**table of contents.**

# Introduction

**Introduction:**
Investigate the factors influencing the proportion of energy demand that each nation satisfies through the use of renewable energy.

**Data Selection:**
The dataset spans across diverse geographical locations.

**Predictors:**
Primary energy consumption per capita, % population's access to electricity (or clean fuel), Energy Intensity (The ratio of energy use to GDP ), GDP per capita

**Problem Statement:**
Our goal is to determine which factors drive nations to use renewable energy at higher rates

**Approach:**
Regression, Association Analysis, Gradient Boosting

**Potential Benefits:**
Our findings have the potential to guide policymakers, businesses, and communities towards more effective strategies for promoting renewable energy deployment .

# *Data* Cleaning Process

1. **Data Import & Cleanup:**
- Import CSV into R Studio.
- Remove columns with blank cells.
- Eliminate rows with blank entries.

2. **Subset and Standardization:**
- Define the 'y2019' subset that only includes observations from 2019.
- Rename columns for clarity and conciseness.
- Fix the 'density' column by removing commas to satisfy a numeric form.

# Regression *Analysis*

### Model Construction

Constructing individual models with "RenewShare" regressed over each predictor, one at a time.

### Significance Testing

Predictors with p-values > 0.05 will be removed to retain only significant variables.

### Performance Evaluation

RMSE will be compared to assess model accuracy.

### Confidence Intervals

Establishing confidence intervals to gauge the uncertainty around our model estimates.

### Graphical Analysis

Using graphical analysis and transformations to make our results easier to understand.

# Regression Analysis

```
```{r Q6}
sd(y2019$RenewShare)
```

[1] 27.90359
```

```
```{r Q7}
w<-lm(RenewShare~gdp_per_capita+EnergyIntensity+PrimaryConsumpCap+
PopAccessCleanFuel+PopAccessElec,data=y2019)
summary(w)
```

Call:
lm(formula = RenewShare ~ gdp_per_capita + EnergyIntensity +
    PrimaryConsumpCap + PopAccessCleanFuel + PopAccessElec, data
= y2019)

Residuals:
    Min      1Q  Median      3Q     Max
-42.675 -12.086  -1.973   9.302  61.444

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         7.135e+01  7.667e+00   9.306  < 2e-16 ***
gdp_per_capita      3.894e-04  1.106e-04   3.521 0.000576 ***
EnergyIntensity     2.143e+00  6.621e-01   3.236 0.001500 **
PrimaryConsumpCap  -2.105e-04  6.861e-05  -3.068 0.002572 **
PopAccessCleanFuel -3.433e-01  7.652e-02  -4.486 1.46e-05 ***
PopAccessElec      -2.984e-01  1.069e-01  -2.791 0.005964 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

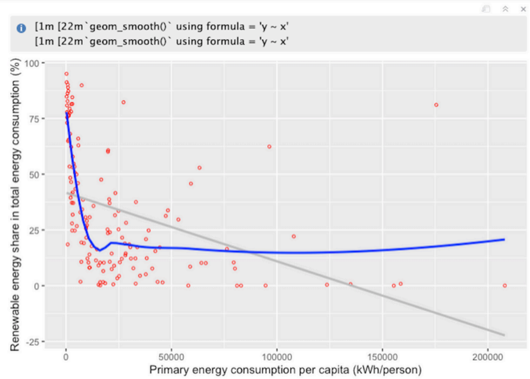
Residual standard error: 17.42 on 145 degrees of freedom
Multiple R-squared:  0.6234,	Adjusted R-squared:  0.6104
F-statistic: 48.01 on 5 and 145 DF,  p-value: < 2.2e-16
```

```
```{r Q9}
confint(c, levels=0.95)
confint(d, levels=0.95)
confint(e, levels=0.95)
confint(g, levels=0.95)
confint(j, levels=0.95)
```

                          2.5 %         97.5 %
(Intercept)        36.5266850365 46.8394764255
PrimaryConsumpCap  -0.0004243646 -0.0001907337
                   2.5 %       97.5 %
(Intercept)   89.9653850 111.4538206
PopAccessElec -0.9185226  -0.6743357
                        2.5 %     97.5 %
(Intercept)        63.5970861 76.0107758
PopAccessCleanFuel -0.6218684 -0.4603304
                   2.5 %   97.5 %
(Intercept)     9.733527 26.86220
EnergyIntensity 1.686824  4.86724
                       2.5 %        97.5 %
(Intercept)     34.4874077693 45.1320475557
gdp_per_capita  -0.0006189993 -0.0002023974
```

