

data_manupulation

2022-06-20

DISCREPTION:

Coronavirus disease (COVID-19) is an infectious disease caused by the SARS-CoV-2 virus. Most people who recover from COVID-19 do not develop serious illness. This data set shows the Details of Statistics showing the Indigenous.Cases.till.25.05.2022, Indi

R glimpse glimpse () function get a glimpse of your data. dplyr package is needed to run this function.

```
library(dplyr)
glimpse(covid19)

## Rows: 42
## Columns: 7
## $ S.No          <chr> "1", "2", "3", "4", "5", "6", "7", "8~
## $ District      <chr> "Ariyalur", "Chengalpattu", "Chennai"~
## $ Indigenous.Cases.till.25.05.2022 <int> 19863, 235806, 752268, 329989, 74065,~
## $ Indigenous.Cases.on.26.05.2022 <int> 0, 15, 33, 3, 0, 0, 0, 0, 0, 3, 0, 0,~
## $ Imported.Cases.till.25.05.2022 <int> 20, 5, 48, 51, 203, 216, 77, 94, 404,~
## $ Imported.Cases.on.26.05.2022 <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ Total.cases.till.26.05.2022 <int> 19883, 235826, 752349, 330043, 74268,~
```

The head () function returns the first part of a vector, Matrix, table, data frame, or function.

```
##   S.No   District Indigenous.Cases.till.25.05.2022
## 1     1   Ariyalur                19863
## 2     2 Chengalpattu            235806
## 3     3   Chennai              752268
## 4     4 Coimbatore             329989
## 5     5 Cuddalore              74065
## 6     6 Dharmapuri             35976
##   Indigenous.Cases.on.26.05.2022 Imported.Cases.till.25.05.2022
## 1                                0                                20
## 2                                15                                5
## 3                                33                                48
## 4                                3                                 51
## 5                                0                               203
## 6                                0                               216
##   Imported.Cases.on.26.05.2022 Total.cases.till.26.05.2022
## 1                                0                19883
## 2                                0                235826
## 3                                0                752349
```

```
## 4 0 330043
## 5 0 74268
## 6 0 36192
```

The `skimr` returns a summary with the name of the dataset and the number of rows and columns

```
library(skimr)
library(janitor)
```

```
##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##   chisq.test, fisher.test

skim_without_charts(covid19)
```

Table 1: Data summary

Name	covid19
Number of rows	42
Number of columns	7
Column type frequency:	
character	2
numeric	5
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
S.No	0	1	1	5	0	42	0
District	0	1	5	36	0	42	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Indigenous.Cases.till.25.05.2022	0	1	164066.48	532857.05	0	33125.50	54705.5	91829.5	3445396
Indigenous.Cases.on.26.05.2022	0	1	2.83	10.46	0	0.00	0.0	0.0	59
Imported.Cases.till.25.05.2022	0	1	459.24	1511.92	3	44.25	99.0	237.0	9644
Imported.Cases.on.26.05.2022	0	1	0.00	0.00	0	0.00	0.0	0.0	0
Total.cases.till.26.05.2022	0	1	164003.26	534404.24	428	30505.00	55714.0	91889.5	3455099

selecting one coloumn

Here we are selecting only one coloumn(District) so We get a dataframe with only the “District” column.

```
covid19%>%select(District)
```

##	District
## 1	Ariyalur
## 2	Chengalpattu
## 3	Chennai
## 4	Coimbatore
## 5	Cuddalore
## 6	Dharmapuri
## 7	Dindigul
## 8	Erode
## 9	Kallakurichi
## 10	Kancheepuram
## 11	Kanniyakumari
## 12	Karur
## 13	Krishnagiri
## 14	Madurai
## 15	Mayiladuthurai
## 16	Nagapattinam
## 17	Namakkal
## 18	Nilgiris
## 19	Perambalur
## 20	Pudukkottai
## 21	Ramanathapuram
## 22	Ranipet
## 23	Salem
## 24	Sivaganga
## 25	Tenkasi
## 26	Thanjavur
## 27	Theni
## 28	Thoothukudi
## 29	Tiruchirappalli
## 30	Tirunelveli
## 31	Tirupathur
## 32	Tiruppur
## 33	Tiruvallur
## 34	Tiruvannamalai
## 35	Tiruvarur
## 36	Vellore
## 37	Villupuram
## 38	Virudhunagar
## 39	Airport Surveillance (International)
## 40	Airport Surveillance (Domestic)
## 41	Railway Surveillance
## 42	Grand Total

selecting all columns other than district

Here we are selecting all column except district so, We get a dataframe without the “Districts” column.

```
covid19%>%select(-District)
```

##	S.No	Indigenous.Cases.till.25.05.2022	Indigenous.Cases.on.26.05.2022
## 1	1	19863	0
## 2	2	235806	15
## 3	3	752268	33
## 4	4	329989	3
## 5	5	74065	0
## 6	6	35976	0
## 7	7	37403	0
## 8	8	132582	0
## 9	9	36118	0
## 10	10	94448	3
## 11	11	86110	0
## 12	12	29711	0
## 13	13	59397	1
## 14	14	90880	3
## 15	15	26459	0
## 16	16	25389	0
## 17	17	67898	0
## 18	18	42091	0
## 19	19	14458	0
## 20	20	34438	0
## 21	21	24547	0
## 22	22	53882	0
## 23	23	126963	0
## 24	24	23721	0
## 25	25	32688	0
## 26	26	92146	0
## 27	27	50558	0
## 28	28	64695	0
## 29	29	94897	0
## 30	30	62348	0
## 31	31	35620	0
## 32	32	129924	0
## 33	33	147517	2
## 34	34	66415	0
## 35	35	47979	0
## 36	36	54997	0
## 37	37	54414	0
## 38	38	56736	0
## 39	39	0	0
## 40	40	0	0
## 41	41	0	0
## 42	Total	3445396	59
##	Imported.Cases.till.25.05.2022	Imported.Cases.on.26.05.2022	
## 1	20	0	
## 2	5	0	

