Modern renewable energy consumption in R

INSTRUCTOR: HAMID RAJAEE

By: SARA KHOSRAVI

March 2021

Overview:

The dataset is taken from Kaggle site.

In this project, a dataset include 5095 observations and 7 variables, The dataset is named “Modern renewable energy consumption”.

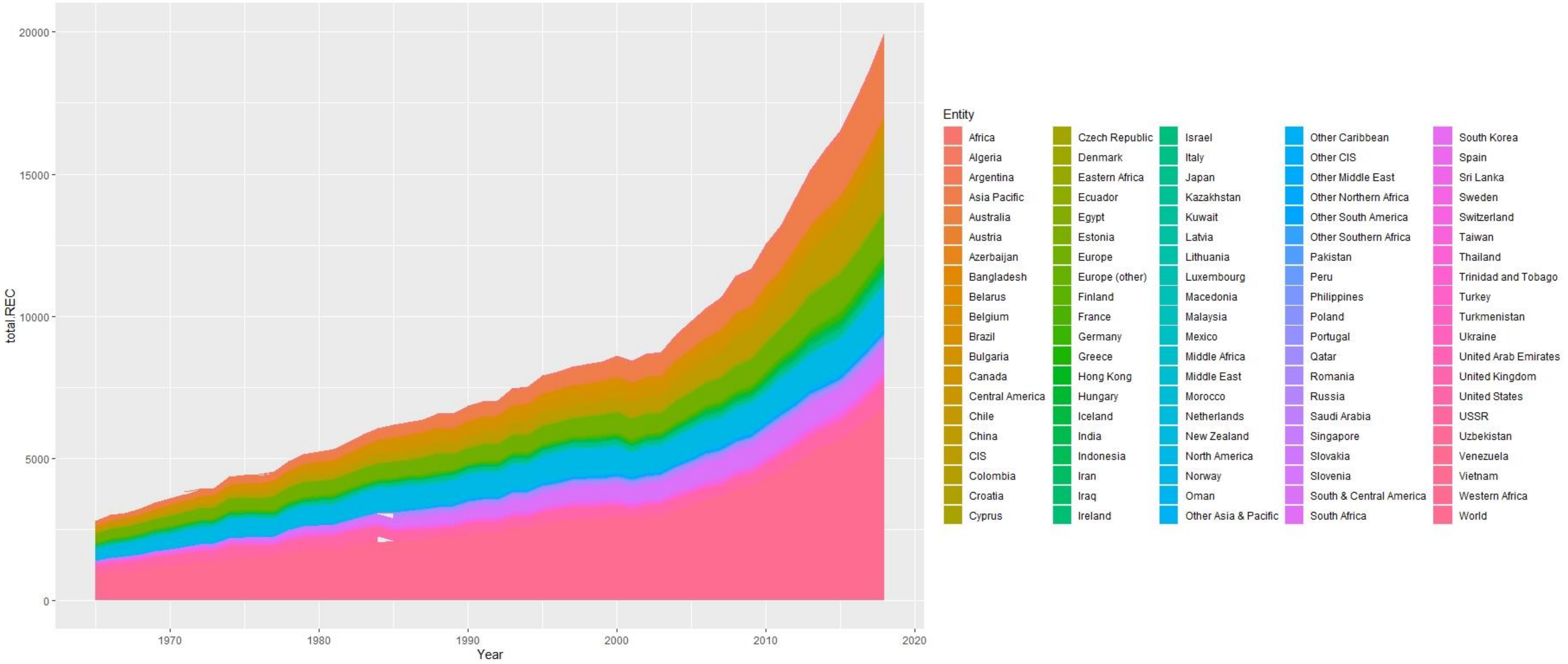
EDA DATA:

Business Understanding

- In this project we looked at, what share renewable technologies collectively accounted for in the energy mix.
- Globally we see that hydropower is by far the largest modern renewable source *[since traditional biomass is not included here]*. But we also see wind and solar power are both growing rapidly.
- The dataset have 7 columns. For understanding the dataset, Analysis and compare the dataset, 3 main columns by calculation have been added the dataset.
- The dataset is taken from Kaggle site.

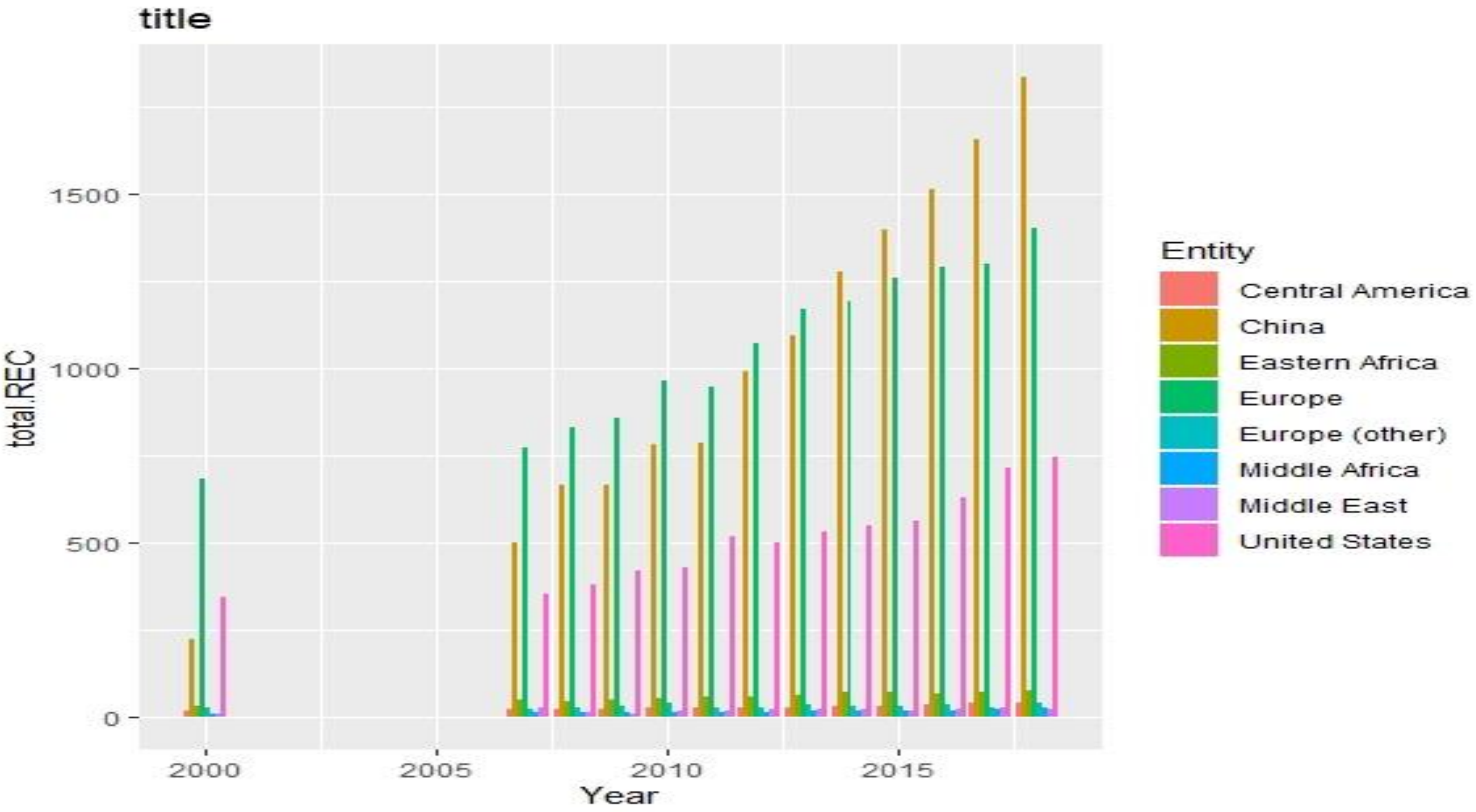
Data Understanding :

