

PR490

KUTA NDZE

8/12/2020

LOADING PACKAGES

```
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```
library(plotly)
```

```
## Warning: package 'plotly' was built under R version 3.6.2
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 3.6.2
```

```
##
```

```
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
```

```
##
```

```
##      last_plot
```

```
## The following object is masked from 'package:stats':
```

```
##
```

```
##      filter
```

```
## The following object is masked from 'package:graphics':
```

```
##
```

```
##      layout
```

```
library(data.table)
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v tibble  3.0.3      v dplyr    1.0.0
```

```
## v tidyr   1.1.1      v stringr 1.4.0
```

```
## v readr   1.3.1      v forcats 0.5.0
```

```
## v purrr   0.3.4
```

```

## Warning: package 'tibble' was built under R version 3.6.2

## Warning: package 'tidyr' was built under R version 3.6.2

## Warning: package 'purrr' was built under R version 3.6.2

## Warning: package 'dplyr' was built under R version 3.6.2

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::between()   masks data.table::between()
## x dplyr::filter()    masks plotly::filter(), stats::filter()
## x dplyr::first()     masks data.table::first()
## x dplyr::lag()       masks stats::lag()
## x dplyr::last()      masks data.table::last()
## x purrr::transpose() masks data.table::transpose()

library(GGally)

## Warning: package 'GGally' was built under R version 3.6.2

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg      ggplot2

library(readxl)
library(car)

## Warning: package 'car' was built under R version 3.6.2

## Loading required package: carData

## Warning: package 'carData' was built under R version 3.6.2

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##   recode

## The following object is masked from 'package:purrr':
##
##   some

library(caret)

## Loading required package: lattice

## Warning: package 'lattice' was built under R version 3.6.2

```

```
##
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':
##
## lift

library(scales)

## Warning: package 'scales' was built under R version 3.6.2

##
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':
##
## discard

## The following object is masked from 'package:readr':
##
## col_factor

library(lmtest)

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 3.6.2

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
## as.Date, as.Date.numeric

library(nnet)

## Warning: package 'nnet' was built under R version 3.6.2

library(glmnet)

## Warning: package 'glmnet' was built under R version 3.6.2

## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':
##
## expand, pack, unpack

## Loaded glmnet 4.0-2
```

```
library(dplyr)
library(ResourceSelection)
```

```
## ResourceSelection 0.3-5    2019-07-22
```

```
library(readxl)
library(ggplot2)
library(ggcorrplot)
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 3.6.2
```

```
##
```

```
## Attaching package: 'reshape2'
```

```
## The following object is masked from 'package:tidyr':
```

```
##
```

```
##      smiths
```

```
## The following objects are masked from 'package:data.table':
```

```
##
```

```
##      dcast, melt
```

```
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 3.6.2
```

```
##
```

```
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
```

```
##
```

```
##      select
```

```
## The following object is masked from 'package:plotly':
```

```
##
```

```
##      select
```

```
library(olsrr)
```

```
##
```

```
## Attaching package: 'olsrr'
```

```
## The following object is masked from 'package:MASS':
```

```
##
```

```
##      cement
```

```
## The following object is masked from 'package:datasets':
```

```
##
```

```
##      rivers
```

```
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 3.6.2
```

READ DATA

```
#Read data
```

```
Covid19=read_xlsx("/Users/celdrick/Desktop/COVID19/covid19.xlsx")
```

```
Covid19
```

```
## # A tibble: 54 x 22
```

```
##   STATE STCD REGION CDHS HOSC ICU PLUF SINC POPD POPS HOML HUMI  
##   <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Alab~ AL South~ 1265 1547 0 16.8 2.19e5 96.9 4.91e6 3261 76.5  
## 2 Alas~ AK West 11 34 0 11.1 4.61e4 1.29 7.34e5 1907 81.5  
## 3 Ariz~ AZ South~ 2443 3094 870 14.1 3.46e5 65.0 7.38e6 10007 79.4  
## 4 Arka~ AR South~ 362 474 0 16.8 1.38e5 58.4 3.04e6 2717 76.9  
## 5 Cali~ CA West 7100 8820 2284 12.8 2.70e6 256. 3.99e7 151278 80.4  
## 6 Colo~ CO West 1643 373 0 9.7 3.59e5 56.4 5.85e6 9619 79.7  
## 7 Conn~ CT North~ 4031 63 0 10.3 2.86e5 736. 3.56e6 3033 79.3  
## 8 Dela~ DE North~ 517 61 7 12.2 5.37e4 504. 9.83e5 921 72.0  
## 9 Dist~ DC Terri~ 644 81 18 16.1 6.08e4 NA NA NA 77.4  
## 10 Flor~ FL South~ 4341 9475 0 13.7 1.14e6 410. 2.20e7 28328 77.0  
## # ... with 44 more rows, and 10 more variables: UNEM <dbl>, MEDA <dbl>,  
## # LEXP <dbl>, ADEP <dbl>, ATEM <dbl>, APRE <dbl>, CIGA <dbl>, OBES <dbl>,  
## # CNCS <dbl>, MRAT <dbl>
```

```
head(Covid19)
```

```
## # A tibble: 6 x 22
```

```
##   STATE STCD REGION CDHS HOSC ICU PLUF SINC POPD POPS HOML HUMI  
##   <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Alab~ AL South~ 1265 1547 0 16.8 2.19e5 96.9 4.91e6 3261 76.5  
## 2 Alas~ AK West 11 34 0 11.1 4.61e4 1.29 7.34e5 1907 81.5  
## 3 Ariz~ AZ South~ 2443 3094 870 14.1 3.46e5 65.0 7.38e6 10007 79.4  
## 4 Arka~ AR South~ 362 474 0 16.8 1.38e5 58.4 3.04e6 2717 76.9  
## 5 Cali~ CA West 7100 8820 2284 12.8 2.70e6 256. 3.99e7 151278 80.4  
## 6 Colo~ CO West 1643 373 0 9.7 3.59e5 56.4 5.85e6 9619 79.7  
## # ... with 10 more variables: UNEM <dbl>, MEDA <dbl>, LEXP <dbl>, ADEP <dbl>,  
## # ATEM <dbl>, APRE <dbl>, CIGA <dbl>, OBES <dbl>, CNCS <dbl>, MRAT <dbl>
```

```
tail(Covid19)
```

```
## # A tibble: 6 x 22
```

```
##   STATE STCD REGION CDHS HOSC ICU PLUF SINC POPD POPS HOML HUMI  
##   <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Verm~ VT North~ 57 14 0 10.7 35805 68.1 6.28e5 1089 76.9  
## 2 Virg~ VA South~ 2068 1157 253 10.7 525430 218. 8.63e6 5783 76.8  
## 3 Wash~ WA West 1235 392 0 10.3 506280 117. 7.80e6 21577 79.1  
## 4 West~ WV South~ 103 78 33 17.4 76129 74.0 1.78e6 1397 78.3  
## 5 Wisc~ WI Midwe~ 825 167 63 11.1 318475 108. 5.85e6 4538 77.1
```

```
## 6 Wyom~ WY      West      25      13      0 10.7 37252  5.84 5.67e5  548 79.5
## # ... with 10 more variables: UNEM <dbl>, MEDA <dbl>, LEXP <dbl>, ADEP <dbl>,
## #   ATEM <dbl>, APRE <dbl>, CIGA <dbl>, OBES <dbl>, CNCS <dbl>, MRAT <dbl>
```

```
#Viewing the data structure
ncol(Covid19)
```

```
## [1] 22
```

```
str(Covid19)
```

```
## tibble [54 x 22] (S3: tbl_df/tbl/data.frame)
## $ STATE : chr [1:54] "Alabama" "Alaska" "Arizona" "Arkansas" ...
## $ STCD : chr [1:54] "AL" "AK" "AZ" "AR" ...
## $ REGION: chr [1:54] "Southeast" "West" "Southwest" "Southeast" ...
## $ CDHS : num [1:54] 1265 11 2443 362 7100 ...
## $ HOSC : num [1:54] 1547 34 3094 474 8820 ...
## $ ICU : num [1:54] 0 0 870 0 2284 ...
## $ PLUF : num [1:54] 16.8 11.1 14.1 16.8 12.8 9.7 10.3 12.2 16.1 13.7 ...
## $ SINC : num [1:54] 219230 46099 346009 137609 2701899 ...
## $ POPD : num [1:54] 96.92 1.29 64.95 58.4 256.37 ...
## $ POPS : num [1:54] 4908620 734002 7378490 3039000 39937500 ...
## $ HOML : num [1:54] 3261 1907 10007 2717 151278 ...
## $ HUMI : num [1:54] 76.5 81.5 79.4 76.9 80.4 ...
## $ UNEM : num [1:54] 7.5 12.4 10 8 14.9 10.5 9.8 12.5 8.6 10.4 ...
## $ MEDA : num [1:54] 20 21 22 27 26 18 21 21 28 18 ...
## $ LEXP : num [1:54] 75.4 78.3 79.5 76 80.8 80 80.8 78.4 NA 79.4 ...
## $ ADEP : num [1:54] 63.1 55.8 67.2 66.4 58.1 56.7 59.8 64 NA 66.3 ...
## $ ATEM : num [1:54] 62.8 26.6 60.3 60.4 59.4 45.1 49 55.3 NA 70.7 ...
## $ APRE : num [1:54] 58.3 22.5 13.6 50.6 22.2 15.9 50.3 45.7 NA 54.5 ...
## $ CIGA : num [1:54] 19.2 19.1 14 22.7 11.2 14.5 12.2 16.5 NA 14.5 ...
## $ OBES : num [1:54] 36.2 29.5 29.5 37.1 25.8 23 27.4 33.5 24.7 30.7 ...
## $ CNCS : num [1:54] 71813 2619 150609 35246 413576 ...
## $ MRAT : num [1:54] 1.76 0.42 1.62 1.03 1.72 ...
```

```
summary(Covid19)
```

```
##      STATE      STCD      REGION      CDHS
## Length:54      Length:54      Length:54      Min.   : 2.0
## Class :character Class :character Class :character 1st Qu.: 228.5
## Mode  :character Mode  :character Mode  :character Median : 809.5
##                                     Mean  : 2036.8
##                                     3rd Qu.: 2664.0
##                                     Max.   :13811.0
##
##      HOSC      ICU      PLUF      SINC
## Min.   : 0.0   Min.   : 0.0   Min.   : 7.60   Min.   : 35805
## 1st Qu.: 64.0   1st Qu.: 0.0   1st Qu.:10.70   1st Qu.: 85847
## Median : 403.5   Median : 13.0   Median :12.80   Median : 226135
## Mean   : 1104.2   Mean   : 193.6   Mean   :12.86   Mean   : 371629
## 3rd Qu.: 1101.0   3rd Qu.: 149.5   3rd Qu.:14.30   3rd Qu.: 507412
## Max.   :10893.0   Max.   :3281.0   Max.   :19.80   Max.   :2701899
```

```
##                                     NA's :3      NA's :3
##      POPD              POPS              HOML              HUMI
## Min.   : 1.286   Min.   : 567025   Min.   : 548   Min.   :71.31
## 1st Qu.: 47.707   1st Qu.: 1857762   1st Qu.: 2315   1st Qu.:75.81
## Median : 107.784   Median : 4572435   Median : 4355   Median :77.14
## Mean   : 203.901   Mean   : 6611970   Mean   : 11113   Mean   :77.57
## 3rd Qu.: 219.567   3rd Qu.: 7692448   3rd Qu.: 9543   3rd Qu.:79.45
## Max.   :1215.198   Max.   :39937500   Max.   :151278   Max.   :82.01
## NA's   :4         NA's   :4         NA's   :4         NA's   :3
##      UNEM              MEDA              LEXP              ADEP
## Min.   : 4.300   Min.   :10.00   Min.   :75.00   Min.   :55.80
## 1st Qu.: 7.600   1st Qu.:17.00   1st Qu.:77.80   1st Qu.:59.92
## Median : 8.700   Median :19.00   Median :78.90   Median :62.20
## Mean   : 9.831   Mean   :20.15   Mean   :78.69   Mean   :62.32
## 3rd Qu.:11.500   3rd Qu.:22.25   3rd Qu.:79.90   3rd Qu.:64.38
## Max.   :17.400   Max.   :48.00   Max.   :82.30   Max.   :69.70
## NA's   :3         NA's   :2         NA's   :4         NA's   :4
##      ATEM              APRE              CIGA              OBES
## Min.   :26.60   Min.   : 2.20   Min.   : 9.00   Min.   :23.0
## 1st Qu.:45.25   1st Qu.:24.52   1st Qu.:14.45   1st Qu.:28.4
## Median :51.20   Median :41.75   Median :16.10   Median :30.9
## Mean   :51.94   Mean   :36.98   Mean   :16.47   Mean   :31.3
## 3rd Qu.:58.65   3rd Qu.:47.90   3rd Qu.:19.05   3rd Qu.:34.3
## Max.   :70.70   Max.   :63.70   Max.   :23.40   Max.   :39.5
## NA's   :4         NA's   :4         NA's   :3         NA's   :1
##      CNCS              MRAT
## Min.   : 38   Min.   :0.420
## 1st Qu.:13226   1st Qu.:1.413
## Median :39225   Median :1.987
## Mean   :73081   Mean   :2.845
## 3rd Qu.:82896   3rd Qu.:3.934
## Max.   :413576   Max.   :8.359
##
```

```
##Checking for Missing Values
colSums(is.na(Covid19))
```

```
## STATE STCD REGION CDHS HOSC ICU PLUF SINC POPD POPS HOML
##      0      0      0      0      0      0      3      3      4      4      4
## HUMI UNEM MEDA LEXP ADEP ATEM APRE CIGA OBES CNCS MRAT
##      3      3      2      4      4      4      4      3      1      0      0
```

```
##Checking for Empty Values
colSums(Covid19=='')
```

```
## STATE STCD REGION CDHS HOSC ICU PLUF SINC POPD POPS HOML
##      0      0      0      0      0      0      NA      NA      NA      NA      NA
## HUMI UNEM MEDA LEXP ADEP ATEM APRE CIGA OBES CNCS MRAT
##      NA      NA      NA      NA      NA      NA      NA      NA      NA      0      0
```

the territories in the original data set with missing entries are automatically deleted

```
#delete missing values
Covid=na.omit(Covid19)
Covid
```

```
## # A tibble: 50 x 22
##   STATE STCD REGION CDHS HOSC ICU PLUF SINC POPD POPS HOML HUMI
##   <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Alab~ AL South~ 1265 1547 0 16.8 2.19e5 96.9 4.91e6 3261 76.5
## 2 Alas~ AK West 11 34 0 11.1 4.61e4 1.29 7.34e5 1907 81.5
## 3 Ariz~ AZ South~ 2443 3094 870 14.1 3.46e5 65.0 7.38e6 10007 79.4
## 4 Arka~ AR South~ 362 474 0 16.8 1.38e5 58.4 3.04e6 2717 76.9
## 5 Cali~ CA West 7100 8820 2284 12.8 2.70e6 256. 3.99e7 151278 80.4
## 6 Colo~ CO West 1643 373 0 9.7 3.59e5 56.4 5.85e6 9619 79.7
## 7 Conn~ CT North~ 4031 63 0 10.3 2.86e5 736. 3.56e6 3033 79.3
## 8 Dela~ DE North~ 517 61 7 12.2 5.37e4 504. 9.83e5 921 72.0
## 9 Flor~ FL South~ 4341 9475 0 13.7 1.14e6 410. 2.20e7 28328 77.0
## 10 Geor~ GA South~ 2547 3179 0 14.5 5.22e5 187. 1.07e7 10443 75.8
## # ... with 40 more rows, and 10 more variables: UNEM <dbl>, MEDA <dbl>,
## # LEXP <dbl>, ADEP <dbl>, ATEM <dbl>, APRE <dbl>, CIGA <dbl>, OBES <dbl>,
## # CNCS <dbl>, MRAT <dbl>
```

```
##Checking for Missing Values
colSums(is.na(Covid))
```

```
## STATE STCD REGION CDHS HOSC ICU PLUF SINC POPD POPS HOML
## 0 0 0 0 0 0 0 0 0 0 0
## HUMI UNEM MEDA LEXP ADEP ATEM APRE CIGA OBES CNCS MRAT
## 0 0 0 0 0 0 0 0 0 0 0
```

```
##Checking for Empty Values
colSums(Covid=='')
```

```
## STATE STCD REGION CDHS HOSC ICU PLUF SINC POPD POPS HOML
## 0 0 0 0 0 0 0 0 0 0 0
## HUMI UNEM MEDA LEXP ADEP ATEM APRE CIGA OBES CNCS MRAT
## 0 0 0 0 0 0 0 0 0 0 0
```

```
summary(Covid)
```

```
## STATE STCD REGION CDHS
## Length:50 Length:50 Length:50 Min. : 11.0
## Class :character Class :character Class :character 1st Qu.: 293.2
## Mode :character Mode :character Mode :character Median : 953.5
## Mean : 2183.1
## 3rd Qu.: 2725.5
## Max. :13811.0
## HOSC ICU PLUF SINC
## Min. : 0.00 Min. : 0.0 Min. : 7.60 Min. : 35805
## 1st Qu.: 69.75 1st Qu.: 0.0 1st Qu.:10.70 1st Qu.: 89311
## Median : 489.50 Median : 20.5 Median :12.65 Median : 227950
## Mean : 1182.58 Mean : 207.8 Mean :12.80 Mean : 377846
```

```
## 3rd Qu.: 1128.25 3rd Qu.: 172.0 3rd Qu.:14.10 3rd Qu.: 507977
## Max. :10893.00 Max. :3281.0 Max. :19.80 Max. :2701899
## POPD POPS HOML HUMI
## Min. : 1.286 Min. : 567025 Min. : 548 Min. :71.31
## 1st Qu.: 47.707 1st Qu.: 1857762 1st Qu.: 2315 1st Qu.:75.78
## Median : 107.784 Median : 4572435 Median : 4355 Median :77.09
## Mean : 203.901 Mean : 6611970 Mean : 11113 Mean :77.57
## 3rd Qu.: 219.567 3rd Qu.: 7692448 3rd Qu.: 9543 3rd Qu.:79.47
## Max. :1215.198 Max. :39937500 Max. :151278 Max. :82.01
## UNEM MEDA LEXP ADEP
## Min. : 4.300 Min. :10.00 Min. :75.00 Min. :55.80
## 1st Qu.: 7.600 1st Qu.:17.00 1st Qu.:77.80 1st Qu.:59.92
## Median : 9.050 Median :19.00 Median :78.90 Median :62.20
## Mean : 9.856 Mean :19.44 Mean :78.69 Mean :62.32
## 3rd Qu.:11.650 3rd Qu.:22.00 3rd Qu.:79.90 3rd Qu.:64.38
## Max. :17.400 Max. :33.00 Max. :82.30 Max. :69.70
## ATEM APRE CIGA OBES
## Min. :26.60 Min. : 2.20 Min. : 9.00 Min. :23.00
## 1st Qu.:45.25 1st Qu.:24.52 1st Qu.:14.43 1st Qu.:28.55
## Median :51.20 Median :41.75 Median :16.05 Median :30.90
## Mean :51.94 Mean :36.98 Mean :16.36 Mean :31.43
## 3rd Qu.:58.65 3rd Qu.:47.90 3rd Qu.:18.98 3rd Qu.:34.38
## Max. :70.70 Max. :63.70 Max. :23.40 Max. :39.50
## CNCS MRAT
## Min. : 1366 Min. :0.420
## 1st Qu.: 16246 1st Qu.:1.333
## Median : 44065 Median :1.987
## Mean : 78429 Mean :2.797
## 3rd Qu.: 84245 3rd Qu.:3.684
## Max. :413576 Max. :8.359
```

```
Covid=Covid%>%
  mutate(STATE=as.factor(STATE), REGION=as.factor(REGION),STCD=as.factor(STCD))
```

