Group three - Assignment 1

Houman Ameri,Sahabe Arbabie, Julio Carneiro

Introduction

This dataset is about Covid analysis

Load dataset

```
covid4=read.csv("C:/Users/h.ameri/Desktop/covid4.csv")
```

Print dataset

```
print(covid4)
```

```
##    submission_date state tot_cases conf_cases prob_cases new_case pnew_case
## 1       01/25/2021    NE    187923         NA         NA      646         0
## 2       06/12/2021    AL    549716     423193     126523       97        31
## 3       06/15/2020    WI     25480      22932       2548      185        11
## 4       07/31/2020    ND      6602       6602          0      133         0
## 5       08/11/2020    GU       449         NA         NA       15         0
## 6       02/02/2020    GU         0         NA         NA        0        NA
## 7       12/27/2020    CA   2319593    2249322      70271    13613       664
## 8       05/15/2020    CT     36085      34059       2026      621       -11
## 9       07/02/2020    NC     76154      76023        131     2070        12
## 10      03/07/2021    NE    203157         NA         NA      374         0
## 11      06/09/2021    ND    110351     101268       9083       53        16
## 12      05/11/2020    CT     33765      31955       1810      211       -30
## 13      09/13/2020    MO    108944         NA         NA      663        18
## 14      03/31/2021    ND    103299      96726       6573      208        49
## 15      03/09/2020    WI         1         NA         NA        0        NA
## 16      03/09/2021    MS    299992     185580     114412      457       301
## 17      03/16/2020    ME        19         NA         NA        5        NA
## 18      08/17/2020    ID     27942      26091       1851      282        32
## 19      02/14/2021    GU      7699         NA         NA       10         0
## 20      03/13/2020    VT         2         NA         NA        0         0
## 21      07/09/2021    MO    631603         NA         NA     1639       472
## 22      08/23/2020    IL    221595     220178       1417     1893         0
## 23      11/19/2020    MI    423621     388797      34824     6304       384
## 24      08/08/2020    NV     55644         NA         NA      887         1
## 25      03/21/2021    VI      2858         NA         NA        6         0
## 26      05/09/2021    MI    974134     873038     101096     1301       175
```

```
## 27      03/08/2021    CA    3673952    3534725     139227      2378        197
## 28      09/22/2020    CA     810080     804380       5700      2204         44
## 29      07/06/2021    WI     678048     613152      64896        40          5
##      tot_death conf_death prob_death new_death pnew_death            created_at
## 1         1894         NA         NA        15          0 01/27/2021 12:00:00 AM
## 2        11355       9053       2302         5          2 06/14/2021 12:00:00 AM
## 3          700        694          6         2          0 06/16/2020 02:10:11 PM
## 4          103         NA         NA         0          0 08/01/2020 02:38:12 PM
## 5            5         NA         NA         0          0 08/12/2020 01:50:14 PM
## 6            0         NA         NA         0         NA 03/26/2020 04:22:39 PM
## 7        29708         NA         NA       570          0 12/28/2020 12:00:00 AM
## 8         3285       2524        761        66          8 05/16/2020 03:46:44 PM
## 9         1479       1479          0        11          0 07/02/2020 12:00:00 AM
## 10        2113         NA         NA         0          0 03/09/2021 12:00:00 AM
## 11        1517         NA         NA         0          0 06/10/2021 02:48:23 PM
## 12        3008       2294        714        41        -10 05/12/2020 05:26:08 PM
## 13        2132         NA         NA        19          0 09/15/2020 12:00:00 AM
## 14        1466         NA         NA         0          0 04/01/2021 02:13:30 PM
## 15           0         NA         NA         0         NA 03/26/2020 04:22:39 PM
## 16        7080       4916       2164         8          5 03/11/2021 12:00:00 AM
## 17           0         NA         NA         0         NA 03/26/2020 04:22:39 PM
## 18         273        245         28         4          0 08/18/2020 02:26:55 PM
## 19         130         NA         NA         0          0 02/15/2021 02:46:24 PM
## 20           0         NA         NA         0          0 03/15/2020 12:00:00 AM
## 21        9387         NA         NA         4          0 07/10/2021 02:17:36 PM
## 22        8089       7880        209         6          0 08/24/2020 01:54:34 PM
## 23        9011       8600        411        93          3 11/20/2020 12:00:00 AM
## 24         977         NA         NA        30          1 08/09/2020 02:03:10 PM
## 25          25         NA         NA         0          0 03/22/2021 01:57:09 PM
## 26       19723      18541       1182        58          1 05/10/2021 12:00:00 AM
## 27       59999         NA         NA       101          0 03/09/2021 12:00:00 AM
## 28       16351         NA         NA        71          0 09/23/2020 12:00:00 AM
## 29        8144       7323        821         0          0 07/07/2021 02:48:49 PM
##      consent_cases consent_deaths
## 1        Not agree      Not agree
## 2            Agree          Agree
## 3            Agree          Agree
## 4            Agree      Not agree
## 5        Not agree      Not agree
## 6        Not agree      Not agree
## 7            Agree      Not agree
## 8            Agree          Agree
## 9            Agree          Agree
## 10       Not agree      Not agree
## 11           Agree      Not agree
## 12           Agree          Agree
## 13       Not agree      Not agree
## 14           Agree      Not agree
## 15           Agree          Agree
## 16           Agree          Agree
## 17           Agree          Agree
## 18           Agree          Agree
## 19       Not agree      Not agree
## 20       Not agree      Not agree
```

```
## 21      Not agree      Not agree
## 22          Agree          Agree
## 23          Agree          Agree
## 24
## 25
## 26          Agree          Agree
## 27          Agree      Not agree
## 28          Agree      Not agree
## 29          Agree          Agree
```

