

Assignment1 Final

R Markdown

```
##Dataset source picked from Kaggle.com of Covid Variants https://www.kaggle.com/pavan9065/covid-variants
```

```
#getwd() #setwd("/Users/ankithdasu/Desktop/Spring 2022/Fundamentals of Machine Learning/Assignment 1/Data")
```

```
##Loading CSV file to a dataframe
```

```
##Data Frame Name: Which has data in it: Initialise it  
##data frame can be initialized using = or <-
```

```
getwd()
```

```
## [1] "/Users/ankithdasu/Desktop/Spring 2022/Fundamentals of Machine Learning/Assignment 1/Data"
```

```
setwd("/Users/ankithdasu/Desktop/Spring 2022/Fundamentals of Machine Learning/Assignment 1/Data")  
Covid_V1 <- read.csv("covid-variants1.csv")  
#select dataframe name and cmdenter to see successful load
```

```
##Descriptive Statistics for Quantitative and Categorical Variables in the data
```

```
##For Quantitative Variable  
mean(Covid_V1$num_sequences_total)
```

```
## [1] 1509.582
```

```
summary(Covid_V1$num_sequences_total)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##         1       12      59    1510     394  146170
```

```
##For Categorical Variable  
table(Covid_V1$location)
```

```
##  
##           Angola           Argentina           Aruba  
##           672           1056           600  
##           Australia           Austria           Bahrain  
##           1056           1032           696  
##           Bangladesh           Belgium           Belize  
##           1080           1080           576
```

##	Benin	Bosnia and Herzegovina	Botswana
##	336	864	816
##	Brazil	Brunei	Bulgaria
##	1056	240	936
##	Cambodia	Cameroon	Canada
##	720	528	1056
##	Chile	Colombia	Costa Rica
##	1056	1056	1008
##	Croatia	Curacao	Cyprus
##	792	792	504
##	Czechia	Denmark	Djibouti
##	1056	1056	720
##	Dominican Republic	Ecuador	Egypt
##	576	1056	984
##	Estonia	Ethiopia	Fiji
##	768	384	336
##	Finland	France	Gambia
##	936	1080	744
##	Georgia	Germany	Ghana
##	864	1056	1008
##	Greece	Guatemala	Hong Kong
##	936	1008	1056
##	Hungary	Iceland	India
##	408	720	1056
##	Indonesia	Iran	Iraq
##	1056	720	528
##	Ireland	Israel	Italy
##	1032	1008	1056
##	Jamaica	Japan	Jordan
##	432	1056	936
##	Kazakhstan	Kenya	Kosovo
##	456	1032	552
##	Kuwait	Latvia	Lebanon
##	432	792	552
##	Liechtenstein	Lithuania	Luxembourg
##	672	960	1008
##	Madagascar	Malawi	Malaysia
##	744	672	1008
##	Maldives	Malta	Mauritius
##	864	456	720
##	Mexico	Moldova	Monaco
##	1080	504	360
##	Mongolia	Montenegro	Morocco
##	456	384	792
##	Mozambique	Nepal	Netherlands
##	888	576	1080
##	New Zealand	Nigeria	North Macedonia
##	1056	1032	816
##	Norway	Oman	Pakistan
##	1056	768	1008
##	Papua New Guinea	Paraguay	Peru
##	816	936	1056
##	Philippines	Poland	Portugal
##	744	984	960

##	Qatar	Romania	Russia
##	936	1056	1056
##	Rwanda	Senegal	Serbia
##	792	912	912
##	Seychelles	Singapore	Sint Maarten (Dutch part)
##	528	1056	552
##	Slovakia	Slovenia	South Africa
##	1008	1008	1056
##	South Korea	Spain	Sri Lanka
##	1056	1056	888
##	Suriname	Sweden	Switzerland
##	792	1056	1056
##	Thailand	Togo	Trinidad and Tobago
##	864	528	792
##	Turkey	Uganda	Ukraine
##	1032	792	816
##	United Arab Emirates	United Kingdom	United States
##	576	1080	1080
##	Uruguay	Vietnam	Zambia
##	576	504	816
##	Zimbabwe		
##	672		

```
summary(Covid_V1$location)
```

