

BÀI TẬP THỰC HÀNH BUỔI 1

Lớp: DS102.N21.2

MSSV: 21520488

Tên: Nguyễn Thị Huyền Trang

Câu a: Code lại các ví dụ trong Phần 3.

```
rm(list=ls())
```

```
coronaData <- read.csv("data/covid_19_data.csv")
```

#Liệt kê số cột trong bảng dữ liệu: Lệnh ncol()

```
ncol(coronaData)
```

→ 8

#Liệt kê số dòng: lệnh nrow()

```
nrow(coronaData)
```

→ 306429

#In ra 10 dòng đầu trong bảng dữ liệu: Lệnh head()

```
head(coronaData,10)
```

	SNo	ObservationDate	Province.State	Country.Region	Last.Update	Confirmed	Deaths	Recovered
1	1	01/22/2020	Anhui	Mainland China	1/22/2020 17:00	1	0	0
2	2	01/22/2020	Beijing	Mainland China	1/22/2020 17:00	14	0	0
3	3	01/22/2020	Chongqing	Mainland China	1/22/2020 17:00	6	0	0
4	4	01/22/2020	Fujian	Mainland China	1/22/2020 17:00	1	0	0
5	5	01/22/2020	Gansu	Mainland China	1/22/2020 17:00	0	0	0
6	6	01/22/2020	Guangdong	Mainland China	1/22/2020 17:00	26	0	0
7	7	01/22/2020	Guangxi	Mainland China	1/22/2020 17:00	2	0	0
8	8	01/22/2020	Guizhou	Mainland China	1/22/2020 17:00	1	0	0
9	9	01/22/2020	Hainan	Mainland China	1/22/2020 17:00	4	0	0
10	10	01/22/2020	Hebei	Mainland China	1/22/2020 17:00	1	0	0

#In ra tên các biến (tên cột) của dữ liệu: lệnh names()

```
names(coronaData)
```

```
[1] "SNo" "ObservationDate" "Province.State" "Country.Region" "Last.Update" "Confirmed" "Deaths" [8] "Recovered"
```

#Tạo biến countryCorona lưu giá trị là các quốc gia có dịch Corona (Cột Country.Region)

```
countryCorona <- coronaData['Country.Region']
```

#Liệt kê số lượng ca lây nhiễm được xác nhận (biến Confirmed) nhiều nhất và lưu vào biến maxConfirmedCases. Sử dụng lệnh max()

```
maxConfirmedCases <- max(coronaData$Confirmed)
```

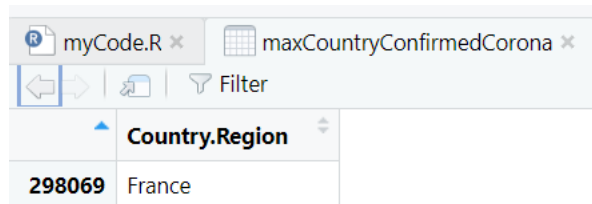
<code>maxConfirmedCases</code>	<code>5863138</code>
--------------------------------	----------------------

#Liệt kê các dữ liệu về covid-19 tại quốc gia Trung Quốc đại lục (Mainland China) và lưu vào biến coronaChina.

```
coronaChina <- coronaData[which(coronaData$Country.Region=='Mainland China'),]
```

#Tìm quốc gia (Country.Region) có số ca lây nhiễm nhiều nhất:

```
maxCountryConfirmedCorona <-  
coronaData[which(coronaData$Confirmed==maxConfirmedCases),]['Country.Region']
```



	Country.Region
298069	France

#Tìm tỉnh (Province.State) có số ca lây nhiễm nhiều nhất

```
maxStateConfirmedCorona <-  
coronaData[which(coronaData$Confirmed==maxConfirmedCases),]['Province.State']
```

#Lấy dữ liệu theo ngày tháng

```
coronaData$ObservationDate <- as.Date(coronaData$ObservationDate, "%m/%d/%Y")
```

#Lấy dữ liệu trong tháng 1/2020: bắt đầu từ 01/01/2020 đến 31/01/2020

```
data_jan <- coronaData[which(coronaData$ObservationDate>= "2020-01-01" &  
coronaData$ObservationDate <= "2020-01-31"), ]
```

