


통계패키지활용 자료분석(2024년 겨울학기)			
담당 교수 : 김 태 수			
강좌 번호	100961-31001	본인의 과제 자체 평가	9/10 점
과제명 : 제 3차 자료시각화 with plotly			

이름	이지원
	

제 출 일	2024년 01월 14일
학 과	산업공학과 ITM전공
학 번	22102009

## 목차

### I. plotly를 이용한 자료시각화 실습

- 1) 실습목표
- 2) 실습내용

### II. 자료분석

- 1) 자료설명
- 2) 자료 상세 확인하기

### III. 기후변화 데이터 시각화 작업

### IV. 결론

### V. 부록

## I. plotly를 이용한 자료시각화 실습

### 1) 실습목표

자신이 분석하고자 하는 데이터셋을 찾아 선택한 후, plotly 패키지를 활용하여 원하는 결론을 도출하기 위한 시각화 작업을 원활하게 수행할 수 있다.

### 2) 실습내용

코로나19 상황 당시 검사자, 확진자, 완치자, 사망자의 수를 담고 있는 데이터셋을 plotly 패키지를 활용해 시각화하고 결론을 도출한다.

## II. 자료분석

### 1) 자료설명

2020년 코로나19 확산 당시 대한민국 질병관리본부가 제공한 데이터를 수집 및 재가공한 데이터셋을 사용하였다. NeurIPS 2020에 이 데이터셋의 일부가 승인되었다. 다양한 데이터들 중 시간에 따른 데이터 추이를 알아보기 위해 Time.csv를 선택하여 분석하였다.

다음은 데이터 셋의 일부이다.

date	time	test	negative	confirmed	released	deceased
#####	16	1	0	1	0	0
#####	16	1	0	1	0	0
#####	16	4	3	1	0	0
#####	16	22	21	1	0	0
#####	16	27	25	2	0	0
#####	16	27	25	2	0	0
#####	16	51	47	3	0	0
#####	16	61	56	4	0	0
#####	16	116	97	4	0	0
#####	16	187	155	4	0	0
#####	16	246	199	6	0	0
#####	16	312	245	11	0	0
#####	16	371	289	12	0	0
#####	16	429	327	15	0	0
#####	16	490	414	15	0	0

## 2) 자료 상세 확인하기

### (1) 개요

```
> summary(df)
```

date		time
Min. :	2020-01-20	Min. : 0.000
1st Qu.:	2020-02-29	1st Qu.: 0.000
Median :	2020-04-10	Median : 0.000
Mean :	2020-04-10	Mean : 4.123
3rd Qu.:	2020-05-20	3rd Qu.:16.000
Max. :	2020-06-30	Max. :16.000

test	negative	confirmed
Min. : 1	Min. : 0	Min. : 1
1st Qu.: 96488	1st Qu.: 58774	1st Qu.: 3443
Median : 503051	Median : 477303	Median :10450
Mean : 497780	Mean : 475484	Mean : 7835
3rd Qu.: 782559	3rd Qu.: 754223	3rd Qu.:11116
Max. :1273766	Max. :1240157	Max. :12800

released	deceased	date_numeric
Min. : 0	Min. : 0.0	Min. : NA
1st Qu.: 29	1st Qu.: 17.5	1st Qu.: NA
Median : 7117	Median :208.0	Median : NA
Mean : 5604	Mean :157.1	Mean :NaN
3rd Qu.:10100	3rd Qu.:263.5	3rd Qu.: NA
Max. :11537	Max. :282.0	Max. : NA
		NA's :163

