Project\_Report

Tomas Salas

4/28/2020

{r setup, include=FALSE} knitr::opts\_chunk$set(echo = TRUE)

# R Markdown

# Before reading this Report, I strongly advice to read Assessment of Coronavirus.docx

### Purpose:

The purpose of this study is to explore how the United states of America (USA) is affected by current epidemic COVID-19 (as of 04/04/2020) and demonstrated how the state of Kentucky can be affected in the near future.

### Methodology:

The data use in this research is the cumulative counts of coronavirus cases in the United States, at the state level. According to the New York Times (NYT), This data set is the product of dozens of journalists working across several time zones to monitor, analyze data releases and seek clarification from public officials on how they categorize cases. The data was compiled by NYT.

### Limitations:

“Because of the patchwork of reporting methods for this data across more than 50 state and territorial governments and hundreds of local health departments, our journalists sometimes had to make difficult interpretations about how to count and record cases” (NYT). For this reason, the data presented in this research will, in some cases, not match exactly with the information reported by states and counties.

### Introduction

In today’s modern world, it is becoming clear that because of globalization, an infection disease can transmit globally very quickly, never in the history of humanity all the continents can be susceptible to be infected with the same type of disease at very fast rate. Examples like the Bubonic plague (also now as the Black Death) spread widely by limited means of transportation which, according to History, it was likely spread by trading ships (History.com Editors, 2010). But not at the same degree of COVIC-19, How are civilizations today prepare to deal with this scenario? Can a country stop at some degree the rate of transmission of an infection disease inside its borders?

The Coronavirus also known as (CODVI-19) it is good subject to study, recently classified as a pandemic (March 11, 2020) by World Health Organization (WHO). A pandemic is the worldwide spread of a new disease (WHO). At this point, it is not a question if we can stop (CODVI-19) from spreading but if we control the spread of it, Dr. Anthony S. Fauci director of the National Institute of Allergy and infectious diseases (NIAID), has stated multiple times not only to news media networks but also in the USA White House Press Room . This research paper focus on (COVID-19), examining the impact that COVID-19 can have in the population by analysis data, using it can predict a pattern of infection across the USA targeting specifically the state of Kentucky. By knowing which parts of the USA are more likely to have a greater degree of infection rate than others, the government at state and federal levels can create an effective guidelines to try and slow rate of infections in the areas that are consider critical, rather than focusing in the entire country as whole, targeting a specific areas based on analysis can give an edge combating the disease. This problem is important because not only it is affecting the whole world but also the United State of America, specifically the state of Kentucky which is where I reside.

### Literate review

According to an article of Business insiders, as of March 26, 2020 The united states of America is the country with most confirm cases of coronavirus “83,507” (Berke 2020). Surpassing that of China “81,782” which according to WHO is where COVID-19 outbreak begun. How is possible that Country that has more 1.4 billion people has less confirmed cases that the USA which has only about a fourth that amount of that of China as its population “330,633,707”. Do the methods applied in both countries to slow down the rate of infection has something to do with this relation? The first person to test positive from COVID-19 reside in Whuhan, China, Dec. 8, 2019 (Department of Defense). The 45th president of the united states on March 13, 2020 declared COVID-19 outbreak a national emergency (Department of Defense). It took about 4 months since the virus had its first confirm case for the USA to start implanting slow down measures to try to put the breaks to the spread of the virus, on March 16, 2020 social distancing guidelines at all level of society where starting to be implemented (Department of Defense). Was this measures impose by the federal government of the USA where to little or late to prevent the spread of COVID-19 in a meaningful way in the short term?

To answer this questions is important to understand better the virus how it is transmitted, according to the CDC, COVID-19 is thought to spread mainly from person to person, mainly through respiratory droplets produced when an infected person coughs or sneezes. These droplets can land in the mouths or noses of people who are nearby or possibly be inhaled into the lungs. Spread is more likely when people are in close contact with one another (within about 6 feet). According to a study conducted by Los Alamos National Laboratory, Los Alamos, New Mexico, USA. States that the R0 is between 2.2–2.7 (R0 is the number of cases, on average, an infected person will cause during their infectious period). In Figure 1 will show a graphical representation of what this means.

### Findings:

The findings of this study suggest that Kentucky and the majority of the Untied State of America will continue to have an increase in reported cases and deaths of COVID-19, if the population does not adherer to social distancing guidelines created by Centers of Disease Control and Prevention (CDC).

