```
R version 4.1.1 (2021-08-10) -- "Kick Things"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> # JRodoni_HW05_script.R
> # C:/Users/jackr/OneDrive/Desktop/Graduate School Courses/
> #   STAT 604 - STAT Computation/Homeworks/JRodoni_HW04_script.R
> # Created By: Jack Rodoni
> # Creation Date: 09/20/2021
> # Purpose: STAT 604 Homework 5
> # Last Executed: 09/21/2021
> Sys.time()
[1] "2021-09-21 13:17:24 CDT"
> 
> ls()
character(0)
> rm(list = ls())
> library()
> search()
[1] ".GlobalEnv"        "package:stats"     "package:graphics" 
[4] "package:grDevices" "package:utils"     "package:datasets" 
[7] "package:methods"   "Autoloads"         "package:base"     
> 
> 
> # 2.) Import the COVID Activity.csv file into an R data frame using the appropriate function. D
O NOT
> #     include code to display the data frame upon creation as it will likely overload the conso
le due to
> #     the amount of data.
> 
> #     (a) Show the structure of the new data frame.
> COVID_Activity <- read.csv("C:/Users/jackr/OneDrive/Desktop/Graduate School Courses/STAT 604 - 
STAT Computation/Rdata/COVID Activity.csv")
> str(COVID_Activity)
'data.frame':   2132949 obs. of  13 variables:
 $ POSITIVE_CASES_COUNT    : int  41851 41928 42025 42188 42309 42309 42309 42686 42760 42862 ...
 $ COUNTY_NAME             : chr  "Guilford" "Guilford" "Guilford" "Guilford" ...
 $ PROVINCE_STATE_NAME     : chr  "North Carolina" "North Carolina" "North Carolina" "North Carol
ina" ...
 $ REPORT_DATE             : chr  "2021-03-22" "2021-03-23" "2021-03-24" "2021-03-25" ...
 $ CONTINENT_NAME          : chr  "America" "America" "America" "America" ...
 $ DATA_SOURCE_NAME        : chr  "New York Times" "New York Times" "New York Times" "New York Ti
mes" ...
 $ DEATH_NEW_COUNT         : int  5 3 8 1 2 0 0 8 0 6 ...
 $ COUNTY_FIPS_NUMBER      : int  37081 37081 37081 37081 37081 37081 37081 37081 37081 37081 ...
 $ COUNTRY_ALPHA_3_CODE    : chr  "USA" "USA" "USA" "USA" ...
 $ COUNTRY_SHORT_NAME      : chr  "United States" "United States" "United States" "United States"
 ...
 $ COUNTRY_ALPHA_2_CODE    : chr  "US" "US" "US" "US" ...
 $ POSITIVE_NEW_CASES_COUNT: int  174 77 97 163 121 0 0 377 74 102 ...
 $ DEATH_COUNT             : int  589 592 600 601 603 603 603 611 611 617 ...
> 
> #     (b) Some of the columns have very long names that could be shortened without any
> #         negative consequences. However, the column order has not always been consistent in
> #         the download of this data so we need to make the changes using a value replacement
> #          You can use the names function to access the column names as a vector that you can
> #          manipulate as you would any other vector. (Remember you are not actually changing
> #          anything unless you use an assignment statement.) Change the columns shown in the
> #          table below:
>
> names(COVID_Activity)[c(1,7,12,13)] = c("TOTAL_CASES", "NEW_DEATHS", "NEW_CASES", "TOTAL_DEATHS
")
>
> #      (c) Display the first 10 rows and all columns of the modified data frame
> COVID_Activity[1:10,]
   TOTAL_CASES COUNTY_NAME PROVINCE_STATE_NAME REPORT_DATE CONTINENT_NAME
1        41851    Guilford      North Carolina  2021-03-22        America
2        41928    Guilford      North Carolina  2021-03-23        America
3        42025    Guilford      North Carolina  2021-03-24        America
4        42188    Guilford      North Carolina  2021-03-25        America
5        42309    Guilford      North Carolina  2021-03-26        America
6        42309    Guilford      North Carolina  2021-03-27        America
7        42309    Guilford      North Carolina  2021-03-28        America
8        42686    Guilford      North Carolina  2021-03-29        America
9        42760    Guilford      North Carolina  2021-03-30        America
10       42862    Guilford      North Carolina  2021-03-31        America
   DATA_SOURCE_NAME NEW_DEATHS COUNTY_FIPS_NUMBER COUNTRY_ALPHA_3_CODE
1    New York Times          5              37081                  USA
2    New York Times          3              37081                  USA
3    New York Times          8              37081                  USA
4    New York Times          1              37081                  USA
5    New York Times          2              37081                  USA
6    New York Times          0              37081                  USA
7    New York Times          0              37081                  USA
8    New York Times          8              37081                  USA
9    New York Times          0              37081                  USA
10   New York Times          6              37081                  USA
   COUNTRY_SHORT_NAME COUNTRY_ALPHA_2_CODE NEW_CASES TOTAL_DEATHS
1       United States                   US       174          589
2       United States                   US        77          592
3       United States                   US        97          600
4       United States                   US       163          601
5       United States                   US       121          603
6       United States                   US         0          603
7       United States                   US         0          603
8       United States                   US       377          611
9       United States                   US        74          611
10      United States                   US       102          617
>
> # 3.)  Create a new data frame that is a subset of the data frame created from the CSV file. Th
e subset
> #      will contain only rows for the state of Texas. Use a list of column numbers in your subs
cript so
> #      the new data frame contains only the following columns in the order shown: COUNTY_NAME,
> #      REPORT_DATE, NEW_CASES, TOTAL_CASES, NEW_DEATHS, TOTAL_DEATHS. Display in the
> #      console the structure of the new data frame.
