```
# Read the data set for the cases in all countries
library(readr)
time_series_covid19_confirmed_global <- read_csv("../input/countries-
data/time_series_covid19_confirmed_global.csv")
time_series_covid19_confirmed_global$t<-as.Date(time_series_covid19_confirmed_global$t,format =
"%m/%d/%Y")
```

```
── Column specification ────────────────────────────────────────────────────────
cols(
  t = col_character(),
  Argentina_C = col_double(),
  Brazil_C = col_double(),
  Mexico_C = col_double(),
  US_C = col_double()
)
```

add Codeadd Markdown

[50]:

```
# Read the index from Brazil
brazil_january_october_2020 <- read_csv("../input/countries-data/brazil_--january-october-2020.csv")
names(brazil_january_october_2020)=c("t","Brazil_Index")
brazil_january_october_2020$t=as.Date(brazil_january_october_2020$t,format = "%m/%d/%Y")
```

```
── Column specification ────────────────────────────────────────────────────────
cols(
  x1 = col_character(),
  x2 = col_number()
)
```

add Codeadd Markdown

[51]:

```
# Read the index from Argentina
argentina_merval_stock_market_index_2020 <- read_csv("../input/countries-data/argentina_-merval-stock-
market-index-2020.csv")
names(argentina_merval_stock_market_index_2020)=c("t","Argentina_Index")
argentina_merval_stock_market_index_2020$t=as.Date(argentina_merval_stock_market_index_2020$t,form
at = "%m/%d/%Y")
```

```
── Column specification ────────────────────────────────────────────────────────
cols(
  x1 = col_character(),
  x2 = col_number()
)
```

add Codeadd Markdown

[52]:

```
# Read the index from Mexico
mexico_ipc_stock_market_index_january_october_2020 <- read_csv("../input/countries-data/mexico_-ipc-
stock-market-index-january-october-2020.csv")
names(mexico_ipc_stock_market_index_january_october_2020)=c("t","Mexico_Index")
mexico_ipc_stock_market_index_january_october_2020$t=as.Date(mexico_ipc_stock_market_index_january_
october_2020$t,format = "%m/%d/%Y")
```

Column specification

```
cols(
  x1 = col_character(),
  x2 = col_number()
)
```

add Codeadd Markdown

[53]:

```
# Read the weekly index from US
weekly_dow_jones_industrial_average_index_performance_2020_2021 <- read_csv("../input/countries-data/weekly-dow-jones-industrial-average-index-performance-2020-2021.csv")
names(weekly_dow_jones_industrial_average_index_performance_2020_2021)=c("t","US_Index")
weekly_dow_jones_industrial_average_index_performance_2020_2021$t=as.Date(weekly_dow_jones_industrial_average_index_performance_2020_2021$t,format = "%m/%d/%Y")
```

Column specification

```
cols(
  x1 = col_character(),
  x2 = col_number()
)
```

add Codeadd Markdown

[54]:

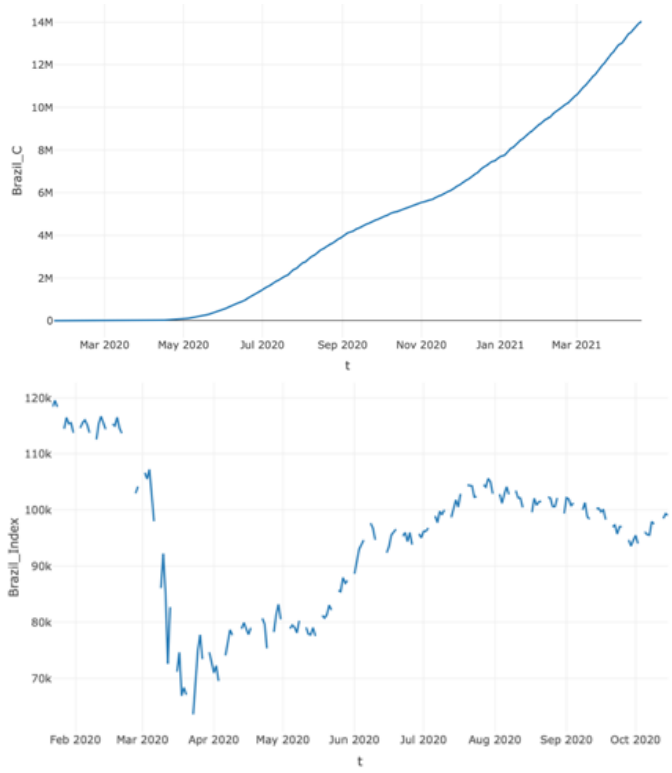
```
# Join the indexes with the data
library(dplyr)
ALL_data=left_join(time_series_covid19_confirmed_global, brazil_january_october_2020,by="t")
ALL_data=left_join(ALL_data, argentina_merval_stock_market_index_2020,by="t")
ALL_data=left_join(ALL_data, mexico_ipc_stock_market_index_january_october_2020,by="t")
ALL_data=left_join(ALL_data, weekly_dow_jones_industrial_average_index_performance_2020_2021,by="t")
```

add Codeadd Markdown

[55]:

```
library(plotly)
fig <- plot_ly(data=ALL_data,x = ~t, y = ~Brazil_C, mode = 'lines')
fig

fig <- plot_ly(data=ALL_data,x = ~t, y = ~Brazil_Index, mode = 'lines')
fig
```