The chart shows this as a stacked area chart, which allows us to more readily see the breakdown of the renewable mix, and relative contribution of each.



Data Understanding:

- This graph shows that 8 top Renewable Energy Consumer(REC) in the world. To obtain this diagram, ***Filter, Subset*** and ***Full Joint*** commands have been used.



Data Understanding:

- ❖ Getting familiar with data for Data Understanding in EDA.
  Data frame has a 5059 observation and 7 columns. The missing value can be seen in the dataset. The important column is Entity, Year , Hydropower, Solar and Wind, So by using slice the column of Code dropped at the dataset.

```
> typeof(REC)
[1] "list"
> # Compactly Display the Structure of an Arbitrary R Object
> str(REC)
'data.frame':   5095 obs. of  7 variables:
 $ Entity         : chr  "Africa" "Africa" "Africa" "Africa" ...
 $ Code           : chr  NA NA NA NA ...
 $ Year           : int  1965 1966 1967 1968 1969 1970 1971 1971 1971 1971 ...
 $ Hydropower     : num  14.3 15.6 16.2 18.6 21.6 ...
 $ Solar          : num  0 0 0 0 0 0 0 0 0 0 ...
 $ Wind           : num  0 0 0 0 0 0 0 0 0 0 ...
 $ OtherRenewables: num  0 0 0 0 0 0 0.164 0.164 0.164 0.164 ...
>
```

Feature Engineering:

❖ For preparation and analysis, the dataset **3 Continues COLUMNS** and **one Categorical Column** are ***added*** to ***dataset*** to make it easy to handle the project.

1. "total.REC“: Total the consumption of Hydropower, Solar, Wind and Other Renewable Energy

```
REC$total.REC <- NA
REC$total.REC <- rowSums(REC[ ,c(3:6)], na.rm=TRUE)
```

2. ” cum_total “: cumulative REC consumption

```
NEWREC$cum_total <- cumsum(NEWREC$total.REC)
```

3. "Growth.rate“: Growth rate per annul

```
RECF <- NEWREC %>% group_by(Entity) %>% mutate(Growth.rate = (total.REC-lag(total.REC))/lag(total.REC))
```

4. “GROUPEntity”:

```
NEWREC$GROUPEntity <- NEWREC %>% group_by(Entity)
```

Data preparation:

Data preparation or Data cleaning is:

1) Handling duplicate data

2) Handling Missing Values

3) Handling outliers

❖ By using frequency in a dataset is observed

That data duplication exists in Africa.

This problem is solved by using the

Duplicated command.

sum(duplicated(REC))

RowDuplicate <- which(duplicated(REC))

REC <- REC[-RowDuplicate,]

❖ For handling Missing value in project is

converted missing value to NA and after that use some command in R to handle that.

REC[REC=='']<-NA # converting Null to Na

sum(is.na(REC))   # 11268 number of missing values

colSums(is.na(REC))