## 3	48	0
## 4	51	0
## 5	203	0
## 6	216	0
## 7	77	0
## 8	94	0
## 9	404	0
## 10	4	0
## 11	126	0
## 12	47	0
## 13	244	0
## 14	174	0
## 15	39	0
## 16	54	0
## 17	112	0
## 18	44	0
## 19	3	0
## 20	35	0
## 21	135	0
## 22	49	0
## 23	438	0
## 24	117	0
## 25	58	0
## 26	22	0
## 27	45	0
## 28	275	0
## 29	75	0
## 30	427	0
## 31	118	0
## 32	16	0
## 33	10	0
## 34	399	0
## 35	38	0
## 36	2326	0
## 37	174	0
## 38	104	0
## 39	1276	0
## 40	1114	0
## 41	428	0
## 42	9644	0
##	Total.cases.till.26.05.2022	
## 1	19883	
## 2	235826	
## 3	752349	
## 4	330043	
## 5	74268	
## 6	36192	
## 7	37480	
## 8	132676	
## 9	14461	
## 10	94457	
## 11	86236	
## 12	29758	
## 13	59642	

```
## 14          91054
## 15          26498
## 16          25443
## 17          68010
## 18          42135
## 19          14461
## 20          34473
## 21          24682
## 22          53931
## 23         127401
## 24          23838
## 25          32746
## 26          92168
## 27          50603
## 28          64970
## 29          94972
## 30          62775
## 31          35738
## 32         129940
## 33         147529
## 34          66814
## 35          48017
## 36          57323
## 37          54588
## 38          56840
## 39           1276
## 40           1114
## 41            428
## 42         3455099
```

rename a variable

rename () function in R Language is used to rename the column names of a data frame, based on the older names, Here we have changed the name of the District into Districts

```
covid19
```

```
##      S.No      District Indigenous.Cases.till.25.05.2022
## 1      1      Ariyalur              19863
## 2      2 Chengalpattu             235806
## 3      3      Chennai             752268
## 4      4    Coimbatore             329989
## 5      5    Cuddalore             74065
## 6      6    Dharmapuri             35976
## 7      7    Dindigul              37403
## 8      8      Erode              132582
## 9      9 Kallakurichi             36118
## 10     10 Kancheepuram           94448
## 11     11 Kanniyakumari          86110
## 12     12      Karur              29711
## 13     13 Krishnagiri           59397
```

## 14	14	Madurai	90880
## 15	15	Mayiladuthurai	26459
## 16	16	Nagapattinam	25389
## 17	17	Namakkal	67898
## 18	18	Nilgiris	42091
## 19	19	Perambalur	14458
## 20	20	Pudukkottai	34438
## 21	21	Ramanathapuram	24547
## 22	22	Ranipet	53882
## 23	23	Salem	126963
## 24	24	Sivaganga	23721
## 25	25	Tenkasi	32688
## 26	26	Thanjavur	92146
## 27	27	Theni	50558
## 28	28	Thoothukudi	64695
## 29	29	Tiruchirappalli	94897
## 30	30	Tirunelveli	62348
## 31	31	Tirupathur	35620
## 32	32	Tiruppur	129924
## 33	33	Tiruvallur	147517
## 34	34	Tiruvannamalai	66415
## 35	35	Tiruvarur	47979
## 36	36	Vellore	54997
## 37	37	Villupuram	54414
## 38	38	Virudhunagar	56736
## 39	39	Airport Surveillance (International)	0
## 40	40	Airport Surveillance (Domestic)	0
## 41	41	Railway Surveillance	0
## 42	Total	Grand Total	3445396
##	Indigenous.Cases.on.26.05.2022 Imported.Cases.till.25.05.2022		
## 1		0	20
## 2		15	5
## 3		33	48
## 4		3	51
## 5		0	203
## 6		0	216
## 7		0	77
## 8		0	94
## 9		0	404
## 10		3	4
## 11		0	126
## 12		0	47
## 13		1	244
## 14		3	174
## 15		0	39
## 16		0	54
## 17		0	112
## 18		0	44
## 19		0	3
## 20		0	35
## 21		0	135
## 22		0	49
## 23		0	438
## 24		0	117

## 25	0	58
## 26	0	22
## 27	0	45
## 28	0	275
## 29	0	75
## 30	0	427
## 31	0	118
## 32	0	16
## 33	2	10
## 34	0	399
## 35	0	38
## 36	0	2326
## 37	0	174
## 38	0	104
## 39	0	1276
## 40	0	1114
## 41	0	428
## 42	59	9644
##	Imported.Cases.on.26.05.2022 Total.cases.till.26.05.2022	
## 1	0	19883
## 2	0	235826
## 3	0	752349
## 4	0	330043
## 5	0	74268
## 6	0	36192
## 7	0	37480
## 8	0	132676
## 9	0	14461
## 10	0	94457
## 11	0	86236
## 12	0	29758
## 13	0	59642
## 14	0	91054
## 15	0	26498
## 16	0	25443
## 17	0	68010
## 18	0	42135
## 19	0	14461
## 20	0	34473
## 21	0	24682
## 22	0	53931
## 23	0	127401
## 24	0	23838
## 25	0	32746
## 26	0	92168
## 27	0	50603
## 28	0	64970
## 29	0	94972
## 30	0	62775
## 31	0	35738
## 32	0	129940
## 33	0	147529
## 34	0	66814
## 35	0	48017


```
## 36          0          57323
## 37          0          54588
## 38          0          56840
## 39          0          1276
## 40          0          1114
## 41          0           428
## 42          0        3455099
```

```
covid19 %>%
  rename("Districts" = "District") %>%
  glimpse()
```

```
## Rows: 42
## Columns: 7
## $ S.No          <chr> "1", "2", "3", "4", "5", "6", "7", "8~
## $ Districts     <chr> "Ariyalur", "Chengalpattu", "Chennai"~
## $ Indigenous.Cases.till.25.05.2022 <int> 19863, 235806, 752268, 329989, 74065,~
## $ Indigenous.Cases.on.26.05.2022 <int> 0, 15, 33, 3, 0, 0, 0, 0, 0, 3, 0, 0,~
## $ Imported.Cases.till.25.05.2022 <int> 20, 5, 48, 51, 203, 216, 77, 94, 404,~
## $ Imported.Cases.on.26.05.2022 <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ Total.cases.till.26.05.2022 <int> 19883, 235826, 752349, 330043, 74268,~
```

clean

The easiest option to replace spaces in column names is with the `clean.names ()` function. This R function creates syntactically correct column names by replacing blanks with an underscore. Here the column names are replaced with Underscores

```
clean_names(covid19)
```

```
##      s_no          district indigenous_cases_till_25_05_2022
## 1      1          Ariyalur          19863
## 2      2      Chengalpattu          235806
## 3      3           Chennai          752268
## 4      4      Coimbatore          329989
## 5      5      Cuddalore          74065
## 6      6      Dharmapuri          35976
## 7      7      Dindigul          37403
## 8      8           Erode          132582
## 9      9      Kallakurichi          36118
## 10     10      Kancheepuram          94448
## 11     11      Kanniyakumari          86110
## 12     12           Karur          29711
## 13     13      Krishnagiri          59397
## 14     14           Madurai          90880
## 15     15      Mayiladuthurai          26459
## 16     16      Nagapattinam          25389
## 17     17          Namakkal          67898
## 18     18          Nilgiris          42091
## 19     19      Perambalur          14458
## 20     20      Pudukkottai          34438
```