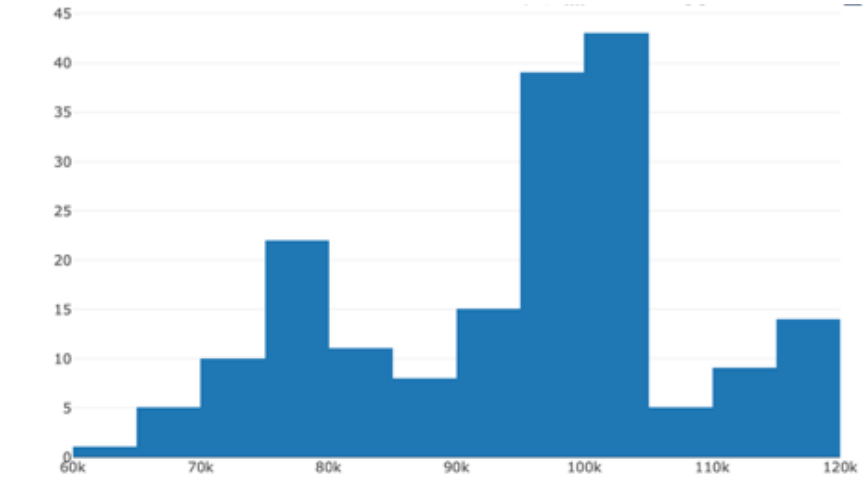
```
with the increase of the COVID cases number.
```



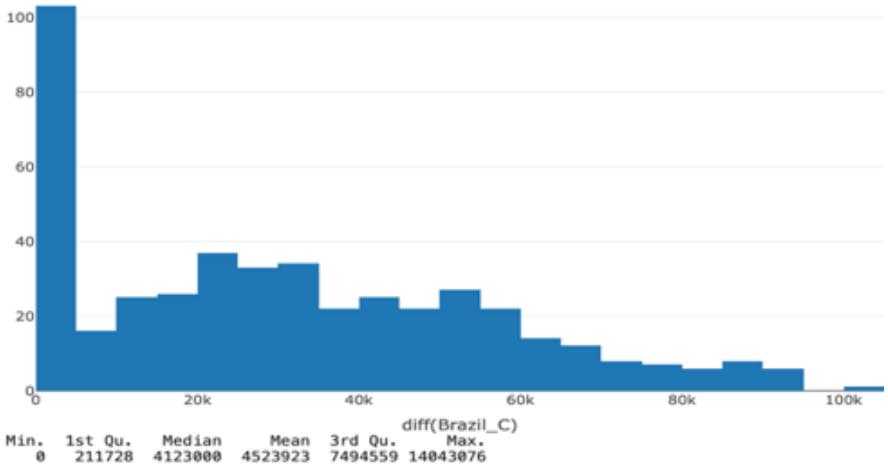
There are missing values for the daily stock market index

```
fig <- plot_ly(data=ALL_data,x = ~Brazil_Index, type = "histogram")
fig

summary(ALL_data$Brazil_Index)
```

```
fig <- plot_ly(data=ALL_data,x = ~ diff(Brazil_C), type = "histogram")
fig

summary(ALL_data$Brazil_C)
```




```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
63570   82299   97006   94640  102218  119528     273
```



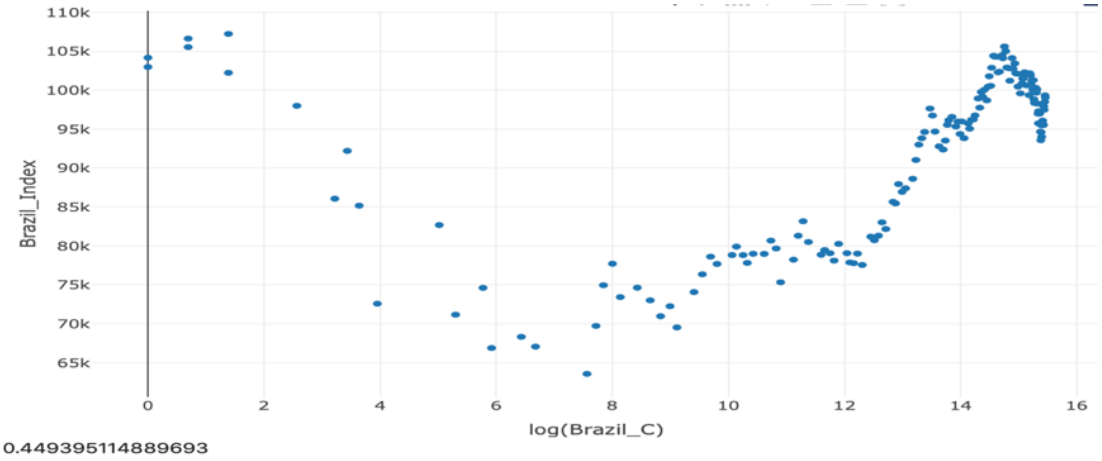

```
Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
   0   211728  4123000  4523923  7494559 14043076
```

For the index, the mean value is 94640 , and it has a multi modal distribution. There are not outliers poins. Casaes is a timeseries, is increasing with time and has an exponential pattern.

As the cases increases in exponetial pattern, we need to discover the behaviour of with the index, so we will take the log transform for the number of cases. And so we are going to exclude the dates of zero cases.

```
fig <- plot_ly(data=ALL_data[ALL_data$Brazil_C>0,],x = ~log(Brazil_C), y = ~Brazil_Index)
fig
```

```
cor(log(ALL_data$Brazil_C[ALL_data$Brazil_C>0]),ALL_data$Brazil_Index[ALL_data$Brazil_C>0],use="complete.obs")
```



0.449395114889693

The correlataion seesms to be positive after a certain value of the cases.

```
mod=lm(data = ALL_data[log(ALL_data$Brazil_C)>8,], formula = Brazil_Index~log(Brazil_C),na.actio
n=na.omit)
summary(mod)

plot(mod)
```

Model:

