Final Project

Alex Coelho, Joel Nithish Kumar Murugan, Kumaresh Pendiyala Venkatesh

2023-11-27

Importing Packages

```
library(ggplot2)
library(tigerstats)
```

```
Loading required package: abd
```

```
Loading required package: nlme
```

```
Loading required package: lattice
```

```
Loading required package: grid
```

```
Loading required package: mosaic
```

```
Registered S3 method overwritten by 'mosaic':
  method                           from
  fortify.SpatialPolygonsDataFrame ggplot2
```

```
The 'mosaic' package masks several functions from core packages in order to add
additional features.  The original behavior of these functions should not be affected by this.
```

```
Attaching package: 'mosaic'
```

```
The following objects are masked from 'package:dplyr':

    count, do, tally
```

```
The following object is masked from 'package:Matrix':

    mean
```

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

    stat
```

```
The following objects are masked from 'package:stats':

    binom.test, cor, cor.test, cov, fivenum, IQR, median, prop.test,
    quantile, sd, t.test, var
```

```
The following objects are masked from 'package:base':

    max, mean, min, prod, range, sample, sum
```

```
Welcome to tigerstats!
To learn more about this package, consult its website:
    http://homerhanumat.github.io/tigerstats
```

```
library(tidyverse)
```

```
── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
✓ forcats   1.0.0     ✓ stringr   1.5.0
✓ lubridate 1.9.2     ✓ tibble    3.2.1
✓ purrr     1.0.2     ✓ tidyr     1.3.0
✓ readr     2.1.4
```

```
── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::collapse() masks nlme::collapse()
✖ mosaic::count()   masks dplyr::count()
✖ purrr::cross()    masks mosaic::cross()
✖ mosaic::do()      masks dplyr::do()
✖ tidyr::expand()   masks Matrix::expand()
✖ dplyr::filter()   masks stats::filter()
✖ dplyr::lag()      masks stats::lag()
✖ tidyr::pack()     masks Matrix::pack()
✖ mosaic::stat()    masks ggplot2::stat()
✖ mosaic::tally()   masks dplyr::tally()
✖ tidyr::unpack()   masks Matrix::unpack()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become e
rrors
```

```
library(MLmetrics)
```

```
Warning: package 'MLmetrics' was built under R version 4.3.2
```

```

Attaching package: 'MLmetrics'

The following object is masked from 'package:base':

    Recall
```

```
library(MASS)
```

```
Attaching package: 'MASS'

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

    select
```

Data Cleaning and Preparation

Importing Data Set

```
Covid <- read.csv("country_wise_latest.csv", sep=",")
head(Covid)
```

```
       Country.Region Confirmed Deaths Recovered Active New.cases New.deaths
1         Afghanistan     36263   1269     25198   9796       106         10
2             Albania      4880    144      2745   1991       117          6
3             Algeria     27973   1163     18837   7973       616          8
4             Andorra       907     52       803     52        10          0
5              Angola       950     41       242    667        18          1
6 Antigua and Barbuda        86      3        65     18         4          0
  New.recovered Deaths...100.Cases Recovered...100.Cases Deaths...100.Recovered
1            18               3.50                 69.49                   5.04
2            63               2.95                 56.25                   5.25
3           749               4.16                 67.34                   6.17
4             0               5.73                 88.53                   6.48
5             0               4.32                 25.47                  16.94
6             5               3.49                 75.58                   4.62
  Confirmed.last.week X1.week.change X1.week...increase            WHO.Region
1               35526            737               2.07 Eastern Mediterranean
2                4171            709              17.00                Europe
3               23691           4282              18.07                Africa
4                 884             23               2.60                Europe
5                 749            201              26.84                Africa
6                  76             10              13.16              Americas
```

```
summary(Covid)
```

```
 Country.Region       Confirmed           Deaths           Recovered
 Length:187         Min.   :     10   Min.   :     0.0   Min.   :      0.0
 Class :character   1st Qu.:   1114   1st Qu.:    18.5   1st Qu.:    626.5
 Mode  :character   Median :   5059   Median :   108.0   Median :   2815.0
                    Mean   :  88131   Mean   :  3497.5   Mean   :  50631.5
                    3rd Qu.:  40460   3rd Qu.:   734.0   3rd Qu.:  22606.0
                    Max.   :4290259   Max.   :148011.0   Max.   :1846641.0
     Active            New.cases         New.deaths      New.recovered
 Min.   :      0.0   Min.   :    0.0   Min.   :   0.00   Min.   :    0.0
 1st Qu.:    141.5   1st Qu.:    4.0   1st Qu.:   0.00   1st Qu.:    0.0
 Median :   1600.0   Median :   49.0   Median :   1.00   Median :   22.0
 Mean   :  34001.9   Mean   : 1223.0   Mean   :  28.96   Mean   :  933.8
 3rd Qu.:   9149.0   3rd Qu.:  419.5   3rd Qu.:   6.00   3rd Qu.:  221.0
 Max.   :2816444.0   Max.   :56336.0   Max.   :1076.00   Max.   :33728.0
 Deaths...100.Cases Recovered...100.Cases Deaths...100.Recovered
 Min.   : 0.000     Min.   :  0.00        Min.   :0.00
 1st Qu.: 0.945     1st Qu.: 48.77        1st Qu.:1.45
 Median : 2.150     Median : 71.32        Median :3.62
 Mean   : 3.020     Mean   : 64.82        Mean   : Inf
 3rd Qu.: 3.875     3rd Qu.: 86.89        3rd Qu.:6.44
 Max.   :28.560     Max.   :100.00        Max.   : Inf
 Confirmed.last.week X1.week.change   X1.week...increase  WHO.Region
 Min.   :     10     Min.   :   -47   Min.   : -3.840    Length:187
 1st Qu.:   1052     1st Qu.:    49   1st Qu.:  2.775    Class :character
 Median :   5020     Median :   432   Median :  6.890    Mode  :character
 Mean   :  78682     Mean   :  9448   Mean   : 13.606
 3rd Qu.:  37080     3rd Qu.:  3172   3rd Qu.: 16.855
 Max.   :3834677     Max.   :455582   Max.   :226.320
```

Checking Data Types

```
str(Covid)
```

```
'data.frame':   187 obs. of  15 variables:
 $ Country.Region        : chr  "Afghanistan" "Albania" "Algeria" "Andorra" ...
 $ Confirmed             : int  36263 4880 27973 907 950 86 167416 37390 15303 20558 ...
 $ Deaths                : int  1269 144 1163 52 41 3 3059 711 167 713 ...
 $ Recovered             : int  25198 2745 18837 803 242 65 72575 26665 9311 18246 ...
 $ Active                : int  9796 1991 7973 52 667 18 91782 10014 5825 1599 ...
 $ New.cases             : int  106 117 616 10 18 4 4890 73 368 86 ...
 $ New.deaths            : int  10 6 8 0 1 0 120 6 6 1 ...
 $ New.recovered         : int  18 63 749 0 0 5 2057 187 137 37 ...
 $ Deaths...100.Cases    : num  3.5 2.95 4.16 5.73 4.32 3.49 1.83 1.9 1.09 3.47 ...
 $ Recovered...100.Cases : num  69.5 56.2 67.3 88.5 25.5 ...
 $ Deaths...100.Recovered: num  5.04 5.25 6.17 6.48 16.94 ...
 $ Confirmed.last.week   : int  35526 4171 23691 884 749 76 130774 34981 12428 19743 ...
 $ X1.week.change        : int  737 709 4282 23 201 10 36642 2409 2875 815 ...
 $ X1.week...increase    : num  2.07 17 18.07 2.6 26.84 ...
 $ WHO.Region            : chr  "Eastern Mediterranean" "Europe" "Africa" "Europe" ...
```

Checking Observations