❖ This project has outlier but this outlier it

is important for analysis of data. Because this

Outlier happened due to the rapid scientific progress

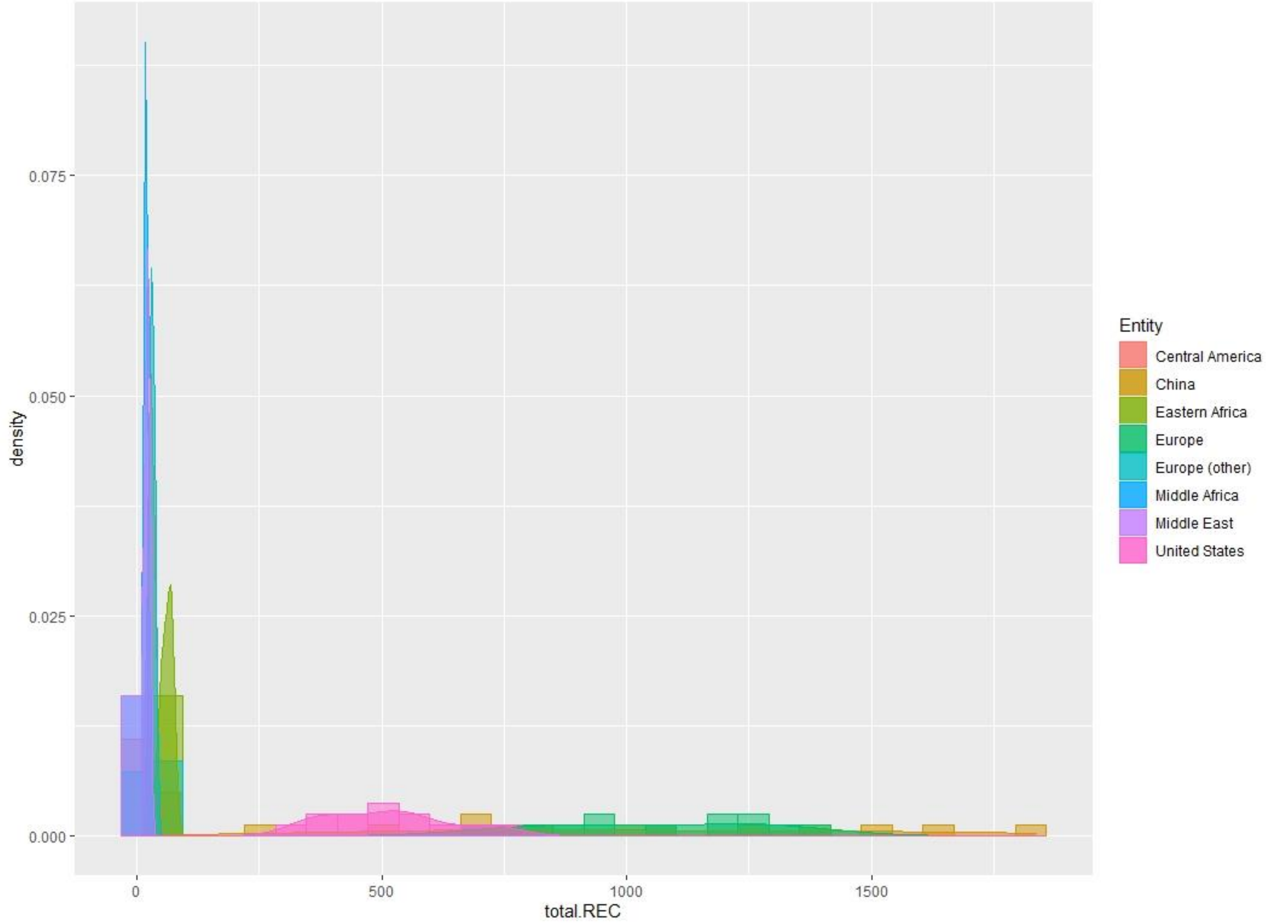
In this field recently.

```
> # since target is categorical variable, in uni-variate Analysis for summarizing I
> # will find frequency and for visualization I plot: pie chart or bar-chart
> tbl<-table(REC$ Entity)
> tbl

               Africa               Algeria             Argentina          Asia Pacific 
                   58                    54                    54                    54 
            Australia               Austria            Azerbaijan            Bangladesh 
                   54                    54                    34                    54 
              Belarus               Belgium                Brazil              Bulgaria 
                   34                    54                    54                    54 
               Canada       Central America                 Chile                 China 
                   54                    54                    54                    54 
                  CIS              Colombia               Croatia                Cyprus 
                   54                    54                    29                    54 
       Czech Republic               Denmark        Eastern Africa               Ecuador 
                   54                    54                    54                    54 
                Egypt               Estonia                Europe       Europe (other) 
                   54                    34                    54                    54 
              Finland                France               Germany                Greece 
                   54                    54                    54                    54 
            Hong Kong               Hungary               Iceland                 India 
                   54                    54                    54                    54 
            Indonesia                  Iran                  Iraq               Ireland 
                   54                    54                    54                    54 
               Israel                 Italy                 Japan            Kazakhstan 
                   54                    54                    54                    34 
               Kuwait                Latvia             Lithuania            Luxembourg 
                   54                    34                    34                    54 
            Macedonia              Malaysia                Mexico         Middle Africa 
                   29                    54                    54                    54 
          Middle East               Morocco           Netherlands           New Zealand 
                   54                    54                    54                    54 
        North America                Norway                  Oman  Other Asia & Pacific 
                   54                    54                    54                    54 
      Other Caribbean             Other CIS    Other Middle East Other Northern Africa 
                   54                    34                    54                    54 
  Other South America Other Southern Africa              Pakistan                  Peru 
                   54                    54                    54                    54 
          Philippines                Poland              Portugal                 Qatar 
                   54                    54                    54                    54 
              Romania                Russia          Saudi Arabia             Singapore 
                   54                    34                    54                    54 
             Slovakia              Slovenia South & Central America         South Africa 
                   54                    29                    54                    54 
          South Korea                 Spain             Sri Lanka                Sweden 
                   54                    54                    54                    54 
          Switzerland                Taiwan              Thailand   Trinidad and Tobago 
                   54                    54                    54                    54 
               Turkey          Turkmenistan               Ukraine  United Arab Emirates 
                   54                    34                    54                    54 
       United Kingdom         United States                  USSR            Uzbekistan 
                   54                    54                    20                    34 
            Venezuela               Vietnam        Western Africa                 World 
                   54                    54                    54                    54 
```

Univariate analysis for Numerical variables :

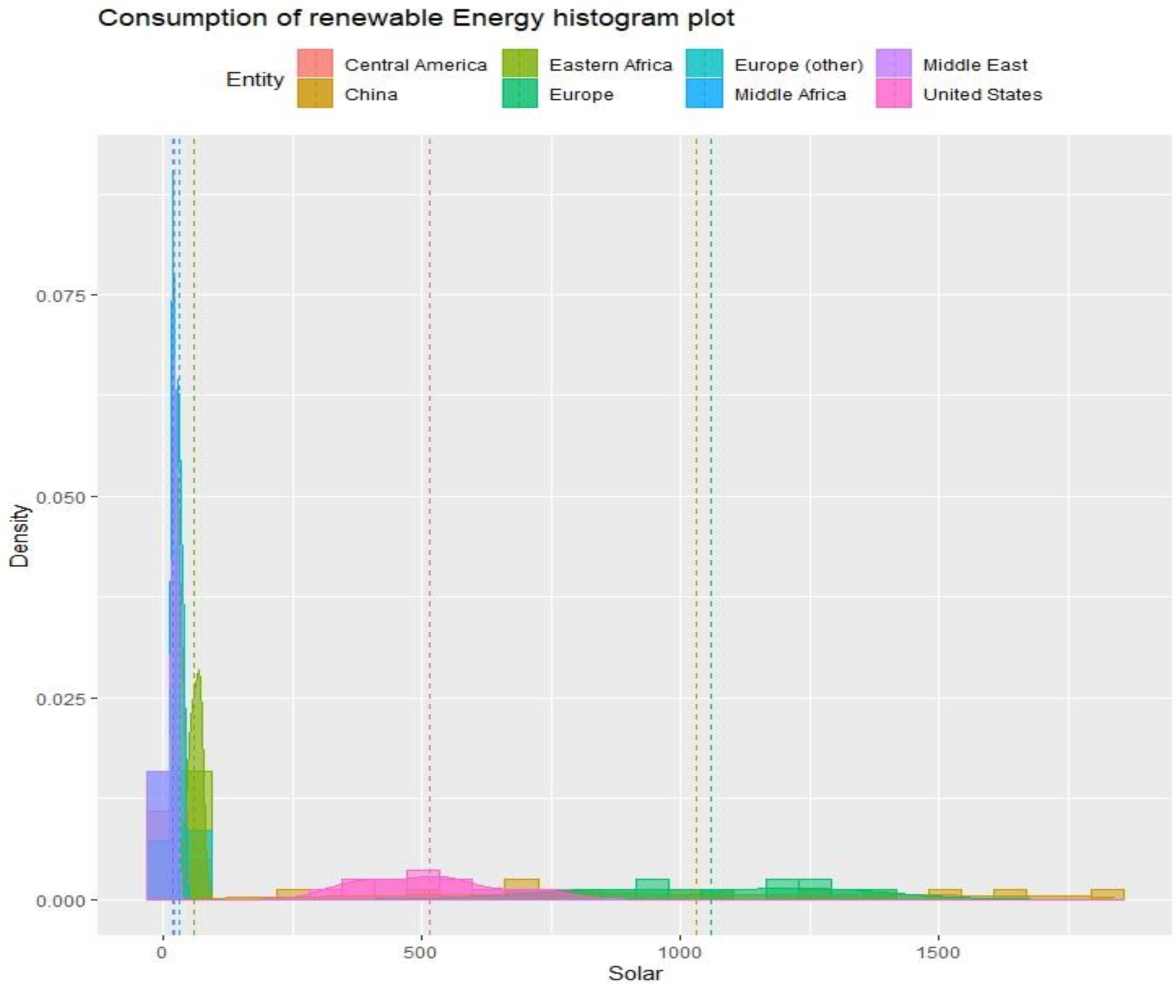
The distribution of “total.REC” shows us , we have mutated recently. Consumption of renewable of energy in the last 10 years. And this graph shows the jump in new energy consumption in recent years.