# Regression Models – ggplots

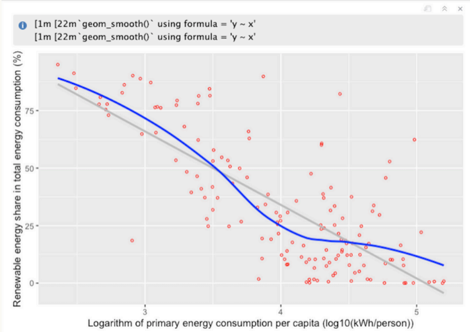
~~~~
```{r Q11}
library(ggplot2)
ggplot(data=y2019, aes(x=PrimaryConsumpCap, y=RenewShare)) +
  xlab("Primary energy consumption per capita (kWh/person)") +
  ylab("Renewable energy share in total energy consumption (%)") +
  geom_point(pch=1, size=1, color="red") +
  geom_smooth(method="lm", color="gray", se=FALSE) +
  geom_smooth(method="loess", color="blue", se=FALSE)
```
~~~~




Non-ellipsoidal so can't use Pearson
Nonmonotonic so can't use Spearman

~~~~
```{r Q13}
y2019outPCC <- y2019[y2019$PrimaryConsumpCap <= 175000, ]

library(ggplot2)
ggplot(data=y2019outPCC, aes(x=log10(PrimaryConsumpCap),
y=RenewShare)) +
  xlab("Logarithm of primary energy consumption per capita
(log10(kWh/person))") +
  ylab("Renewable energy share in total energy consumption (%)") +
  geom_point(pch=1, size=1, color="red") +
  geom_smooth(method="lm", color="gray", se=FALSE) +
  geom_smooth(method="loess", color="blue", se=FALSE)
```
~~~~

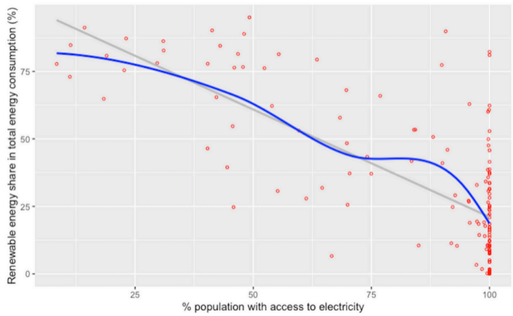



Pearson & Spearman OK

# Regression Models – ggplots

````
```{r Q14}
library(ggplot2)
ggplot(data=y2019, aes(x=PopAccessElec, y=RenewShare)) +
  xlab("% population with access to electricity") +
  ylab("Renewable energy share in total energy consumption (%)") +
  geom_point(pch=1, size=1, color="red") +
  geom_smooth(method="lm", color="gray", se=FALSE) +
  geom_smooth(method="loess", color="blue", se=FALSE)
```
````

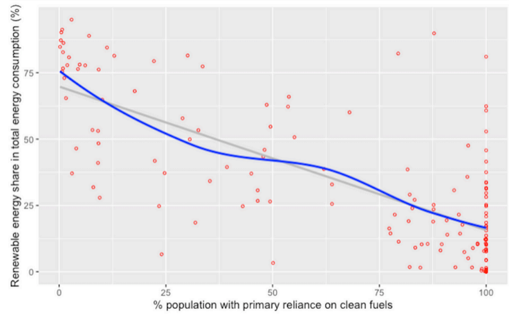
[1m [22m`geom_smooth()` using formula = 'y ~ x'
[1m [22m`geom_smooth()` using formula = 'y ~ x'




Non-ellipsoidal so can't use Pearson
Nominally monotonic so CAN use Spearman

````
```{r Q15}
library(ggplot2)
ggplot(data=y2019, aes(x=PopAccessCleanFuel, y=RenewShare)) +
  xlab("% population with primary reliance on clean fuels") +
  ylab("Renewable energy share in total energy consumption (%)") +
  geom_point(pch=1, size=1, color="red") +
  geom_smooth(method="lm", color="gray", se=FALSE) +
  geom_smooth(method="loess", color="blue", se=FALSE)
```
````