```
dim(Covid)
```

```
[1] 187  15
```

Checking for Missing Values

```
sum(is.na(Covid))
```

```
[1] 0
```

Checking for Duplicate Values

```
sum(Covid[duplicated(Covid), ])
```

```
[1] 0
```

Standardizing Data

```
sum(is.infinite(Covid$Deaths...100.Recovered))
```

```
[1] 5
```

```
Covid <- Covid[is.finite(Covid$`Deaths...100.Recovered`), ]

sum(is.infinite(Covid$Deaths...100.Recovered))
```

```
[1] 0
```

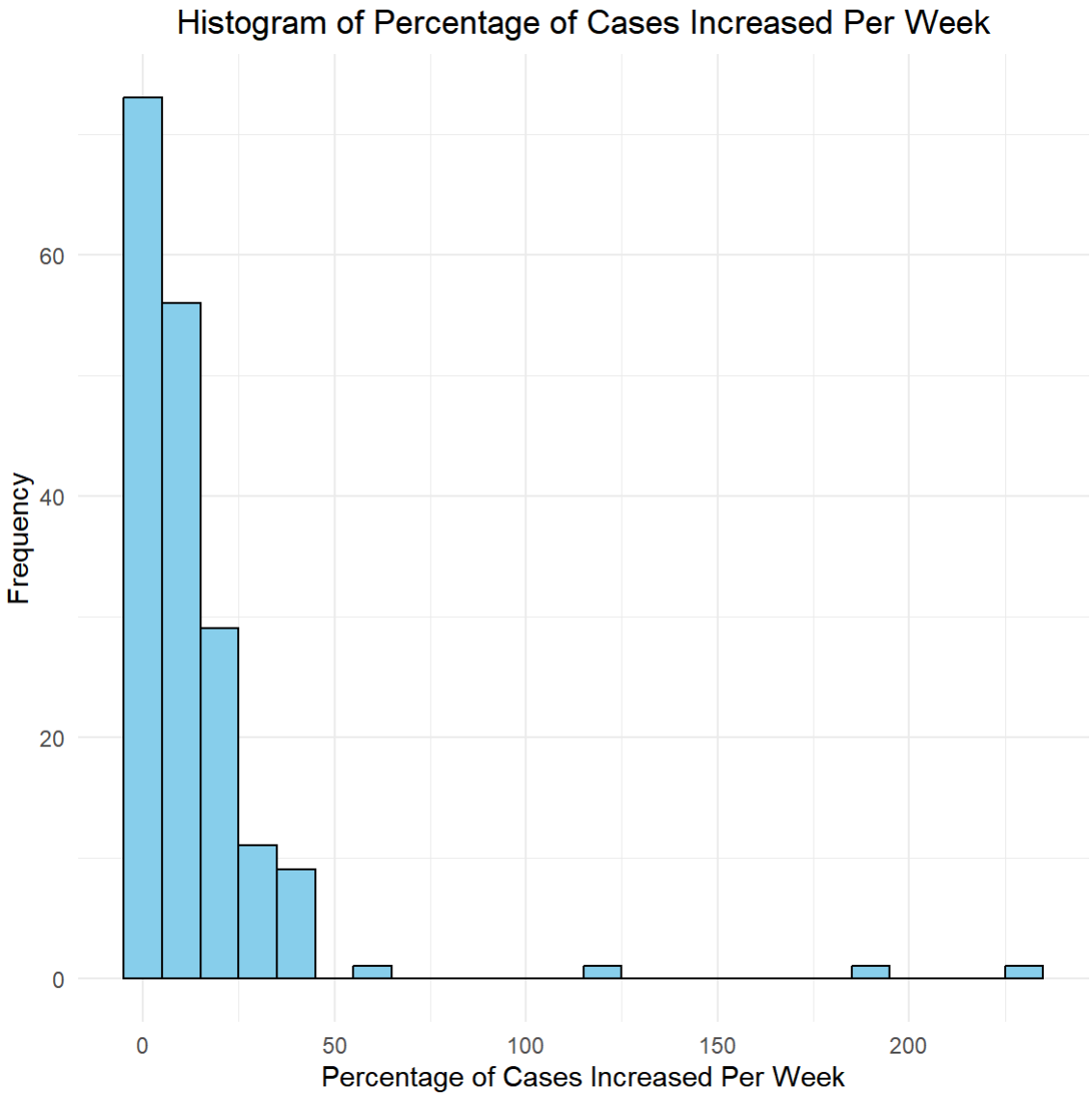
```
Covid <- Covid[, c(2:14,1,15)]

str(Covid)
```

```
'data.frame':   182 obs. of  15 variables:
 $ Confirmed             : int  36263 4880 27973 907 950 86 167416 37390 15303 20558 ...
 $ Deaths                : int  1269 144 1163 52 41 3 3059 711 167 713 ...
 $ Recovered             : int  25198 2745 18837 803 242 65 72575 26665 9311 18246 ...
 $ Active                : int  9796 1991 7973 52 667 18 91782 10014 5825 1599 ...
 $ New.cases             : int  106 117 616 10 18 4 4890 73 368 86 ...
 $ New.deaths            : int  10 6 8 0 1 0 120 6 6 1 ...
 $ New.recovered         : int  18 63 749 0 0 5 2057 187 137 37 ...
 $ Deaths...100.Cases    : num  3.5 2.95 4.16 5.73 4.32 3.49 1.83 1.9 1.09 3.47 ...
 $ Recovered...100.Cases : num  69.5 56.2 67.3 88.5 25.5 ...
 $ Deaths...100.Recovered: num  5.04 5.25 6.17 6.48 16.94 ...
 $ Confirmed.last.week   : int  35526 4171 23691 884 749 76 130774 34981 12428 19743 ...
 $ X1.week.change        : int  737 709 4282 23 201 10 36642 2409 2875 815 ...
 $ X1.week...increase    : num  2.07 17 18.07 2.6 26.84 ...
 $ Country.Region        : chr  "Afghanistan" "Albania" "Algeria" "Andorra" ...
 $ WHO.Region            : chr  "Eastern Mediterranean" "Europe" "Africa" "Europe" ...
```

Exploratory Data Analysis

Histogram

```
ggplot(data = Covid, aes(x = X1.week...increase )) +
  geom_histogram(binwidth = 10, fill = "skyblue", color = "black") +
  labs(title = "Histogram of Percentage of Cases Increased Per Week", x = "Percentage of Cases I
ncreased Per Week", y = "Frequency") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```

