Accidental drug related deaths 2012-2018

Analyzing using R and RStudio

By Mustafa Sadiq

05/05/2020

The opioid epidemic (also known as the opioid crisis) is a topic that is very current and sensitive. With many laws, drugs are one of the most controlled substance in the United States but still manages to be abused and used in illegal manners and activities.

From 1999 to 2017, more than 399,000 people died from drug overdoses that involved prescription and illicit opioids. In 2017 alone, there were 70,237 recorded drug overdose deaths, and of those deaths, 47,600 involved an opioid. Currently, an estimated 130 people every day in the United States die from an opioid-related drug overdose.
https://www.cdc.gov/mmwr/volumes/67/wr/mm675152e1.htm

For my final project for Data 101 at Rutgers University, I am going to investigate a dataset provided by the State of Connecticut which provides with an extensive look into accidental drug related deaths.

The dataset

Found at data.ct.gov, provided by the Office of the Chief Medical Examiner, “Accidental Drug Related Deaths 2012-2018” is a dataset with 5,105 entries and 41 attributes. Each row is a death and includes date, age, sex, race, city, description of injury, and names of drugs among other many attributes.

	ID	Date	DateType	Age	Sex	Race	ResidenceCity	ResidenceCounty	ResidenceState	DeathCity	Death
1	14-0273	06/28/2014 12:00:00 AM	DateReported	NA							
2	13-0102	03/21/2013 12:00:00 AM	DateofDeath	48	Male	Black	NORWALK			NORWALK	FAIRFI
3	16-0165	03/13/2016 12:00:00 AM	DateofDeath	30	Female	White	SANDY HOOK	FAIRFIELD	CT	DANBURY	
4	16-0208	03/31/2016 12:00:00 AM	DateofDeath	23	Male	White	RYE	WESTCHESTER	NY	GREENWICH	
5	13-0052	02/13/2013 12:00:00 AM	DateofDeath	22	Male	Asian, Other	FLUSHING	QUEENS		GREENWICH	FAIRFI
6	14-0277	06/29/2014 12:00:00 AM	DateofDeath	23	Male	White	BRISTOL			BRISTOL	HARTI
7	12-0205	08/12/2012 12:00:00 AM	DateofDeath	21	Female	White	WEST HARTFORD	HARTFORD		WEST HARTFORD	HARTI
8	13-0404	11/10/2013 12:00:00 AM	DateofDeath	40	Female	White	HAMDEN	NEW HAVEN		NEW HAVEN	NEW I
9	12-0107	04/25/2012 12:00:00 AM	DateofDeath	40	Male	White	EAST HARTFORD	HARTFORD		EAST HARTFORD	HARTI
10	13-0161	05/15/2013 12:00:00 AM	DateofDeath	50	Male	White	MONTVILLE	NEW LONDON		MONTVILLE	NEW I
11	12-0218	08/23/2012 12:00:00 AM	DateofDeath	26	Female	Hispanic, White				STRATFORD	FAIRFI

https://data.ct.gov/Health-and-Human-Services/Accidental-Drug-Related-Deaths-2012-2018/rybz-nyjw

Over the years

In 2018, Connecticut medical providers wrote 43.0 opioid prescriptions for every 100 persons compared to the average U.S. rate of 51.4 opioid prescriptions. Since 2012, this represents a 66 percent decline.

https://www.drugabuse.gov/drugs-abuse/opioids/opioid-summaries-by-state/connecticut-opioid-summary