EXPLORATORY ANALYSIS

```
#Before exploring the data variables, it is important to inspect and change the class of variables into
Covid.1=Covid[ ,-c(1:4)] #This data set will not contain the CDHS column for the CNCS analysis
str(Covid.1)
```

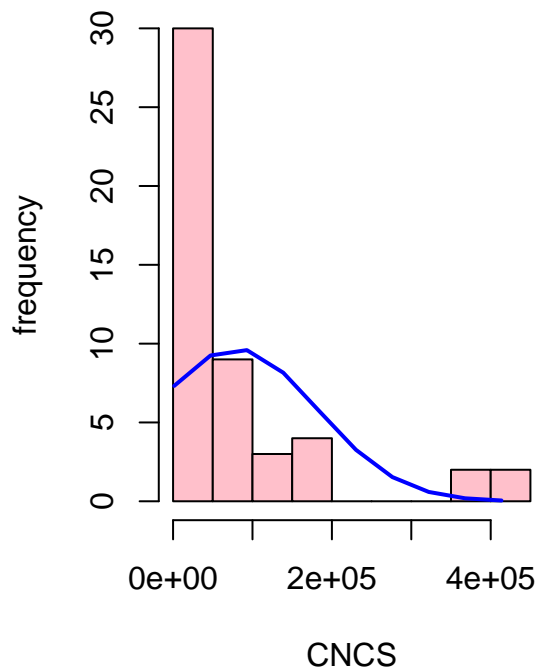
```
## tibble [50 x 18] (S3: tbl_df/tbl/data.frame)
## $ HOSC: num [1:50] 1547 34 3094 474 8820 ...
## $ ICU : num [1:50] 0 0 870 0 2284 ...
## $ PLUF: num [1:50] 16.8 11.1 14.1 16.8 12.8 9.7 10.3 12.2 13.7 14.5 ...
## $ SINC: num [1:50] 219230 46099 346009 137609 2701899 ...
## $ POPD: num [1:50] 96.92 1.29 64.95 58.4 256.37 ...
## $ POPS: num [1:50] 4908620 734002 7378490 3039000 39937500 ...
## $ HOML: num [1:50] 3261 1907 10007 2717 151278 ...
## $ HUMI: num [1:50] 76.5 81.5 79.4 76.9 80.4 ...
## $ UNEM: num [1:50] 7.5 12.4 10 8 14.9 10.5 9.8 12.5 10.4 7.6 ...
## $ MEDA: num [1:50] 20 21 22 27 26 18 21 21 18 17 ...
## $ LEXP: num [1:50] 75.4 78.3 79.5 76 80.8 80 80.8 78.4 79.4 77.2 ...
## $ ADEP: num [1:50] 63.1 55.8 67.2 66.4 58.1 56.7 59.8 64 66.3 59.8 ...
```

```
## $ ATEM: num [1:50] 62.8 26.6 60.3 60.4 59.4 45.1 49 55.3 70.7 63.5 ...
## $ APRE: num [1:50] 58.3 22.5 13.6 50.6 22.2 15.9 50.3 45.7 54.5 50.7 ...
## $ CIGA: num [1:50] 19.2 19.1 14 22.7 11.2 14.5 12.2 16.5 14.5 16.1 ...
## $ OBES: num [1:50] 36.2 29.5 29.5 37.1 25.8 23 27.4 33.5 30.7 32.5 ...
## $ CNCS: num [1:50] 71813 2619 150609 35246 413576 ...
## $ MRAT: num [1:50] 1.76 0.42 1.62 1.03 1.72 ...
## - attr(*, "na.action")= 'omit' Named int [1:4] 9 12 37 42
## ..- attr(*, "names")= chr [1:4] "9" "12" "37" "42"
```

#Distribution of the dependent variable

```
y_1= Covid.1$CNCS
par(mfrow=c(1,2))
h=hist(y_1, col = "pink",main = "distribution of confirm cases", xlab = "CNCS", ylab="frequency")
xfit=seq(min(y_1), max(y_1),length=10)
yfit=dnorm(xfit,mean = mean(y_1),sd = sd(y_1))
yfit=yfit*diff(h$mids[1:2])*length(y_1)
lines(xfit,yfit,col="blue",lwd=2)
```

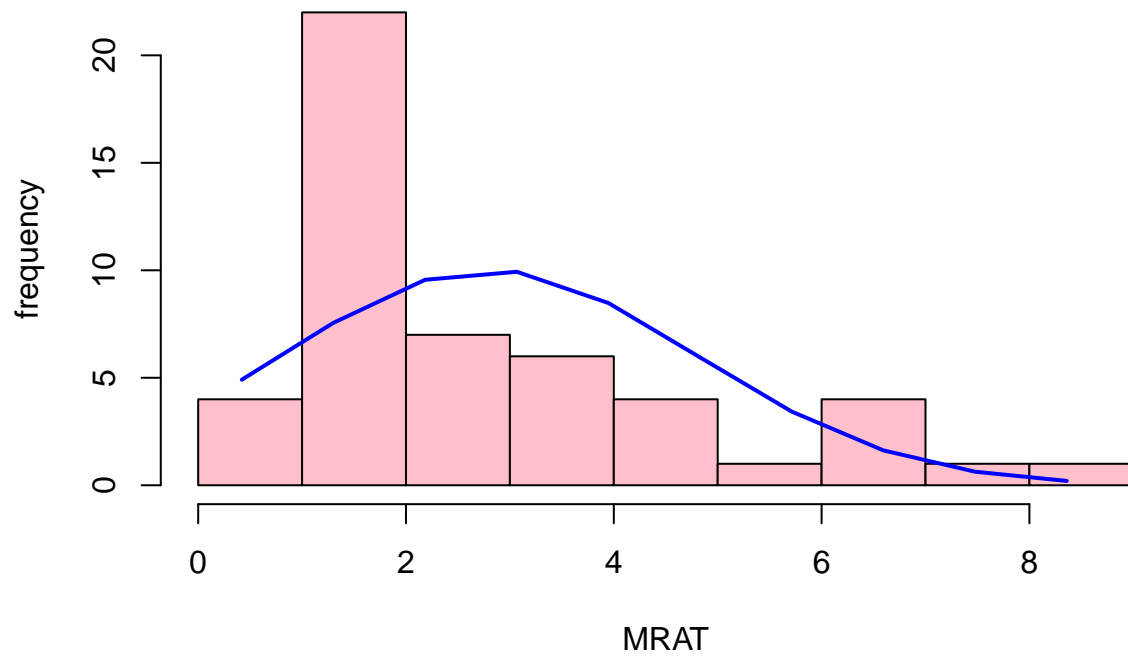
distribution of confirm cases



#Distribution of the dependent variable

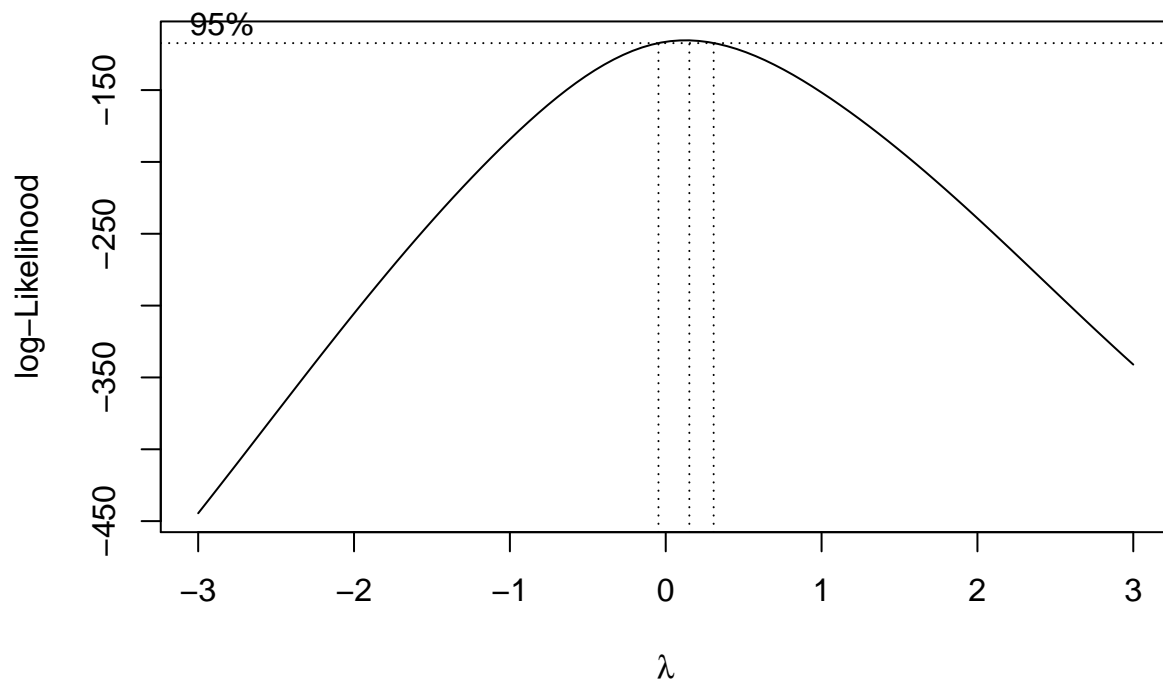
```
y_2= Covid.1$MRAT
h=hist(y_2, col = "pink",main = "Mortality rate", xlab = "MRAT", ylab="frequency")
xfit=seq(min(y_2), max(y_2),length=10)
yfit=dnorm(xfit,mean = mean(y_2),sd = sd(y_2))
yfit=yfit*diff(h$mids[1:2])*length(y_2)
lines(xfit,yfit,col="blue",lwd=2)
```

Mortality rate



will apply `box_cox` transformation to normalize *CNCS* and *MRAT* response

```
set.seed(8)
#par(mfrow=c(2,1))
#box cox transformation
bc=boxcox(Covid.1$CNCS~1,lambda = seq(-3,3))
```

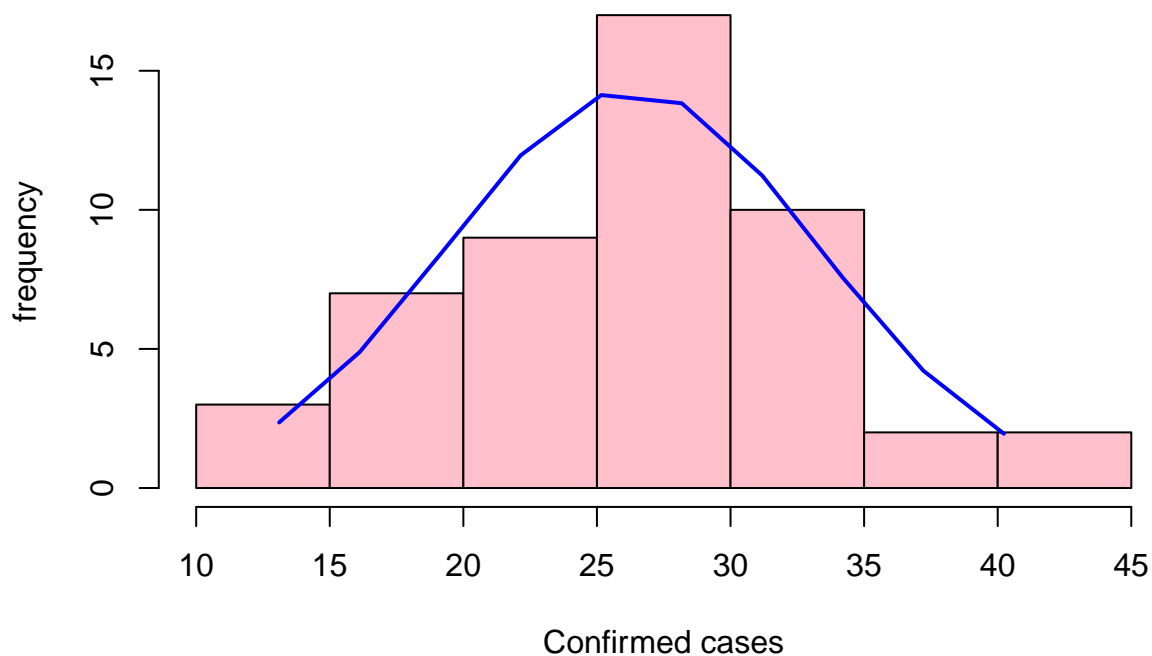


```
cox=data.frame(bc$x,bc$y) #create a data frame with the result
cox2 = cox[with(cox, order(-cox$bc.y)),] # Order the new data frame by decreasing y
cox2[1,] # Display the lambda with the greatest log likelihood
```

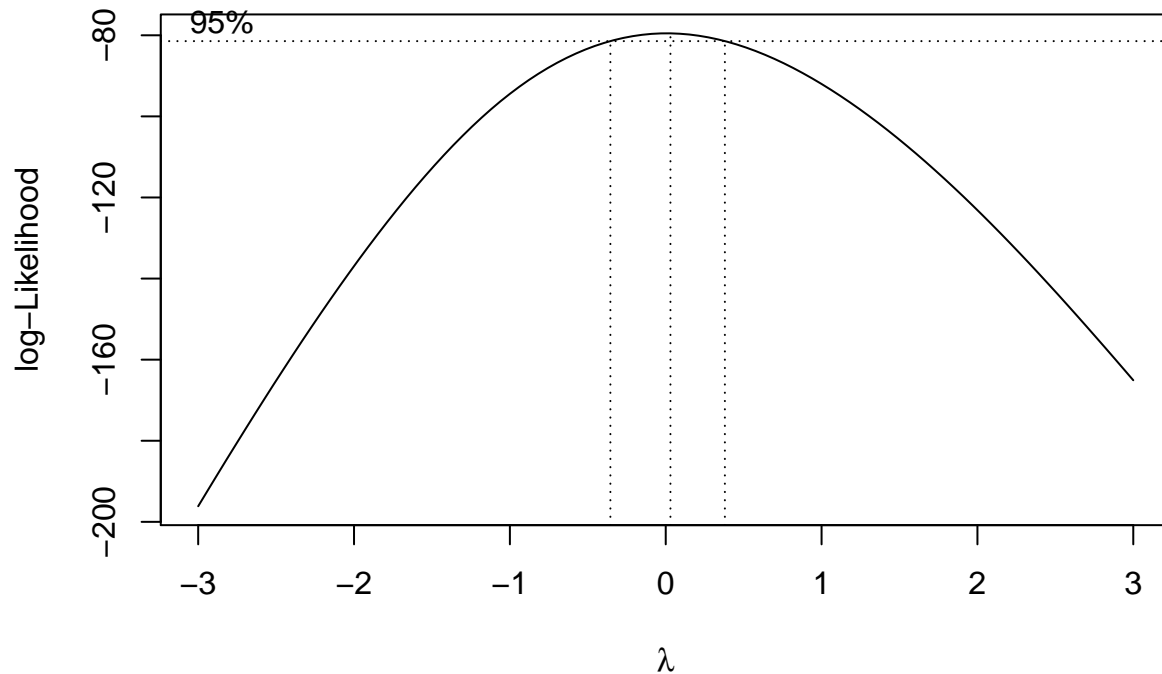
```
##          bc.x      bc.y
## 53 0.1515152 -115.3866
```

```
best.lam=bc$x[which(bc$y==max(bc$y))] #extract the lamnda
lambda = cox2[1, "bc.x"]
T_box = (Covid.1$CNCS ^ lambda - 1)/lambda
h=hist(T_box, col = "pink",main = "distribution of confirmed cases", xlab = "Confirmed cases", ylab="frequency")
xfit=seq(min(T_box), max(T_box),length=10)
yfit=dnorm(xfit,mean = mean(T_box),sd = sd(T_box))
yfit=yfit*diff(h$mids[1:2])*length(T_box)
lines(xfit,yfit,col="blue",lwd=2)
```

distribution of confirmed cases



```
set.seed(8)
#par(mfrow=c(2,1))
#box cox transformation
Box=boxcox(Covid.1$MRAT~1,lambda = seq(-3,3))
```

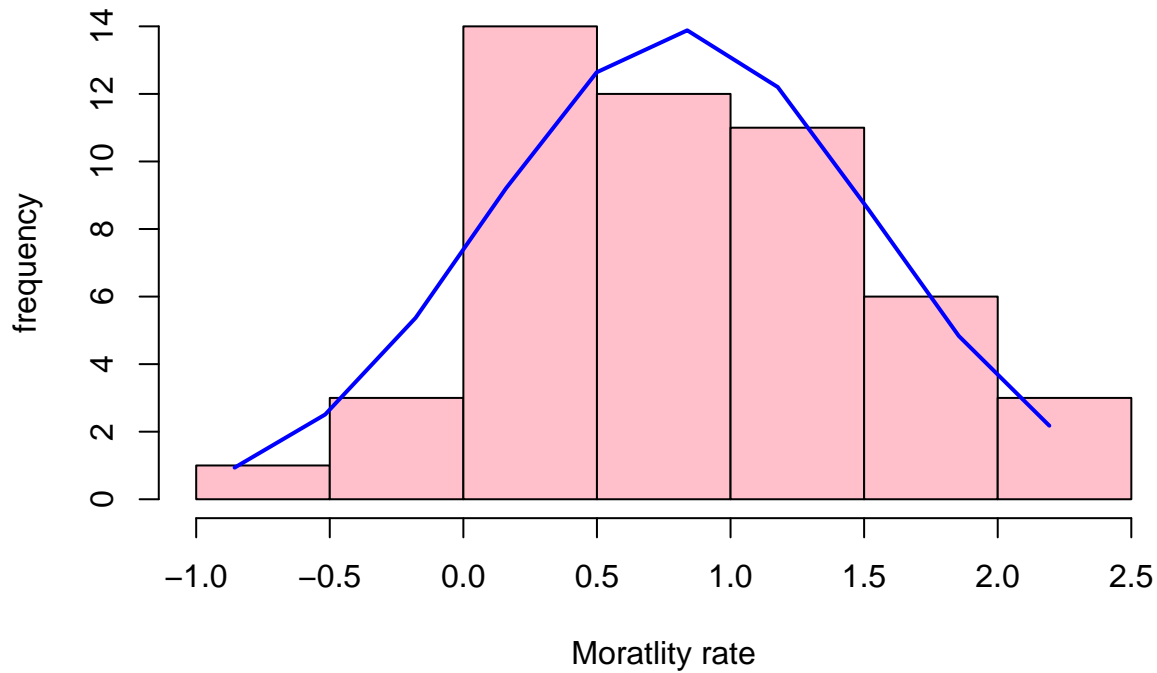


```
cox.1=data.frame(Box$x,Box$y) #create a data frame with the result
cox2.1 = cox.1[with(cox.1, order(-cox.1$Box.y)),] # Order the new data frame by decreasing y
cox2.1[1,] # Display the lambda with the greatest log likelihood
```

```
##          Box.x      Box.y
## 51 0.03030303 -79.51957
```

```
best.lam.1=bc$x[which(bc$y==max(bc$y))] #extract the lamnda
lambda.1 = cox2.1[1, "Box.x"]
M_box = (Covid.1$MRAT ^ lambda.1 - 1)/lambda.1
h=hist(M_box, col = "pink",main = "distribution of mortality rate", xlab = "Moratlity rate", ylab="frequency")
xfit=seq(min(M_box), max(M_box),length=10)
yfit=dnorm(xfit,mean = mean(M_box),sd = sd(M_box))
yfit=yfit*diff(h$mids[1:2])*length(M_box)
lines(xfit,yfit,col="blue",lwd=2)
```

distribution of mortality rate



```
CNCS=T_box
Covid.1$CNCS=CNCS
```

```
MRAT=M_box
Covid.1$MRAT=MRAT
```

Multicollinearity negative correlations are in blue color and positive correlations in red.