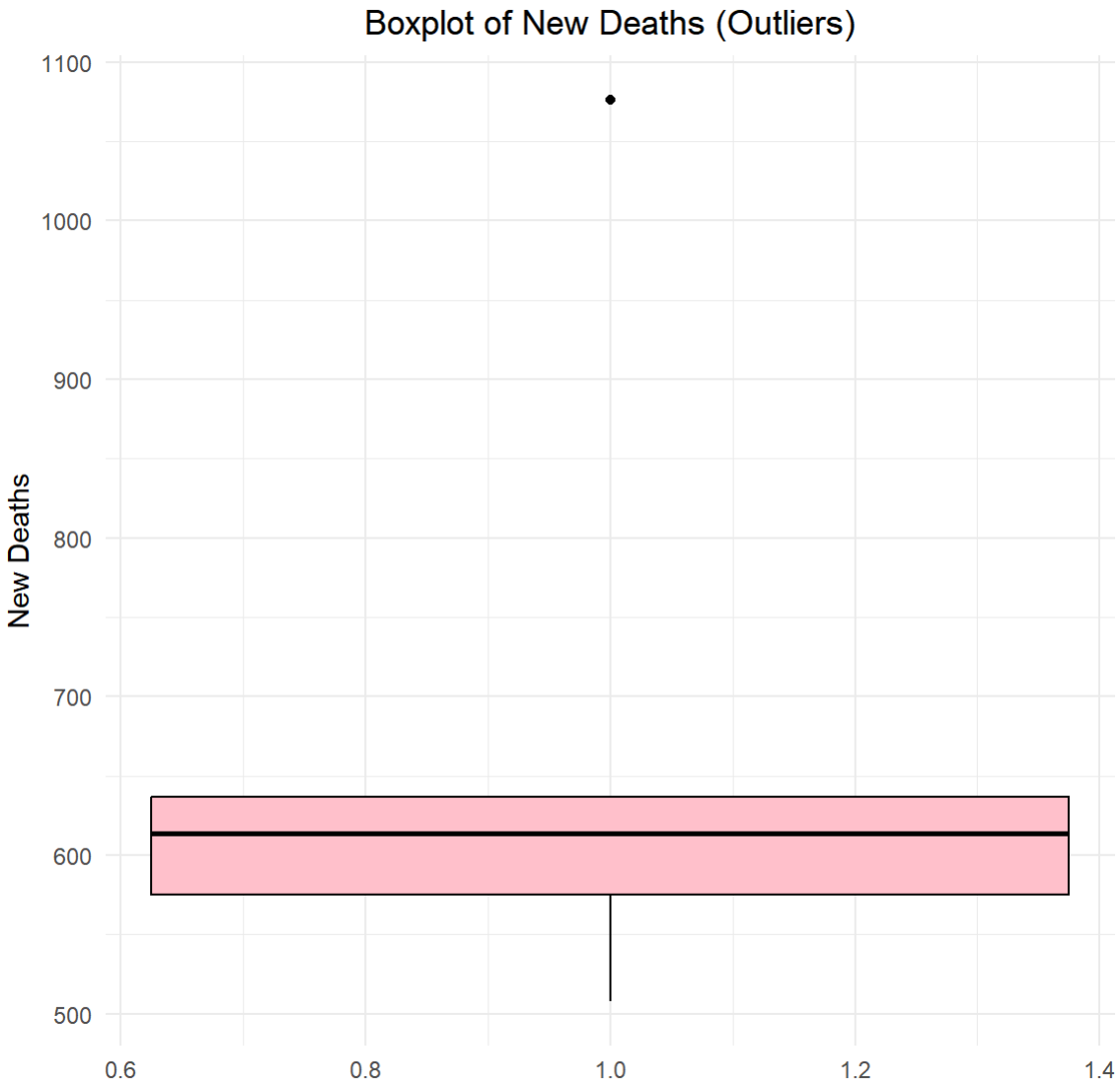


Boxplot

```
Covid_a <- scale(Covid$New.deaths)
Covid_outliers <- Covid[Covid_a>3,]

Covid_outliers
```

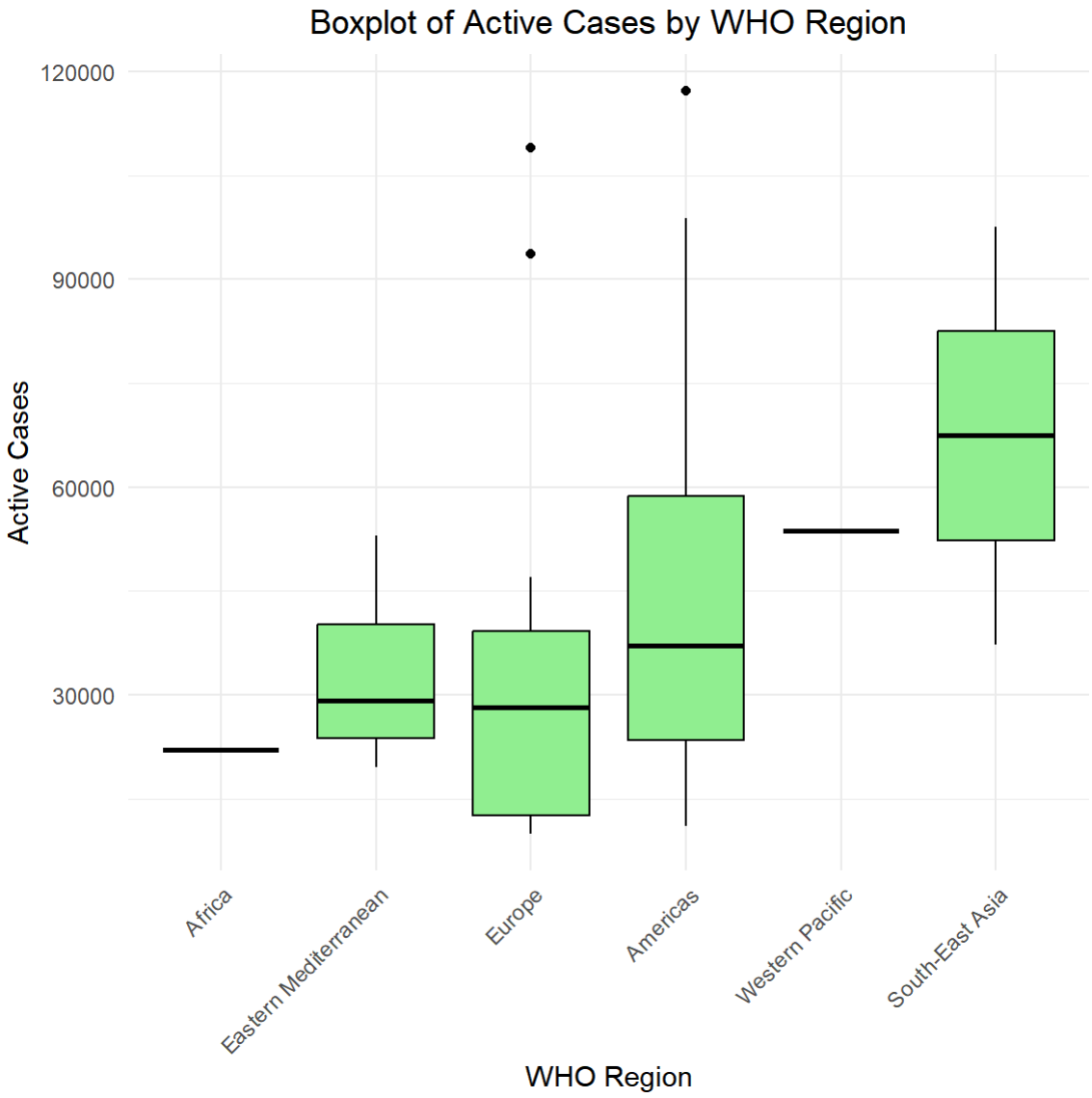
```
    Confirmed Deaths Recovered  Active New.cases New.deaths New.recovered
24    2442375  87618   1846641  508116     23284        614         33728
38     257101   8777    131161  117163     16306        508         11494
80    1480073  33408    951166  495499     44457        637         33598
133    389717  18418    272547   98752     13756        575          4697
174   4290259 148011   1325804 2816444     56336       1076         27941
    Deaths...100.Cases Recovered...100.Cases Deaths...100.Recovered
24                3.59                 75.61                   4.74
38                3.41                 51.02                   6.69
80                2.26                 64.26                   3.51
133               4.73                 69.93                   6.76
174               3.45                 30.90                  11.16
    Confirmed.last.week X1.week.change X1.week...increase Country.Region
24              2118646         323729              15.28         Brazil
38               204005          53096              26.03       Colombia
80              1155338         324735              28.11          India
133              357681          32036               8.96           Peru
174             3834677         455582              11.88             US
         WHO.Region
24         Americas
38         Americas
80  South-East Asia
133        Americas
174        Americas
```

```
ggplot(data = Covid_outliers, aes(x = 1, y = New.deaths)) +
  geom_boxplot(fill = "pink", color = "black") +
  labs(title = "Boxplot of New Deaths (Outliers)", x = "", y = "New Deaths") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```



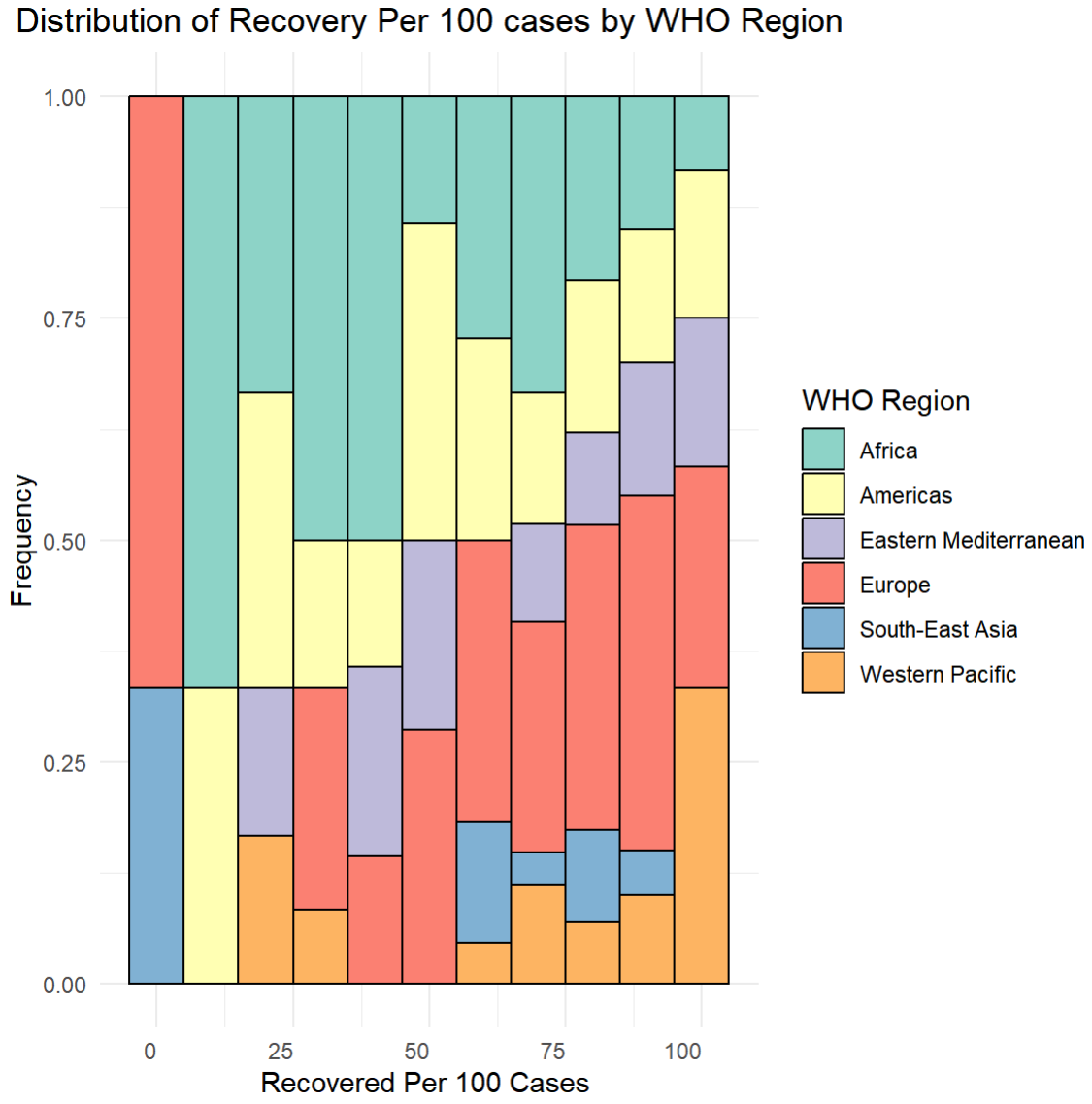
Overlay Boxplot