## 21	21	Ramanathapuram	24547
## 22	22	Ranipet	53882
## 23	23	Salem	126963
## 24	24	Sivaganga	23721
## 25	25	Tenkasi	32688
## 26	26	Thanjavur	92146
## 27	27	Theni	50558
## 28	28	Thoothukudi	64695
## 29	29	Tiruchirappalli	94897
## 30	30	Tirunelveli	62348
## 31	31	Tirupathur	35620
## 32	32	Tiruppur	129924
## 33	33	Tiruvallur	147517
## 34	34	Tiruvannamalai	66415
## 35	35	Tiruvarur	47979
## 36	36	Vellore	54997
## 37	37	Villupuram	54414
## 38	38	Virudhunagar	56736
## 39	39	Airport Surveillance (International)	0
## 40	40	Airport Surveillance (Domestic)	0
## 41	41	Railway Surveillance	0
## 42	Total	Grand Total	3445396
##	indigenous_cases_on_26_05_2022	imported_cases_till_25_05_2022	
## 1	0	20	
## 2	15	5	
## 3	33	48	
## 4	3	51	
## 5	0	203	
## 6	0	216	
## 7	0	77	
## 8	0	94	
## 9	0	404	
## 10	3	4	
## 11	0	126	
## 12	0	47	
## 13	1	244	
## 14	3	174	
## 15	0	39	
## 16	0	54	
## 17	0	112	
## 18	0	44	
## 19	0	3	
## 20	0	35	
## 21	0	135	
## 22	0	49	
## 23	0	438	
## 24	0	117	
## 25	0	58	
## 26	0	22	
## 27	0	45	
## 28	0	275	
## 29	0	75	
## 30	0	427	
## 31	0	118	

## 32	0	16
## 33	2	10
## 34	0	399
## 35	0	38
## 36	0	2326
## 37	0	174
## 38	0	104
## 39	0	1276
## 40	0	1114
## 41	0	428
## 42	59	9644
##	imported_cases_on_26_05_2022	total_cases_till_26_05_2022
## 1	0	19883
## 2	0	235826
## 3	0	752349
## 4	0	330043
## 5	0	74268
## 6	0	36192
## 7	0	37480
## 8	0	132676
## 9	0	14461
## 10	0	94457
## 11	0	86236
## 12	0	29758
## 13	0	59642
## 14	0	91054
## 15	0	26498
## 16	0	25443
## 17	0	68010
## 18	0	42135
## 19	0	14461
## 20	0	34473
## 21	0	24682
## 22	0	53931
## 23	0	127401
## 24	0	23838
## 25	0	32746
## 26	0	92168
## 27	0	50603
## 28	0	64970
## 29	0	94972
## 30	0	62775
## 31	0	35738
## 32	0	129940
## 33	0	147529
## 34	0	66814
## 35	0	48017
## 36	0	57323
## 37	0	54588
## 38	0	56840
## 39	0	1276
## 40	0	1114
## 41	0	428
## 42	0	3455099

unique

The `unique ()` function in R is used to eliminate or delete the duplicate values or the rows present in the vector, data frame, or matrix as well. In `Kancheepuram` and `Perambalur Imported.Cases.till.25.05.2022` is 3 duplicate value is deleted using `unique` function.

```
unique(covid19$Imported.Cases.till.25.05.2022)
```

```
## [1] 20 5 48 51 203 216 77 94 404 4 126 47 244 174 39
## [16] 54 112 44 3 35 135 49 438 117 58 22 45 275 75 427
## [31] 118 16 10 399 38 2326 104 1276 1114 428 9644
```

ORGANISING THE DATA (`filter()`,`group_by()`,`arrange()`,`summarise()`)

filter

The `filter ()` function is used to produce a subset of the data frame, retaining all rows that satisfy the specified conditions. In this data set in column `Indigenous.Cases.till.25.05.2022` we filter the cases with 0. There are three districts with 0 cases.

```
covid19%>%filter(Indigenous.Cases.till.25.05.2022 == "0")
```

```
## S.No District Indigenous.Cases.till.25.05.2022
## 1 39 Airport Surveillance (International) 0
## 2 40 Airport Surveillance (Domestic) 0
## 3 41 Railway Surveillance 0
## Indigenous.Cases.on.26.05.2022 Imported.Cases.till.25.05.2022
## 1 0 1276
## 2 0 1114
## 3 0 428
## Imported.Cases.on.26.05.2022 Total.cases.till.26.05.2022
## 1 0 1276
## 2 0 1114
## 3 0 428
```

filter

In this data set in column `Indigenous.Cases.on.26.05.2022` we filter the cases with 3. There are two districts with 3 cases.

```
covid19%>%filter(Indigenous.Cases.on.26.05.2022 == "3")
```

```
## S.No District Indigenous.Cases.till.25.05.2022
## 1 4 Coimbatore 329989
## 2 10 Kancheepuram 94448
## 3 14 Madurai 90880
## Indigenous.Cases.on.26.05.2022 Imported.Cases.till.25.05.2022
## 1 3 51
```

```
## 2          3          4
## 3          3         174
## Imported.Cases.on.26.05.2022 Total.cases.till.26.05.2022
## 1          0        330043
## 2          0        94457
## 3          0        91054
```

In this data set in column Imported.Cases.till.25.05.2022 we filter the cases with 3. There are two districts with 3 cases.

```
covid19%>%filter(Imported.Cases.till.25.05.2022=="3")

## S.No District Indigenous.Cases.till.25.05.2022
## 1 19 Perambalur 14458
## Indigenous.Cases.on.26.05.2022 Imported.Cases.till.25.05.2022
## 1 0 3
## Imported.Cases.on.26.05.2022 Total.cases.till.26.05.2022
## 1 0 14461
```

Like wise we filter Total.cases.till.26.05.2022 as 14461, There are two districts with 14458 cases.

```
covid19%>%filter(Total.cases.till.26.05.2022=="14461")

## S.No District Indigenous.Cases.till.25.05.2022
## 1 9 Kallakurichi 36118
## 2 19 Perambalur 14458
## Indigenous.Cases.on.26.05.2022 Imported.Cases.till.25.05.2022
## 1 0 404
## 2 0 3
## Imported.Cases.on.26.05.2022 Total.cases.till.26.05.2022
## 1 0 14461
## 2 0 14461
```

group_by & summarise

group_by allows you to group by a one or more variables. summarize: summarize creates a new data.frame containing calculated summary information about a grouped variable. Here the data set is grouped by district and summarize information about Imported.Cases.till.25.05.2022

```
covid19%>%group_by(District)%>%drop_na()%>%summarise(Imported.Cases.till.25.05.2022)

## # A tibble: 42 x 2
## District Imported.Cases.till.25.05.2022
## <chr> <int>
## 1 Airport Surveillance (Domestic) 1114
```

```
## 2 Airport Surveillance (International) 1276
## 3 Ariyalur 20
## 4 Chengalpattu 5
## 5 Chennai 48
## 6 Coimbatore 51
## 7 Cuddalore 203
## 8 Dharmapuri 216
## 9 Dindigul 77
## 10 Erode 94
## # ... with 32 more rows
```