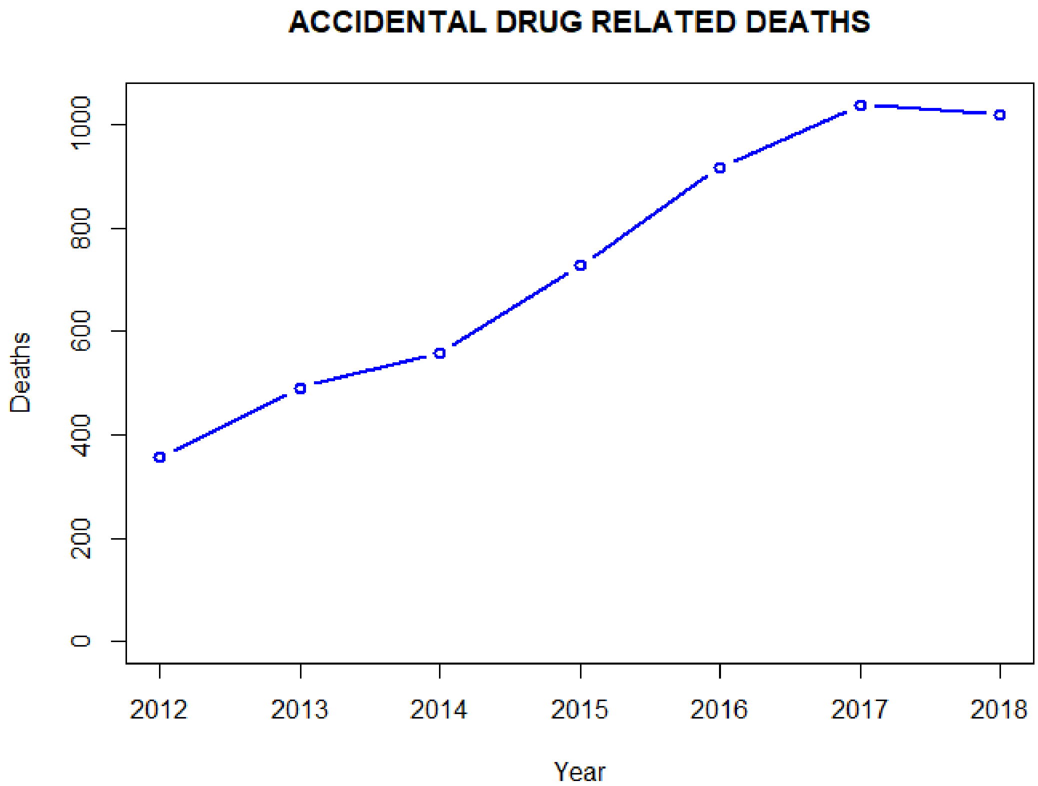
https://www.cdc.gov/drugoverdose/maps/rxrate-maps.html

To look at how accidental drug related deaths varied over the years 2012 to 2018 we plot a simple graph using R and RStudio.

```
> Accidental_Drug_Related_Deaths_2012.2018$Year <- substr(Accidental_Drug_Rel
ated_Deaths_2012.2018$Date,7,10)
> yearcount <- table(Accidental_Drug_Related_Deaths_2012.2018$Year, exclude =
"")
> yearcount
```

```
2012 2013 2014 2015 2016 2017 2018
 355  490  558  727  917 1038 1018
```

```
> plot(yearcount, main='ACCIDENTAL DRUG RELATED DEATHS', xlab = 'Year', ylab=
'Deaths', col='blue', type='b')
```



We can see from our plot that there is positive gradient from 2012 to 2017 with a decline from 2017 to 2018. With opioid prescription on the decline we should keep in mind drug abuse through illegal means.

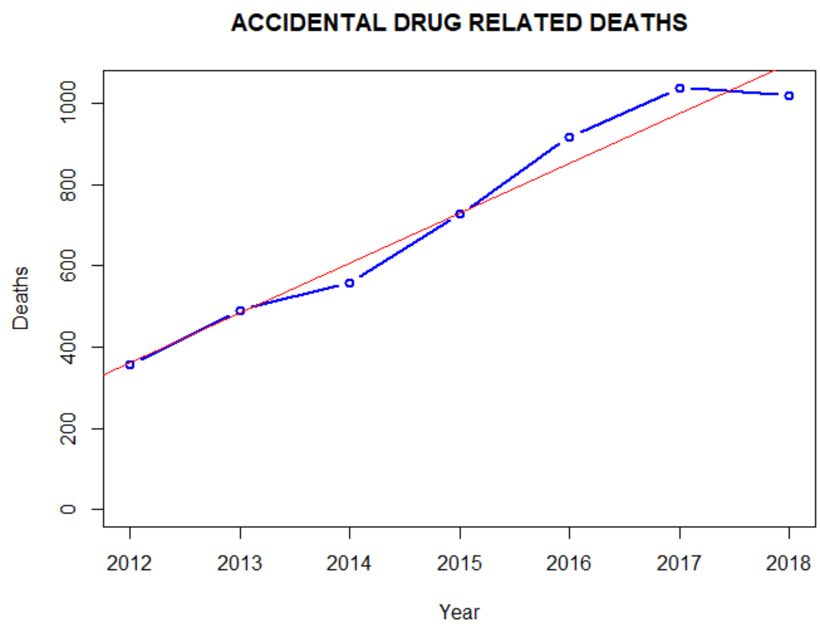
Let us predict this trend into the future:

First, we create a data frame:

```
> Year <- c(2012, 2013, 2014, 2015, 2016, 2017, 2018)
> Deaths <- c(355,  490,  558,  727,  917, 1038, 1018)
> yearly.data <- data.frame(Years, Deaths)
```

Using this data frame, we train a linear model, and use it to predict into the future:

```
> yearly.lm = lm(Deaths~Year, data=yearly.data)
> abline(yearly.lm, col="red")
```




```
> predict(yearly.lm, data.frame(Year=c(2019:2025)))
   1    2    3    4    5    6    7
1221 1344 1467 1590 1713 1836 1959
```

For the year 2019, our model predicts 1221 accidental drug related deaths.

To look into if our prediction model was right, I looked up numbers for the year 2019 from Hartford Courant, the largest daily newspaper in the U.S. state of Connecticut, and often recognized as the oldest continuously published newspaper in the United States. They quote Dr. James Gill, the state's Chief Medical Examiner giving a total of 1,200 deaths for the year 2019.

https://www.ctpost.com/local/article/Fatal-drug-overdoses-rising-in-CT-Town-by-town-15066922.php

As a data skeptic, I then tried to look for any data directly from the Office of the Chief Medical Examiner. On their website they do not have an updated dataset on the statistics subpage yet. I tried to do a search using their search tool and luckily found a document with yearly totals last updated on 2/14/2020. The total for year 2019 was in fact 1,200.

https://portal.ct.gov/-/media/OCME/Statistics/Calendar-Years-2012-to-2019-final.pdf?la=en

Let us add this number to a new dataset which includes values for 2019:

```
> yearly.data.with2019 <- yearly.data
> yearly.data.with2019[nrow(yearly.data.with2019) + 1,] = c(2019, 1200)
```

With actual data now available let us find out the error in prediction for 2019:

```
> DMwR::regr.eval(predicted[1], yearly.data.with2019[8,]$Deaths)
```

```
        mae          mse         rmse         mape
21.00000000 441.00000000  21.00000000   0.01719902
```

Prescriptions over the years

To improve our prediction, lets account for prescriptions over the years. I looked up on the Connecticut Prescription Monitoring and Reporting System (CPMRS) which has statistics for drug prescriptions, including opioid and non-opioid. Since our dataset includes all type of controlled drug related deaths, I compiled the following data:

Year	# of controlled substance prescriptions
2012	NA
2013	5,990,233
2014	6,064,563
2015	6,249,637
2016	6,545,550

2017	6,724,447
2018	6,908,152
2019	7,089,918

https://portal.ct.gov/DCP/Prescription-Monitoring-Program/CTPMP-Statistics

Let us add this data to our dataset:

```
> Prescriptions <- c(NA, 5990233, 6064563, 6249637, 6545550, 6724447, 6908152
)
> yearly.data$Prescriptions <- Prescriptions
> Prescriptionswith2019 <- c(NA, 5990233, 6064563, 6249637, 6545550, 6724447,
6908152, 7089918)
> yearly.data.with2019$Predcriptions <- Prescriptionswith2019
```

Now lets see how our predictions for 2019 looks if we include prescriptions in our training set:

```
> prescriptions.lm = lm(Deaths~Prescriptions, data=yearly.data, na.action=na.
omit)
> predicted1 <- predict(prescriptions.lm, data.frame(Year=c(2019), Prescripti
ons=c(7089918)))
> predicted1
       1
1209.809
```

This prediction, 1,209, is much closer to our actual value of 1,200 for the year 2019!

