CMSC320 Final Project (Covid-19)

Luobin Chen, Yufei Zhang

Loading and Plotting Data

We find the data from Kaggle (https://www.kaggle.com/imdevskp/corona-virus-report?select=covid_19_clean_complete.csv (https://www.kaggle.com/imdevskp/corona-virus-report?select=covid_19_clean_complete.csv)). Now, We download the data and perform some simple plots to analyze the data. The first step is extracting the csv file then passing that to our data frame. Since we want to use the data for nucleic acids test(NAT) for further analysis, we download the data test.csv from Kaggle. (https://www.kaggle.com/imdevskp/corona-virus-report?select=tests.csv (https://www.kaggle.com/imdevskp/corona-virus-report?select=tests.csv)) We read the file using read_csv after importing essential library “tidyverse”. We renamed “Country/Region” to “Region” and “Province/States” to “States” to be easier to reference.

```
library(tidyverse)
df <- read_csv("covid_19_clean_complete.csv")
names(df)[1] <- "State"
names(df)[2] <- "Region"
head(df)
```

```
## # A tibble: 6 x 8
##   State Region                Lat   Long Date    Confirmed Deaths Recovered
##   <chr> <chr>               <dbl>  <dbl> <chr>       <dbl>  <dbl>     <dbl>
## 1 <NA>  Afghanistan          33     65    1/22/20         0      0         0
## 2 <NA>  Albania              41.2   20.2  1/22/20         0      0         0
## 3 <NA>  Algeria              28.0    1.66 1/22/20         0      0         0
## 4 <NA>  Andorra              42.5    1.52 1/22/20         0      0         0
## 5 <NA>  Angola              -11.2   17.9  1/22/20         0      0         0
## 6 <NA>  Antigua and Barbuda  17.1 -61.8  1/22/20         0      0         0
```

```
test <- read_csv("tests.csv")
head(test)
```

```
## # A tibble: 6 x 5
##   Country `Cases per 1M pop` `Deaths per 1M pop` `Total Tests` `Tests per 1M po…
##   <chr>                <dbl>               <dbl>         <dbl>             <dbl>
## 1 USA                   4256                 252       9935720             30017
## 2 Spain                 5765                 576       2467761             52781
## 3 Russia                1591                  14       5805404             39781
## 4 UK                    3336                 482       2007146             29566
## 5 Italy                 3659                 511       2673655             44221
## 6 France                2730                 414       1384633             21213
```

Now we have two essential data set for our analysis. First data set “df” has the information of confirmed cases, deaths, and recovered cases for every countries over the world from 1/22/20 to 5/09/2020. Additionally, it includes lat and long attributes for our interaction visualization. Second data set “test” has the information of ratio of confirmed cases, deaths, and NAT per 1 million population for each country.

Now let’s make some plot to have a general view of the data set. Initially, we want to plot all the confirmed cases for each country until 5/9/2020. However, because of a large amount of data, we can’t plot all the region attribute as color. There are too many data points which will make the x and y axis hard to see and analyze. Therefore, we first use arrange to sort data points by their confirmed cases. Then, we use slice command to slice out the top 15 countries that have the highest

number of confirmed cases. We use group_by and sum command to get confirmed case for different regions/countries. The ungroup command is necessary for arrange command. We create a new data frame called “confirmed” which contains the information for the top 15 countries on the purpose of making the plot.

The ggplot command helps us to make the plot. The data is from the first data set “df”. Here we map the Region attribute to the x position in the plot and the Confirmed_cases_for_country attribute to the y position in the plot. The ggplot contains the aes call. Here we choose points as the geometric representations of our chosen graphical characteristics using the geom_point function.

```
# Confirmed case for top 20
confirmed <-  df %>%
  filter(Date=="5/9/20") %>%
  group_by(Region) %>%
  mutate(Confirmed_cases_for_country=sum(Confirmed)) %>%
  ungroup(Region) %>%
  arrange(desc(Confirmed_cases_for_country))


confirmed <- unique(data.frame(Region=confirmed$Region,Confirmed_cases_for_country=confirmed$Co
nfirmed_cases_for_country))

confirmed
```

```
##                            Region Confirmed_cases_for_country
## 1                              US                      1309550
## 2                           Spain                       223578
## 3                           Italy                       218268
## 4                  United Kingdom                       216525
## 15                         Russia                       198676
## 16                         France                       176782
## 27                        Germany                       171324
## 28                         Brazil                       156061
## 29                         Turkey                       137115
## 30                           Iran                       106220
## 31                          China                        83990
## 64                         Canada                        68918
## 78                           Peru                        65015
## 79                          India                        62808
## 80                        Belgium                        52596
## 81                    Netherlands                        42575
## 85                   Saudi Arabia                        37136
## 86                         Mexico                        33460
## 87                    Switzerland                        30251
## 88                        Ecuador                        29071
## 89                       Pakistan                        28736
## 90                       Portugal                        27406
## 91                          Chile                        27219
## 92                         Sweden                        25921
## 93                        Ireland                        22760
## 94                      Singapore                        22460
## 95                        Belarus                        22052
## 96                          Qatar                        21331
## 97            United Arab Emirates                        17417
## 98                         Israel                        16454
## 99                         Austria                        15833
## 100                          Japan                        15663
## 101                         Poland                        15651
## 102                        Romania                        15131
## 103                        Ukraine                        14710
## 104                     Bangladesh                        13770
## 105                      Indonesia                        13645
## 106                    South Korea                        10874
## 107                    Philippines                        10610
## 108                        Denmark                        10517
## 111                       Colombia                        10495
## 112                         Serbia                        10032
## 113             Dominican Republic                         9882
## 114                   South Africa                         9420
## 115                          Egypt                         8964
## 116                         Panama                         8282
## 117                         Norway                         8099
## 118                        Czechia                         8095
## 119                         Kuwait                         7623
## 120                      Australia                         6939
## 128                       Malaysia                         6589
## 129                        Morocco                         5910
## 130                        Finland                         5880
## 131                      Argentina                         5776
## 132                        Algeria                         5558
## 133                     Kazakhstan                         4975
## 134                        Moldova                         4867
## 135                        Bahrain                         4774
```