Câu b: Tìm dữ liệu về số ca lây nhiễm tại Vietnam (Country.Region == 'Vietnam') và lưu vào biến coronaVietnam.

```
coronaVietnam <- coronaData[which(coronaData$Country.Region == 'Vietnam'),]
```

Câu c: In ra số ca lây nhiễm nhiều nhất tại Việt Nam (Sử dụng lệnh print() trong R):

```
max_Confirmed_Cases_VN <- max(coronaVietnam$Confirmed)  
print(max_Confirmed_Cases_VN)
```

➔ 6908

Câu d: Tìm dữ liệu về số ca lây nhiễm tại Việt Nam trong tháng 02:

```
coronaVietnam_t2<-coronaVietnam[which(coronaVietnam$ObservationDate>='2021-02-01' &  
coronaVietnam$ObservationDate<='2021-02-28'),]
```

Câu e: In ra số dữ liệu về ca lây nhiễm nhiều nhất trong tháng 01 và 02 tại Việt Nam (Lấy năm 2021):

```
coronaVietNam_t1_t2 <- coronaData[which(coronaData$ObservationDate>= "2021-01-01" &
coronaData$ObservationDate <= "2021-02-28" & coronaData$Country.Region=='Vietnam'),]
max_coronaVietnam_t1_t2_Confirmed <- max(coronaVietNam_t1_t2$Confirmed)
print(max_coronaVietnam_t1_t2_Confirmed)
➔ 2448
```

Câu f: Thực hiện tương tự câu e) cho Indonesia và Philipine

```
# Indo
coronaIndonesia_t1_t2 <- coronaData[which(coronaData$ObservationDate>= "2021-01-01" &
coronaData$ObservationDate <= "2021-02-28" & coronaData$Country.Region=='Indonesia'),]
max_coronaIndonesia_t1_t2_Confirmed <- max(coronaIndonesia_t1_t2$Confirmed)
print(max_coronaIndonesia_t1_t2_Confirmed)
➔ 1334634

#Philippines
coronaPhilippines_t1_t2 <- coronaData[which(coronaData$ObservationDate>= "2021-01-01" &
coronaData$ObservationDate <= "2021-02-28" & coronaData$Country.Region=='Philippines'),]
max_coronaPhilippines_t1_t2_Confirmed <- max(coronaPhilippines_t1_t2$Confirmed)
print(max_coronaPhilippines_t1_t2_Confirmed)
➔ 576352
```

Câu g: In ra dữ liệu về ca nhiễm ghi nhận (Confirmed) của Trung Quốc trong khoảng thời gian từ 01/02/2021 cho đến 15/02/2021. In ra màn hình sử dụng lệnh print()

```
coronaChina_1_15_t2 <- coronaData[which(coronaData$ObservationDate>= "2021-02-01" &
coronaData$ObservationDate <= "2021-02-15" & coronaData$Country.Region=='Mainland
China'),]
print(coronaChina_1_15_t2["Confirmed"])
```

Câu h: Thống kê số lượng record theo từng tỉnh của Trung Quốc trong tháng 02/2021

```
coronaChina$ObservationDate <- as.Date(coronaChina$ObservationDate, "%m/%d/%Y")
table(coronaChina[which(coronaChina$ObservationDate >='2021-02-01'&
coronaChina$ObservationDate <='2021-02-28'),]$Province.State)
```

Anhui	Beijing	Chongqing	Fujian	Gansu	Guangdong	Guangxi	Guizhou	Hainan
28	28	28	28	28	28	28	28	28
Hebei	Heilongjiang	Henan	Hubei	Hunan	Inner Mongolia	Jiangsu	Jiangxi	Jilin
28	28	28	28	28	28	28	28	28
Liaoning	Ningxia	Qinghai	Shaanxi	Shandong	Shanghai	Shanxi	Sichuan	Tianjin
28	28	28	28	28	28	28	28	28
Tibet	Unknown	Xinjiang	Yunnan	Zhejiang				
28	28	28	28	28				

