지수와 지도의 시각화

이번 장에서는 지금까지 살펴본 시각화외에도 plotly에서 추가적으로 제공하는 시각화 중 지수(index)의 시가화와 지도의 시각화를 살펴보도록 하겠다.

지수(Index)는 데이터 들을 특징일 종합하여 하나의 수치로 표현한 것을 말한다. 평균, 표준편차와 같은 수치도 가장 흔하게 만날 수 있는 지수 중에 하나이고 매일 매일 주식 시장 전체의 가격 변동을 종합한 KOPIS나 KOSDAQ와 같은 주가 종합 지수도 있다. 이렇게 하나의 수치를 시각화로 표현하는 방법을 plotly에서 제공한다.

지도(Map)의 시각화는 매우 흔하게 사용하는 시각화로 보통 행정 구역과 같은 공간 단위별로 수집된 데이터에 대해 무늬나 색상의 단계로 데이터를 표현하는 단계 구분도(Choropleth map)나 지도위에 위경도 좌표를 지닌 객체들을 표현하는 분산 맵(Scatter map) 등이 사용된다.

# 인디케이터(indicator)

인디케이터의 사전적 의미는 ‘가리키는 것’이다. 특정한 지시, 징후, 징표를 가리키는 것이라는 의미가 적절할 것이다. plotly에서 인디케이터의 의미는 ’value’ 속성으로 지정된 단일 값을 시각화하는 것이다. 인디케이터의 속성은 다음의 네 가지 속성이 첫번째 레벨의 속성이다.

* value: 시각화 할 값
* mode: 인디케이터의 종류
* align: 수치와 델터 인디케이터의 정렬(left, center, right)
* domain: 인디케이터의 위치 범위

plotly에서는 인디케이터를 시각화하기 위해 indicator 트레이스를 제공한다. indicator 트레이스는 인디케이터로 표현할 수치를 수치, 델타, 게이지의 세 가지 방법을 사용하여 시각화하는 것이다. 이 세가지 방법은 ‘mode’ 속성을 통해 설정이 가능한데, 다음은 indicator 트레이스에서 사용하는 주요 속성이다.



## 수치(number) 인디케이터

수치 인디케이터는 단일 수치값을 시각화하는 방법이다. 이 수치 인디케이터는 indicator 트레이스 속성 중에 ‘number’ 속성에 의해 설정이 가능하다. 다음은 수치 인디케이터의 설정에 사용하는 ’number’의 주요 속성이다.



수치 인디케이터는 전체 인디케이터에서 표현되는 부분 중에 수치값을 표현하는 부분으로 수치 인디케이터가 단독으로 사용되기 보다는 뒤에 설명할 ’delta’나 ’guage’와 함께 전체 인디케이터를 구성하는 일부분으로 사용되는 경우가 일반적이다. 하지만 단순한 현재 수치를 표현할 때는 단독으로 사용될 수 있다.

다음은 주요 5개국의 코로나 19 사망자수 선 그래프에 한국의 코로나 사망자수를 indicator 트레이스로 추가하는 R과 python 코드이다. 한국의 코로나 사망자수를 그리고 나서 그래프 영역의 특정 부분(X축의 0.4~0.6, Y축의 0,85~0.95)에 수치 속성만을 가지는 indicator 트레이스를 추가하였다. indicator 트레이스의 ‘value’를 설정하기 위해 앞서 코로나 사망자수를 그리는데 사용했던 total\_deaths\_5\_nations\_by\_day 데이터프레임에서 한국의 10만명당 사망자수 데이터를 산출하고 이 데이터를 사용하여 indicator 트레이스를 생성하였다. indicator 트레이스의 위치를 설정하는데 ’domain’ 속성을 사용한다.

* R

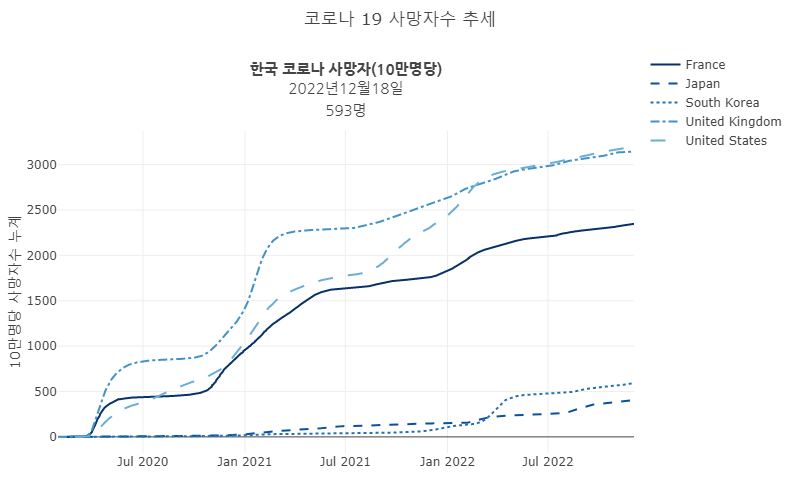
R에서 수치 인디케이터를 만들기 위해서는 indicator 트레이스를 만드는데, add\_trace(type = 'indicator', ...)를 사용한다. 수치 인디케이터는 ‘mode’ 속성을 “number”로 설정하고 ‘number’ 속성을 설정하여 만들 수 있다.

number\_KOR <- total\_deaths\_5\_nations\_by\_day |>  
 filter(date == max(date), iso\_code == 'KOR') |>  
 select(total\_deaths\_per\_million) |> pull()  
  
fig1 <- total\_deaths\_5\_nations\_by\_day |>  
 plot\_ly() |>  
 ## scatter 트레이스 생성  
 add\_trace(type = 'scatter', mode = 'lines',   
 x = ~date, y = ~total\_deaths\_per\_million ,   
 linetype = ~location, connectgaps = T) |>  
 ## layout의 제목, 축제목, 여백 속성 설정  
 layout(title = list(text = '코로나 19 사망자수 추세', pad = list(b = 5)),   
 xaxis = list(title = ''),   
 yaxis = list(title = '10만명당 사망자수 누계', domain = c(0, 0.8)),   
 margin = margins\_R)  
  
fig1 |>  
 add\_trace(type = 'indicator', mode = 'number',   
 value = number\_KOR,   
 title = list(text = paste0('<b>한국 코로나 사망자(10만명당)</b>\n', year(today()), '년', month(today()), '월', day(today()), '일'),  
 font = list(family = '나눔고딕', size = 15)  
 ),  
 number = list(font = list(family = '나눔고딕',   
 size = 15),   
 suffix = '명'),  
 domain = list(x = c(0.4, 0.6), y = c(0.8, 0.9)))

* python

python에서 수치 인디케이터를 만들기 위해서는 indicator 트레이스를 사용하고, add\_trace()에 plotly.graph\_objects.Indicator()를 사용한다. 수치 인디케이터는 ‘mode’ 속성을 “number”로 설정하고 ‘number’ 속성을 설정하여 만들 수 있다.

from datetime import datetime, timedelta  
  
today = datetime.today()  
  
number\_KOR = total\_deaths\_5\_nations\_by\_day.loc[(total\_deaths\_5\_nations\_by\_day['date']==total\_deaths\_5\_nations\_by\_day['date'].max()) & (total\_deaths\_5\_nations\_by\_day['iso\_code']=='KOR')]  
number\_KOR = number\_KOR.loc[:,'total\_deaths\_per\_million'].values[0]  
  
fig = go.Figure()  
for location, group in total\_deaths\_5\_nations\_by\_day.groupby('location'):  
 fig.add\_trace(go.Scatter(  
 mode = 'lines',   
 x = group['date'],   
 y = group['total\_deaths\_per\_million'],   
 line = dict(dash = nations[location]),   
 name = location,  
 connectgaps = True  
 ))  
  
fig.update\_layout(title = dict(text = '코로나 19 사망자수 추세', x = 0.5),   
 xaxis = dict(title = ''),   
 yaxis = dict(title = '10만명당 사망자수 누계', domain = (0, 0.8)))   
  
fig.add\_trace(go.Indicator(  
 mode = 'number', value = number\_KOR,  
 title = dict(text = '<b>한국 코로나 사망자(10만명당)</b><br>' + str(today.year) + '년' + str(today.month) + '월' + str(today.day) + '일\n',  
 font = dict(family = '나눔고딕', size = 15)),  
 number = dict(font = dict(family = '나눔고딕', size = 15),  
 suffix = '명'),  
 domain = dict(x = (0.4, 0.6), y = (0.8, 0.9))  
))  
  
fig.show()



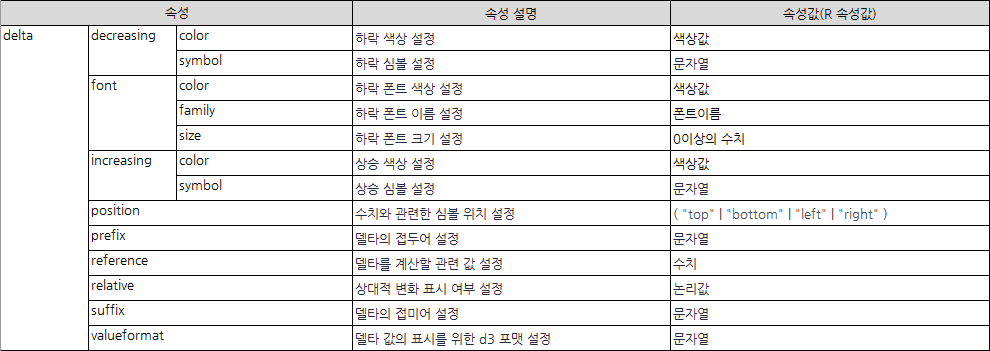
실행결과 VI-1. 수치 인디케이터가 포함된 R의 선 그래프

## 델타(delta) 인디케이터

델타 인디케이터는 일반적으로 수치 인디케이터와 병행하여 사용하여 수치 인디케이터에 표현된 값이 비교 기준값과의 차이를 기호와 증감값을 사용하여 보여주는 인디케이터를 말한다. 델타 인디케이터에는 수치 인디케이터에 비해 다음과 같은 몇 가지 요소들이 필요하다.

* reference : 값의 증감을 비교할 기준 값
* relative : 값의 증감이 절대적인지, 상대적인지를 결정
* increasing : 값이 증가할 경우 표시할 속성
* decreasing : 값이 감소할 경우 표시할 속성

다음은 델타 인디케이터의 설정에 사용하는 ’delta’의 주요 속성이다.