List Variables

```
names(covid4)
```

```
##  [1] "submission_date" "state"           "tot_cases"       "conf_cases"
##  [5] "prob_cases"      "new_case"        "pnew_case"       "tot_death"
##  [9] "conf_death"      "prob_death"      "new_death"       "pnew_death"
## [13] "created_at"      "consent_cases"   "consent_deaths"
```

Print top 15 rows

```
head(covid4, n=15)
```

```
##     submission_date state tot_cases conf_cases prob_cases new_case pnew_case
## 1        01/25/2021    NE    187923         NA         NA      646         0
## 2        06/12/2021    AL    549716     423193     126523       97        31
## 3        06/15/2020    WI     25480      22932       2548      185        11
## 4        07/31/2020    ND      6602       6602          0      133         0
## 5        08/11/2020    GU       449         NA         NA       15         0
## 6        02/02/2020    GU         0         NA         NA        0        NA
## 7        12/27/2020    CA   2319593    2249322      70271    13613       664
## 8        05/15/2020    CT     36085      34059       2026      621       -11
## 9        07/02/2020    NC     76154      76023        131     2070        12
## 10       03/07/2021    NE    203157         NA         NA      374         0
## 11       06/09/2021    ND    110351     101268       9083       53        16
## 12       05/11/2020    CT     33765      31955       1810      211       -30
## 13       09/13/2020    MO    108944         NA         NA      663        18
## 14       03/31/2021    ND    103299      96726       6573      208        49
## 15       03/09/2020    WI         1         NA         NA        0        NA
##     tot_death conf_death prob_death new_death pnew_death          created_at
## 1        1894         NA         NA        15          0 01/27/2021 12:00:00 AM
## 2       11355       9053       2302         5          2 06/14/2021 12:00:00 AM
## 3         700        694          6         2          0 06/16/2020 02:10:11 PM
## 4         103         NA         NA         0          0 08/01/2020 02:38:12 PM
## 5           5         NA         NA         0          0 08/12/2020 01:50:14 PM
## 6           0         NA         NA         0         NA 03/26/2020 04:22:39 PM
## 7       29708         NA         NA       570          0 12/28/2020 12:00:00 AM
## 8        3285       2524        761        66          8 05/16/2020 03:46:44 PM
## 9        1479       1479          0        11          0 07/02/2020 12:00:00 AM
```

```
## 10      2113       NA       NA        0         0 03/09/2021 12:00:00 AM
## 11      1517       NA       NA        0         0 06/10/2021 02:48:23 PM
## 12      3008     2294      714       41       -10 05/12/2020 05:26:08 PM
## 13      2132       NA       NA       19         0 09/15/2020 12:00:00 AM
## 14      1466       NA       NA        0         0 04/01/2021 02:13:30 PM
## 15         0       NA       NA        0        NA 03/26/2020 04:22:39 PM
##    consent_cases consent_deaths
## 1     Not agree      Not agree
## 2         Agree          Agree
## 3         Agree          Agree
## 4         Agree      Not agree
## 5     Not agree      Not agree
## 6     Not agree      Not agree
## 7         Agree      Not agree
## 8         Agree          Agree
## 9         Agree          Agree
## 10    Not agree      Not agree
## 11        Agree      Not agree
## 12        Agree          Agree
## 13    Not agree      Not agree
## 14        Agree      Not agree
## 15        Agree          Agree
```

User defined function on dataset

```
testfunction <- function(x,y) {
x+y
}
a=covid4$tot_cases[1]
b=covid4$tot_cases[2]
testfunction(a,b)
```

```
## [1] 737639
```

Filter rows based on any logical criteria

```
covid5=as.data.frame(filter(covid4,covid4$new_death>5))
print(covid5)
```

```
##   submission_date state tot_cases conf_cases prob_cases new_case pnew_case
## 1      01/25/2021    NE    187923         NA         NA      646         0
## 2      12/27/2020    CA   2319593    2249322      70271    13613       664
## 3      05/15/2020    CT     36085      34059       2026      621       -11
## 4      07/02/2020    NC     76154      76023        131     2070        12
## 5      05/11/2020    CT     33765      31955       1810      211       -30
## 6      09/13/2020    MO    108944         NA         NA      663        18
## 7      03/09/2021    MS    299992     185580     114412      457       301
## 8      08/23/2020    IL    221595     220178       1417     1893         0
## 9      11/19/2020    MI    423621     388797      34824     6304       384
```

```
## 10        08/08/2020     NV      55644          NA           NA         887           1
## 11        05/09/2021     MI     974134       873038       101096        1301         175
## 12        03/08/2021     CA    3673952      3534725       139227        2378         197
## 13        09/22/2020     CA     810080       804380         5700        2204          44
##      tot_death conf_death prob_death new_death pnew_death          created_at
## 1         1894         NA         NA        15          0 01/27/2021 12:00:00 AM
## 2        29708         NA         NA       570          0 12/28/2020 12:00:00 AM
## 3         3285       2524        761        66          8 05/16/2020 03:46:44 PM
## 4         1479       1479          0        11          0 07/02/2020 12:00:00 AM
## 5         3008       2294        714        41        -10 05/12/2020 05:26:08 PM
## 6         2132         NA         NA        19          0 09/15/2020 12:00:00 AM
## 7         7080       4916       2164         8          5 03/11/2021 12:00:00 AM
## 8         8089       7880        209         6          0 08/24/2020 01:54:34 PM
## 9         9011       8600        411        93          3 11/20/2020 12:00:00 AM
## 10         977         NA         NA        30          1 08/09/2020 02:03:10 PM
## 11       19723      18541       1182        58          1 05/10/2021 12:00:00 AM
## 12       59999         NA         NA       101          0 03/09/2021 12:00:00 AM
## 13       16351         NA         NA        71          0 09/23/2020 12:00:00 AM
##      consent_cases consent_deaths
## 1        Not agree       Not agree
## 2            Agree       Not agree
## 3            Agree           Agree
## 4            Agree           Agree
## 5            Agree           Agree
## 6        Not agree       Not agree
## 7            Agree           Agree
## 8            Agree           Agree
## 9            Agree           Agree
## 10
## 11           Agree           Agree
## 12           Agree       Not agree
## 13           Agree       Not agree
```