Here the data set is grouped by district and summarize information about `Total.cases.till.26.05.2022`

```
covid19%>%group_by(District)%>%drop_na()%>%summarise(Total.cases.till.26.05.2022)
```

```
## # A tibble: 42 x 2
##   District Total.cases.till.26.05.2022
##   <chr> <int>
## 1 Airport Surveillance (Domestic) 1114
## 2 Airport Surveillance (International) 1276
## 3 Ariyalur 19883
## 4 Chengalpattu 235826
## 5 Chennai 752349
## 6 Coimbatore 330043
## 7 Cuddalore 74268
## 8 Dharmapuri 36192
## 9 Dindigul 37480
## 10 Erode 132676
## # ... with 32 more rows
```

arrange

The `arrange ()` function is used to rearrange rows in ascending or descending order. In this dataset `District` column is arranged in descending order.

```
covid19%>%arrange(desc(District))
```

```
##   S.No District Indigenous.Cases.till.25.05.2022
## 1 38 Virudhunagar 56736
## 2 37 Villupuram 54414
## 3 36 Vellore 54997
## 4 35 Tiruvarur 47979
## 5 34 Tiruvannamalai 66415
## 6 33 Tiruvallur 147517
## 7 32 Tiruppur 129924
## 8 31 Tirupathur 35620
## 9 30 Tirunelveli 62348
## 10 29 Tiruchirappalli 94897
```

## 11	28	Thoothukudi	64695
## 12	27	Theni	50558
## 13	26	Thanjavur	92146
## 14	25	Tenkasi	32688
## 15	24	Sivaganga	23721
## 16	23	Salem	126963
## 17	22	Ranipet	53882
## 18	21	Ramanathapuram	24547
## 19	41	Railway Surveillance	0
## 20	20	Pudukkottai	34438
## 21	19	Perambalur	14458
## 22	18	Nilgiris	42091
## 23	17	Namakkal	67898
## 24	16	Nagapattinam	25389
## 25	15	Mayiladuthurai	26459
## 26	14	Madurai	90880
## 27	13	Krishnagiri	59397
## 28	12	Karur	29711
## 29	11	Kanniyakumari	86110
## 30	10	Kancheepuram	94448
## 31	9	Kallakurichi	36118
## 32	Total	Grand Total	3445396
## 33	8	Erode	132582
## 34	7	Dindigul	37403
## 35	6	Dharmapuri	35976
## 36	5	Cuddalore	74065
## 37	4	Coimbatore	329989
## 38	3	Chennai	752268
## 39	2	Chengalpattu	235806
## 40	1	Ariyalur	19863
## 41	39	Airport Surveillance (International)	0
## 42	40	Airport Surveillance (Domestic)	0
##	Indigenous.Cases.on.26.05.2022 Imported.Cases.till.25.05.2022		
## 1		0	104
## 2		0	174
## 3		0	2326
## 4		0	38
## 5		0	399
## 6		2	10
## 7		0	16
## 8		0	118
## 9		0	427
## 10		0	75
## 11		0	275
## 12		0	45
## 13		0	22
## 14		0	58
## 15		0	117
## 16		0	438
## 17		0	49
## 18		0	135
## 19		0	428
## 20		0	35
## 21		0	3

## 22	0	44
## 23	0	112
## 24	0	54
## 25	0	39
## 26	3	174
## 27	1	244
## 28	0	47
## 29	0	126
## 30	3	4
## 31	0	404
## 32	59	9644
## 33	0	94
## 34	0	77
## 35	0	216
## 36	0	203
## 37	3	51
## 38	33	48
## 39	15	5
## 40	0	20
## 41	0	1276
## 42	0	1114
##	Imported.Cases.on.26.05.2022	Total.cases.till.26.05.2022
## 1	0	56840
## 2	0	54588
## 3	0	57323
## 4	0	48017
## 5	0	66814
## 6	0	147529
## 7	0	129940
## 8	0	35738
## 9	0	62775
## 10	0	94972
## 11	0	64970
## 12	0	50603
## 13	0	92168
## 14	0	32746
## 15	0	23838
## 16	0	127401
## 17	0	53931
## 18	0	24682
## 19	0	428
## 20	0	34473
## 21	0	14461
## 22	0	42135
## 23	0	68010
## 24	0	25443
## 25	0	26498
## 26	0	91054
## 27	0	59642
## 28	0	29758
## 29	0	86236
## 30	0	94457
## 31	0	14461
## 32	0	3455099

## 33	0	132676
## 34	0	37480
## 35	0	36192
## 36	0	74268
## 37	0	330043
## 38	0	752349
## 39	0	235826
## 40	0	19883
## 41	0	1276
## 42	0	1114

arrange by group

R arrange () orders the rows of a data frame by the values of selected columns. , In this dataset Indigenous.Cases.on.26.05.2022 arranged in decending order.

```
covid19%>%arrange(desc(Indigenous.Cases.on.26.05.2022), .by_group = TRUE)
```

##	S.No	District	Indigenous.Cases.till.25.05.2022
## 1	Total	Grand Total	3445396
## 2	3	Chennai	752268
## 3	2	Chengalpattu	235806
## 4	4	Coimbatore	329989
## 5	10	Kancheepuram	94448
## 6	14	Madurai	90880
## 7	33	Tiruvallur	147517
## 8	13	Krishnagiri	59397
## 9	1	Ariyalur	19863
## 10	5	Cuddalore	74065
## 11	6	Dharmapuri	35976
## 12	7	Dindigul	37403
## 13	8	Erode	132582
## 14	9	Kallakurichi	36118
## 15	11	Kanniyakumari	86110
## 16	12	Karur	29711
## 17	15	Mayiladuthurai	26459
## 18	16	Nagapattinam	25389
## 19	17	Namakkal	67898
## 20	18	Nilgiris	42091
## 21	19	Perambalur	14458
## 22	20	Pudukkottai	34438
## 23	21	Ramanathapuram	24547
## 24	22	Ranipet	53882
## 25	23	Salem	126963
## 26	24	Sivaganga	23721
## 27	25	Tenkasi	32688
## 28	26	Thanjavur	92146
## 29	27	Theni	50558
## 30	28	Thoothukudi	64695
## 31	29	Tiruchirappalli	94897
## 32	30	Tirunelveli	62348