* R

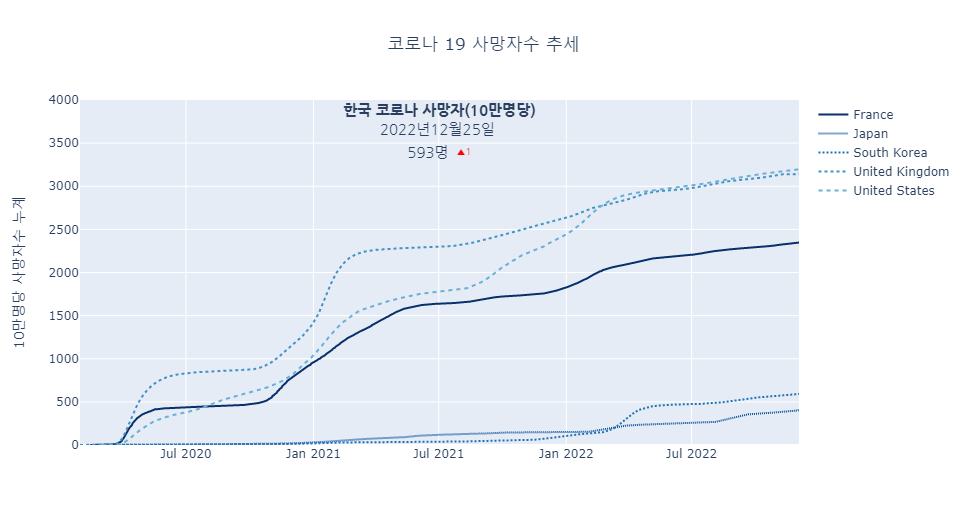
R에서 델타 인디케이터를 만들기 위해서는 indicator 트레이스를 사용하는데, add\_trace(type = 'indicator', ...)를 사용한다. 수치 인디케이터는 ‘mode’ 속성을 “delta”로 설정하고 ‘delta’ 속성을 설정하여 만들 수 있다.

number1\_KOR <- total\_deaths\_5\_nations\_by\_day |>  
 filter(date == max(date)-1, iso\_code == 'KOR') |>  
 select(total\_deaths\_per\_million) |> pull()  
  
fig1 <- total\_deaths\_5\_nations\_by\_day |>  
 plot\_ly() |>  
 ## scatter 트레이스 생성  
 add\_trace(type = 'scatter', mode = 'lines',   
 x = ~date, y = ~total\_deaths\_per\_million ,   
 linetype = ~location, connectgaps = T) |>  
 ## layout의 제목, 축제목, 여백 속성 설정  
 layout(title = list(text = '코로나 19 사망자수 추세', pad = list(b = 5)),   
 xaxis = list(title = ''),   
 yaxis = list(title = '10만명당 사망자수 누계', domain = c(0, 0.8)),   
 margin = margins\_R)  
  
fig1 |>  
 add\_trace(type = 'indicator', mode = 'number+delta',   
 value = number\_KOR,   
 title = list(text = paste0('<b>한국 코로나 사망자(10만명당)</b>\n', year(today()), '년', month(today()), '월', day(today()), '일'),  
 font = list(family = '나눔고딕', size = 15)  
 ),  
 number = list(font = list(family = '나눔고딕',   
 size = 15),   
 suffix = '명'),  
 delta = list(reference = number1\_KOR, position = 'right',   
 increasing = list(color = 'red'),  
 decreasing = list(color = 'blue'),  
 font = list(family = '나눔고딕', size = 10)),  
 domain = list(x = c(0.4, 0.6), y = c(0.8, 0.9)))

* python

python에서 델타 인디케이터를 만들기 위해서는 indicator 트레이스를 사용하고, add\_trace()에 plotly.graph\_objects.Indicator()를 사용한다. 델타 인디케이터는 ‘mode’ 속성을 “delta”로 설정하고 ‘delta’ 속성을 설정하여 만들 수 있다.

number1\_KOR = total\_deaths\_5\_nations\_by\_day.loc[(total\_deaths\_5\_nations\_by\_day['date']==total\_deaths\_5\_nations\_by\_day['date'].max()- timedelta(days = 1)) & (total\_deaths\_5\_nations\_by\_day['iso\_code']=='KOR')]  
number1\_KOR = number1\_KOR.loc[:,'total\_deaths\_per\_million'].values[0]  
  
fig = go.Figure()  
for location, group in total\_deaths\_5\_nations\_by\_day.groupby('location'):  
 fig.add\_trace(go.Scatter(  
 mode = 'lines',   
 x = group['date'],   
 y = group['total\_deaths\_per\_million'],   
 line = dict(dash = nations[location]),   
 name = location,  
 connectgaps = True  
 ))  
  
fig.update\_layout(title = dict(text = '코로나 19 사망자수 추세', x = 0.5),   
 xaxis = dict(title = ''),   
 yaxis = dict(title = '10만명당 사망자수 누계', domain = (0, 0.8)))   
  
fig.add\_trace(go.Indicator(  
 mode = 'number+delta', value = number\_KOR,  
 title = dict(text = '<b>한국 코로나 사망자(10만명당)</b><br>' + str(today.year) + '년' + str(today.month) + '월' + str(today.day) + '일\n',  
 font = dict(family = '나눔고딕', size = 15)),  
 number = dict(font = dict(family = '나눔고딕', size = 15),  
 suffix = '명'),  
 delta = dict(reference = number1\_KOR, position = 'right',   
 increasing = dict(color = 'red'),  
 decreasing = dict(color = 'blue'),  
 font = dict(family = '나눔고딕', size = 10)),  
 domain = dict(x = (0.4, 0.6), y = (0.8, 0.9))  
))  
  
fig.show()

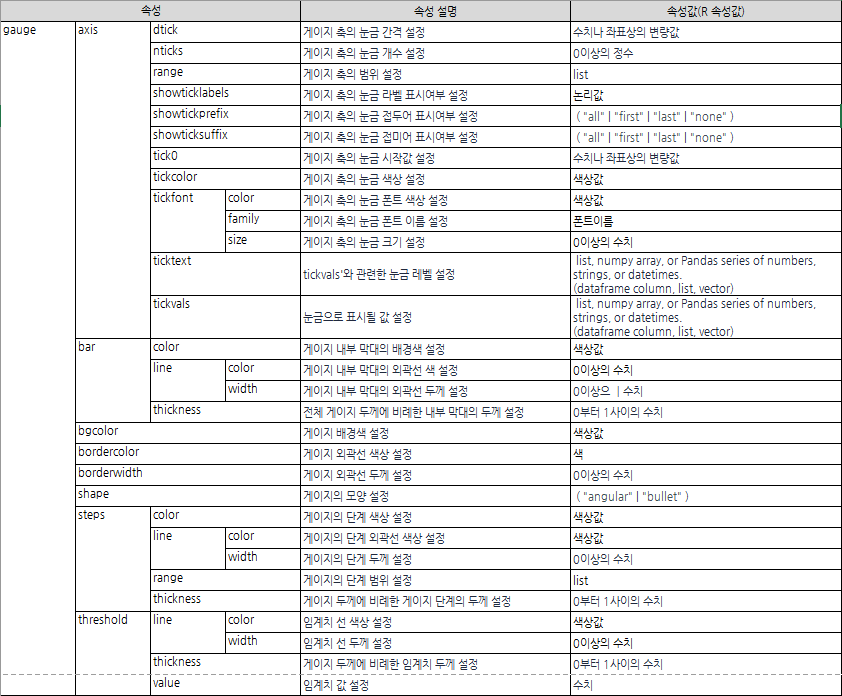


실행결과 VI-2. 델타 인디케이터가 표시된 python 선 그래프

## 게이지(gauge) 인디케이터

게이지 인디케이터는 둥근 반원의 형태의 틀에 둥글게 그려지는 막대의 각도에 의해 최소값과 최대값 사이에서의 현재의 지수를 표현할 때 사용하는 시각화이다. 이 시각화를 구성하는데에는 주 표현값인 ‘value’ 값 이외에 ‘steps’, ‘threshold’, ‘delta’의 속성을 사용한다. 게이지 인디케이터는 indicator 트레이스에 ’mode’ 속성에 “gauge”를 포함하여 만들 수 있다. “gauge”만을 사용하거나, “number+gauge”를 사용하여 수치와 함꼐 게이지를 사용하거나, “number+gauge+delta”를 사용하여 수치, 게이지, 증감 기호를 모두 사용할 수도 있다.

다음은 게이지 인디케이터의 설정에 사용하는 ’gauge’의 주요 속성이다.



* R

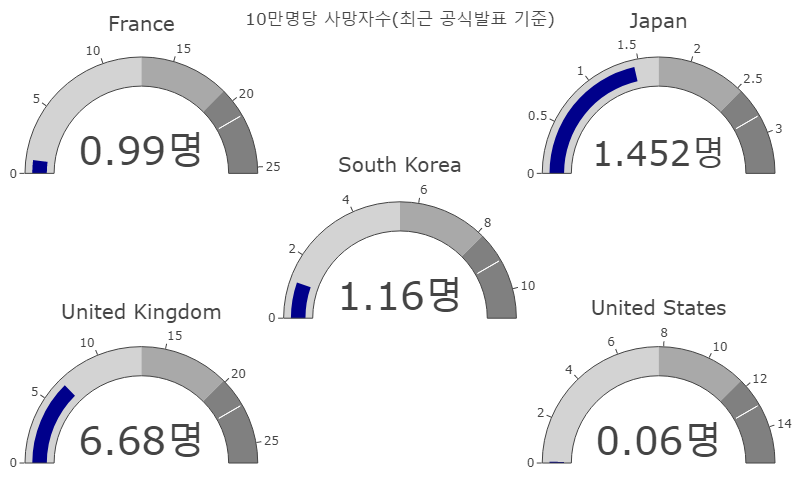
R에서 게이지 인디케이터를 만들기 위해서는 indicator 트레이스를 사용하는데, add\_trace(type = 'indicator', ...)를 사용한다. 수치 인디케이터는 ‘mode’ 속성을 “gauge”로 설정하고 ‘gauge’ 속성을 설정하여 만들 수 있다.