>
> Covid_Texas = subset(COVID_Activity[,c(2,4,12,1,7,13)], COVID_Activity$PROVINCE_STATE_NAME == "
Texas")
>
> # 4.) Write an expression to import the txt file into a data frame. You may spread the expressi
on
> #     across multiple lines in your script so it does not get cut off when you convert the scri
pt to pdf if
> #     you will insert your breaks between elements of the expression or function.
>
>
> PopTable <- read.table("C:/Users/jackr/OneDrive/Desktop/Graduate School Courses/STAT 604 - STAT
 Computation/RData/Master Location Pop Table.txt",
+                         header = TRUE, sep = ":", quote = "\"")
>
> #      (a) Display the structure of the new data frame
> str(PopTable)
'data.frame':   3483 obs. of  10 variables:
 $ ï..COUNTRY_SHORT_NAME      : chr  "Afghanistan" "Albania" "Algeria" "Andorra" ...
 $ COUNTRY_ALPHA_3_CODE       : chr  "AFG" "ALB" "DZA" "AND" ...
 $ COUNTRY_ALPHA_2_CODE       : chr  "AF" "AL" "DZ" "AD" ...
 $ PROVINCE_STATE_NAME        : chr  "" "" "" "" ...
 $ COUNTY_NAME                : chr  "" "" "" "" ...
 $ COUNTY_FIPS_NUMBER         : int  NA NA NA NA NA NA NA NA NA NA ...
 $ GEO_LATITUDE               : num  34 40.7 28.6 42.5 -12.8 ...
 $ GEO_LONGITUDE              : num  65.53 20.08 2.64 1.59 17.81 ...
 $ GEO_REGION_POPULATION_COUNT: int  38041757 2880913 43053054 77146 31825299 14872 97115 4478067
5 2957728 106310 ...
 $ DATA_SOURCE_NAME           : chr  "United Nations - 2019 Median" "United Nations - 2019 Median
" "United Nations - 2019 Median" "United Nations - 2019 Median" ...
>
> #     (b) Change the name of the column that contains population data to POPULATION to be more
concise
> names(PopTable)[9] = "POPULATION"
>
> #     (c) Display the structure again showing the modifications
> str(PopTable)
'data.frame':   3483 obs. of  10 variables:
 $ ï..COUNTRY_SHORT_NAME: chr  "Afghanistan" "Albania" "Algeria" "Andorra" ...
 $ COUNTRY_ALPHA_3_CODE : chr  "AFG" "ALB" "DZA" "AND" ...
 $ COUNTRY_ALPHA_2_CODE : chr  "AF" "AL" "DZ" "AD" ...
 $ PROVINCE_STATE_NAME  : chr  "" "" "" "" ...
 $ COUNTY_NAME          : chr  "" "" "" "" ...
 $ COUNTY_FIPS_NUMBER   : int  NA NA NA NA NA NA NA NA NA NA ...
 $ GEO_LATITUDE         : num  34 40.7 28.6 42.5 -12.8 ...
 $ GEO_LONGITUDE        : num  65.53 20.08 2.64 1.59 17.81 ...
 $ POPULATION           : int  38041757 2880913 43053054 77146 31825299 14872 97115 44780675 2957
728 106310 ...
 $ DATA_SOURCE_NAME     : chr  "United Nations - 2019 Median" "United Nations - 2019 Median" "Uni
ted Nations - 2019 Median" "United Nations - 2019 Median" ...
>
> #     (d) Display the first 10 rows of the modified data frame
> head(PopTable, n = 10)
   ï..COUNTRY_SHORT_NAME COUNTRY_ALPHA_3_CODE COUNTRY_ALPHA_2_CODE
1            Afghanistan                  AFG                   AF
2                Albania                  ALB                   AL
3                Algeria                  DZA                   DZ
4                Andorra                  AND                   AD
5                 Angola                  AGO                   AO
6               Anguilla                  AIA                   AI
7    Antigua and Barbuda                  ATG                   AG
8              Argentina                  ARG                   AR
9                Armenia                  ARM                   AM
10                 Aruba                  ABW                   AW
   PROVINCE_STATE_NAME COUNTY_NAME COUNTY_FIPS_NUMBER GEO_LATITUDE
1                                                  NA      34.0230
2                                                  NA      40.6540
3                                                  NA      28.6045
4                                                  NA      42.5425
5                                                  NA     -12.8360
6                                                  NA      18.2177
7                                                  NA      17.6250
8                                                  NA     -33.1660
9                                                  NA      40.5600
10                                                 NA      12.5176
   GEO_LONGITUDE POPULATION             DATA_SOURCE_NAME
1        65.5267   38041757 United Nations - 2019 Median
2        20.0760    2880913 United Nations - 2019 Median
3         2.6400   43053054 United Nations - 2019 Median
4         1.5893      77146 United Nations - 2019 Median
5        17.8080   31825299 United Nations - 2019 Median
6       -63.0406      14872 United Nations - 2019 Median
7       -61.7860      97115 United Nations - 2019 Median
8       -64.3100   44780675 United Nations - 2019 Median
9        44.4490    2957728 United Nations - 2019 Median
10      -69.9818     106310 United Nations - 2019 Median
>
> # 5.) Create a new data frame by combining the “Texas” data frame with the “population” data fr
ame
```

```
> #     that you created in the previous step. When the “population” data frame is referenced in 
your
> #     expression to combine the data frames, use expressions for the rows and columns so that o
nly
> #     rows from Texas are selected and only the COUNTY_NAME and POPULATION columns. Include
> #     non-matches in the resulting data frame. The new data frame should have 153,255 rows
> 
> Merged_df = merge(Covid_Texas, 
+             subset(PopTable[,c("POPULATION", "COUNTY_NAME")], PopTable$PROVINCE_STATE_NAME == "
Texas"),
+             all = TRUE)
> 
> #     (a) Display a summary of the new data frame
> summary(Merged_df)
 COUNTY_NAME        REPORT_DATE          NEW_CASES        TOTAL_CASES    
 Length:153255      Length:153255      Min.   :-1222.0   Min.   :     0  
 Class :character   Class :character   1st Qu.:    0.0   1st Qu.:    23  
 Mode  :character   Mode  :character   Median :    0.0   Median :   440  