## 33	31	Tirupathur	35620
## 34	32	Tiruppur	129924
## 35	34	Tiruvannamalai	66415
## 36	35	Tiruvarur	47979
## 37	36	Vellore	54997
## 38	37	Villupuram	54414
## 39	38	Virudhunagar	56736
## 40	39	Airport Surveillance (International)	0
## 41	40	Airport Surveillance (Domestic)	0
## 42	41	Railway Surveillance	0
##	Indigenous.Cases.on.26.05.2022 Imported.Cases.till.25.05.2022		
## 1	59	9644	
## 2	33	48	
## 3	15	5	
## 4	3	51	
## 5	3	4	
## 6	3	174	
## 7	2	10	
## 8	1	244	
## 9	0	20	
## 10	0	203	
## 11	0	216	
## 12	0	77	
## 13	0	94	
## 14	0	404	
## 15	0	126	
## 16	0	47	
## 17	0	39	
## 18	0	54	
## 19	0	112	
## 20	0	44	
## 21	0	3	
## 22	0	35	
## 23	0	135	
## 24	0	49	
## 25	0	438	
## 26	0	117	
## 27	0	58	
## 28	0	22	
## 29	0	45	
## 30	0	275	
## 31	0	75	
## 32	0	427	
## 33	0	118	
## 34	0	16	
## 35	0	399	
## 36	0	38	
## 37	0	2326	
## 38	0	174	
## 39	0	104	
## 40	0	1276	
## 41	0	1114	
## 42	0	428	
##	Imported.Cases.on.26.05.2022 Total.cases.till.26.05.2022		

## 1	0	3455099
## 2	0	752349
## 3	0	235826
## 4	0	330043
## 5	0	94457
## 6	0	91054
## 7	0	147529
## 8	0	59642
## 9	0	19883
## 10	0	74268
## 11	0	36192
## 12	0	37480
## 13	0	132676
## 14	0	14461
## 15	0	86236
## 16	0	29758
## 17	0	26498
## 18	0	25443
## 19	0	68010
## 20	0	42135
## 21	0	14461
## 22	0	34473
## 23	0	24682
## 24	0	53931
## 25	0	127401
## 26	0	23838
## 27	0	32746
## 28	0	92168
## 29	0	50603
## 30	0	64970
## 31	0	94972
## 32	0	62775
## 33	0	35738
## 34	0	129940
## 35	0	66814
## 36	0	48017
## 37	0	57323
## 38	0	54588
## 39	0	56840
## 40	0	1276
## 41	0	1114
## 42	0	428

```
covid19%>%arrange(across(starts_with("T")))
```

##	S.No	District	Indigenous.Cases.till.25.05.2022
## 1	41	Railway Surveillance	0
## 2	40	Airport Surveillance (Domestic)	0
## 3	39	Airport Surveillance (International)	0
## 4	9	Kallakurichi	36118
## 5	19	Perambalur	14458
## 6	1	Ariyalur	19863
## 7	24	Sivaganga	23721
## 8	21	Ramanathapuram	24547
## 9	16	Nagapattinam	25389

## 10	15	Mayiladuthurai	26459
## 11	12	Karur	29711
## 12	25	Tenkasi	32688
## 13	20	Pudukkottai	34438
## 14	31	Tirupathur	35620
## 15	6	Dharmapuri	35976
## 16	7	Dindigul	37403
## 17	18	Nilgiris	42091
## 18	35	Tiruvavarur	47979
## 19	27	Theni	50558
## 20	22	Ranipet	53882
## 21	37	Villupuram	54414
## 22	38	Virudhunagar	56736
## 23	36	Vellore	54997
## 24	13	Krishnagiri	59397
## 25	30	Tirunelveli	62348
## 26	28	Thoothukudi	64695
## 27	34	Tiruvannamalai	66415
## 28	17	Namakkal	67898
## 29	5	Cuddalore	74065
## 30	11	Kanniyakumari	86110
## 31	14	Madurai	90880
## 32	26	Thanjavur	92146
## 33	10	Kancheepuram	94448
## 34	29	Tiruchirappalli	94897
## 35	23	Salem	126963
## 36	32	Tiruppur	129924
## 37	8	Erode	132582
## 38	33	Tiruvallur	147517
## 39	2	Chengalpattu	235806
## 40	4	Coimbatore	329989
## 41	3	Chennai	752268
## 42	Total	Grand Total	3445396
##	Indigenous.Cases.on.26.05.2022 Imported.Cases.till.25.05.2022		
## 1	0	428	
## 2	0	1114	
## 3	0	1276	
## 4	0	404	
## 5	0	3	
## 6	0	20	
## 7	0	117	
## 8	0	135	
## 9	0	54	
## 10	0	39	
## 11	0	47	
## 12	0	58	
## 13	0	35	
## 14	0	118	
## 15	0	216	
## 16	0	77	
## 17	0	44	
## 18	0	38	
## 19	0	45	
## 20	0	49	

## 21	0	174
## 22	0	104
## 23	0	2326
## 24	1	244
## 25	0	427
## 26	0	275
## 27	0	399
## 28	0	112
## 29	0	203
## 30	0	126
## 31	3	174
## 32	0	22
## 33	3	4
## 34	0	75
## 35	0	438
## 36	0	16
## 37	0	94
## 38	2	10
## 39	15	5
## 40	3	51
## 41	33	48
## 42	59	9644
##	Imported.Cases.on.26.05.2022 Total.cases.till.26.05.2022	
## 1	0	428
## 2	0	1114
## 3	0	1276
## 4	0	14461
## 5	0	14461
## 6	0	19883
## 7	0	23838
## 8	0	24682
## 9	0	25443
## 10	0	26498
## 11	0	29758
## 12	0	32746
## 13	0	34473
## 14	0	35738
## 15	0	36192
## 16	0	37480
## 17	0	42135
## 18	0	48017
## 19	0	50603
## 20	0	53931
## 21	0	54588
## 22	0	56840
## 23	0	57323
## 24	0	59642
## 25	0	62775
## 26	0	64970
## 27	0	66814
## 28	0	68010
## 29	0	74268
## 30	0	86236
## 31	0	91054

```
## 32          0          92168
## 33          0          94457
## 34          0          94972
## 35          0         127401
## 36          0         129940
## 37          0         132676
## 38          0         147529
## 39          0         235826
## 40          0         330043
## 41          0         752349
## 42          0        3455099
```

slice

Description `slice()` lets you index rows by their (integer) locations. It allows you to select, remove, and duplicate rows.

```
covid19%>%slice(1L)
```

```
##   S.No District Indigenous.Cases.till.25.05.2022 Indigenous.Cases.on.26.05.2022
## 1     1 Ariyalur                               19863                               0
##   Imported.Cases.till.25.05.2022 Imported.Cases.on.26.05.2022
## 1                               20                               0
##   Total.cases.till.26.05.2022
## 1                               19883
```

```
covid19%>%slice_min(District,n=15)
```

```
##   S.No District Indigenous.Cases.till.25.05.2022
## 1    40 Airport Surveillance (Domestic)          0
## 2    39 Airport Surveillance (International)      0
## 3     1 Ariyalur                               19863
## 4     2 Chengalpattu                          235806
## 5     3 Chennai                             752268
## 6     4 Coimbatore                          329989
## 7     5 Cuddalore                           74065
## 8     6 Dharmapuri                          35976
## 9     7 Dindigul                           37403
## 10    8 Erode                             132582
## 11 Total Grand Total                        3445396
## 12     9 Kallakurichi                        36118
## 13    10 Kancheepuram                      94448
## 14    11 Kanniyakumari                     86110
## 15    12 Karur                             29711
##   Indigenous.Cases.on.26.05.2022 Imported.Cases.till.25.05.2022
## 1                               0          1114
## 2                               0          1276
## 3                               0           20
## 4                              15           5
## 5                              33          48
## 6                               3          51
## 7                               0         203
## 8                               0         216
## 9                               0          77
## 10                              0          94
```

## 11	59	9644
## 12	0	404
## 13	3	4
## 14	0	126
## 15	0	47
##	Imported.Cases.on.26.05.2022	Total.cases.till.26.05.2022
## 1	0	1114
## 2	0	1276
## 3	0	19883
## 4	0	235826
## 5	0	752349
## 6	0	330043
## 7	0	74268
## 8	0	36192
## 9	0	37480
## 10	0	132676
## 11	0	3455099
## 12	0	14461
## 13	0	94457
## 14	0	86236
## 15	0	29758

TRANSFORMING DATA mutate(),unite()

mutate

The mutate () function is a function for creating new variables.