Univariate analysis for Numerical variables :

For **Visualization** this **Numerical variable** (Total of Renewable Energy) **plot density** is chosen.

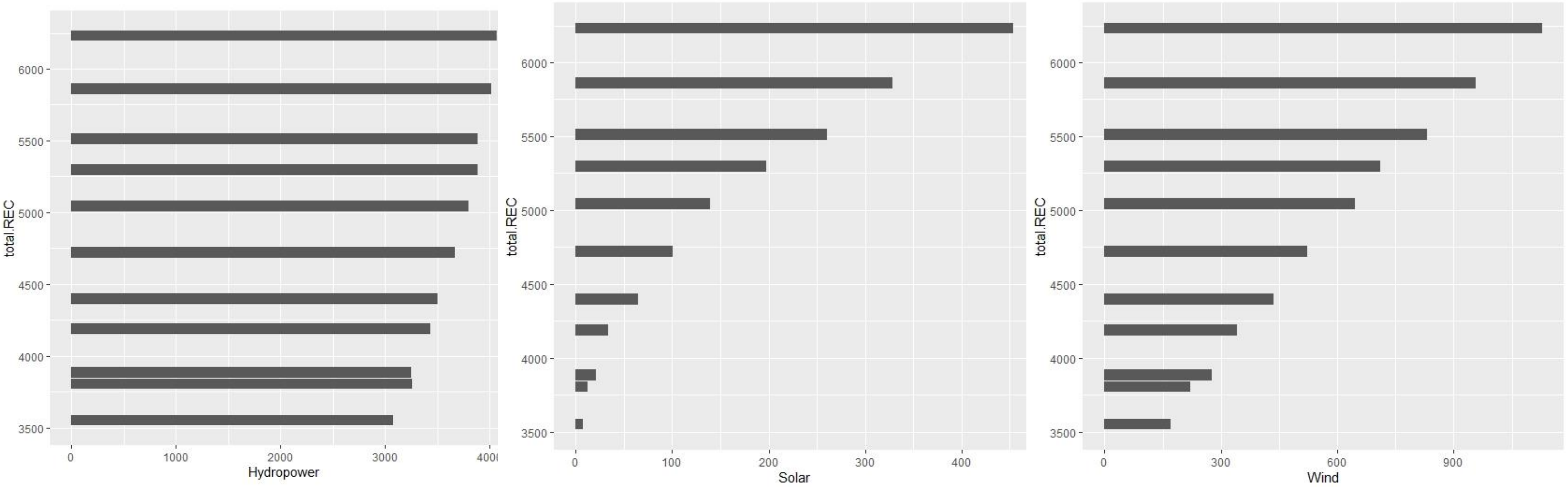
This graph shows , the Solar Energy versus density.



Bi-variate Analysis for Continuous Vs. Continuous :

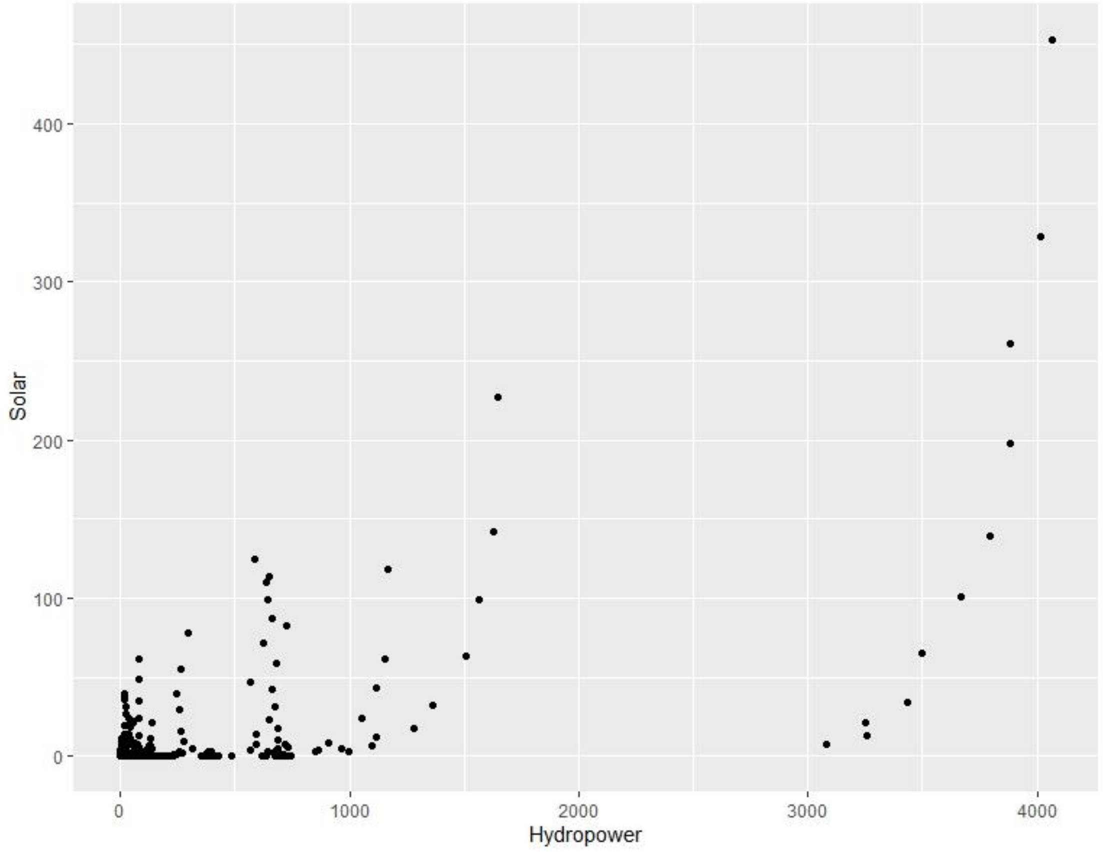
The amount of Consumption Hydropower, Wind and Solar energy of the total of energy.