max\_deaths\_per\_million\_by\_day <- total\_deaths\_5\_nations\_by\_day |> group\_by(location) |>  
 summarise(최대사망자 = max(new\_deaths\_per\_million, na.rm = TRUE))  
  
deaths\_per\_million\_in\_lateast <- total\_deaths\_5\_nations\_by\_day |> group\_by(location) |>  
 filter(is.na(new\_deaths\_per\_million) == FALSE) |>  
 filter(date == max(date)) |>  
 select(iso\_code, date, new\_deaths\_per\_million)  
  
  
df\_gauge <- left\_join(max\_deaths\_per\_million\_by\_day, deaths\_per\_million\_in\_lateast, by = 'location') |> arrange(location)  
  
fig\_gauge <- df\_gauge |> plot\_ly() |>  
 add\_trace(type = 'indicator', mode = "gauge+number", title = pull(df\_gauge[3, 1]),  
 domain = list(row = 1, column = 1), value = pull(df\_gauge[3, 5]),   
 gauge = list(axis = list(  
 range = list(NULL, pull(df\_gauge[3, 2])\*1.2)),  
 steps = list(  
 list(range = c(0, pull(df\_gauge[3, 2])\*1.2\*0.5), color = "lightgray"),  
 list(range = c(pull(df\_gauge[3, 2])\*1.2\*0.5, pull(df\_gauge[3, 2])\*1.2\*0.75), color = "darkgray"),  
 list(range = c(pull(df\_gauge[3, 2])\*1.2\*0.75, pull(df\_gauge[3, 2])\*1.2), color = "gray")),  
 threshold = list(line = list(color = 'white'),  
 value = pull(df\_gauge[3, 2])),   
 bar = list(color = "darkblue")),   
 number = list(suffix = '명'))  
  
fig\_gauge <- fig\_gauge |>  
 add\_trace(type = 'indicator', mode = "gauge+number", title = pull(df\_gauge[1, 1]),  
 domain = list(row = 0, column = 0), value = pull(df\_gauge[1, 5]),   
 gauge = list(axis = list(  
 range = list(NULL, pull(df\_gauge[1, 2])\*1.2)),  
 steps = list(  
 list(range = c(0, pull(df\_gauge[1, 2])\*1.2\*0.5), color = "lightgray"),  
 list(range = c(pull(df\_gauge[1, 2])\*1.2\*0.5, pull(df\_gauge[1, 2])\*1.2\*0.75), color = "darkgray"),  
 list(range = c(pull(df\_gauge[1, 2])\*1.2\*0.75, pull(df\_gauge[1, 2])\*1.2), color = "gray")),  
 threshold = list(line = list(color = 'white'),  
 value = pull(df\_gauge[1, 2])),   
 bar = list(color = "darkblue")),   
 number = list(suffix = '명'))  
  
fig\_gauge <- fig\_gauge |>  
 add\_trace(type = 'indicator', mode = "gauge+number", title = pull(df\_gauge[2, 1]),  
 domain = list(row = 0, column = 2), value = pull(df\_gauge[2, 5]),   
 gauge = list(axis = list(  
 range = list(NULL, pull(df\_gauge[2, 2])\*1.2)),  
 steps = list(  
 list(range = c(0, pull(df\_gauge[2, 2])\*1.2\*0.5), color = "lightgray"),  
 list(range = c(pull(df\_gauge[2, 2])\*1.2\*0.5, pull(df\_gauge[2, 2])\*1.2\*0.75), color = "darkgray"),  
 list(range = c(pull(df\_gauge[2, 2])\*1.2\*0.75, pull(df\_gauge[2, 2])\*1.2), color = "gray")),  
 threshold = list(line = list(color = 'white'),  
 value = pull(df\_gauge[2, 2])),   
 bar = list(color = "darkblue")),   
 number = list(suffix = '명'))  
  
fig\_gauge <- fig\_gauge |>  
 add\_trace(type = 'indicator', mode = "gauge+number", title = pull(df\_gauge[4, 1]),  
 domain = list(row = 2, column = 0), value = pull(df\_gauge[4, 5]),   
 gauge = list(axis = list(  
 range = list(NULL, pull(df\_gauge[4, 2])\*1.2)),  
 steps = list(  
 list(range = c(0, pull(df\_gauge[4, 2])\*1.2\*0.5), color = "lightgray"),  
 list(range = c(pull(df\_gauge[4, 2])\*1.2\*0.5, pull(df\_gauge[4, 2])\*1.2\*0.75), color = "darkgray"),  
 list(range = c(pull(df\_gauge[4, 2])\*1.2\*0.75, pull(df\_gauge[4, 2])\*1.2), color = "gray")),  
 threshold = list(line = list(color = 'white'),  
 value = pull(df\_gauge[4, 2])),   
 bar = list(color = "darkblue")),   
 number = list(suffix = '명'))  
  
fig\_gauge <- fig\_gauge |>  
 add\_trace(type = 'indicator', mode = "gauge+number", title = pull(df\_gauge[5, 1]),  
 domain = list(row = 2, column = 2), value = pull(df\_gauge[5, 5]),   
 gauge = list(axis = list(  
 range = list(NULL, pull(df\_gauge[5, 2])\*1.2)),  
 steps = list(  
 list(range = c(0, pull(df\_gauge[5, 2])\*1.2\*0.5), color = "lightgray"),  
 list(range = c(pull(df\_gauge[5, 2])\*1.2\*0.5, pull(df\_gauge[5, 2])\*1.2\*0.75), color = "darkgray"),  
 list(range = c(pull(df\_gauge[5, 2])\*1.2\*0.75, pull(df\_gauge[5, 2])\*1.2), color = "gray")),  
 threshold = list(line = list(color = 'white'),  
 value = pull(df\_gauge[5, 2])),   
 bar = list(color = "darkblue")),   
 number = list(suffix = '명'))  
  
fig\_gauge |> layout(grid=list(rows=3, columns=3),   
 margin = margins\_R,   
 title = '10만명당 사망자수(최근 공식발표 기준)')

* python

python에서 게이지 인디케이터를 만들기 위해서는 indicator 트레이스를 사용하고, add\_trace()에 plotly.graph\_objects.Indicator()를 사용한다. 게이지 인디케이터는 ‘mode’ 속성을 “gauge”로 설정하고 ‘gauge’ 속성을 설정하여 만들 수 있다.

deaths\_per\_million\_in\_lateast = total\_deaths\_5\_nations\_by\_day[total\_deaths\_5\_nations\_by\_day['new\_deaths\_per\_million'].isna() == False]  
deaths\_per\_million\_in\_lateast = pd.merge(deaths\_per\_million\_in\_lateast.groupby('location')['date'].max(), deaths\_per\_million\_in\_lateast, on = ("location", 'date'))[['iso\_code', 'location', 'date', 'new\_deaths\_per\_million']]  
df\_gauge = pd.merge(deaths\_per\_million\_in\_lateast, total\_deaths\_5\_nations\_by\_day.groupby('location')['new\_deaths\_per\_million'].max().reset\_index(), on = 'location').sort\_values('location')  
df\_gauge.columns = ('iso\_code', 'location', 'date', '최근사망자', '최대사망자')  
  
fig = go.Figure()  
fig.add\_trace(go.Indicator(  
 type = 'indicator', mode = "gauge+number", title = df\_gauge.iloc[2, 1],  
 domain = dict(row = 1, column = 1), value = df\_gauge.iloc[2, 3],   
 gauge = dict(axis = dict(  
 range = (0, df\_gauge.iloc[2, 4]\*1.2)),  
 steps = [  
 dict(range = (0, (df\_gauge.iloc[2, 4])\*1.2\*0.5), color = "lightgray"),  
 dict(range = ((df\_gauge.iloc[2, 4])\*1.2\*0.5, (df\_gauge.iloc[2, 4])\*1.2\*0.75), color = "darkgray"),   
 dict(range = ((df\_gauge.iloc[2, 4])\*1.2\*0.75, (df\_gauge.iloc[2, 4])\*1.2), color = "gray")],  
 threshold = dict(  
 line = dict(color = 'white'),  
 value = df\_gauge.iloc[2, 4]),   
 bar = dict(color = "darkblue")),   
 number = dict(suffix = '명')  
))  
  
fig.add\_trace(go.Indicator(  
 type = 'indicator', mode = "gauge+number", title = df\_gauge.iloc[0, 1],  
 domain = dict(row = 0, column = 0), value = df\_gauge.iloc[0, 3],   
 gauge = dict(axis = dict(  
 range = (0, df\_gauge.iloc[0, 4]\*1.2)),  
 steps = [  
 dict(range = (0, (df\_gauge.iloc[0, 4])\*1.2\*0.5), color = "lightgray"),  
 dict(range = ((df\_gauge.iloc[0, 4])\*1.2\*0.5, (df\_gauge.iloc[0, 4])\*1.2\*0.75), color = "darkgray"),  
 dict(range = ((df\_gauge.iloc[0, 4])\*1.2\*0.75, (df\_gauge.iloc[0, 4])\*1.2), color = "gray")],  
 threshold = dict(  
 line = dict(color = 'white'),  
 value = df\_gauge.iloc[0, 4]),   
 bar = dict(color = "darkblue")),   
 number = dict(suffix = '명')))  
  
fig.add\_trace(go.Indicator(  
 type = 'indicator', mode = "gauge+number", title = df\_gauge.iloc[1, 1],  
 domain = dict(row = 0, column = 2), value = df\_gauge.iloc[1, 3],   
 gauge = dict(axis = dict(  
 range = (0, df\_gauge.iloc[1, 4]\*1.2)),  
 steps = [  
 dict(range = (0, (df\_gauge.iloc[1, 4])\*1.2\*0.5), color = "lightgray"),  
 dict(range = ((df\_gauge.iloc[1, 4])\*1.2\*0.5, (df\_gauge.iloc[1, 4])\*1.2\*0.75), color = "darkgray"),  
 dict(range = ((df\_gauge.iloc[1, 4])\*1.2\*0.75, (df\_gauge.iloc[1, 4])\*1.2), color = "gray")],  
 threshold = dict(  
 line = dict(color = 'white'),  
 value = df\_gauge.iloc[1, 4]),   
 bar = dict(color = "darkblue")),   
 number = dict(suffix = '명')))  
  
fig.add\_trace(go.Indicator(  
 type = 'indicator', mode = "gauge+number", title = df\_gauge.iloc[3, 1],  
 domain = dict(row = 2, column = 0), value = df\_gauge.iloc[3, 3],   
 gauge = dict(axis = dict(  
 range = (0, df\_gauge.iloc[3, 4]\*1.2)),  
 steps = [  
 dict(range = (0, (df\_gauge.iloc[3, 4])\*1.2\*0.5), color = "lightgray"),  
 dict(range = ((df\_gauge.iloc[3, 4])\*1.2\*0.5, (df\_gauge.iloc[3, 4])\*1.2\*0.75), color = "darkgray"),  
 dict(range = ((df\_gauge.iloc[3, 4])\*1.2\*0.75, (df\_gauge.iloc[3, 4])\*1.2), color = "gray")],  
 threshold = dict(  
 line = dict(color = 'white'),  
 value = df\_gauge.iloc[3, 4]),   
 bar = dict(color = "darkblue")),   
 number = dict(suffix = '명')))  
  
fig.add\_trace(go.Indicator(  
 type = 'indicator', mode = "gauge+number", title = df\_gauge.iloc[4, 1],  
 domain = dict(row = 2, column = 2), value = df\_gauge.iloc[4, 3],   
 gauge = dict(axis = dict(  
 range = (0, df\_gauge.iloc[4, 4]\*1.2)),  
 steps = [  
 dict(range = (0, (df\_gauge.iloc[4, 4])\*1.2\*0.5), color = "lightgray"),  
 dict(range = ((df\_gauge.iloc[4, 4])\*1.2\*0.5, (df\_gauge.iloc[4, 4])\*1.2\*0.75), color = "darkgray"),  
 dict(range = ((df\_gauge.iloc[4, 4])\*1.2\*0.75, (df\_gauge.iloc[4, 4])\*1.2), color = "gray")],  
 threshold = dict(  
 line = dict(color = 'white'),  
 value = df\_gauge.iloc[4, 4]),   
 bar = dict(color = "darkblue")),   
 number = dict(suffix = '명')))  
  
fig.update\_layout(grid=dict(rows=3, columns=3),  
 title = dict(text = '10만명당 사망자수(최근 공식발표 기준)', x = 0.5))  
  
fig.show()



