COVID-19 in Poland

Joyce Lee

12/09/2020

In this post, we will be looking at how the Corona virus has been developed in Poland, a country located in Central Europe, with a total popolation of nearly 40 millions. Then, we will try to estimate the current situation

First, let's import neccessary libraries and the Covid-19 data from the European CDC. Notice that the original data contains the information of all european countries. Therefore, the first step is to get the data only for Poland. This is done by subsetting the original dataset with the criteria (countriesAndTerritories == "Poland").

```
#Uncomment the following code if any package is not installed locally
#install.packages("plotly")
#install.packages("magrittr")
#install.packages("dplyr")
#install.packages("lubridate")
library(tidyverse)
library(utils)
library(data.table)
library(plotly)
library(shiny)
library(magrittr)
library(dplyr)
library(lubridate)

Original_DT <- read.csv("https://opendata.ecdc.europa.eu/covid19/casedistribution/csv", na.strings = ""

Poland_DT<-as.data.table(subset(Original_DT, countriesAndTerritories == "Poland"))
```

Data Story 1: How has Corona virus been developed in Poland?

Unavoidly, Poland also suffers from the pandemic. But nearly eight months since the pandemic, we would like to know: how has the Corona virus been developed in this country?

To answer this question, the most intuitive way of knowing the outbreak is to look at the daily new cases graph, which records daily new cases from March 4th, 2020 to the current day (September 13, 2020). We can observe that in the frist month of the outbreak, the virus spreaded expotentially and then the curve seems to be flattened in the next few months until late July. In other words, during this period, everyday there are a number of new cases but the spread was still under control intead of growing exponentially. However, a second bounce was spotted in August given the largest number (903) of new cases on August 22 since the pandemic. This also corresponds to what the another graph (Cumulative number for 14 days per 100 thousands population) implies. We all know for this corona virus that its incubation period could be 14 days. Therefore, it is worth to investigate into his graph. From the graph, we can find that it shares similar curve pattern with the daily new cases graph where two waves of spread happened in March and August respectively. When the curve extends in the horizontal direction or has a downward trend, it indicates that