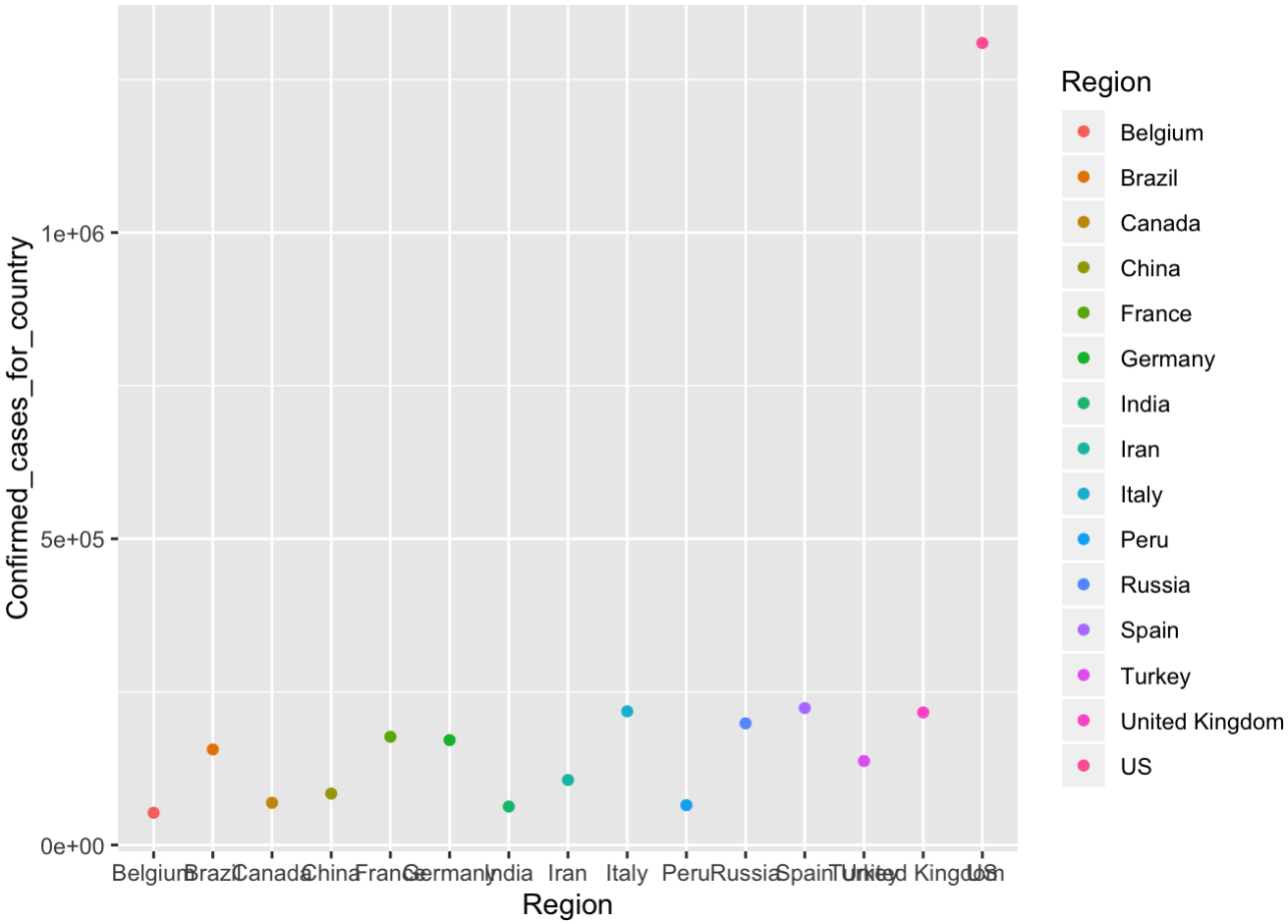
/

```
## 136                    Ghana                    4263
## 137                  Nigeria                    4151
## 138              Afghanistan                    4033
## 139               Luxembourg                    3877
## 140                     Oman                    3224
## 141                  Hungary                    3213
## 142                  Armenia                    3175
## 143                 Thailand                    3004
## 144                   Greece                    2710
## 145                     Iraq                    2679
## 146                  Bolivia                    2437
## 147               Azerbaijan                    2422
## 148               Uzbekistan                    2349
## 149                 Cameroon                    2274
## 150                  Croatia                    2176
## 151 Bosnia and Herzegovina                    2090
## 152                   Guinea                    2042
## 153                 Bulgaria                    1921
## 154                 Honduras                    1830
## 155                  Iceland                    1801
## 156                     Cuba                    1754
## 157                  Estonia                    1733
## 158            Cote d'Ivoire                    1667
## 159                  Senegal                    1634
## 160           North Macedonia                    1622
## 161              New Zealand                    1494
## 162                 Slovakia                    1455
## 163                 Slovenia                    1454
## 164                Lithuania                    1444
## 165                 Djibouti                    1189
## 166                    Sudan                    1164
## 167                  Tunisia                    1032
## 168                  Somalia                     997
## 169                Guatemala                     967
## 170          Congo (Kinshasa)                     937
## 171                Kyrgyzstan                     931
## 172                    Latvia                     930
## 173                    Cyprus                     892
## 174                    Kosovo                     862
## 175                   Albania                     856
## 176                 Sri Lanka                     847
## 177                     Niger                     815
## 178                   Lebanon                     809
## 179                  Maldives                     790
## 180               El Salvador                     784
## 181                Costa Rica                     780
## 182                   Andorra                     754
## 183              Burkina Faso                     748
## 184          Diamond Princess                     712
## 185                   Uruguay                     702
## 186                      Mali                     692
## 187                  Paraguay                     689
## 188                     Gabon                     661
## 189                     Kenya                     649
## 190             Guinea-Bissau                     641
## 191                San Marino                     637
## 192                   Georgia                     626
## 193                 Tajikistan                     612
## 194                    Jordan                     522
## 195                  Tanzania                     509
```

```