실행결과 VI-3. R의 게이지 인디테이터

## 불릿(bullet) 인디케이터

불릿 인디케이터는 막대 그래프를 변형하여 특정 값의 현재 상태를 표시하는 시각화의 방법이다. 온도계나 진행률의 표시에 사용되는 방식을 차용하여 만드는 불릿 인디케이터는 둥글게 표시되는 게이지 인디케이터의 대용으로 대시보드에서 많이 사용된다. 불릿 인디케이터는 단계를 나타내는 질적 단계(step)와 양적 막대(value), 임계치의 표시의 세 가지 요소로 구성한다. 단계는 일반적으로 배열의 형태로 단계 값들을 지정하여 정의하고, 막대는 표시할 수치의 현재 값을 나타내고, 임계값은 수치의 목표치나 한계치에 대한 목표 지점을 표시함으로서 나타낸다.

불릿 인디케이터는 indicator 트레이스를 사용하여 만들수 있고 게이지 인디케이터의 속성 중 ’shape’를 “bullet”으로 설정함으로써 만들 수 있다. 또 게이지 인디케이터의 속성들을 공유한다.

* R

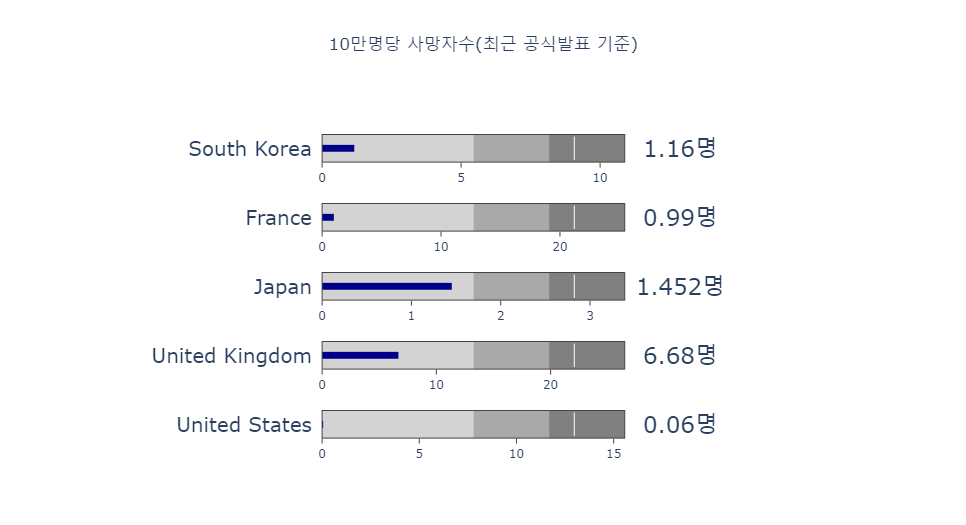
R에서 불릿 인디케이터를 만들기 위해서는 indicator 트레이스를 사용하는데, add\_trace(type = 'indicator', ...)를 사용한다. 불릿 인디케이터는 ‘mode’ 속성을 “gauge”로, ‘gauge’의 ’shape’ 속성을 “bullet”으로 설정하고 ‘gauge’ 속성을 설정하여 만들 수 있다.

fig\_gauge <- df\_gauge |> plot\_ly() |>  
 add\_trace(type = 'indicator', mode = "gauge+number", title = pull(df\_gauge[3, 1]),  
 domain = list(x = c(0.3,0.8), y = c(0.82, 0.9)),  
 value = pull(df\_gauge[3, 5]),   
 gauge = list(axis = list(  
 range = list(NULL, pull(df\_gauge[3, 2])\*1.2)),  
 steps = list(  
 list(range = c(0, pull(df\_gauge[3, 2])\*1.2\*0.5), color = "lightgray"),  
 list(range = c(pull(df\_gauge[3, 2])\*1.2\*0.5, pull(df\_gauge[3, 2])\*1.2\*0.75), color = "darkgray"),  
 list(range = c(pull(df\_gauge[3, 2])\*1.2\*0.75, pull(df\_gauge[3, 2])\*1.2), color = "gray")),  
 shape = "bullet",  
 threshold = list(  
 line = list(color = 'white'),  
 value = pull(df\_gauge[3, 2])),   
 bar = list(color = "darkblue")),   
 number = list(suffix = '명'))  
  
fig\_gauge <- fig\_gauge |>  
 add\_trace(type = 'indicator', mode = "gauge+number", title = pull(df\_gauge[1, 1]),  
 domain = list(x = c(0.3,0.8), y = c(0.62, 0.7)),  
 value = pull(df\_gauge[1, 5]),   
 gauge = list(axis = list(  
 shape = "bullet",  
 range = list(NULL, pull(df\_gauge[1, 2])\*1.2)),  
 steps = list(  
 list(range = c(0, pull(df\_gauge[1, 2])\*1.2\*0.5), color = "lightgray"),  
 list(range = c(pull(df\_gauge[1, 2])\*1.2\*0.5, pull(df\_gauge[1, 2])\*1.2\*0.75), color = "darkgray"),  
 list(range = c(pull(df\_gauge[1, 2])\*1.2\*0.75, pull(df\_gauge[1, 2])\*1.2), color = "gray")),  
 shape = "bullet",  
 threshold = list(  
 line = list(color = 'white'),  
 value = pull(df\_gauge[1, 2])),   
 bar = list(color = "darkblue")),   
 number = list(suffix = '명'))  
  
fig\_gauge <- fig\_gauge |>  
 add\_trace(type = 'indicator', mode = "gauge+number", title = pull(df\_gauge[2, 1]),  
 domain = list(x = c(0.3,0.8), y = c(0.42, 0.5)),  
 value = pull(df\_gauge[2, 5]),   
 gauge = list(axis = list(  
 shape = "bullet",  
 range = list(NULL, pull(df\_gauge[2, 2])\*1.2)),  
 steps = list(  
 list(range = c(0, pull(df\_gauge[2, 2])\*1.2\*0.5), color = "lightgray"),  
 list(range = c(pull(df\_gauge[2, 2])\*1.2\*0.5, pull(df\_gauge[2, 2])\*1.2\*0.75), color = "darkgray"),  
 list(range = c(pull(df\_gauge[2, 2])\*1.2\*0.75, pull(df\_gauge[2, 2])\*1.2), color = "gray")),  
 shape = "bullet",  
 threshold = list(  
 line = list(color = 'white'),  
 value = pull(df\_gauge[2, 2])),   
 bar = list(color = "darkblue")),   
 number = list(suffix = '명'))  
  
fig\_gauge <- fig\_gauge |>  
 add\_trace(type = 'indicator', mode = "gauge+number", title = pull(df\_gauge[4, 1]),  
 domain = list(x = c(0.3,0.8), y = c(0.22, 0.3)),  
 value = pull(df\_gauge[4, 5]),   
 gauge = list(axis = list(  
 shape = "bullet",  
 range = list(NULL, pull(df\_gauge[4, 2])\*1.2)),  
 steps = list(  
 list(range = c(0, pull(df\_gauge[4, 2])\*1.2\*0.5), color = "lightgray"),  
 list(range = c(pull(df\_gauge[4, 2])\*1.2\*0.5, pull(df\_gauge[4, 2])\*1.2\*0.75), color = "darkgray"),  
 list(range = c(pull(df\_gauge[4, 2])\*1.2\*0.75, pull(df\_gauge[4, 2])\*1.2), color = "gray")),  
 shape = "bullet",  
 threshold = list(  
 line = list(color = 'white'),  
 value = pull(df\_gauge[4, 2])),   
 bar = list(color = "darkblue")),   
 number = list(suffix = '명'))  
  
fig\_gauge <- fig\_gauge |>  
 add\_trace(type = 'indicator', mode = "gauge+number", title = pull(df\_gauge[5, 1]),  
 domain = list(x = c(0.3,0.8), y = c(0.02, 0.1)),  
 value = pull(df\_gauge[5, 5]),   
 gauge = list(axis = list(  
 shape = "bullet",  
 range = list(NULL, pull(df\_gauge[5, 2])\*1.2)),  
 steps = list(  
 list(range = c(0, pull(df\_gauge[5, 2])\*1.2\*0.5), color = "lightgray"),  
 list(range = c(pull(df\_gauge[5, 2])\*1.2\*0.5, pull(df\_gauge[5, 2])\*1.2\*0.75), color = "darkgray"),  
 list(range = c(pull(df\_gauge[5, 2])\*1.2\*0.75, pull(df\_gauge[5, 2])\*1.2), color = "gray")),  
 shape = "bullet",  
 threshold = list(  
 line = list(color = 'white'),  
 value = pull(df\_gauge[5, 2])),   
 bar = list(color = "darkblue")),   
 number = list(suffix = '명'))  
  
fig\_gauge |> layout(margin = margins\_R,   
 title = '10만명당 사망자수(최근 공식발표 기준)')

* python

python에서 불릿 인디케이터를 만들기 위해서는 indicator 트레이스를 사용하고, add\_trace()에 plotly.graph\_objects.Indicator()를 사용한다. 불릿 인디케이터는 ‘mode’ 속성을 “gauge”로, ‘gauge’의 ’shape’ 속성을 “bullet”로 설정하고 ‘gauge’ 속성을 설정하여 만들 수 있다.