```
Covid_segment <- Covid[Covid$Active > 10000 & Covid$Active < 150000, ]

ggplot(data = Covid_segment, aes(x = reorder(WHO.Region, Active), y = Active)) +
  geom_boxplot(fill = "lightgreen", color = "black") +
  labs(
    title = "Boxplot of Active Cases by WHO Region",
    x = "WHO Region",
    y = "Active Cases"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  theme(plot.title = element_text(hjust = 0.5))
```



Overlay Histogram

```
ggplot(data = Covid, aes(x = Recovered...100.Cases )) +
  geom_histogram(binwidth = 10, aes(fill = WHO.Region), color = "black", position = "fill") +
  labs(
    x = "Recovered Per 100 Cases ",
    y = "Frequency",
    fill = "WHO Region",
    title = "Distribution of Recovery Per 100 cases by WHO Region"
  ) +
  theme_minimal() +
  scale_fill_brewer(palette = "Set3") +
  theme(axis.text.x = element_text( hjust = 1,), plot.title = element_text(hjust = 0.5))
```



Hypotheses

```
Covid_Asia <- Covid$Recovered...100.Cases[Covid$WHO.Region == "South-East Asia"]

Covid_Europe <- Covid$Recovered...100.Cases[Covid$WHO.Region == "Europe"]
```

Parameter:
σ^2_{CA} = Variance of the Recovered Per 100 Cases in Asia.

σ^2_{CE} = Variance of the Recovered Per 100 Cases in Europe.

Hypotheses:

$$H_0 : \sigma^2_{CA}/\sigma^2_{CE} = 1,$$
$$H_1 : \sigma^2_{CA}/\sigma^2_{CE} \neq 1,$$

Variance Test

```
var.test(Covid_Asia, Covid_Europe)
```

```
    F test to compare two variances

data:  Covid_Asia and Covid_Europe
F = 1.2433, num df = 9, denom df = 53, p-value = 0.5792
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.525703 4.307022
sample estimates:
ratio of variances
          1.243271
```

Parameter:
$\mu_{CA} =$ mean of the Recovered Per 100 Cases in Asia.

$\mu_{CE} =$ mean of the Recovered Per 100 Cases in Europe.

Hypotheses:

$$H_0 : \mu_{CA} = \mu_{CE},$$
$$H_1 : \mu_{CA} \neq \mu_{CE}.$$

Hypotheses Test

```
t.test(Covid_Asia, Covid_Europe, var.equal = TRUE)
```

```
    Two Sample t-test

data:  Covid_Asia and Covid_Europe
t = -0.53776, df = 62, p-value = 0.5927
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -21.10028  12.15421
sample estimates:
mean of x mean of y
 66.70400  71.17704
```

Linear Regression

```
i <- sample(2, nrow(Covid), replace = TRUE, prob = c(0.8, 0.2))
CovidTraining <- Covid[i == 1, ]
CovidTest <- Covid[i == 2, ]
summary(CovidTraining)
```