#Pearson Correlation visualization: the pearson correlation shows strong correlation with some of the

```
cormat <- round(cor(Covid.1),2)
head(cormat)
```

```
##      HOSC  ICU  PLUF  SINC  POPD  POPS  HOML  HUMI  UNEM  MEDA  LEXP  ADEP  ATEM
## HOSC 1.00  0.78  0.22  0.78  0.05  0.85  0.55 -0.05  0.11  0.06 -0.01 -0.02  0.53
## ICU  0.78  1.00  0.13  0.72 -0.03  0.76  0.54  0.04  0.12  0.06  0.06 -0.09  0.29
## PLUF 0.22  0.13  1.00  0.03 -0.25  0.10  0.01 -0.22 -0.18  0.60 -0.81  0.32  0.51
## SINC 0.78  0.72  0.03  1.00  0.22  0.98  0.89 -0.07  0.41  0.16  0.21 -0.29  0.28
## POPD 0.05 -0.03 -0.25  0.22  1.00  0.17  0.13 -0.50  0.55  0.12  0.29 -0.43  0.11
## POPS 0.85  0.76  0.10  0.98  0.17  1.00  0.82 -0.09  0.34  0.14  0.13 -0.22  0.36
##      APRE  CIGA  OBES  CNCS  MRAT
## HOSC 0.18  0.00  0.03  0.64 -0.20
## ICU  0.03  0.04  0.01  0.47 -0.14
## PLUF 0.21  0.45  0.69  0.21 -0.18
## SINC 0.12 -0.19 -0.20  0.76  0.06
## POPD 0.38 -0.45 -0.35  0.35  0.58
## POPS 0.16 -0.10 -0.12  0.79  0.01
```

```
melted_cormat <- melt(cormat) #the reshape library is used to melt the correlation matrix
head(melted_cormat)
```

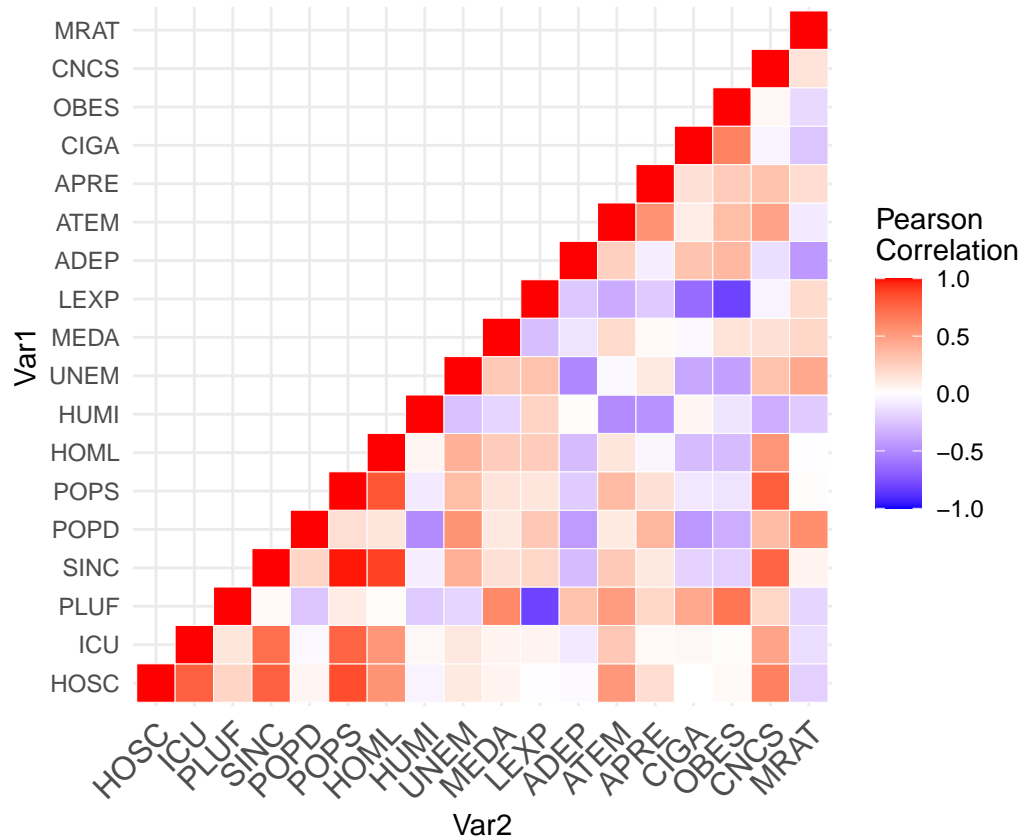
```
##   Var1 Var2 value
## 1 HOSC HOSC  1.00
## 2 ICU  HOSC  0.78
## 3 PLUF HOSC  0.22
## 4 SINC HOSC  0.78
## 5 POPD HOSC  0.05
## 6 POPS HOSC  0.85
```

```
# Get lower triangle of the correlation matrix
get_lower_tri<-function(cormat){
  cormat[upper.tri(cormat)] <- NA
  return(cormat)
}
# Get upper triangle of the correlation matrix
get_upper_tri <- function(cormat){
  cormat[lower.tri(cormat)]<- NA
  return(cormat)
}
upper_tri = get_upper_tri(cormat)
upper_tri
```

```
##      HOSC  ICU  PLUF  SINC  POPD  POPS  HOML  HUMI  UNEM  MEDA  LEXP  ADEP  ATEM
## HOSC      1 0.78 0.22 0.78  0.05 0.85 0.55 -0.05  0.11  0.06 -0.01 -0.02  0.53
## ICU      NA 1.00 0.13 0.72 -0.03 0.76 0.54  0.04  0.12  0.06  0.06 -0.09  0.29
## PLUF      NA  NA  1.00 0.03 -0.25 0.10 0.01 -0.22 -0.18  0.60 -0.81  0.32  0.51
## SINC      NA  NA  NA  1.00  0.22 0.98 0.89 -0.07  0.41  0.16  0.21 -0.29  0.28
## POPD      NA  NA  NA  NA  1.00 0.17 0.13 -0.50  0.55  0.12  0.29 -0.43  0.11
## POPS      NA  NA  NA  NA  NA  1.00 0.82 -0.09  0.34  0.14  0.13 -0.22  0.36
## HOML      NA  NA  NA  NA  NA  NA  1.00  0.05  0.41  0.27  0.27 -0.29  0.13
## HUMI      NA  NA  NA  NA  NA  NA  NA  1.00 -0.27 -0.18  0.23  0.01 -0.50
## UNEM      NA  NA  NA  NA  NA  NA  NA  NA  1.00  0.28  0.32 -0.52 -0.03
## MEDA      NA  NA  NA  NA  NA  NA  NA  NA  NA  1.00 -0.28 -0.12  0.19
## LEXP      NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  1.00 -0.24 -0.36
## ADEP      NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  1.00  0.25
## ATEM      NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  1.00
## APRE      NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA
## CIGA      NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA
## OBES      NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA
## CNCS      NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA
## MRAT      NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA
##      APRE  CIGA  OBES  CNCS  MRAT
## HOSC  0.18  0.00  0.03  0.64 -0.20
## ICU   0.03  0.04  0.01  0.47 -0.14
## PLUF  0.21  0.45  0.69  0.21 -0.18
## SINC  0.12 -0.19 -0.20  0.76  0.06
## POPD  0.38 -0.45 -0.35  0.35  0.58
## POPS  0.16 -0.10 -0.12  0.79  0.01
## HOML -0.04 -0.29 -0.29  0.54 -0.01
## HUMI -0.47  0.05 -0.11 -0.35 -0.22
## UNEM  0.11 -0.38 -0.41  0.32  0.45
```

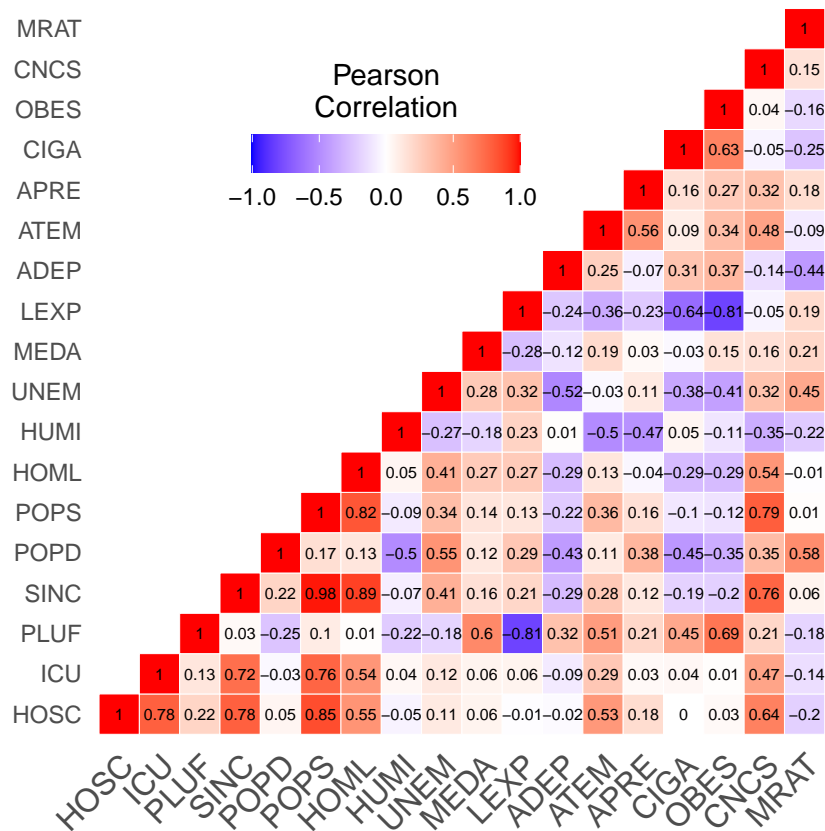
```
## MEDA 0.03 -0.03 0.15 0.16 0.21
## LEXP -0.23 -0.64 -0.81 -0.05 0.19
## ADEP -0.07 0.31 0.37 -0.14 -0.44
## ATEM 0.56 0.09 0.34 0.48 -0.09
## APRE 1.00 0.16 0.27 0.32 0.18
## CIGA NA 1.00 0.63 -0.05 -0.25
## OBES NA NA 1.00 0.04 -0.16
## CNCS NA NA NA 1.00 0.15
## MRAT NA NA NA NA 1.00
```

```
melted_cormat <- melt(upper_tri, na.rm = TRUE)
# Create a ggheatmap
ggheatmap <- ggplot(melted_cormat, aes(Var2, Var1, fill = value))+
  geom_tile(color = "white")+
  scale_fill_gradient2(low = "blue", high = "red", mid = "white",
    midpoint = 0, limit = c(-1,1), space = "Lab",
    name="Pearson\nCorrelation") +
  theme_minimal()+ # minimal theme
  theme(axis.text.x = element_text(angle = 45, vjust = 1,
    size = 12, hjust = 1))+
  coord_fixed()
# Print the heatmap
print(ggheatmap)
```



```
ggheatmap +
  geom_text(aes(label = value), color = "black", size = 2) +
```

```
theme(
  axis.title.x = element_blank(),
  axis.title.y = element_blank(),
  panel.grid.major = element_blank(),
  panel.border = element_blank(),
  panel.background = element_blank(),
  axis.ticks = element_blank(),
  legend.justification = c(1, 0),
  legend.position = c(0.6, 0.7),
  legend.direction = "horizontal")+
  guides(fill = guide_colorbar(barwidth = 7, barheight = 1,
    title.position = "top", title.hjust = 0.5))
```



```
cormat <- round(cor(Covid.1),2)
upper.tri(cormat)
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
## [1,] FALSE TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [2,] FALSE FALSE TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [3,] FALSE FALSE FALSE TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [4,] FALSE FALSE FALSE FALSE TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [5,] FALSE FALSE FALSE FALSE FALSE TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [6,] FALSE FALSE FALSE FALSE FALSE FALSE TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [7,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE  TRUE  TRUE  TRUE  TRUE
## [8,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE  TRUE  TRUE  TRUE
## [9,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE  TRUE  TRUE
## [10,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE  TRUE  TRUE
## [11,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE  TRUE
## [12,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE
```

```

## [10,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE
## [11,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE
## [12,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [13,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [14,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [15,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [16,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [17,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [18,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##      [,13] [,14] [,15] [,16] [,17] [,18]
## [1,]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [2,]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [3,]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [4,]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [5,]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [6,]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [7,]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [8,]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [9,]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [10,] TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [11,] TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [12,] TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [13,] FALSE TRUE  TRUE  TRUE  TRUE  TRUE
## [14,] FALSE FALSE TRUE  TRUE  TRUE  TRUE
## [15,] FALSE FALSE FALSE TRUE  TRUE  TRUE
## [16,] FALSE FALSE FALSE FALSE TRUE  TRUE
## [17,] FALSE FALSE FALSE FALSE FALSE TRUE
## [18,] FALSE FALSE FALSE FALSE FALSE FALSE

```

```

upper<-cormat
upper[upper.tri(cormat)]<-""
upper<-as.data.frame(upper)
upper

```

```

##      HOSC  ICU  PLUF  SINC  POPD  POPS  HOML  HUMI  UNEM  MEDA  LEXP  ADEP
## HOSC      1
## ICU    0.78      1
## PLUF   0.22  0.13      1
## SINC   0.78  0.72  0.03      1
## POPD   0.05 -0.03 -0.25  0.22      1
## POPS   0.85  0.76  0.1  0.98  0.17      1
## HOML   0.55  0.54  0.01  0.89  0.13  0.82      1
## HUMI  -0.05  0.04 -0.22 -0.07 -0.5 -0.09  0.05      1
## UNEM   0.11  0.12 -0.18  0.41  0.55  0.34  0.41 -0.27      1
## MEDA   0.06  0.06  0.6  0.16  0.12  0.14  0.27 -0.18  0.28      1
## LEXP  -0.01  0.06 -0.81  0.21  0.29  0.13  0.27  0.23  0.32 -0.28      1
## ADEP  -0.02 -0.09  0.32 -0.29 -0.43 -0.22 -0.29  0.01 -0.52 -0.12 -0.24      1
## ATEM   0.53  0.29  0.51  0.28  0.11  0.36  0.13 -0.5 -0.03  0.19 -0.36  0.25
## APRE   0.18  0.03  0.21  0.12  0.38  0.16 -0.04 -0.47  0.11  0.03 -0.23 -0.07
## CIGA    0  0.04  0.45 -0.19 -0.45 -0.1 -0.29  0.05 -0.38 -0.03 -0.64  0.31
## OBES   0.03  0.01  0.69 -0.2 -0.35 -0.12 -0.29 -0.11 -0.41  0.15 -0.81  0.37
## CNCS   0.64  0.47  0.21  0.76  0.35  0.79  0.54 -0.35  0.32  0.16 -0.05 -0.14
## MRAT  -0.2 -0.14 -0.18  0.06  0.58  0.01 -0.01 -0.22  0.45  0.21  0.19 -0.44
##      ATEM APRE  CIGA  OBES CNCS MRAT