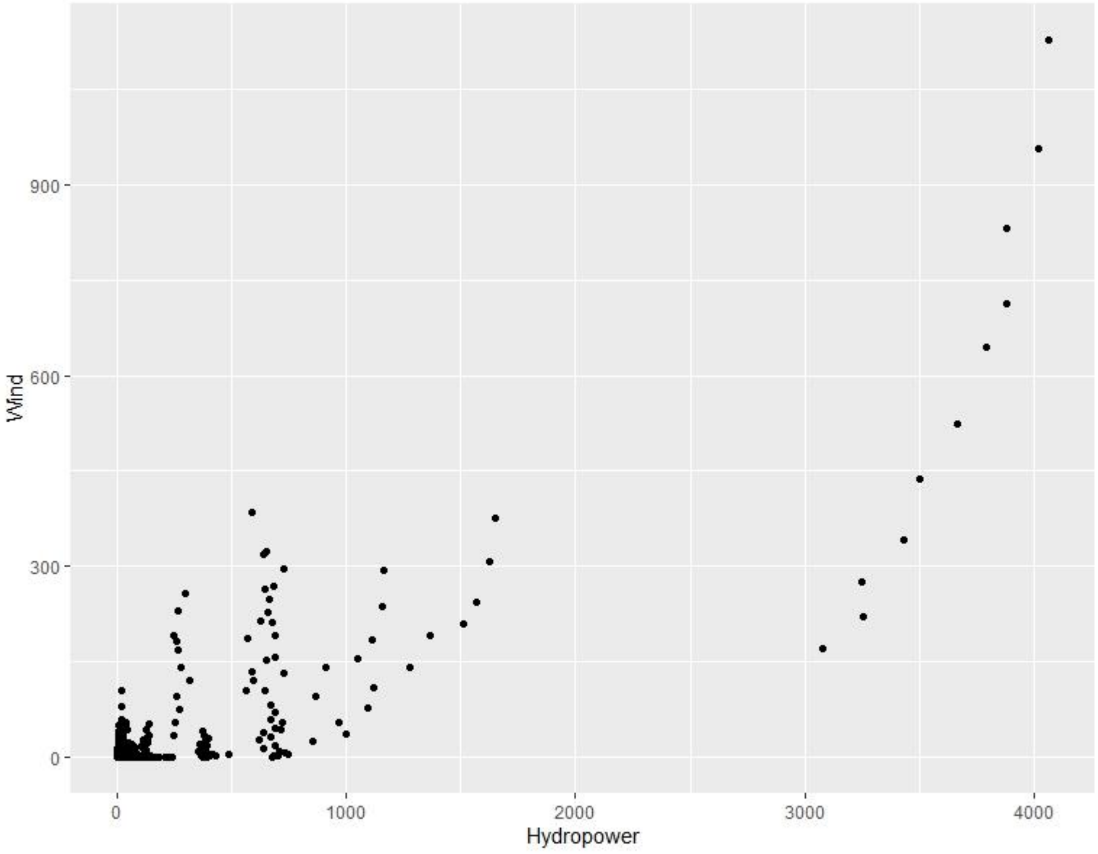
In charts shown here we look at the breakdown of renewable technologies by their individual components – hydropower, solar, wind, and others.



Bi-variate Analysis for Continuous Vs. Continuous:

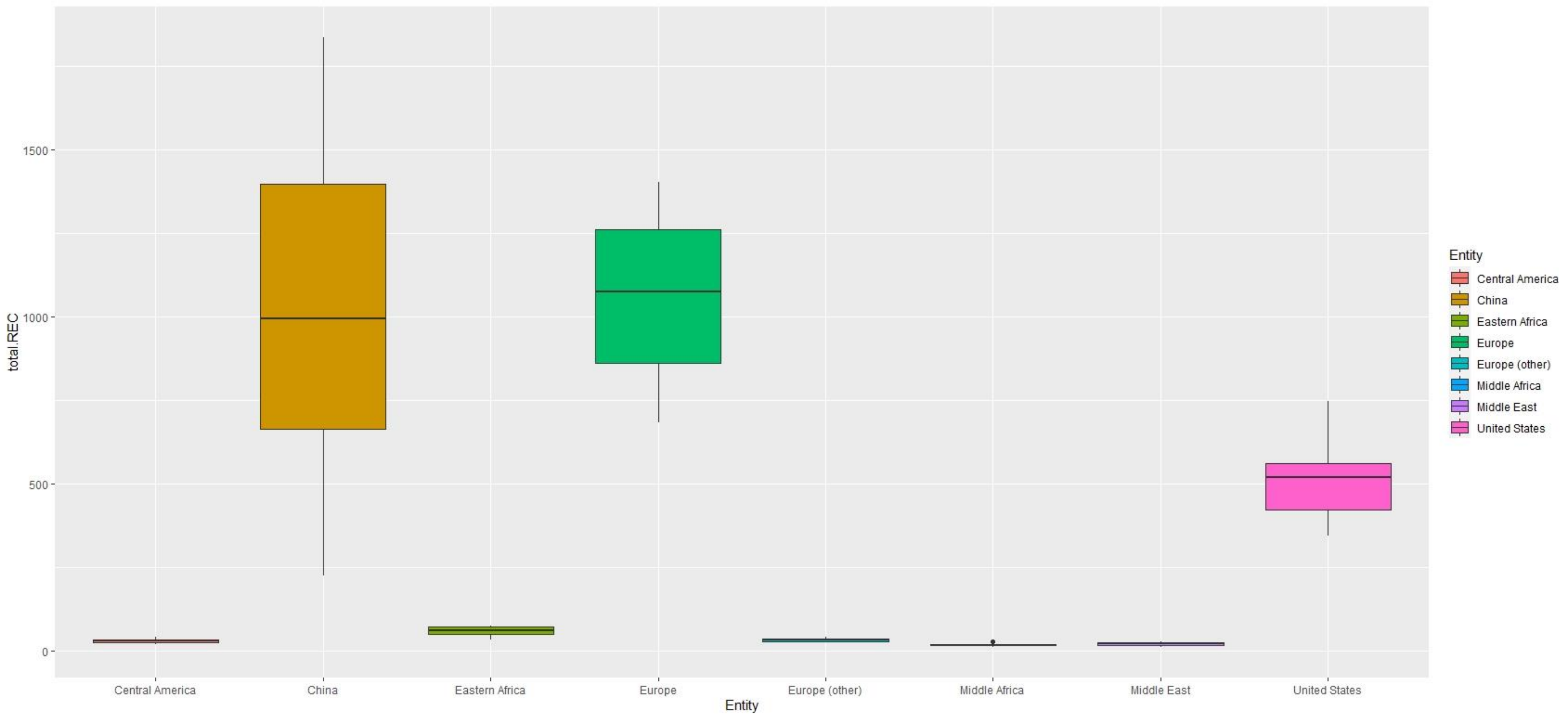
Using ***scatter plot*** for showing the relationship between solar versus Hydropower. Also, relationship between Hydropower versus Solar .