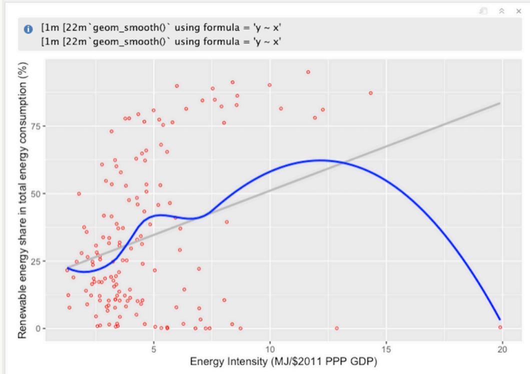
[1m [22m`geom_smooth()` using formula = 'y ~ x'
[1m [22m`geom_smooth()` using formula = 'y ~ x'




Non-ellipsoidal so can't use Pearson
Nominally monotonic so CAN use Spearman

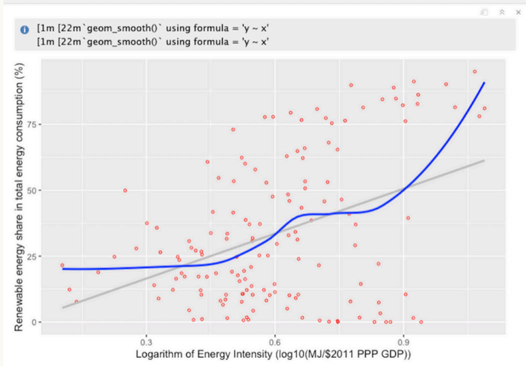
# Regression Models – ggplots

````
```{r Q16}
library(ggplot2)
ggplot(data=y2019, aes(x=EnergyIntensity, y=RenewShare)) +
  xlab("Energy Intensity (MJ/$2011 PPP GDP)") +
  ylab("Renewable energy share in total energy consumption (%)") +
  geom_point(pch=1, size=1, color="red") +
  geom_smooth(method="lm", color="gray", se=FALSE) +
  geom_smooth(method="loess", color="blue", se=FALSE)
```
````




Non-ellipsoidal so can't use Pearson
Nonmonotonic so can't use Spearman

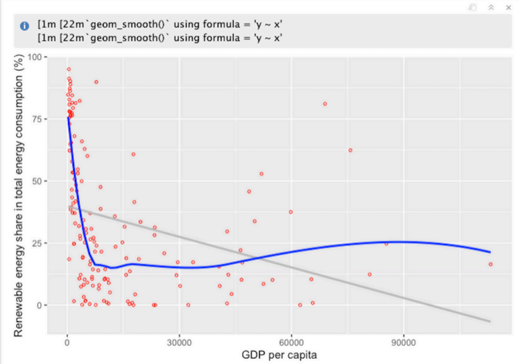
````
```{r Q18}
y2019outEI <- y2019[y2019$EnergyIntensity <= 12.5, ]
library(ggplot2)
ggplot(data=y2019outEI, aes(x=log10(EnergyIntensity), y=RenewShare)) +
  xlab("Logarithm of Energy Intensity (log10(MJ/$2011 PPP GDP))") +
  ylab("Renewable energy share in total energy consumption (%)") +
  geom_point(pch=1, size=1, color="red") +
  geom_smooth(method="lm", color="gray", se=FALSE) +
  geom_smooth(method="loess", color="blue", se=FALSE)
```
````




Pearson & Spearman OK

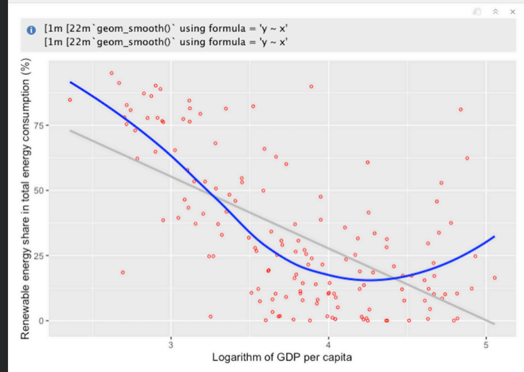
# Regression Models – ggplots

```{r Q19}
library(ggplot2)
ggplot(data=y2019, aes(x=gdp_per_capita, y=RenewShare)) +
  xlab("GDP per capita") +
  ylab("Renewable energy share in total energy consumption (%)") +
  geom_point(pch=1, size=1, color="red") +
  geom_smooth(method="lm", color="gray", se=FALSE) +
  geom_smooth(method="loess", color="blue", se=FALSE)
```




Non-ellipsoidal so can't use Pearson
Nonmonotonic so can't use Spearman

```{r Q20}
library(ggplot2)
ggplot(data=y2019, aes(x=log10(gdp_per_capita), y=RenewShare)) +
  xlab("Logarithm of GDP per capita") +
  ylab("Renewable energy share in total energy consumption (%)") +
  geom_point(pch=1, size=1, color="red") +
  geom_smooth(method="lm", color="gray", se=FALSE) +
  geom_smooth(method="loess", color="blue", se=FALSE)
```