```
  Confirmed           Deaths          Recovered           Active
 Min.   :     10   Min.   :     0   Min.   :      0   Min.   :      0.0
 1st Qu.:   1132   1st Qu.:    21   1st Qu.:    803   1st Qu.:    128.5
 Median :   6321   Median :   115   Median :   3824   Median :   1920.0
 Mean   : 102503   Mean   :  4149   Mean   :  59720   Mean   :  38633.7
 3rd Qu.:  44652   3rd Qu.:  1040   3rd Qu.:  26120   3rd Qu.:   8877.0
 Max.   :4290259   Max.   :148011   Max.   :1846641   Max.   :2816444.0
  New.cases       New.deaths      New.recovered   Deaths...100.Cases
 Min.   :    0   Min.   :   0.00   Min.   :    0   Min.   : 0.000
 1st Qu.:    5   1st Qu.:   0.00   1st Qu.:    0   1st Qu.: 0.925
 Median :   49   Median :   1.00   Median :   24   Median : 2.110
 Mean   : 1337   Mean   :  30.01   Mean   : 1007   Mean   : 3.037
 3rd Qu.:  468   3rd Qu.:   6.00   3rd Qu.:  252   3rd Qu.: 3.885
 Max.   :56336   Max.   :1076.00   Max.   :33728   Max.   :28.560
 Recovered...100.Cases Deaths...100.Recovered Confirmed.last.week
 Min.   :  0.00        Min.   :   0.000       Min.   :     10
 1st Qu.: 53.38        1st Qu.:   1.295       1st Qu.:   1064
 Median : 73.35        Median :   3.510       Median :   5639
 Mean   : 67.66        Mean   :  48.953       Mean   :  91817
 3rd Qu.: 87.63        3rd Qu.:   6.155       3rd Qu.:  39261
 Max.   :100.00        Max.   :3259.260       Max.   :3834677
 X1.week.change   X1.week...increase Country.Region      WHO.Region
 Min.   :   -47   Min.   : -3.84     Length:147         Length:147
 1st Qu.:    49   1st Qu.:  2.83     Class :character   Class :character
 Median :   465   Median :  6.89     Mode  :character   Mode  :character
 Mean   : 10686   Mean   : 11.53
 3rd Qu.:  3490   3rd Qu.: 15.31
 Max.   :455582   Max.   :191.07
```

```
summary(CovidTest)
```

```
   Confirmed            Deaths          Recovered            Active
 Min.   :    23.0   Min.   :   0.0   Min.   :    11.0   Min.   :     0.0
 1st Qu.:   784.5   1st Qu.:  11.5   1st Qu.:   237.5   1st Qu.:   234.5
 Median :  4448.0   Median :  69.0   Median :  2905.0   Median :  1399.0
 Mean   : 34005.7   Mean   : 824.4   Mean   : 19693.5   Mean   : 13487.8
 3rd Qu.: 19266.5   3rd Qu.: 535.0   3rd Qu.: 13572.0   3rd Qu.:  6556.5
 Max.   :452529.0   Max.   :8777.0   Max.   :274925.0   Max.   :170537.0
   New.cases          New.deaths     New.recovered     Deaths...100.Cases
 Min.   :    0.0   Min.   :  0.00   Min.   :    0.0   Min.   : 0.000
 1st Qu.:    0.0   1st Qu.:  0.00   1st Qu.:    0.0   1st Qu.: 1.055
 Median :   44.0   Median :  0.00   Median :   24.0   Median : 2.290
 Mean   :  874.5   Mean   : 27.94   Mean   :  760.0   Mean   : 2.699
 3rd Qu.:  318.0   3rd Qu.:  4.00   3rd Qu.:  153.5   3rd Qu.: 3.460
 Max.   :16306.0   Max.   :508.00   Max.   :11494.0   Max.   :13.400
 Recovered...100.Cases Deaths...100.Recovered Confirmed.last.week
 Min.   :  5.48        Min.   : 0.000         Min.   :    19.0
 1st Qu.: 42.99        1st Qu.: 2.065         1st Qu.:   773.5
 Median : 64.60        Median : 3.650         Median :  3748.0
 Mean   : 62.15        Mean   : 5.300         Mean   : 28637.1
 3rd Qu.: 85.21        3rd Qu.: 7.085         3rd Qu.: 17950.0
 Max.   :100.00        Max.   :23.140         Max.   :373628.0
 X1.week.change  X1.week...increase Country.Region      WHO.Region
 Min.   :    0   Min.   :  0.000    Length:35          Length:35
 1st Qu.:   35   1st Qu.:  2.695    Class :character   Class :character
 Median :  401   Median :  6.420    Mode  :character   Mode  :character
 Mean   : 5369   Mean   : 22.531
 3rd Qu.: 2260   3rd Qu.: 24.535
 Max.   :78901   Max.   :226.320
```

```
model_1 <- lm(New.deaths ~ Active , data = CovidTraining[, 1:13])
summary(model_1)
```

```


Call:
lm(formula = New.deaths ~ Active, data = CovidTraining[, 1:13])

Residuals:
    Min      1Q  Median      3Q     Max
-150.90  -13.59  -13.37  -11.74  519.08

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.337e+01  6.175e+00   2.165    0.032 *
Active      4.309e-04  2.550e-05  16.894   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 73.9 on 145 degrees of freedom
Multiple R-squared:  0.6631,    Adjusted R-squared:  0.6608
F-statistic: 285.4 on 1 and 145 DF,  p-value: < 2.2e-16
```

```
intercept_only <- lm(New.deaths  ~ 1, data=CovidTraining[, 1:13])

all <- lm(New.deaths ~. , data = CovidTraining[, 1:13])

model_2 <- stepAIC (intercept_only, direction='forward',scope = formula(all))
```

```
Start:  AIC=1424.93
New.deaths ~ 1

                         Df Sum of Sq     RSS    AIC
+ New.cases               1   2082884  267838 1107.6
+ X1.week.change          1   1980913  369809 1155.1
+ Confirmed               1   1930012  420710 1174.0
+ Confirmed.last.week     1   1895844  454877 1185.5
+ New.recovered           1   1842292  508430 1201.8
+ Deaths                  1   1712065  638657 1235.4
+ Recovered               1   1695802  654920 1239.1
+ Active                  1   1558773  791949 1267.0
<none>                               2350722 1424.9
+ Recovered...100.Cases   1     16306 2334416 1425.9
+ Deaths...100.Cases      1     11217 2339505 1426.2
+ X1.week...increase      1      4136 2346586 1426.7
+ Deaths...100.Recovered  1      1208 2349514 1426.9

Step:  AIC=1107.63
New.deaths ~ New.cases

                         Df Sum of Sq    RSS    AIC
+ Deaths                  1     49138 218700 1079.8
+ Recovered               1     37331 230507 1087.6
+ Confirmed.last.week     1     25415 242423 1095.0
+ Confirmed               1     23991 243847 1095.8
+ New.recovered           1      9140 258698 1104.5
+ Deaths...100.Cases      1      5428 262410 1106.6
<none>                               267838 1107.6
+ X1.week.change          1      1951 265887 1108.6
+ Recovered...100.Cases   1       681 267157 1109.3
+ X1.week...increase      1       257 267581 1109.5
+ Deaths...100.Recovered  1       228 267610 1109.5
+ Active                  1        53 267785 1109.6

Step:  AIC=1079.84
New.deaths ~ New.cases + Deaths

                         Df Sum of Sq    RSS    AIC
+ Active                  1   21075.9 197624 1066.9
+ Recovered               1    9628.6 209072 1075.2
+ Deaths...100.Recovered  1    7648.4 211052 1076.6
+ X1.week.change          1    4371.5 214329 1078.9
+ New.recovered           1    4262.9 214438 1079.0
<none>                               218700 1079.8
+ Recovered...100.Cases   1    2778.3 215922 1080.0
+ Confirmed               1    1866.3 216834 1080.6
+ Confirmed.last.week     1    1400.1 217300 1080.9
+ Deaths...100.Cases      1     466.6 218234 1081.5
+ X1.week...increase      1     108.7 218592 1081.8

