Data Vizualization HomeWork 4

Isaiah Thompson Ocansey

2023-02-27

```
rm(list = ls())
library(ggplot2)
library(dplyr)
library(scales)
```

```
df<-read.csv(file="owid-covid-data.csv",sep=",", header = TRUE)
dim(df)
```

```
## [1] 259146     67
```

The covid 19 data has 259146 observations and 67 variables ‘

```
index <- df$location == 'Bahamas'
Total_cases_per_million<-df[index,]$total_cases_per_million
total_deaths_per_million<-df[index,]$total_deaths_per_million
diabetes_prevalence<-df[index,]$diabetes_prevalence
dat<-cbind.data.frame(Total_cases_per_million,diabetes_prevalence,total_deaths_per_million)
dim(dat)
```

```
## [1] 1073    3
```

I selected the total number of cases per million in the Bahamas and looked at it’s distribution and then plot it against total number of deaths to see whether an increased number of cases translates into an increased number of deaths.

```
library(EnvStats)
```

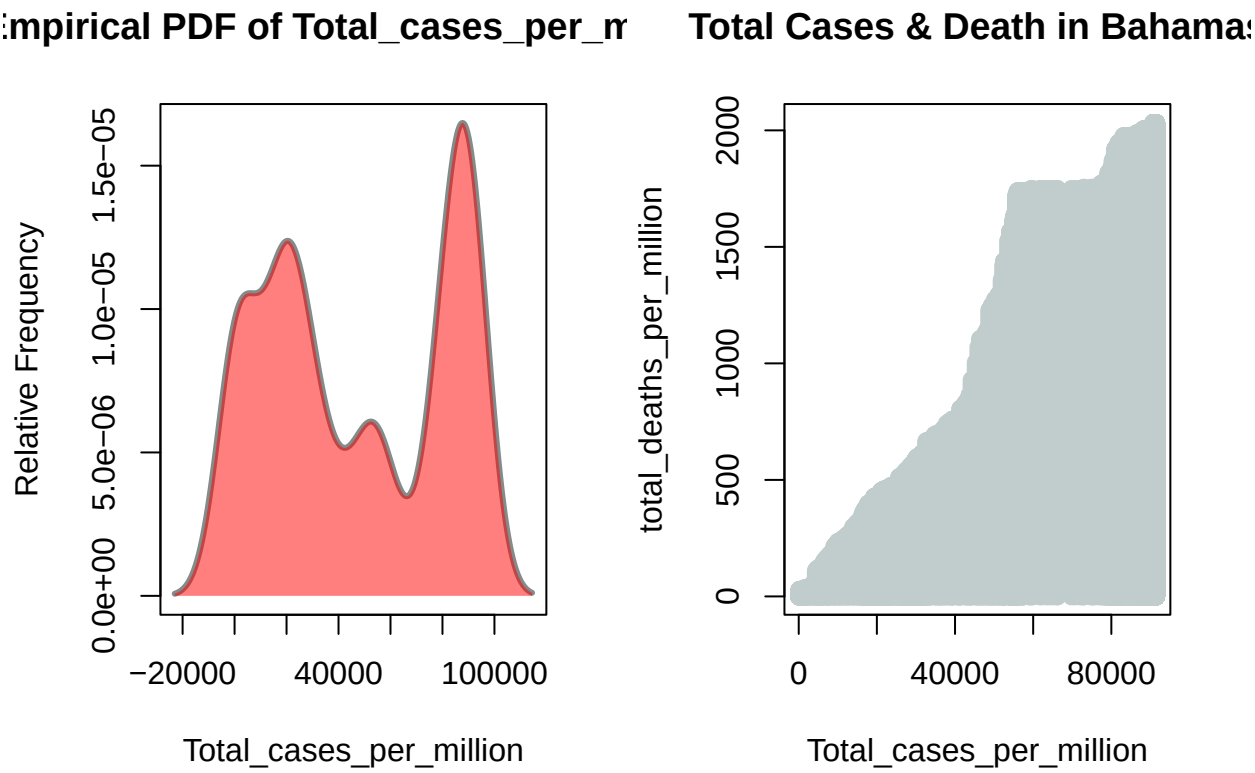
```
##
## Attaching package: 'EnvStats'
```

```
## The following objects are masked from 'package:stats':
##
##     predict, predict.lm
```

```
## The following object is masked from 'package:base':
##
##     print.default
```

```r
par(mfrow=c(1,2))

epdfPlot(Total_cases_per_million, # Vector with data
         curve.fill = TRUE, # Fill the area
         curve.fill.col = rgb(1, 0, 0, alpha = 0.5), # Area color
         epdf.col = "azure4") # Line color
plot(Total_cases_per_million,total_deaths_per_million,main="Total Cases & Death in Bahamas",
type="h",
col="azure3",lwd = 9,mex = 0.2, cex = 0.2)
```



The first visualization on the left depicts the empirical pdf of the total number of cases per million in the Bahamas. We can observe that the total cases per million in the Bahamas is not normally distributed. The second visualization on the right seeks to ascertain if an increase in the number of total cases translates into an increase number of deaths. Clearly, we can observe that as total number of cases increases, there is a marginal increase in the total number of deaths.