                                       Mean   :   24.9   Mean   :  5803  
                                       3rd Qu.:    5.0   3rd Qu.:  2168  
                                       Max.   :14129.0   Max.   :526158  
                                                                         
   NEW_DEATHS        TOTAL_DEATHS       POPULATION     
 Min.   :-21.0000   Min.   :   0.00   Min.   :    169  
 1st Qu.:  0.0000   1st Qu.:   0.00   1st Qu.:   6704  
 Median :  0.0000   Median :  12.00   Median :  18695  
 Mean   :  0.3938   Mean   :  99.31   Mean   : 114157  
 3rd Qu.:  0.0000   3rd Qu.:  51.00   3rd Qu.:  52600  
 Max.   :455.0000   Max.   :7636.00   Max.   :4713325  
                                      NA's   :601      
> 
> #     (b) Display the first 50 rows of the new data frame
> head(Merged_df, n = 50)
   COUNTY_NAME REPORT_DATE NEW_CASES TOTAL_CASES NEW_DEATHS TOTAL_DEATHS
1     Anderson  2021-02-06         2        5968          1           93
2     Anderson  2021-03-17         0        6089          0          112
3     Anderson  2021-03-16         0        6089          0          112
4     Anderson  2020-11-10         7        3028          1           42
5     Anderson  2020-12-23        17        4236          0           57
6     Anderson  2021-03-18       -12        6077          0          112
7     Anderson  2020-11-11         7        3035          0           42
8     Anderson  2021-03-14         4        6089          0          112
9     Anderson  2021-03-13         9        6085          1          112
10    Anderson  2020-12-24        40        4276          0           57
11    Anderson  2020-03-03         0           0          0            0
12    Anderson  2020-11-13        -4        3041          0           43
13    Anderson  2020-11-12        10        3045          1           43
14    Anderson  2020-12-22        45        4219          0           57
15    Anderson  2020-08-08        23        2402          2           12
16    Anderson  2021-03-15         0        6089          0          112
17    Anderson  2020-08-03        98        2307          1            9
18    Anderson  2021-08-04         0        6252          0          134
19    Anderson  2021-02-05        42        5966          0           92
20    Anderson  2020-03-04         0           0          0            0
21    Anderson  2020-08-07        26        2379          0           10
22    Anderson  2020-03-05         0           0          0            0
23    Anderson  2021-02-09         2        5990          2           95
24    Anderson  2021-07-27        -7        6205          0          133
25    Anderson  2021-07-28        17        6222          0          133
26    Anderson  2020-03-09         0           0          0            0
27    Anderson  2021-03-12         6        6076          1          111
28    Anderson  2020-11-07        10        3018          0           41
29    Anderson  2021-02-07         0        5968          0           93
30    Anderson  2021-03-09         2        6075          0          109
31    Anderson  2020-08-20        -7        2416          0           17
32    Anderson  2020-03-06         0           0          0            0
33    Anderson  2020-03-02         0           0          0            0
34    Anderson  2020-12-21         0        4174          0           57
35    Anderson  2020-11-30         0        3167          0           49
36    Anderson  2021-02-08        20        5988          0           93
```

```
37    Anderson  2020-11-16         0        3051         0           45
38    Anderson  2020-03-08         0           0         0            0
39    Anderson  2020-12-30         5        4510         0           63
40    Anderson  2020-12-31        28        4538         0           63
41    Anderson  2021-07-25         0        6212         0          133
42    Anderson  2020-11-20        -9        3092         0           46
43    Anderson  2021-07-26         0        6212         0          133
44    Anderson  2020-11-22         0        3127         0           46
45    Anderson  2020-08-21         1        2417         1           18
46    Anderson  2021-02-04        -1        5924         1           92
47    Anderson  2020-11-29         0        3167         0           49
48    Anderson  2020-12-26         0        4270         0           57
49    Anderson  2021-08-02         0        6244         0          134
50    Anderson  2020-08-04        22        2329         1           10
   POPULATION
1       57735
2       57735
3       57735
4       57735
5       57735
6       57735
7       57735
8       57735
9       57735
10      57735
11      57735
12      57735
13      57735
14      57735
15      57735
16      57735
17      57735
18      57735
19      57735
20      57735
21      57735
22      57735
23      57735
24      57735
25      57735
26      57735
27      57735
28      57735
29      57735
30      57735
31      57735
32      57735
33      57735
34      57735
35      57735
36      57735
37      57735
38      57735
39      57735
40      57735
41      57735
42      57735
43      57735
44      57735
45      57735
46      57735
47      57735
48      57735
49      57735
50      57735
>
> # 6.) Execute a function that will make the columns of the data frame available to R directly b
y
> #     column name to simplify coding in the modifications described below:
>
```

```
> attach(Merged_df)
> 
> #     (a) Use a function to convert REPORT_DATE to an actual R date value and assign it to a ne
w
> #         column in the data frame. Display a summary of the new date column. Note: You
> #         cannot refer to this column only by name because it did not exist when you executed
> #         the function to make the columns available.