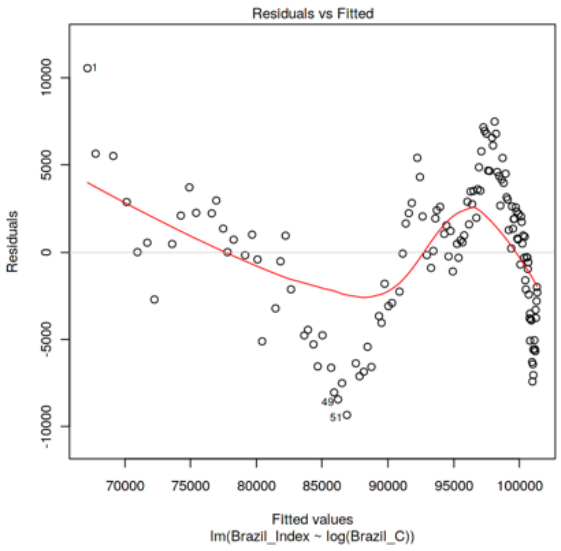
Call:
lm(formula = Brazil_Index ~ log(Brazil_C), data = ALL_data[log(ALL_data$Brazil_C) >
  8,], na.action = na.omit)

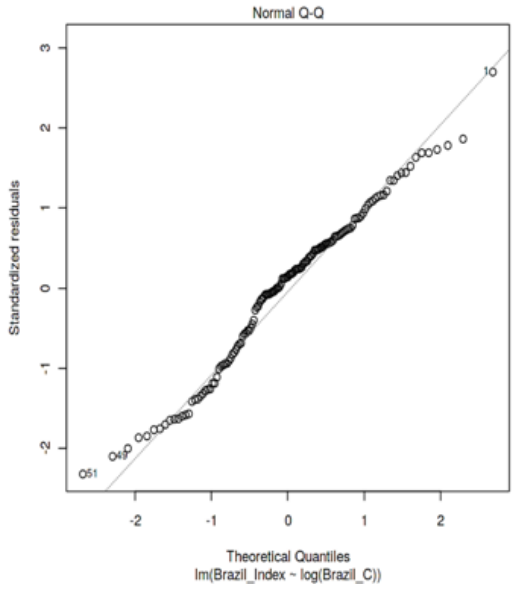
Residuals:
   Min   1Q Median   3Q   Max
-9330.4 -3002.0   561.2 2643.5 10546.0

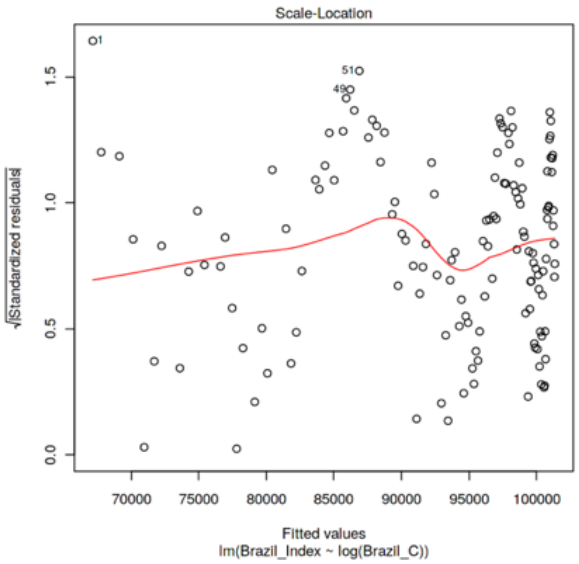
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   30474.7   2372.9   12.84 <2e-16 ***
log(Brazil_C) 4585.4   173.4   26.44 <2e-16 ***

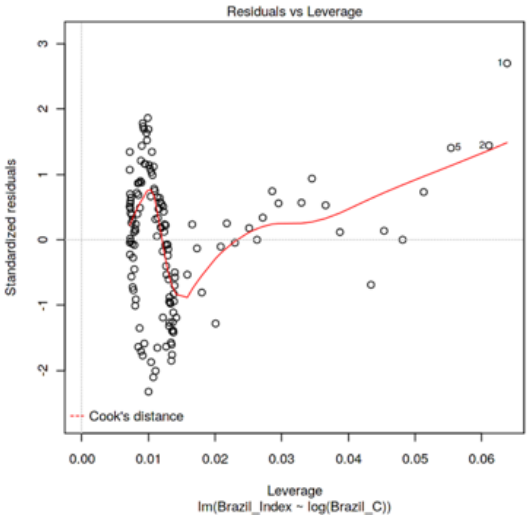
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4036 on 137 degrees of freedom
  (252 observations deleted due to missingness)
Multiple R-squared: 0.8361,   Adjusted R-squared: 0.8349
F-statistic: 699.1 on 1 and 137 DF, p-value: < 2.2e-16