## 196                          Jamaica                    490
## 197                            Malta                    490
## 198                          Taiwan*                    440
## 199                Equatorial Guinea                    439
## 200                        Venezuela                    402
## 201               West Bank and Gaza                    375
## 202                        Mauritius                    332
## 203                       Montenegro                    324
## 204                             Chad                    322
## 205                     Sierra Leone                    291
## 206                          Vietnam                    288
## 207                            Benin                    284
## 208                           Rwanda                    280
## 209             Congo (Brazzaville)                    274
## 210                           Zambia                    252
## 211                       Cabo Verde                    236
## 212                         Ethiopia                    210
## 213           Sao Tome and Principe                    208
## 214                          Liberia                    199
## 215                       Madagascar                    193
## 216                            Burma                    178
## 217                         Eswatini                    163
## 218                             Togo                    153
## 219                            Haiti                    151
## 220         Central African Republic                    143
## 221                           Brunei                    141
## 222                         Cambodia                    122
## 223                      South Sudan                    120
## 224              Trinidad and Tobago                    116
## 225                           Uganda                    116
## 226                            Nepal                    110
## 227                           Monaco                     96
## 228                           Guyana                     94
## 229                          Bahamas                     92
## 230                       Mozambique                     87
## 231                         Barbados                     84
## 232                     Liechtenstein                     82
## 233                            Libya                     64
## 234                           Malawi                     56
## 235                            Syria                     47
## 236                           Angola                     43
## 237                         Mongolia                     42
## 238                          Eritrea                     39
## 239                         Zimbabwe                     35
## 240                            Yemen                     34
## 241              Antigua and Barbuda                     25
## 242                      Timor-Leste                     24
## 243                         Botswana                     23
## 244                          Grenada                     21
## 245                           Gambia                     20
## 246                             Laos                     19
## 247                             Fiji                     18
## 248                      Saint Lucia                     18
## 249                           Belize                     18
## 250 Saint Vincent and the Grenadines                     17
## 251                          Namibia                     16
## 252                        Nicaragua                     16
## 253                         Dominica                     16
## 254             Saint Kitts and Nevis                     15
## 255                          Burundi                     15
```

```
## 256                    Holy See                   12
## 257                  Seychelles                   11
## 258                     Comoros                   11
## 259                    Suriname                   10
## 260                  MS Zaandam                    9
## 261                  Mauritania                    8
## 262           Papua New Guinea                     8
## 263                      Bhutan                    7
## 264             Western Sahara                     6
```

```
confirmed %>%
  slice(1:15) %>%
  ggplot(mapping=aes(y=Confirmed_cases_for_country,x=Region,color=Region)) +
  geom_point()
```



