COVID-19 & Diet

DATA 622 FINAL PROJECT

Abdellah Ait Elmouden | Gabriel Abreu | Jered Ataky

CUNY School of
Professional Studies

The Dataset

The dataset combined data of :

- Different types of food,
- World population obesity and
- undernourished rate, and
- Global COVID-19 cases count from around the world

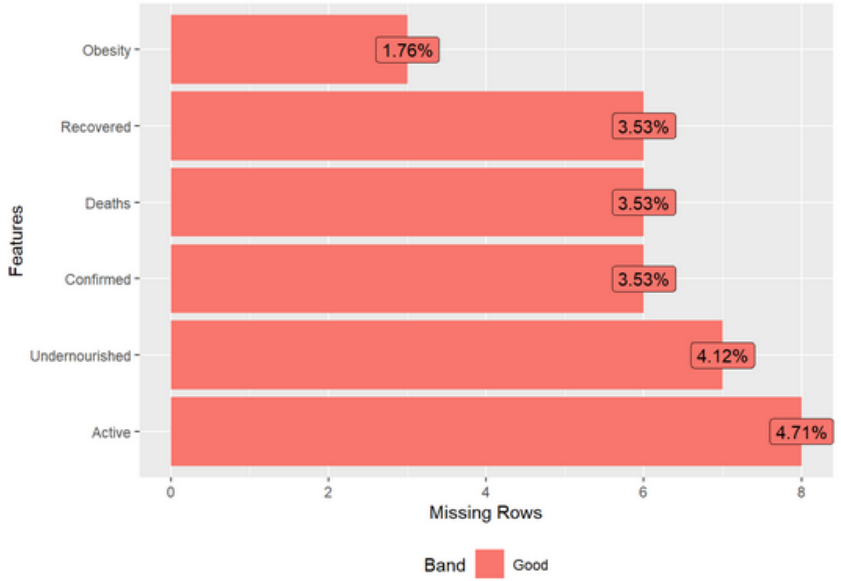
Methodology

- Exploratory Data Analysis (EDA)
- Machine Learning :
  - ✓ PCA analysis
  - ✓ K-mean Clustering
  - ✓ Random forest,
  - ✓ Gradient Boosting
  - ✓ Cubist
  - ✓ Support Vector Regression

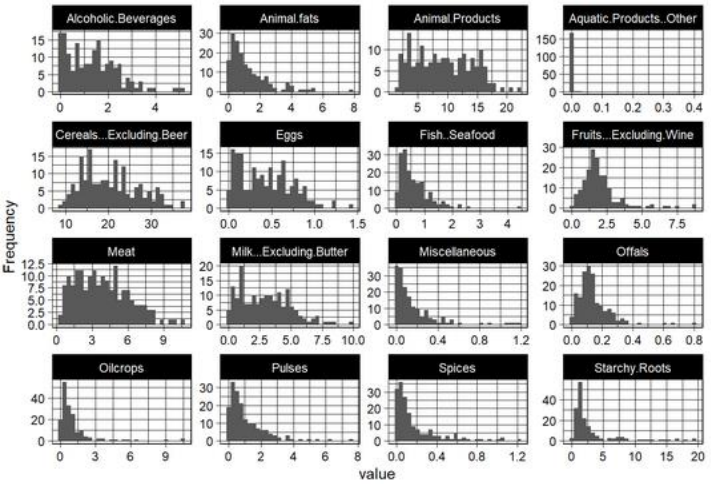
Data Exploration and Processing

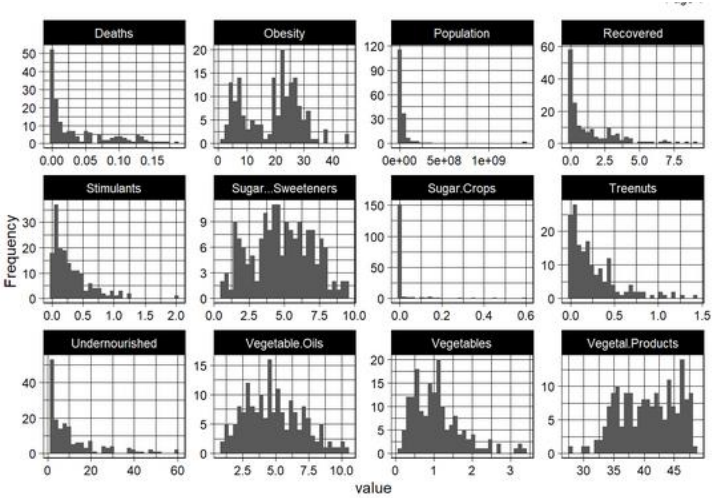
```
##  [1] "Country"                    "Alcoholic.Beverages"
##  [3] "Animal.Products"            "Animal.fats"
##  [5] "Aquatic.Products..Other"    "Cereals...Excluding.Beer"
##  [7] "Eggs"                       "Fish..Seafood"
##  [9] "Fruits...Excluding.Wine"    "Meat"
## [11] "Milk...Excluding.Butter"    "Miscellaneous"
## [13] "Offals"                     "Oilcrops"
## [15] "Pulses"                     "Spices"
## [17] "Starchy.Roots"              "Stimulants"
## [19] "Sugar.Crops"                "Sugar...Sweeteners"
## [21] "Treenuts"                   "Vegetal.Products"
## [23] "Vegetable.Oils"             "Vegetables"
## [25] "Obesity"                    "Undernourished"
## [27] "Confirmed"                  "Deaths"
## [29] "Recovered"                  "Active"
## [31] "Population"                 "Unit..all.except.Population."
```