```

```
## HOSC
## ICU
## PLUF
## SINC
## POPD
## POPS
## HOML
## HUMI
## UNEM
## MEDA
## LEXP
## ADEP
## ATEM      1
## APRE  0.56      1
## CIGA  0.09 0.16      1
## OBES  0.34 0.27  0.63      1
## CNCS  0.48 0.32 -0.05  0.04      1
## MRAT -0.09 0.18 -0.25 -0.16 0.15      1
```

```
lower<-cormat
lower[lower.tri(cormat, diag=TRUE)]<-""
lower<-as.data.frame(lower)
lower
```

```
##      HOSC  ICU PLUF SINC  POPD POPS HOML  HUMI  UNEM  MEDA  LEXP  ADEP  ATEM
## HOSC      0.78 0.22 0.78  0.05 0.85 0.55 -0.05  0.11  0.06 -0.01 -0.02  0.53
## ICU          0.13 0.72 -0.03 0.76 0.54  0.04  0.12  0.06  0.06 -0.09  0.29
## PLUF              0.03 -0.25  0.1 0.01 -0.22 -0.18  0.6 -0.81  0.32  0.51
## SINC                  0.22 0.98 0.89 -0.07  0.41  0.16  0.21 -0.29  0.28
## POPD                        0.17 0.13 -0.5  0.55  0.12  0.29 -0.43  0.11
## POPS                              0.82 -0.09 0.34  0.14  0.13 -0.22  0.36
## HOML                                    0.05  0.41  0.27  0.27 -0.29  0.13
## HUMI                                          -0.27 -0.18  0.23  0.01 -0.5
## UNEM                                              0.28  0.32 -0.52 -0.03
## MEDA                                                  -0.28 -0.12  0.19
## LEXP                                                         -0.24 -0.36
## ADEP                                                                0.25
## ATEM
## APRE
## CIGA
## OBES
## CNCS
## MRAT
##      APRE  CIGA  OBES  CNCS  MRAT
## HOSC  0.18      0  0.03  0.64  -0.2
## ICU   0.03  0.04  0.01  0.47 -0.14
## PLUF  0.21  0.45  0.69  0.21 -0.18
## SINC  0.12 -0.19 -0.2  0.76  0.06
## POPD  0.38 -0.45 -0.35  0.35  0.58
## POPS  0.16 -0.1 -0.12  0.79  0.01
## HOML -0.04 -0.29 -0.29  0.54 -0.01
## HUMI -0.47  0.05 -0.11 -0.35 -0.22
## UNEM  0.11 -0.38 -0.41  0.32  0.45
## MEDA  0.03 -0.03  0.15  0.16  0.21
```

```
## LEXP -0.23 -0.64 -0.81 -0.05 0.19
## ADEP -0.07 0.31 0.37 -0.14 -0.44
## ATEM 0.56 0.09 0.34 0.48 -0.09
## APRE      0.16 0.27 0.32 0.18
## CIGA      0.63 -0.05 -0.25
## OBES      0.04 -0.16
## CNCS      0.15
## MRAT
```

```
library(xtable)
print(xtable(upper), type="html")
```

```
## <!-- html table generated in R 3.6.1 by xtable 1.8-4 package -->
## <!-- Fri Dec 4 09:46:33 2020 -->
## <table border=1>
## <tr> <th> </th> <th> HOSC </th> <th> ICU </th> <th> PLUF </th> <th> SINC </th> <th> POPD </th> <th>
## <tr> <td align="right"> HOSC </td> <td> 1 </td> <td> </td> <td> </td> <td> </td> <td> </td> <td>
## <tr> <td align="right"> ICU </td> <td> 0.78 </td> <td> 1 </td> <td> </td> <td> </td> <td> </td>
## <tr> <td align="right"> PLUF </td> <td> 0.22 </td> <td> 0.13 </td> <td> 1 </td> <td> </td> <td>
## <tr> <td align="right"> SINC </td> <td> 0.78 </td> <td> 0.72 </td> <td> 0.03 </td> <td> 1 </td> <td>
## <tr> <td align="right"> POPD </td> <td> 0.05 </td> <td> -0.03 </td> <td> -0.25 </td> <td> 0.22 </td>
## <tr> <td align="right"> POPS </td> <td> 0.85 </td> <td> 0.76 </td> <td> 0.1 </td> <td> 0.98 </td>
## <tr> <td align="right"> HOML </td> <td> 0.55 </td> <td> 0.54 </td> <td> 0.01 </td> <td> 0.89 </td>
## <tr> <td align="right"> HUMI </td> <td> -0.05 </td> <td> 0.04 </td> <td> -0.22 </td> <td> -0.07 </td>
## <tr> <td align="right"> UNEM </td> <td> 0.11 </td> <td> 0.12 </td> <td> -0.18 </td> <td> 0.41 </td>
## <tr> <td align="right"> MEDA </td> <td> 0.06 </td> <td> 0.06 </td> <td> 0.6 </td> <td> 0.16 </td>
## <tr> <td align="right"> LEXP </td> <td> -0.01 </td> <td> 0.06 </td> <td> -0.81 </td> <td> 0.21 </td>
## <tr> <td align="right"> ADEP </td> <td> -0.02 </td> <td> -0.09 </td> <td> 0.32 </td> <td> -0.29 </td>
## <tr> <td align="right"> ATEM </td> <td> 0.53 </td> <td> 0.29 </td> <td> 0.51 </td> <td> 0.28 </td>
## <tr> <td align="right"> APRE </td> <td> 0.18 </td> <td> 0.03 </td> <td> 0.21 </td> <td> 0.12 </td>
## <tr> <td align="right"> CIGA </td> <td> 0 </td> <td> 0.04 </td> <td> 0.45 </td> <td> -0.19 </td>
## <tr> <td align="right"> OBES </td> <td> 0.03 </td> <td> 0.01 </td> <td> 0.69 </td> <td> -0.2 </td>
## <tr> <td align="right"> CNCS </td> <td> 0.64 </td> <td> 0.47 </td> <td> 0.21 </td> <td> 0.76 </td>
## <tr> <td align="right"> MRAT </td> <td> -0.2 </td> <td> -0.14 </td> <td> -0.18 </td> <td> 0.06 </td>
## </table>
```

```
#Checking for Multicollinearity
M=lm(CNCS~., data = Covid.1)
vif(M)
```

```
##      HOSC      ICU      PLUF      SINC      POPD      POPS      HOML
## 10.700764  3.431623 10.749846 330.340090  3.777993 243.306640 24.022492
##      HUMI      UNEM      MEDA      LEXP      ADEP      ATEM      APRE
##  2.521634  2.235707  3.654130  8.438558  2.503346  4.098009  2.553535
##      CIGA      OBES      MRAT
##  2.687573  4.173875  2.563724
```

```
M.1=lm(CNCS~HOSC+ICU+PLUF+POPD+POPS+HOML+HUMI+UNEM+MEDA+LEXP+ADEP+ATEM+APRE+CIGA+OBES+MRAT , data = Covid.1)
vif(M.1)
```

```
##      HOSC      ICU      PLUF      POPD      POPS      HOML      HUMI      UNEM
##  9.676887  3.179175 10.195093  3.328914 15.716942  6.780408  2.506555  2.235533
##      MEDA      LEXP      ADEP      ATEM      APRE      CIGA      OBES      MRAT
##  3.652616  8.133886  2.486222  4.014495  2.525483  2.626393  4.144656  2.423495
```

```
M.2=lm(CNCS~HOSC+ICU+POPD+PLUF+HOML+HUMI+UNEM+MEDA+LEXP+ADEP+ATEM+APRE+CIGA+OBES+MRAT , data = Covid.1)
vif(M.2)
```

```
##      HOSC      ICU      POPD      PLUF      HOML      HUMI      UNEM      MEDA
## 4.448164 3.103431 3.303580 10.173497 2.349533 2.333318 2.225700 3.586526
##      LEXP      ADEP      ATEM      APRE      CIGA      OBES      MRAT
## 8.119858 2.485841 3.878572 2.482580 2.565382 4.138495 2.092709
```

```
M.3=lm(CNCS~HOSC+ICU+POPD+HOML+HOML+HUMI+UNEM+MEDA+LEXP+ADEP+ATEM+APRE+CIGA+OBES+MRAT , data = Covid.1)
vif(M.3)
```

```
##      HOSC      ICU      POPD      HOML      HUMI      UNEM      MEDA      LEXP
## 4.345368 3.090481 3.235879 2.335767 2.333180 2.225688 1.633163 4.866170
##      ADEP      ATEM      APRE      CIGA      OBES      MRAT
## 2.078795 3.832165 2.457136 2.555387 4.108408 2.092614
```

#Checking for Multicollinearity

```
M.M=lm(MRAT~., data = Covid.1)
vif(M)
```

```
##      HOSC      ICU      PLUF      SINC      POPD      POPS      HOML
## 10.700764 3.431623 10.749846 330.340090 3.777993 243.306640 24.022492
##      HUMI      UNEM      MEDA      LEXP      ADEP      ATEM      APRE
## 2.521634 2.235707 3.654130 8.438558 2.503346 4.098009 2.553535
##      CIGA      OBES      MRAT
## 2.687573 4.173875 2.563724
```

#Checking for Multicollinearity

```
M.M.1=lm(MRAT~HOSC+ICU+PLUF+POPD+POPS+HOML+HUMI+UNEM+MEDA+LEXP+ADEP+ATEM+APRE+CIGA+OBES+CNCS, data = Covid.1)
vif(M.M.1)
```

```
##      HOSC      ICU      PLUF      POPD      POPS      HOML      HUMI      UNEM
## 8.762553 3.444791 10.443484 3.157989 24.409632 6.552879 2.442612 2.197414
##      MEDA      LEXP      ADEP      ATEM      APRE      CIGA      OBES      CNCS
## 3.596364 7.877608 2.338937 4.073121 2.562379 2.627710 3.967740 4.986504
```

#Checking for Multicollinearity

```
M.M.2=lm(MRAT~HOSC+ICU+PLUF+POPD+HOML+HUMI+UNEM+MEDA+LEXP+ADEP+ATEM+APRE+CIGA+OBES+CNCS, data = Covid.1)
vif(M.M.2)
```

```
##      HOSC      ICU      PLUF      POPD      HOML      HUMI      UNEM      MEDA
## 4.764292 3.134645 10.249328 3.068346 2.721137 2.390437 2.177096 3.594820
##      LEXP      ADEP      ATEM      APRE      CIGA      OBES      CNCS
## 7.875344 2.328438 3.849205 2.479276 2.596572 3.960677 2.772488
```

#Checking for Multicollinearity

```
M.M.3=lm(MRAT~HOSC+ICU+POPD+HOML+HUMI+UNEM+MEDA+LEXP+ADEP+ATEM+APRE+CIGA+OBES+CNCS, data = Covid.1)
vif(M.M.3)
```

```
##      HOSC      ICU      POPD      HOML      HUMI      UNEM      MEDA      LEXP
## 4.696145 3.117848 2.974132 2.717161 2.389302 2.177028 1.596513 4.576902
##      ADEP      ATEM      APRE      CIGA      OBES      CNCS
## 1.911764 3.808672 2.451080 2.582431 3.940187 2.751850
```

- 3 variables dropped POPS,SINC,PLUF
- This is the same for the CNCS response

MODEL

```
#Separating the data into two datasets the train and testing datasets.
set.seed(333)
Covid.2=Covid.1[,-18]
Covid.3=Covid.1[,-17]
id = sample(nrow(Covid.2), nrow(Covid.2)*0.7)
train =Covid.2[id,]
test =Covid.2[-id,]
x=sample(nrow(Covid.3),nrow(Covid.3)*0.7)
train.1=Covid.3[x,]
test.1=Covid.3[-x,]

#Linear regression Model on the the confirmed cases
lm.fit1=lm(CNCS~HOSC+ICU+POPD+HOML+HUMI+UNEM+MEDA+LEXP+ADEP+ATEM+APRE+CIGA+OBES,train)
summary(lm.fit1)
```

```
##
## Call:
## lm(formula = CNCS ~ HOSC + ICU + POPD + HOML + HUMI + UNEM +
##      MEDA + LEXP + ADEP + ATEM + APRE + CIGA + OBES, data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.4505 -2.1008 -0.4406  1.2592  6.8516
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.218e+01  9.640e+01  -0.126  0.900654
## HOSC         3.783e-03  1.631e-03   2.319  0.030554 *
## ICU         -8.777e-03  5.177e-03  -1.695  0.104767
## POPD         1.182e-02  4.709e-03   2.511  0.020288 *
## HOML         2.418e-04  6.109e-05   3.958  0.000719 ***
## HUMI        -2.135e-01  5.310e-01  -0.402  0.691646
## UNEM         3.189e-01  3.529e-01   0.904  0.376360
## MEDA        -1.780e-01  1.916e-01  -0.929  0.363487
## LEXP         1.387e-01  9.657e-01   0.144  0.887162
## ADEP         2.541e-01  3.571e-01   0.712  0.484581
## ATEM         1.467e-01  1.675e-01   0.875  0.391279
## APRE        -1.705e-01  8.548e-02  -1.994  0.059291 .
## CIGA         5.379e-01  3.513e-01   1.531  0.140667
## OBES         3.640e-01  3.696e-01   0.985  0.335995
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 4.443 on 21 degrees of freedom
## Multiple R-squared: 0.76, Adjusted R-squared: 0.6114
## F-statistic: 5.115 on 13 and 21 DF, p-value: 0.0004908
```

*model accuracy for CNCS is 61%

*The summary of the lm.fit1 shows that the adjusted R-squared value of 0.6114. Here we could see that there are ten variables with $\Pr(>|t|) > 0.05$ which show that the variables have no significant effect toward the model.

#Linear regression Model on the the Mortality rate

```
lm.fit2=lm(MRAT~HOSC+ICU+POPD+HOML+HUMI+UNEM+MEDA+LEXP+ADEP+ATEM+APRE+CIGA+OBES,train.1)
summary(lm.fit2)
```

```
##
## Call:
## lm(formula = MRAT ~ HOSC + ICU + POPD + HOML + HUMI + UNEM +
##      MEDA + LEXP + ADEP + ATEM + APRE + CIGA + OBES, data = train.1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.89434 -0.20089 -0.06739  0.17169  1.12393
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.355e+01  1.323e+01  -1.780  0.08960 .
## HOSC         -2.151e-04  7.561e-05  -2.844  0.00971 **
## ICU          9.319e-05  2.440e-04   0.382  0.70637
## POPD         1.171e-03  6.718e-04   1.744  0.09584 .
## HOML        -6.254e-06  7.344e-06  -0.852  0.40406
## HUMI         4.573e-02  5.792e-02   0.790  0.43862
## UNEM         1.337e-02  3.987e-02   0.335  0.74061
## MEDA         3.709e-02  2.195e-02   1.690  0.10585
## LEXP         2.613e-01  1.682e-01   1.554  0.13522
## ADEP        -8.992e-02  3.435e-02  -2.618  0.01608 *
## ATEM         7.200e-02  2.576e-02   2.795  0.01086 *
## APRE        -3.015e-03  9.362e-03  -0.322  0.75061
## CIGA         6.042e-02  6.864e-02   0.880  0.38868
## OBES         1.675e-02  4.198e-02   0.399  0.69390
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4909 on 21 degrees of freedom
## Multiple R-squared: 0.7242, Adjusted R-squared: 0.5534
## F-statistic: 4.241 on 13 and 21 DF, p-value: 0.001676
```

*model accuracy for CNCS is 55%

*The summary of the lm.fit2 shows that the adjusted R-squared value of 0.5534. Here we could see that there are ten variables with $\Pr(>|t|) > 0.05$ which show that the variables have no significant effect toward the model.

For comparing, we would use step-wise regression with backward elimination method*

```
#stepwise regression both direction for lm.fit1
stats::step(lm.fit1, direction = "both")
```

```
## Start:  AIC=114.52
## CNCS ~ HOSC + ICU + POPD + HOML + HUMI + UNEM + MEDA + LEXP +
##      ADEP + ATEM + APRE + CIGA + OBES
##
##      Df Sum of Sq  RSS   AIC
## - LEXP  1      0.407 414.96 112.55
## - HUMI  1      3.192 417.75 112.78
## - ADEP  1      9.994 424.55 113.35
## - ATEM  1     15.127 429.68 113.77
## - UNEM  1     16.125 430.68 113.85
## - MEDA  1     17.034 431.59 113.92
## - OBES  1     19.140 433.70 114.09
## <none>                 414.56 114.52
## - CIGA  1     46.279 460.84 116.22
## - ICU   1     56.746 471.30 117.00
## - APRE  1     78.495 493.05 118.58
## - HOSC  1    106.170 520.73 120.50
## - POPD  1    124.472 539.03 121.70
## - HOML  1    309.225 723.78 132.02
##
## Step:  AIC=112.55
## CNCS ~ HOSC + ICU + POPD + HOML + HUMI + UNEM + MEDA + ADEP +
##      ATEM + APRE + CIGA + OBES
##
##      Df Sum of Sq  RSS   AIC
## - HUMI  1      2.92 417.89 110.80
## - ADEP  1     10.37 425.33 111.41
## - ATEM  1     15.04 430.00 111.80
## - UNEM  1     16.23 431.19 111.89
## - MEDA  1     21.86 436.82 112.35
## <none>                 414.96 112.55
## - OBES  1     27.01 441.97 112.76
## + LEXP  1      0.41 414.56 114.52
## - CIGA  1     51.23 466.20 114.62
## - ICU   1     60.01 474.98 115.28
## - APRE  1     78.32 493.29 116.60
## - HOSC  1    111.31 526.27 118.87
## - POPD  1    124.69 539.66 119.75
## - HOML  1    328.67 743.64 130.97
##
## Step:  AIC=110.8
## CNCS ~ HOSC + ICU + POPD + HOML + UNEM + MEDA + ADEP + ATEM +
##      APRE + CIGA + OBES
##
##      Df Sum of Sq  RSS   AIC
## - ADEP  1     14.83 432.72 110.02
## - UNEM  1     19.66 437.55 110.40
## - MEDA  1     20.75 438.64 110.49
## - ATEM  1     21.87 439.76 110.58
## <none>                 417.89 110.80
```

```

## - OBES 1      27.01 444.90 110.99
## + HUMI 1       2.92 414.96 112.55
## + LEXP 1       0.14 417.75 112.78
## - CIGA 1      53.28 471.17 113.00
## - ICU 1       68.99 486.88 114.14
## - APRE 1      79.03 496.92 114.86
## - HOSC 1     121.63 539.52 117.74
## - POPD 1     189.46 607.35 121.88
## - HOML 1     348.46 766.35 130.02
##
## Step: AIC=110.02
## CNCS ~ HOSC + ICU + POPD + HOML + UNEM + MEDA + ATEM + APRE +
##      CIGA + OBES
##
##      Df Sum of Sq    RSS    AIC
## - UNEM 1      12.03 444.75 108.98
## <none>          432.72 110.02
## - MEDA 1      28.52 461.23 110.25
## - OBES 1      33.82 466.54 110.65
## + ADEP 1      14.83 417.89 110.80
## + HUMI 1       7.39 425.33 111.41
## + LEXP 1       0.25 432.46 112.00
## - ICU 1      55.15 487.86 112.21
## - CIGA 1      58.81 491.52 112.48
## - ATEM 1      70.47 503.19 113.30
## - APRE 1     105.81 538.53 115.67
## - HOSC 1     107.77 540.49 115.80
## - POPD 1     180.60 613.31 120.22
## - HOML 1     349.78 782.49 128.75
##
## Step: AIC=108.98
## CNCS ~ HOSC + ICU + POPD + HOML + MEDA + ATEM + APRE + CIGA +
##      OBES
##
##      Df Sum of Sq    RSS    AIC
## - MEDA 1      22.40 467.15 108.70
## <none>          444.75 108.98
## - OBES 1      27.01 471.76 109.04
## + UNEM 1      12.03 432.72 110.02
## + HUMI 1       9.48 435.26 110.22
## + ADEP 1       7.20 437.55 110.40
## - ICU 1      50.58 495.33 110.75
## + LEXP 1       0.19 444.56 110.96
## - CIGA 1      54.19 498.94 111.00
## - ATEM 1      64.08 508.83 111.69
## - HOSC 1     103.12 547.87 114.27
## - APRE 1     103.25 548.00 114.28
## - POPD 1     268.34 713.09 123.50
## - HOML 1     400.11 844.86 129.43
##
## Step: AIC=108.7
## CNCS ~ HOSC + ICU + POPD + HOML + ATEM + APRE + CIGA + OBES
##
##      Df Sum of Sq    RSS    AIC

```

```
## - OBES 1 16.13 483.28 107.88
## <none> 467.15 108.70
## + MEDA 1 22.40 444.75 108.98
## + ADEP 1 13.79 453.36 109.65
## + HUMI 1 8.10 459.05 110.08
## + UNEM 1 5.92 461.23 110.25
## + LEXP 1 5.02 462.13 110.32
## - ICU 1 51.86 519.01 110.38
## - ATEM 1 51.92 519.07 110.38
## - CIGA 1 57.37 524.52 110.75
## - APRE 1 84.97 552.11 112.54
## - HOSC 1 112.37 579.52 114.24
## - POPD 1 246.53 713.68 121.53
## - HOML 1 383.79 850.94 127.69
##
## Step: AIC=107.88
## CNCS ~ HOSC + ICU + POPD + HOML + ATEM + APRE + CIGA
##
## Df Sum of Sq RSS AIC
## <none> 483.28 107.88
## + ADEP 1 17.93 465.35 108.56
## + OBES 1 16.13 467.15 108.70
## + MEDA 1 11.52 471.76 109.04
## + HUMI 1 8.87 474.41 109.23
## - ICU 1 52.85 536.13 109.52
## + UNEM 1 3.10 480.18 109.66
## + LEXP 1 1.25 482.03 109.79
## - ATEM 1 69.13 552.41 110.56
## - APRE 1 80.70 563.98 111.29
## - CIGA 1 103.53 586.81 112.68
## - HOSC 1 113.84 597.11 113.29
## - POPD 1 230.88 714.16 119.55
## - HOML 1 370.75 854.02 125.81
##
##
## Call:
## lm(formula = CNCS ~ HOSC + ICU + POPD + HOML + ATEM + APRE +
## CIGA, data = train)
##
## Coefficients:
## (Intercept) HOSC ICU POPD HOML ATEM
## 2.5647518 0.0033186 -0.0072846 0.0120294 0.0002311 0.2303244
## APRE CIGA
## -0.1584887 0.6604312
```