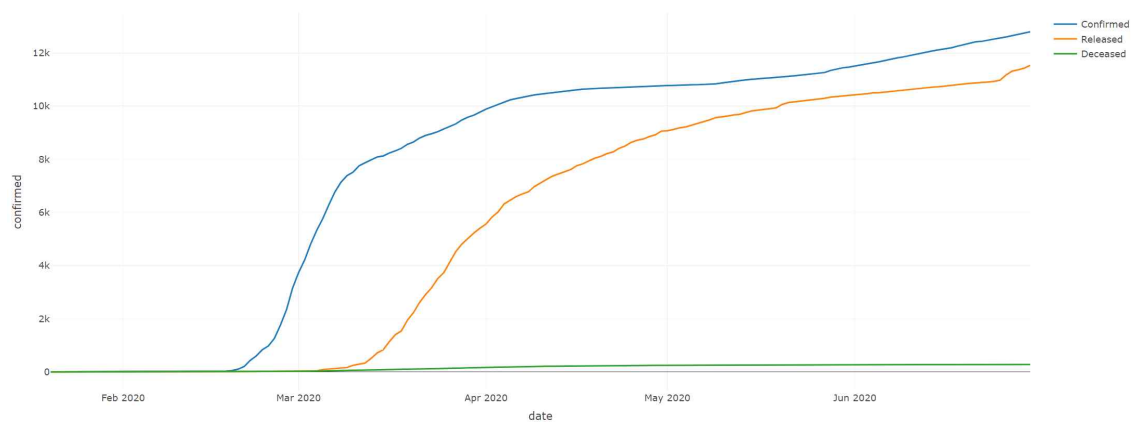
### (2) 데이터구조

```
> str(df)
```

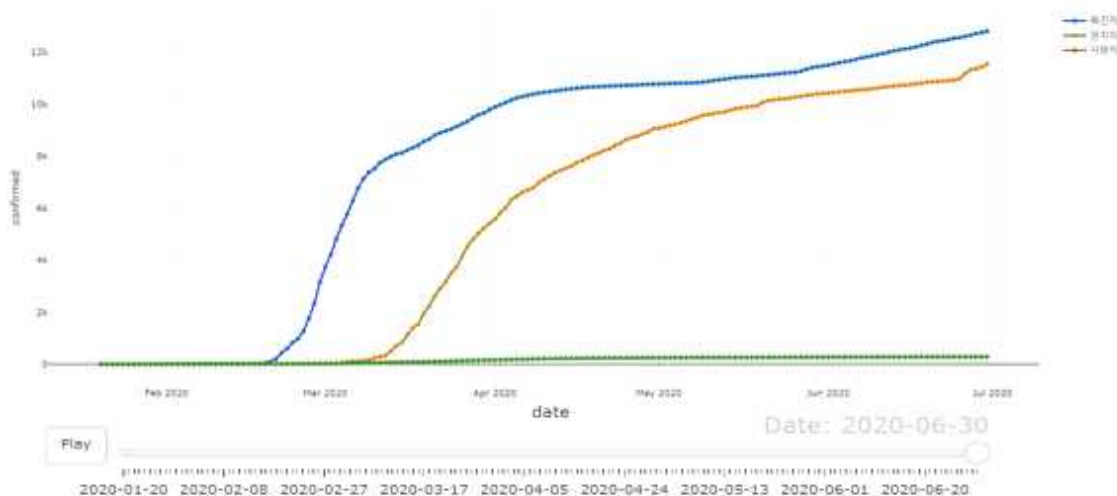
```
'data.frame': 163 obs. of 8 variables:
 $ date      : Date, format: "2020-01-20" "2020-01-21" "2020-01-22" "2020-01-23" ...
 $ time      : int  16 16 16 16 16 16 16 16 16 16 ...
 $ test      : int  1 1 4 22 27 27 51 61 116 187 ...
 $ negative   : int  0 0 3 21 25 25 47 56 97 155 ...
 $ confirmed  : int  1 1 1 1 2 2 3 4 4 4 ...
 $ released   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ deceased   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ date_numeric: num  NA NA NA NA NA NA NA NA NA NA ...
```

### III. 기후변화 데이터 시각화 작업

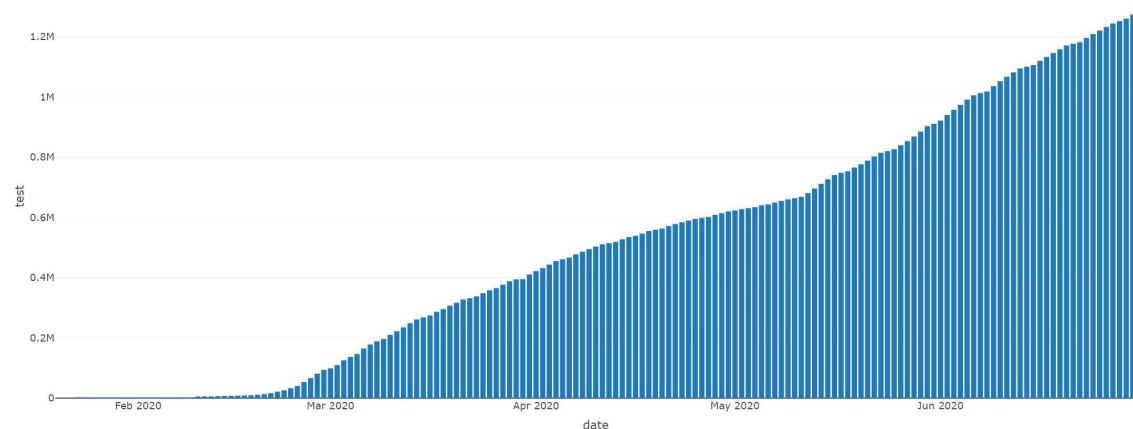
1) 선 그래프를 이용한 확진자, 완치자, 사망자 수의 시간에 따른 변화 추이 시각화