Bi-variate Analysis for Continuous Vs. Categorical:

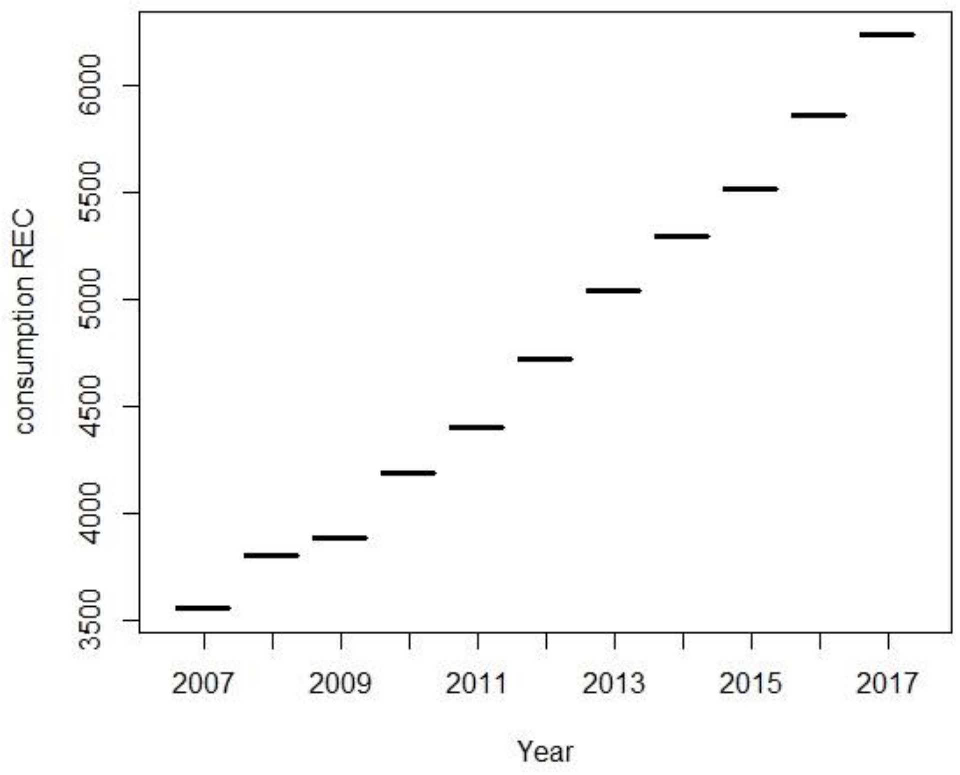
visualization: Grouped **box plot**
Bi-variate Analysis for continuous(total.REC) Vs. categorical (Entity)



Bi-variate Analysis for Continuous Vs. Categorical:

Consumption Renewable Energy during the 2007-2017 in the word. Target **Year as a categorical** variable in this project.

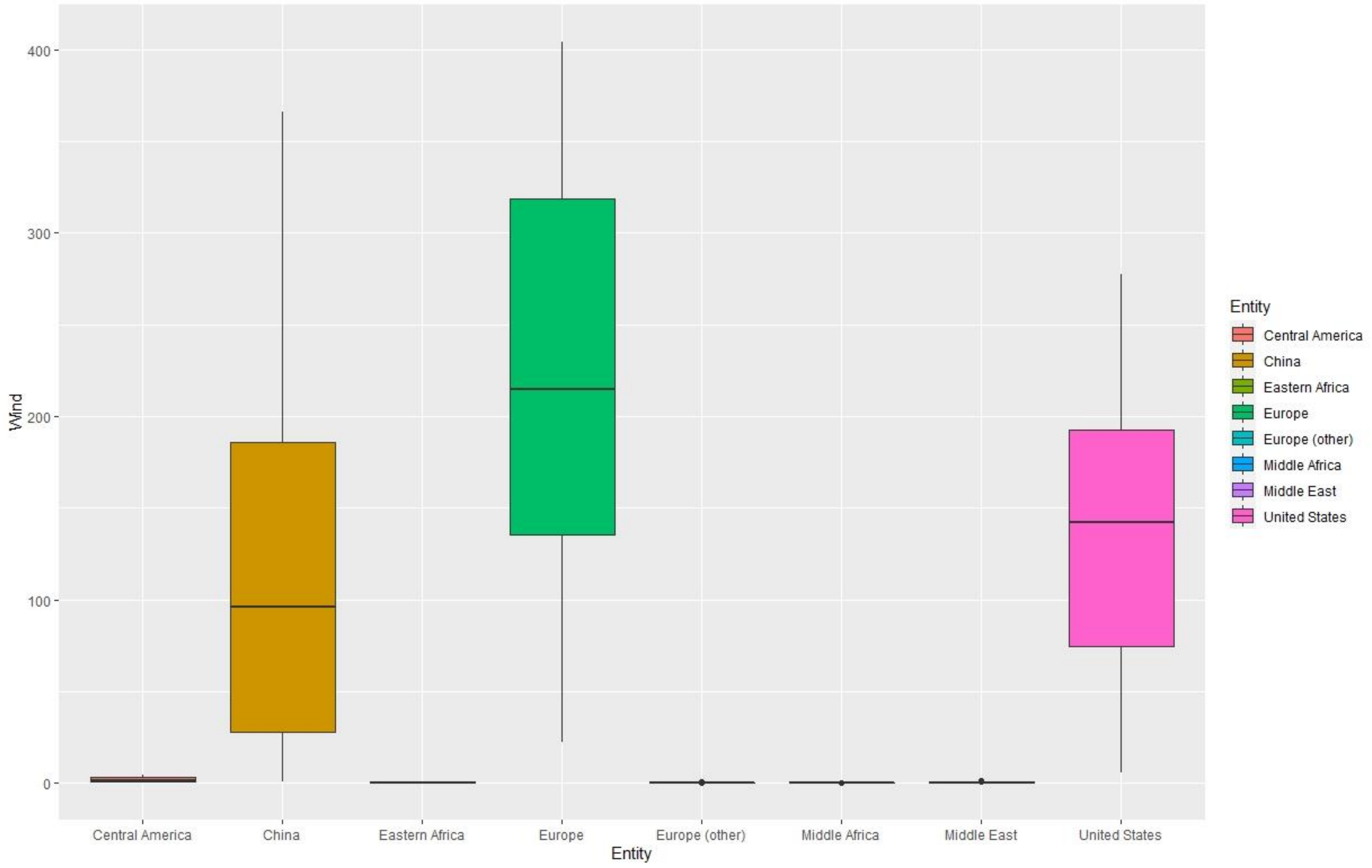
Treating year as a **categorical variable** will calculate effect of each individual **year** - i.e., what impact on the target **variable** was in average each year. On the other hand, including t as **numerical variable** says what happens on average two **years** later.