Missing Data

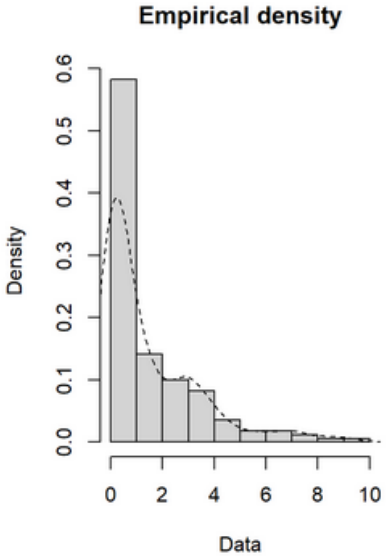


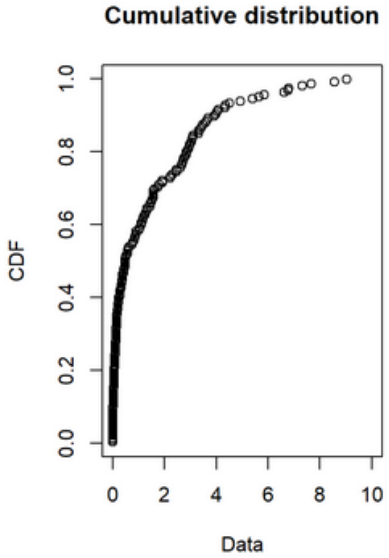
Statistical Analyses





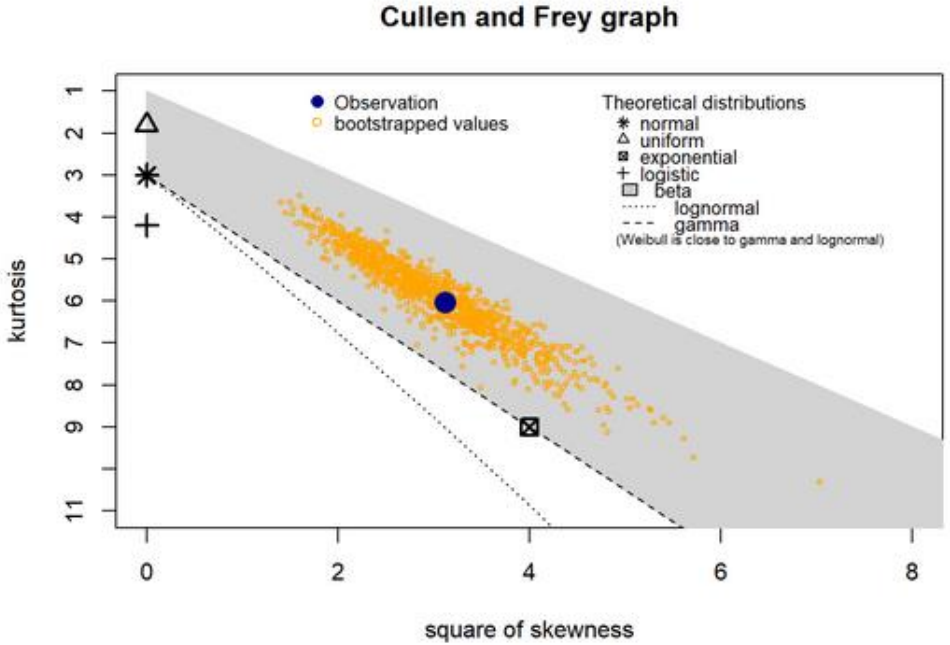
Statistical Analyses



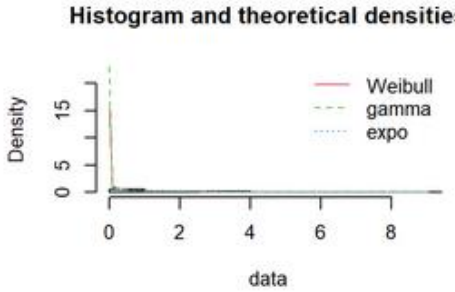


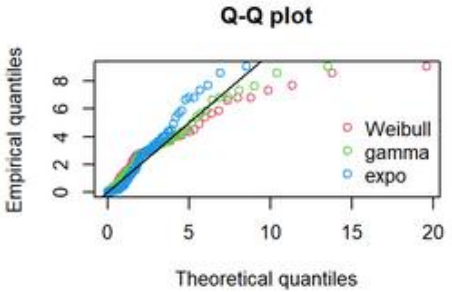
From the empirical

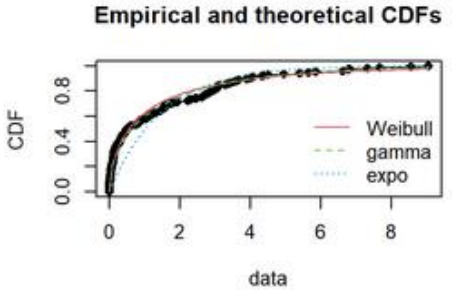
Statistical Analyses

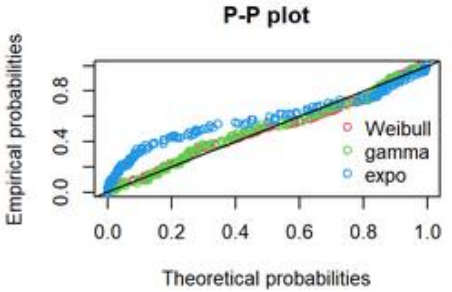


Statistical Analyses

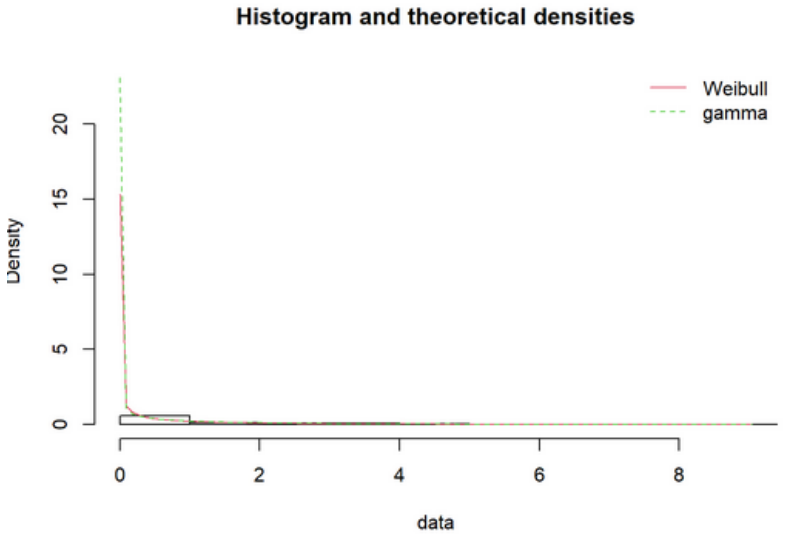








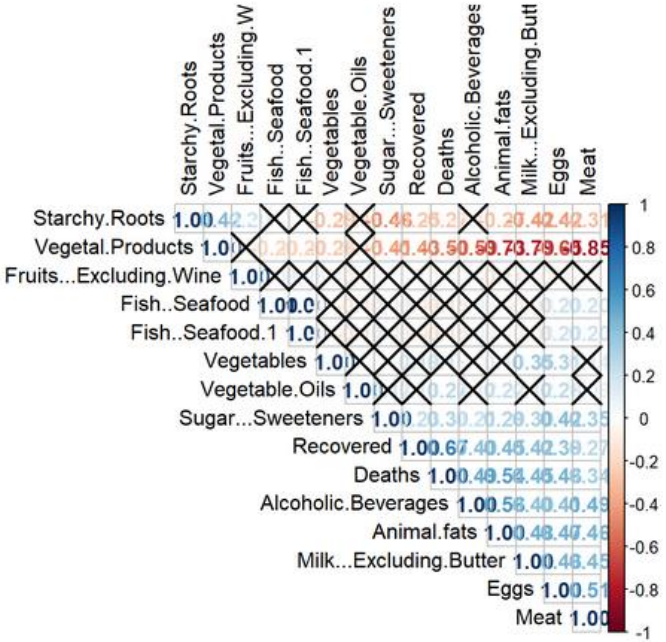
Statistical Analyses



It seems that still both

PCA

Code



From the plot we can

PCA

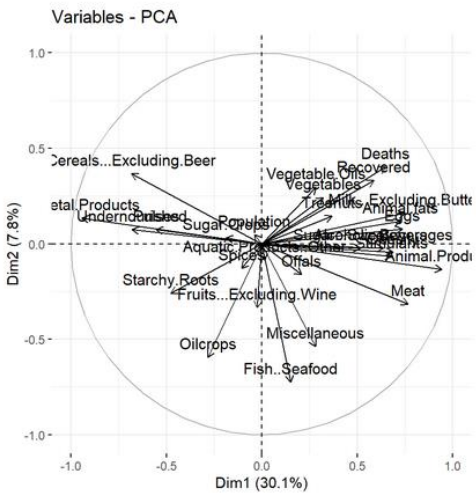
PCA Summary

Code

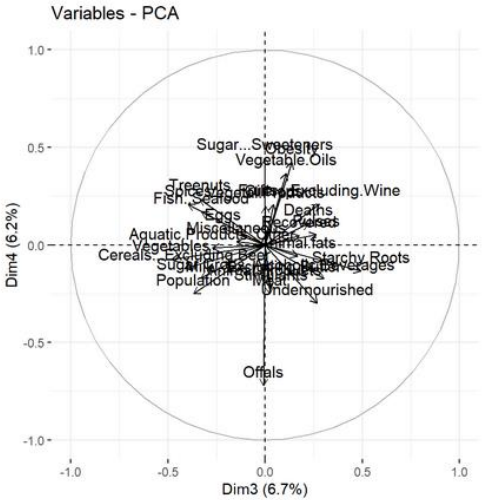
```
## Importance of components:
##                           PC1     PC2     PC3     PC4    PC5     PC6     PC7
## Standard deviation     2.9034 1.47529 1.36569 1.32158 1.1833 1.15260 1.07568
## Proportion of Variance 0.3011 0.07773 0.06661 0.06238 0.0500 0.04745 0.04132
## Cumulative Proportion  0.3011 0.37880 0.44541 0.50779 0.5578 0.60524 0.64657
##                            