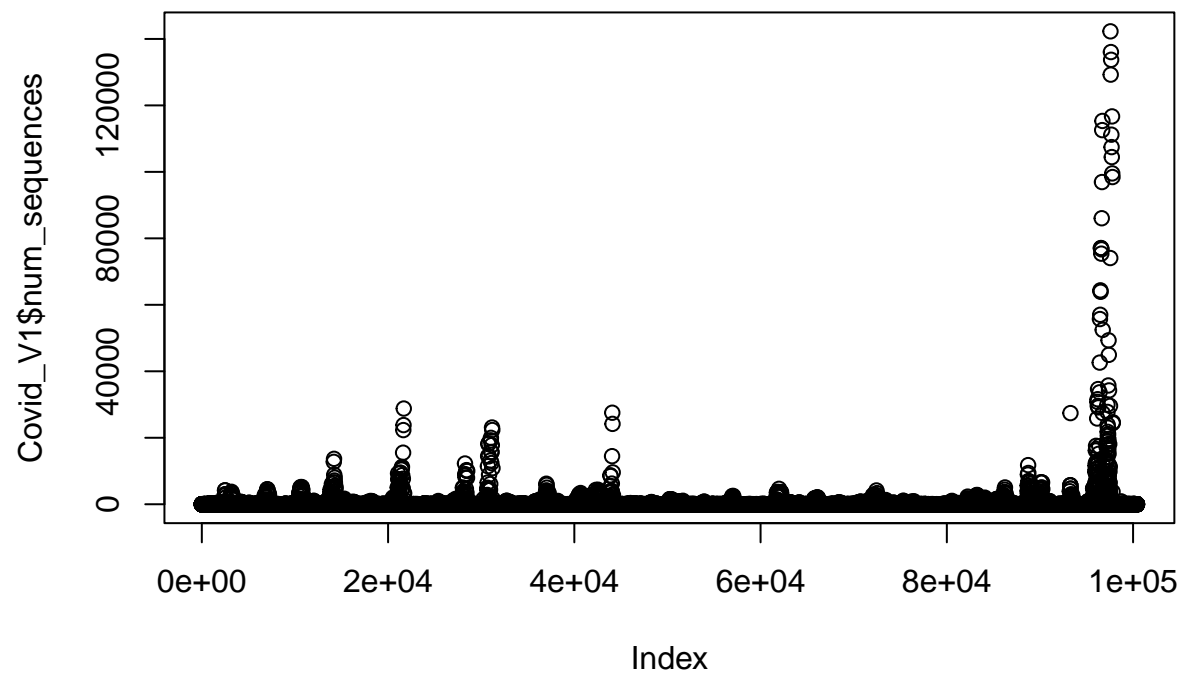
```
##      Length      Class      Mode
## 100416 character character
```

```
##Transformation of variables
```

```
##log transformation
log(Covid_V1$num_sequences)
```

```
##Plotting of the data
```

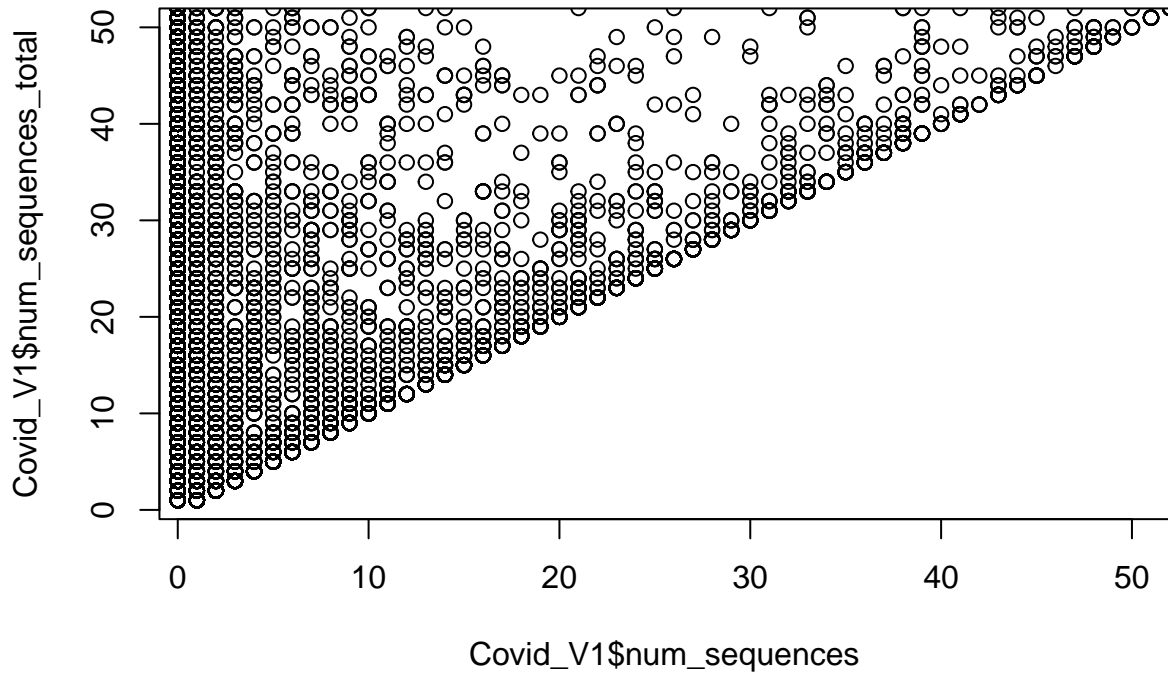
```
##Plot
plot(Covid_V1$num_sequences)
```



```
##Scatter Plot
```

```
plot(Covid_V1$num_sequences,Covid_V1$num_sequences_total,main = "Sample ScatterPlot", xlim = c(1,50), y
```

Sample ScatterPlot



##Completed