Reshaping techniques

```
confirmedcases=covid4$conf_cases

probablecases= covid4$prob_cases

totalcases=cbind(confirmedcases,probablecases)

names(totalcases)[1] = "total cases"
names(totalcases)[2] = "probable cases"


print(totalcases)
```

```
##      confirmedcases probablecases
## [1,]             NA            NA
## [2,]         423193        126523
## [3,]          22932          2548
## [4,]           6602             0
```

```
##  [5,]             NA              NA
##  [6,]             NA              NA
##  [7,]        2249322           70271
##  [8,]          34059            2026
##  [9,]          76023             131
## [10,]             NA              NA
## [11,]         101268            9083
## [12,]          31955            1810
## [13,]             NA              NA
## [14,]          96726            6573
## [15,]             NA              NA
## [16,]         185580          114412
## [17,]             NA              NA
## [18,]          26091            1851
## [19,]             NA              NA
## [20,]             NA              NA
## [21,]             NA              NA
## [22,]         220178            1417
## [23,]         388797           34824
## [24,]             NA              NA
## [25,]             NA              NA
## [26,]         873038          101096
## [27,]        3534725          139227
## [28,]         804380            5700
## [29,]         613152           64896
## attr(,"names")
##  [1] "total cases"    "probable cases" NA               NA
##  [5] NA               NA               NA               NA
##  [9] NA               NA               NA               NA
## [13] NA               NA               NA               NA
## [17] NA               NA               NA               NA
## [21] NA               NA               NA               NA
## [25] NA               NA               NA               NA
## [29] NA               NA               NA               NA
## [33] NA               NA               NA               NA
## [37] NA               NA               NA               NA
## [41] NA               NA               NA               NA
## [45] NA               NA               NA               NA
## [49] NA               NA               NA               NA
## [53] NA               NA               NA               NA
## [57] NA               NA
```

