Group Project: Covid Origins

Name: Sal Figueroa
Partner: Kylie Stearns

2025-03-30

Contents

Background

The World Health Organization has recently employed a new data science initiative, *CSIT-165*, that uses data science to characterize pandemic diseases. *CSIT-165* disseminates data driven analyses to global decision makers.

CSIT-165 is a conglomerate comprised of two fabricated entities: *Global Health Union (GHU)* and *Private Diagnostic Laboratories (PDL)*. Your and your partner's role is to play a data scientist from one of these two entities.

Data

2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by John Hopkins CSSE

Data for 2019 Novel Corona virus is operated by the John Hopkins University Center for Systems Science and Engineering (JHU CSSE). Data includes daily time series CSV summary tables, including confirmations, recoveries, and deaths. Country/region are countries/regions hat conform to World Health Organization (WHO). Lat and Long refer to coordinates references for the user. Date fields are stored in MM/DD/YYYY format.

Project Objectives

Objective 1

Predict where the origin started based on the area with the greatest number of confirmations and deaths on the first recorded day in the data set. Show this is the origin using an if statement.

```
#Most deaths and confirmed cases on x1.22.20 (date)
#find max value in column 5 (10.22.20) and check to see how many common max values there is.
max_value <- 0 #holds the updated max_value
RowMax <- 0 #this holds row (i) position in Data frame

#For loop starting it column 5 to final column using nrow().
```

```r
for(i in 1:nrow(deaths_global[,]))
{
num <- deaths_global[i,5] + confirmed_global[i,5]#Holds value to be tested against the current max value

  if (num > max_value) #confitional if loop that updates the max value as iterates through column 5.
    {
    max_value <- num
    RowMax <- i #Holds the row position where the max value is located
    }
}

#holds info. for ob3
ob1Loc_PS <- deaths_global[RowMax,1]
ob1Loc_CR <- deaths_global[RowMax,2]
ob1Loc_Lat <- deaths_global[RowMax,3]
ob1Loc_Long  <- deaths_global[RowMax,4]

#Holds average
deathNconfimred <- deaths_global[RowMax,5] + confirmed_global[RowMax,5]
cat("Province/State:",deaths_global[RowMax,1],"Country/Region:",deaths_global[RowMax,2],"\n")
```

Province/State: Hubei Country/Region: China

```r
cat("Latitude:",deaths_global[RowMax,3],",Longitude:",deaths_global[RowMax,4],"\n")
```

Latitude: 30.9756 ,Longitude: 112.2707

```r
cat("Deaths:",deaths_global[RowMax,5],",confirmed cases:",confirmed_global[RowMax,5],"\n")
```

Deaths: 17 ,confirmed cases: 444

```r
cat("Predicted area of Origin had the most initial deaths and confirmed cases (01.22.2020) at:",deathNc
```

Predicted area of Origin had the most initial deaths and confirmed cases (01.22.2020) at: 461

Objective 2

Where is the most recent area to have a first confirmed case? To do this, you will need to use a for loop, if statement, and subsets.

```r
#find max value in column 5 (10.22.20) and check to see how many max location there is

RowMax <- nrow(confirmed_global[,]) #Qty of Rows 289
ColMax <- ncol(confirmed_global[,]) #QTy of Columns 1147
max_value <- as.numeric(ncol(confirmed_global[,]) * nrow(confirmed_global[,])) #serves as while loop li
num <- 0 #Holds the rotating confirmed_cases value in loop

ColName <- names(confirmed_global) #Loads the the column names into a vector

#initiates the loop counter and used for pulling data from data frame.
i <- 5 #Start the loop at column 5
j <- 1 #Start the loop at column 1 instead of zero, adjust later

#The while loop stops running when the num variable loads a value > 0
while(j < max_value && num == 0)
 {
  num <- confirmed_global[j,i]
```

```r
    if( j %% RowMax == 0) #This conditional statement allows the while loop to irritate-
    {                        #through the each column before looping to next col.
    i <- i + 1 #Iterate through columns 1-1147
    }
  j <- j + 1  #Iterate through rows 1-289
 }

j <- j - 1 #This accounts for the final counter increment when exiting the while loop.

#holds info. for ob3
ob2Loc_PS <- confirmed_global[j,1]
ob2Loc_CR <- confirmed_global[j,2]
ob2Loc_Lat <- confirmed_global[j,3]
ob2Loc_Long  <- confirmed_global[j,4]
```

```r
cat("Province/State: ",confirmed_global[j,1],"Province/State: ",confirmed_global[j,2],"\n")
```

```
## Province/State:  Anhui Province/State:  China
```

```r
cat("Latitude: ",confirmed_global[j,3]," ,Longitude:",confirmed_global[j,4],"\n")
```

```
## Latitude:  31.8257   ,Longitude: 117.2264
```

```r
cat("Predicted area of Origin had the most initial deaths on", ColName[i] ,"at:",confirmed_global[j,i],
```

```
## Predicted area of Origin had the most initial deaths on X1.22.20 at: 1
```