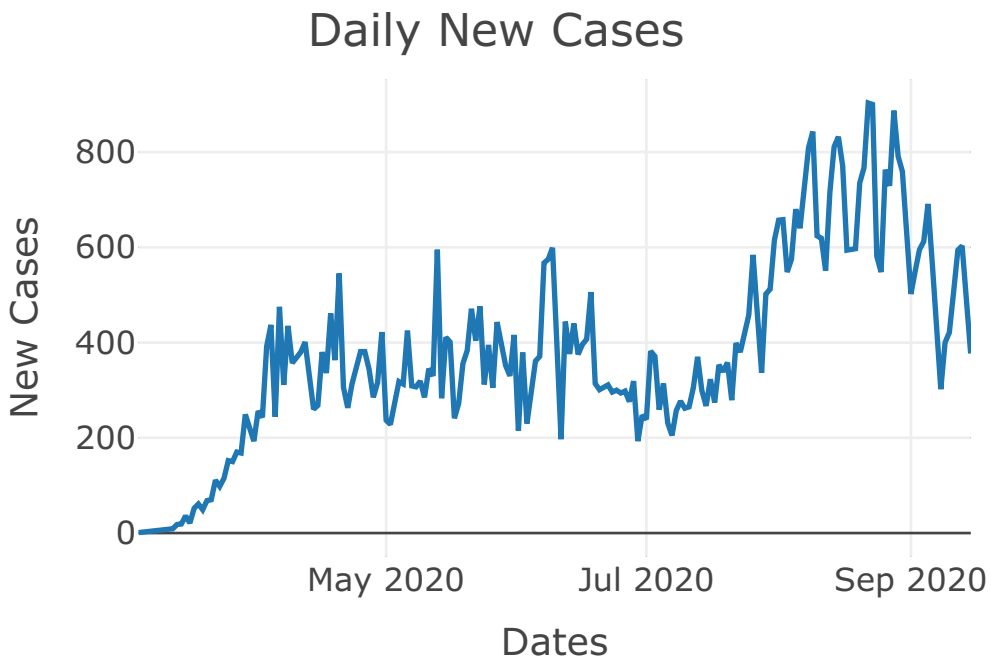
the transmission of the virus is effectively controlled and the epidemic is developing in a good direction. Therefore, when we see the curve goes down from sepetember, we are relatively confident to say that the pandemic is turning to a good direction in this country.

```
Dates_Vector<-rev(Poland_DT[, dateRep])
Dates_Trans<-as.Date(Dates_Vector, format="%d/%m/%y")
Cases_Vector<-rev(Poland_DT[, cases])
Death_Vector<-rev(Poland_DT[, deaths])
CumuNum_14days_Vector<-rev(Poland_DT[, Cumulative_number_for_14_days_of_COVID.19_cases_per_100000])

Cases_df<-data.frame(Dates_Trans, Cases_Vector, Death_Vector, CumuNum_14days_Vector)

#plot_ly(data = Cases_df, x = ~Dates_Vector, y = ~Cases_Vector)
fig1<-plot_ly(data = Cases_df, x = ~Dates_Trans, y = ~Cases_Vector, type = 'scatter', mode = 'lines') %

  layout(title = 'Daily New Cases',
         xaxis = list(title = 'Dates',
                      zeroline = TRUE),
         yaxis = list(title = 'New Cases'))
fig1
```




```
fig2<-plot_ly(data = Cases_df, x = ~Dates_Trans, y = ~CumuNum_14days_Vector, type = 'scatter', mode = '

  layout(title = 'Cumulative_Number_14Days_Per100KPopulation',
         xaxis = list(title = 'Dates',
                      zeroline = TRUE),
         yaxis = list(title = 'Cumu Cases'))
fig2
```

Data Story 2: Is the situation really turning to a good direction now?

If we only look at Figure.1 and Figure.2, it seems that there is a turning point from Sepetember. We might easily come to the conclusion that the pandemic is turning to a good direction, the spread is slowing down.