Remove missing values

```
is.na(covid4)
```

```
##      submission_date state tot_cases conf_cases prob_cases new_case pnew_case
## [1,]           FALSE FALSE     FALSE       TRUE       TRUE    FALSE     FALSE
## [2,]           FALSE FALSE     FALSE      FALSE      FALSE    FALSE     FALSE
## [3,]           FALSE FALSE     FALSE      FALSE      FALSE    FALSE     FALSE
## [4,]           FALSE FALSE     FALSE      FALSE      FALSE    FALSE     FALSE
## [5,]           FALSE FALSE     FALSE       TRUE       TRUE    FALSE     FALSE
```

```
##  [6,]            FALSE FALSE      FALSE       TRUE       TRUE      FALSE       TRUE
##  [7,]            FALSE FALSE      FALSE      FALSE      FALSE      FALSE      FALSE
##  [8,]            FALSE FALSE      FALSE      FALSE      FALSE      FALSE      FALSE
##  [9,]            FALSE FALSE      FALSE      FALSE      FALSE      FALSE      FALSE
## [10,]            FALSE FALSE      FALSE       TRUE       TRUE      FALSE      FALSE
## [11,]            FALSE FALSE      FALSE      FALSE      FALSE      FALSE      FALSE
## [12,]            FALSE FALSE      FALSE      FALSE      FALSE      FALSE      FALSE
## [13,]            FALSE FALSE      FALSE       TRUE       TRUE      FALSE      FALSE
## [14,]            FALSE FALSE      FALSE      FALSE      FALSE      FALSE      FALSE
## [15,]            FALSE FALSE      FALSE       TRUE       TRUE      FALSE       TRUE
## [16,]            FALSE FALSE      FALSE      FALSE      FALSE      FALSE      FALSE
## [17,]            FALSE FALSE      FALSE       TRUE       TRUE      FALSE       TRUE
## [18,]            FALSE FALSE      FALSE      FALSE      FALSE      FALSE      FALSE
## [19,]            FALSE FALSE      FALSE       TRUE       TRUE      FALSE      FALSE
## [20,]            FALSE FALSE      FALSE       TRUE       TRUE      FALSE      FALSE
## [21,]            FALSE FALSE      FALSE       TRUE       TRUE      FALSE      FALSE
## [22,]            FALSE FALSE      FALSE      FALSE      FALSE      FALSE      FALSE
## [23,]            FALSE FALSE      FALSE      FALSE      FALSE      FALSE      FALSE
## [24,]            FALSE FALSE      FALSE       TRUE       TRUE      FALSE      FALSE
## [25,]            FALSE FALSE      FALSE       TRUE       TRUE      FALSE      FALSE
## [26,]            FALSE FALSE      FALSE      FALSE      FALSE      FALSE      FALSE
## [27,]            FALSE FALSE      FALSE      FALSE      FALSE      FALSE      FALSE
## [28,]            FALSE FALSE      FALSE      FALSE      FALSE      FALSE      FALSE
## [29,]            FALSE FALSE      FALSE      FALSE      FALSE      FALSE      FALSE
##       tot_death conf_death prob_death new_death pnew_death created_at
##  [1,]     FALSE       TRUE       TRUE     FALSE      FALSE      FALSE
##  [2,]     FALSE      FALSE      FALSE     FALSE      FALSE      FALSE
##  [3,]     FALSE      FALSE      FALSE     FALSE      FALSE      FALSE
##  [4,]     FALSE       TRUE       TRUE     FALSE      FALSE      FALSE
##  [5,]     FALSE       TRUE       TRUE     FALSE      FALSE      FALSE
##  [6,]     FALSE       TRUE       TRUE     FALSE       TRUE      FALSE
##  [7,]     FALSE       TRUE       TRUE     FALSE      FALSE      FALSE
##  [8,]     FALSE      FALSE      FALSE     FALSE      FALSE      FALSE
##  [9,]     FALSE      FALSE      FALSE     FALSE      FALSE      FALSE
## [10,]     FALSE       TRUE       TRUE     FALSE      FALSE      FALSE
## [11,]     FALSE       TRUE       TRUE     FALSE      FALSE      FALSE
## [12,]     FALSE      FALSE      FALSE     FALSE      FALSE      FALSE
## [13,]     FALSE       TRUE       TRUE     FALSE      FALSE      FALSE
## [14,]     FALSE       TRUE       TRUE     FALSE      FALSE      FALSE
## [15,]     FALSE       TRUE       TRUE     FALSE       TRUE      FALSE
## [16,]     FALSE      FALSE      FALSE     FALSE      FALSE      FALSE
## [17,]     FALSE       TRUE       TRUE     FALSE       TRUE      FALSE
## [18,]     FALSE      FALSE      FALSE     FALSE      FALSE      FALSE
## [19,]     FALSE       TRUE       TRUE     FALSE      FALSE      FALSE
## [20,]     FALSE       TRUE       TRUE     FALSE      FALSE      FALSE
## [21,]     FALSE       TRUE       TRUE     FALSE      FALSE      FALSE
## [22,]     FALSE      FALSE      FALSE     FALSE      FALSE      FALSE
## [23,]     FALSE      FALSE      FALSE     FALSE      FALSE      FALSE
## [24,]     FALSE       TRUE       TRUE     FALSE      FALSE      FALSE
## [25,]     FALSE       TRUE       TRUE     FALSE      FALSE      FALSE
## [26,]     FALSE      FALSE      FALSE     FALSE      FALSE      FALSE
## [27,]     FALSE       TRUE       TRUE     FALSE      FALSE      FALSE
## [28,]     FALSE       TRUE       TRUE     FALSE      FALSE      FALSE
## [29,]     FALSE      FALSE      FALSE     FALSE      FALSE      FALSE
```

```
##        consent_cases consent_deaths
##  [1,]          FALSE          FALSE
##  [2,]          FALSE          FALSE
##  [3,]          FALSE          FALSE
##  [4,]          FALSE          FALSE
##  [5,]          FALSE          FALSE
##  [6,]          FALSE          FALSE
##  [7,]          FALSE          FALSE
##  [8,]          FALSE          FALSE
##  [9,]          FALSE          FALSE
## [10,]          FALSE          FALSE
## [11,]          FALSE          FALSE
## [12,]          FALSE          FALSE
## [13,]          FALSE          FALSE
## [14,]          FALSE          FALSE
## [15,]          FALSE          FALSE
## [16,]          FALSE          FALSE
## [17,]          FALSE          FALSE
## [18,]          FALSE          FALSE
## [19,]          FALSE          FALSE
## [20,]          FALSE          FALSE
## [21,]          FALSE          FALSE
## [22,]          FALSE          FALSE
## [23,]          FALSE          FALSE
## [24,]          FALSE          FALSE
## [25,]          FALSE          FALSE
## [26,]          FALSE          FALSE
## [27,]          FALSE          FALSE
## [28,]          FALSE          FALSE
## [29,]          FALSE          FALSE
```

```
covid6=as.data.frame(filter(covid4,!is.na(covid4$conf_cases)))
is.na(covid6$conf_cases)
```

```
##  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [13] FALSE FALSE FALSE FALSE FALSE
```

Identify duplicated data

```
duplicated(covid4$new_death)
```

```
##  [1] FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE  TRUE FALSE
## [13] FALSE  TRUE  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE
## [25]  TRUE FALSE FALSE FALSE  TRUE
```

Remove duplicated data

```
covid7=as.data.frame(covid4 %>% distinct(covid4$new_case))
```

```
covid7
```

```
##    covid4$new_case
## 1              646
## 2               97
## 3              185
## 4              133
## 5               15
## 6                0
## 7            13613
## 8              621
## 9             2070
## 10             374
## 11              53
## 12             211
## 13             663
## 14             208
## 15             457
## 16               5
## 17             282
## 18              10
## 19            1639
## 20            1893
## 21            6304
## 22             887
## 23               6
## 24            1301
## 25            2378
## 26            2204
## 27              40
```

Reorder multiple rows in descending order

```
covid9=as.data.frame(covid7 %>% arrange(desc(covid7)))
covid9
```

```
##    covid4$new_case
## 1            13613
## 2             6304
## 3             2378
## 4             2204
## 5             2070
## 6             1893
## 7             1639
## 8             1301
## 9              887
## 10             663
## 11             646
## 12             621
## 13             457
## 14             374
## 15             282
## 16             211
## 17             208
```

```
## 18                    185
## 19                    133
## 20                     97
## 21                     53
## 22                     40
## 23                     15
## 24                     10
## 25                      6
## 26                      5
## 27                      0
```