Câu i: Đếm số lượng ca nhiễm theo từng tỉnh của Trung Quốc trong tháng 02/2021.

```
library(dplyr)
```

```
coronaChina_feb <- coronaData[which(coronaData$ObservationDate>='2021-02-01' &
coronaData$ObservationDate<='2021-02-15' & coronaData$Country.Region=="Mainland
China"),]
```

```
table1 <- coronaChina_feb %>% group_by(coronaChina_feb$Province.State)
```

```
table2 <- table1 %>% summarise(last(Confirmed)-first(Confirmed))
```

```
table2
```

```
# A tibble: 32 × 2
  `coronaChina_feb$Province.State` `last(Confirmed) - first(Confirmed)`
  <chr>                                <dbl>
1 Anhui                                0
2 Beijing                              7
3 Chongqing                            0
4 Fujian                               7
5 Gansu                                0
6 Guangdong                           44
7 Guangxi                              0
8 Guizhou                              0
9 Hainan                               0
10 Hebei                               4
```

Câu k: Tìm dữ liệu ca tử vong của Trung Quốc trong khoảng thời gian từ 01/02/2021 cho đến 15/02/2021. In ra màn hình sử dụng lệnh print()

```
china_data<-coronaData[which(coronaData$Country.Region=="Mainland China"),]
```

```
china_feb=china_data[which(china_data$ObservationDate>='2021-02-01' &
china_data$ObservationDate<='2021-02-15'),]
```

```
dataDeaths_t2 <- china_feb[c('Country.Region', 'ObservationDate', 'Deaths')]
```

```
print(dataDeaths_t2)
```

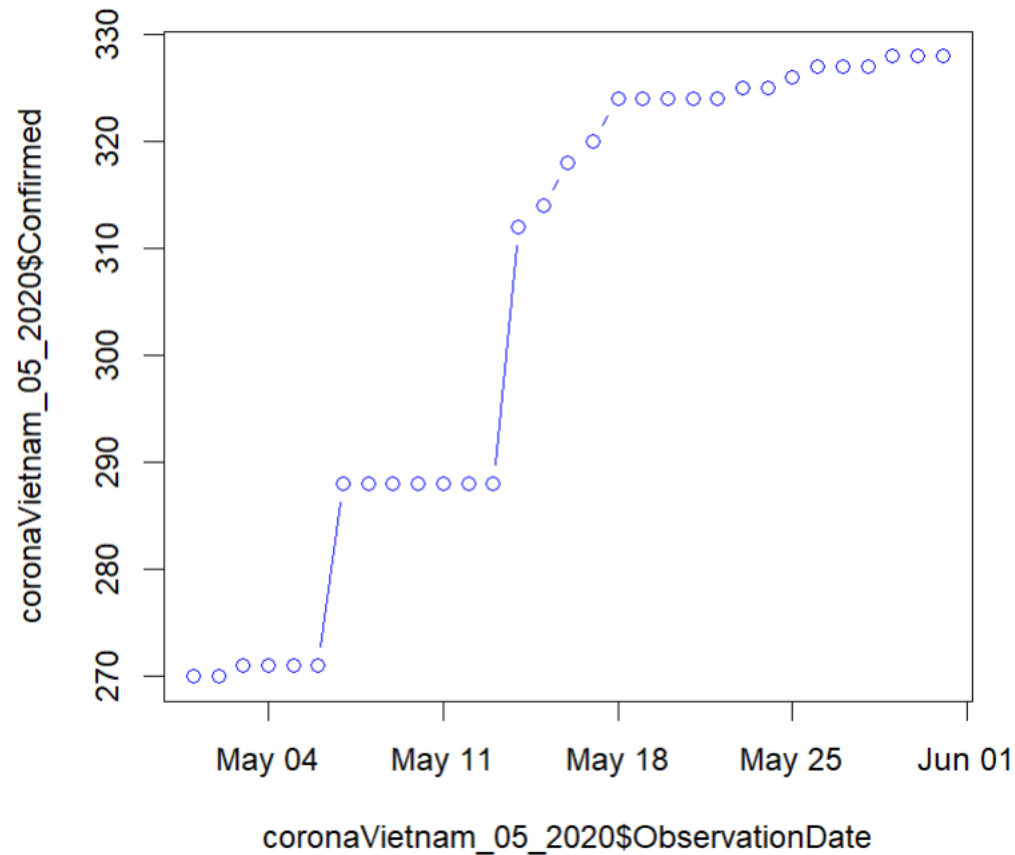
Câu l: *Có nhận xét gì về số ca nhiễm mới tại Việt Nam giữa tháng 05/2020 và tháng

05/2021. Vẽ biểu đồ đường thể hiện số ca nhiễm mới trong 2 tháng trên. Gợi ý: Dùng hàm plot() trong R.