Here I have added New column named Result to show the result if Total.cases.till.26.05.2022 > 5000 then “Many cases”, if not more then 5000 then “Not many cases”

```
covid19 %>%
mutate(Result= if_else(Total.cases.till.26.05.2022 > 5000, "Many cases", "Not many cases"))
```

##	S.No	District	Indigenous.Cases.till.25.05.2022
## 1	1	Ariyalur	19863
## 2	2	Chengalpattu	235806
## 3	3	Chennai	752268
## 4	4	Coimbatore	329989
## 5	5	Cuddalore	74065
## 6	6	Dharmapuri	35976
## 7	7	Dindigul	37403
## 8	8	Erode	132582
## 9	9	Kallakurichi	36118
## 10	10	Kancheepuram	94448
## 11	11	Kanniyakumari	86110
## 12	12	Karur	29711
## 13	13	Krishnagiri	59397
## 14	14	Madurai	90880
## 15	15	Mayiladuthurai	26459
## 16	16	Nagapattinam	25389
## 17	17	Namakkal	67898
## 18	18	Nilgiris	42091
## 19	19	Perambalur	14458

## 20	20	Pudukkottai	34438
## 21	21	Ramanathapuram	24547
## 22	22	Ranipet	53882
## 23	23	Salem	126963
## 24	24	Sivaganga	23721
## 25	25	Tenkasi	32688
## 26	26	Thanjavur	92146
## 27	27	Theni	50558
## 28	28	Thoothukudi	64695
## 29	29	Tiruchirappalli	94897
## 30	30	Tirunelveli	62348
## 31	31	Tirupathur	35620
## 32	32	Tiruppur	129924
## 33	33	Tiruvallur	147517
## 34	34	Tiruvannamalai	66415
## 35	35	Tiruvarur	47979
## 36	36	Vellore	54997
## 37	37	Villupuram	54414
## 38	38	Virudhunagar	56736
## 39	39	Airport Surveillance (International)	0
## 40	40	Airport Surveillance (Domestic)	0
## 41	41	Railway Surveillance	0
## 42	Total	Grand Total	3445396
##	Indigenous.Cases.on.26.05.2022 Imported.Cases.till.25.05.2022		
## 1		0	20
## 2		15	5
## 3		33	48
## 4		3	51
## 5		0	203
## 6		0	216
## 7		0	77
## 8		0	94
## 9		0	404
## 10		3	4
## 11		0	126
## 12		0	47
## 13		1	244
## 14		3	174
## 15		0	39
## 16		0	54
## 17		0	112
## 18		0	44
## 19		0	3
## 20		0	35
## 21		0	135
## 22		0	49
## 23		0	438
## 24		0	117
## 25		0	58
## 26		0	22
## 27		0	45
## 28		0	275
## 29		0	75
## 30		0	427

## 31	0	118	
## 32	0	16	
## 33	2	10	
## 34	0	399	
## 35	0	38	
## 36	0	2326	
## 37	0	174	
## 38	0	104	
## 39	0	1276	
## 40	0	1114	
## 41	0	428	
## 42	59	9644	
##	Imported.Cases.on.26.05.2022	Total.cases.till.26.05.2022	Result
## 1	0	19883	Many cases
## 2	0	235826	Many cases
## 3	0	752349	Many cases
## 4	0	330043	Many cases
## 5	0	74268	Many cases
## 6	0	36192	Many cases
## 7	0	37480	Many cases
## 8	0	132676	Many cases
## 9	0	14461	Many cases
## 10	0	94457	Many cases
## 11	0	86236	Many cases
## 12	0	29758	Many cases
## 13	0	59642	Many cases
## 14	0	91054	Many cases
## 15	0	26498	Many cases
## 16	0	25443	Many cases
## 17	0	68010	Many cases
## 18	0	42135	Many cases
## 19	0	14461	Many cases
## 20	0	34473	Many cases
## 21	0	24682	Many cases
## 22	0	53931	Many cases
## 23	0	127401	Many cases
## 24	0	23838	Many cases
## 25	0	32746	Many cases
## 26	0	92168	Many cases
## 27	0	50603	Many cases
## 28	0	64970	Many cases
## 29	0	94972	Many cases
## 30	0	62775	Many cases
## 31	0	35738	Many cases
## 32	0	129940	Many cases
## 33	0	147529	Many cases
## 34	0	66814	Many cases
## 35	0	48017	Many cases
## 36	0	57323	Many cases
## 37	0	54588	Many cases
## 38	0	56840	Many cases
## 39	0	1276	Not many cases
## 40	0	1114	Not many cases
## 41	0	428	Not many cases

```
## 42                                0                3455099      Many cases
```

unite

The unite () function from the tidyr package can be used to unite multiple data frame columns into a single column. Here in this data set i have choosed columns District and Total.cases.till.26.05.2022.

```
covid19%>%unite("a",District:Total.cases.till.26.05.2022,remove=FALSE)
```