Rename some of the column names

```
names(covid4)[1] = "name1"
names(covid4)[2] = "name2"
```

```
print(covid4)
```

```
##          name1 name2 tot_cases conf_cases prob_cases new_case pnew_case
## 1  01/25/2021    NE    187923         NA         NA      646         0
## 2  06/12/2021    AL    549716     423193     126523       97        31
## 3  06/15/2020    WI     25480      22932       2548      185        11
## 4  07/31/2020    ND      6602       6602          0      133         0
## 5  08/11/2020    GU       449         NA         NA       15         0
## 6  02/02/2020    GU         0         NA         NA        0        NA
## 7  12/27/2020    CA   2319593    2249322      70271    13613       664
## 8  05/15/2020    CT     36085      34059       2026      621       -11
## 9  07/02/2020    NC     76154      76023        131     2070        12
## 10 03/07/2021    NE    203157         NA         NA      374         0
## 11 06/09/2021    ND    110351     101268       9083       53        16
## 12 05/11/2020    CT     33765      31955       1810      211       -30
## 13 09/13/2020    MO    108944         NA         NA      663        18
## 14 03/31/2021    ND    103299      96726       6573      208        49
## 15 03/09/2020    WI         1         NA         NA        0        NA
## 16 03/09/2021    MS    299992     185580     114412      457       301
## 17 03/16/2020    ME        19         NA         NA        5        NA
## 18 08/17/2020    ID     27942      26091       1851      282        32
## 19 02/14/2021    GU      7699         NA         NA       10         0
## 20 03/13/2020    VT         2         NA         NA        0         0
## 21 07/09/2021    MO    631603         NA         NA     1639       472
## 22 08/23/2020    IL    221595     220178       1417     1893         0
## 23 11/19/2020    MI    423621     388797      34824     6304       384
## 24 08/08/2020    NV     55644         NA         NA      887         1
## 25 03/21/2021    VI      2858         NA         NA        6         0
## 26 05/09/2021    MI    974134     873038     101096     1301       175
## 27 03/08/2021    CA   3673952    3534725     139227     2378       197
## 28 09/22/2020    CA    810080     804380       5700     2204        44
## 29 07/06/2021    WI    678048     613152      64896       40         5
##    tot_death conf_death prob_death new_death pnew_death              created_at
## 1       1894         NA         NA        15          0 01/27/2021 12:00:00 AM
## 2      11355       9053       2302         5          2 06/14/2021 12:00:00 AM
```

```
## 3      700     694       6       2     0 06/16/2020 02:10:11 PM
## 4      103      NA      NA       0     0 08/01/2020 02:38:12 PM
## 5        5      NA      NA       0     0 08/12/2020 01:50:14 PM
## 6        0      NA      NA       0    NA 03/26/2020 04:22:39 PM
## 7    29708      NA      NA     570     0 12/28/2020 12:00:00 AM
## 8     3285    2524     761      66     8 05/16/2020 03:46:44 PM
## 9     1479    1479       0      11     0 07/02/2020 12:00:00 AM
## 10    2113      NA      NA       0     0 03/09/2021 12:00:00 AM
## 11    1517      NA      NA       0     0 06/10/2021 02:48:23 PM
## 12    3008    2294     714      41   -10 05/12/2020 05:26:08 PM
## 13    2132      NA      NA      19     0 09/15/2020 12:00:00 AM
## 14    1466      NA      NA       0     0 04/01/2021 02:13:30 PM
## 15       0      NA      NA       0    NA 03/26/2020 04:22:39 PM
## 16    7080    4916    2164       8     5 03/11/2021 12:00:00 AM
## 17       0      NA      NA       0    NA 03/26/2020 04:22:39 PM
## 18     273     245      28       4     0 08/18/2020 02:26:55 PM
## 19     130      NA      NA       0     0 02/15/2021 02:46:24 PM
## 20       0      NA      NA       0     0 03/15/2020 12:00:00 AM
## 21    9387      NA      NA       4     0 07/10/2021 02:17:36 PM
## 22    8089    7880     209       6     0 08/24/2020 01:54:34 PM
## 23    9011    8600     411      93     3 11/20/2020 12:00:00 AM
## 24     977      NA      NA      30     1 08/09/2020 02:03:10 PM
## 25      25      NA      NA       0     0 03/22/2021 01:57:09 PM
## 26   19723   18541    1182      58     1 05/10/2021 12:00:00 AM
## 27   59999      NA      NA     101     0 03/09/2021 12:00:00 AM
## 28   16351      NA      NA      71     0 09/23/2020 12:00:00 AM
## 29    8144    7323     821       0     0 07/07/2021 02:48:49 PM
##    consent_cases consent_deaths
## 1     Not agree      Not agree
## 2         Agree          Agree
## 3         Agree          Agree
## 4         Agree      Not agree
## 5     Not agree      Not agree
## 6     Not agree      Not agree
## 7         Agree      Not agree
## 8         Agree          Agree
## 9         Agree          Agree
## 10    Not agree      Not agree
## 11        Agree      Not agree
## 12        Agree          Agree
## 13    Not agree      Not agree
## 14        Agree      Not agree
## 15        Agree          Agree
## 16        Agree          Agree
## 17        Agree          Agree
## 18        Agree          Agree
## 19    Not agree      Not agree
## 20    Not agree      Not agree
## 21    Not agree      Not agree
## 22        Agree          Agree
## 23        Agree          Agree
## 24
## 25
## 26        Agree          Agree
```

```
## 27        Agree      Not agree
## 28        Agree      Not agree
## 29        Agree          Agree
```