- 6 dropped and 7 left

```
stats::step(lm.fit2, direction = "both")
```

```
## Start: AIC=-39.68
## MRAT ~ HOSC + ICU + POPD + HOML + HUMI + UNEM + MEDA + LEXP +
## ADEP + ATEM + APRE + CIGA + OBES
##
```

```

##          Df Sum of Sq    RSS      AIC
## - APRE   1   0.02500  5.0864 -41.507
## - UNEM   1   0.02712  5.0885 -41.492
## - ICU    1   0.03515  5.0966 -41.437
## - OBES   1   0.03838  5.0998 -41.415
## - HUMI   1   0.15025  5.2117 -40.656
## - HOML   1   0.17478  5.2362 -40.491
## - CIGA   1   0.18676  5.2482 -40.411
## <none>                5.0614 -39.679
## - LEXP   1   0.58176  5.6432 -37.871
## - MEDA   1   0.68826  5.7497 -37.217
## - POPD   1   0.73279  5.7942 -36.947
## - ADEP   1   1.65160  6.7130 -31.795
## - ATEM   1   1.88252  6.9440 -30.612
## - HOSC   1   1.94984  7.0113 -30.274
##
## Step:  AIC=-41.51
## MRAT ~ HOSC + ICU + POPD + HOML + HUMI + UNEM + MEDA + LEXP +
##        ADEP + ATEM + CIGA + OBES
##
##          Df Sum of Sq    RSS      AIC
## - OBES   1   0.02611  5.1125 -43.328
## - UNEM   1   0.03659  5.1230 -43.256
## - ICU    1   0.04211  5.1285 -43.218
## - HUMI   1   0.15702  5.2434 -42.443
## - CIGA   1   0.17382  5.2602 -42.331
## - HOML   1   0.23949  5.3259 -41.897
## <none>                5.0864 -41.507
## + APRE   1   0.02500  5.0614 -39.679
## - LEXP   1   0.62534  5.7118 -39.449
## - POPD   1   0.79364  5.8801 -38.432
## - MEDA   1   0.84388  5.9303 -38.135
## - ADEP   1   1.62774  6.7142 -33.789
## - ATEM   1   1.85753  6.9440 -32.612
## - HOSC   1   2.01674  7.1032 -31.818
##
## Step:  AIC=-43.33
## MRAT ~ HOSC + ICU + POPD + HOML + HUMI + UNEM + MEDA + LEXP +
##        ADEP + ATEM + CIGA
##
##          Df Sum of Sq    RSS      AIC
## - UNEM   1   0.02994  5.1425 -45.124
## - ICU    1   0.04746  5.1600 -45.004
## - HUMI   1   0.15444  5.2670 -44.286
## - HOML   1   0.24016  5.3527 -43.721
## - CIGA   1   0.27291  5.3854 -43.508
## <none>                5.1125 -43.328
## + OBES   1   0.02611  5.0864 -41.507
## + APRE   1   0.01273  5.0998 -41.415
## - LEXP   1   0.62046  5.7330 -41.319
## - POPD   1   0.76951  5.8820 -40.421
## - MEDA   1   0.86184  5.9744 -39.875
## - ADEP   1   1.65513  6.7677 -35.512
## - HOSC   1   2.27115  7.3837 -32.463

```

```

## - ATEM 1 2.36915 7.4817 -32.001
##
## Step: AIC=-45.12
## MRAT ~ HOSC + ICU + POPD + HOML + HUMI + MEDA + LEXP + ADEP +
## ATEM + CIGA
##
##      Df Sum of Sq    RSS    AIC
## - ICU 1 0.05210 5.1946 -46.771
## - HUMI 1 0.15748 5.2999 -46.068
## - HOML 1 0.21088 5.3534 -45.717
## - CIGA 1 0.25200 5.3945 -45.449
## <none>      5.1425 -45.124
## + UNEM 1 0.02994 5.1125 -43.328
## - LEXP 1 0.59682 5.7393 -43.280
## + APRE 1 0.02016 5.1223 -43.261
## + OBES 1 0.01946 5.1230 -43.256
## - MEDA 1 0.93301 6.0755 -41.288
## - POPD 1 1.06526 6.2077 -40.534
## - ADEP 1 1.76362 6.9061 -36.803
## - HOSC 1 2.24402 7.3865 -34.449
## - ATEM 1 2.35661 7.4991 -33.920
##
## Step: AIC=-46.77
## MRAT ~ HOSC + POPD + HOML + HUMI + MEDA + LEXP + ADEP + ATEM +
## CIGA
##
##      Df Sum of Sq    RSS    AIC
## - HUMI 1 0.12030 5.3149 -47.969
## - HOML 1 0.24373 5.4383 -47.166
## <none>      5.1946 -46.771
## - CIGA 1 0.40031 5.5949 -46.172
## + ICU 1 0.05210 5.1425 -45.124
## + UNEM 1 0.03458 5.1600 -45.004
## + APRE 1 0.02641 5.1682 -44.949
## + OBES 1 0.02387 5.1707 -44.932
## - LEXP 1 0.80282 5.9974 -43.741
## - POPD 1 1.01795 6.2125 -42.507
## - MEDA 1 1.03348 6.2280 -42.420
## - ADEP 1 2.01300 7.2076 -37.308
## - ATEM 1 2.40189 7.5965 -35.468
## - HOSC 1 2.67346 7.8680 -34.239
##
## Step: AIC=-47.97
## MRAT ~ HOSC + POPD + HOML + MEDA + LEXP + ADEP + ATEM + CIGA
##
##      Df Sum of Sq    RSS    AIC
## <none>      5.3149 -47.969
## - HOML 1 0.33084 5.6457 -47.856
## + HUMI 1 0.12030 5.1946 -46.771
## - CIGA 1 0.57607 5.8909 -46.368
## + UNEM 1 0.03531 5.2796 -46.203
## + APRE 1 0.03025 5.2846 -46.169
## + OBES 1 0.01938 5.2955 -46.097
## + ICU 1 0.01492 5.2999 -46.068

```

```
## - POPD 1 0.98663 6.3015 -44.010
## - MEDA 1 1.12789 6.4428 -43.234
## - LEXP 1 1.12845 6.4433 -43.231
## - ADEP 1 2.11293 7.4278 -38.254
## - ATEM 1 2.29595 7.6108 -37.402
## - HOSC 1 2.59167 7.9065 -36.068
```

```
##
## Call:
## lm(formula = MRAT ~ HOSC + POPD + HOML + MEDA + LEXP + ADEP +
##     ATEM + CIGA, data = train.1)
##
## Coefficients:
## (Intercept)      HOSC      POPD      HOML      MEDA      LEXP
## -2.410e+01  -1.994e-04   8.410e-04  -7.584e-06   4.385e-02   3.208e-01
##      ADEP      ATEM      CIGA
## -9.687e-02   7.079e-02   8.542e-02
```

- 5 dropped and 8 left

```
Model1=lm(formula = CNCS ~ HOSC + ICU + POPD + HOML + ATEM + APRE + CIGA, data = train)
summary(Model1)
```

```
##
## Call:
## lm(formula = CNCS ~ HOSC + ICU + POPD + HOML + ATEM + APRE +
##     CIGA, data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.3508 -2.3831 -0.1628  2.2221  8.2031
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.565e+00  7.050e+00   0.364  0.718850
## HOSC         3.319e-03  1.316e-03   2.522  0.017879 *
## ICU         -7.285e-03  4.239e-03  -1.718  0.097175 .
## POPD         1.203e-02  3.349e-03   3.592  0.001290 **
## HOML         2.311e-04  5.078e-05   4.551  0.000102 ***
## ATEM         2.303e-01  1.172e-01   1.965  0.059750 .
## APRE        -1.585e-01  7.464e-02  -2.123  0.043034 *
## CIGA         6.604e-01  2.746e-01   2.405  0.023298 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.231 on 27 degrees of freedom
## Multiple R-squared:  0.7202, Adjusted R-squared:  0.6477
## F-statistic: 9.928 on 7 and 27 DF,  p-value: 4.374e-06
```

```
confint(Model1)
```

```
##              2.5 %          97.5 %
```

```
## (Intercept) -1.190093e+01 17.0304331877
## HOSC        6.185357e-04 0.0060186581
## ICU         -1.598283e-02 0.0014135579
## POPD        5.157030e-03 0.0189017952
## HOML        1.269251e-04 0.0003353244
## ATEM        -1.014667e-02 0.4707954589
## APRE        -3.116371e-01 -0.0053402785
## CIGA        9.697759e-02 1.2238847824
```

- The stepwise regression throw away the 6 variables that does not have any significant effect on the model. To check if this variables have significant effect on the model, we will run an annova test.

```
Model2=lm(formula = MRAT ~ HOSC + POPD + HOML + MEDA + LEXP + ADEP + ATEM + CIGA, data = train.1)
summary(Model2)
```

```
##
## Call:
## lm(formula = MRAT ~ HOSC + POPD + HOML + MEDA + LEXP + ADEP +
##     ATEM + CIGA, data = train.1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.87796 -0.23464 -0.05722  0.18138  1.17415
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.410e+01  1.148e+01  -2.100  0.04556 *
## HOSC         -1.994e-04  5.599e-05  -3.561  0.00145 **
## POPD         8.410e-04  3.828e-04   2.197  0.03714 *
## HOML        -7.584e-06  5.962e-06  -1.272  0.21457
## MEDA         4.385e-02  1.867e-02   2.349  0.02670 *
## LEXP         3.208e-01  1.365e-01   2.350  0.02667 *
## ADEP        -9.687e-02  3.013e-02  -3.215  0.00347 **
## ATEM         7.079e-02  2.112e-02   3.351  0.00247 **
## CIGA         8.542e-02  5.089e-02   1.679  0.10519
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4521 on 26 degrees of freedom
## Multiple R-squared:  0.7104, Adjusted R-squared:  0.6213
## F-statistic: 7.971 on 8 and 26 DF,  p-value: 2.213e-05
```

```
confint(Model2)
```

```
##              2.5 %      97.5 %
## (Intercept) -4.769501e+01 -5.124365e-01
## HOSC        -3.144338e-04 -8.426768e-05
## POPD         5.413350e-05  1.627962e-03
## HOML        -1.983909e-05  4.670136e-06
## MEDA         5.477767e-03  8.222555e-02
## LEXP         4.013710e-02  6.013847e-01
## ADEP        -1.588040e-01 -3.493577e-02
## ATEM         2.737018e-02  1.142028e-01
## CIGA        -1.917443e-02  1.900241e-01
```

*model accuracy for MRAT is 62%

*The summary of the `lm.fit2` shows that the adjusted R-squared value of 0.6213. Here we could see that there are ten variables with $\Pr(>|t|) > 0.05$ which show that the variables have no significant effect toward the model.

```
#check if the two models are the same or differennt.
anova(lm.fit1,Model1)
```

```
## Analysis of Variance Table
##
## Model 1: CNCS ~ HOSC + ICU + POPD + HOML + HUMI + UNEM + MEDA + LEXP +
##      ADEP + ATEM + APRE + CIGA + OBES
## Model 2: CNCS ~ HOSC + ICU + POPD + HOML + ATEM + APRE + CIGA
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1      21 414.56
## 2      27 483.28 -6   -68.722 0.5802  0.742
```

- F-value show that there is enough statistical evident that the models are the same, so we reject the null hypothesis and conclude that there is enough evidence that the two models are the same. Hence Model1 is preferred because of parsimony.

```
# calculate the rmse and r2 of each model
lm.fit1_rmse <- RMSE(lm.fit1$fitted.values, obs = train$CNCS)
Model1_rmse <- RMSE(Model1$fitted.values, obs = train$CNCS)
model_rmse_table <- data.frame(lm.fit1_rmse, Model1_rmse)
model_rmse_table
```

```
##   lm.fit1_rmse Model1_rmse
## 1      3.44158      3.715906
```

```
# calculate the rsq value of trained model
lm.fit1_rsqr <- cor(lm.fit1$fitted.values, train$CNCS)^2
Model1_rsqr <- cor(Model1$fitted.values, train$CNCS)^2
model_rsqr_table <- data.frame(lm.fit1_rsqr, Model1_rsqr )
model_rsqr_table
```

```
##   lm.fit1_rsqr Model1_rsqr
## 1    0.7599845  0.7201966
```

```
#check if the two models are the same or differennt for the MRAT.
anova(lm.fit2,Model2)
```

```
## Analysis of Variance Table
##
## Model 1: MRAT ~ HOSC + ICU + POPD + HOML + HUMI + UNEM + MEDA + LEXP +
##      ADEP + ATEM + APRE + CIGA + OBES
## Model 2: MRAT ~ HOSC + POPD + HOML + MEDA + LEXP + ADEP + ATEM + CIGA
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1      21 5.0614
## 2      26 5.3149 -5   -0.25344 0.2103  0.9544
```

- F-value show that there is enough statistical evident that the models are the same, so we reject the null hypothesis and conclude that there is enough evidence that the two models are the same. Hence Model2 is preferred because of parsimony.

```
#see how well the model predict the variable for CNCS
Model1.test.pred = predict(Model1, newdata=test)
Model1.test.rsq = cor(Model1.test.pred, test$CNCS)^2
Model1.test.rsq
```

```
## [1] 0.6146801
```

```
Model1.test.rmse=RMSE(Model1.test.pred, obs = test$CNCS)
Model1.test.rmse
```

```
## [1] 10.87199
```

```
test_rsqs_table =data.frame(Model1.test.rsq, Model1.test.rmse)
test_rsqs_table
```

```
##      Model1.test.rsq Model1.test.rmse
## 1      0.6146801      10.87199
```

```
#see how well the model predict the variable for MRAT
Model2_rsqs = cor(Model2$fitted.values, train.1$MRAT)^2
Model2_rsqs
```

```
## [1] 0.7103718
```

```
Model2.test.1.pred = predict(Model2, newdata=test.1)
Model2.test.1_rsqs = cor(Model2.test.1.pred, test.1$MRAT)^2
Model2.test.1_rsqs
```

```
## [1] 0.008462034
```

```
test.1_rsqs_table =data.frame(Model2_rsqs, Model2.test.1_rsqs )
test.1_rsqs_table
```

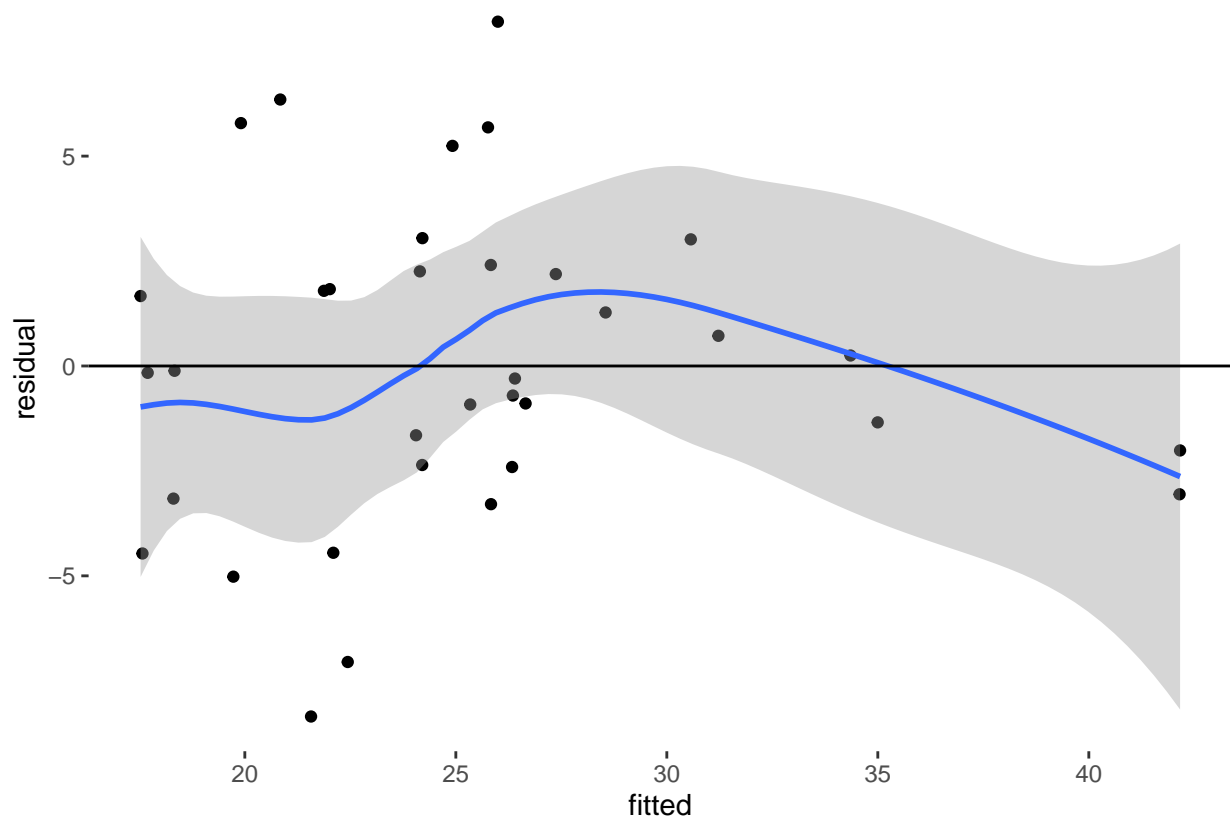
```
##      Model2_rsqs Model2.test.1_rsqs
## 1  0.7103718      0.008462034
```

EVALUATION OF MODEL

```
#Linearity for CNCS
lin <- data.frame(residual = Model1$residuals, fitted = Model1$fitted.values)

lin %>% ggplot(aes(fitted, residual)) + geom_point() + geom_smooth() + geom_hline(aes(yintercept = 0)) +
  theme(panel.grid = element_blank(), panel.background = element_blank())
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



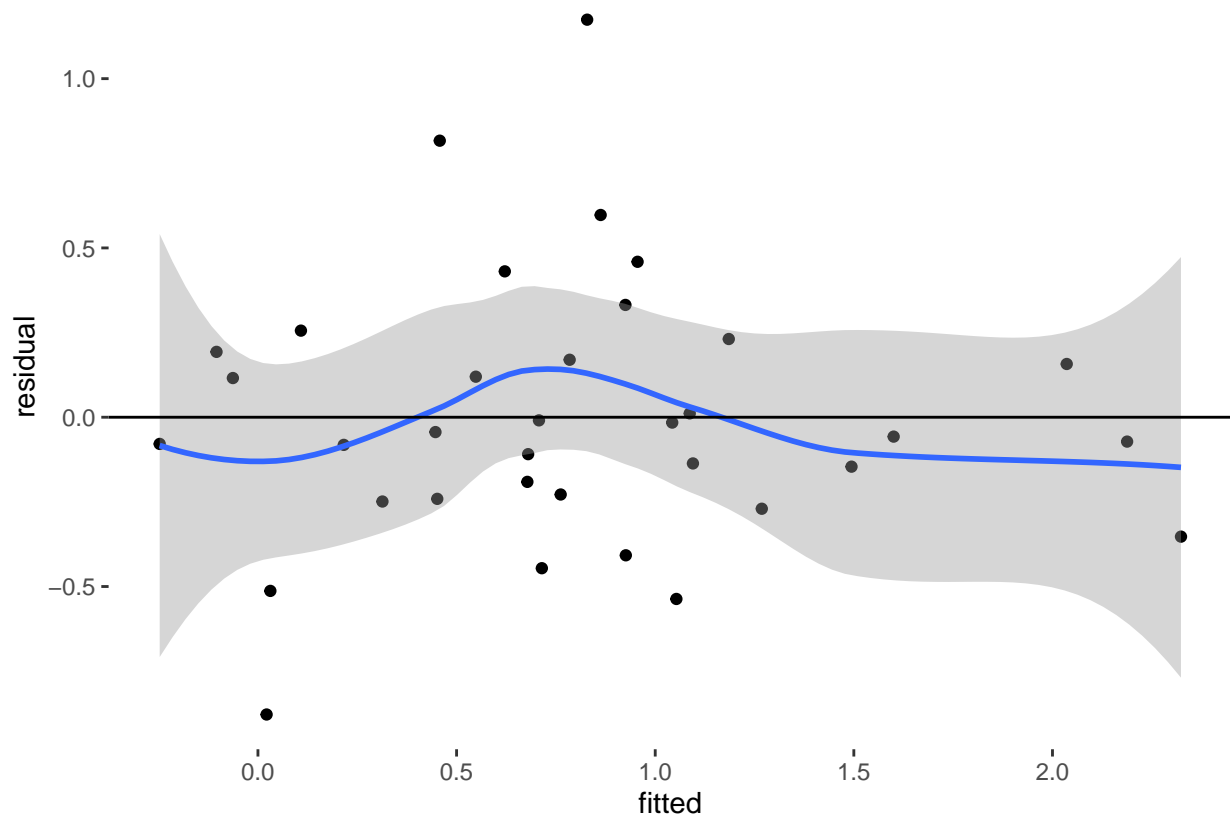
- No visible pattern of the plot hence linearity for Model 1

#Linearity