PC8     PC9   PC10    PC11   PC12    PC13    PC14
## Standard deviation     1.03489 0.99988 0.9742 0.96614 0.8663 0.83569 0.77697
## Proportion of Variance 0.03825 0.03571 0.0339 0.03334 0.0268 0.02494 0.02156
## Cumulative Proportion  0.68481 0.72052 0.7544 0.78775 0.8145 0.83950 0.86106
##                           PC15    PC16    PC17    PC18    PC19   PC20    PC21
## Standard deviation     0.73837 0.69250 0.65579 0.63972 0.63625 0.6033 0.57680
## Proportion of Variance 0.01947 0.01713 0.01536 0.01462 0.01446 0.0130 0.01188
## Cumulative Proportion  0.88053 0.89765 0.91301 0.92763 0.94209 0.9551 0.96697
##                           PC22    PC23    PC24    PC25    PC26     PC27
## Standard deviation     0.53723 0.50452 0.45911 0.41348 0.00224 0.001914
## Proportion of Variance 0.01031 0.00909 0.00753 0.00611 0.00000 0.000000
## Cumulative Proportion  0.97728 0.98637 0.99389 1.00000 1.00000 1.000000
##                           PC28
## Standard deviation     1.047e-05
## Proportion of Variance 0.000e+00
## Cumulative Proportion  1.000e+00
```