Adding new variables

```
covid10=as.data.frame(covid4 %>% mutate(newcolumn = new_death+4))
covid10
```

```
##          name1 name2 tot_cases conf_cases prob_cases new_case pnew_case
## 1  01/25/2021    NE    187923         NA         NA      646         0
## 2  06/12/2021    AL    549716     423193     126523       97        31
## 3  06/15/2020    WI     25480      22932       2548      185        11
## 4  07/31/2020    ND      6602       6602          0      133         0
## 5  08/11/2020    GU       449         NA         NA       15         0
## 6  02/02/2020    GU         0         NA         NA        0        NA
## 7  12/27/2020    CA   2319593    2249322      70271    13613       664
## 8  05/15/2020    CT     36085      34059       2026      621       -11
## 9  07/02/2020    NC     76154      76023        131     2070        12
## 10 03/07/2021    NE    203157         NA         NA      374         0
## 11 06/09/2021    ND    110351     101268       9083       53        16
## 12 05/11/2020    CT     33765      31955       1810      211       -30
## 13 09/13/2020    MO    108944         NA         NA      663        18
## 14 03/31/2021    ND    103299      96726       6573      208        49
## 15 03/09/2020    WI         1         NA         NA        0        NA
## 16 03/09/2021    MS    299992     185580     114412      457       301
## 17 03/16/2020    ME        19         NA         NA        5        NA
## 18 08/17/2020    ID     27942      26091       1851      282        32
## 19 02/14/2021    GU      7699         NA         NA       10         0
## 20 03/13/2020    VT         2         NA         NA        0         0
## 21 07/09/2021    MO    631603         NA         NA     1639       472
## 22 08/23/2020    IL    221595     220178       1417     1893         0
## 23 11/19/2020    MI    423621     388797      34824     6304       384
## 24 08/08/2020    NV     55644         NA         NA      887         1
## 25 03/21/2021    VI      2858         NA         NA        6         0
## 26 05/09/2021    MI    974134     873038     101096     1301       175
## 27 03/08/2021    CA   3673952    3534725     139227     2378       197
## 28 09/22/2020    CA    810080     804380       5700     2204        44
## 29 07/06/2021    WI    678048     613152      64896       40         5
##    tot_death conf_death prob_death new_death pnew_death           created_at
## 1       1894         NA         NA        15          0 01/27/2021 12:00:00 AM
## 2      11355       9053       2302         5          2 06/14/2021 12:00:00 AM
## 3        700        694          6         2          0 06/16/2020 02:10:11 PM
## 4        103         NA         NA         0          0 08/01/2020 02:38:12 PM
## 5          5         NA         NA         0          0 08/12/2020 01:50:14 PM
## 6          0         NA         NA         0         NA 03/26/2020 04:22:39 PM
## 7      29708         NA         NA       570          0 12/28/2020 12:00:00 AM
## 8       3285       2524        761        66          8 05/16/2020 03:46:44 PM
## 9       1479       1479          0        11          0 07/02/2020 12:00:00 AM
## 10      2113         NA         NA         0          0 03/09/2021 12:00:00 AM
## 11      1517         NA         NA         0          0 06/10/2021 02:48:23 PM
## 12      3008       2294        714        41        -10 05/12/2020 05:26:08 PM
```

```
## 13     2132        NA        NA       19        0 09/15/2020 12:00:00 AM
## 14     1466        NA        NA        0        0 04/01/2021 02:13:30 PM
## 15        0        NA        NA        0       NA 03/26/2020 04:22:39 PM
## 16     7080      4916      2164        8        5 03/11/2021 12:00:00 AM
## 17        0        NA        NA        0       NA 03/26/2020 04:22:39 PM
## 18      273       245        28        4        0 08/18/2020 02:26:55 PM
## 19      130        NA        NA        0        0 02/15/2021 02:46:24 PM
## 20        0        NA        NA        0        0 03/15/2020 12:00:00 AM
## 21     9387        NA        NA        4        0 07/10/2021 02:17:36 PM
## 22     8089      7880       209        6        0 08/24/2020 01:54:34 PM
## 23     9011      8600       411       93        3 11/20/2020 12:00:00 AM
## 24      977        NA        NA       30        1 08/09/2020 02:03:10 PM
## 25       25        NA        NA        0        0 03/22/2021 01:57:09 PM
## 26    19723     18541      1182       58        1 05/10/2021 12:00:00 AM
## 27    59999        NA        NA      101        0 03/09/2021 12:00:00 AM
## 28    16351        NA        NA       71        0 09/23/2020 12:00:00 AM
## 29     8144      7323       821        0        0 07/07/2021 02:48:49 PM
##     consent_cases consent_deaths newcolumn
## 1      Not agree      Not agree        19
## 2          Agree          Agree         9
## 3          Agree          Agree         6
## 4          Agree      Not agree         4
## 5      Not agree      Not agree         4
## 6      Not agree      Not agree         4
## 7          Agree      Not agree       574
## 8          Agree          Agree        70
## 9          Agree          Agree        15
## 10     Not agree      Not agree         4
## 11         Agree      Not agree         4
## 12         Agree          Agree        45
## 13     Not agree      Not agree        23
## 14         Agree      Not agree         4
## 15         Agree          Agree         4
## 16         Agree          Agree        12
## 17         Agree          Agree         4
## 18         Agree          Agree         8
## 19     Not agree      Not agree         4
## 20     Not agree      Not agree         4
## 21     Not agree      Not agree         8
## 22         Agree          Agree        10
## 23         Agree          Agree        97
## 24                                     34
## 25                                      4
## 26         Agree          Agree        62
## 27         Agree      Not agree       105
## 28         Agree      Not agree        75
## 29         Agree          Agree         4
```

