HW2 - Exercise 3 (Covid-19)

MATH 345-545

Spring 2022

Question 1

Import the data:

```
covid <- read.table("world-covid-2022-02-04.txt", header = TRUE)
```

Note: the argument `header = TRUE` tells R that the first line of the data file contains the variables' names -not data.

Attach the data frame for easier manipulation:

```
attach(covid)
```

Question 2

Number of rows (cases) and columns (variables) in dataset:

```
dim(covid)
```

```
## [1] 225  14
```

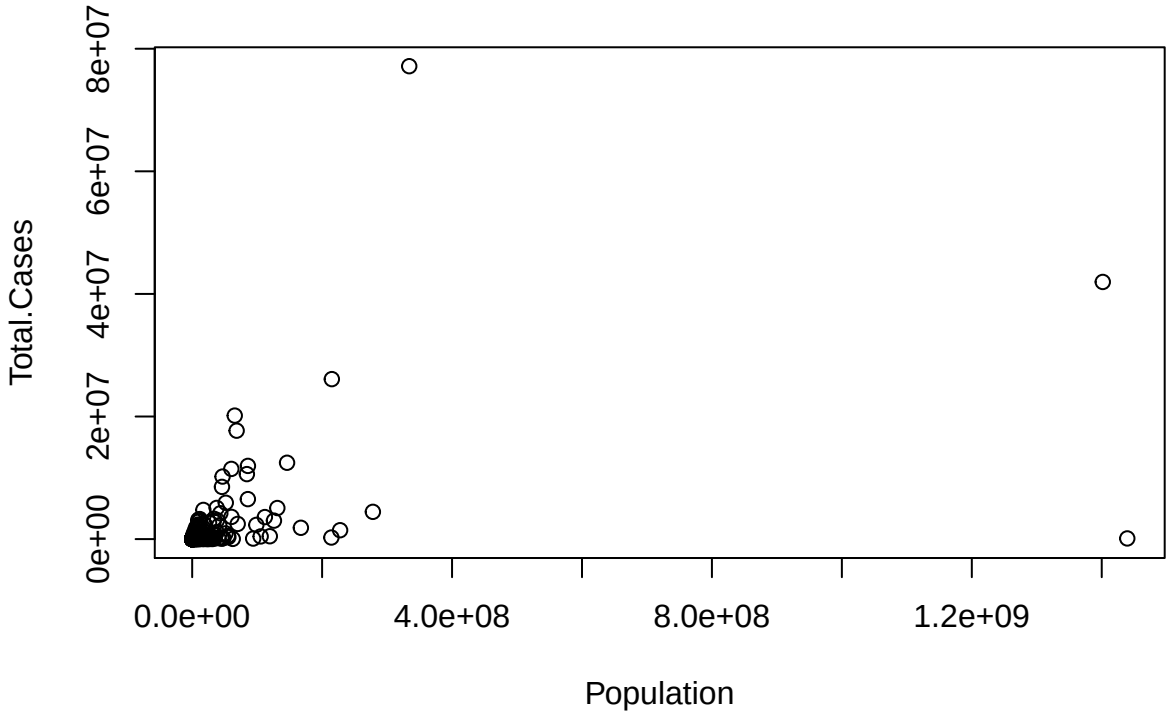
```
str(covid) # to get information on variable names and types
```

```
## 'data.frame':    225 obs. of  14 variables:
##  $ Country          : chr  "USA" "India" "Brazil" "France" ...
##  $ Total.Cases      : int  77150412 41952712 26099735 20147341 17689885 12452765 11940695 11449601 10
##  $ New.Cases        : num  NA NA NA NA 84.1 ...
##  $ Total.Deaths     : num  921 500 630 132 158 ...
##  $ New.Deaths       : int  NA NA NA NA 254 682 NA 433 26 NA ...
##  $ Total.Recovered  : int  47313736 40017088 22602506 13437823 14977096 10449181 11173406 9083090 795
##  $ New.Recovered    : int  NA NA NA NA 145655 58449 NA 210353 84000 NA ...
##  $ Active.Cases     : int  28915847 1435537 2867228 6577666 2554805 1669545 679225 2218344 2526949 30
##  $ Serious..Critical: num  21.64 8.94 8.32 3.64 478 ...
##  $ Tot.Cases.1M.pop : num  230.9 29.9 121.4 307.6 258.4 ...
##  $ Deaths.1M.pop    : num  2.76 357 2.93 2.01 2.31 ...
##  $ Total.Tests      : num  9.10e+08 7.36e+08 6.38e+07 2.32e+08 4.58e+08 ...
##  $ Tests.1M.pop     : int  2724296 524986 296687 3536455 6697523 1695496 1546631 2895261 1064272 1415
##  $ Population       : int  334088290 1401569003 214961303 65503153 68452931 146033957 85781430 60320
```