Step:  AIC=1066.94
```

```
New.deaths ~ New.cases + Deaths + Active

                         Df Sum of Sq    RSS    AIC
+ X1.week.change          1    9675.8 187949 1061.6
+ Deaths...100.Recovered  1    9060.0 188565 1062.0
+ New.recovered           1    3505.6 194119 1066.3
+ Deaths...100.Cases      1    3219.2 194405 1066.5
<none>                                197624 1066.9
+ Confirmed.last.week     1    1577.7 196047 1067.8
+ Recovered...100.Cases   1    1562.0 196063 1067.8
+ Confirmed               1     533.0 197092 1068.5
+ Recovered               1     533.0 197092 1068.5
+ X1.week...increase      1      87.2 197537 1068.9

Step:  AIC=1061.56
New.deaths ~ New.cases + Deaths + Active + X1.week.change

                         Df Sum of Sq    RSS    AIC
+ Confirmed               1   18355.1 169594 1048.5
+ Recovered               1   18355.1 169594 1048.5
+ Confirmed.last.week     1   18355.1 169594 1048.5
+ Deaths...100.Recovered  1   13695.1 174254 1052.4
+ New.recovered           1   10706.9 177242 1054.9
+ Deaths...100.Cases      1    8657.1 179292 1056.6
<none>                                187949 1061.6
+ Recovered...100.Cases   1    2431.3 185517 1061.7
+ X1.week...increase      1     235.7 187713 1063.4

Step:  AIC=1048.46
New.deaths ~ New.cases + Deaths + Active + X1.week.change + Confirmed

                         Df Sum of Sq    RSS    AIC
+ Deaths...100.Recovered  1    7787.3 161806 1043.5
+ New.recovered           1    6612.7 162981 1044.6
+ Deaths...100.Cases      1    4347.8 165246 1046.6
<none>                                169594 1048.5
+ X1.week...increase      1     834.0 168760 1049.7
+ Recovered...100.Cases   1     253.9 169340 1050.2

Step:  AIC=1043.55
New.deaths ~ New.cases + Deaths + Active + X1.week.change + Confirmed +
    Deaths...100.Recovered

                        Df Sum of Sq    RSS    AIC
+ New.recovered          1    8086.3 153720 1038.0
+ Deaths...100.Cases     1    2752.5 159054 1043.0
<none>                               161806 1043.5
+ X1.week...increase     1     654.6 161152 1045.0
+ Recovered...100.Cases  1      87.2 161719 1045.5

Step:  AIC=1038.01
New.deaths ~ New.cases + Deaths + Active + X1.week.change + Confirmed +
```

```
    Deaths...100.Recovered + New.recovered

                        Df Sum of Sq    RSS    AIC
<none>                                  153720 1038.0
+ Deaths...100.Cases     1   1733.99 151986 1038.3
+ X1.week...increase     1    777.80 152942 1039.3
+ Recovered...100.Cases  1    278.13 153442 1039.7
```

```
summary(model_2)
```

```
Call:
lm(formula = New.deaths ~ New.cases + Deaths + Active + X1.week.change +
    Confirmed + Deaths...100.Recovered + New.recovered, data = CovidTraining[,
    1:13])

Residuals:
     Min       1Q   Median       3Q      Max
-119.706   -0.268    2.754    3.284  200.974

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)            -2.934e+00  2.969e+00  -0.988 0.324699
New.cases               2.246e-02  2.187e-03  10.270  < 2e-16 ***
Deaths                  2.321e-03  6.121e-04   3.793 0.000222 ***
Active                 -1.875e-05  8.814e-05  -0.213 0.831874
X1.week.change         -3.452e-03  7.391e-04  -4.670 7.01e-06 ***
Confirmed               1.317e-04  4.734e-05   2.783 0.006142 **
Deaths...100.Recovered -2.357e-02  8.147e-03  -2.894 0.004421 **
New.recovered           1.842e-02  6.814e-03   2.704 0.007706 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 33.26 on 139 degrees of freedom
Multiple R-squared:  0.9346,    Adjusted R-squared:  0.9313
F-statistic: 283.8 on 7 and 139 DF,  p-value: < 2.2e-16
```

```
model_3 <- stepAIC (all, direction='backward')
```

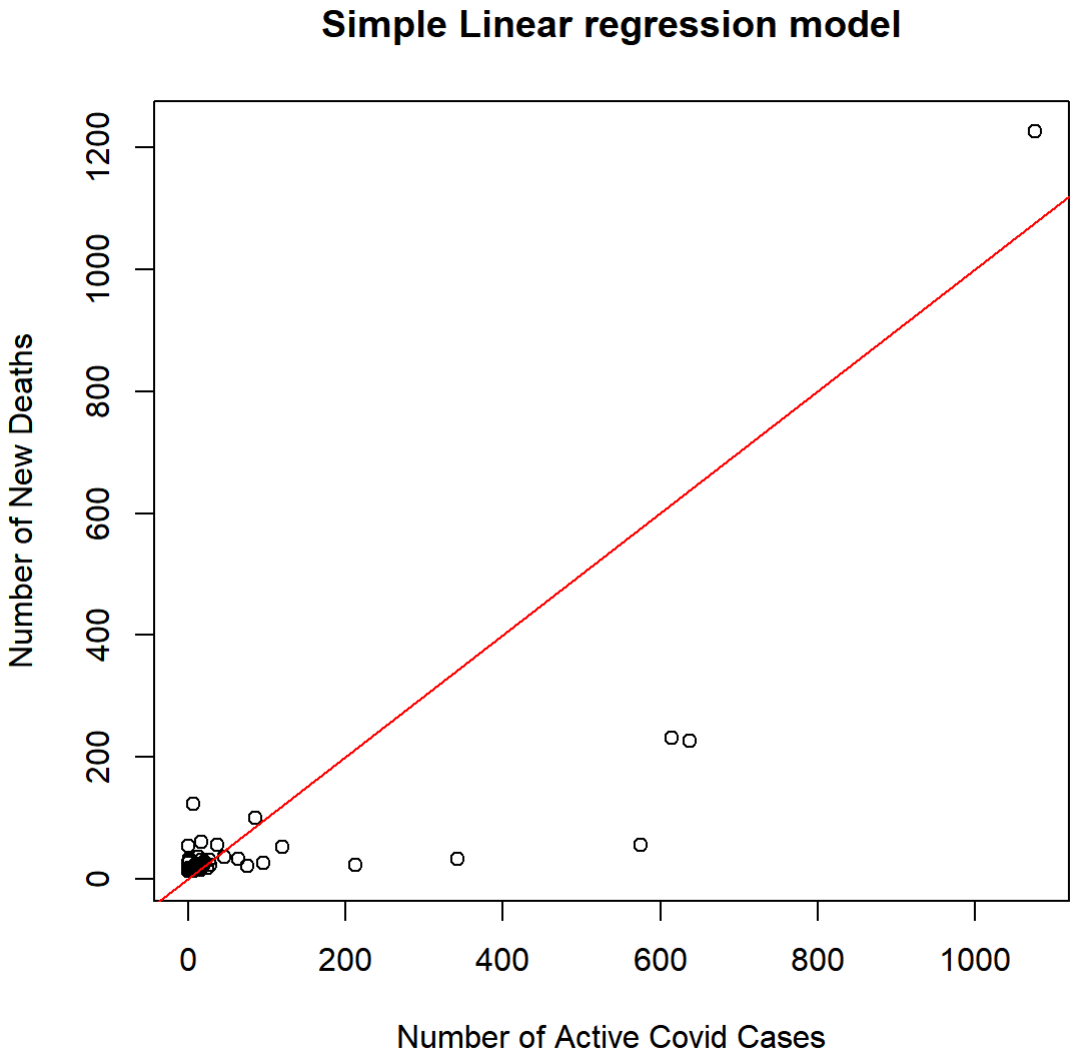
```
Start:  AIC=1041.63
New.deaths ~ Confirmed + Deaths + Recovered + Active + New.cases +
    New.recovered + Deaths...100.Cases + Recovered...100.Cases +
    Deaths...100.Recovered + Confirmed.last.week + X1.week.change +
    X1.week...increase