From the plot, we can see the top 15 countries that have the highest number of confirmed cases clearly. They are US, Spain, Italy, United Kingdom, Russia, France, Germany, Brazil, Turkey, Iran, China, Canada, Peru, India, Belgium. We can see from the plot that the rate of confirmed cases in the United States is much higher than in other countries.

Now, we want to take advantage of the interaction visualization to make a more intuitive graph. We use the data form our data set “df”. We select all the data points from the last date which is 5/9/20. We use Leaflet library to generate the graph for each states or provinces.

We learned from this website https://rstudio.github.io/leaflet/markers.html (https://rstudio.github.io/leaflet/markers.html) to generate useful icons. Base on our condition, states with less than 10000 confirmed cases have green popup icons. States with less than 50000 confirmed cases have orange popup icons. States will more than 50000 confirmed cases have red popup icons. These icons give us a better understanding of the distribution of corona virus. Additionally, it’s easier to see which area has the highest number of confirmed cases.

```
df <- df %>%
  filter(Date=="5/9/20")

head(df)
```

```
## # A tibble: 6 x 8
##   State Region                Lat   Long Date   Confirmed Deaths Recovered
##   <chr> <chr>               <dbl>  <dbl> <chr>      <dbl>  <dbl>     <dbl>
## 1 <NA>  Afghanistan          33     65    5/9/20      4033    115       502
## 2 <NA>  Albania              41.2   20.2  5/9/20       856     31       627
## 3 <NA>  Algeria              28.0    1.66 5/9/20      5558    494      2546
## 4 <NA>  Andorra              42.5    1.52 5/9/20       754     48       545
## 5 <NA>  Angola              -11.2   17.9  5/9/20        43      2        13
## 6 <NA>  Antigua and Barbuda  17.1  -61.8  5/9/20        25      3        19
```

```
library(leaflet)
getColor <- function(df) {
  sapply(df$Confirmed, function(Confirmed) {
  if(Confirmed <= 10000) {
    "green"
  } else if(Confirmed <= 50000) {
    "orange"
  } else {
    "red"
  } })
}


icons <- awesomeIcons(
  icon = 'ios-close',
  iconColor = 'black',
  library = 'ion',
  markerColor = getColor(df)
)

map <- leaflet(df) %>%
  addTiles() %>%
  addAwesomeMarkers(~Long,~Lat,popup=~as.character(Confirmed),icon=icons,label=~as.character(Co
nfirmed))

map
```

Now let's look at our second data set. Initially, we want to ask if tests is enough for each country. Does confirmed cases affected by the number of tests? Is there any people with symptoms in the country still haven't recieve test?

First, we chose to change the name for each attributes for easier reference. We arrange the data frame base on their total tests. Then we compute the ratio of Cases_per_1M_pop/Tests_per_1M_pop.

If the ratio is very small, it's good. This means we have enough test to examine more people in the country. However, if the ration is very large, it means that it's possible there are still a lot of people who haven't been tested. We select top 50 countries which has large ratio. Base on ratio, we can conclude that countries which has ratio larger than 0.1 don't have enough NAT.

```
names(test)[2] <- "Cases_per_1M_pop"
names(test)[3] <- "Deaths_per_1M_pop"
names(test)[4] <- "Total_Tests"
names(test)[5] <- "Tests_per_1M_pop"

test %>%
  mutate(ratio=Cases_per_1M_pop/Tests_per_1M_pop)  %>%
  arrange(desc(ratio)) %>%
  slice(1:50)
```

```
## # A tibble: 50 x 6
##    Country  Cases_per_1M_pop Deaths_per_1M_p… Total_Tests Tests_per_1M_pop ratio
##    <chr>               <dbl>            <dbl>       <dbl>             <dbl> <dbl>
##  1 Sao Tom…              949               23         175               799 1.19
##  2 Algeria               138               12        6500               148 0.932
##  3 Guinea-…              417                2        1500               762 0.547
##  4 Equator…              313                3         854               609 0.514
##  5 Yemen                   2              0.3         120                 4 0.5
##  6 Andorra              9810              621        1673             21653 0.453
##  7 Bolivia               254               11        7651               655 0.388
##  8 Honduras              210               12        5653               571 0.368
##  9 Ecuador              1724              132       85223              4830 0.357
## 10 Cabo Ve…              480                4         791              1423 0.337
## # … with 40 more rows
```