Question 3

Plot population size (x axis) versus total number of cases (y axis):

```r
plot(Population, Total.Cases)
```

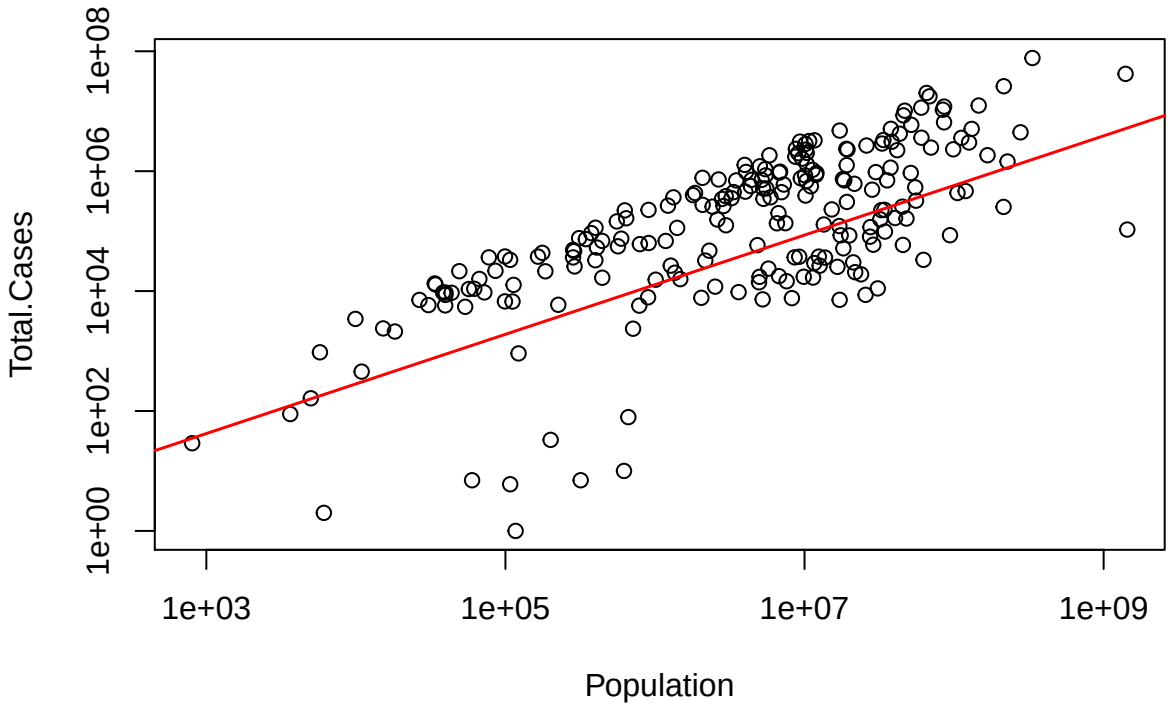


Because of very large values, this scatter plot is a little difficult to interpret. Plot it again on a logarithmic scale:

```r
plot(Population, Total.Cases,  log = "xy")
## Add a regression line
reg <- lm(log(Total.Cases) ~ log(Population))
summary(reg)
```