Create a training set using random number generator engine

```
set.seed(143)
covid11=as.data.frame(covid4 %>% sample_n(10, replace = FALSE) )
```

```
covid11
```

```
##         name1 name2 tot_cases conf_cases prob_cases new_case pnew_case
## 1  08/08/2020    NV     55644         NA         NA      887         1
## 2  02/02/2020    GU         0         NA         NA        0        NA
## 3  03/13/2020    VT         2         NA         NA        0         0
## 4  05/15/2020    CT     36085      34059       2026      621       -11
## 5  01/25/2021    NE    187923         NA         NA      646         0
## 6  05/11/2020    CT     33765      31955       1810      211       -30
## 7  08/23/2020    IL    221595     220178       1417     1893         0
## 8  03/21/2021    VI      2858         NA         NA        6         0
## 9  05/09/2021    MI    974134     873038     101096     1301       175
## 10 07/02/2020    NC     76154      76023        131     2070        12
##    tot_death conf_death prob_death new_death pnew_death          created_at
## 1        977         NA         NA        30          1 08/09/2020 02:03:10 PM
## 2          0         NA         NA         0         NA 03/26/2020 04:22:39 PM
## 3          0         NA         NA         0          0 03/15/2020 12:00:00 AM
## 4       3285       2524        761        66          8 05/16/2020 03:46:44 PM
## 5       1894         NA         NA        15          0 01/27/2021 12:00:00 AM
## 6       3008       2294        714        41        -10 05/12/2020 05:26:08 PM
## 7       8089       7880        209         6          0 08/24/2020 01:54:34 PM
## 8         25         NA         NA         0          0 03/22/2021 01:57:09 PM
## 9      19723      18541       1182        58          1 05/10/2021 12:00:00 AM
## 10      1479       1479          0        11          0 07/02/2020 12:00:00 AM
##    consent_cases consent_deaths
## 1
## 2      Not agree      Not agree
## 3      Not agree      Not agree
## 4          Agree          Agree
## 5      Not agree      Not agree
## 6          Agree          Agree
## 7          Agree          Agree
## 8
## 9          Agree          Agree
## 10         Agree          Agree
```

Print the summary statistics

```
summary(covid4)
```

```
##     name1               name2              tot_cases          conf_cases
##  Length:29          Length:29          Min.   :      0   Min.   :   6602
##  Class :character   Class :character   1st Qu.:   7699   1st Qu.:  34059
##  Mode  :character   Mode  :character   Median :  103299  Median : 185580
##                                        Mean   : 398921   Mean   : 569884
##                                        3rd Qu.: 423621   3rd Qu.: 613152
##                                        Max.   :3673952   Max.   :3534725
##                                                          NA's   :12
##    prob_cases        new_case          pnew_case         tot_death
##  Min.   :     0   Min.   :      0   Min.   :-30.00    Min.   :     0
##  1st Qu.:  1851   1st Qu.:     40   1st Qu.:  0.00    1st Qu.:   130
```

```
##   Median :  6573    Median :   282    Median : 11.50    Median : 1894
##   Mean   : 40140    Mean   :  1252    Mean   : 91.19    Mean   : 6826
##   3rd Qu.: 70271    3rd Qu.:  1301    3rd Qu.: 47.75    3rd Qu.: 8144
##   Max.   :139227    Max.   : 13613    Max.   :664.00    Max.   :59999
##   NA's   :12                         NA's   :3
##    conf_death       prob_death       new_death        pnew_death
##   Min.   :  245    Min.   :   0.0    Min.   :  0.00    Min.   :-10.0000
##   1st Qu.: 1886    1st Qu.: 118.5    1st Qu.:  0.00    1st Qu.:  0.0000
##   Median : 4916    Median : 714.0    Median :  4.00    Median :  0.0000
##   Mean   : 5777    Mean   : 781.6    Mean   : 38.07    Mean   :  0.3846
##   3rd Qu.: 8240    3rd Qu.:1001.5    3rd Qu.: 30.00    3rd Qu.:  0.0000
##   Max.   :18541    Max.   :2302.0    Max.   :570.00    Max.   :  8.0000
##   NA's   :18       NA's   :18                          NA's   :3
##    created_at       consent_cases     consent_deaths
##   Length:29        Length:29         Length:29
##   Class :character Class :character  Class :character
##   Mode  :character Mode  :character  Mode  :character
##
##
##
##
```

Calculating Mean

```
mean(covid4$new_case)
```

```
## [1] 1251.552
```

Calculating Median

```
median(covid4$new_case)
```

```
## [1] 282
```

Calculating Mode

```
mode(covid4$new_case)
```
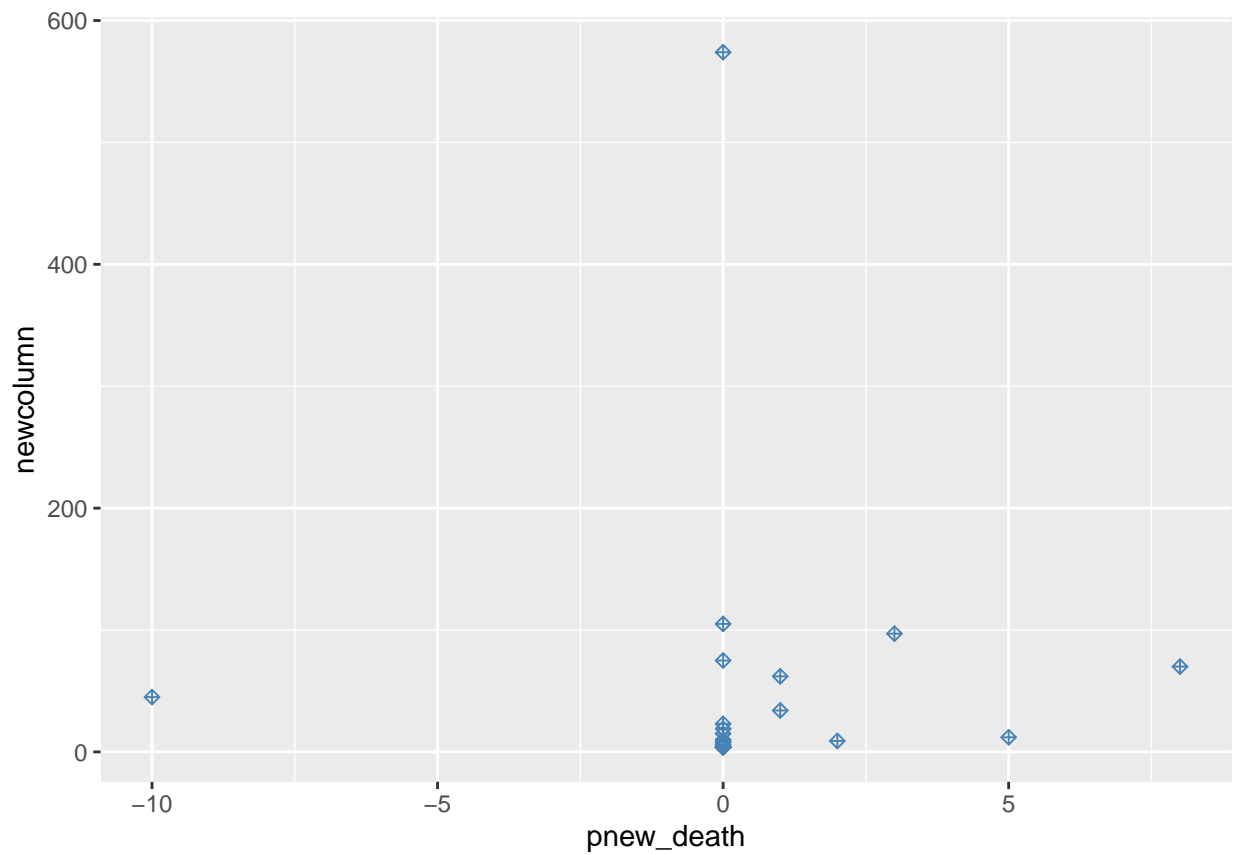
```
## [1] "numeric"
```