PCA

PC1-PC2

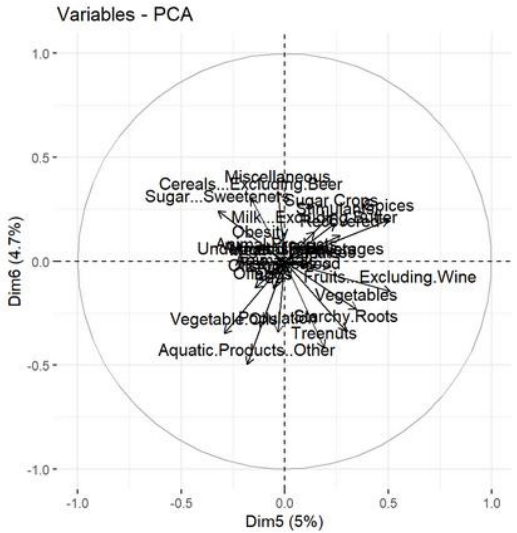


PC3-PC4

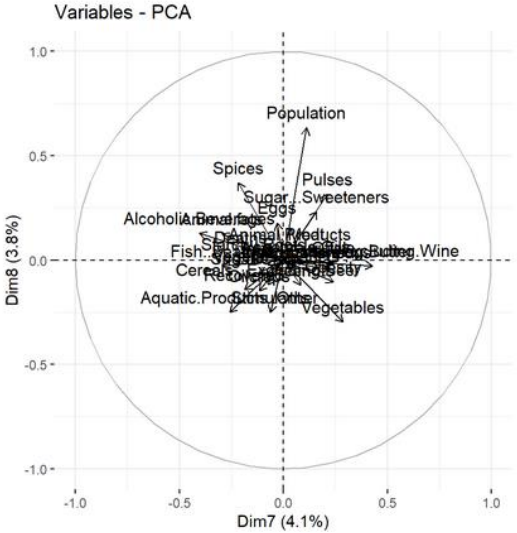


PCA

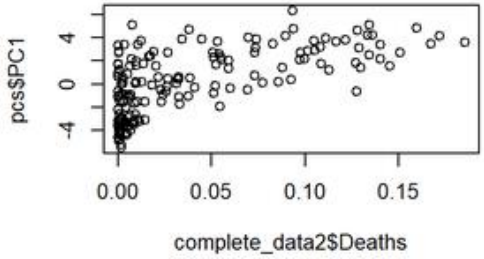
PC5-PC6

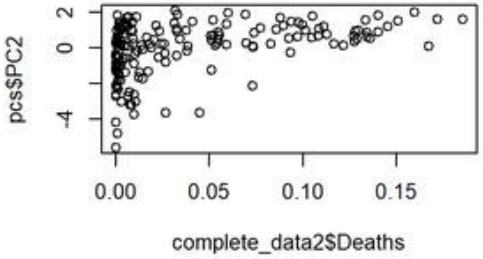


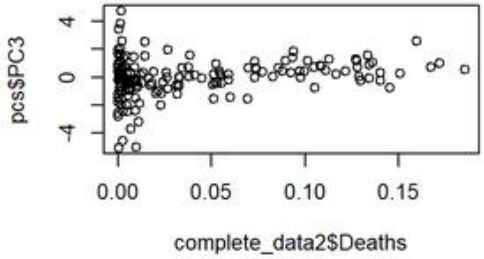
PC7-PC8

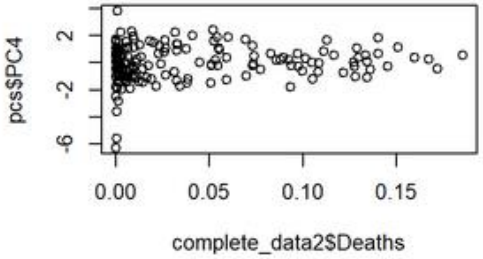


Linear regression with principal components

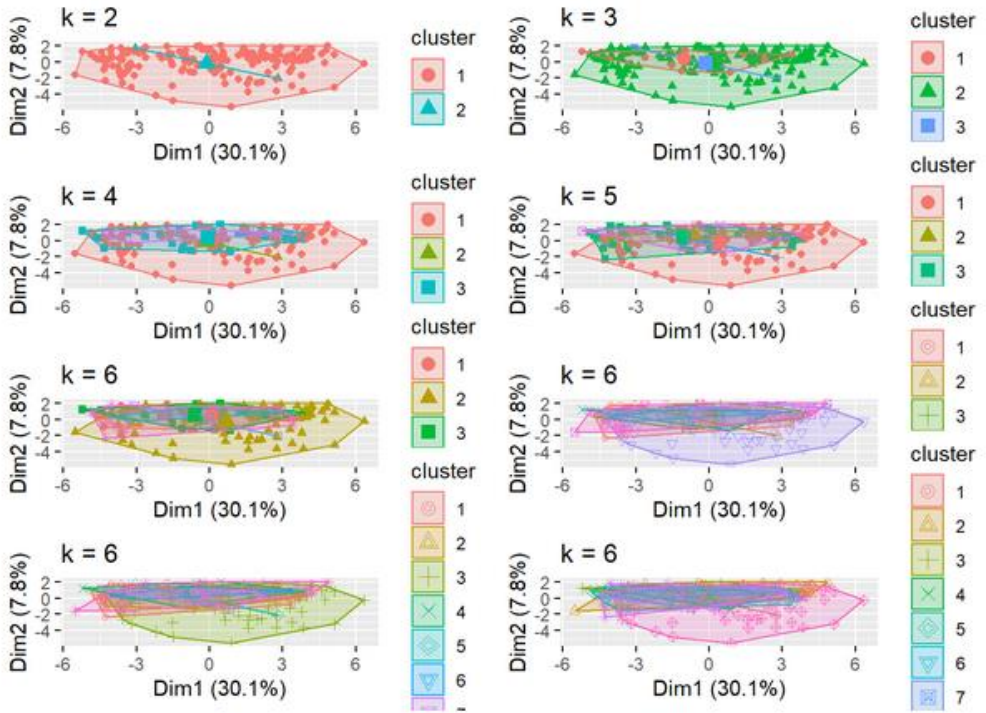






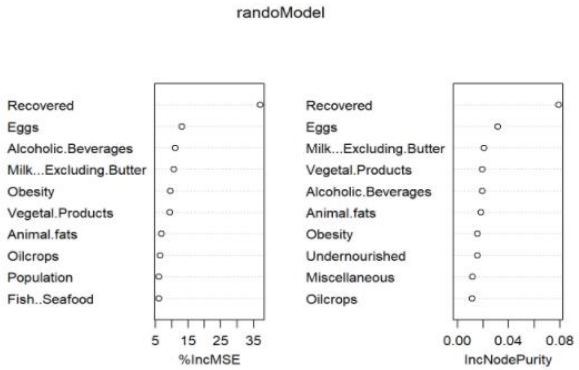


K-mean

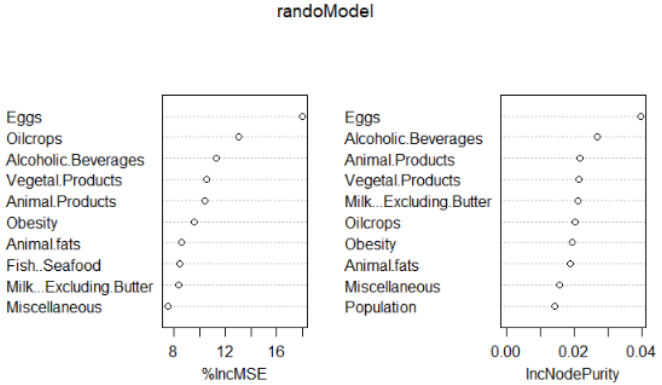