Step:  AIC=1041.63
New.deaths ~ Confirmed + Deaths + Recovered + Active + New.cases +
    New.recovered + Deaths...100.Cases + Recovered...100.Cases +
    Deaths...100.Recovered + Confirmed.last.week + X1.week...increase


Step:  AIC=1041.63
New.deaths ~ Confirmed + Deaths + Recovered + New.cases + New.recovered +
    Deaths...100.Cases + Recovered...100.Cases + Deaths...100.Recovered +
    Confirmed.last.week + X1.week...increase

                         Df Sum of Sq    RSS    AIC
- Recovered...100.Cases   1       130 151377 1039.8
- Recovered               1       195 151441 1039.8
- X1.week...increase      1       285 151532 1039.9
- Deaths...100.Cases      1      1671 152918 1041.2
<none>                               151247 1041.6
- New.recovered           1      7294 158541 1046.5
- Deaths...100.Recovered  1      7461 158708 1046.7
- Deaths                  1     15573 166820 1054.0
- Confirmed.last.week     1     23706 174953 1061.0
- Confirmed               1     27137 178384 1063.9
- New.cases               1    118377 269624 1124.6

Step:  AIC=1039.75
New.deaths ~ Confirmed + Deaths + Recovered + New.cases + New.recovered +
    Deaths...100.Cases + Deaths...100.Recovered + Confirmed.last.week +
    X1.week...increase

                         Df Sum of Sq    RSS    AIC
- Recovered               1       182 151559 1037.9
- X1.week...increase      1       609 151986 1038.3
- Deaths...100.Cases      1      1565 152942 1039.3
<none>                               151377 1039.8
- New.recovered           1      7206 158583 1044.6
- Deaths...100.Recovered  1      7586 158963 1044.9
- Deaths                  1     15686 167063 1052.2
- Confirmed.last.week     1     23582 174959 1059.0
- Confirmed               1     27026 178403 1061.9
- New.cases               1    118262 269639 1122.6

Step:  AIC=1037.93
New.deaths ~ Confirmed + Deaths + New.cases + New.recovered +
    Deaths...100.Cases + Deaths...100.Recovered + Confirmed.last.week +
    X1.week...increase
```

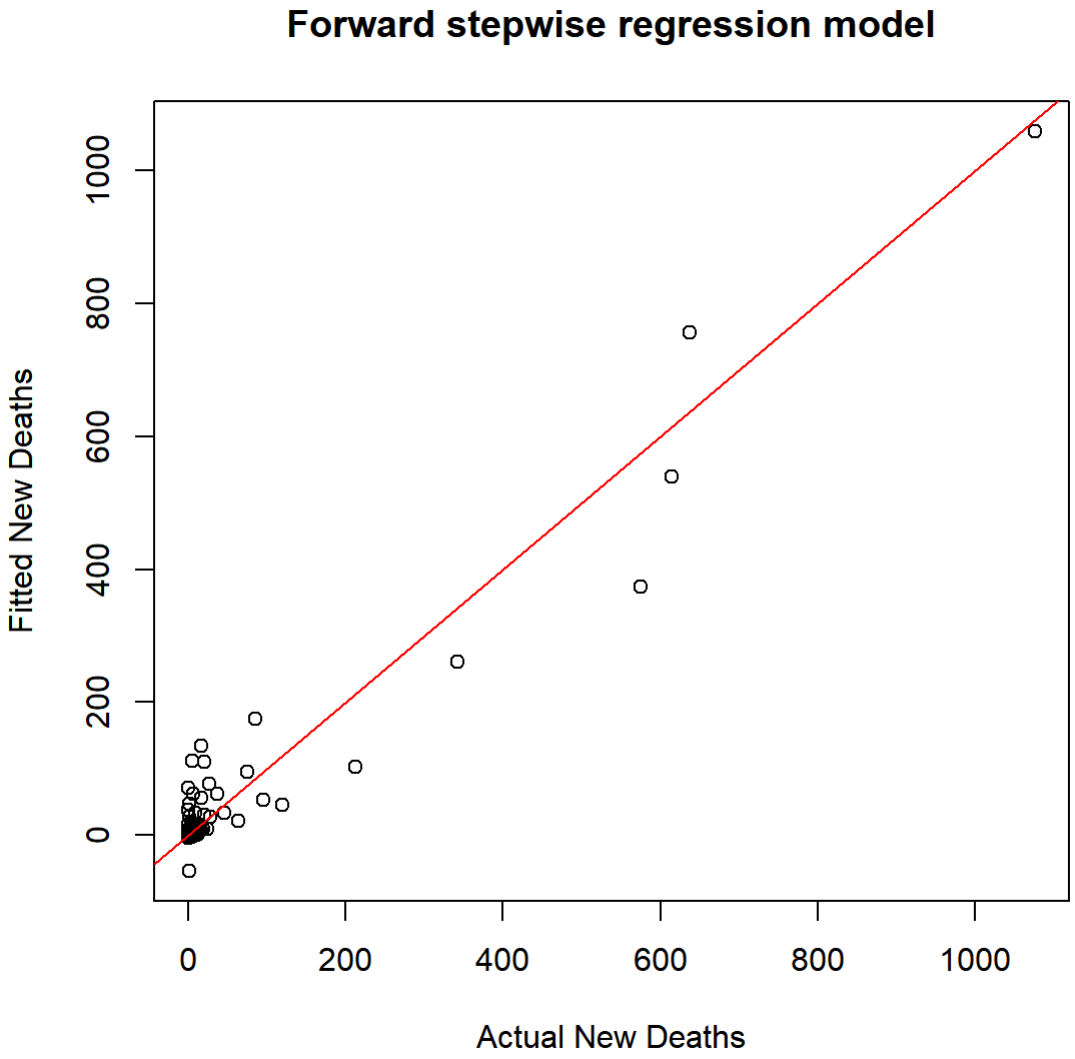
```
                         Df Sum of Sq    RSS    AIC
- X1.week...increase      1       611 152170 1036.5
- Deaths...100.Cases      1      1437 152996 1037.3
<none>                                151559 1037.9
- Deaths...100.Recovered  1      8101 159660 1043.6
- Deaths                  1     17360 168919 1051.9
- Confirmed.last.week     1     49140 200699 1077.2
- New.recovered           1     50432 201991 1078.2
- Confirmed               1     51251 202810 1078.8
- New.cases               1    162767 314325 1143.2

Step:  AIC=1036.52
New.deaths ~ Confirmed + Deaths + New.cases + New.recovered + 
    Deaths...100.Cases + Deaths...100.Recovered + Confirmed.last.week

                         Df Sum of Sq    RSS    AIC
- Deaths...100.Cases      1      1600 153770 1036.1
<none>                                152170 1036.5
- Deaths...100.Recovered  1      8202 160372 1042.2
- Deaths                  1     17619 169789 1050.6
- Confirmed.last.week     1     48530 200700 1075.2
- New.recovered           1     49887 202057 1076.2
- Confirmed               1     50641 202812 1076.8
- New.cases               1    162570 314740 1141.3

Step:  AIC=1036.06
New.deaths ~ Confirmed + Deaths + New.cases + New.recovered + 
    Deaths...100.Recovered + Confirmed.last.week