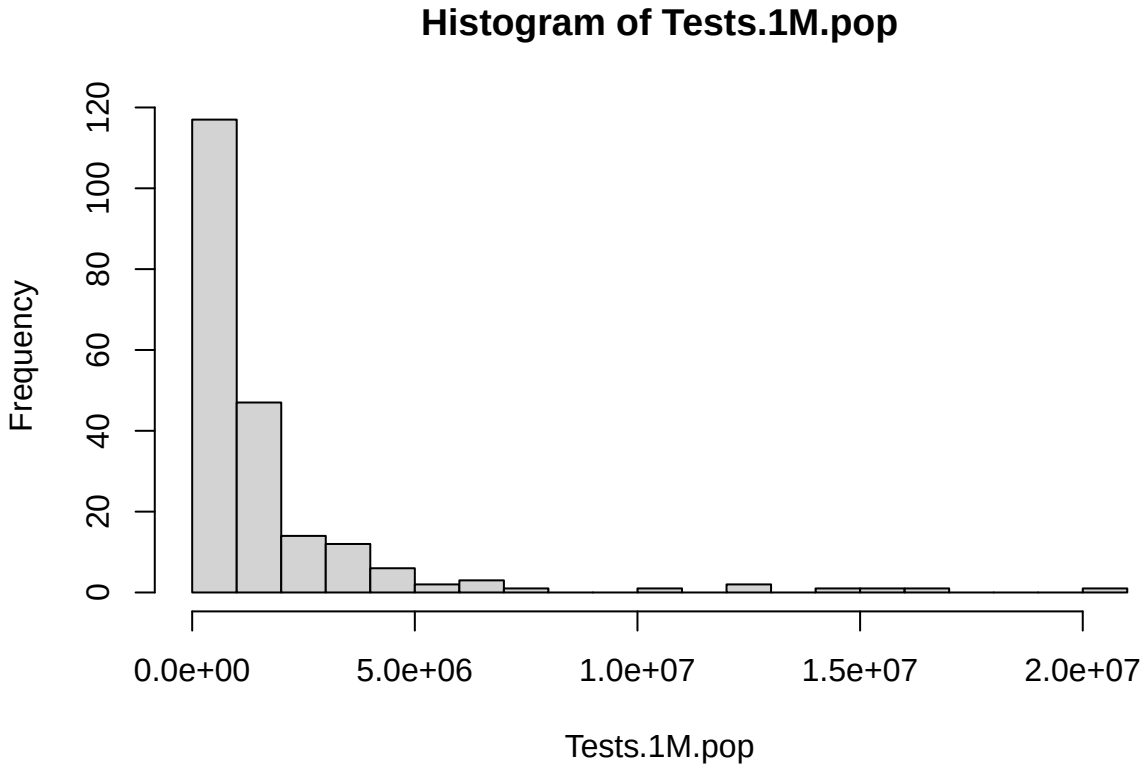
```
##
## Call:
## lm(formula = log(Total.Cases) ~ log(Population))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.7976 -1.2914  0.7599  1.5552  2.7845
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -0.8587     0.8332  -1.031    0.304
## log(Population)   0.8275     0.0545  15.182   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.162 on 221 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.5105, Adjusted R-squared:  0.5083
## F-statistic: 230.5 on 1 and 221 DF,  p-value: < 2.2e-16
```

```r
abline(reg, col = "red", lwd = 1.5)
```