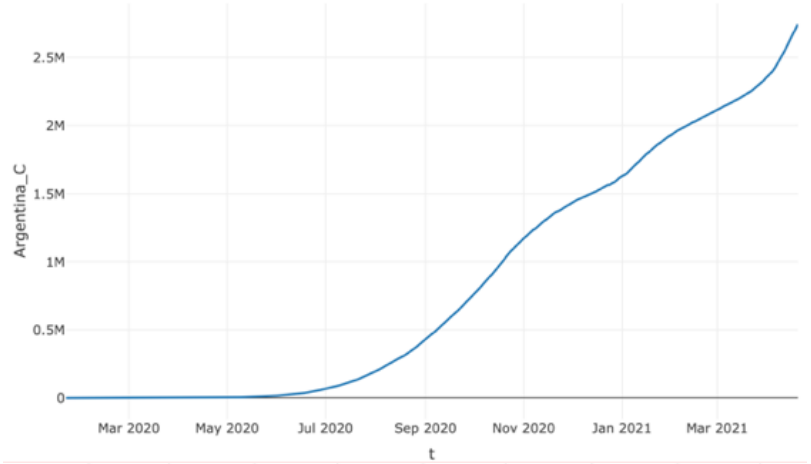


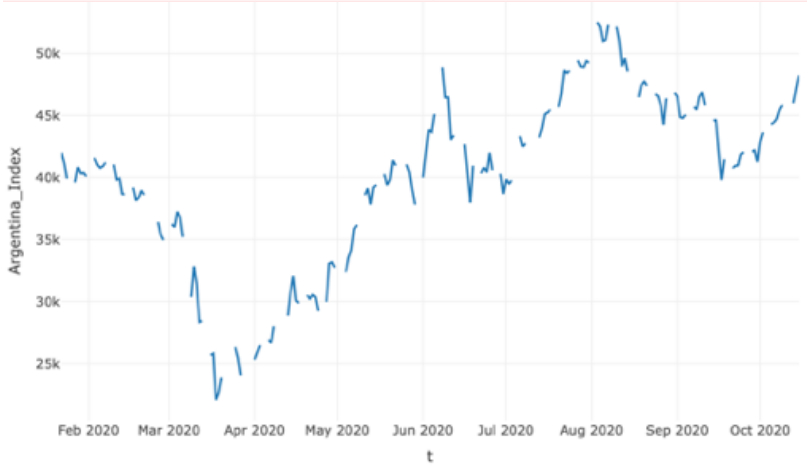




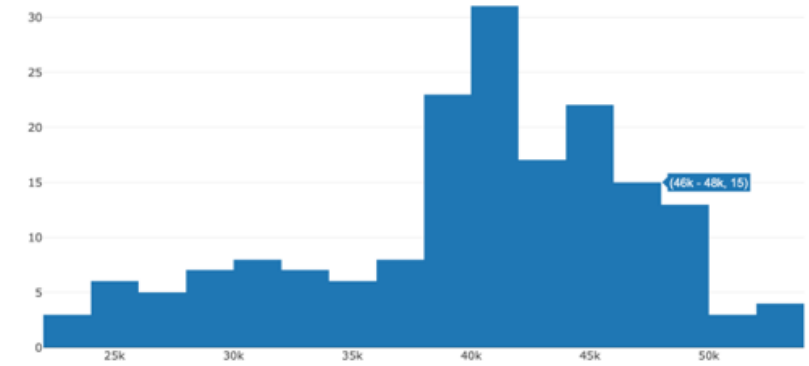
Data analysis for Argentina

```
library(plotly)
fig <- plot_ly(data=ALL_data,x = ~t, y = ~Argentina_C, mode = 'lines')
fig

fig <- plot_ly(data=ALL_data,x = ~t, y = ~Argentina_Index, mode = 'lines')
fig
```

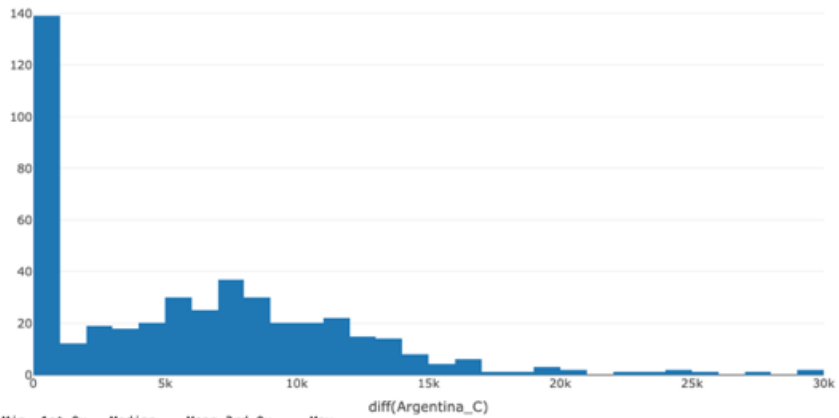





```
fig <- plot_ly(data=ALL_data,x = ~Argentina_Index, type = "histogram")
fig
summary(ALL_data$Argentina_Index)

fig <- plot_ly(data=ALL_data,x = ~diff(Argentina_C), type = "histogram")
fig
summary(ALL_data$Argentina_C)
```



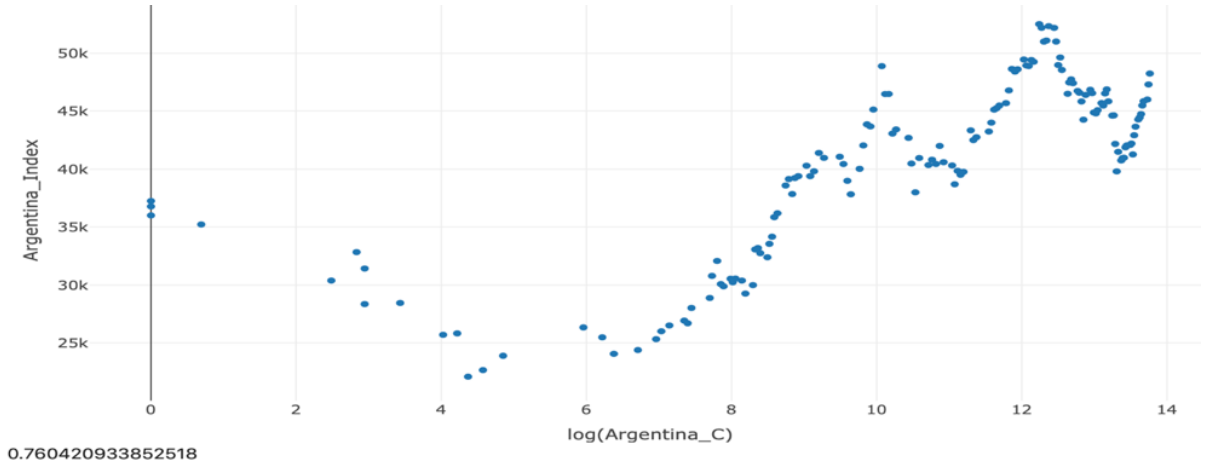

```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 22087   36510   40925   40037   45216   52504     277
```




```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0    7306  471806  834658 1586905 2743620
```

```
fig <- plot_ly(data=ALL_data[ALL_data$Argentina_C>0,],x = ~log(Argentina_C), y =
~Argentina_Index)
fig
```

```
cor(log(ALL_data$Argentina_C[ALL_data$Argentina_C>0]),ALL_data$Argentina_Index[ALL_data$Argentin
a_C>0],use="complete.obs")
```