> 
> ReportDate = as.Date(REPORT_DATE)
> Merged_df = cbind(Merged_df, ReportDate)
> summary(Merged_df$ReportDate)
        Min.      1st Qu.       Median         Mean      3rd Qu.         Max. 
"2020-01-21" "2020-06-19" "2020-11-16" "2020-11-16" "2021-04-15" "2021-09-12" 
> 
> #     (b) The COVID activity statistics are contained in four columns whose names were changed
as
> #         instructed earlier in the assignment.  Create four new columns in the data frame that
> #         represent each of the statistics as a percentage of the population of that county. Th
is is
> #         done by dividing the original column by the POPULATION column. Include PCT in the
> #         names of your new columns to differentiate them from the originals. Leave the
> #         percentage values in their raw format of a value between 0 and 1. You will notice tha
t
> #         some of the percentages are so small they are displayed in exponential notation
> 
> Merged_df = cbind(Merged_df, PCT_Total_CASES = Merged_df$TOTAL_CASES/Merged_df$POPULATION,
+                              PCT_NEW_DEATHS  = Merged_df$NEW_DEATHS/Merged_df$POPULATION,
+                              PCT_NEW_CASES   = Merged_df$NEW_CASES/Merged_df$POPULATION,
+                              PCT_TOTAL_DEATHS= Merged_df$TOTAL_DEATHS/Merged_df$POPULATION)
> 
> #     (c) Display the structure of the updated data frame and its first 20 rows.
> 
> str(Merged_df)
'data.frame':   153255 obs. of  12 variables:
 $ COUNTY_NAME     : chr  "Anderson" "Anderson" "Anderson" "Anderson" ...
 $ REPORT_DATE     : chr  "2021-02-06" "2021-03-17" "2021-03-16" "2020-11-10" ...
 $ NEW_CASES       : int  2 0 0 7 17 -12 7 4 9 40 ...
 $ TOTAL_CASES     : int  5968 6089 6089 3028 4236 6077 3035 6089 6085 4276 ...
 $ NEW_DEATHS      : int  1 0 0 1 0 0 0 0 1 0 ...
 $ TOTAL_DEATHS    : int  93 112 112 42 57 112 42 112 112 57 ...
 $ POPULATION      : int  57735 57735 57735 57735 57735 57735 57735 57735 57735 57735 ...
 $ ReportDate      : Date, format: "2021-02-06" "2021-03-17" ...
 $ PCT_Total_CASES : num  0.1034 0.1055 0.1055 0.0524 0.0734 ...
 $ PCT_NEW_DEATHS  : num  1.73e-05 0.00 0.00 1.73e-05 0.00 ...
 $ PCT_NEW_CASES   : num  3.46e-05 0.00 0.00 1.21e-04 2.94e-04 ...
 $ PCT_TOTAL_DEATHS: num  0.001611 0.00194 0.00194 0.000727 0.000987 ...
> head(Merged_df, n = 20)
   COUNTY_NAME REPORT_DATE NEW_CASES TOTAL_CASES NEW_DEATHS TOTAL_DEATHS
1     Anderson  2021-02-06         2        5968          1           93
2     Anderson  2021-03-17         0        6089          0          112
3     Anderson  2021-03-16         0        6089          0          112
4     Anderson  2020-11-10         7        3028          1           42
5     Anderson  2020-12-23        17        4236          0           57
6     Anderson  2021-03-18       -12        6077          0          112
7     Anderson  2020-11-11         7        3035          0           42
8     Anderson  2021-03-14         4        6089          0          112
9     Anderson  2021-03-13         9        6085          1          112
10    Anderson  2020-12-24        40        4276          0           57
11    Anderson  2020-03-03         0           0          0            0
12    Anderson  2020-11-13        -4        3041          0           43
13    Anderson  2020-11-12        10        3045          1           43
14    Anderson  2020-12-22        45        4219          0           57
15    Anderson  2020-08-08        23        2402          2           12
16    Anderson  2021-03-15         0        6089          0          112
17    Anderson  2020-08-03        98        2307          1            9
18    Anderson  2021-08-04         0        6252          0          134
19    Anderson  2021-02-05        42        5966          0           92
20    Anderson  2020-03-04         0           0          0            0
   POPULATION ReportDate PCT_Total_CASES PCT_NEW_DEATHS PCT_NEW_CASES
1       57735 2021-02-06      0.10336884   1.732052e-05  3.464103e-05
```

```
2       57735 2021-03-17      0.10546462   0.000000e+00  0.000000e+00
3       57735 2021-03-16      0.10546462   0.000000e+00  0.000000e+00
4       57735 2020-11-10      0.05244652   1.732052e-05  1.212436e-04
5       57735 2020-12-23      0.07336971   0.000000e+00  2.944488e-04
6       57735 2021-03-18      0.10525678   0.000000e+00 -2.078462e-04
7       57735 2020-11-11      0.05256777   0.000000e+00  1.212436e-04
8       57735 2021-03-14      0.10546462   0.000000e+00  6.928206e-05
9       57735 2021-03-13      0.10539534   1.732052e-05  1.558846e-04
10      57735 2020-12-24      0.07406253   0.000000e+00  6.928206e-04
11      57735 2020-03-03      0.00000000   0.000000e+00  0.000000e+00
12      57735 2020-11-13      0.05267169   0.000000e+00 -6.928206e-05
13      57735 2020-11-12      0.05274097   1.732052e-05  1.732052e-04
14      57735 2020-12-22      0.07307526   0.000000e+00  7.794232e-04
15      57735 2020-08-08      0.04160388   3.464103e-05  3.983719e-04
16      57735 2021-03-15      0.10546462   0.000000e+00  0.000000e+00
17      57735 2020-08-03      0.03995843   1.732052e-05  1.697411e-03
18      57735 2021-08-04      0.10828787   0.000000e+00  0.000000e+00
19      57735 2021-02-05      0.10333420   0.000000e+00  7.274617e-04
20      57735 2020-03-04      0.00000000   0.000000e+00  0.000000e+00
   PCT_TOTAL_DEATHS
1      0.0016108080
2      0.0019398978
3      0.0019398978
4      0.0007274617
5      0.0009872694
6      0.0019398978
7      0.0007274617
8      0.0019398978
9      0.0019398978
10     0.0009872694
11     0.0000000000
12     0.0007447822
13     0.0007447822
14     0.0009872694
15     0.0002078462
16     0.0019398978
17     0.0001558846
18     0.0023209492
19     0.0015934875
20     0.0000000000