Pearson OK

# Association *Analysis*

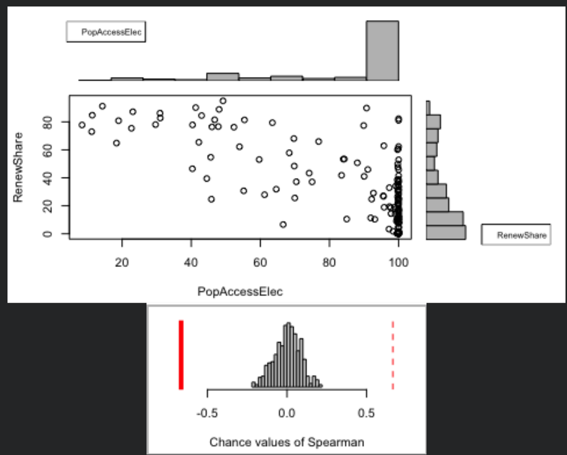
**Correlation Analysis :**
Assessing Pearson and/or Spearman correlations for each relationship based on graphical analysis.

**Significance Testing :**
Correlations for each predictor had p-value ranges of 0 - 0.007 < 0.05

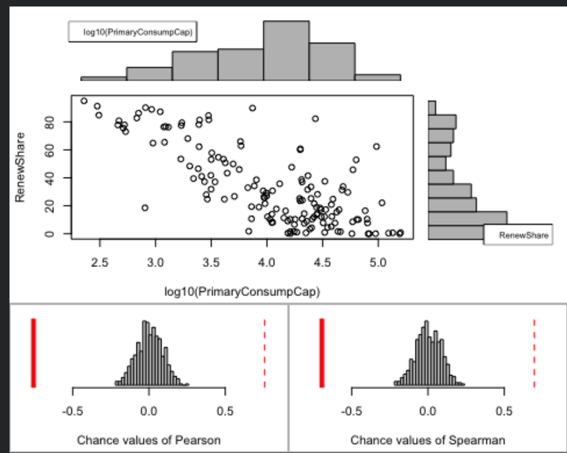
**Chance Values Plots :**
SSE reductions far beyond what happens "by chance"

**Strength Assessment :**
R-squared if Pearson was appropriate, otherwise, defer to graphical analysis

# Association Analysis

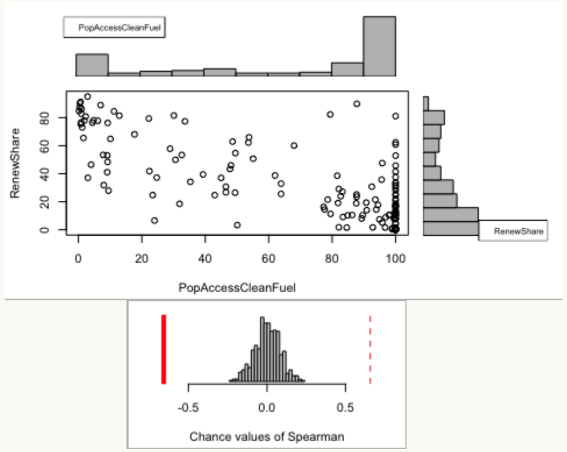



RenewShare~PopAccessElec

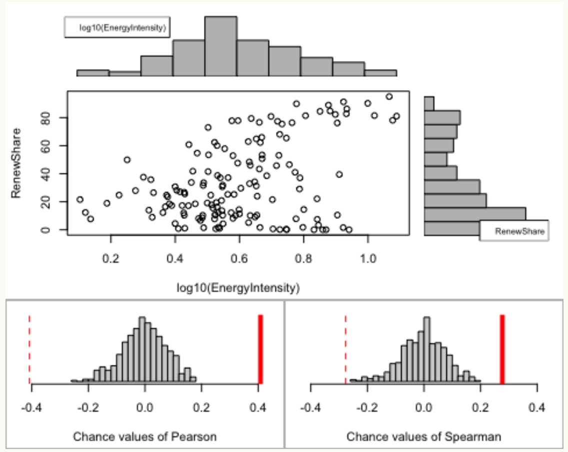



RenewShare~
log10(PrimaryConsumpCap)