### Conclusion:

It is the utmost importance to control the spread coronavirus in the USA, millions of people can be affected by It if the disease is left untouched. Some places are more affected than others, for example, New York State and New Jersey. The spread of the virus in these states is growing exponentially. If this pattern is not changed, the ability of the states to control the outbreak can become very limited. New York State can become the epicenter of the virus if the state keeps adding new cases to an already alarming number of cases in the state. Kentucky, on the other hand, does not have nearly as big a problem as the other two states mention, with only 755 reported cases. But the utmost attention must be put by its government to ensure the spread of COVID-19 becomes very limited, as models show, if the state does not take the necessary precaution, it will become a tough situation to manage adequately. COVID-19 has proven without a thought that our current civilization is not prepared for dealing with this kind of scenario, especially the USA. There is still much to learn and revised to ensure the risk of another pandemic like COVID-19 will be less in the future.

###Data Analysis and Results

### Methodology for R Code

##### Data was uploaded using (read.CSV and modified with dplyr):

require(dplyr)  
CoronaVirusCasesByStates <- read.csv(file="states.csv",header=TRUE, sep=",")  
CoronaVirusCasesBycounties <- read.csv(file="counties.csv",header=TRUE, sep=",")  
  
#is.data.frame(CoronaVirusCasesByStates)  
#is.data.frame(CoronaVirusCasesByCounty)  
  
CoronaVirusCasesByStates$date <- as.Date(CoronaVirusCasesByStates$date)  
  
NY <- subset(CoronaVirusCasesByStates,state == "New York")  
NY$date <- as.Date(NY$date)  
  
NJ <- subset(CoronaVirusCasesByStates,state == "New Jersey")  
NJ$date <- as.Date(NJ$date)  
  
K <- subset(CoronaVirusCasesBycounties,state == "Kentucky")  
K$date <- as.Date(K$date)  
  
KSpecificJ <- subset(K, county == "Jefferson")

##### Data visualization

require(ggplot2)  
#require(plotly)  
options(scipen=999)  
##graph 1  
 CasesVsState <- ggplot(CoronaVirusCasesByStates, aes(x=state, y=cases)) +   
 geom\_bar(stat="identity", width=.5, fill="tomato3") +   
  
 labs(title="state by cases",   
 subtitle=" In The United States of America",   
 caption="source: New York Times") +   
 theme(axis.text.x = element\_text(angle=65, vjust=0.6))  
  
#CasesVsState<- ggplotly(CasesVsState)  
  
CasesVsState  
  
  
#plotCasesStates <- ggplotly(plotCasesStates)  
  
#graph 2  
DateVsCases <- ggplot(CoronaVirusCasesByStates, aes(x=date, y=cases)) +   
 geom\_bar(stat="identity", width=.5, fill="tomato3") +   
  
 labs(title="state by cases",   
 subtitle=" In The United States of America",   
 caption="source: New York Times") +   
 theme(axis.text.x = element\_text(angle=65, vjust=0.6))  
  
#dateVsCases <- ggplotly(dateVsCases)  
  
#graph 3  
  
  
DateVsCases  
  
ggNY <- ggplot(NY, aes(x=date, y=cases)) +   
 geom\_point() +   
 geom\_smooth(method="loess", se=F) +  
 labs(subtitle="date vs cases",   
 y="cases",   
 x="date",   
 title="New York. Cases by date",   
 caption = "Source: New york Times") +  
 theme(axis.text.x = element\_text(angle=90, vjust=0.6))  
  
plot(ggNY)  
  
#graph 4  
  
ggNJ <- ggplot(NJ, aes(x=date, y=cases)) +   
 geom\_point() +   
 geom\_smooth(method="loess", se=F) +  
 labs(subtitle="date vs cases",   
 y="cases",   
 x="date",   
 title="New Jersey. Cases by date",   
 caption = "Source: New york Times") +  
 theme(axis.text.x = element\_text(angle=90, vjust=0.6))  
  
plot(ggNJ)  
  
  
#Graph 5  
KG <-ggplot(K, aes(x=county, y=cases)) +   
 geom\_bar(stat="identity", width=.3, fill="tomato3") +   
 labs(title="CODVIC-19 cases by County in Kentucky",   
 subtitle="Kentucky",   
 caption="source: New York Time") +   
 theme(axis.text.x = element\_text(angle=65, vjust=0.6))  
  
##KG<-ggplotly(KG)  
  