#Năm 05/2020:

```
coronaVietnam_05_2020<-coronaVietnam[which(coronaVietnam$ObservationDate>="2020-05-01"&coronaVietnam$ObservationDate<="2020-05-31"),]
```

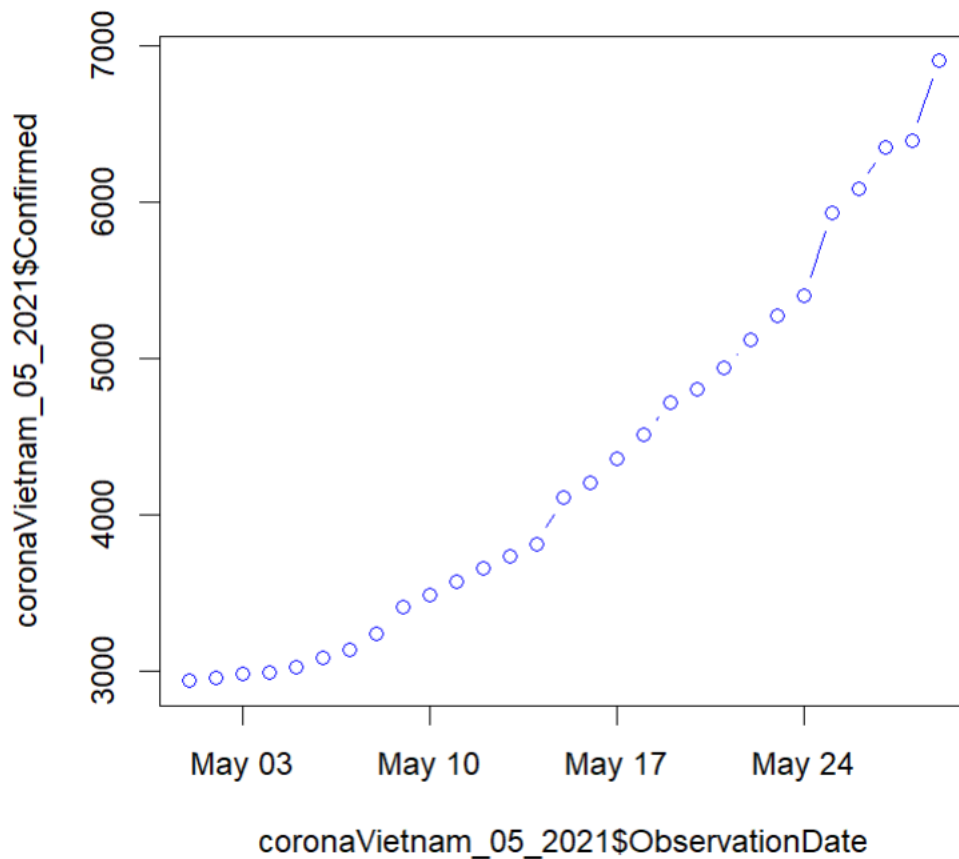
```
plot(coronaVietnam_05_2020$ObservationDate,coronaVietnam_05_2020$Confirmed,type="b",col="blue")
```



#Năm 05/2021

```
coronaVietnam_05_2021<-coronaVietnam[which(coronaVietnam$ObservationDate>="2021-05-01"&coronaVietnam$ObservationDate<="2021-05-31"),]
```

```
plot(coronaVietnam_05_2021$ObservationDate,coronaVietnam_05_2021$Confirmed,type="b",col="blue")
```



Nhật xét: ở năm 2021 độ dốc tăng số ca tuy ít dốc hơn 2020 nhưng mà số lượng tăng là tới con số hơn 4000 ca

Câu m: * Vẽ biểu đồ về số ca lây nhiễm nhiều nhất của 3 quốc gia: Vietnam, Indonesia và Philippine trong 2 tháng gồm 01 và tháng 02 năm 2021.

```
#Vietnam
```

```
coronaVietnam_t1<-coronaVietnam[which(coronaVietnam$ObservationDate>='2021-01-01' &
coronaVietnam$ObservationDate<='2021-01-31'),]
```

```
case_t1<-max(coronaVietnam_t1$Confirmed)
```

```
coronaVietnamcase_t1<-coronaVietnam_t1[which(coronaVietnam_t1$Confirmed==case_t1),]
```