# Association Analysis

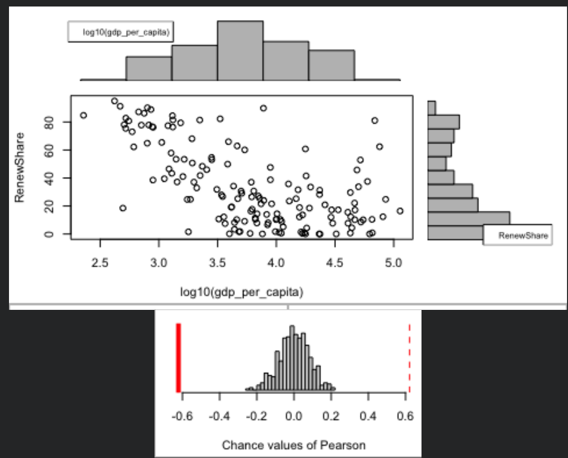



RenewShare~PopAccessCleanFuel




RenewShare ~
log10(EnergyIntensity)

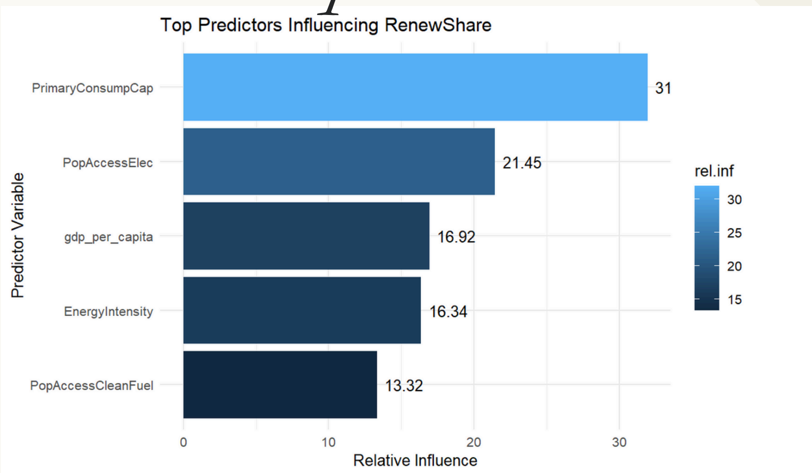
# Association Analysis




RenewShare~log10(GDP_Per_Capita)

# Other Technique: Gradient Boosting Machines (GBM)

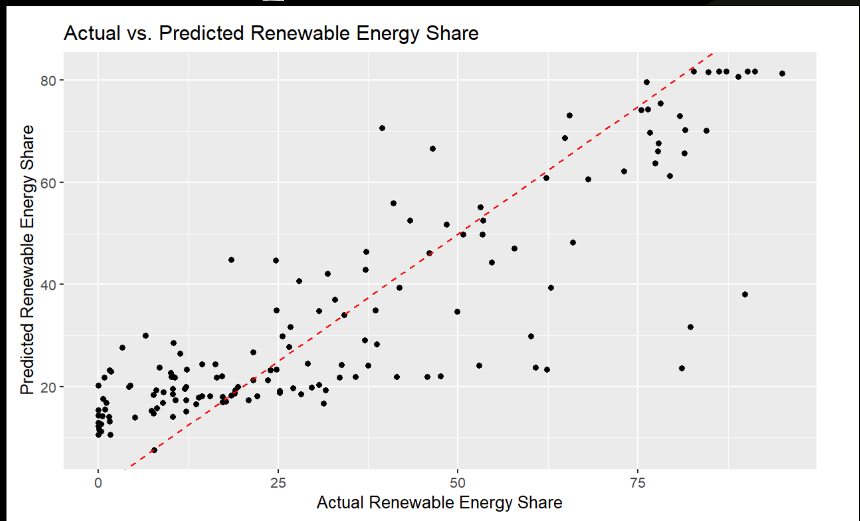
## *Predictor Importance*




- Powerful machine learning algorithm
- Aims to identify the most influential predictors for our Response Variable
- By evaluating how often each predictor is used in splitting data and if it significantly decreases prediction error when used for splitting.

# Gradient Boosting Machines (GBM)

## *Model Importance*




- **Capturing the variation in RenewShare:** It means how accurately the model predicts changes in the response variable (RenewShare) using the top predictor variables.

- **R-squared:** 0.7256 model explains 72.56% of the variance in Renewable Energy Share (RenewShare) -which further means that the model can reliably predict changes in RenewShare based on the selected predictors.

# Conclusion

GBM: Primary energy consumption per capita (kWh/person): Regions with high energy consumption per capita often have the infrastructure, resources, and policies to support renewable energy, leading to a greater share of its usage.