final project

Essraa A.

25/12/2020

Loading libiraries:

library(tidyverse)

library(gdata)

Loading and preview the main data:

main\_data<- read.csv(choose.files(),header = T, stringsAsFactors = T)  
head(main\_data)

## SNo ObservationDate Province.State Country.Region Last.Update Confirmed  
## 1 1 01/22/2020 Anhui Mainland China 1/22/2020 17:00 1  
## 2 2 01/22/2020 Beijing Mainland China 1/22/2020 17:00 14  
## 3 3 01/22/2020 Chongqing Mainland China 1/22/2020 17:00 6  
## 4 4 01/22/2020 Fujian Mainland China 1/22/2020 17:00 1  
## 5 5 01/22/2020 Gansu Mainland China 1/22/2020 17:00 0  
## 6 6 01/22/2020 Guangdong Mainland China 1/22/2020 17:00 26  
## Deaths Recovered  
## 1 0 0  
## 2 0 0  
## 3 0 0  
## 4 0 0  
## 5 0 0  
## 6 0 0

Dropping no needed variables and filter the china data only:

filterd\_data<- main\_data %>%   
 remove.vars(names=c("SNo","Province.State","Last.Update"),info=T) %>%   
 filter(Country.Region=="Mainland China"|Country.Region=="Hong Kong")

##   
## Changing in .   
## Dropping variables: SNo, Province.State, Last.Update

head(filterd\_data)

## ObservationDate Country.Region Confirmed Deaths Recovered  
## 1 01/22/2020 Mainland China 1 0 0  
## 2 01/22/2020 Mainland China 14 0 0  
## 3 01/22/2020 Mainland China 6 0 0  
## 4 01/22/2020 Mainland China 1 0 0  
## 5 01/22/2020 Mainland China 0 0 0  
## 6 01/22/2020 Mainland China 26 0 0

Changing Mainland China & Hong Kong into China:

china\_data <- mutate(filterd\_data,Country.Region= "China")  
head(china\_data)

## ObservationDate Country.Region Confirmed Deaths Recovered  
## 1 01/22/2020 China 1 0 0  
## 2 01/22/2020 China 14 0 0  
## 3 01/22/2020 China 6 0 0  
## 4 01/22/2020 China 1 0 0  
## 5 01/22/2020 China 0 0 0  
## 6 01/22/2020 China 26 0 0

Aggregate data by ObservationDate,with renaming the columns again:

final\_china\_data<- aggregate(list(Confirmed=china\_data$Confirmed,Deaths=china\_data$Deaths,Recovered=china\_data$Recovered),by= list(ObservationDate=china\_data$ObservationDate,Country.Region  
=china\_data$Country.Region),sum)  
head(final\_china\_data)

## ObservationDate Country.Region Confirmed Deaths Recovered  
## 1 01/22/2020 China 547 17 28  
## 2 01/23/2020 China 641 18 30  
## 3 01/24/2020 China 918 26 36  
## 4 01/25/2020 China 1404 42 39  
## 5 01/26/2020 China 2070 56 49  
## 6 01/27/2020 China 2871 82 58

Saving the data in CSV file:

write.table(final\_china\_data,file="final china data.csv",sep=",",row.names = F)