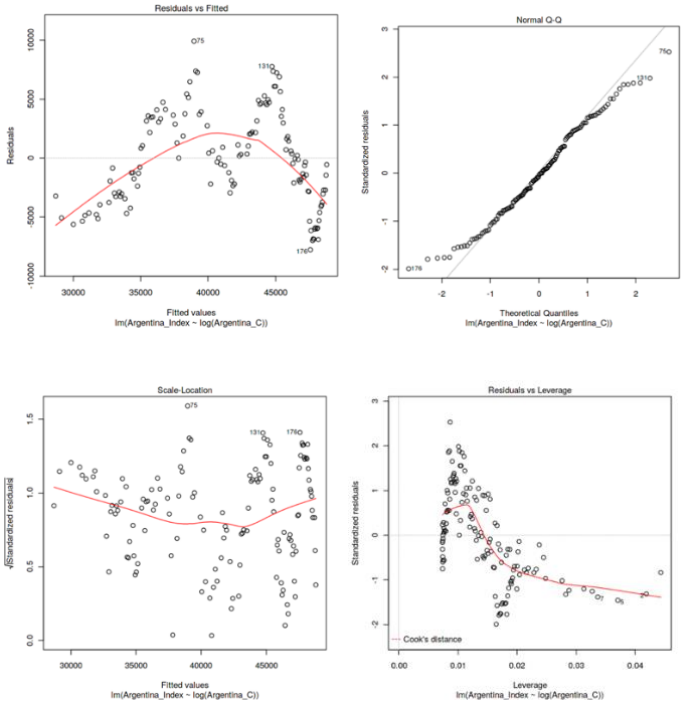

```
0.760420933852518
```

```
mod=lm(data = ALL_data[log(ALL_data$Argentina_C)>6,], formula =
Argentina_Index~log(Argentina_C),na.action=na.omit)
summary(mod)
```

```
plot(mod)
```

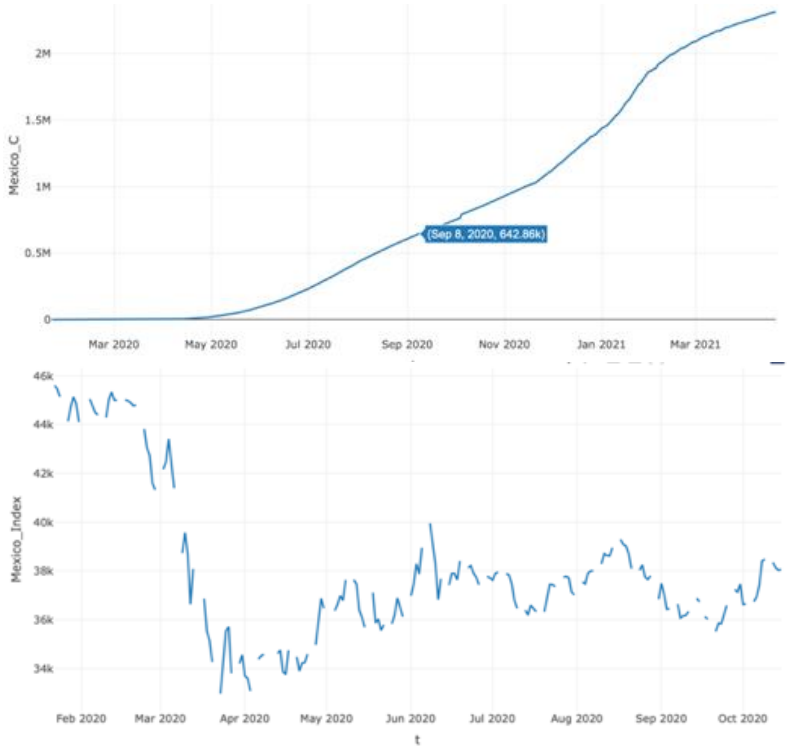
Model:

```
Call:
lm(formula = Argentina_Index ~ log(Argentina_C), data = ALL_data[log(ALL_data$Argentina_C) >
    6, ], na.action = na.omit)

Residuals:
    Min      1Q  Median      3Q     Max
-7778.2 -2951.7  -225.2  3206.3  9925.8

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)        12137       1786   6.798 3.19e-10 ***
log(Argentina_C)    2663        160  16.644  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3944 on 134 degrees of freedom
  (255 observations deleted due to missingness)
Multiple R-squared:  0.674,     Adjusted R-squared:  0.6716
F-statistic:   277 on 1 and 134 DF,  p-value: < 2.2e-16
```