Question 4

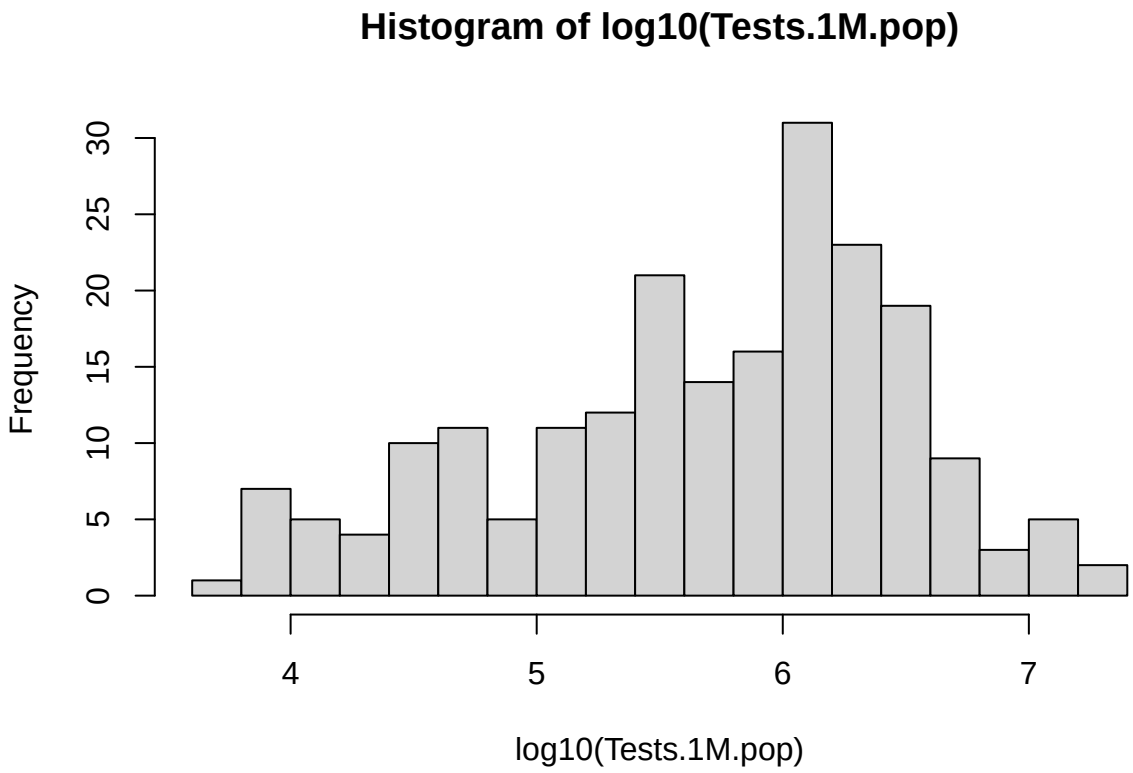
```
hist(Tests.1M.pop, breaks = 20)
```



This histogram is skewed to the right (it has a long right tail). While most countries have under 5 millions

tests per million people, some have over 20 million tests per million people.

```
hist(log10(Tests.1M.pop), breaks = 20)
```



This histogram shows the same data on a logarithmic scale base 10). As a result, it shows much more detail on the lower values (say, under 1M test per M people) which make up most of the data. Here, the histogram appears skewed to the left (long left tail, shorter right tail).