> 
> #     (d) Execute a function so that the column names of the data frame are no longer available

> #         in the R search path
> 
> detach(Merged_df)
> 
> # 7.) Create and display a new data frame that is a subset of the data frame created in the pre
vious
> #     step. Use a logical test to subset the rows to only those where the REPORT_DATE is the la
st
> #     available and POPULATION is not missing. Determine the last date value based on the summa
ry
> #     of the Date column from the previous step. Hard code this value into your expression. Dis
play
> #     the structure of the new data frame.
> 
> Merged_df_Latest_NAsRemoved = subset(Merged_df, Merged_df$REPORT_DATE == "2021-09-12" & is.na(M
erged_df$POPULATION) == FALSE)
> 
> # 8.) Use the colSums function to display the statewide totals of each of the columns containin
g the
> #     original Covid count statistics. Use the apply function to make the same calculation. Inc
lude an
> #     argument on your functions so that you will get a total even if there are missing values 
for some
> #     counties.
> 
> 
> colSums(Merged_df_Latest_NAsRemoved[,c("TOTAL_CASES", "NEW_DEATHS", "NEW_CASES", "TOTAL_DEATHS"
)])
 TOTAL_CASES   NEW_DEATHS    NEW_CASES TOTAL_DEATHS 
     3815818          136         2499        60357 
> apply(Merged_df_Latest_NAsRemoved[,c("TOTAL_CASES", "NEW_DEATHS", "NEW_CASES", "TOTAL_DEATHS")]
, MARGIN = 2, FUN = sum)
 TOTAL_CASES   NEW_DEATHS    NEW_CASES TOTAL_DEATHS 
     3815818          136         2499        60357 
> 
> # 9.) Using the last data frame created, display a list of County names, TOTAL_CASES, POPULATIO
N,
> #     and percent of TOTAL_CASES, listed from the highest percentage to the lowest.
> 
> Merged_df_Latest_NAsRemoved[order(Merged_df_Latest_NAsRemoved$PCT_Total_CASES, decreasing = TRU
E),
+                               c("COUNTY_NAME","TOTAL_CASES", "POPULATION","PCT_Total_CASES")]
         COUNTY_NAME TOTAL_CASES POPULATION PCT_Total_CASES
38252         Dimmit        3619      10124      0.35746740
28761         Concho         669       2726      0.24541453
76732         Karnes        3501      15601      0.22440869
84503           Lamb        2813      12893      0.21818041
56971           Hale        7150      33406      0.21403341
31805       Crockett         727       3464      0.20987298
140005        Uvalde        5597      26741      0.20930406
95073       Maverick       12285      58722      0.20920609
140188     Val Verde        9908      49025      0.20210097
147614       Willacy        4313      21358      0.20193838
144427          Webb       54530     276652      0.19710683
22388      Childress        1433       7306      0.19614016
153195        Zavala        2316      11840      0.19560811
117109        Reeves        3121      15976      0.19535553
91142        Lubbock       60296     310569      0.19414687
135357     Tom Green       22996     119200      0.19291946
58661       Hansford        1028       5399      0.19040563
46234          Floyd        1043       5712      0.18259804
112419        Potter       21344     117415      0.18178257
24374           Coke         612       3387      0.18069088
124869        Scurry        3008      16703      0.18008741
21426       Chambers        7737      43837      0.17649474
57268           Hall         518       2964      0.17476383
32562      Culberson         379       2171      0.17457393
16429       Caldwell        7622      43664      0.17456028
49236           Frio        3537      20306      0.17418497
35256     Deaf Smith        3215      18546      0.17335274
128121         Starr       11056      64633      0.17105813
42049        El Paso      143199     839238      0.17062979
50027      Galveston       58267     342139      0.17030213
71593        Jackson        2466      14760      0.16707317
16957        Calhoun        3551      21290      0.16679192
63119       Hemphill         636       3819      0.16653574
55481         Grimes        4797      28880      0.16610111
14449          Brown        6238      37864      0.16474752
65869        Hockley        3788      23021      0.16454542
74190       Jim Hogg         855       5200      0.16442308
92232        Madison        2343      14284      0.16402968
75064      Jim Wells        6616      40482      0.16343066
59644         Hardin        9400      57602      0.16318878
39456          Duval        1812      11157      0.16240925
147245     Wilbarger        2069      12769      0.16203305
15314       Burleson        2968      18443      0.16092827
99999          Mills         783       4873      0.16068131
118927     Robertson        2735      17074      0.16018508
20491         Castro        1202       7530      0.15962815
106706        Nueces       57750     362294      0.15940093
117724       Refugio        1107       6948      0.15932642
42113          Ellis       29160     184826      0.15777001
53395       Gonzales        3281      20837      0.15746029
31994         Crosby         903       5737      0.15739934
130809        Sutton         593       3776      0.15704449
86415            Lee        2705      17239      0.15691165
134782         Titus        5113      32750      0.15612214
```

```
34717         Dawson        1981      12728      0.15564111
104601       Navarro        7796      50113      0.15556842
133949         Terry        1918      12337      0.15546729
120108       Runnels        1593      10264      0.15520265
77174        Kaufman       21117     136154      0.15509643
114784       Randall       21316     137713      0.15478568
30876          Crane         740       4797      0.15426308
102704         Moore        3228      20940      0.15415473
83635          Lamar        7665      49859      0.15373353