Trees – Random Forest, Cubist, Gradient Boost

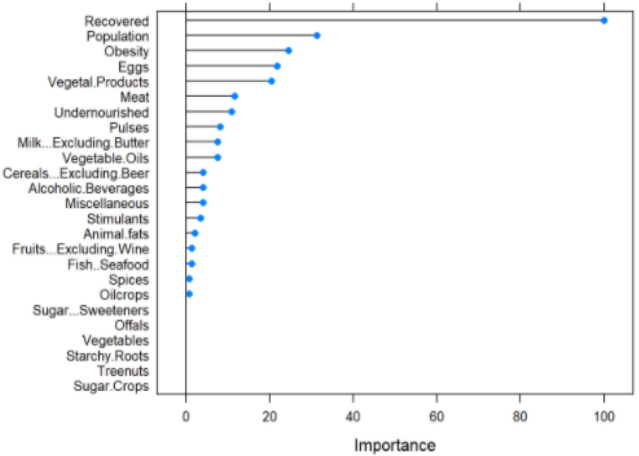
Random Forest: Food Only

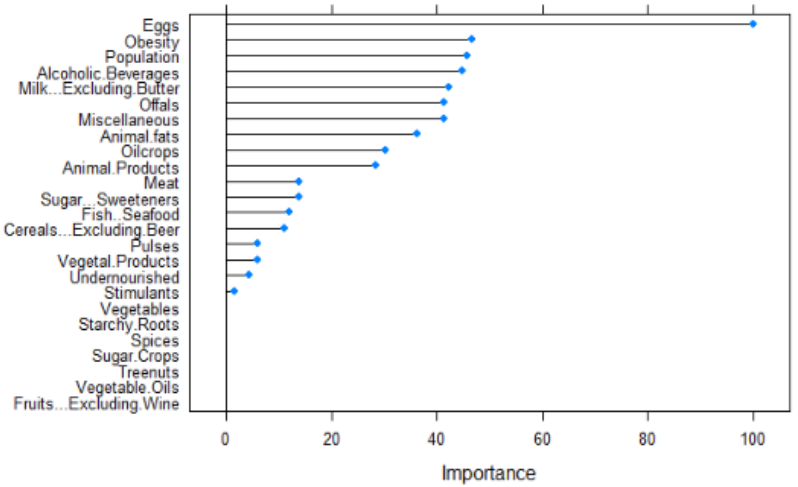


Random Forest

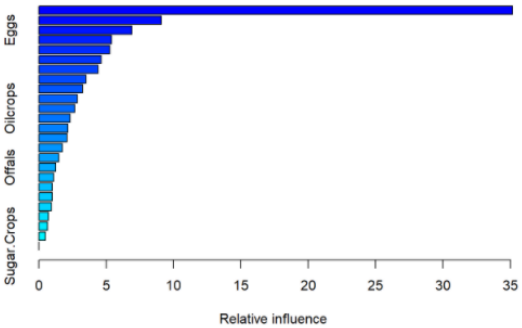


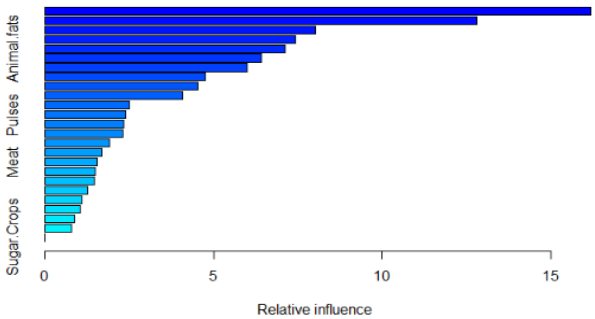
Cubist


Recovered
Population
Obesity
Eggs
Vegetal.Products
Meat
Undernourished
Pulses
Milk...Excluding.Butter
Vegetable.Oils
Cereals...Excluding.Beer
Alcoholic.Beverages
Miscellaneous
Stimulants
Animal.fats
Fruits...Excluding.Wine
Fish..Seafood
Spices
Oilcrops
Sugar...Sweeteners
Offals
Vegetables
Starchy.Roots
Treenuts
Sugar.Crops
0
20
40
60
80
100
Importance


Eggs
Obesity
Population
Alcoholic.Beverages
Milk...Excluding.Butter
Offals
Miscellaneous
Animal.fats
Oilcrops
Animal.Products
Meat
Sugar...Sweeteners
Fish..Seafood
Cereals...Excluding.Beer
Pulses
Vegetal.Products
Undernourished
Stimulants
Vegetables
Starchy.Roots
Spices
Sugar.Crops
Treenuts
Vegetable.Oils
Fruits...Excluding.Wine
0
20
40
60
80
100
Importance