fig = go.Figure()  
fig.add\_trace(go.Indicator(  
 type = 'indicator', mode = "gauge+number", title = df\_gauge.iloc[2, 1],  
 domain = dict(x = (0.3,0.8), y = (0.82, 0.9)),  
 value = df\_gauge.iloc[2, 3],   
 gauge = dict(axis = dict(  
 range = (0, df\_gauge.iloc[2, 4]\*1.2)),  
 shape = "bullet",  
 steps = [  
 dict(range = (0, (df\_gauge.iloc[2, 4])\*1.2\*0.5), color = "lightgray"),  
 dict(range = ((df\_gauge.iloc[2, 4])\*1.2\*0.5, (df\_gauge.iloc[2, 4])\*1.2\*0.75), color = "darkgray"),  
 dict(range = ((df\_gauge.iloc[2, 4])\*1.2\*0.75, (df\_gauge.iloc[2, 4])\*1.2), color = "gray")],  
 threshold = dict(  
 line = dict(color = 'white'),  
 value = df\_gauge.iloc[2, 4]),   
 bar = dict(color = "darkblue")),   
 number = dict(suffix = '명')))  
  
fig.add\_trace(go.Indicator(  
 type = 'indicator', mode = "gauge+number", title = df\_gauge.iloc[0, 1],  
 domain = dict(x = (0.3,0.8), y = (0.62, 0.7)),  
 value = df\_gauge.iloc[0, 3],   
 gauge = dict(axis = dict(  
 range = (0, df\_gauge.iloc[0, 4]\*1.2)),  
 shape = "bullet",  
 steps = [  
 dict(range = (0, (df\_gauge.iloc[0, 4])\*1.2\*0.5), color = "lightgray"),  
 dict(range = ((df\_gauge.iloc[0, 4])\*1.2\*0.5, (df\_gauge.iloc[0, 4])\*1.2\*0.75), color = "darkgray"),  
 dict(range = ((df\_gauge.iloc[0, 4])\*1.2\*0.75, (df\_gauge.iloc[0, 4])\*1.2), color = "gray")],  
 threshold = dict(  
 line = dict(color = 'white'),  
 value = df\_gauge.iloc[0, 4]),   
 bar = dict(color = "darkblue")),   
 number = dict(suffix = '명')))  
  
fig.add\_trace(go.Indicator(  
 type = 'indicator', mode = "gauge+number", title = df\_gauge.iloc[1, 1],  
 domain = dict(x = (0.3,0.8), y = (0.42, 0.5)),  
 value = df\_gauge.iloc[1, 3],   
 gauge = dict(axis = dict(  
 range = (0, df\_gauge.iloc[1, 4]\*1.2)),  
 shape = "bullet",  
 steps = [  
 dict(range = (0, (df\_gauge.iloc[1, 4])\*1.2\*0.5), color = "lightgray"),  
 dict(range = ((df\_gauge.iloc[1, 4])\*1.2\*0.5, (df\_gauge.iloc[1, 4])\*1.2\*0.75), color = "darkgray"),  
 dict(range = ((df\_gauge.iloc[1, 4])\*1.2\*0.75, (df\_gauge.iloc[1, 4])\*1.2), color = "gray")],  
 threshold = dict(  
 line = dict(color = 'white'),  
 value = df\_gauge.iloc[1, 4]),   
 bar = dict(color = "darkblue")),   
 number = dict(suffix = '명')))  
  
fig.add\_trace(go.Indicator(  
 type = 'indicator', mode = "gauge+number", title = df\_gauge.iloc[3, 1],  
 domain = dict(x = (0.3,0.8), y = (0.22, 0.3)),  
 value = df\_gauge.iloc[3, 3],   
 gauge = dict(axis = dict(  
 range = (0, df\_gauge.iloc[3, 4]\*1.2)),  
 shape = "bullet",  
 steps = [  
 dict(range = (0, (df\_gauge.iloc[3, 4])\*1.2\*0.5), color = "lightgray"),  
 dict(range = ((df\_gauge.iloc[3, 4])\*1.2\*0.5, (df\_gauge.iloc[3, 4])\*1.2\*0.75), color = "darkgray"),  
 dict(range = ((df\_gauge.iloc[3, 4])\*1.2\*0.75, (df\_gauge.iloc[3, 4])\*1.2), color = "gray")],  
 threshold = dict(  
 line = dict(color = 'white'),  
 value = df\_gauge.iloc[3, 4]),   
 bar = dict(color = "darkblue")),   
 number = dict(suffix = '명')))  
  
fig.add\_trace(go.Indicator(  
 type = 'indicator', mode = "gauge+number", title = df\_gauge.iloc[4, 1],  
 domain = dict(x = (0.3,0.8), y = (0.02, 0.1)),  
 value = df\_gauge.iloc[4, 3],   
 gauge = dict(axis = dict(  
 range = (0, df\_gauge.iloc[4, 4]\*1.2)),  
 shape = "bullet",  
 steps = [  
 dict(range = (0, (df\_gauge.iloc[4, 4])\*1.2\*0.5), color = "lightgray"),  
 dict(range = ((df\_gauge.iloc[4, 4])\*1.2\*0.5, (df\_gauge.iloc[4, 4])\*1.2\*0.75), color = "darkgray"),  
 dict(range = ((df\_gauge.iloc[4, 4])\*1.2\*0.75, (df\_gauge.iloc[4, 4])\*1.2), color = "gray")],  
 threshold = dict(  
 line = dict(color = 'white'),  
 value = df\_gauge.iloc[4, 4]),   
 bar = dict(color = "darkblue")),   
 number = dict(suffix = '명')))  
  
fig.update\_layout(grid=dict(rows=3, columns=3),  
 title = dict(text = '10만명당 사망자수(최근 공식발표 기준)', x = 0.5))  
  
fig.show()



실행결과 VI-4. python의 불릿 인디케이터

# 지도의 시각화

지도의 시각화는 plotly에서 쉽게 할수 있는 매우 강력한 시각화 방법이다. 사실 지도의 시각화는 엑셀이나 파워포인트, 포토샾이나 일러스트 같은 대중적으로 많이 사용되는 툴에서도 가능하겠지만 매우 번거로운 작업임에 틀림없다. 또 단순한 지도의 시각화라면 다소의 시간과 많은 정신적 스트레스를 감소하고라도 할 수 있겠지만 만약 200개가 넘는 시군구 지역별 시각화를 해야한다면 이는 단순히 시간과 정신적 스트레스과 바꿀수 있는 작업이 아닐 것이다. 이때는 전문가를 찾아 의뢰하는 것이 해결책이겠지만 항상 전문가들은 바쁘기때문에 시간이 많이 소요되고 비싸다. plolty는 지도의 시각화에 생각보다 많은 기능이 지원된다.

지도를 시각화하기 위해서는 먼저 몇가지 알아야하는 개념이 있다.

## 지도 시각화의 기초 개념

### 지형 데이터

지형 데이터는 지도를 표현하기 위한 지형의 경계선에 대한 폴리곤 정보를 가지고 있는 데이터를 말한다. 지형 데이터는 다음의 두 가지 형태의 데이터 파일로 제공된다.

* shape 파일

Shape 파일은 대부분의 GIS(Geographical Information System)에서 사용되는 지형 벡터 데이터 파일 포맷이다. 확장자를 ‘shp’로 설정하는데 선, 점, 다각형(Polygon)으로 지형을 벡터의 형태로 표현하는 텍스트 파일이다. 하지만 Shape 파일은 보통’\*.shp’, ‘\*.shx’, ‘\*.dbf’, ‘\*.kml’, ‘\*.prj’ 등의 파일이 한 세트처럼 제공되는 것이 일반적이다. 그렇다고 이 네개의 파일을 모두 쓰는 것은 아니고 사용하는 응용에 따라 선택하여 사용할 수 있다.

* geojson 파일

GeoJSON은 위치정보를 갖는 점을 기반으로 체계적으로 지형을 표현하기 위해 설계된 개방형 공개 표준 형식이다. 지형 정보를 자바스그립트 오브젝트 노테이션(JAVA Script Object Notation, JSON)을 활용하여 사용하는 파일 포맷으로 OpenLayers, Leaflet, MapServer, Geoforge 소프트웨어, GeoServer, GeoDjango, GDAL, Safe Software FME 등 많은 매핑 및 GIS 소프트웨어 패키지에서 지원하고있다[[1]](#footnote-49)

geojson 파일의 확장자는 ’\*.json’으로 붙는다. geojson 파일은 앞서 설명한 Shape 파일에서부터 변환하여 생성할 수 있다.

geojson 파일은 Shape 파일에 비해 파일 사이즈가 작고 처리속도도 Shape 파일보다 매우 빠르다는 장점이 있다.

### 지도 투영법

지구는 3차원의 구형태이지만 우리가 보는 지도는 2차원의 평면이다. 따라서 3차원을 2차원에 표현해야 하기 때문에 이 표현의 방법이 여러가지인데 이를 투영법(Projection)이라고 한다. plotly에서는 이 투영법에 따라 지도가 화면에 표현되는 평태가 달라진다. plolty에서 지원하는 투영법은 정각도법(equirectangular), 원통도법(mercator), 정사도법(orthographic) 등 총 22개의 투영법을 지원한다.

### 좌표계의 종류

일반적으로 지도의 좌표라고 하면 북위, 남위와 같이 적도를 기준으로 북쪽으로 얼마나 많이 떨어졌는지, 남쪽으로 얼마나 많이 떨어졌는지를 표시하는 위도(latitude)와 동경, 서경과 같이 영국 그리니치 천문대를 기준으로 동쪽으로의 위치, 서쪽으로의 위치를 표시하는 경도(longitude)를 생각한다. 이를 좌표계라고 하는데 이 좌표계도 몇 가지 종류가 있다. 이 좌표계가 어떻게 설정되어 있는가에 따라 같은 위도와 경도의 위치도 달라질 수 있다.