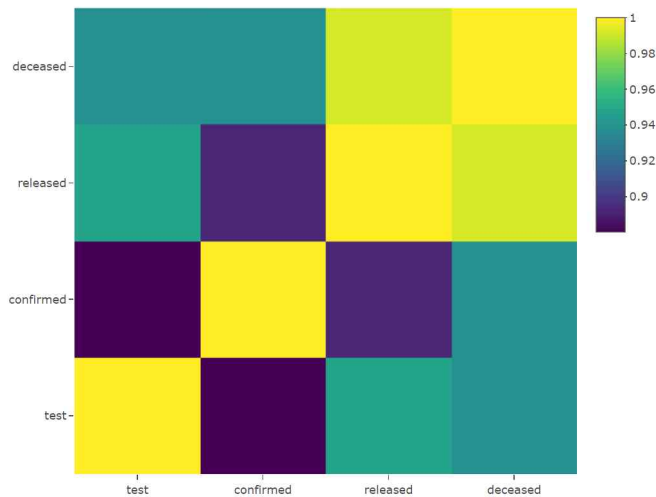
2) 1)에 animation 및 silde 적용



3) 막대 그래프를 이용해 일일 테스트 수의 변화 시각화



4) Heatmap을 이용한 날짜별 테스트, 확진자, 완치자, 사망자 수의 상관관계 시각화



#### IV. 결론

- 2020년 코로나19 확진자는 2월 말쯤 급격하게 증가했고 완치자는 그로부터 약 한 달 뒤인 3월 중순쯤 급격하게 증가하였다. 사망자 추이에서 급격한 변화는 없었다.
- 일일 테스트 수 는 2020년 상반기, 3월을 기점으로 급격히 증가하였다. 이는 코로나19가 본격적으로 유행하기 시작한 시점과 거의 일치한다.
- 사망자와 완치자 사이가 양의 상관관계가 가장 강하고 확진자와 테스트 수가 가장 약한 양의 상관관계를 보였다.

#### V. 부록: R스크립트

```
plot_ly(data) %>%
  + add_trace(x = ~date, y = ~confirmed, name =
'Confirmed', type = 'scatter', mode = 'lines') %>%
  + add_trace(x = ~date, y = ~released, name = 'Released',
type = 'scatter', mode = 'lines') %>%
  + add_trace(x = ~date, y = ~deceased, name = 'Deceased',
type = 'scatter', mode = 'lines')
data$date_str <- as.character(data$date)
```

```

p <- data %>% plot_ly()
p <- p %>% add_trace(x = ~date, y = ~confirmed, name = '확진
자', type = 'scatter', mode = 'lines', frame = ~date_str)
p <- p %>% add_trace(x = ~date, y = ~released, name = '완치자',
type = 'scatter', mode = 'lines', frame = ~date_str)
p <- p %>% add_trace(x = ~date, y = ~deceased, name = '사망자
', type = 'scatter', mode = 'lines', frame = ~date_str)

```

```

p <- p %>% animation_opts(1000) %>%
animation_slider(currentvalue = list(prefix = "Date: ", font =
list(size = 20)))
p

```

```

plot_ly(df, x = ~date, y = ~test, type = 'bar')fig <- plot_ly(data, x
= ~date, y = ~confirmed, name = 'Confirmed', type = 'scatter',
mode = 'lines') %>%
  add_trace(y = ~released, name = 'Released', mode = 'lines')
%>%
  add_trace(y = ~deceased, name = 'Deceased', mode = 'lines')
%>%
  layout(title = 'COVID-19 Cases Over Time',
    xaxis = list(title = 'Date'),
    yaxis = list(title = 'Number of Cases'),
    updatemenus = list(
      list(type = "buttons",
        x = 1.05,
        xanchor = "left",
        y = 0.8,
        yanchor = "bottom",
        buttons = list(

```

```

        list(label = "Play",
              method = "animate",
              args = list(list(NULL),
                           list(frame = list(duration =
100, redraw = TRUE), fromcurrent = TRUE))),
        list(label = "Pause",
              method = "animate",
              args = list(list(NULL),
                           list(frame = list(duration = 0,
redraw = TRUE), fromcurrent = TRUE)))
    )
  )
),
sliders = list(
  list(y = -0.2,
        len = 0.1,
        pad = list(b = 10, t = 10),
        steps = list(
          list(args = list("frame", list(duration = 100,
redraw = TRUE), fromcurrent = TRUE),
              label = 'Play',
              method = 'animate'),
          list(args = list("frame", list(duration = 0, redraw
= TRUE), fromcurrent = TRUE),
              label = 'Pause',
              method = 'animate')
        )
      )
)
)
plot_ly(df, x = ~date, y = ~test, type = 'bar')

```



```
correlation_matrix <- cor(data[c('test', 'confirmed', 'released',  
'deceased')])  
> plot_ly(x = colnames(correlation_matrix), y =  
rownames(correlation_matrix), z = correlation_matrix, type =  
"heatmap")
```