Let us find out the error for this prediction:

```
> DMwR::regr.eval(predicted1[1], yearly.data.with2019[8,]$Deaths)

         mae          mse         rmse         mape
 9.808566779 96.207982252  9.808566779  0.008107536
```

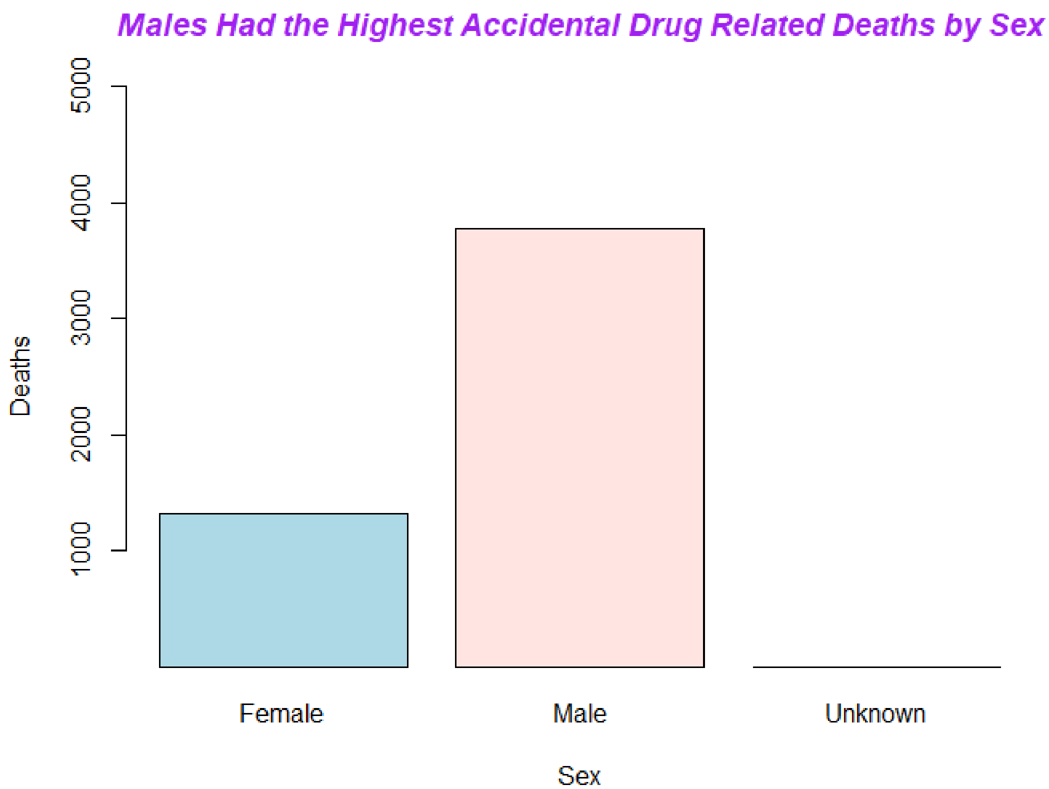
We can say that as prescriptions increase there are more accidental drug related deaths which makes sense because drugs which are heavily controlled by laws, are usually obtained through medical prescriptions or illegal means.

Demographics and other informational attributes

With many other attributes available in the dataset, let us look at some and find if they have any importance:

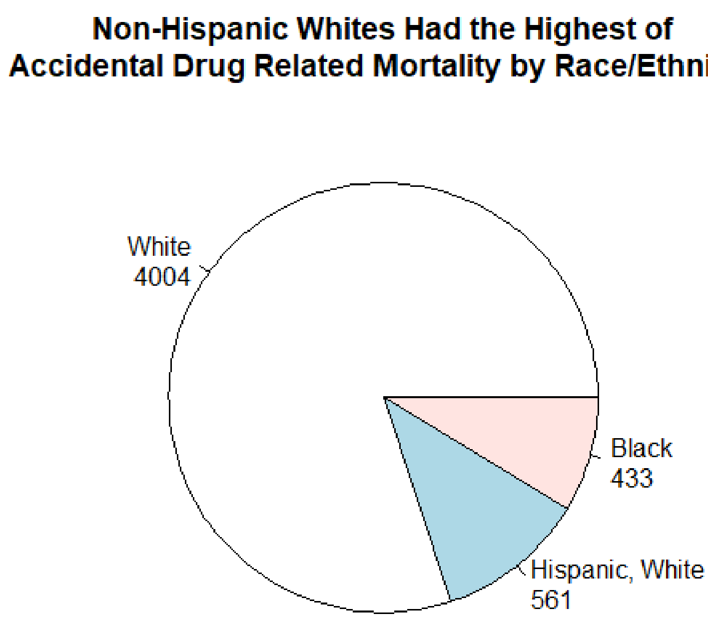
Sex and Accidental Drug Related Deaths

```
> sexcount <- table(Accidental_Drug_Related_Deaths_2012.2018$Sex, exclude = "
")
> barplot(sexcount, ylim = range(1:5000), main = "Males Had the Highest Accid
ental Drug Related Deaths by Sex", font.main = 4, col.main = "purple", xlab =
"Sex", ylab = "Deaths", col = c("lightblue", "mistyrose"))
```



