

Covid Cases - EDA

Katie Schilling - 501130072

07/02/2022

Required tools to be loaded

```
library(dlookr)
```

```
##  
## Attaching package: 'dlookr'  
  
## The following object is masked from 'package:base':  
##  
##   transform
```

```
library(lubridate)
```

```
##  
## Attaching package: 'lubridate'  
  
## The following objects are masked from 'package:base':  
##  
##   date, intersect, setdiff, union
```

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

```
##  
## Attaching package: 'data.table'  
  
## The following objects are masked from 'package:lubridate':  
##  
##   hour, isoweek, mday, minute, month, quarter, second, wday, week,  
##   yday, year
```

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'  
  
## The following objects are masked from 'package:data.table':  
##  
##   between, first, last
```

```
## The following objects are masked from 'package:stats':  
##  
##   filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library(moments)
```

```
##  
## Attaching package: 'moments'
```

```
## The following objects are masked from 'package:dlookr':  
##  
##   kurtosis, skewness
```

```
library(ggpubr)
```

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

```
library(smooth)
```

```
## Loading required package: greybox
```

```
## Package "greybox", v1.0.4 loaded.
```

```
##  
## Attaching package: 'greybox'
```

```
## The following object is masked from 'package:lubridate':  
##  
##   hm
```

```
## This is package "smooth", v3.1.5
```

```
library(greybox)  
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':  
##   method           from  
##   as.zoo.data.frame zoo
```

```
##  
## Attaching package: 'forecast'
```

```
## The following object is masked from 'package:greybox':  
##  
##   forecast
```

```
## The following object is masked from 'package:ggpubr':  
##  
##     gghistogram
```

```
library(funModeling)
```

```
## Loading required package: Hmisc
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
##
```

```
## Attaching package: 'Hmisc'
```

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

```
##
```

```
##     src, summarize
```

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

```
##
```

```
##     describe
```

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

```
##
```

```
##     format.pval, units
```

```
## funModeling v.1.9.4 :)
```

```
## Examples and tutorials at livebook.datascienceheroes.com
```

```
## / Now in Spanish: librovivodecienciadedatos.ai
```

Import dataset in regards to COVID-19

```
Covid_monthly <- read.csv("C:/Users/Katie Schilling/Downloads/covid_monthly_clean.csv")
```

View the data to see what information is present

```
head(Covid_monthly)
```

```
##           Date Covid  
## 1 2020-02-1    11  
## 2 2020-03-1   8533  
## 3 2020-04-1  44688  
## 4 2020-05-1  37711  
## 5 2020-06-1  13257  
## 6 2020-07-1  12108
```

Check for any missing values