```
lin <- data.frame(residual = Model2$residuals, fitted = Model2$fitted.values)
```

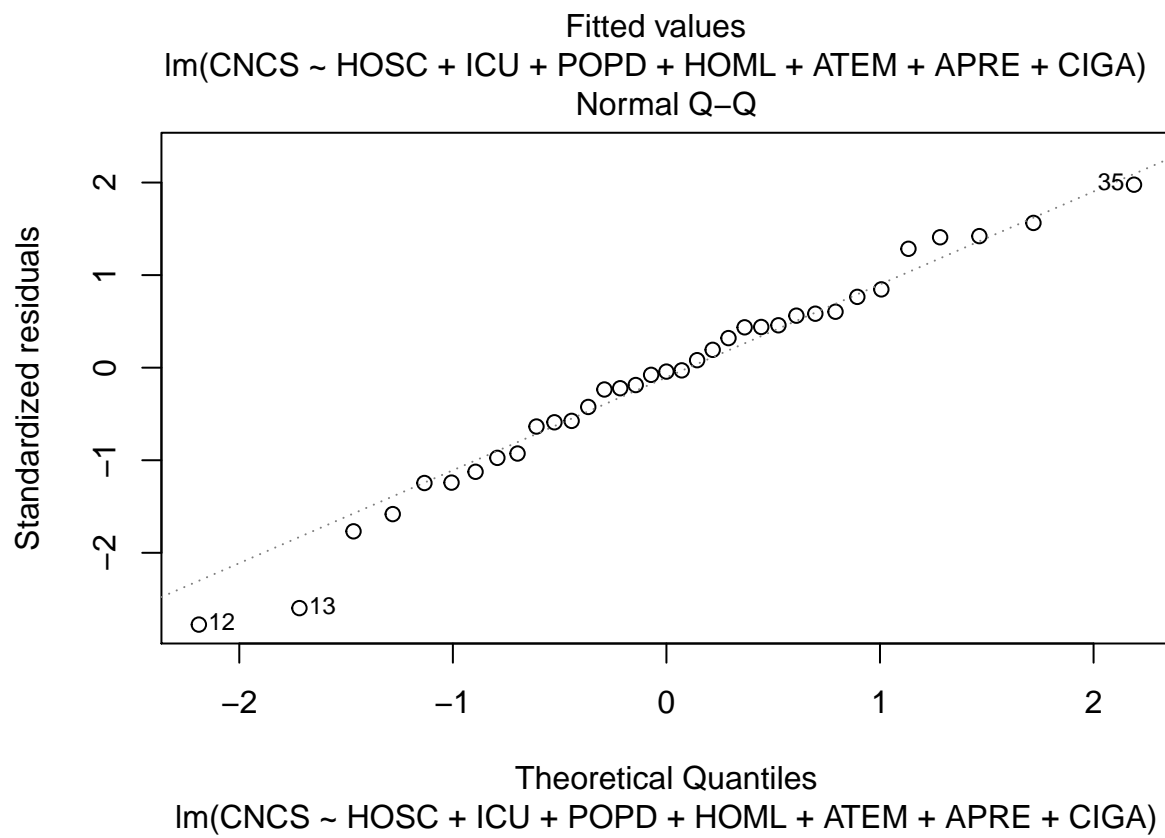
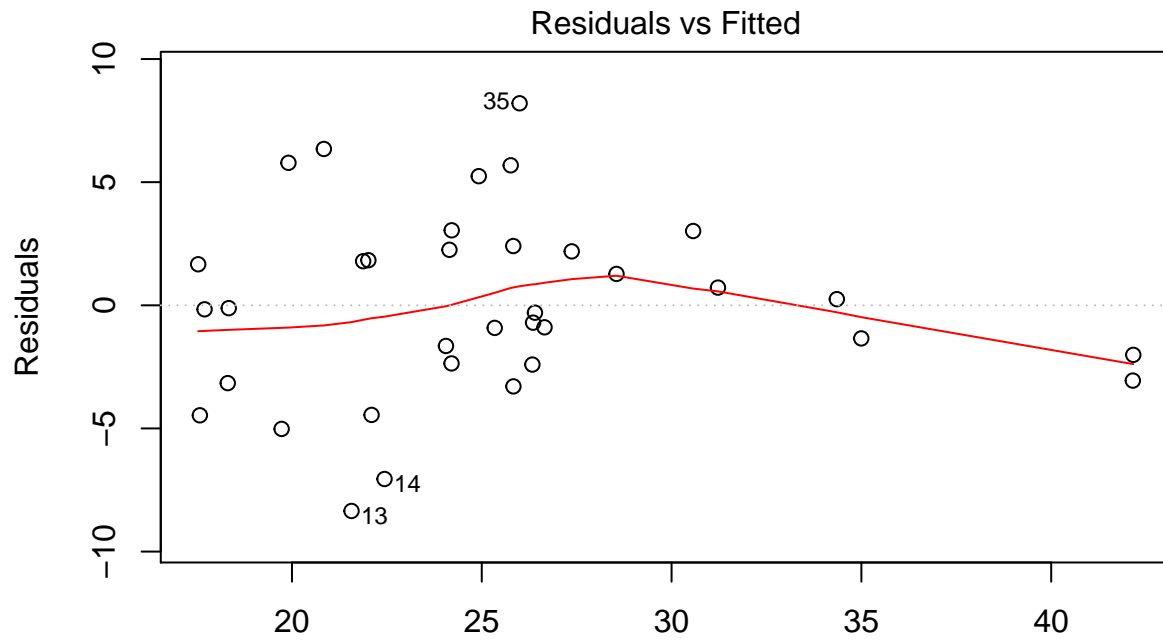
```
lin %>% ggplot(aes(fitted, residual)) + geom_point() + geom_smooth() + geom_hline(aes(yintercept = 0)) +  
  theme(panel.grid = element_blank(), panel.background = element_blank())
```

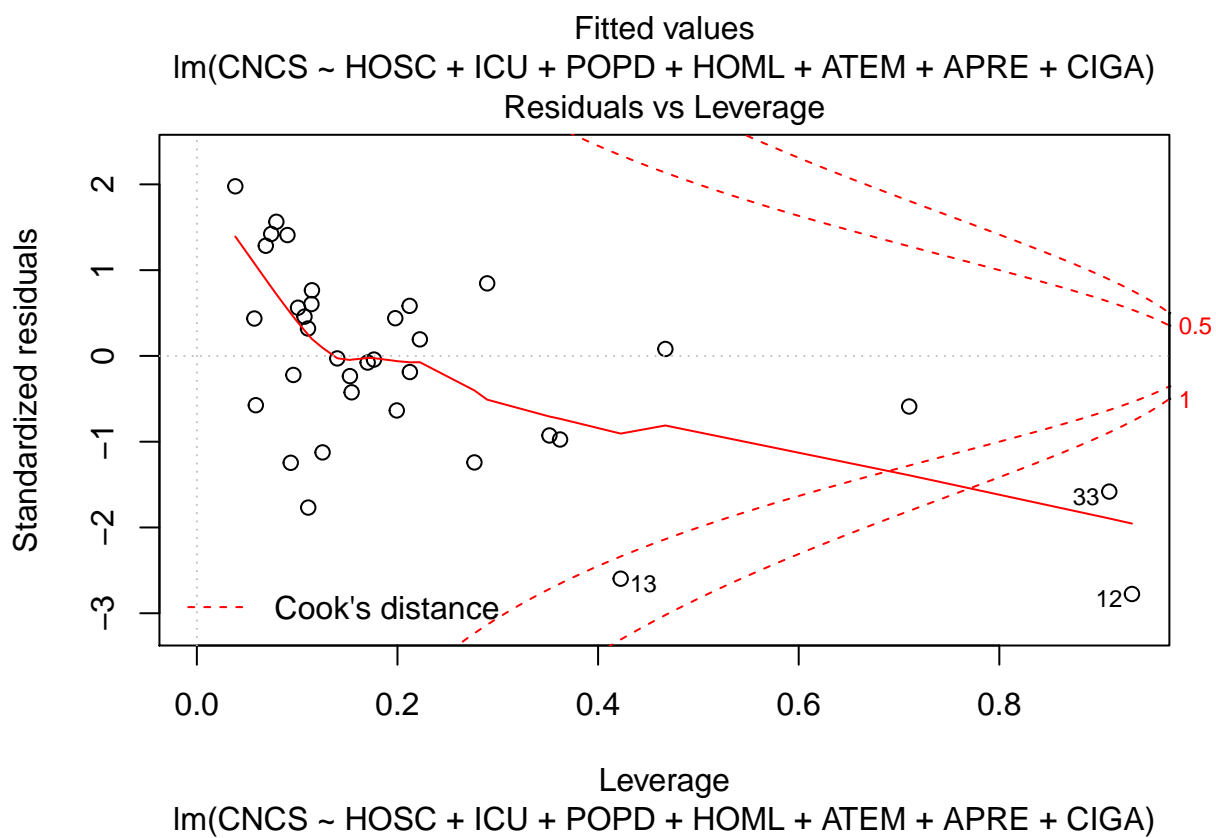
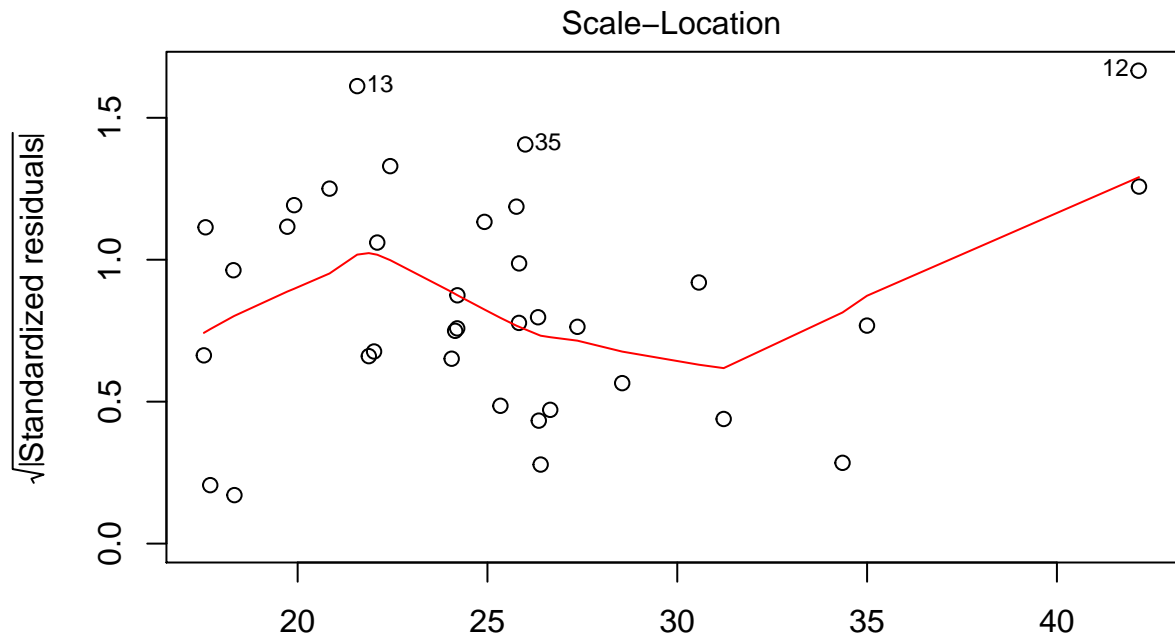
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



- No visible clear pattern for Model 2 hence linearity

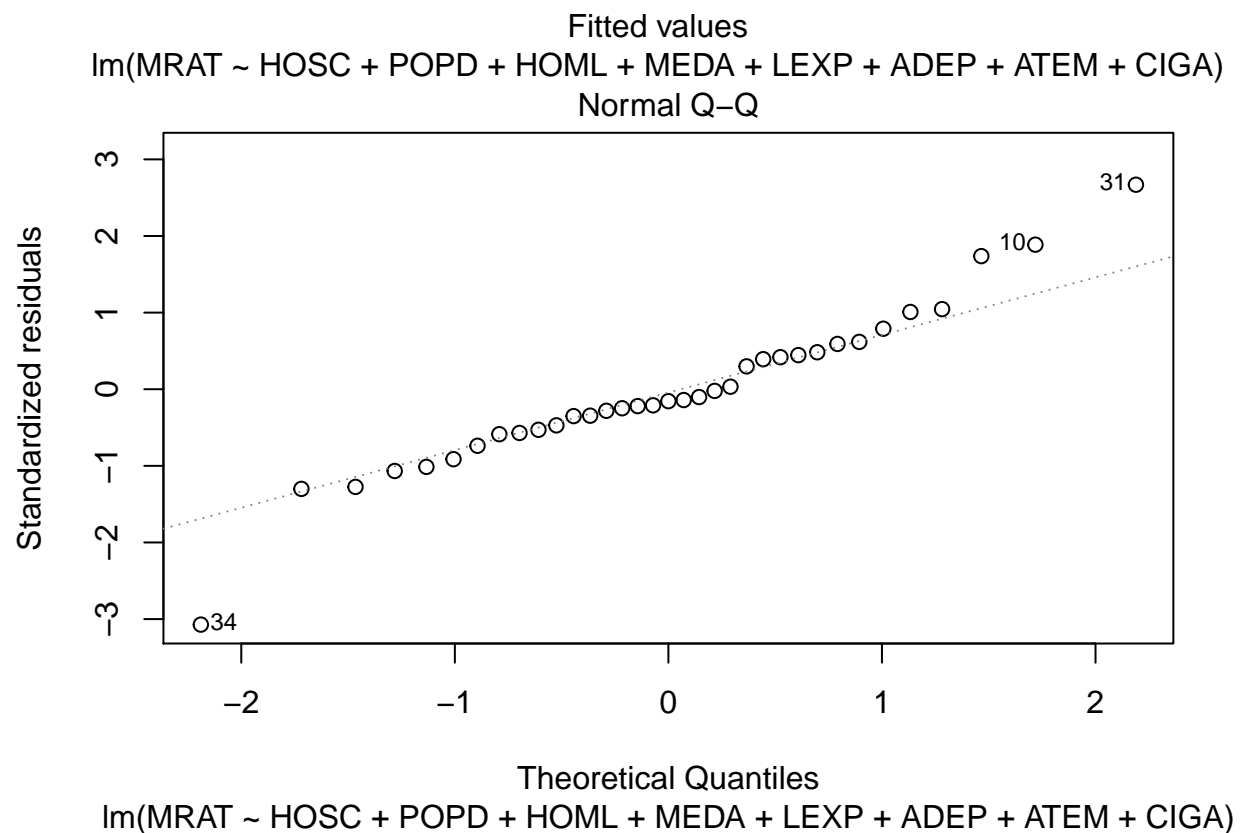
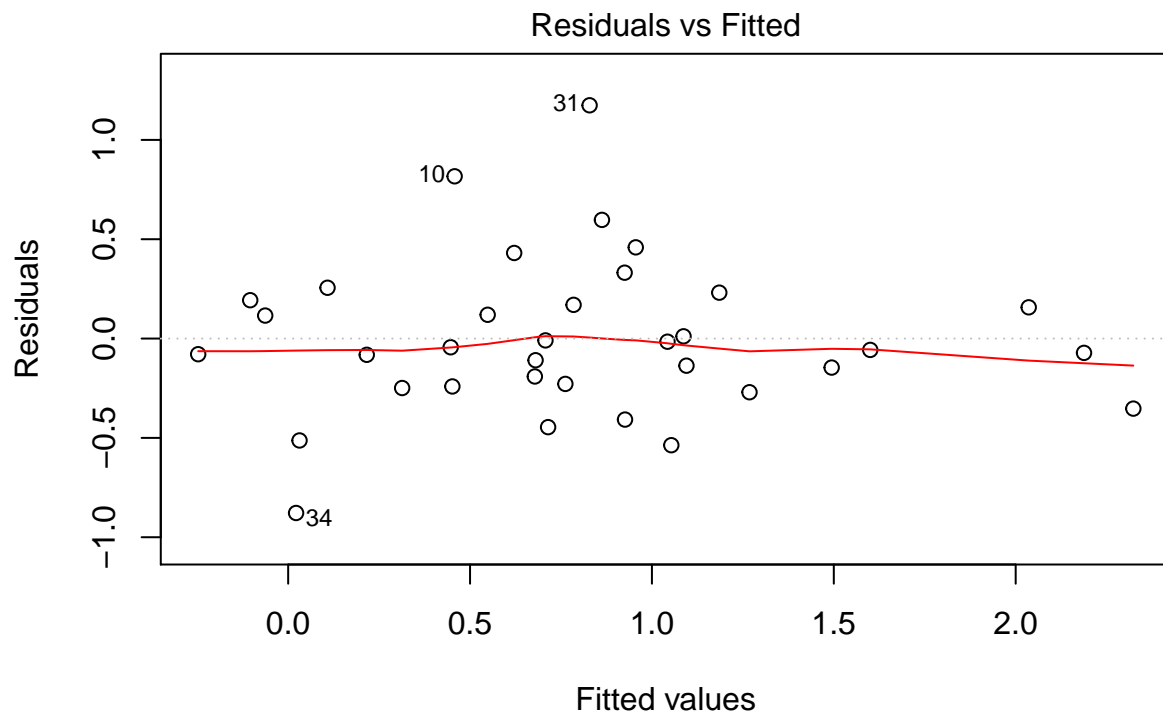
```
#par(mfrow=c(2,2))  
plot(Model1)
```

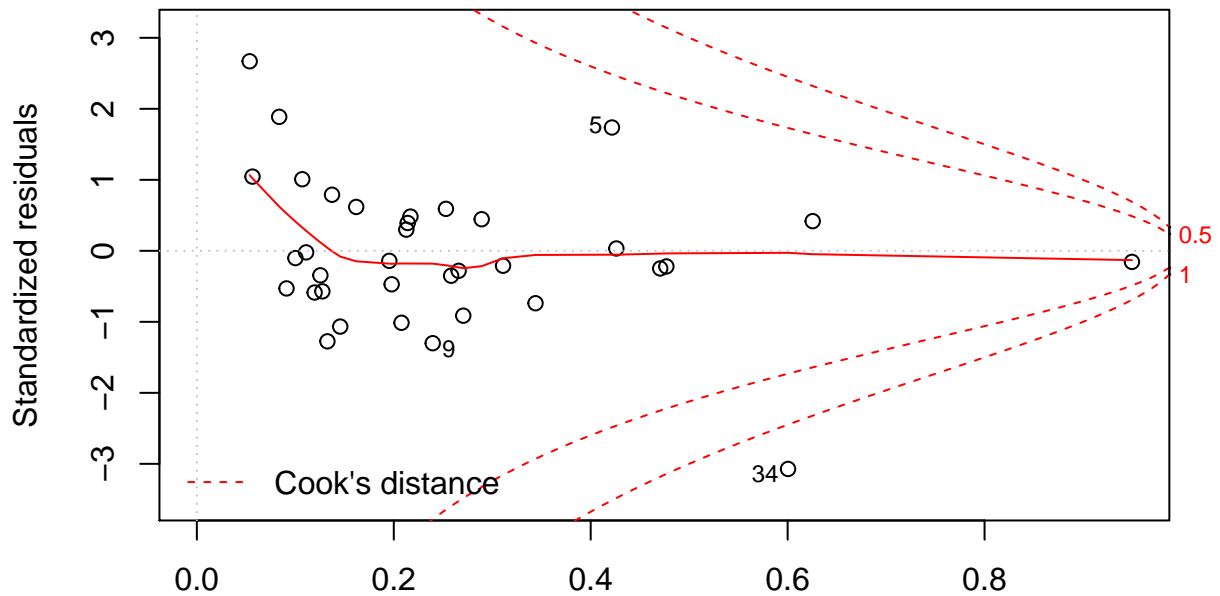
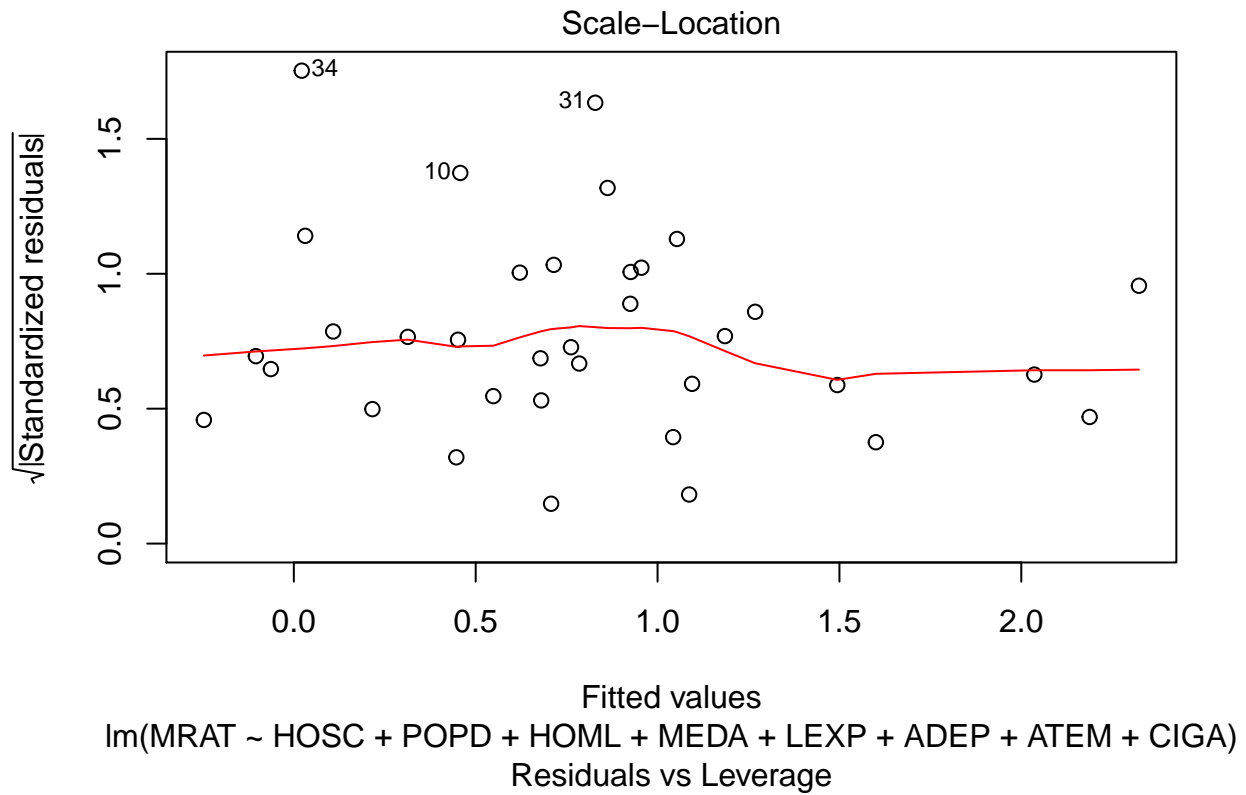




- Residual Vs fitted plot shows that there appears to be a linear relationship between the fitted line and the residual value creating a mostly horizontal line in the representation of their relationship

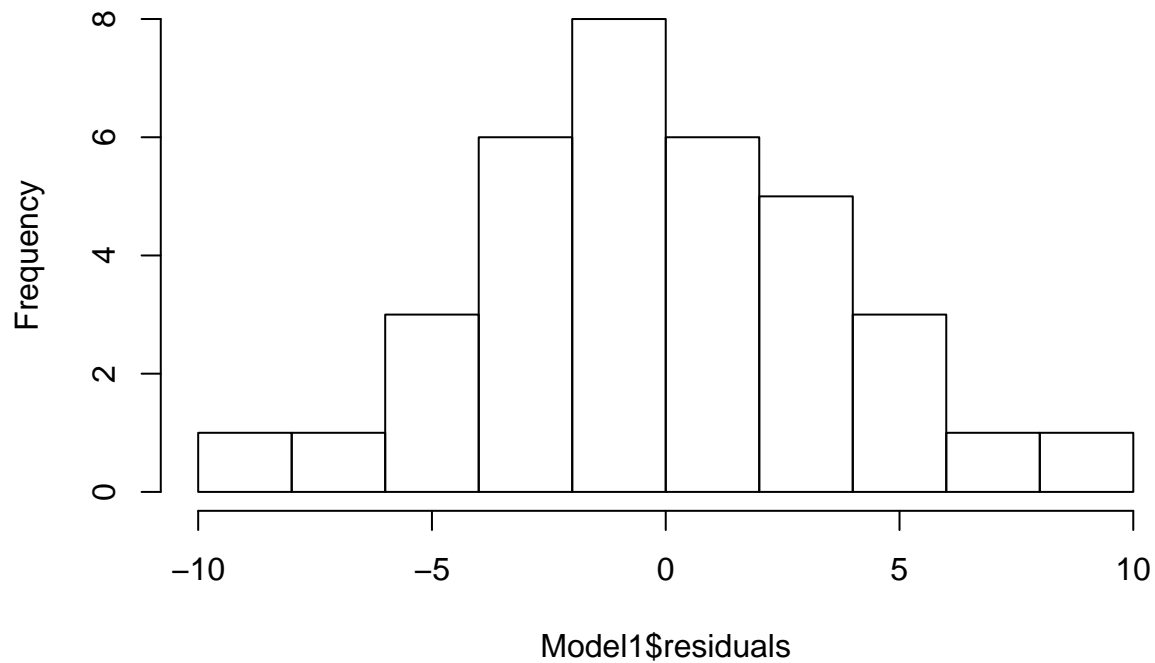
```
#par(mfrow=c(2,2))
plot(Model2)
```