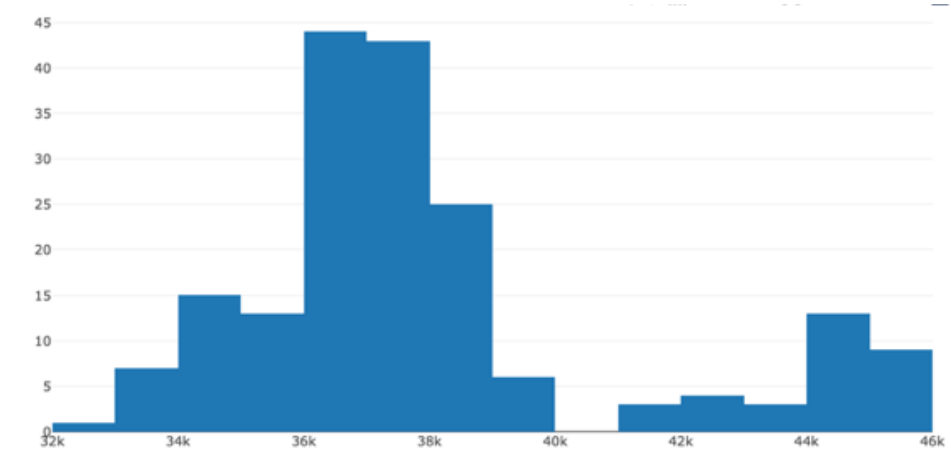


Data analysis for Mexico

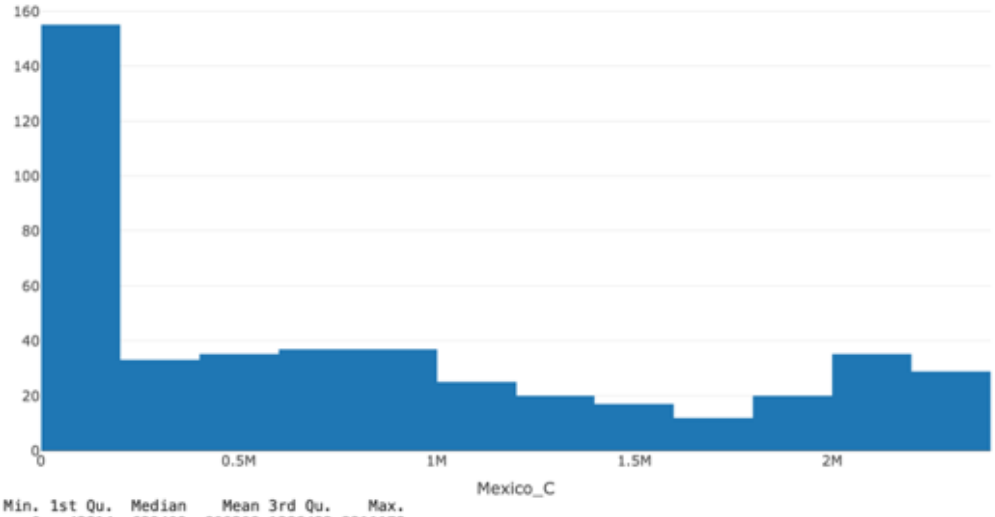
```
library(plotly)
fig <- plot_ly(data=ALL_data,x = ~t, y = ~Mexico_C, mode = 'lines')
fig

fig <- plot_ly(data=ALL_data,x = ~t, y = ~Mexico_Index, mode = 'lines')
fig
```




```
fig <- plot_ly(data=ALL_data,x = ~Mexico_Index, type = "histogram")
fig
summary(ALL_data$Mexico_Index)

fig <- plot_ly(data=ALL_data,x = ~Mexico_C, type = "histogram")
fig
summary(ALL_data$Mexico_C)
```



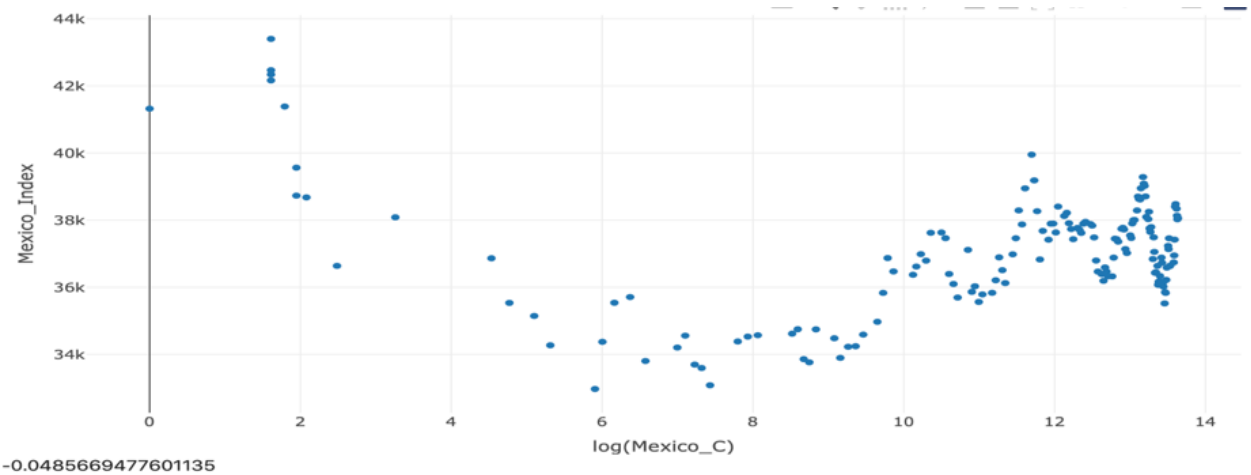

```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
32964   36244   37460   38064   38460   45604     269
```




```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0   43814  629409  809208 1386432 2311172
```

```
fig <- plot_ly(data=ALL_data[ALL_data$Mexico_C>0,],x = ~log(Mexico_C), y = ~Mexico_Index)
fig
```

```
cor(log(ALL_data$Mexico_C[ALL_data$Mexico_C>0]),ALL_data$Mexico_Index[ALL_data$Mexico_C>0],use="
complete.obs")
```




```
-0.0485669477601135
```

```
mod=lm(data = ALL_data[(ALL_data$Mexico_C)>6,], formula =
Mexico_Index~log(Mexico_C),na.action=na.omit)
summary(mod)

plot(mod)

Model:
```