* WGS84(EPSG:4326): 미국에서 군사용으로 GPS 시스템를 이용하면서 만든 좌표계(경도와 위도)
* Bessel 1841(EPSG:4004): 3차원 원형 지구를 2차원으로 투영한 결과에 대한 좌표계로 우리나라에서 많이 사용되는 좌표계
* GRS80 UTM-K(EPSG:5179): 한반도 전체를 하나의 좌표계로, 전국을경도 127.5를 기준으로 설계한 지도로 네이버에서 사용
* GRS80 중부원점(EPSG:5181): 과거 지리원 좌표계를 타원체로 수정한 좌표계. Kakao, 공공데이터포탈에서 사용
* Web mercator projection(Pseudo-Mercator, EPSG:3857) : WGS84 타원체의 장반경을 반지름으로 하는 좌표계로 구글, 빙, 야후 등 에서 사용
* Albers projection(EPSG:9822) : 미국 지질 조사국에서 사용하는 좌표계

## 지도의 데이터 표현

plotly는 지도를 그리기 위해 사용하는 패키지가 아니다. 즉 plotly에서의 지도의 사용은 지역적 데이터를 표현하기 지도를 사용하여 시각화하기 위해 사용하는 것이다. 그렇다면 데이터를 어떻게 지도위에 표현할 것인가가 가장 중요하다. 지도위에 데이터를 표현하는 방법은 지역의 색을 사용하여 각각의 지역의 특성을 시각화하는 단계 구분도(Choropleth Map)과 지도의 특정 위치에 점을 표기하는 방법(Scatter Map)이 많이 사용된다.

### 지도 트레이스

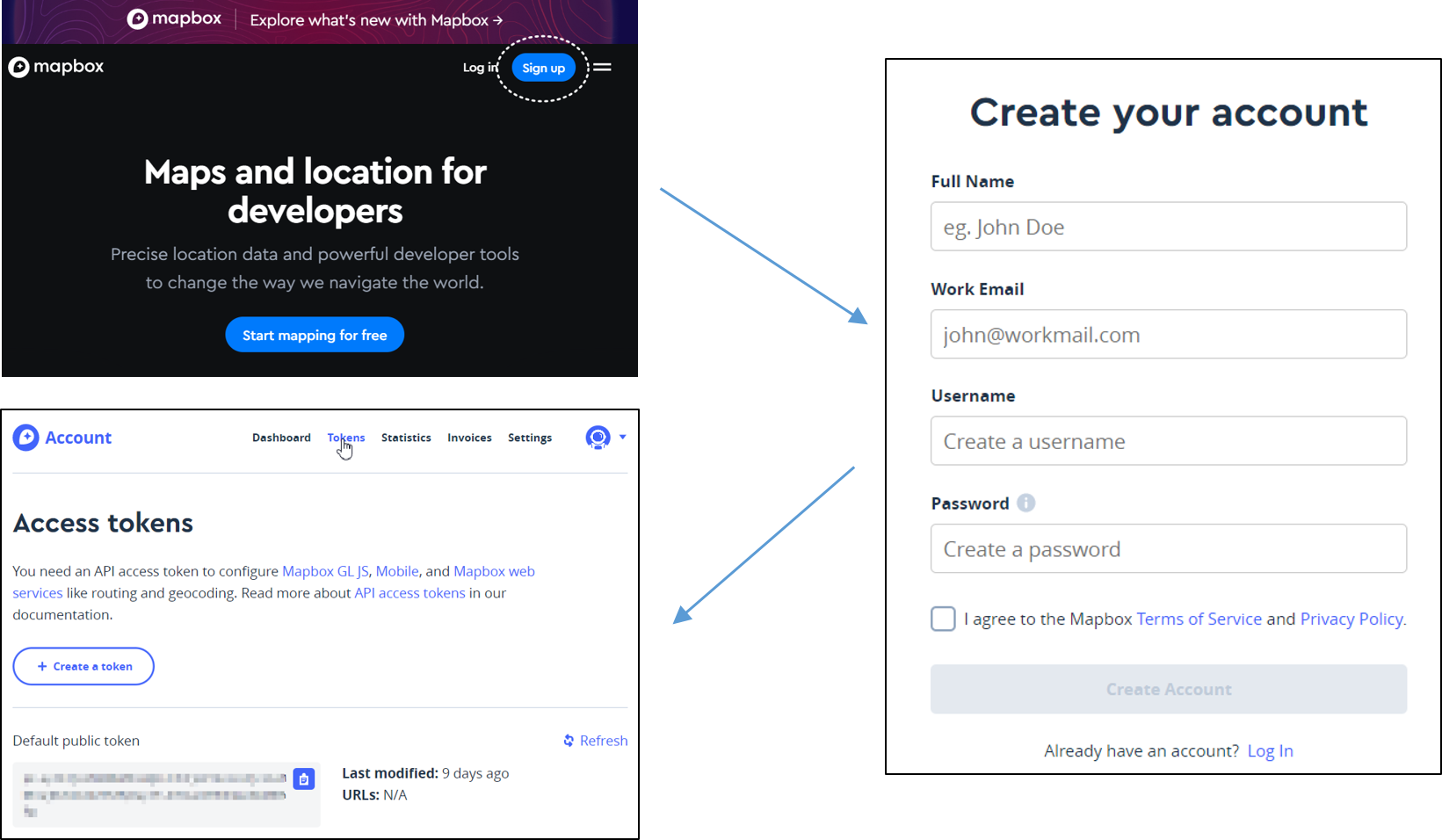
plotly에서 지도를 그리기 위해서는 앞서 설명한 두 가지 데이터 타입(shape, geojson) 중 geojson 타입을 기본적으로 사용한다. 이 geojson 데이터를 사용하여 지도에 데이터를 표현하는 방법에 따라 다음과 같이 다섯 가지 트레이스를 제공한다.

* scatter geo 트레이스 : geojson 지형 정보 위에 위도와 경도로 표현된 scatter 트레이스의 점이나 선을 표시하는 트레이스
* choropleth 트레이스 : geojson 지형 정보의 지형 다각형(polygon)에 ’z’로 매핑되는 내부 색을 통해 데이터를 표현하는 트레이스
* scattermapbox 트레이스 : mapbox.com에서 제공하는 맵위에 위도와 경도로 표현된 scatter 트레이스의 점이나 선을 표시하는 트레이스
* choroplethmapbox 트레이스 : mapbox.com에서 제공하는 맵위에 지형 다각형(polygon)에 ’z’로 매핑되는 내부 색을 통해 데이터를 표현하는 트레이스
* densitymapbox 트레이스 : 색상을 사용하여 위도, 경도 좌표와 ’z’에 매핑된 값에 따른 밀도 커널 데이터를 표현하는 트레이스

plotly에서 지원하는 지도 관련 트레이스는 다섯가지가 있고 사용자가 표시하기 원하는 지도의 형태에 따라 선택하여 사용하는 것이 원칙이나 R에서는 지도를 위한 R에 최적화된 몇 가지 추가적 함수를 제공한다. 따라서 R과 python을 따로 설명하도록 하는데, 단계 구분도(choropleth)와 맵박스 산점도(scattermapbox)를 그리는 방법을 설명한다.

단계 구분도는 지역적으로 구분된 데이터를 지도에 표시하는 방법으로 가장 많이 사용되는 것은 색으로 지도의 지형에 데이터를 표시하는 방법이다. 각각의 다각형(Polygon)으로 구성되어 있는 지역의 내부 색을 데이터에 따라 다르게 함으로서 지역간 데이터의 차이를 시각화 하는 방법이다. 결국 색이 칠해진 다각형을 배치함으로 데이터를 표현하는 것이다.

맵박스 산점도는 mapbox.com에서 제공하는 지도를 기반으로 해서 위도와 경도상의 위치에 특정 데이터를 점이나 선, 텍스트를 표시해주는 지도이다. mapbox.com의 지도를 사용하기 위해서는 먼저 mapbox.com에 가입하고 mapbox.com의 지도를 가져오기 위한 액세스 토큰를 받아야 한다.



위의 두 가지 지도를 그리기 위해 두 가지 추가적인 데이터를 사용한다. 첫 번째 데이터는 전국의 대학 충원률 데이터이다.[[2]](#footnote-58) 두 번째 데이터는 서울의 일부 대학의 위도, 경도 데이터이다.[[3]](#footnote-60)

plotly의 지도 관련 트레이스의 레이아웃을 설정하는데 사용하는 속성은 ‘geo’ 속성이다. 다음은 ’geo’의 주요 속성이다.

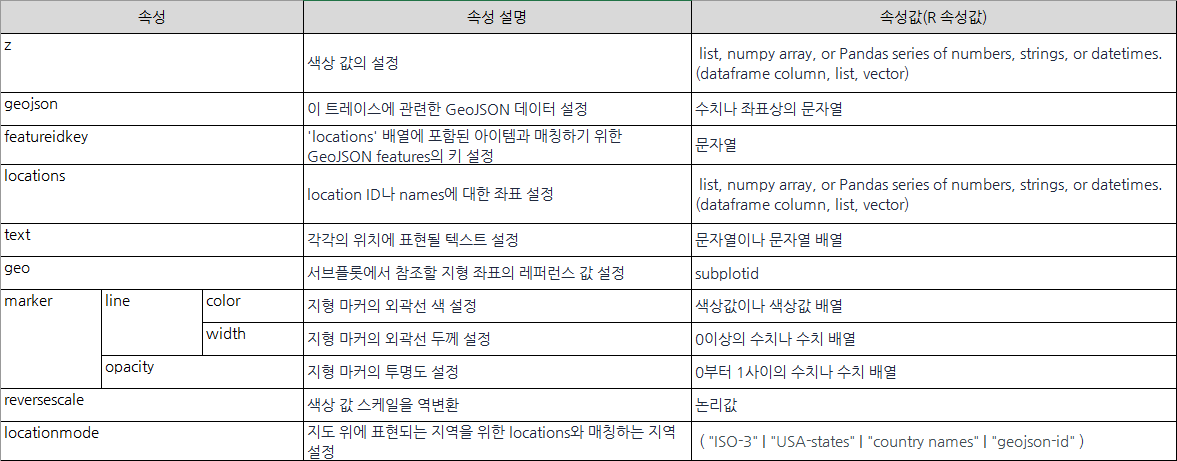
### python

python에서는 지도를 그리는 것은 지금까지 plotly를 그리는 방식과 같이 초기화된 plotly 객체에 지도를 위한 트레이스를 추가하는 방식으로 그릴 수 있다. plotly의 다섯 가지 지도 트레이스 중 choropleth 트레이스와 mapboxscatter 트레이스를 만드는 방법을 알아보겠다.

* 단계 구분도(choropleth)

python에서 choropleth 트레이스를 추가하지 위해서는 add\_trace()에 plotly.graph\_objects.Choropleth()를 사용하여 추가한다.

다음은 choropleth 트레이스에서 사용되는 주요 속성들이다.