Calculating Median

```
median(covid4$new_case)
```
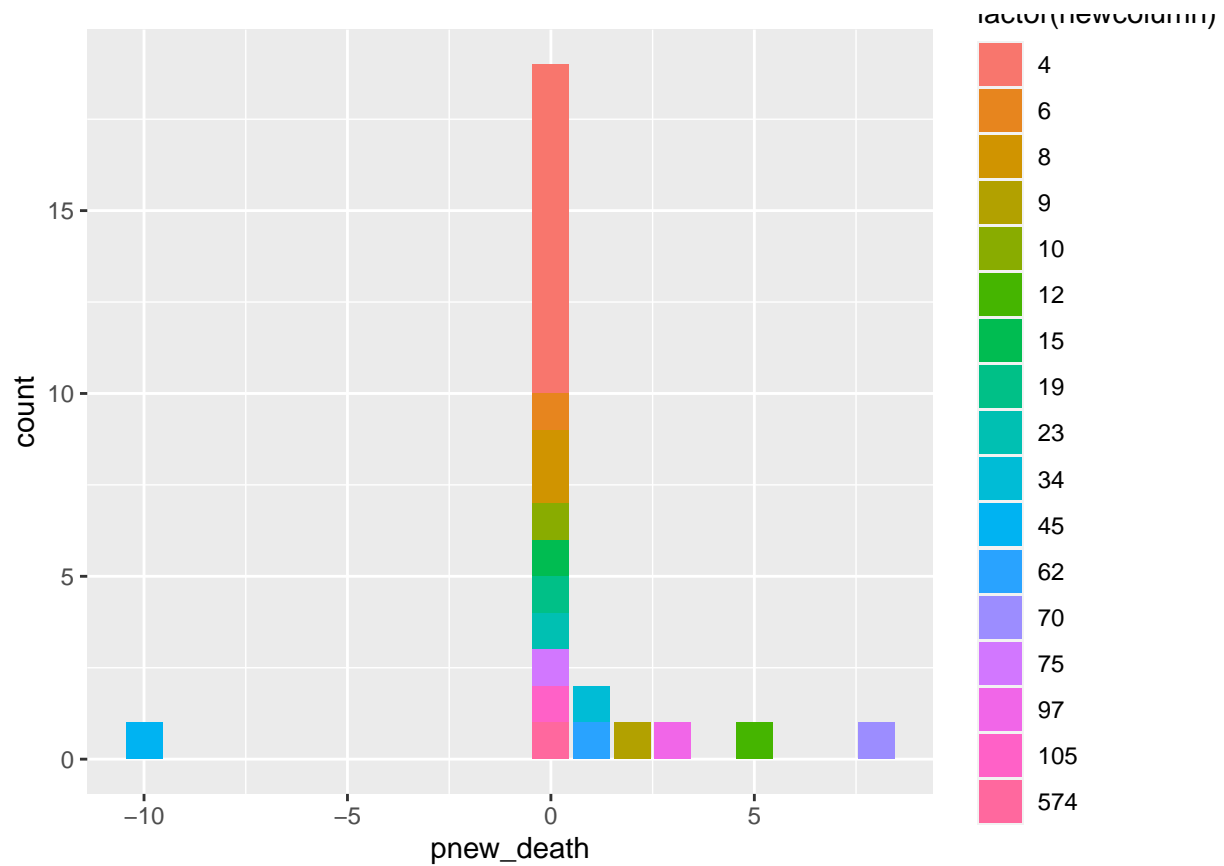
```
## [1] 282
```

Scatter plot for any 2 variables

```
covid10=as.data.frame(filter(covid10,!is.na(covid10$pnew_death)))
ggplot(data = covid10, aes(x=pnew_death,y=newcolumn))+geom_point(color="steelblue",shape=9)
```


Scatter bar plot for any 2 variables

```
ggplot(data=covid10,aes(x=pnew_death,fill=factor(newcolumn)))+geom_bar()
```

factor(newcolumn)

4
6
8
9
10
12
15
19
23
34
45
62
70
75
97
105
574

count

pnew_death

Finding the correlation between any 2 variables

```
x=covid4[,"tot_death"]
y=covid4[,"new_death"]
xycorr<-cor(x,y,method="pearson")
xycorr
```

```
## [1] 0.5203182
```

Conclusion

Based on above calculation x and y are correlated