```
sum(is.na(Covid_monthly))
```

```
## [1] 0
```

Create function to view basic EDA

```
basic_eda <- function(data)
{
  glimpse(data)
  print(status(data))
  freq(data)
  print(profiling_num(data))
  plot_num(data)
  describe(data)
}
```

produce and view the results of the EDA function

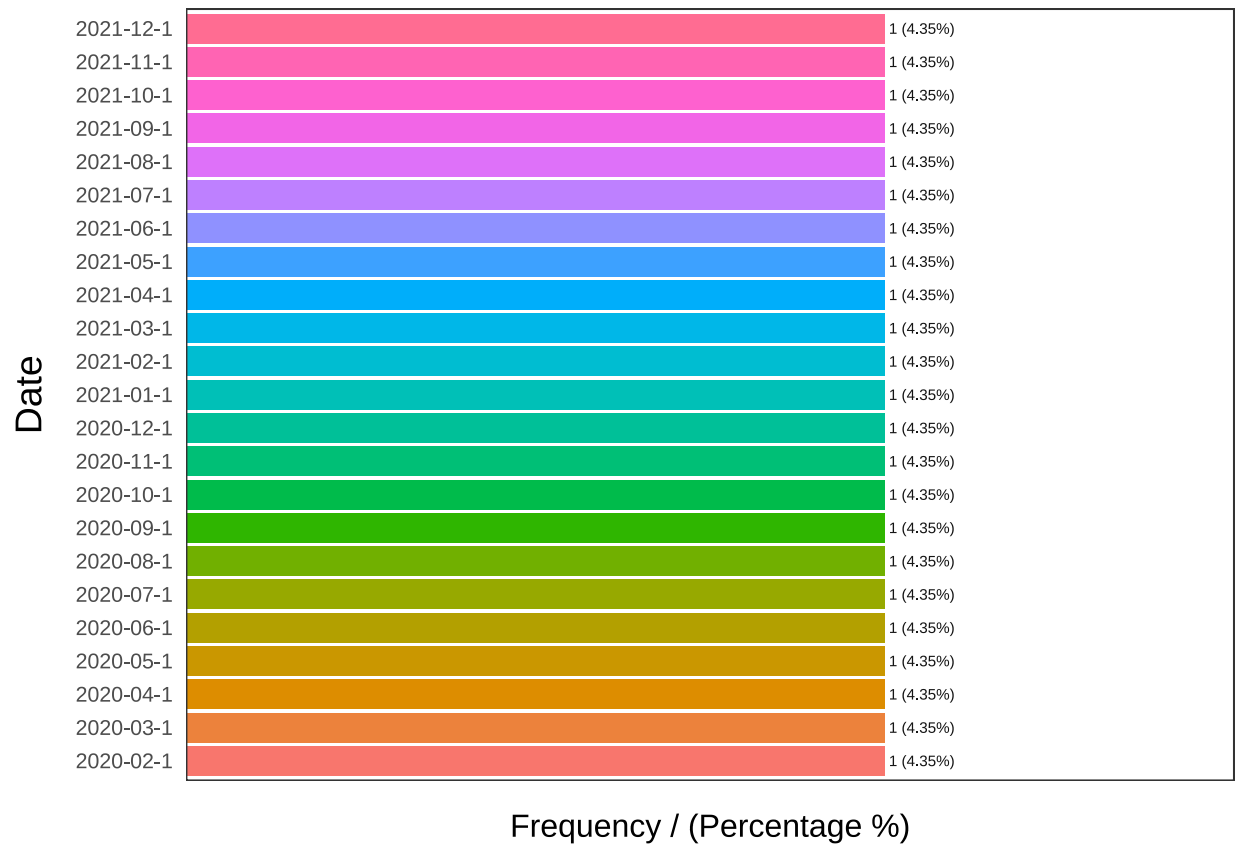
```
basic_eda(Covid_monthly)
```

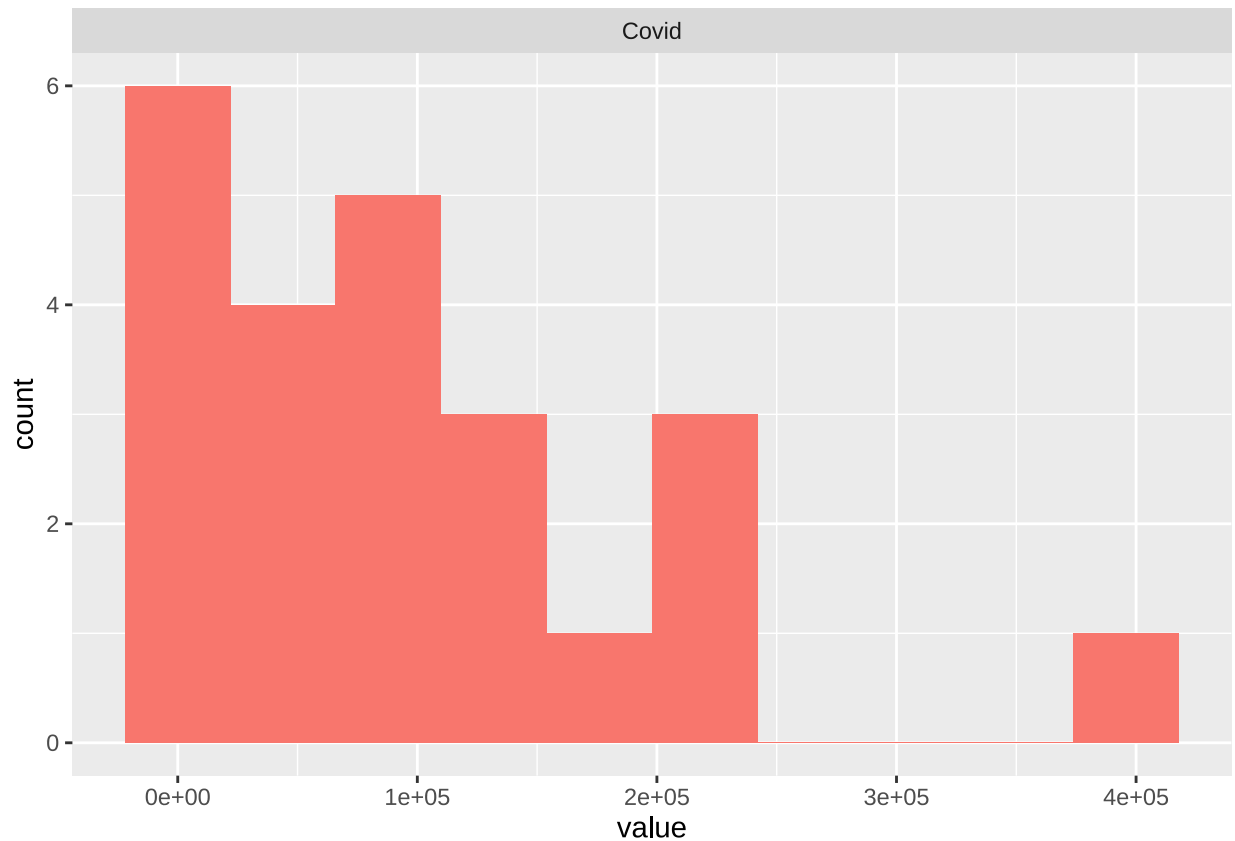
```
## Rows: 23
## Columns: 2
## $ Date <chr> "2020-02-1", "2020-03-1", "2020-04-1", "2020-05-1", "2020-06-1", ~
## $ Covid <int> 11, 8533, 44688, 37711, 13257, 12108, 12636, 29810, 76686, 14269~
##      variable q_zeros p_zeros q_na p_na q_inf p_inf      type unique
## Date      Date      0      0    0    0    0    0 character    23
## Covid     Covid     0      0    0    0    0    0   integer    23
```