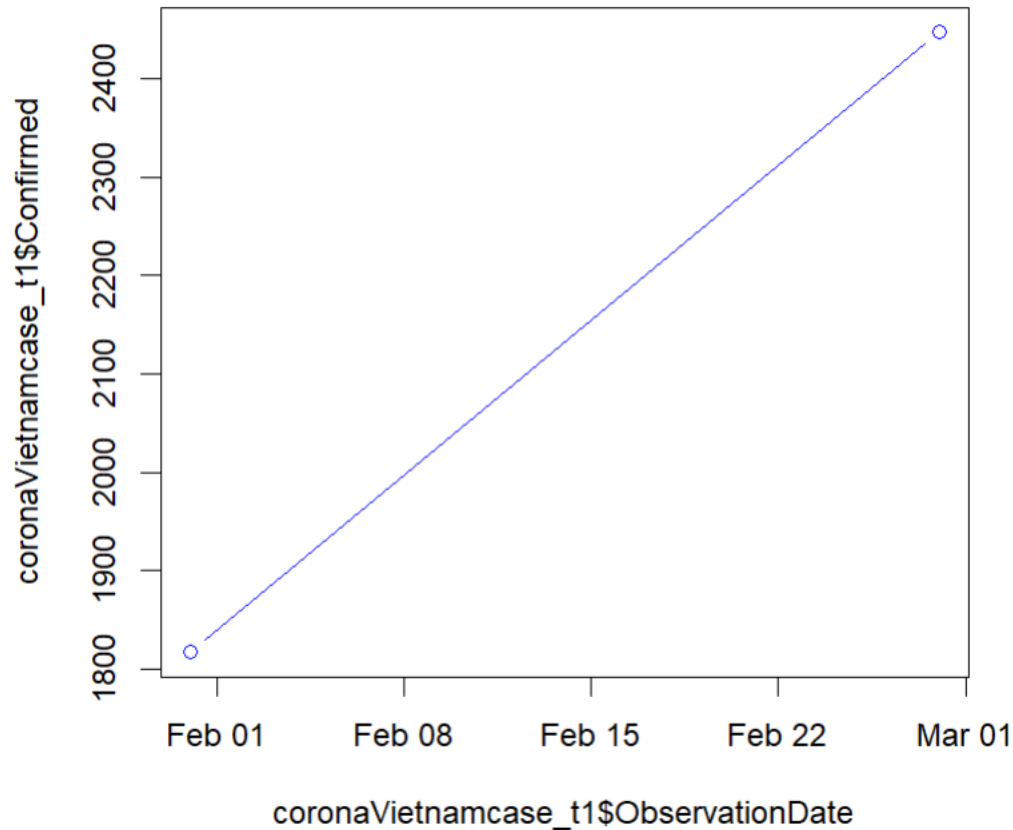
```
coronaVietnam_t2<-coronaVietnam[which(coronaVietnam$ObservationDate>='2021-02-01' &
coronaVietnam$ObservationDate<='2021-02-28'),]
```

```
case_t2<-max(coronaVietnam_t2$Confirmed)
```

```
coronaVietnamcase_t2<-coronaVietnam_t2[which(coronaVietnam_t2$Confirmed==case_t2),]
```

```
coronaVietnamcase_t1[nrow(coronaVietnamcase_t1) + 1,]=coronaVietnamcase_t2
```

```
plot(coronaVietnamcase_t1$ObservationDate,coronaVietnamcase_t1$Confirmed,type="b",col="blue")
```



#Indonesia

```
coronaIndo_t1<-coronaData[which(coronaData$ObservationDate>='2021-01-01' &  
coronaData$ObservationDate<='2021-01-31' & coronaData$Country.Region=='Indonesia' ),]
```

```
case_Indo_t1<-max(coronaIndo_t1$Confirmed)
```

```
coronaIndocase_t1<-coronaIndo_t1[which(coronaIndo_t1$Confirmed==case_Indo_t1),]
```