```
#Normality test for CNCS
hist(Model1$residuals, breaks = 10)
```

Histogram of Model1\$residuals



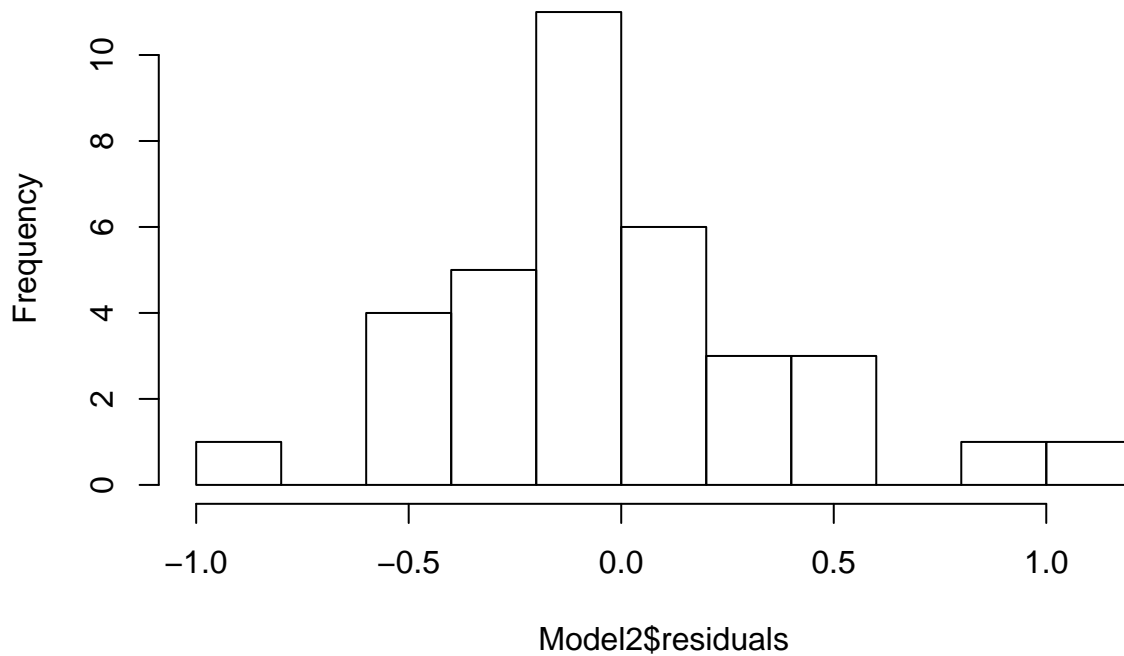
```
s=shapiro.test(Model1$residuals)
s
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Model1$residuals
## W = 0.98973, p-value = 0.9819
```

- From the Normal Q-Q plot and shapiro test the p-value > 0.05 which means that we fail to reject the null hypothesis and conclude that the residuals are following a normal distribution, double check this with the normal QQ plot.

```
#Normality test for MRAT
hist(Model2$residuals, breaks = 10)
```

Histogram of Model2\$residuals



```
t=shapiro.test(Model2$residuals)
t
```

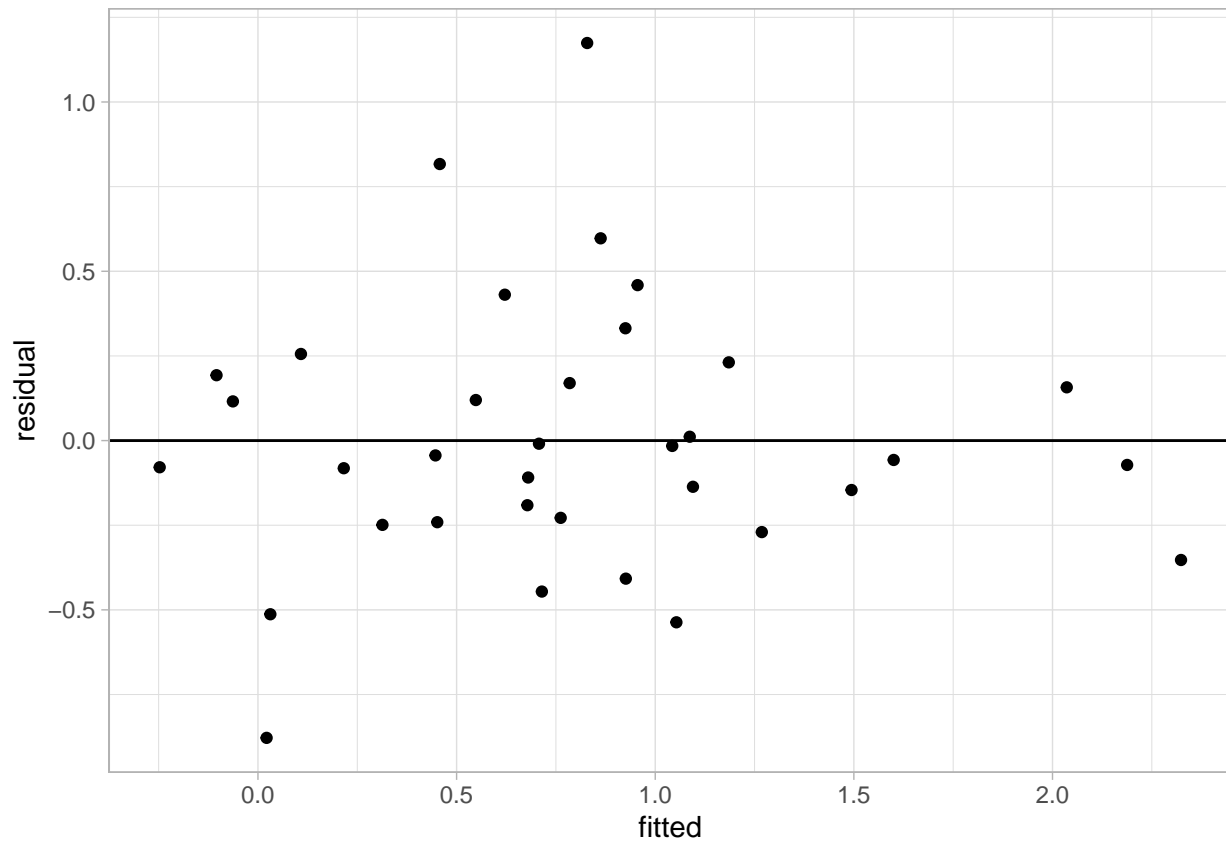
```
##
##  Shapiro-Wilk normality test
##
## data:  Model2$residuals
## W = 0.96272, p-value = 0.2756
```

- From the Normal Q-Q plot and shapiro test the p-value > 0.05 which means that we fail to reject the null hypothesis and conclude that the residuals are following a normal distribution, double check this with the normal QQ plot.

```
#Heteroscedasticity for CNCS
library(lmtest)
bptest(Model1)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  Model1
## BP = 4.6325, df = 7, p-value = 0.7047
```

```
lin %>% ggplot(aes(fitted, residual)) + geom_point() + theme_light() + geom_hline(aes(yintercept = 0))
```



- From the residual (studentized) vs fitted value plot and Using the Breusch-Pagan test, the model shows $p\text{-value} > 0.05$, so it can be concluded that heteroscedasticity is not present in our model
- with the residuals also appearing to be largely equal in variance with a slight increase towards the left of the plot.

```
#Heteroscedasticity for MRAT
```

```
library(lmtest)
```

```
bptest(Model2)
```

```
##
```

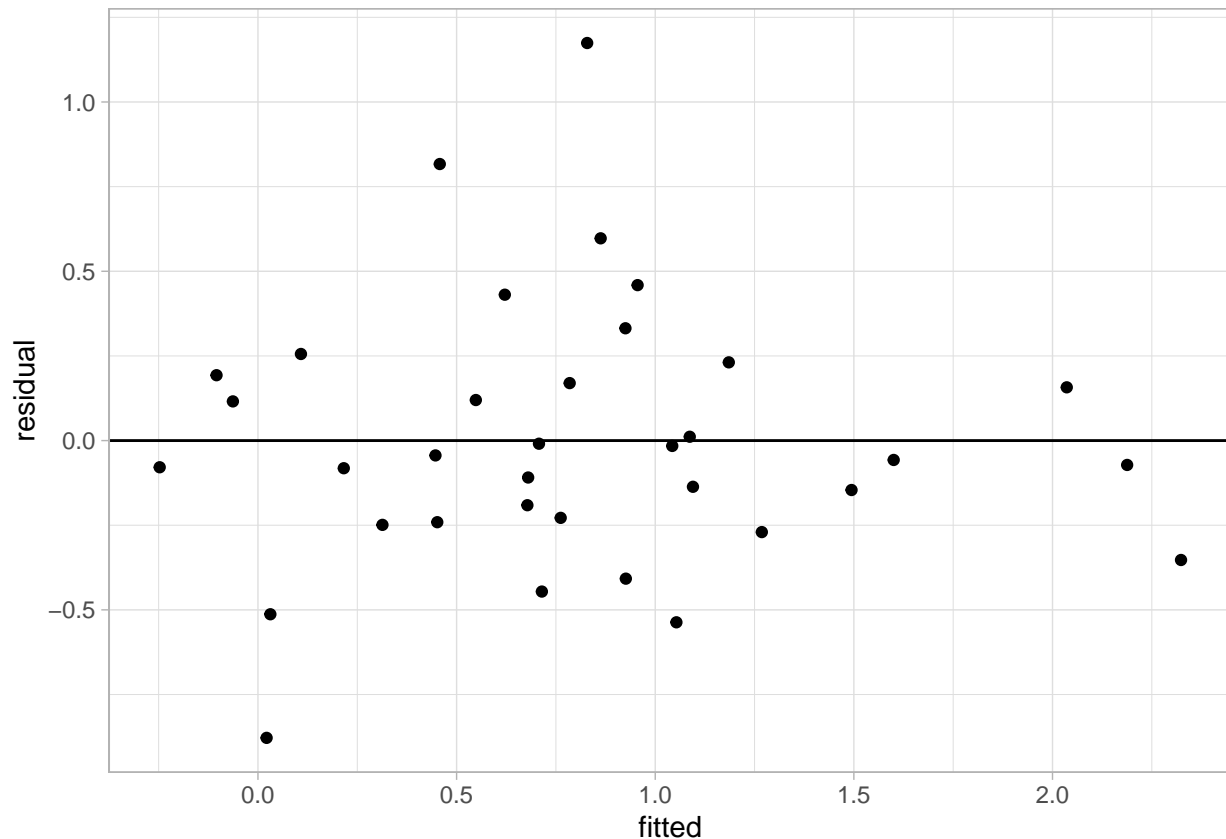
```
## studentized Breusch-Pagan test
```

```
##
```

```
## data: Model2
```

```
## BP = 6.7875, df = 8, p-value = 0.5597
```

```
lin %>% ggplot(aes(fitted, residual)) + geom_point() + theme_light() + geom_hline(aes(yintercept = 0))
```



- From the residual (studentized) vs fitted value plot and Using the Breusch-Pagan test, the model shows $p\text{-value} > 0.05$, so it can be concluded that heteroscedasticity is not present in our model
- with the residuals also appearing to be largely equal in variance with a slight increase towards the left of the plot.