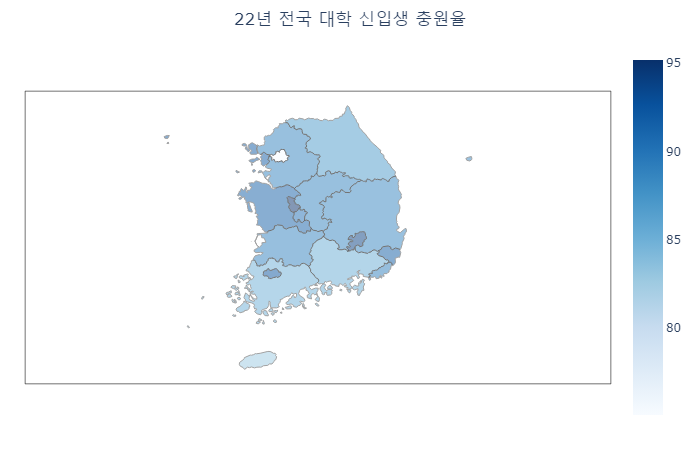
plotly에서는 미국 지도를 바로 지원하지만 한국 지도를 지원하지 않는다. 따라서 제일 먼저 한국 지형을 geoJSON으로 불러 읽어들이고 이를 ’geojson’으로 설정한다. 우리나라의 geoJSON 파일은 인터넷에서 쉽게 구할 수 있다. 파일을 다운 받을때는 시도 단위의 데이터인지 시군구 단위의 데이터인지를 잘 구분하여 다운해야 한다.[[4]](#footnote-65) geoJSON 파일의 데이터를 읽어오기 위해서 json 라이브러리를 임포트하고 json.load()를 사용하여 geoJSON 데이터를 불러온다.

다음으로 전국 대학 신입생 충원율 데이터를 불러오고 이 데이터 중 필요한 데이터 5개를 필터링하고 열 이름을 적절히 설정한다.

불러들여온 전국 대학 신입생 충원율 데이터에는 지역을 표현하는데 한글 시도명을 사용한다. 반면 geoJSON 데이터에서는 properties의 CTPRVN\_CD 속성으로 정의되는데 시도를 수치로 표현하고 있다. 그래서 이를 매칭시키기 위한 함수를 정의 하였고 이 함수를 사용하여 대학 신입생 충원율 데이터에 매핑된 수치 열을 만들어 주었다.

choropleth 트레이스를 만들기 위한 데이터 전처리가 끝났으면 plotly를 초기화하고 plotly.graph\_objects.Choropleth()를 사용한다. 앞에서 불러들어온 geoJSON 데이터를 ‘geojson’ 속성에 설정하고 충원율 데이터와 매칭할 키인 “properties.CTPRVN\_CD”를 ‘featureidkey’에 설정한다. 이번에는 ’featureidkey’에 매칭될 충원율 데이터의 열을 ’locations’에 설정하고 색상으로 표현해 줄 데이터인 충원율 열을 ’z’ 속성으로 설정한다. 전체적 색상 스케일 속성인 ’colorscale’을 푸른색 팔레트(“blues”)로 설정하고 지형의 경계선, 컬러바, 표현 텍스트, 호버를 설정하였다.

import json  
  
def cat(row):  
 key = row['지역']  
 value = {'강원' : '42','경기' : '41','경남' : '48','경북' : '47','광주' : '29','대구' : '27','대전' : '30','부산' : '26','서울' : '11','세종' : '36','울산' : '31','인천' : '28','전남' : '46','전북' : '45','제주' : '50','충남' : '44','충북' : '43'}.get(key)  
 return value  
  
geometry = json.load(open('D:/R/git/datavisualization/plotly/RnPy/chap6/TL\_SCCO\_CTPRVN.json', encoding='utf-8'))  
  
df\_충원율 = pd.read\_excel("D:/R/git/datavisualization/plotly/RnPy/chap6/고등 주요 01-시도별 신입생 충원율(2010-2022)\_220825y.xlsx",   
 sheet\_name = 'Sheet1', skiprows=(6), header = 0)  
   
df\_충원율 = df\_충원율.iloc[:, 0:5]  
df\_충원율.columns = ('연도', '지역', '정원내모집인원', '정원내입학생수', '신입생충원율')  
df\_충원율['지역코드']= df\_충원율.apply(cat, axis=1)  
df\_충원율 = df\_충원율[(df\_충원율['지역'] != '전국')&(df\_충원율['연도'] == 2022)]  
  
fig = go.Figure()  
  
fig.add\_trace(go.Choropleth(  
 geojson=geometry,   
 featureidkey='properties.CTPRVN\_CD',   
 locations = df\_충원율['지역코드'],   
 z = df\_충원율['신입생충원율'],   
 colorscale="Blues",   
# text = df\_충원율['신입생충원율'],  
 marker=dict(line=dict(width=1), opacity = 0.5),  
 colorbar = dict(y = 0.5, yanchor = 'middle'),  
 text = df\_충원율['지역'] + '\n' + round(df\_충원율['신입생충원율'], 2).astype(str) + '%',  
# hoveron = "fills",  
 hoverinfo = "text"  
))  
  
fig.update\_layout(  
 title = dict(text = '한국지도', x = 0.5),  
 geo=dict(  
 showframe = True,  
 fitbounds = "locations",  
 visible = False,  
 center=dict(lon=126.98, lat=37.56)  
 ),  
 autosize=False,  
 margin = dict(t = 50, b = 25, l = 25, r = 25,  
 pad=4,  
 autoexpand=True  
 ))

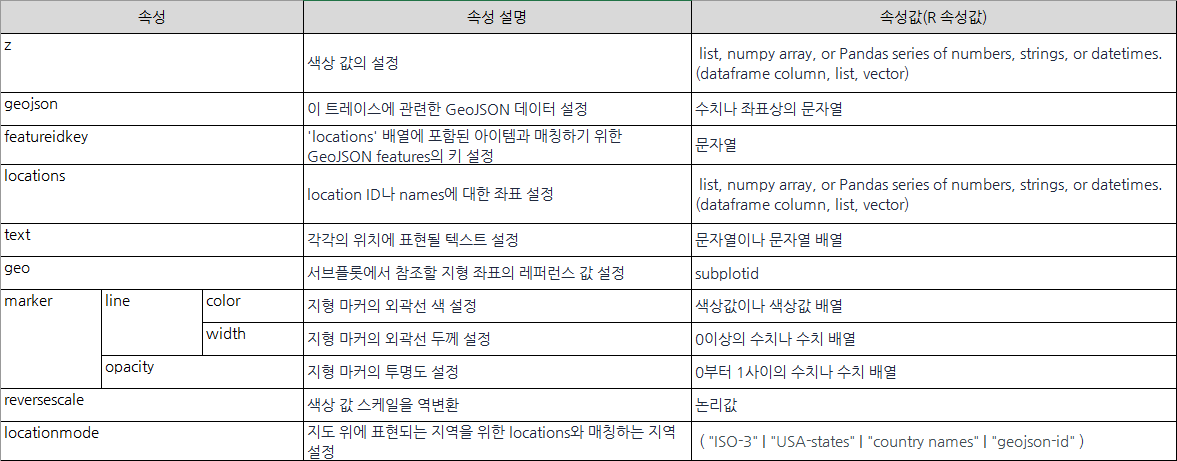


실행결과 VI-5. python의 단계구분도

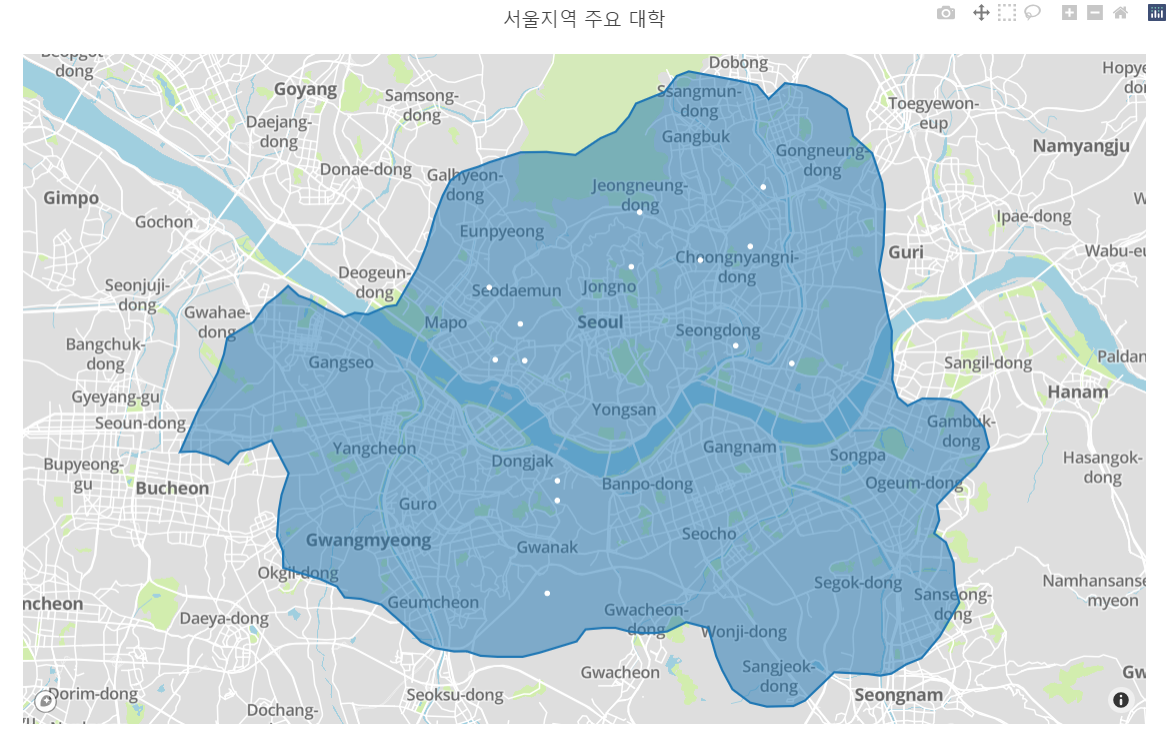
* 스캐터 맵박스 지도(Scatter Mapbox)

python에서 scattermapbox 트레이스를 추가하지 위해서는 add\_trace()에 plotly.graph\_objects.Scattermapbox()를 사용하여 추가한다.