7803             Bee        4978      32565      0.15286350
111080        Parmer        1468       9605      0.15283706
83084       La Salle        1148       7520      0.15265957
108111        Oldham         322       2112      0.15246212
3939        Atascosa        7791      51153      0.15230778
78740           Kent         116        762      0.15223097
41239        Edwards         294       1932      0.15217391
142087        Walker       11083      72971      0.15188225
14100         Brooks        1077       7093      0.15183984
131647       Tarrant      319204    2102515      0.15182008
94828      Matagorda        5547      36643      0.15137953
33390         Dallam        1103       7287      0.15136545
108437        Orange       12608      83396      0.15118231
30424         Cottle         209       1398      0.14949928
85476         Lavaca        3012      20154      0.14944924
152639        Zapata        2119      14179      0.14944636
84958       Lampasas        3189      21428      0.14882397
78639         Kenedy          60        404      0.14851485
8547           Bexar      296585    2003554      0.14802945
68251         Howard        5415      36664      0.14769256
6517         Bastrop       13091      88723      0.14754911
146133       Wichita       19487     132230      0.14737200
46798          Foard         170       1155      0.14718615
65230           Hill        5391      36649      0.14709815
119541      Rockwall       15422     104915      0.14699519
127904     Somervell        1340       9128      0.14680105
110582        Parker       20919     142878      0.14641162
96370       McLennan       37492     256623      0.14609758
105851         Nolan        2148      14714      0.14598342
76308          Jones        2929      20083      0.14584474
69995     Hutchinson        3046      20938      0.14547712
143721    Washington        5218      35882      0.14542110
87823      Limestone        3401      23437      0.14511243
18556        Cameron       61192     423163      0.14460622
39060         Donley         473       3278      0.14429530
123657      San Saba         871       6055      0.14384806
98222         Menard         306       2138      0.14312442
150110          Wise        9991      69984      0.14276120
11750       Brazoria       53293     374264      0.14239414
12546         Brazos       32524     229211      0.14189546
27792       Comanche        1933      13635      0.14176751
53820           Gray        3097      21886      0.14150599
97182       McMullen         105        743      0.14131898
66651           Hood        8697      61643      0.14108658
109159    Palo Pinto        4111      29189      0.14084073
19155           Camp        1842      13094      0.14067512
132471        Taylor       19315     138034      0.13992929
16220         Burnet        6724      48155      0.13963244
43275          Falls        2410      17297      0.13933052
23842        Cochran         397       2853      0.13915177
151732         Young        2506      18010      0.13914492
75229        Johnson       24463     175817      0.13913899
1728        Angelina       12056      86715      0.13903016
34036         Dallas      366278    2635516      0.13897772
99683          Milam        3443      24823      0.13870201
40755          Ector       23027     166223      0.13853077
61872        Hartley         770       5576      0.13809182
131186       Swisher        1021       7397      0.13802893
93689         Martin         792       5771      0.13723791
44573        Fayette        3441      25346      0.13576107
143627          Ward        1626      11998      0.13552259
```

```
58218       Hamilton        1145       8461      0.13532679
55190          Gregg       16759     123945      0.13521320
107544     Ochiltree        1325       9836      0.13470923
37210         DeWitt        2711      20160      0.13447421
101230      Montague        2656      19818      0.13401958
139146         Upton         490       3657      0.13398961
145581       Wheeler         677       5056      0.13390032
141809      Victoria       12313      92084      0.13371487
99149        Midland       23620     176832      0.13357311
129898     Stonewall         180       1350      0.13333333
113406      Presidio         886       6704      0.13215990
145369       Wharton        5489      41556      0.13208682
6874          Baylor         463       3509      0.13194642
5342          Bailey         922       7000      0.13171429
82274        Kleberg        4035      30680      0.13151890
87541        Liberty       11556      88219      0.13099219
13379        Briscoe         202       1546      0.13065977
127220         Smith       30322     232751      0.13027656
120715          Rusk        7086      54406      0.13024299
101987    Montgomery       78976     607391      0.13002498
3573       Armstrong         245       1887      0.12983572
42849          Erath        5534      42698      0.12960794
89086       Live Oak        1580      12207      0.12943393
62911           Hays       29632     230191      0.12872788
148752        Wilson        6567      51070      0.12858821
48199      Freestone        2532      19717      0.12841710
66971        Hopkins        4742      37084      0.12787186
56489      Guadalupe       21244     166847      0.12732623
87081           Leon        2215      17404      0.12726959
111445         Pecos        2013      15823      0.12721987
115032        Reagan         489       3849      0.12704599
115775          Real         432       3452      0.12514484
29961        Coryell        9494      75951      0.12500165
19998           Cass        3738      30026      0.12449211
64833        Hidalgo      108111     868707      0.12445048