```
set.seed(1)
durbinWatsonTest(Model1) #Autocorrelation for CNCS
```

```
## lag Autocorrelation D-W Statistic p-value
## 1 0.3797416 1.089293 0.004
## Alternative hypothesis: rho != 0
```

```
set.seed(1)
durbinWatsonTest(Model2) #Autocorrelation for MRAT
```

```
## lag Autocorrelation D-W Statistic p-value
## 1 0.0207532 1.905094 0.874
## Alternative hypothesis: rho != 0
```

```
#Multicollinearity
vif(Model1)
```

```
## HOSC ICU POPD HOML ATEM APRE CIGA
## 11.859874 10.847496 1.727649 1.203060 1.883680 2.141862 1.592828
```

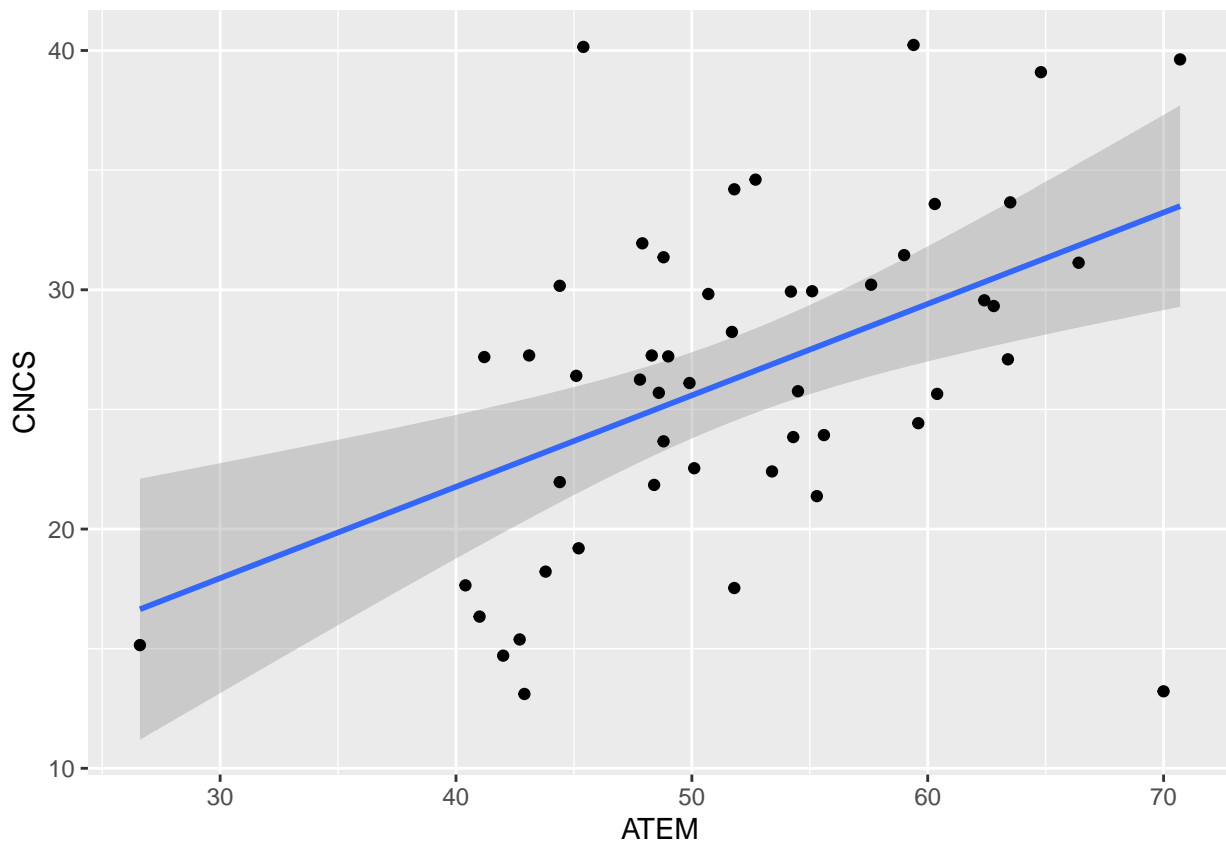
```
#Multicollinearity
vif(Model2)
```

```
##      HOSC      POPD      HOML      MEDA      LEXP      ADEP      ATEM      CIGA
## 2.967099 1.959338 1.506315 1.328379 5.934120 1.950079 5.421643 3.767684
```

- Multicollinearity could indicate correlation between the independent variables/predictors. A VIF value that exceeds 5 or 10 indicates a problematic amount of collinearity. In the model, all the VIF value is under 10 so correlation between predictors is weak.

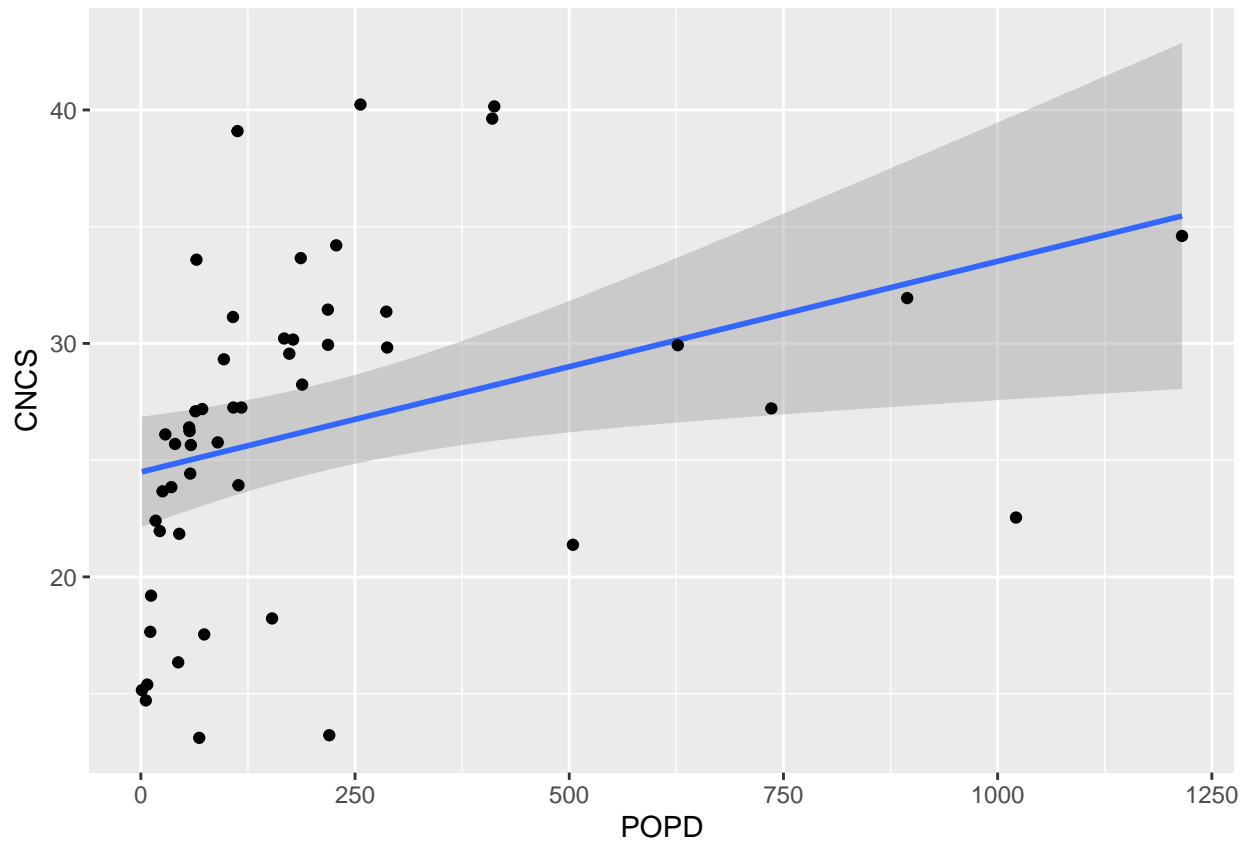
```
#EDA CCASES
par(mfrow=c(2,2))
a <- ggplot(Covid.1, aes(x = ATEM, y = CNCS))
d <- ggplot(Covid.1, aes(x = POPD, y = CNCS))
e <- ggplot(Covid.1, aes(x = APRE, y = CNCS))
# Scatter plot with regression line
a + #geom_point()+
  geom_smooth(method = "lm") +geom_point()
```

```
## `geom_smooth()` using formula 'y ~ x'
```



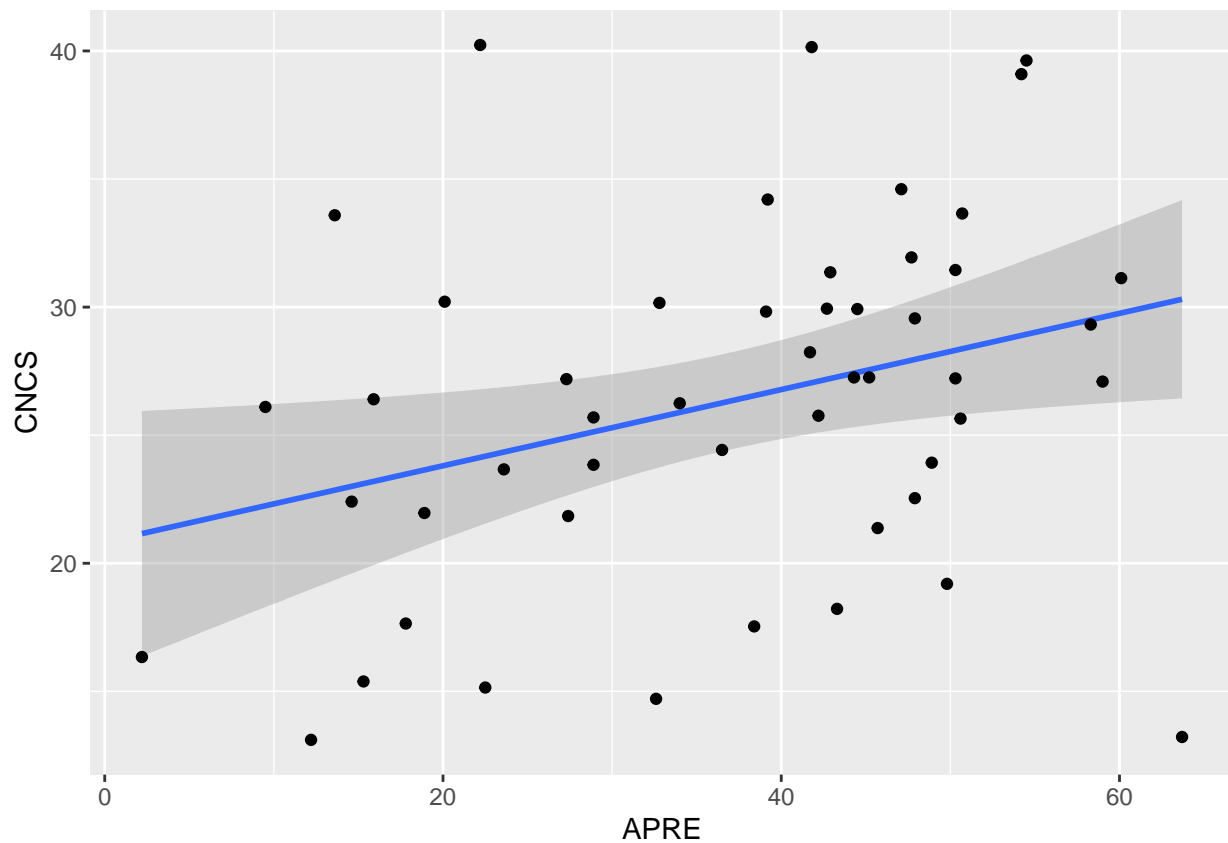
```
d + #geom_point()+
  geom_smooth(method = "lm") +geom_point()
```

```
## `geom_smooth()` using formula 'y ~ x'
```



```
e + #geom_point()+  
  geom_smooth(method = "lm") +geom_point()
```

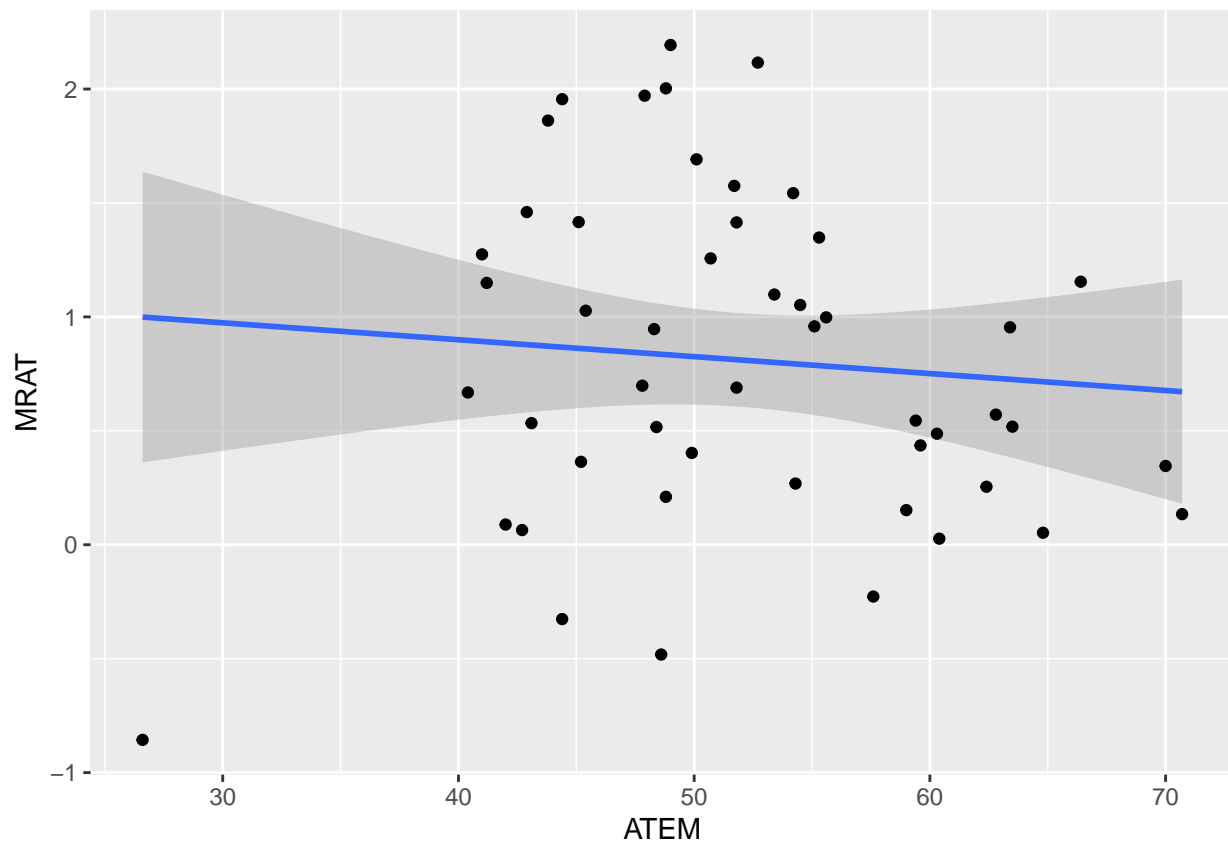
```
## `geom_smooth()` using formula 'y ~ x'
```



```
# Add a loess smoothed fit curve
#b + geom_point()+
  #geom_smooth(method = "loess")
```

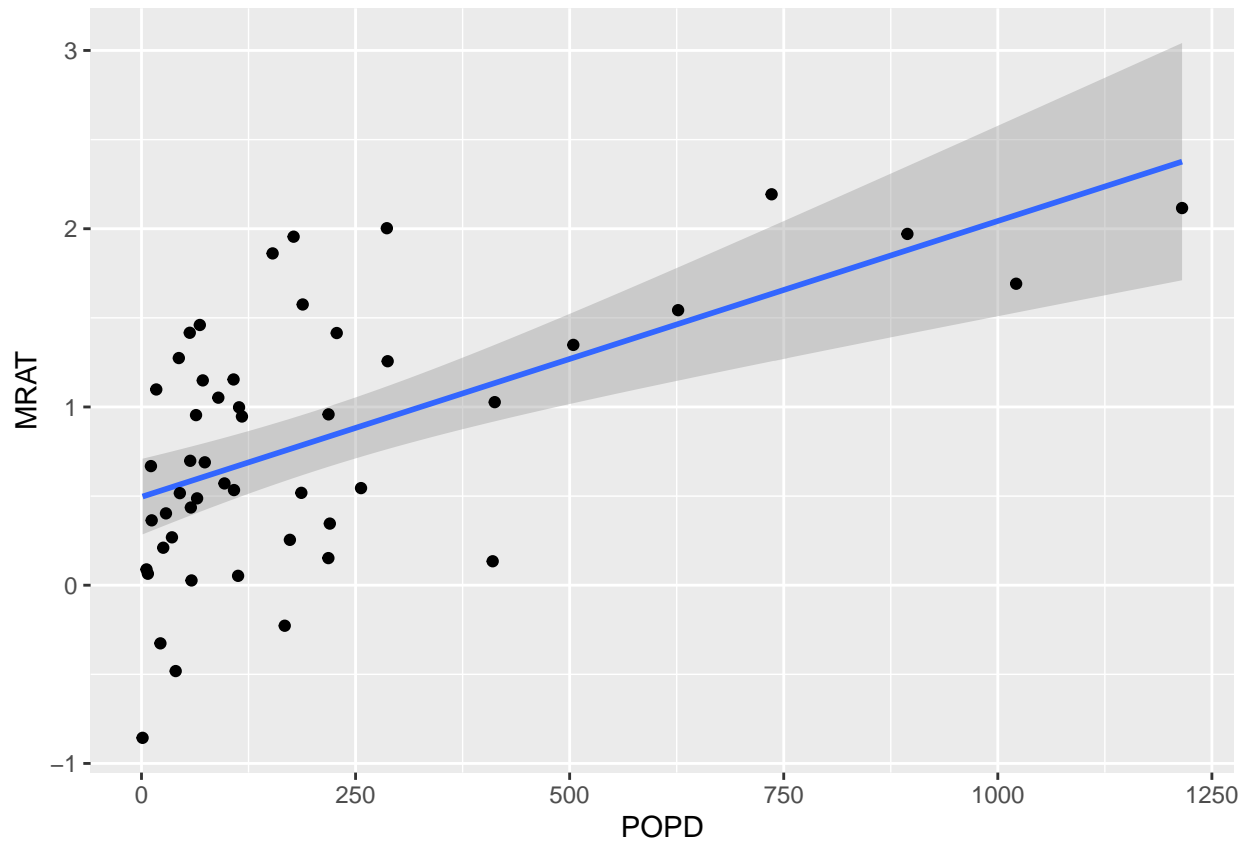
```
#EDA MRAT
par(mfrow=c(2,2))
b <- ggplot(Covid.1, aes(x = ATEM, y = MRAT))
c <- ggplot(Covid.1, aes(x = POPD, y = MRAT))
# Scatter plot with regression line
b + #geom_point()+
  geom_smooth(method = "lm") +geom_point()
```

```
## `geom_smooth()` using formula 'y ~ x'
```



```
c + #geom_point()+  
  geom_smooth(method = "lm") + geom_point()
```

```
## `geom_smooth()` using formula 'y ~ x'
```



```
# Add a loess smoothed fit curve  
#b + geom_point()+  
#geom_smooth(method = "loess")
```