Gradient Boost






```
##                                          var    rel.inf
## Recovered                          Recovered 35.1600955
## Alcoholic.Beverages  Alcoholic.Beverages  9.0825219
## Eggs                                  Eggs  6.9010083
## Milk...Excluding.Butter  Milk...Excluding.Butter  5.3945097
## Vegetal.Products            Vegetal.Products  5.2604965
## Animal.fats                      Animal.fats  4.6215363
## Undernourished              Undernourished  4.4023647
## Miscellaneous                Miscellaneous  3.5102497
## Population                      Population  3.2351117
## Vegetable.Oils              Vegetable.Oils  2.8498912
## Oilcrops                        Oilcrops  2.6793660
## Pulses                            Pulses  2.3313082
## Obesity                          Obesity  2.1397723
## Treenuts                        Treenuts  2.0770761
## Cereals...Excluding.Beer Cereals...Excluding.Beer  1.7274599
## Fish..Seafood                Fish..Seafood  1.4909817
## Offals                            Offals  1.2420637
## Stimulants                    Stimulants  1.1062169
## Meat                                Meat  0.9992101
## Vegetables                    Vegetables  0.9882710
## Spices                            Spices  0.9373554
## Fruits...Excluding.Wine  Fruits...Excluding.Wine  0.7179831
## Sugar...Sweeteners          Sugar...Sweeteners  0.6502176
## Starchy.Roots                Starchy.Roots  0.4949325
## Sugar.Crops                    Sugar.Crops  0.0000000
```

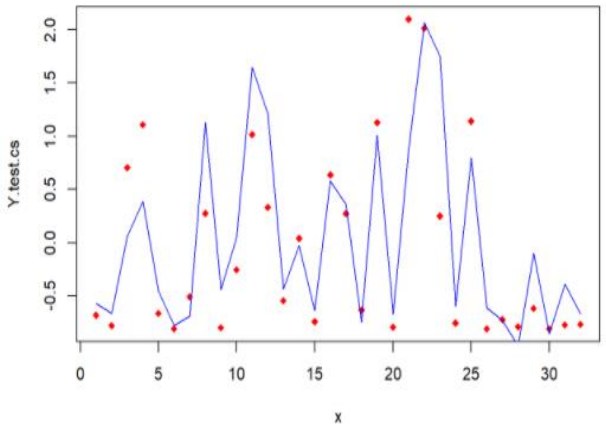
	var <fctr>	rel.inf <dbl>
Eggs	Eggs	16.1816192
Alcoholic.Beverages	Alcoholic.Beverages	12.7966105
Oilcrops	Oilcrops	8.0096502
Milk...Excluding.Butter	Milk...Excluding.Butter	7.4407244
Animal.fats	Animal.fats	7.1066959
Animal.Products	Animal.Products	6.4278695
Obesity	Obesity	5.9832168
Fish..Seafood	Fish..Seafood	4.7433491
Miscellaneous	Miscellaneous	4.5398486
Population	Population	4.0877509

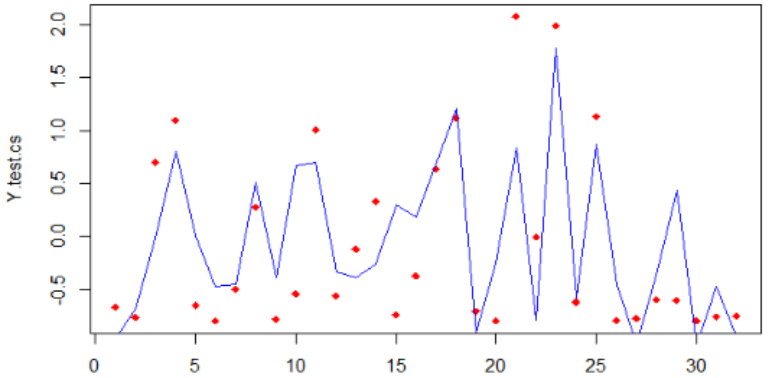
Tree Prediction Results

	RMSE	Rsquared	MAE
Random Forest	0.0224314	0.7182063	0.0153533
Gradient Boosted Tree	0.0228734	0.7221370	0.0165111
Cubist	0.0278355	0.6732597	0.0170036

	RMSE	Rsquared	MAE
Random Forest	0.0295002	0.5064067	0.0214034
Gradient Boosted Tree	0.0303363	0.4756502	0.0217214
Cubist	0.0338986	0.3475386	0.0216508