91617           Lynn         735       5951      0.12350865
298         Anderson        7121      57735      0.12333940
138249        Upshur        5117      41753      0.12255407
26375  Collingsworth         357       2920      0.12226027
151096        Yoakum        1063       8713      0.12200161
1179         Andrews        2280      18705      0.12189254
23361           Clay        1275      10471      0.12176487
81714         Kinney         446       3667      0.12162531
129537      Sterling         157       1291      0.12161115
133073       Terrell          94        776      0.12113402
97585         Medina        6237      51584      0.12090958
61105       Harrison        8030      66553      0.12065572
11402          Bowie       11176      93245      0.11985629
19747         Carson         707       5926      0.11930476
103503        Motley         143       1200      0.11916667
51514      Gillespie        3214      26988      0.11908997
10802         Bosque        2222      18685      0.11891892
43931         Fannin        4201      35514      0.11829138
124320    Schleicher         330       2793      0.11815252
149273       Winkler         946       8010      0.11810237
150394          Wood        5367      45539      0.11785503
2910          Archer        1004       8553      0.11738571
69032       Hudspeth         568       4886      0.11625051
22224       Cherokee        6113      52646      0.11611518
102883        Morris        1437      12388      0.11599935
140914     Van Zandt        6517      56590      0.11516169
25224        Coleman         933       8175      0.11412844
109599        Panola        2645      23194      0.11403811
148088    Williamson       67141     590551      0.11369213
88357       Lipscomb         367       3233      0.11351686
73468      Jefferson       28518     251565      0.11336235
25403         Collin      117227    1034730      0.11329236
67706        Houston        2582      22968      0.11241728
94350          Mason         480       4274      0.11230697
60398         Harris      526158    4713325      0.11163202
```

```
29404          Cooke        4601     41257      0.11152047
17585       Callahan        1552     13943      0.11131033
70694          Irion         170      1536      0.11067708
47303      Fort Bend       88936    811688      0.10956919
79636           Kerr        5756     52600      0.10942966
27300          Comal       16976    156209      0.10867492
12785       Brewster         999      9203      0.10855156
26798       Colorado        2324     21493      0.10812823
89906          Llano        2351     21795      0.10786878
129154      Stephens        1006      9366      0.10740978
59106       Hardeman         421      3933      0.10704297
36105         Denton       93227    887207      0.10507920
47775       Franklin        1123     10725      0.10470862
37777        Dickens         230      2211      0.10402533
137508         Tyler        2243     21672      0.10349760
125546   Shackelford         337      3265      0.10321593
103980   Nacogdoches        6728     65204      0.10318385
142772        Waller        5688     55246      0.10295768
64016      Henderson        8517     82737      0.10294064
72229         Jasper        3647     35529      0.10264854
50991          Garza         638      6229      0.10242415
125634        Shelby        2573     25274      0.10180423
95591      McCulloch         811      7984      0.10157816
112226          Polk        5196     51353      0.10118201
54638        Grayson       13773    136212      0.10111444
45676         Fisher         382      3830      0.09973890
4579          Austin        2987     30032      0.09946058
40012       Eastland        1804     18360      0.09825708
93090         Marion         967      9854      0.09813274
2185         Aransas        2301     23510      0.09787325
114084         Rains        1218     12514      0.09733099
36026          Delta         515      5331      0.09660476
52257      Glasscock         136      1409      0.09652236
100885      Mitchell         821      8545      0.09607958
118328       Roberts          82       854      0.09601874
116509     Red River        1144     12023      0.09515096
77611        Kendall        4482     47431      0.09449516
136779       Trinity        1382     14651      0.09432803
8241            Bell       34205    362924      0.09424838
69181           Hunt        9266     98594      0.09398138
121931 San Augustine         772      8237      0.09372344
122780  San Patricio        6205     66730      0.09298666
126598       Sherman         279      3022      0.09232296
80225         Kimble         395      4337      0.09107678
9193          Blanco        1085     11931      0.09093957
62422        Haskell         503      5658      0.08890067
5877         Bandera        2034     23112      0.08800623
73265     Jeff Davis         199      2274      0.08751099
49800         Gaines        1856     21492      0.08635771
135959        Travis      109645   1273954      0.08606669
122083   San Jacinto        2445     28859      0.08472227
71074           Jack         749      8935      0.08382764
121076        Sabine         833     10542      0.07901726
82381           Knox         288      3664      0.07860262
52830         Goliad         568      7658      0.07417080
105749        Newton         865     13595      0.06362633
134273  Throckmorton          90      1501      0.05996003
9811          Borden          35       654      0.05351682
90689         Loving           7       169      0.04142012
80907           King          11       272      0.04044118
>
> # 10.) Display all data for counties whose names contain the letter V, ignoring case.