Objective 3

How far away are the areas from objective 2 from where the first confirmed case(s) occurred? Please provide answer(s) in terms of miles. Use the function distm from the R package geosphere to calculate the distance between two coordinates in meters (geosphere::distm). You will need to convert the value returned by distm from meters to miles (this conversion is simple and can be found online). Please use a table or printed statement to describe what Province/State and Country/Region first confirmed cases occurred as well as the distance (in miles) away from the origin. Please print the following: {recent region} is {distance in miles} away from {origin city, origin country}.

```r
#Calculate the distance using distm()
distance_meters <- distm(c(ob1Loc_Long, ob1Loc_Lat), c(ob2Loc_Long, ob2Loc_Lat), fun = distHaversine)
cat("From (ob1):\n")
```

```
## From (ob1):
```

```r
cat("Province/State:",ob1Loc_PS,"Country/Region:",ob1Loc_CR,"\n")
```

```
## Province/State: Hubei Country/Region: China
```

```r
cat("Latitude:",ob1Loc_Lat,",Longitude:",ob1Loc_Long,"\n\n")
```

```
## Latitude: 30.9756 ,Longitude: 112.2707
```

```r
cat("To (ob2):\n")
```

```
## To (ob2):
```

```r
cat("Province/State:",ob2Loc_PS,"Country/Region:",ob2Loc_CR,"\n")
```

```
## Province/State: Anhui Country/Region: China
cat("Latitude:",ob2Loc_Lat,",Longitude:",ob2Loc_Long,"\n\n")

## Latitude: 31.8257 ,Longitude: 117.2264
cat("Distance:",round(distance_meters), "- meters\n") #Print the distance in meters

## Distance: 480239 - meters
cat("Distance:",round(distance_meters/1000),"- Kilometers\n" ) #Print the distance in Kilometers

## Distance: 480 - Kilometers
cat("Distance:",round(distance_meters/1609.34),"- miles\n" ) #1 mile = 1609.34 meters

## Distance: 298 - miles
```

You are asked to make two tables with the top 5 countries that have the most COVID-19 related confirmations and and deaths. Make sure to include all of the counts for the country, not just the counts for one area in the country. To do this we will need to sum all of the values for each country, create new data frames from these values, and use the package kable to convert those data frames into tables.

Hint: Sum each country's counts by subsetting the data frame using a list of countries available in the data set. Use a for loop to iterate through the data frame using the list of countries. For each country, calculate the count sum and assign this value to a list.

GitHub Log

```
git log --pretty=format:"%nSubject: %s%nAuthor: %aN%nDate: %aD%nBody: %b"
```

```
##
## Subject: code and Comments for ob2 updated
## Author: Sal - Figgs0bit
## Date: Sun, 30 Mar 2025 08:09:05 -0700
## Body:
##
## Subject: Updateding notes, ob3 and ob 5 needed.
## Author: Sal - Figgs0bit
## Date: Sun, 30 Mar 2025 05:49:15 -0700
## Body:
##
## Subject: Finished Code for Ob2, need code comments
## Author: Sal - Figgs0bit
## Date: Fri, 28 Mar 2025 12:23:15 -0700
## Body:
##
## Subject: Ob1 code done, need to update code comments
## Author: Sal - Figgs0bit
## Date: Fri, 28 Mar 2025 03:30:43 -0700
## Body:
##
## Subject: R ob1 chunk
## Author: Sal - Figgs0bit
## Date: Fri, 28 Mar 2025 02:12:28 -0700
## Body:
##
## Subject: Objective 1, Deaths finished. Need Confirmation for loops.
```

```
## Author: Sal - Figgs0bit
## Date: Mon, 24 Mar 2025 20:39:21 -0700
## Body:
##
## Subject: Merge branch 'main' of github.com:Figgs0bit/CSIT165-CovidGroupProj
## Author: Sal - Figgs0bit
## Date: Mon, 24 Mar 2025 17:34:18 -0700
## Body:
##
## Subject: added {r Data setup} chunk. Downloads csv from repository to dataframe
## Author: Sal - Figgs0bit
## Date: Mon, 24 Mar 2025 17:34:05 -0700
## Body:
##
## Subject: Uploaded data
## Author: Figgs0bit
## Date: Mon, 24 Mar 2025 17:14:29 -0700
## Body: time_series_covid19 data (global deaths, global recovered)
##
## Subject: Merge branch 'main' of github.com:Figgs0bit/CSIT165-CovidGroupProj
## Author: Sal - Figgs0bit
## Date: Mon, 24 Mar 2025 17:12:01 -0700
## Body:
##
## Subject: Template knited to PDF w/o (\usepackage{tabu}, library(kableExtra))
## Author: Sal - Figgs0bit
## Date: Mon, 24 Mar 2025 17:11:54 -0700
## Body:
##
## Subject: Update README.md
## Author: Figgs0bit
## Date: Mon, 24 Mar 2025 17:10:19 -0700
## Body: Contains copy of Group Contract
##
## Subject: First Commit: setting up Github repository need to add template
## Author: Sal - Figgs0bit
## Date: Mon, 24 Mar 2025 17:05:25 -0700
## Body:
##
## Subject: Initial commit
## Author: Figgs0bit
## Date: Mon, 24 Mar 2025 17:00:21 -0700
## Body:
```