##	S.No	a
## 1	1	Ariyalur_19863_0_20_0_19883
## 2	2	Chengalpattu_235806_15_5_0_235826
## 3	3	Chennai_752268_33_48_0_752349
## 4	4	Coimbatore_329989_3_51_0_330043
## 5	5	Cuddalore_74065_0_203_0_74268
## 6	6	Dharmapuri_35976_0_216_0_36192
## 7	7	Dindigul_37403_0_77_0_37480
## 8	8	Erode_132582_0_94_0_132676
## 9	9	Kallakurichi_36118_0_404_0_14461
## 10	10	Kancheepuram_94448_3_4_0_94457
## 11	11	Kanniyakumari_86110_0_126_0_86236
## 12	12	Karur_29711_0_47_0_29758
## 13	13	Krishnagiri_59397_1_244_0_59642
## 14	14	Madurai_90880_3_174_0_91054
## 15	15	Mayiladuthurai_26459_0_39_0_26498
## 16	16	Nagapattinam_25389_0_54_0_25443
## 17	17	Namakkal_67898_0_112_0_68010
## 18	18	Nilgiris_42091_0_44_0_42135
## 19	19	Perambalur_14458_0_3_0_14461
## 20	20	Pudukkottai_34438_0_35_0_34473
## 21	21	Ramanathapuram_24547_0_135_0_24682
## 22	22	Ranipet_53882_0_49_0_53931
## 23	23	Salem_126963_0_438_0_127401
## 24	24	Sivaganga_23721_0_117_0_23838
## 25	25	Tenkasi_32688_0_58_0_32746
## 26	26	Thanjavur_92146_0_22_0_92168
## 27	27	Theni_50558_0_45_0_50603
## 28	28	Thoothukudi_64695_0_275_0_64970
## 29	29	Tiruchirappalli_94897_0_75_0_94972
## 30	30	Tirunelveli_62348_0_427_0_62775
## 31	31	Tirupathur_35620_0_118_0_35738
## 32	32	Tiruppur_129924_0_16_0_129940
## 33	33	Tiruvallur_147517_2_10_0_147529
## 34	34	Tiruvannamalai_66415_0_399_0_66814
## 35	35	Tiruvarur_47979_0_38_0_48017
## 36	36	Vellore_54997_0_2326_0_57323
## 37	37	Villupuram_54414_0_174_0_54588
## 38	38	Virudhunagar_56736_0_104_0_56840
## 39	39	Airport Surveillance (International)_0_0_1276_0_1276
## 40	40	Airport Surveillance (Domestic)_0_0_1114_0_1114
## 41	41	Railway Surveillance_0_0_428_0_428
## 42	Total	Grand Total_3445396_59_9644_0_3455099
##		District Indigenous.Cases.till.25.05.2022
## 1		Ariyalur 19863

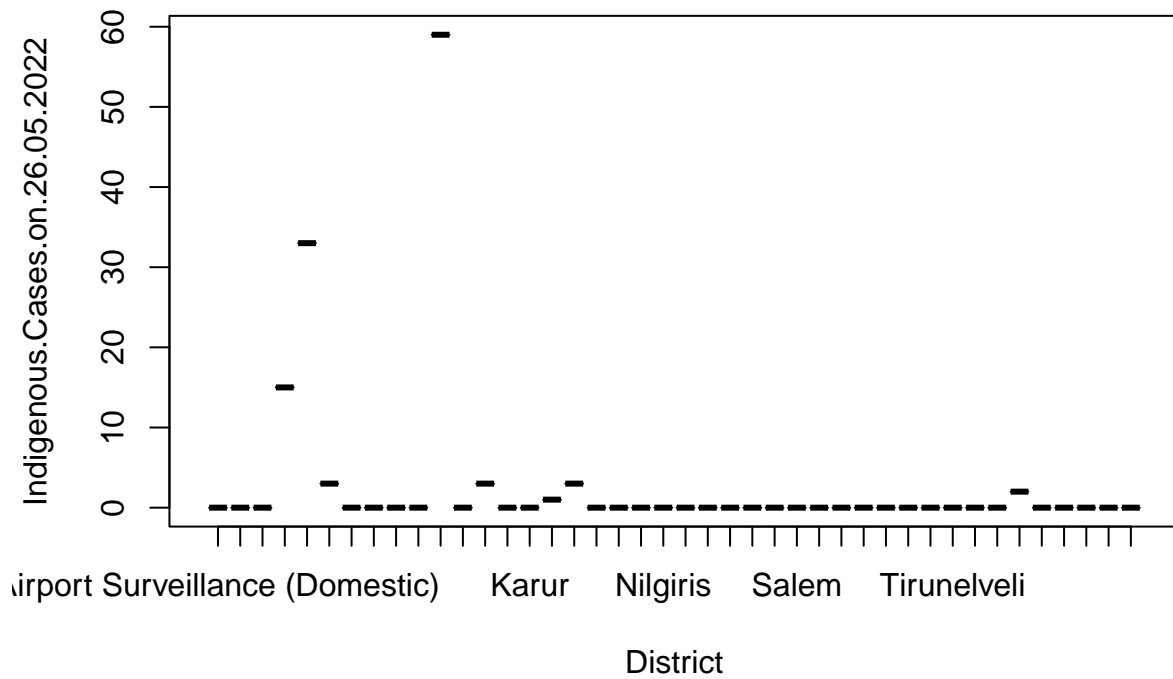
## 2	Chengalpattu	235806
## 3	Chennai	752268
## 4	Coimbatore	329989
## 5	Cuddalore	74065
## 6	Dharmapuri	35976
## 7	Dindigul	37403
## 8	Erode	132582
## 9	Kallakurichi	36118
## 10	Kancheepuram	94448
## 11	Kanniyakumari	86110
## 12	Karur	29711
## 13	Krishnagiri	59397
## 14	Madurai	90880
## 15	Mayiladuthurai	26459
## 16	Nagapattinam	25389
## 17	Namakkal	67898
## 18	Nilgiris	42091
## 19	Perambalur	14458
## 20	Pudukkottai	34438
## 21	Ramanathapuram	24547
## 22	Ranipet	53882
## 23	Salem	126963
## 24	Sivaganga	23721
## 25	Tenkasi	32688
## 26	Thanjavur	92146
## 27	Theni	50558
## 28	Thoothukudi	64695
## 29	Tiruchirappalli	94897
## 30	Tirunelveli	62348
## 31	Tirupathur	35620
## 32	Tiruppur	129924
## 33	Tiruvallur	147517
## 34	Tiruvannamalai	66415
## 35	Tiruvarur	47979
## 36	Vellore	54997
## 37	Villupuram	54414
## 38	Virudhunagar	56736
## 39	Airport Surveillance (International)	0
## 40	Airport Surveillance (Domestic)	0
## 41	Railway Surveillance	0
## 42	Grand Total	3445396
##	Indigenous.Cases.on.26.05.2022 Imported.Cases.till.25.05.2022	
## 1	0	20
## 2	15	5
## 3	33	48
## 4	3	51
## 5	0	203
## 6	0	216
## 7	0	77
## 8	0	94
## 9	0	404
## 10	3	4
## 11	0	126
## 12	0	47

## 13	1	244
## 14	3	174
## 15	0	39
## 16	0	54
## 17	0	112
## 18	0	44
## 19	0	3
## 20	0	35
## 21	0	135
## 22	0	49
## 23	0	438
## 24	0	117
## 25	0	58
## 26	0	22
## 27	0	45
## 28	0	275
## 29	0	75
## 30	0	427
## 31	0	118
## 32	0	16
## 33	2	10
## 34	0	399
## 35	0	38
## 36	0	2326
## 37	0	174
## 38	0	104
## 39	0	1276
## 40	0	1114
## 41	0	428
## 42	59	9644
##	Imported.Cases.on.26.05.2022 Total.cases.till.26.05.2022	
## 1	0	19883
## 2	0	235826
## 3	0	752349
## 4	0	330043
## 5	0	74268
## 6	0	36192
## 7	0	37480
## 8	0	132676
## 9	0	14461
## 10	0	94457
## 11	0	86236
## 12	0	29758
## 13	0	59642
## 14	0	91054
## 15	0	26498
## 16	0	25443
## 17	0	68010
## 18	0	42135
## 19	0	14461
## 20	0	34473
## 21	0	24682
## 22	0	53931
## 23	0	127401