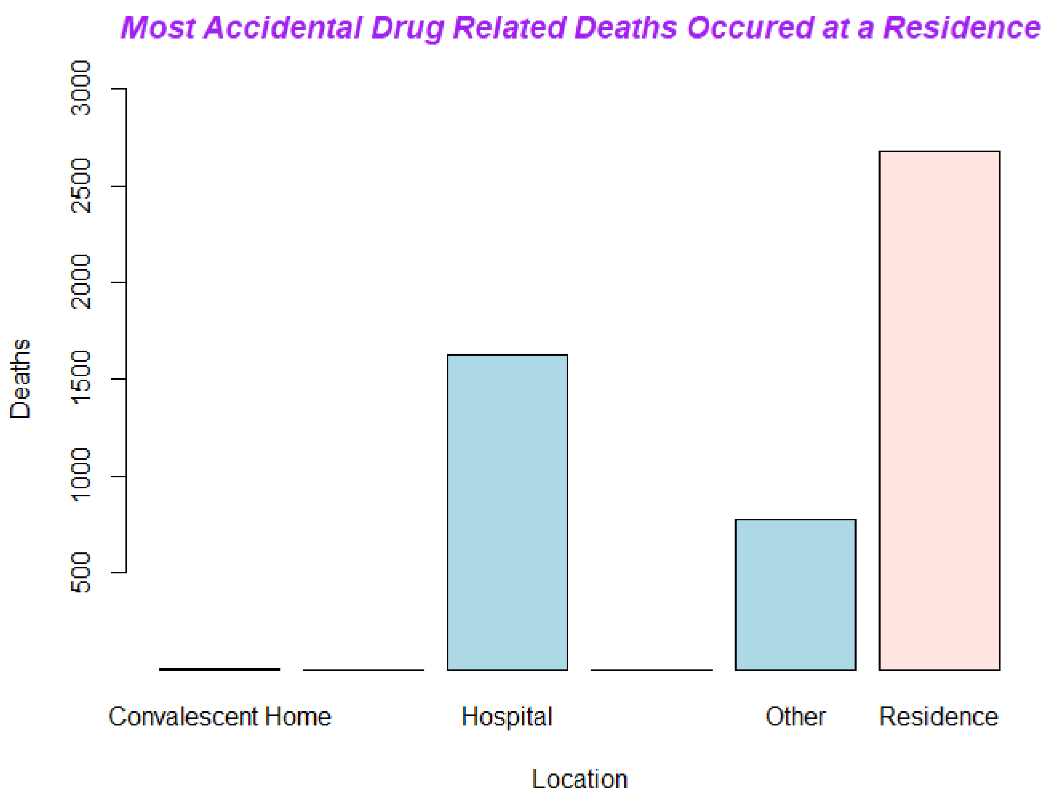
Race and Accidental Drug Related Deaths

```
> racecount <- table(Accidental_Drug_Related_Deaths_2012.2018$Race, exclude =
"")
> racecount <- sort(racecount, decreasing = TRUE)
> lbls <- paste(names(racecount), "\n", racecount, sep="")
> pie(racecount[1:3], labels = lbls,
+     main="Non-Hispanic Whites Had the Highest of\n Accidental Drug Related
Mortality by Race/Ethnicity")
```



Location

```
> locationcount <- table(Accidental_Drug_Related_Deaths_2012.2018$Location, e
xclude = "")
> barplot(locationcount, ylim = range(1:3000), main = "Most Accidental Drug R
elated Deaths Occured at a Residence", font.main = 4, col.main = "purple", xl
ab = "Location", ylab = "Deaths", col = c("lightblue", "mistyrose"))
```