List the countries by increasing order of testing (not asked in the homework):

```
o <- order(Tests.1M.pop)
data.frame(Country[o], Tests.1M.pop[o])
```

```
##                Country.o. Tests.1M.pop.o.
## 1                 Algeria            5118
## 2                 Eritrea            6537
## 3         Solomon Islands            7170
## 4                   Macao            7864
## 5                   Syria            8048
## 6                   Niger            8507
## 7                   Yemen            8593
## 8                     DRC            9019
## 9                    Chad           11141
## 10           Burkina Faso           11201
## 11                  Haiti           11393
## 12                  Sudan           12379
## 13             Madagascar           12859
## 14                    CAR           16379
## 15                Nigeria           19213
## 16            Afghanistan           21382
## 17                Somalia           24129
## 18                Liberia           26651
```

```
## 19                 Malawi          26691
## 20       Papua New Guinea          27034
## 21                Burundi          27764
## 22                   Mali          28340
## 23            South Sudan          28446
## 24           Sierra Leone          31569
## 25                  Egypt          35041
## 26               Ethiopia          36556
## 27             Mozambique          37734
## 28                 Angola          40235
## 29             Uzbekistan          40243
## 30                 Guinea          42594
## 31                  Benin          47857
## 32                 Uganda          49301
## 33            Ivory Coast          49620
## 34                 Gambia          53718
## 35                Senegal          56434
## 36                  Kenya          58355
## 37          Guinea-Bissau          58492
## 38                  Congo          60661
## 39               Cameroon          63482
## 40                  Ghana          69822
## 41                Vanuatu          72204
## 42             Bangladesh          75629
## 43                   Togo          77624
## 44                 Mexico         108984
## 45               Pakistan         110927
## 46                  China         111163
## 47               Honduras         117175
## 48                Myanmar         117654
## 49              Venezuela         118660
## 50  Sao Tome and Principe         128692
## 51                   Laos         130960
## 52               Zimbabwe         131458
## 53                Ecuador         136710
## 54             Mauritania         143469
## 55               Cambodia         164071
## 56                 Zambia         166554
## 57            Timor-Leste         166656
## 58                Lesotho         175534
## 59                  Nepal         175641
## 60      Equatorial Guinea         185240
## 61    Antigua and Barbuda         190500
## 62              Guatemala         197023
## 63                Bolivia         222470
## 64            Philippines         244381
## 65               Thailand         246449
## 66            El Salvador         248980
## 67              Indonesia         264852
## 68                Jamaica         269181
## 69     Dominican Republic         269806
## 70                Djibouti         272370
## 71                   Japan         273567
## 72              Mauritius         281286
```

```
## 73              Kyrgyzstan         284860
## 74               Sri Lanka         285573
## 75                  Brazil         296687
## 76                 Morocco         298810
## 77                S. Korea         307835
## 78                   Libya         322184
## 79                Paraguay         328991
## 80                  Rwanda         336671
## 81                 Tunisia         340600
## 82           New Caledonia         341420
## 83                 Namibia         354571
## 84                Suriname         366763
## 85            South Africa         369841
## 86                Eswatini         389941
## 87     Trinidad and Tobago         395907
## 88                    Iraq         420942
## 89                 Ukraine         424888
## 90                  Taiwan         441895
## 91  Bosnia and Herzegovina         505870
## 92                    Iran         521513
## 93                   India         524986
## 94                 Bahamas         528017
## 95                    Fiji         543665
## 96                Palestine         559385
## 97                 Albania         569936
## 98              Kazakhstan         604956
## 99              Azerbaijan         605564
## 100               Mayotte         624568
## 101              Colombia         624806
## 102             Costa Rica         643265
## 103                 Guyana         645963
## 104                  Gabon         669338
## 105                Moldova         690386
## 106            Saint Lucia         691281
## 107                Lebanon         707642
## 108             Cabo Verde         708970
## 109              Argentina         716011
## 110                   Peru         761160
## 111                Vietnam         782670
## 112                 Poland         825760
## 113               Botswana         835132
## 114        North Macedonia         865723
## 115 St. Vincent Grenadines         872054
## 116                Armenia         929658
## 117                 Serbia         953068
## 118                Romania        1026867
## 119               Slovakia        1047026
## 120                Croatia        1058281
## 121                Hungary        1060916
## 122                Germany        1064272
## 123           Saudi Arabia        1077570
## 124                   Cuba        1126937
## 125                Grenada        1128520
## 126  Caribbean Netherlands        1132131
```