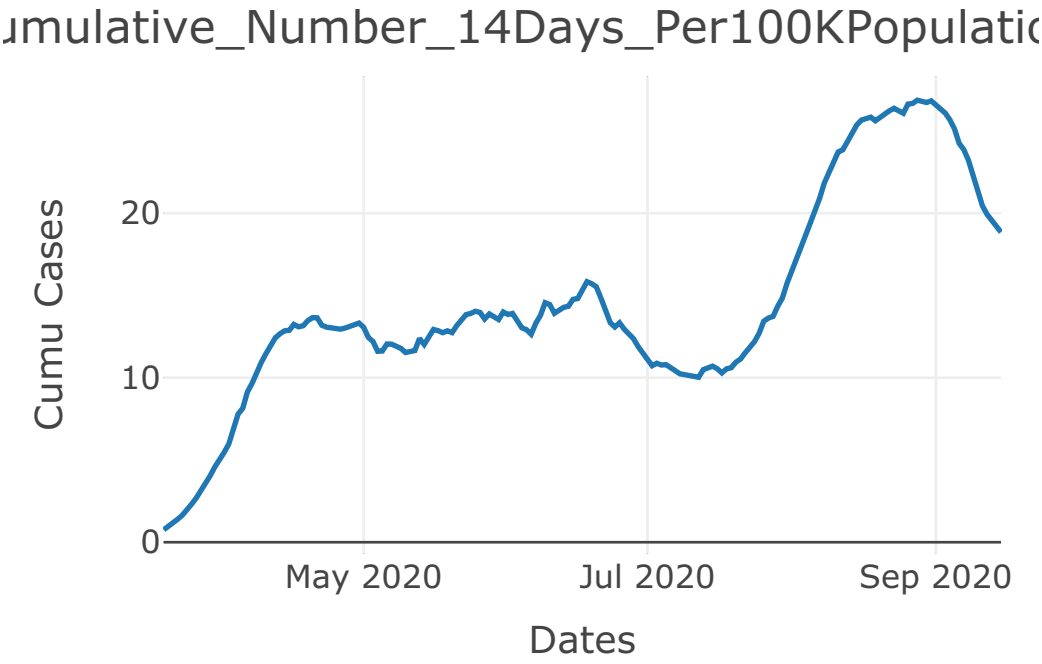
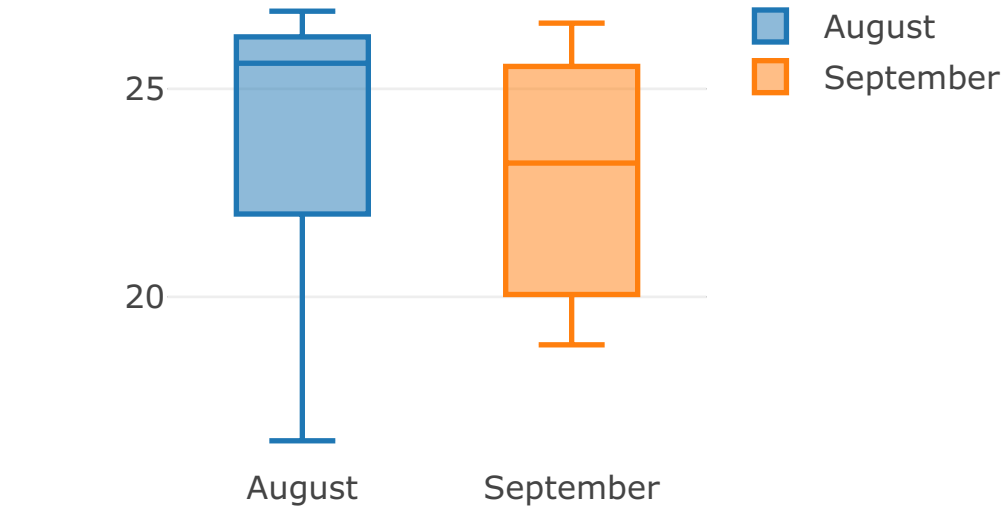


Figure 1: Figure2: 14-Days Cumulative Cases Per 100K Population Curve

Hold on... Are we sure? To further check whether our conclusion is valid or not, we can take the 14-days cumulative number of cases data to plot boxplots for both Sepetember and August. From the boxplot, it shows that the median 14-days cumulative number of cases for August is greater than that of September with a difference of nearly 2. However, is this difference significantly evident? Or in other word, does this difference significantly indicate the spread is slowing down? In order to test this hypothsis, we need to conduct an unpaired t-test. From the following R output, the p-value of the test is 0.66, which is much greater than the significant level alpha = 0.05. We can conclude that there is no significant evidence to indicate that the mean 14-days cumulative number of cases for August is greater than that of Sepetember. Unfortunately, we can't say for sure that the situation is significantly turning to a good direction.

```
CumCaseSept<-subset(Poland_DT$Cumulative_number_for_14_days_of_COVID.19_cases_per_100000, Poland_DT$mon
CumCaseAugust<-subset(Poland_DT$Cumulative_number_for_14_days_of_COVID.19_cases_per_100000, Poland_DT$m

fig3 <- plot_ly(y = CumCaseAugust, name = "August", type = "box") %>% add_trace(y = CumCaseSept, name =
fig3
```




```
#Unpaired t-test
Compare <- t.test(CumCaseAugust, CumCaseSept, var.equal = TRUE)
Compare
```

```
##
##  Two Sample t-test
##
## data:  CumCaseAugust and CumCaseSept
## t = 1.0621, df = 44, p-value = 0.294
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.9234839  2.9815155
## sample estimates:
## mean of x mean of y
##  23.88712  22.85811
```