Bi-variate Analysis for Continuous Vs. Categorical:

visualization: Grouped **box plot**

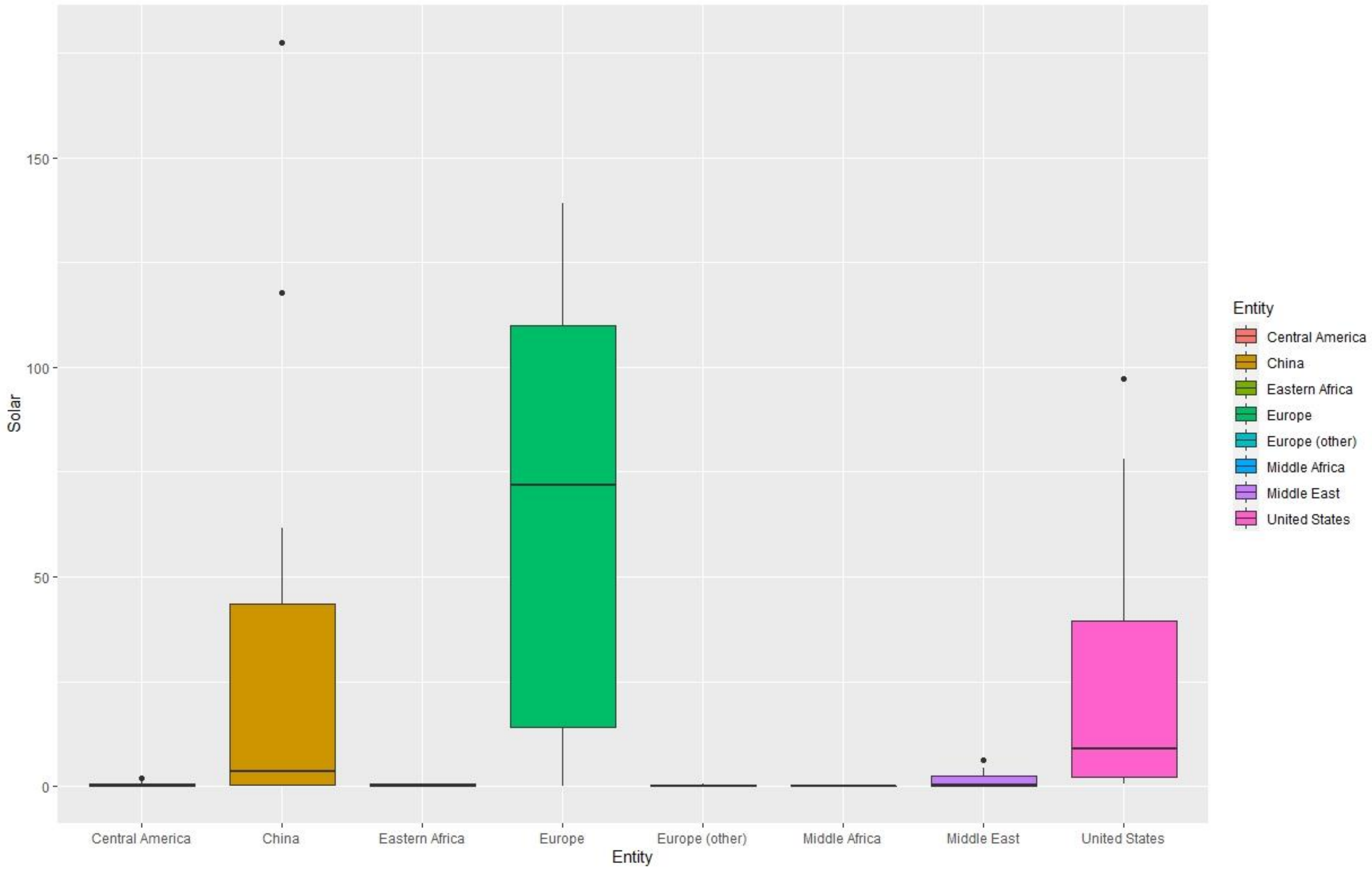
Bi-variate Analysis for continuous (Wind) Vs. categorical (Entity)



Bi-variate Analysis for Continuous Vs. Categorical:

visualization: Grouped **box plot**

Bi-variate Analysis for continuous (Wind) Vs. categorical (Entity)



Bi-variate Analysis for Continuous Vs. Categorical:

visualization: Grouped **box plot**

By using aggregation function compare the numerical variable is comfortable.

```
agg2 <- cbind(aggregate( total.REC ~ Entity , REC , min),
              aggregate( total.REC ~ Entity , REC  , max)[,2],
              aggregate( total.REC ~ Entity , REC  , mean)[,2])

names(agg2) <- c("total.REC","min_REC","max_REC","mean_REC")
agg2

write.table(agg2, file = "agg2.csv",
            sep = "\t", row.names = F)
```

Bi-variate Analysis for Continuous Vs. Categorical :

Test of independence: Anova

Perform the ANOVA test:

- One-way ANOVA

In the one-way ANOVA example, we are modeling crop total.REC as a function of the type of Entity used. First, we will use aov() to run the model, then we will use summary() to print the summary of the model.

```
one.way <- aov(total.REC~Entity, data = REC.ORGIN)
summary(one.way)
```

Bi-variate Analysis for Continuous Vs. Categorical :

Test of independence: Anova

- Two-way ANOVA

In the two-way ANOVA example, we are modeling crop total.REC as a function of type of Entity and Year. First, we use aov() to run the model, then we use summary() to print the summary of the model.

```
two.way <- aov(total.REC~Entity + Year, data = REC.ORGIN)
summary(two.way)
```

- Adding interactions between variables

Sometimes you have reason to think that two of your independent variables have an interaction effect rather than an additive effect.

```
interaction <- aov(total.REC~Entity * Year, data =  REC.ORGIN)
summary(interaction
```

Bi-variate Analysis for Continuous Vs. Categorical :

Test of independence: Anova

❖ Adding a Solaring variable

If you have grouped your experimental treatments in some way, or if you have a confounding variable that might affect the relationship you are interested in testing, you should include that element in the model as a Solaring variable. The simplest way to do this is just to add the variable into the

model with a '+'.

```
Solaring <- aov(total.REC~Entity + Year + Solar, data =  REC.ORGIN)
summary(Solaring)
```

❖ Find the best-fit model:

There are now four different ANOVA models to explain the data. How do you decide which one to use? Usually, you will want to use the 'best-fit' model -

the model that best explains the variation in the dependent variable.