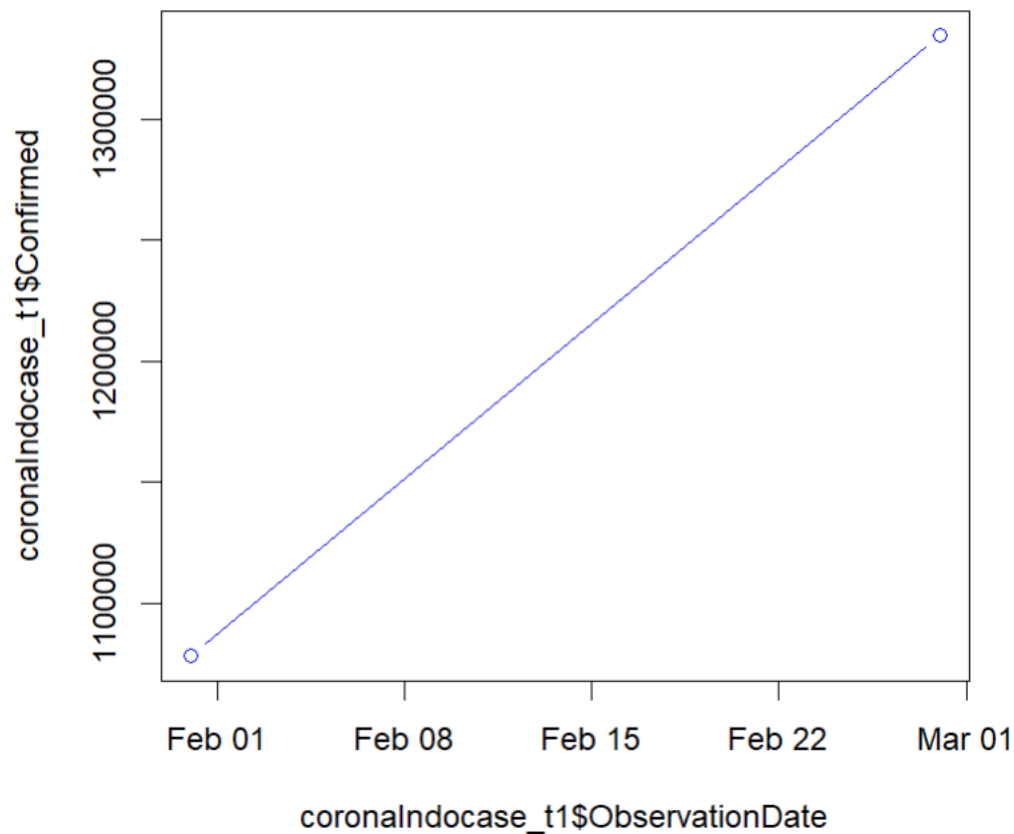
```
coronaIndo_t2<-coronaData[which(coronaData$ObservationDate>='2021-02-01' &  
coronaData$ObservationDate<='2021-02-28' & coronaData$Country.Region=='Indonesia' ),]
```

```
case_Indo_t2<-max(coronaIndo_t2$Confirmed)
```

```
coronaIndocase_t2<-coronaIndo_t2[which(coronaIndo_t2$Confirmed==case_Indo_t2),]
```

```
coronaIndocase_t1[nrow(coronaIndocase_t1) + 1,]=coronaIndocase_t2
```

```
plot(coronaIndocase_t1$ObservationDate,coronaIndocase_t1$Confirmed,type="b",col="blue")
```



#Philippines

```
coronaPhilip_t1<-coronaData[which(coronaData$ObservationDate>='2021-01-01' &
coronaData$ObservationDate<='2021-01-31' & coronaData$Country.Region=='Philippines' ),]
```

```
case_Philip_t1<-max(coronaPhilip_t1$Confirmed)
```

```
coronaPhilicasep_t1<-coronaPhilip_t1[which(coronaPhilip_t1$Confirmed==case_Philip_t1),]
```

```
coronaPhilip_t2<-coronaData[which(coronaData$ObservationDate>='2021-02-01' &
coronaData$ObservationDate<='2021-02-28' & coronaData$Country.Region=='Philippines' ),]
```

```
case_Philip_t2<-max(coronaPhilip_t2$Confirmed)
```

```
coronaPhilicasep_t2<-coronaPhilip_t2[which(coronaPhilip_t2$Confirmed==case_Philip_t2),]
```

```
coronaPhilicasep_t2[nrow(coronaPhilicasep_t2) + 1,]=coronaPhilicasep_t2
```

```
plot(coronaPhilip_t1$ObservationDate,coronaPhilip_t1$Confirmed,type="b",col="blue")
```