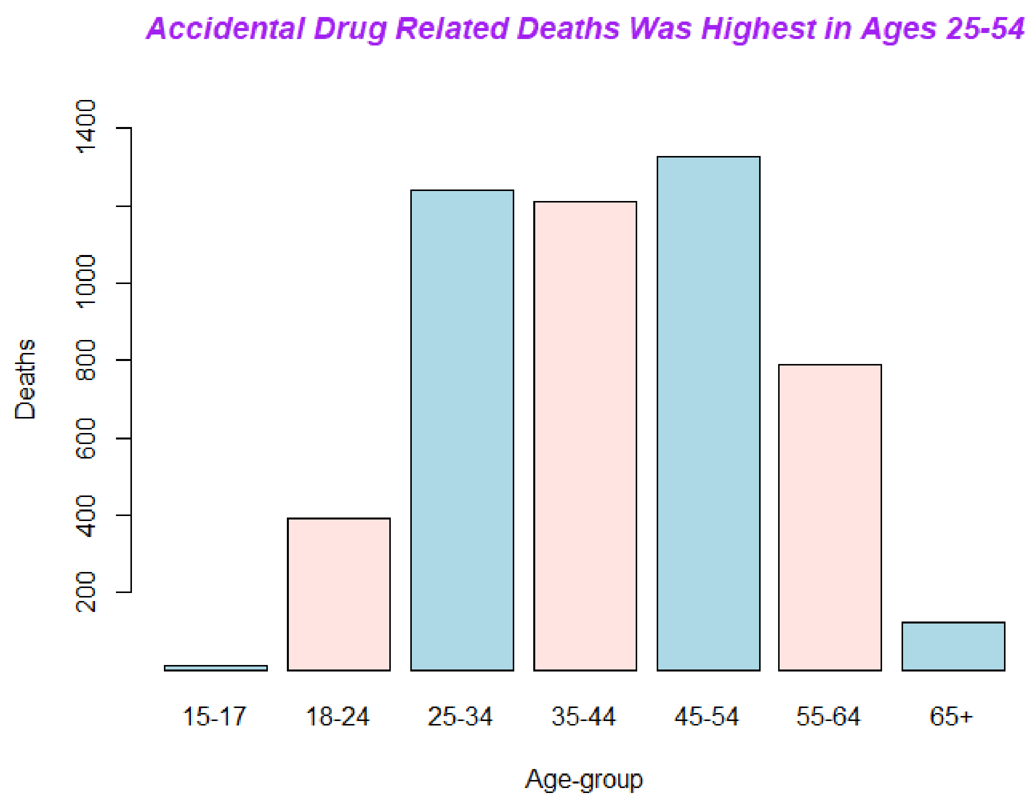
Age

To look at ages we will first categorize every entry into an age group:

```
> attach(Accidental_Drug_Related_Deaths_2012.2018)
> Accidental_Drug_Related_Deaths_2012.2018$agecat[Age >= 15 & Age <= 17] <- "
15-17"
> Accidental_Drug_Related_Deaths_2012.2018$agecat[Age >= 18 & Age <= 24] <- "
18-24"
> Accidental_Drug_Related_Deaths_2012.2018$agecat[Age >= 25 & Age <= 34] <- "
25-34"
> Accidental_Drug_Related_Deaths_2012.2018$agecat[Age >= 35 & Age <= 44] <- "
35-44"
> Accidental_Drug_Related_Deaths_2012.2018$agecat[Age >= 45 & Age <= 54] <- "
45-54"
> Accidental_Drug_Related_Deaths_2012.2018$agecat[Age >= 55 & Age <= 64] <- "
55-64"
> Accidental_Drug_Related_Deaths_2012.2018$agecat[Age >= 65] <- "65+"
> detach(Accidental_Drug_Related_Deaths_2012.2018)
```

Now lets plot a graph

```
> agecount <- table(Accidental_Drug_Related_Deaths_2012.2018$agecat, exclude
= "")
> barplot(agecount[1:7], ylim = range(1:1500), main = "Accidental Drug Relate
d Deaths Was Highest in Ages 25-54", font.main = 4, col.main = "purple", xlab
= "Age-group", ylab = "Deaths", col = c("lightblue", "mistyrose"))
```



It looks like there is a major highlight in every attribute. This can be used by lawmakers and scientists/researchers to better understand drug abuse and make future plans to manage drugs and decrease accidental drug related deaths.

With our data study and results consistent with reality, I think it is very much possible that data can be used in any global crisis as a tool to fight against it.