KG

##### Specific data was used for futher analysis

#Jefferson County  
ggplotRegression <- function (fit) {  
   
 require(ggplot2)  
   
 ggplot(fit$model, aes\_string(x = names(fit$model)[2], y = names(fit$model)[1])) +   
 geom\_point() +  
 stat\_smooth(method = "loess", se=F, col= "blue") +  
 labs(title = paste("Adj R2 = ",signif(summary(fit)$adj.r.squared, 5),  
 "Intercept =",signif(fit$coef[[1]],5 ),  
 " Slope =",signif(fit$coef[[2]], 5),  
 " P =",signif(summary(fit)$coef[2,4], 5)))  
}  
  
  
fit1 <-lm(deaths ~ cases, data = KSpecificJ)  
ggplotRegression(fit1)  
  
myplot <- qplot(0,0)  
myplotXscale <- scale\_x\_continuous(limits = c(0, 900))  
myplotYscale <- scale\_y\_continuous(limits = c(0, 900))  
myplotXlabel <- xlab("Cases in Jefferson")  
myplotYlabel <- ylab("Deaths in Jefferson")  
myplotTitle <- ggtitle("Prediction Model Jefferson")  
  
myplot <- myplot + myplotXscale + myplotYscale + myplotXlabel + myplotYlabel  
  
print(myplot)  
  
myplotHorizontalLine <- geom\_hline(aes(yintercept=0), colour="green")  
myplot <- myplot + myplotHorizontalLine  
print(myplot)  
  
myplotVerticalLine <- geom\_vline(aes(xintercept=0), colour="green")  
myplot <- myplot + myplotVerticalLine  
print(myplot)  
  
myline <- geom\_abline(slope = 0.042477, intercept = -0.087627, color="red")  
  
myplot <- myplot + myline  
print(myplot)  
  
##Kentucky expoential model using python  
  
#GlowScript 2.9 VPython  
  
# tgraph=graph(xtitle="Time [Days]",ytitle="Number Infected", width=475, height=300)  
# f1=gcurve(color=color.blue, markers=False)  
# f2=gcurve(color=color.red)  
#   
# #number of infected humans at start  
# #N=1  
#   
# #infection rate - TRY CHANGING THIS  
# a=.2  
#   
#   
# #starting time  
# t=0  
# #time step  
# dt=.1  
#   
#   
# #this is a loop for 40 days  
# while t<60:  
#   
# #add data to the graph  
#   
# #update number of infected humans  
# N=N+a\*N\*dt  
#   
#   
#   
# #update time  
# t=t+dt  
#   
# f1.plot(t,N)  
#   
#   
# #print("N for a = 0.2 = ",N," people")  
# #print("N for a = 0.19 = ",N2," people")  
  
  
  
ggplotRegression2 <- function (fit) {  
   
 require(ggplot2)  
   
 ggplot(fit$model, aes\_string(x = names(fit$model)[2], y = names(fit$model)[1])) +   
 geom\_point() +  
 stat\_smooth(method = "loess", se=F, col= "blue") +  
 labs(title = paste("Adj R2 = ",signif(summary(fit)$adj.r.squared, 5),  
 "Intercept =",signif(fit$coef[[1]],5 ),  
 " Slope =",signif(fit$coef[[2]], 5),  
 " P =",signif(summary(fit)$coef[2,4], 5)))  
}  
  
  
#fit2 <-lm(deaths ~ cases, data = k)  
#ggplotRegression2(fit2)  
  
  
myplot <- qplot(0,0)  
myplotXscale <- scale\_x\_continuous(limits = c(0, 150000))  
myplotYscale <- scale\_y\_continuous(limits = c(0, 150000))  
myplotXlabel <- xlab("Kentucky Cases")  
myplotYlabel <- ylab("Kentucky Deaths")  
myplotTitle <- ggtitle("Prediction Model Kentucky")  
  
myplot <- myplot + myplotXscale + myplotYscale + myplotXlabel + myplotYlabel  
  
print(myplot)  
  
myplotHorizontalLine <- geom\_hline(aes(yintercept=0), colour="green")  
myplot <- myplot + myplotHorizontalLine  
print(myplot)  
  
myplotVerticalLine <- geom\_vline(aes(xintercept=0), colour="green")  
myplot <- myplot + myplotVerticalLine  
print(myplot)  
  
myline <- geom\_abline(slope = 0.032633, intercept = -0.07065, color="red")  
  
myplot <- myplot + myline  
print(myplot)