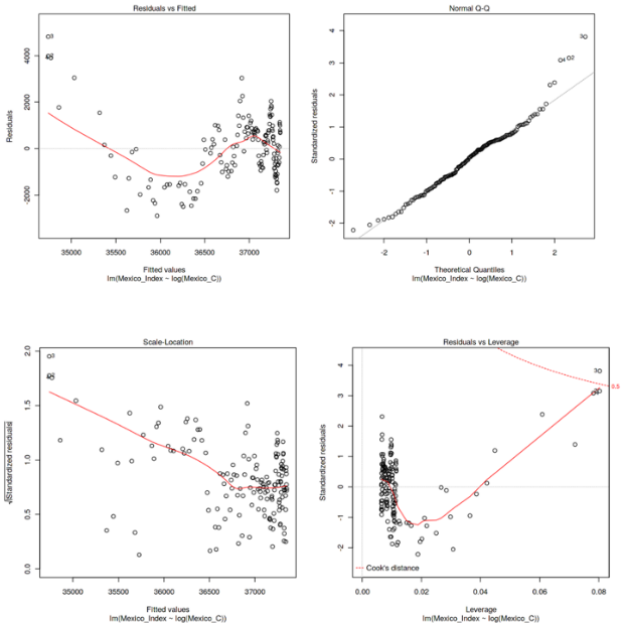
Call:
lm(formula = Mexico_Index ~ log(Mexico_C), data = ALL_data[(ALL_data$Mexico_C) >
  6,], na.action = na.omit)

Residuals:
  Min 1Q Median 3Q Max
-2889.0 -869.4 56.7 787.3 4823.4

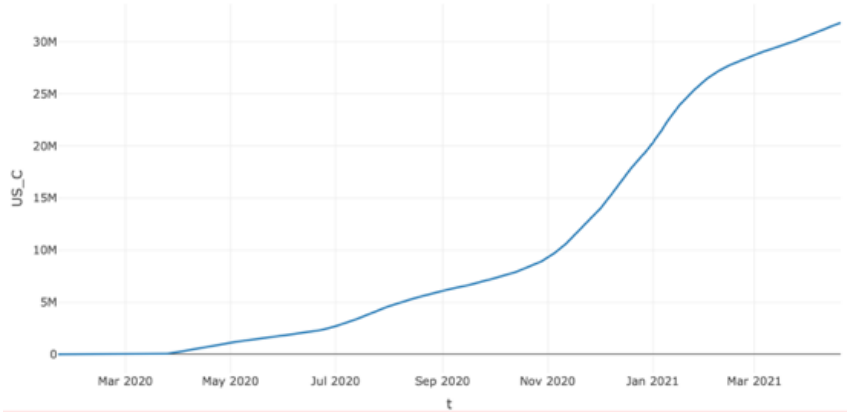
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 34308.46 446.33 76.867 < 2e-16 ***
log(Mexico_C) 222.83 38.64 5.767 4.37e-08 ***

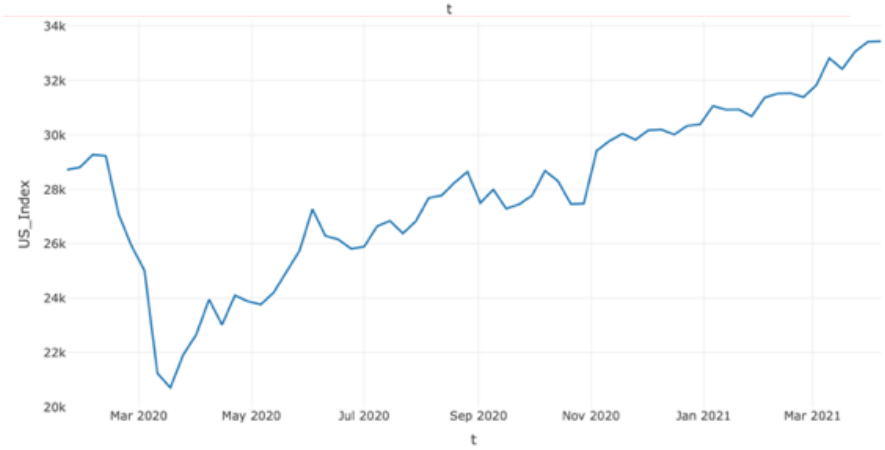
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1319 on 152 degrees of freedom
  (255 observations deleted due to missingness)
Multiple R-squared: 0.1795, Adjusted R-squared: 0.1741
F-statistic: 33.26 on 1 and 152 DF, p-value: 4.368e-08