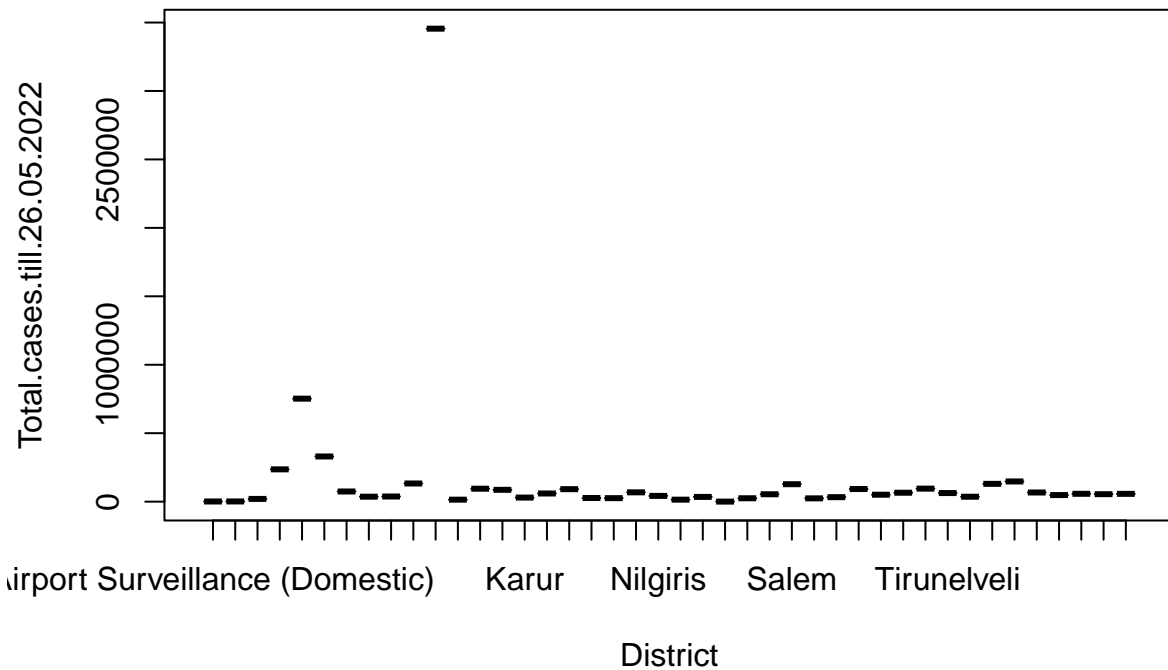
## 24	0	23838
## 25	0	32746
## 26	0	92168
## 27	0	50603
## 28	0	64970
## 29	0	94972
## 30	0	62775
## 31	0	35738
## 32	0	129940
## 33	0	147529
## 34	0	66814
## 35	0	48017
## 36	0	57323
## 37	0	54588
## 38	0	56840
## 39	0	1276
## 40	0	1114
## 41	0	428
## 42	0	3455099

Plot representation

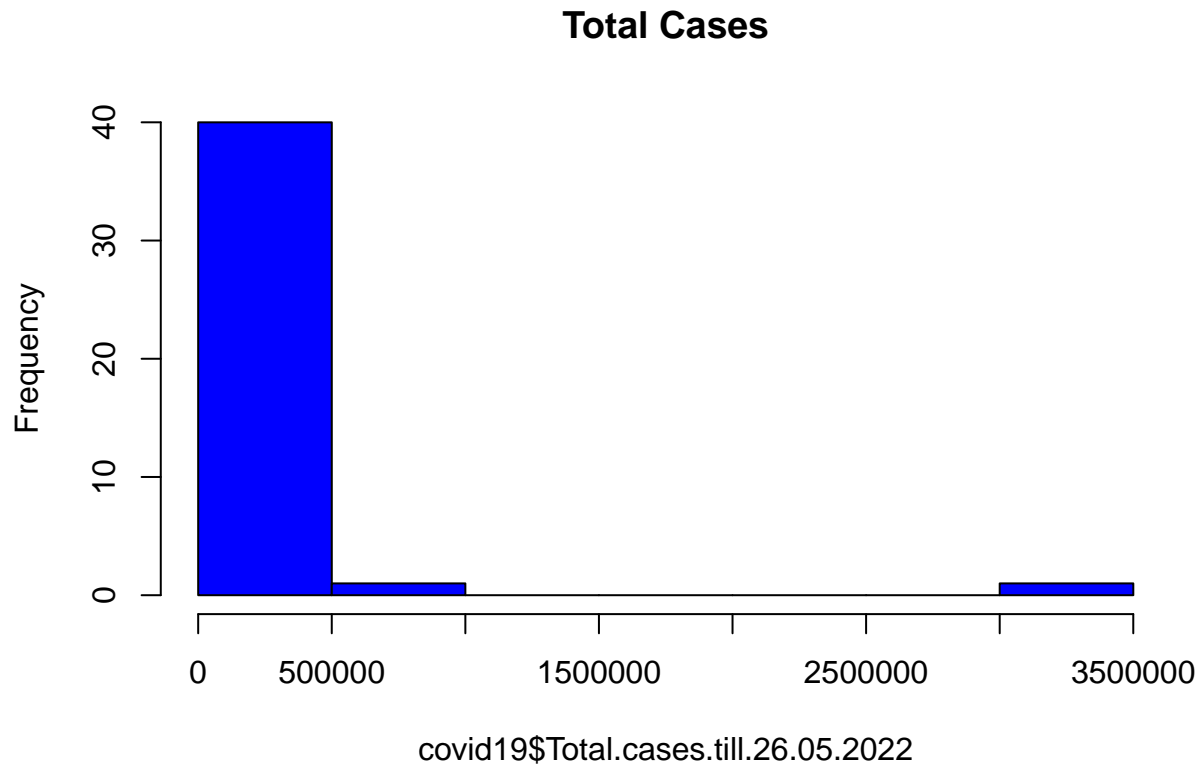
```
boxplot(Indigenous.Cases.on.26.05.2022~District)
```



```
boxplot(Total.cases.till.26.05.2022~District)
```



```
hist(covid19$Total.cases.till.26.05.2022,col = "blue", main="Total Cases")
```



INSIGHTS :

In this Data set we come to know about the Total number of cases till 26.5.2022 District wise by using this dataset we have performed Data manipulation. Through Data manipulation we can perform various functions using that function we can filter the particular column we need arrange them in Ascending and descending.

By filter we can find similar number of cases in Different Districts