +0.45





CONCLUSION

- Global hectares of forest land is the best predictor of amount of carbon emissions from each country
 - Followed by crop land, grazing land, fishing grounds, and built up land in descending order
- Increasing forest land use is increasing the amount of carbon emissions
- Using more plants and trees = less plants and trees to absorb carbon dioxide (greenhouse gases)

REFERENCES

- <https://blog.minitab.com/blog/adventures-in-statistics-2/regression-analysis-how-do-i-interpret-r-squared-and-assess-the-goodness-of-fit>
- <https://climate.nasa.gov/causes/>
- <https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions>
- <https://www.footprintnetwork.org>
- <https://sebastianraschka.com/faq/docs/scale-training-test.html>
- <https://www.statisticallysignificantconsulting.com/RegressionAnalysis.htm>