```
## Warning: 'guides(<scale> = FALSE)' is deprecated. Please use 'guides(<scale> =
## "none")' instead.
```

```
##   variable    mean std_dev variation_coef    p_01    p_05    p_25    p_50    p_75
## 1 Covid 95143.7 95264.01      1.001265 1885.84 8890.5 22889 75935 134127.5
##   p_95    p_99 skewness kurtosis    iqr      range_98
## 1 233906 360943.5 1.52201 5.322721 111238.5 [1885.84, 360943.46]
##      range_80
## 1 [12213.6, 202315.6]
```

```
## Warning: 'guides(<scale> = FALSE)' is deprecated. Please use 'guides(<scale> =
## "none")' instead.
```



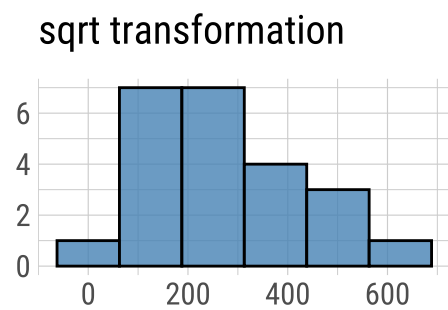
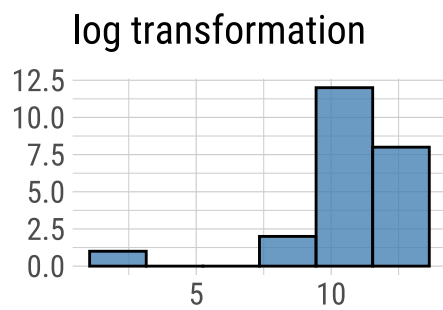
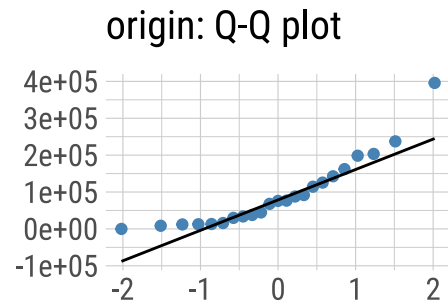
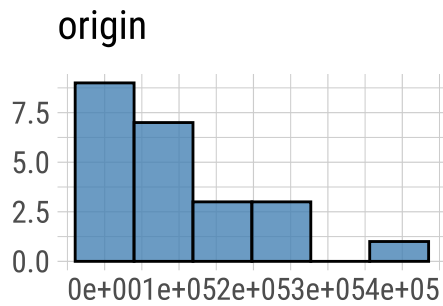


```
## data
##
## 2 Variables      23 Observations
## -----
## Date
##      n missing distinct
##      23      0      23
##
## lowest : 2020-02-1 2020-03-1 2020-04-1 2020-05-1 2020-06-1
## highest: 2021-08-1 2021-09-1 2021-10-1 2021-11-1 2021-12-1
## -----
## Covid
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      23      0      23          1    95144    100476    8890    12214
##      .25      .50      .75      .90      .95
##      22889    75935    134128    202316    233906
##
## lowest :      11    8533    12108    12636    13257, highest: 162155 198426 203288 237308 395815
## -----
```

Visual analysis of the normality of the data.

