Visualizing Uncertainty

Md Easin Hasan

4/09/2021

data<-read.csv(file="owid-covid-data\_new1.csv",header=TRUE)  
dim(data)

## [1] 80745 59

colnames(data)

## [1] "iso\_code"   
## [2] "continent"   
## [3] "location"   
## [4] "date"   
## [5] "total\_cases"   
## [6] "new\_cases"   
## [7] "new\_cases\_smoothed"   
## [8] "total\_deaths"   
## [9] "new\_deaths"   
## [10] "new\_deaths\_smoothed"   
## [11] "total\_cases\_per\_million"   
## [12] "new\_cases\_per\_million"   
## [13] "new\_cases\_smoothed\_per\_million"   
## [14] "total\_deaths\_per\_million"   
## [15] "new\_deaths\_per\_million"   
## [16] "new\_deaths\_smoothed\_per\_million"   
## [17] "reproduction\_rate"   
## [18] "icu\_patients"   
## [19] "icu\_patients\_per\_million"   
## [20] "hosp\_patients"   
## [21] "hosp\_patients\_per\_million"   
## [22] "weekly\_icu\_admissions"   
## [23] "weekly\_icu\_admissions\_per\_million"   
## [24] "weekly\_hosp\_admissions"   
## [25] "weekly\_hosp\_admissions\_per\_million"   
## [26] "new\_tests"   
## [27] "total\_tests"   
## [28] "total\_tests\_per\_thousand"   
## [29] "new\_tests\_per\_thousand"   
## [30] "new\_tests\_smoothed"   
## [31] "new\_tests\_smoothed\_per\_thousand"   
## [32] "positive\_rate"   
## [33] "tests\_per\_case"   
## [34] "tests\_units"   
## [35] "total\_vaccinations"   
## [36] "people\_vaccinated"   
## [37] "people\_fully\_vaccinated"   
## [38] "new\_vaccinations"   
## [39] "new\_vaccinations\_smoothed"   
## [40] "total\_vaccinations\_per\_hundred"   
## [41] "people\_vaccinated\_per\_hundred"   
## [42] "people\_fully\_vaccinated\_per\_hundred"   
## [43] "new\_vaccinations\_smoothed\_per\_million"  
## [44] "stringency\_index"   
## [45] "population"   
## [46] "population\_density"   
## [47] "median\_age"   
## [48] "aged\_65\_older"   
## [49] "aged\_70\_older"   
## [50] "gdp\_per\_capita"   
## [51] "extreme\_poverty"   
## [52] "cardiovasc\_death\_rate"   
## [53] "diabetes\_prevalence"   
## [54] "female\_smokers"   
## [55] "male\_smokers"   
## [56] "handwashing\_facilities"   
## [57] "hospital\_beds\_per\_thousand"   
## [58] "life\_expectancy"   
## [59] "human\_development\_index"

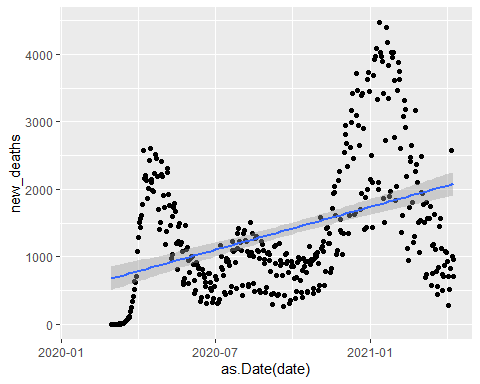
data1 <- data[data$iso\_code == "USA", ]   
dim(data1)

## [1] 445 59

We have obtained the data-set on COVID-19 (coronavirus) by Our World in Data. They have up-to-date data on confirmed cases, deaths, hospitalizations, testing, and vaccinations, throughout the duration of the COVID-19 pandemic. The dimension of this data-set is 80745 by 59.

library(ggplot2)  
library(reshape2)  
library(scales)  
ggplot(data1, aes(x = as.Date(date), y = new\_deaths)) +   
 geom\_point() +   
 geom\_smooth(method = "lm")

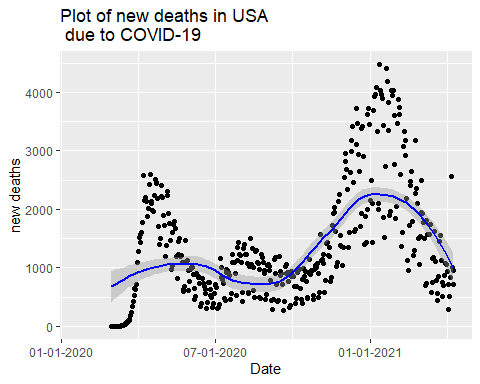
## `geom\_smooth()` using formula 'y ~ x'



# Visualizing Uncertainity Improvement:

ggplot(data1, aes(x = as.Date(date), y = new\_deaths)) +   
 geom\_point()+  
 geom\_smooth(method = loess, se=TRUE, color='blue')+  
 scale\_x\_date(labels = date\_format("%m-%d-%Y"))+  
 labs(x = "Date")+  
 labs(y = "new deaths")+  
 ggtitle("Plot of new deaths in USA \n due to COVID-19")

## `geom\_smooth()` using formula 'y ~ x'



From this visualization we can see the trends and uncertainty of new deaths in USA due to COVID-19. Here we used locally estimated scatter-plot smoothing (LOESS) instead of lm method. It’s pretty clear that the new death rate is decreasing in USA last couple of days since the LOESS line is decreasing and the higher death rate happened at the beginning of 2021.