```
## 127                 Belize         1146226
## 128                  Qatar         1181507
## 129                 Panama         1190909
## 130               Mongolia         1198647
## 131  Saint Kitts and Nevis         1210867
## 132               Slovenia         1212250
## 133            New Zealand         1222964
## 134            Netherlands         1227520
## 135               Bulgaria         1251978
## 136                Belarus         1256917
## 137               Malaysia         1355655
## 138                 Monaco         1384941
## 139                  Palau         1387088
## 140                  Spain         1415323
## 141            Sint Maarten         1421672
## 142                 Jordan         1443794
## 143                 Canada         1481925
## 144                Uruguay         1518256
## 145                 Turkey         1546631
## 146             Montserrat         1560136
## 147                  Chile         1570045
## 148                Réunion         1572987
## 149                 Brunei         1588150
## 150                 Kuwait         1605687
## 151             Martinique         1608153
## 152                  Aruba         1654944
## 153                 Sweden         1676740
## 154                 Russia         1695496
## 155                Finland         1715985
## 156             Isle of Man         1758093
## 157              Montenegro         1787323
## 158                  Norway         1794999
## 159           French Guiana         1836863
## 160       Wallis and Futuna         1876991
## 161                Barbados         1892389
## 162                  Bhutan         1914241
## 163            Saint Martin         1973259
## 164             Switzerland         1982050
## 165              Guadeloupe         2075094
## 166                Dominica         2078027
## 167           Liechtenstein         2099447
## 168                 Estonia         2168517
## 169                 Ireland         2191939
## 170               Australia         2367673
## 171        Falkland Islands         2370777
## 172                 Belgium         2616465
## 173                     USA         2724296
## 174                  Malta         2732117
## 175                Curaçao         2757683
## 176              Lithuania         2763855
## 177                  Italy         2895261
## 178              Greenland         2897149
## 179  Saint Pierre Miquelon         3004523
## 180 British Virgin Islands         3215836
```

```
## 181                Andorra         3225298
## 182          Cayman Islands         3225784
## 183               Portugal         3291729
## 184                 Latvia         3348829
## 185               Anguilla         3376840
## 186                 France         3536455
## 187             San Marino         3675831
## 188                Georgia         3715505
## 189               Maldives         3745915
## 190              Singapore         3845655
## 191              Hong Kong         4252404
## 192                 Israel         4436346
## 193                Iceland         4705311
## 194                   Oman         4705606
## 195                Czechia         4726332
## 196                Bahrain         4942417
## 197                 Greece         5647290
## 198              St. Barth         5691788
## 199             Luxembourg         6225329
## 200                     UK         6697523
## 201        Channel Islands         6742260
## 202                 Cyprus         7759687
## 203       Turks and Caicos        10614478
## 204                Bermuda        12346040
## 205                    UAE        12566172
## 206              Gibraltar        14651314
## 207          Faeroe Islands        15317954
## 208                Austria        16039782
## 209                Denmark        20762507
## 210       French Polynesia              NA
## 211             Seychelles              NA
## 212               Tanzania              NA
## 213              Nicaragua              NA
## 214             Tajikistan              NA
## 215                Comoros              NA
## 216               Kiribati              NA
## 217       Diamond Princess              NA
## 218                  Samoa              NA
## 219           Vatican City              NA
## 220         Western Sahara              NA
## 221             MS Zaandam              NA
## 222       Marshall Islands              NA
## 223                  Tonga              NA
## 224           Saint Helena              NA
## 225             Micronesia              NA
```

Question 5

Calculate the ratio of active cases over population size:

```
r <- Active.Cases / Population
```

What is the largest value of this ratio?

```
max(r) # does not work because of missing values
```

```
## [1] NA
```

```
max(r, na.rm = TRUE) # this work but we don't which country this corresponds to
```

```
## [1] 0.2796086
```

Find the corresponding country:

```
idx <- which.max(r)
Country[idx]
```

```
## [1] "Faeroe Islands"
```

In light of the fact that Denmark has the highest testing rate (question 4), it makes sense that the Faeroe Islands, an autonomous territory of Denmark, has the highest rate of active infections. This is because more testing implies more detection of the virus, all other things being equal.

Finally, don't forget to detach the data frame when you are done with your analysis:

```
detach(covid)
```