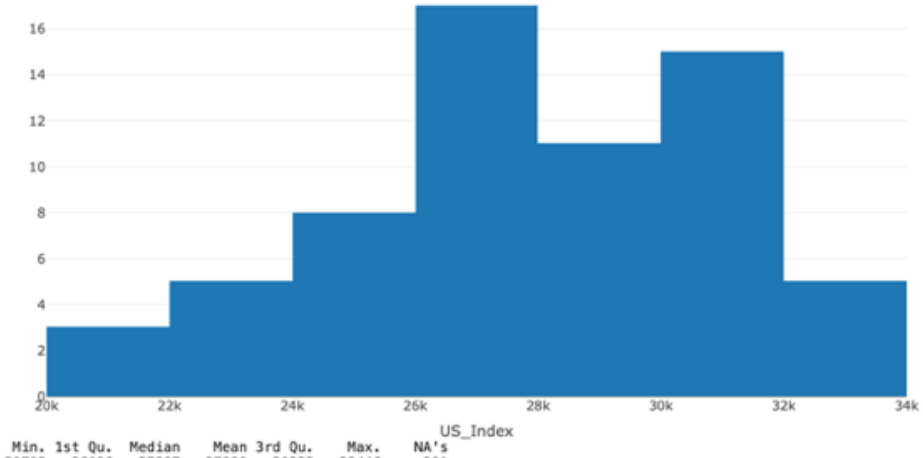



```
# Data analysis for US

library(plotly)
fig <- plot_ly(data=ALL_data,x = ~t, y = ~US_C, mode = 'lines')
fig

fig <- plot_ly(data=ALL_data[!is.na(ALL_data$US_Index),],x = ~t, y =
~US_Index, mode = 'lines')
fig
```

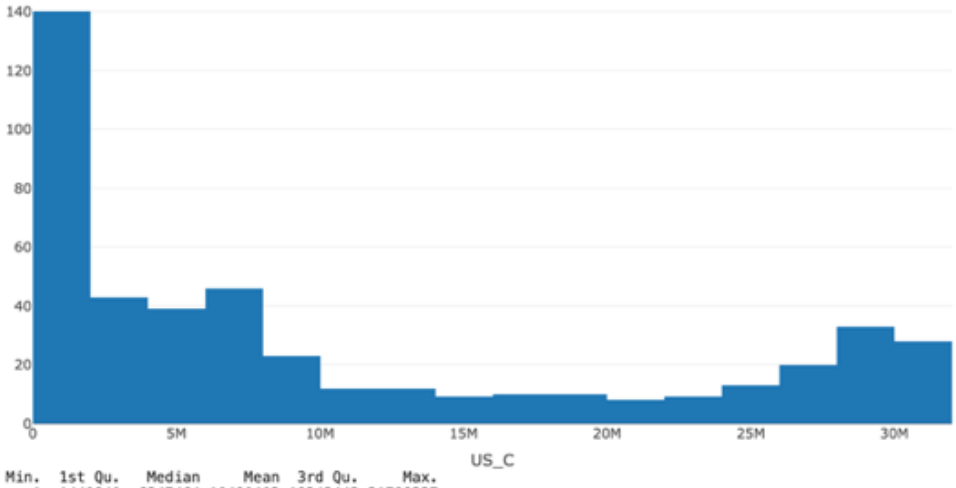





```
fig <- plot_ly(data=ALL_data,x = ~US_Index, type = "histogram")
fig
summary(ALL_data$US_Index)

fig <- plot_ly(data=ALL_data,x = ~US_C, type = "histogram")
fig
summary(ALL_data$US_C)
```



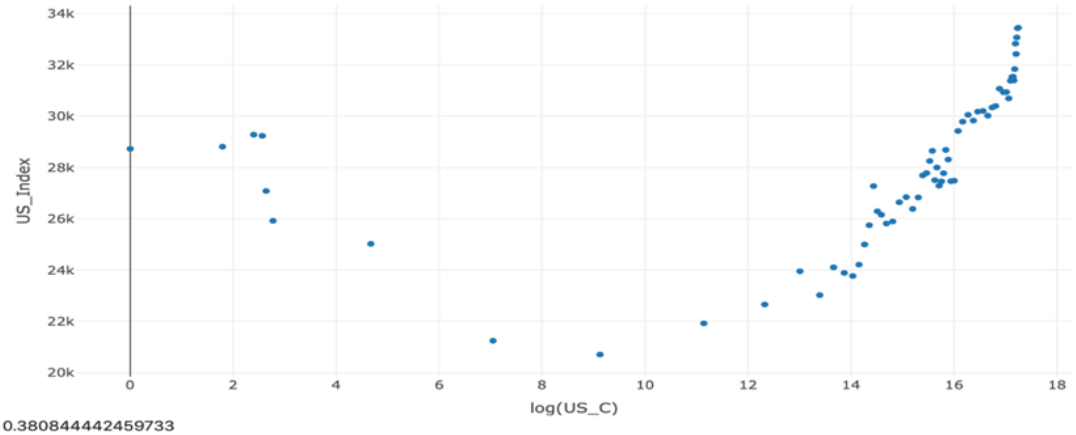

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  20705   26096   27887   27989   30233   33446     391
```




```
   Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
      1  1440840  6247464 10490403 19342443 31799237
```

```
fig <- plot_ly(data=ALL_data[ALL_data$US_C>0,],x = ~log(US_C), y = ~US_Index)
fig
```

```
cor(log(ALL_data$US_C[ALL_data$US_C>0]),ALL_data$US_Index[ALL_data$US_C>0],use="complete
.obs")
```




```
0.380844442459733
```

```
mod=lm(data = ALL_data[(ALL_data$US_C)>8,], formula = US_Index~log(US_C),na.action=na.omit)
summary(mod)

plot(mod)
Model:
Call:
lm(formula = US_Index ~ log(US_C), data = ALL_data[(ALL_data$US_C) >
   8, ], na.action = na.omit)

Residuals:
   Min     1Q Median     3Q    Max
-5258.5 -1850.3  -195.3  1975.2  5878.2

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 22483.54    1350.61  16.647  < 2e-16 ***
log(US_C)     381.42      90.79   4.201 8.95e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2755 on 60 degrees of freedom
  (381 observations deleted due to missingness)
Multiple R-squared:  0.2273,    Adjusted R-squared:  0.2144
F-statistic: 17.65 on 1 and 60 DF,  p-value: 8.946e-05
```