다음은 Scattermapbox 트레이스에서 사용되는 주요 속성들이다.



mapbox\_access\_token = open(".mapbox\_token").read()  
  
fig = go.Figure(go.Scattermapbox(  
 lat=df\_univ['lat'],  
 lon=df\_univ['lon'],  
 mode='markers',  
 marker=go.scattermapbox.Marker(  
 size=9  
 ),  
 text=df\_univ['학교명'],  
 ))  
  
fig.update\_layout(  
 autosize=True,  
 hovermode='closest',  
 mapbox=dict(  
 accesstoken= mapbox\_access\_token,  
 bearing=0,  
 center=dict(lon=126.98, lat=37.56),  
 pitch=0,  
 zoom=10  
 ),  
)  
  
fig.show()



실행결과 VI-7. 맵박스를 사용한 R의 scatter 지도

### R

R에서 지도를 그리기 위해서는 하나 알아두어야 하는 것이 sf 클래스 데이터이다. 는데 sf 클래스는 R의 sf 패키지에서 제공하는 클래스 데이터 타입으로 ’simple feature’의 준말이다. sf 클래스는 지형 데이터를 R에서 기본적으로 사용되는 data.frame이나 tibble 형태로 저장하여 R에서 다루기가 쉽게 사용되는 클래스로 R을 종합적으로 관장하는 R consortium에서 관리하는 데이터 타입이다. 그 때문에 R에서 가장 쉽게 지형 데이터를 관리하고 처리할 수 있는 데이터 타입이기 때문에 plolty에서도 sf 데이터 타입을 지원하도록 설계되어 있다.

그래서 R에서는 plotly의 지도 전용 트레이스를 사용하기 보다는 sf 클래스를 활용한 R 전용 함수를 사용하는 방법이 지도의 품질상 더 효과적이다. sf 클래스는 plot\_ly(), plot\_geo()와 plot\_mapbox()로 사용이 가능하다.

plot\_ly()는 지금까지 plotly 객체를 초기화하는데 사용했던 바로 그 함수이다. plot\_ly()에 바인딩되는 데이터프레임이 sf 클래스면 plot\_ly()는 지형 데이터를 처리하도록 동작한다. plot\_ly()에서 사용하는 주요 매개변수는 다음과 같다.

plot\_ly(data = data.frame(), …, type = NULL, name, color, colors = NULL, alpha = NULL, stroke, strokes = NULL, alpha\_stroke = 1, size, sizes = c(10, 100), span, spans = c(1, 20), symbol, symbols = NULL, linetype, linetypes = NULL, split, frame, width = NULL, height = NULL, source = “A”)

* data : 시각화에 사용될 지형 데이터(sf 타입)
* color : 내부에 채워지는 색상 매핑으로 사용될 데이터
* colors : color에 사용될 색상 팔레트
* alpha : 투명도를 설정하는 색상의 알파 채널값
* stroke : 컬러와 유사하지만 색이 채워진 폴리곤에서의 외곽선 색
* split : 다중 트레이스로 구분될 값 설정

plot\_geo()는 R의 plotly에서 지도 객체의 초기화를 위해 사용하는 함수로 plot\_ly() 대신 사용할 수 있는 함수이다. plot\_geo()는 plotly의 주로 shape 파일이나 geoJSON으로 지도를 그리고 데이터를 사용하여 지도를 그리는데 지도 트레이스 중 scattermap 트레이스로 지도를 그린다. plot\_geo()의 매개변수는 ‘offline’ 매개변수를 제외하고는 plot\_ly()의 매개변수와 같다

plot\_geo(data = data.frame(), …, type = NULL, name, color, colors = NULL, alpha = NULL, stroke, strokes = NULL, alpha\_stroke = 1, size, sizes = c(10, 100), span, spans = c(1, 20), symbol, symbols = NULL, linetype, linetypes = NULL, split, frame, width = NULL, height = NULL, source = “A”, offline = FALSE)

* offline : 인터넷 연결 여부에 관계없이 지도를 볼 수 있도록 지형 데이터를 포함시킬지를 결정

plot\_mapbox()는 mapbox.com의 지도를 사용하여 지도 그리는 함수로 앞에서 언급했듯이 mapbox.com의 API key가 필요하다. plot\_mapbox()는 지도 트레이스 중 scattermapbox 트레이스로 지도를 그린다. 사용법은 plot\_ly()와 동일하다.

다음은 ggplot2의 geom\_sf()와 plot\_ly(), plot\_geo(), plot\_mapbox()로 그린 우리나라의 지도이다.

map\_data <- getData("GADM", country = "KOR", level = 1, type = "sf")  
st\_crs(map\_data) <- st\_crs(map\_data)  
  
subplot(nrows = 2,  
 ggplot(map\_data) + geom\_sf(),  
 plot\_ly(map\_data, color = I('skyBlue')),  
 plot\_geo(map\_data, color = I('skyBlue')),  
 plot\_mapbox(map\_data, color = I('skyBlue')), margin = 0.05  
) %>% hide\_legend()

* 단계 구분도(Choropleth Map)

처음으로 우리나라의 지도 데이터를 가져온다. R에서 단계 구분도를 그리기 위한 지형 데이터를 가져오는 방법은 여러가지가 있지만 여기서는 raster 패키지의 getData()를 통해 제공되는 한국의 지형 데이터를 사용하였다. getData()는 특정 국가의 지형 데이터를 가져오는 함수인데 ’level’에 따라 지도의 지형 레벨을 결정할 수 있는데 한국 지형 데이터의 경우 ’level = 1’이면 시도 단위의 지형 데이터를 의미한다. 이 데이터를 sf 데이터 타입으로 가져왔다.

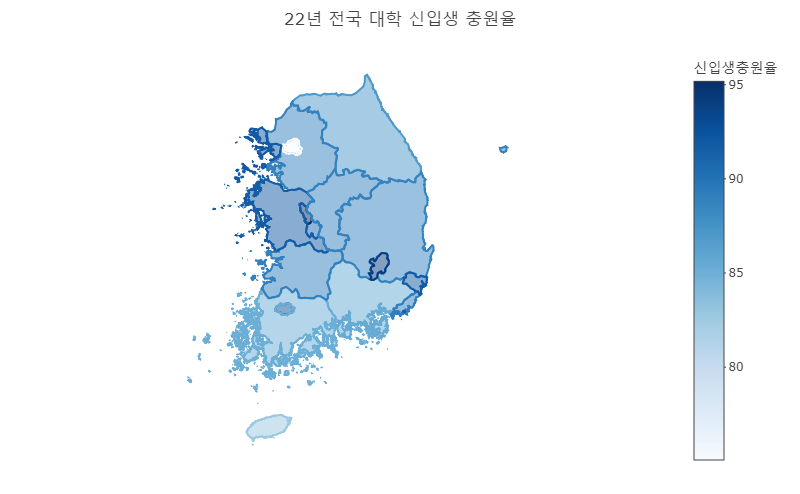
다음으로 지도에 표시할 데이터인 전국의 대학 신입생 충원율 데이터를 가져온다. 이 데이터 중 사용할 5개의 열만을 선택해서 데이터를 저장하고 열 이름을 설정한다.

R에서 단계 구분도를 그리기 위해 중요한 것은 지형 데이터와 표시할 데이터를 가진 데이터프레임을 조인하여 하나의 객체로 만들어야 한다는 것이다. 이를 위해서 지역을 표현하는 지형 데이터의 키와 표현할 데이터의 지역 키를 일치하도록 전처리 해야한다는 것이다. 전국 대학 충원률 데이터에서는 지역이 한글로 표현되어 있고 지형 데이터에서는 영문 키로 저장되어 있기 때문에 지형 데이터의 키와 맞추기 위해 충원률 데이터의 지역을 지형 데이터의 영문 키로 변환하는 과정을 거쳤다.

이제 충원율이 포함된 sf 클래스 데이터프레임을 사용하여 단계 구분도를 그린다. 단계 구분도를 그리기 위해서 plot\_ly(), plot\_geo(), plot\_mapbox() 중에 plot\_ly()를 사용하였다.

그리고 전체 플롯의 제목과 여백을 설정하기 위해 layout()을 사용하였다.

다음은 한국지도에 전국의 대학 신입생 충원율을 색상으로 표시하는 단계구분도를 만드는 R 코드이다.



실행결과 VI-5. R의 add\_sf()로 구현한 단계 구분도

* 스캐터 맵(Scatter Map)

스캐터 맵은 지도 위에 위도, 경도로 표현된 데이터를 표시하여 그 상대적 위치를 시각화하는 지도이다. 우리가 네이버나 카카오에서 특정 장소를 찾을 때 쓰는 지도가 바로 스캐터 맵이다. R에서 스캐터 맵을 구현하는 것은 앞에서 설명한 sf 클래스의 지형 데이터나 맵박스로 그린 지도 위에 위도와 경도 축에 표현되는 scatter 트레이스를 그리는 방법으로 구현할 수 있다.

다음은 서울의 주요 대학의 위도와 경도 데이터를 읽어들이고, raster패키지에서 우리나라 지도를 시군구 레벨(‘level = 2’)로 불러들였다. 이 지형 데이터에서 서울 지역 데이터만 필터링하고 plot\_mapbox()에 바인딩하고 add\_sf()를 사용하여 서울의 외곽선 만을 그렸다.

이 지도위에 scatter 트레이스로 마커와 텍스트로 각 대학의 위치를 표시하였다.

df\_univ <- read\_excel("D:/R/git/datavisualization/plotly/RnPy/chap6/university.xlsx",   
 col\_types = c('text', 'numeric', 'numeric'))  
   
map\_data\_lev2 <- getData("GADM", country = "KOR", level = 2, type = "sf")  
  
plot\_dat\_seoul <- plot\_dat |> filter(GID\_1 == 'KOR.16\_1')  
  
plot\_mapbox(plot\_dat\_seoul) |>  
 add\_sf(alpha = 0) |>  
 add\_trace(data = df\_univ, type = 'scattermapbox', mode = 'markers+text',   
 x = ~lon, y = ~lat,   
 marker = list(color = 'blue', symbol = 'college', size = 10, allowoverlap = T),  
 text = ~학교명, textposition = 'top center',   
 textfont = list(color = 'blue')) |>  
 layout(title = '서울지역 주요 대학',   
 margin = margins\_R,   
 showlegend = FALSE)

1. <https://ko.wikipedia.org/wiki/GeoJSON> [↑](#footnote-ref-49)
2. 해당 파일은 교육통계 홈페이지([https://kess.kedi.re.kr)에서](https://kess.kedi.re.kr)에서){.uri} 다운로드 받을 수 있으며 필자의 블로그에서도 다운받을 수 있다. [↑](#footnote-ref-58)
3. 해당 파일은 필자의 블로그에서도 다운받을 수 있다. [↑](#footnote-ref-60)
4. 우리나라 시도 단위의 geoJSON 파일은 필자의 블로그(2stndard.tistory.com)에서 다운 받을 수 있다. [↑](#footnote-ref-65)