```
plot_normality(Covid_monthly )
```

Normality Diagnosis Plot (Covid)



```
eda_web_report(Covid_monthly)
```

```
##
##
## processing file: eda_temp.Rmd

## |
## ordinary text without R code
##
## |
## label: setup (with options)
## List of 3
## $ echo : logi FALSE
## $ warning: logi FALSE
## $ message: logi FALSE
##
## |
## ordinary text without R code
##
## |
## label: load_packages
## |
## ordinary text without R code
##
## |
```

```

## label: unnamed-chunk-9 (with options)
## List of 2
## $ echo : logi FALSE
## $ engine: chr "css"
##
## | .....
## ordinary text without R code
##
## | .....
## label: udf (with options)
## List of 3
## $ echo : logi FALSE
## $ warning: logi FALSE
## $ message: logi FALSE
##
## | .....
## ordinary text without R code
##
## | .....
## label: check_variables (with options)
## List of 4
## $ echo : logi FALSE
## $ warning: logi FALSE
## $ message: logi FALSE
## $ comment: chr ""
##
## | .....
## ordinary text without R code
##
## | .....
## label: create-overview
## | .....
## ordinary text without R code
##
## | .....
## label: overview (with options)
## List of 1
## $ results: chr "asis"
##
## | .....
## ordinary text without R code
##
## | .....
## label: overview-pre (with options)
## List of 1
## $ results: chr "asis"
##
## | .....
## ordinary text without R code
##
## | .....
## label: unnamed-chunk-10 (with options)
## List of 1
## $ results: chr "asis"

```



```

##
## | .....
## ordinary text without R code
##
## | .....
## label: unnamed-chunk-11 (with options)
## List of 1
## $ results: chr "asis"
##
## | .....
## ordinary text without R code
##
## | .....
## label: variables (with options)
## List of 1
## $ results: chr "asis"

## | .....
## ordinary text without R code
##
## | .....
## label: normality (with options)
## List of 1
## $ results: chr "asis"
##
## | .....
## ordinary text without R code
##
## | .....
## label: normality-list (with options)
## List of 2
## $ comment: chr ""
## $ results: chr "asis"

## | .....
## ordinary text without R code
##
## | .....
## label: unnamed-chunk-12 (with options)
## List of 1
## $ results: chr "asis"
##
## | .....
## ordinary text without R code
##
## | .....
## label: unnamed-chunk-13 (with options)
## List of 1
## $ results: chr "asis"
##
## | .....
## ordinary text without R code
##
## | .....

```

```

## label: compare_numerical (with options)
## List of 1
## $ results: chr "asis"
##
## | .....
## ordinary text without R code
##
## | .....
## label: unnamed-chunk-14 (with options)
## List of 1
## $ results: chr "asis"
##
## | .....
## ordinary text without R code
##
## | .....
## label: compare-category (with options)
## List of 1
## $ results: chr "asis"
##
## | .....
## ordinary text without R code
##
## | .....
## label: unnamed-chunk-15 (with options)
## List of 1
## $ results: chr "asis"
##
## | .....
## ordinary text without R code
##
## | .....
## label: unnamed-chunk-16 (with options)
## List of 1
## $ results: chr "asis"
##
## | .....
## ordinary text without R code
##
## | .....
## label: unnamed-chunk-17 (with options)
## List of 1
## $ results: chr "asis"
##
## | .....
## ordinary text without R code
##
## | .....
## label: correlation (with options)
## List of 1
## $ results: chr "asis"
##
## | .....
## ordinary text without R code

```

```

##
## | .....
## label: unnamed-chunk-18 (with options)
## List of 1
## $ results: chr "asis"
##
## | .....
## ordinary text without R code
##
## | .....
## label: plot-correlation (with options)
## List of 1
## $ results: chr "asis"
##
## | .....
## ordinary text without R code
##
## | .....
## label: unnamed-chunk-19 (with options)
## List of 1
## $ results: chr "asis"
##
## | .....
## ordinary text without R code
##
## | .....
## label: unnamed-chunk-20 (with options)
## List of 1
## $ results: chr "asis"
##
## | .....
## ordinary text without R code
##
## | .....
## label: unnamed-chunk-21 (with options)
## List of 1
## $ results: chr "asis"
##
## | .....
## ordinary text without R code
##
## | .....
## label: group-numerical (with options)
## List of 1
## $ results: chr "asis"
##
## | .....
## ordinary text without R code
##
## | .....
## label: group-categorical (with options)
## List of 1
## $ results: chr "asis"
##

```

```

## | .....
## ordinary text without R code
##
## | .....
## label: unnamed-chunk-22 (with options)
## List of 1
## $ results: chr "asis"
##
## | .....
## ordinary text without R code
##
## | .....
## label: group-correlation (with options)
## List of 1
## $ results: chr "asis"
##
## | .....
## ordinary text without R code

## output file: eda_temp.knit.md

## "C:/Program Files/RStudio/bin/pandoc/pandoc" +RTS -K512m -RTS eda_temp.knit.md --to html4 --from mar

##
## Output created: C:\Users\KATIES~1\AppData\Local\Temp\Rtmp0KizRI\EDA_Report.html

```