>
> Merged_df_Latest_NAsRemoved[grep("v",Merged_df_Latest_NAsRemoved$COUNTY_NAME, ignore.case = TRU
E),]
       COUNTY_NAME REPORT_DATE NEW_CASES TOTAL_CASES NEW_DEATHS TOTAL_DEATHS
39456        Duval  2021-09-12         8        1812          0           50
50027    Galveston  2021-09-12       335       58267          2          577
73265   Jeff Davis  2021-09-12         0         199          0            6
85476       Lavaca  2021-09-12         0        3012          3           86
```

```
89086     Live Oak  2021-09-12         0         1580         0           29
90689       Loving  2021-09-12         0            7         0            0
95073     Maverick  2021-09-12         0        12285         0          377
104601     Navarro  2021-09-12        20         7796         0          159
116509   Red River  2021-09-12         0         1144         0           42
117109      Reeves  2021-09-12         2         3121         0           47
127904   Somervell  2021-09-12         0         1340         0           17
135959      Travis  2021-09-12         0       109645         5         1221
140005      Uvalde  2021-09-12         0         5597         0           86
140188   Val Verde  2021-09-12         0         9908         0          233
140914   Van Zandt  2021-09-12         0         6517         1          161
141809    Victoria  2021-09-12         0        12313         0          277
153195      Zavala  2021-09-12         0         2316         0           50
       POPULATION ReportDate PCT_Total_CASES PCT_NEW_DEATHS PCT_NEW_CASES
39456       11157 2021-09-12      0.16240925   0.000000e+00  0.0007170386
50027      342139 2021-09-12      0.17030213   5.845577e-06  0.0009791342
73265        2274 2021-09-12      0.08751099   0.000000e+00  0.0000000000
85476       20154 2021-09-12      0.14944924   1.488538e-04  0.0000000000
89086       12207 2021-09-12      0.12943393   0.000000e+00  0.0000000000
90689         169 2021-09-12      0.04142012   0.000000e+00  0.0000000000
95073       58722 2021-09-12      0.20920609   0.000000e+00  0.0000000000
104601      50113 2021-09-12      0.15556842   0.000000e+00  0.0003990980
116509      12023 2021-09-12      0.09515096   0.000000e+00  0.0000000000
117109      15976 2021-09-12      0.19535553   0.000000e+00  0.0001251878
127904       9128 2021-09-12      0.14680105   0.000000e+00  0.0000000000
135959    1273954 2021-09-12      0.08606669   3.924788e-06  0.0000000000
140005      26741 2021-09-12      0.20930406   0.000000e+00  0.0000000000
140188      49025 2021-09-12      0.20210097   0.000000e+00  0.0000000000
140914      56590 2021-09-12      0.11516169   1.767097e-05  0.0000000000
141809      92084 2021-09-12      0.13371487   0.000000e+00  0.0000000000
153195      11840 2021-09-12      0.19560811   0.000000e+00  0.0000000000
       PCT_TOTAL_DEATHS
39456      0.0044814914
50027      0.0016864491
73265      0.0026385224
85476      0.0042671430
89086      0.0023756861
90689      0.0000000000
95073      0.0064200811
104601     0.0031728294
116509     0.0034933045
117109     0.0029419129
127904     0.0018624014
135959     0.0009584334
140005     0.0032160353
140188     0.0047526772
140914     0.0028450256
141809     0.0030081230
153195     0.0042229730
>
> # 11.) Display the contents of the workspace
>
> ls()
[1] "COVID_Activity"              "Covid_Texas"
[3] "Merged_df"                   "Merged_df_Latest_NAsRemoved"
[5] "PopTable"                    "ReportDate"
>
> # 12.) Remove everything from the workspace except the data frame created beginning in step 5
> #      above and the data frame created in step 7. Display the contents of the workspace again.
>
> rm(list = setdiff(ls(), c("Merged_df", "Merged_df_Latest_NAsRemoved")))
>
> # 13.)  Save the workspace in case we want to use it in the next assignment. Name it HW05.RData
.
> #       You may save it initially using the R GUI but your script must contain code to save the
 workspace
> #       in case you submit the script again.
>
> #### ASK MARK
>
```

```
> # 14.) After you have debugged your program and successfully executed it in a new R session, us
e the
> #      information in your console to answer the questions below in comment lines at the bottom
 of
> #      your script:
>
> #     (a) How many observations were loaded from the CSV file?
>
> #         2132949
>
> #     (b) How many observations and variables are in the data frame loaded from the txt file?
>
> #         3483 observations of 10 variables
>
> #     (c) What is one possible explanation for the minimum value of NEW_CASES shown in the
> #         summary from step 5a and what is your reaction to this value as an analyst?
>
> #         The minimum value could represent an adjustment to the previous entry's number of new
 cases.
> #         In other words, the new cases, minus adjustments made to the previous entry is -1222.
> #         As an analyst my first reaction would be to investigate this further.
>
> #     (d) Explain the difference in the summaries of the two date columns. What are the
> #         minimum and maximum dates in the data frame?
>
> #         The original date column is a character vector, so the entries are not interpreted by

> #         r as dates, thus there are no numerical summaries available for the original date col
umn.
> #         The minimum and maximum dates in the data frame are 01/21/2020 & 09/12/2021 respectiv
ely.
>
> #     (e) What is the total number of COVID cases and deaths in the state of Texas on the last
> #         date reported?
>
> #         Total Cases = 3815818, Total Deaths = 60357
>
> #     (f) What is the name and population of the county with the lowest percentage of cases as
> #         of the last date reported?
>
> #         County Name: King, Population: 272
>
>
>
>
>
```