Bi-variate Analysis for Continuous Vs. Categorical :

Test of independence: Anova

```
install.packages("AICcmodavg")
library("AICcmodavg")

model.set <- list(one.way, two.way, interaction, Solaring)
model.names <- c("one.way", "two.way", "interaction", "Solaring")

aictab(model.set, modnames = model.names)
```

❖ Check for homoscedasticity

To check whether the model fits the assumption of homoscedasticity, look at the model diagnostic plots in R using the plot() function:

```
par(mfrow=c(2,2))
plot(two.way)
par(mfrow=c(1,1))
```

Bi-variate Analysis for Continuous Vs. Categorical :

Test of independence: Anova

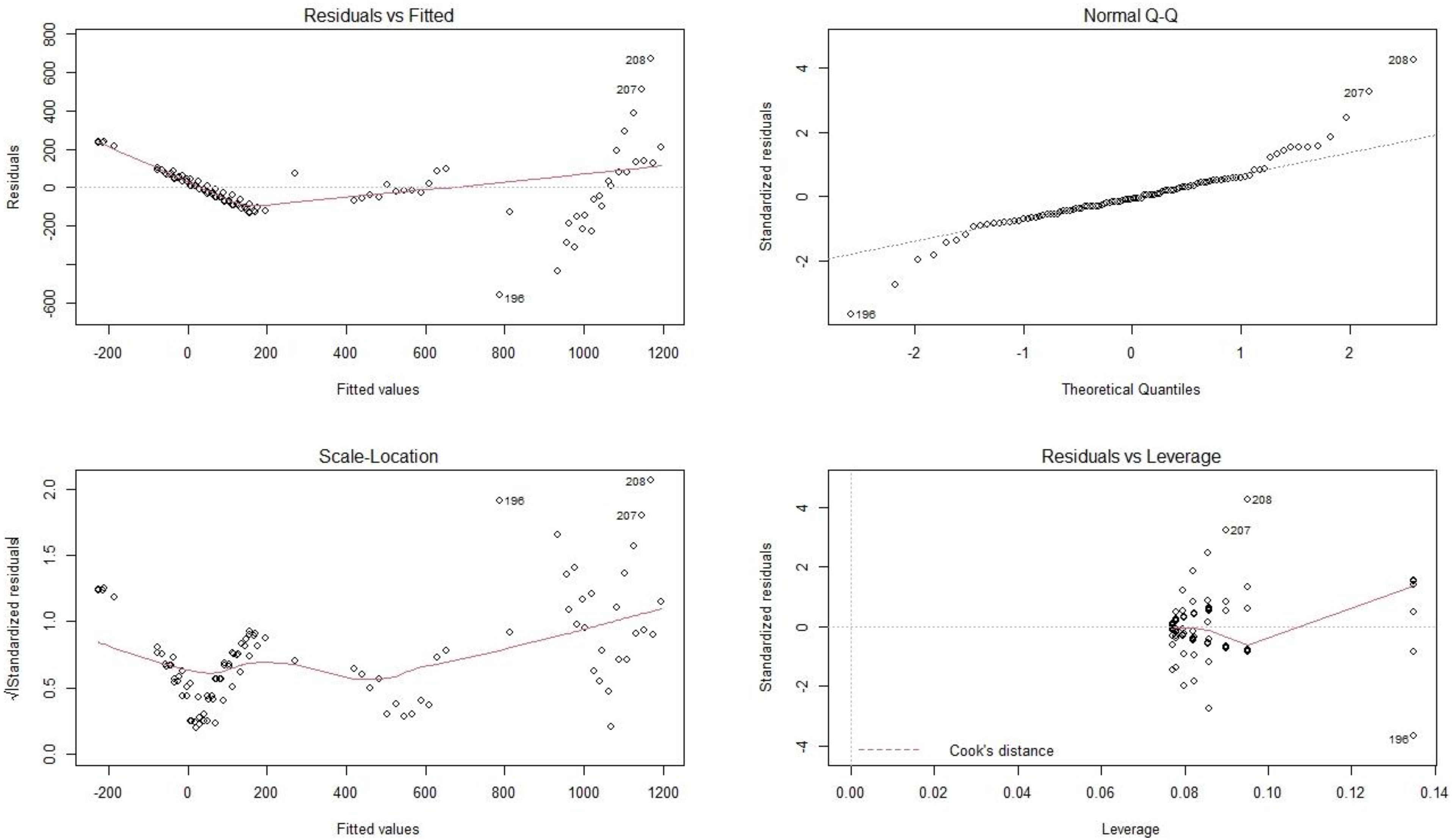
Focus on the column: the probability that F is greater than the listed value from the previous column. This is often called the *p value*. In most cases you put significance at the alpha=.05 level, or *we require the P value to be less then .05* to be considered statistically significant.

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Solaring <- aov(total.REC~Entity + Year + Solar, data =  REC.ORGIN)
> summary(Solaring)
            Df   Sum Sq Mean Sq F value   Pr(>F)    
Entity       7 19461739 2780248  258.00  < 2e-16 ***
Year         1  1032737 1032737   95.83 5.15e-16 ***
Solar        1  1589781 1589781  147.53  < 2e-16 ***
Residuals   94  1012963   10776                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> install.packages("AICcmodavg")
```

- **5.115e-16 < 0.05**
- **Therefore, we fail to reject the null hypothesis**

Bi-variate Analysis for Continuous Vs. Categorical :

Test of independence: Anova



Conclusion:

- We see in this Project the rapid growth of renewable technologies in the World
- This interactive chart shows the amount of energy generated from solar power each year.
- Solar generation at scale – compared to hydropower, for example – is a relatively modern renewable energy source but is growing quickly in many countries across the world.

Thank you for your attention!