Support Vector Regression Plots



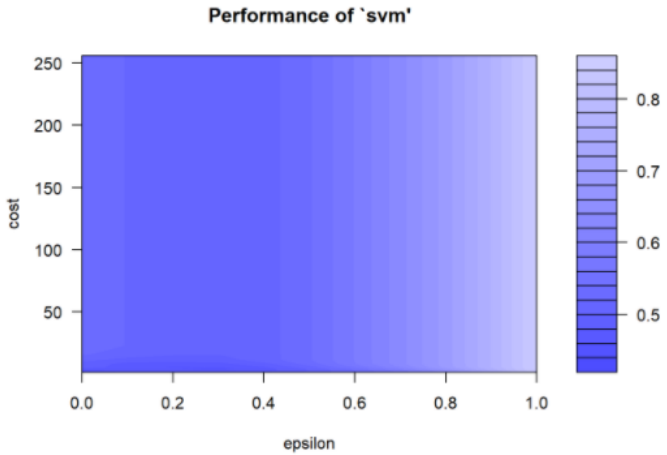


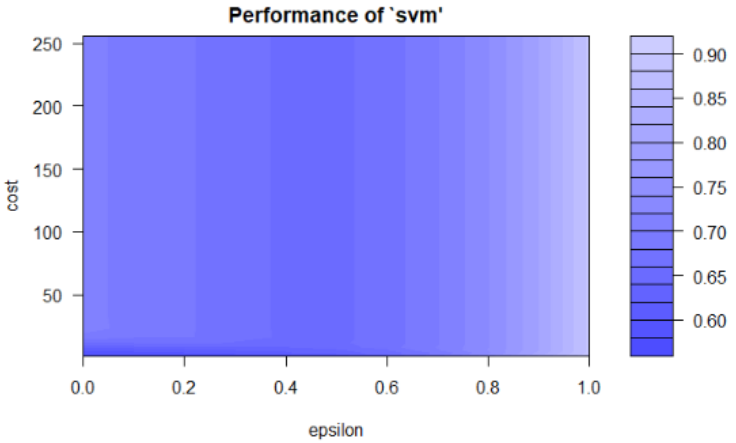
Code

```
##  MAE: 0.3288479
##  MSE 0.2374864
##  RMSE: 0.4873258
##  R-Squared: 0.673371
```

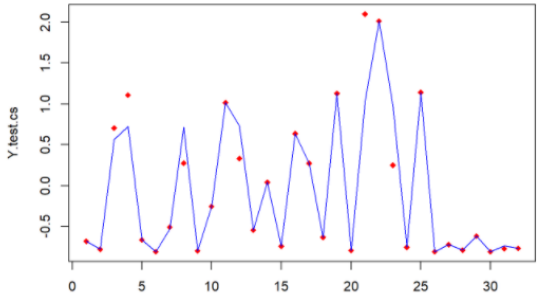
```
MAE:  0.4139295
MSE 0.2809483
RMSE:  0.5300456
R-Squared:  0.4499807
```

Tuned Grid Plots



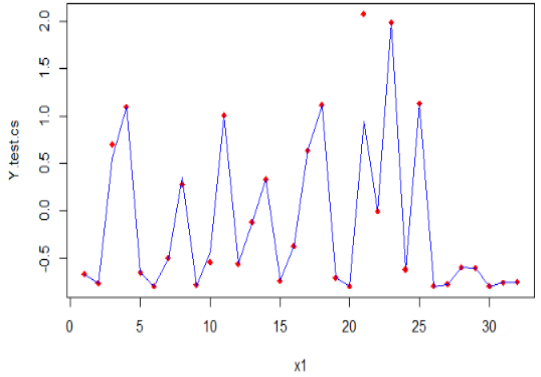


Tuned SVR Plots and Results




```
##  MAE: 0.1001404
##  MSE 0.06744831
##  RMSE: 0.2597081
##  R-Squared: 0.8997206
```

```
##                    Recovered                Alcoholic.Beverages                        Animal.fats
##                  28.2607712                        13.4241681                          8.3170989
##                   Stimulants                    Vegetable.Oils                             Obesity
##                    7.5151268                        7.3795706                          6.6640575
##             Vegetal.Products                    Fish..Seafood           Milk...Excluding.Butter
##                    6.6186768                        6.6145386                          6.4610278
##                    Oilcrops                             Meat                 Sugar...Sweeteners
##                    6.4095564                        6.3819978                          5.7479848
## Cereals...Excluding.Beer                    Starchy.Roots                      Undernourished
##                    5.6192914                        5.3327217                          5.2363520
##                       Offals                          Pulses                               Eggs
##                    4.7150095                        4.3883377                          4.2977024
##                  Sugar.Crops          Fruits...Excluding.Wine                          Treenuts
##                    4.2012758                        3.3345405                          3.1727651
##                   Vegetables                    Miscellaneous                            Spices
##                    2.2829954                        0.9078458                          0.6570656
##                   Population
##                    0.4452825
```




```
MAE: 0.04570103
MSE 0.04143534
RMSE: 0.2035567
R-Squared: 0.9316657
```

```
Alcoholic.Beverages                Animal.fats                 Stimulants                   Oilcrops                    Obesity
          16.4929304                 11.1736033                  9.6760724                  9.2417270                  8.2925872
     Vegetable.Oils             Vegetal.Products          Animal.Products  Milk...Excluding.Butter             Starchy.Roots
          6.5122790                  6.4161505                  6.3133139                  6.2089923                  6.1961900
                Eggs          Sugar...Sweeteners                       Meat Cereals...Excluding.Beer          Undernourished
          6.1794657                  6.1202952                  5.0946005                  4.8510191                  4.7400024
      Sugar.Crops             Fish..Seafood                  Treenuts                      Pulses  Fruits...Excluding.Wine
          4.2676529                  4.0257740                  3.8684046                  3.4483294                  3.1342840
           Offals             Miscellaneous                 Population                     Spices                 Vegetables
          2.9662306                  1.3472979                  0.9580128                  0.5983853                  0.2748476
```

Conclusion & Findings

We used several models to study the correlation between countries' diet and COVID-19 mortality rate
We found out that food cultures is very important key as in predicting the mortality rate of COVID-19
The models show foods associated with overall poor health increases the mortality rate of COVID-19 in each country
A population which has a healthy diet (food based on vegetal products, cereals,..) has a lower death rate in comparison with a population which has a higher obesity rate and consumes more animal products

Thank you!