                         Df Sum of Sq    RSS    AIC
<none>                                153770 1036.1
- Deaths...100.Recovered  1      9568 163338 1042.9
- Deaths                  1     16665 170435 1049.2
- Confirmed.last.week     1     47730 201500 1073.8
- New.recovered           1     49198 202968 1074.9
- Confirmed               1     49563 203333 1075.1
- New.cases               1    160975 314745 1139.4
```

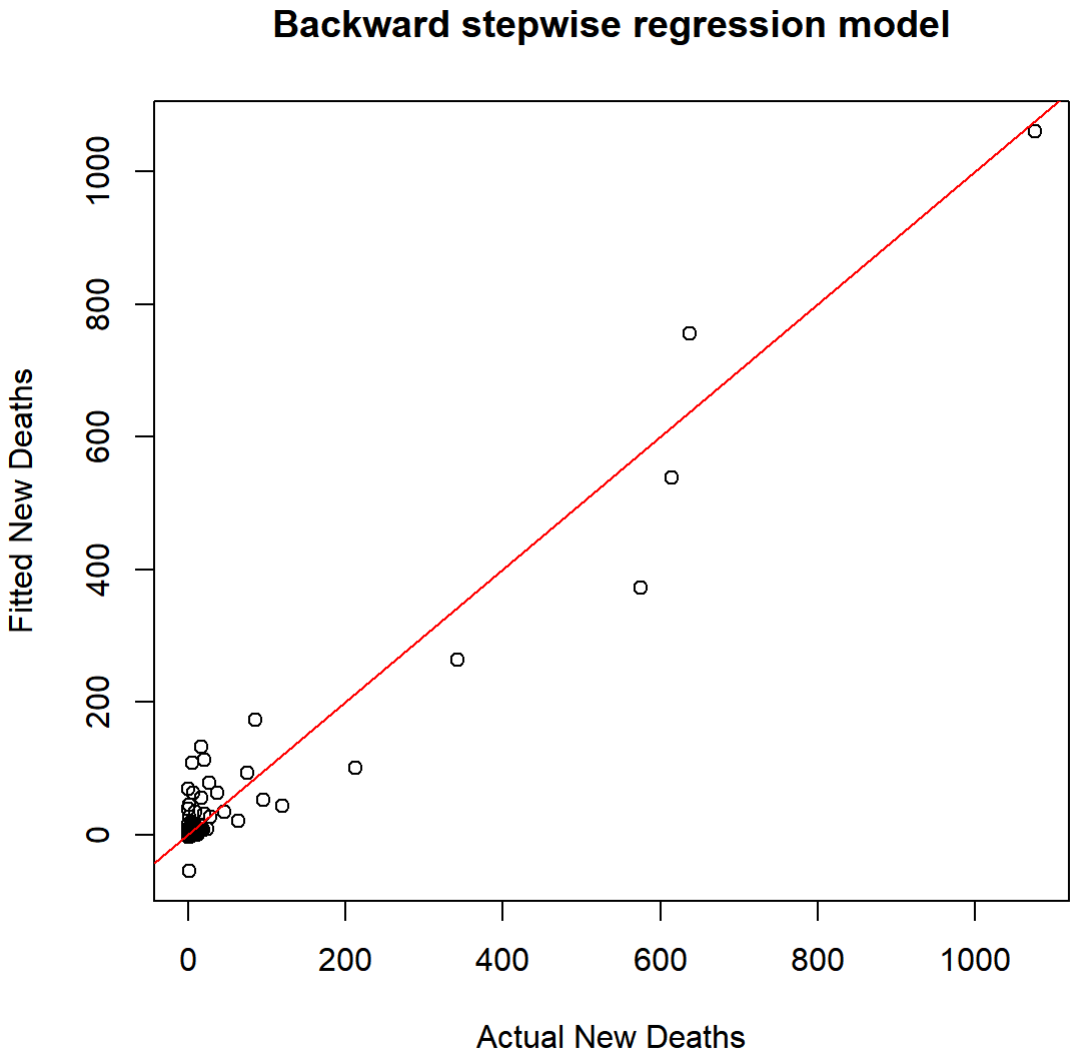
```
summary(model_3)
```

```
Call:
lm(formula = New.deaths ~ Confirmed + Deaths + New.cases + New.recovered +
    Deaths...100.Recovered + Confirmed.last.week, data = CovidTraining[,
    1:13])

Residuals:
     Min       1Q   Median       3Q      Max
-119.040   -0.223    2.734    3.230  202.250

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)            -2.8654480  2.9410619  -0.974 0.331593
Confirmed              -0.0034316  0.0005108  -6.717 4.29e-10 ***
Deaths                  0.0022858  0.0005868   3.895 0.000151 ***
New.cases               0.0222062  0.0018343  12.106  < 2e-16 ***
New.recovered           0.0197297  0.0029479   6.693 4.87e-10 ***
Deaths...100.Recovered -0.0237851  0.0080589  -2.951 0.003710 **
Confirmed.last.week     0.0035587  0.0005398   6.592 8.19e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 33.14 on 140 degrees of freedom
Multiple R-squared:  0.9346,    Adjusted R-squared:  0.9318
F-statistic: 333.4 on 6 and 140 DF,  p-value: < 2.2e-16
```

```
{
plot(CovidTraining$New.deaths,fitted(model_1),xlab = "Number of Active Covid Cases", ylab = "Num
ber of New Deaths", main = "Simple Linear regression model")

abline(0, 1, col = "red")

plot(CovidTraining$New.deaths,fitted(model_2),xlab = "Actual New Deaths", ylab = "Fitted New Dea
ths", main = "Forward stepwise regression model")

abline(0, 1, col = "red")

plot(CovidTraining$New.deaths,fitted(model_3),xlab = "Actual New Deaths", ylab = "Fitted New Dea
ths", main = "Backward stepwise regression model")

abline(0, 1, col = "red")
}
```


Simple Linear regression model
Number of New Deaths
0
200
400
600
800
1000
1200
Number of Active Covid Cases
0
200
400
600
800
1000


Forward stepwise regression model
Fitted New Deaths
0
200
400
600
800
1000
Actual New Deaths
0
200
400
600
800
1000


Backward stepwise regression model
Fitted New Deaths
0
200
400
600
800
1000
Actual New Deaths
0
200
400
600
800
1000

```
aic_values <- c(
  AIC(model_1),
  AIC(model_2),
  AIC(model_3)
)

adjusted_r_squared <- c(
  summary(model_1)$adj.r.squared,
  summary(model_2)$adj.r.squared,
  summary(model_3)$adj.r.squared
)

predictions_model_1 <- predict(model_1, newdata = CovidTest)
predictions_model_2 <- predict(model_2, newdata = CovidTest)
predictions_model_3 <- predict(model_3, newdata = CovidTest)

mae_values <- c(
  MAE(y_pred = predictions_model_1, y_true = CovidTest$New.deaths),
  MAE(y_pred = predictions_model_2, y_true = CovidTest$New.deaths),
  MAE(y_pred = predictions_model_3, y_true = CovidTest$New.deaths)
)

mse_values <- c(
  MSE(y_pred = predictions_model_1, y_true = CovidTest$New.deaths),
  MSE(y_pred = predictions_model_2, y_true = CovidTest$New.deaths),
  MSE(y_pred = predictions_model_3, y_true = CovidTest$New.deaths)
)

evaluation_table <- data.frame(
  Model = c("Simple Linear Regression", "Forward Stepwise", "Backward Stepwise"),
  AIC = aic_values,
  Adjusted_R_Squared = adjusted_r_squared,
  MAE = mae_values,
  MSE = mse_values
)
evaluation_table
```

```
                    Model      AIC Adjusted_R_Squared      MAE       MSE
1 Simple Linear Regression 1686.165          0.6607806 32.03055 7113.5543
2         Forward Stepwise 1457.179          0.9313142 12.38794  965.9000
3        Backward Stepwise 1455.227          0.9317826 12.08118  912.4351
```