



عنوان

مصورسازی شیوع ویروس کرونا در جهان

تهیه کننده

سارا معصومی

خرداد ۱۴۰۲

مقدمه

در این مقاله قصد داریم همه‌گیری ویروس کرونا را در سراسر جهان به صورت بصری بررسی کنیم. بنابر جدیدترین آمار جهانی کرونا، مجموع مبتلایان به بیماری کرونا در دنیا تاکنون به ۶۶۴ میلیون و ۷۲۶ هزار و ۳۴۶ نفر رسیده و مرگ شش میلیون و ۶۹۶ هزار و ۷۲۲ بیمار نیز تایید شده است.

ابتدا مجموعه داده‌ی مورد نیاز را از آدرس <https://covid19.who.int/data> دانلود می‌کنیم. این مجموعه داده شامل اطلاعات ثبت شده‌ای از تعداد مبتلایان جدید و تعداد فوتی‌های جدید در هر روز برای تمامی کشورها می‌باشد. یکی از ستون‌های موجود در داده‌ها، ستونی به نام WHO_region است که مقادیر این ستون تقسیم بندی نقشه جغرافیایی بر اساس سازمان بهداشت جهانی می‌باشد و مقادیر آن به شرح زیر است:

❖ AFRO : دفتر منطقه ای آفریقایی

❖ AMRO : دفتر منطقه ای برای کشورهای آمریکایی

❖ EMRO : دفتر منطقه ای برای کشورهای مدیترانه شرقی

❖ EURO : دفتر منطقه ای کشورهای اروپایی

❖ SEARO : دفتر منطقه ای کشورهای آسیای جنوب شرقی

❖ WPRO : دفتر منطقه ای کشورهای پاسیفیک غربی (غرب اقیانوس کبیر)

شایان ذکر است جهت ترسیم نمودارها، زبان پایتون و کتابخانه‌های pandas seaborn numpy و plotly.express مورد استفاده قرار گرفته‌اند.

آماده‌سازی داده

ابتدا لازم است کتابخانه‌های مورد نیاز را فراخوانی کنیم:

```
import pandas as pd
```

```
import numpy as np
```

```
import seaborn as sbn
```

```
import plotly.express as px
```

حال لازم است داده‌ها را دانلود کنیم، سپس به کمک Pandas آنها را فراخوانی کنیم و ستون تاریخ را به عنوان index معرفی کنیم:

```
data3 = pd.read_csv("ProjectData/WHO-COVID-19-global-data.csv", parse_dates = ["Date_reported"],
index_col = "Date_reported")
```

```
data3.head(10)
```

Out[307]:

| | Country_code | Country | WHO_region | New_cases | Cumulative_cases | New_deaths | Cumulative_deaths |
|---------------|--------------|-------------|------------|-----------|------------------|------------|-------------------|
| Date_reported | | | | | | | |
| 2020-01-03 | AF | Afghanistan | EMRO | 0 | 0 | 0 | 0 |
| 2020-01-04 | AF | Afghanistan | EMRO | 0 | 0 | 0 | 0 |
| 2020-01-05 | AF | Afghanistan | EMRO | 0 | 0 | 0 | 0 |
| 2020-01-06 | AF | Afghanistan | EMRO | 0 | 0 | 0 | 0 |
| 2020-01-07 | AF | Afghanistan | EMRO | 0 | 0 | 0 | 0 |
| 2020-01-08 | AF | Afghanistan | EMRO | 0 | 0 | 0 | 0 |
| 2020-01-09 | AF | Afghanistan | EMRO | 0 | 0 | 0 | 0 |
| 2020-01-10 | AF | Afghanistan | EMRO | 0 | 0 | 0 | 0 |
| 2020-01-11 | AF | Afghanistan | EMRO | 0 | 0 | 0 | 0 |
| 2020-01-12 | AF | Afghanistan | EMRO | 0 | 0 | 0 | 0 |

داده‌های فوق به ترتیب از سمت چپ شامل ستون‌های تاریخ ثبت اطلاعات، نام اختصاری کشورها، نام کامل کشورها، نام منطقه بهداشتی متناظر آنها، تعداد مبتلایان جدید، فراوانی تجمعی مبتلایان جدید، تعداد

فوتی‌های جدید و فراوانی تجمعی فوتی‌های جدید می‌باشد. که ستون تاریخ ثبت به عنوان index در نظر گرفته شده است. جهت مرتب سازی داده‌ها آنها را بر اساس تاریخ ثبت مرتب می‌کنیم:

```
data3.sort_values(by=['Date_reported'],inplace=True)
```

```
data3.head()
```

Out[398]:

| | Date_reported | Country_code | Country | WHO_region | New_cases | Cumulative_cases | New_deaths | Cumulative_deaths |
|--|---------------|--------------|----------------------------------|------------|-----------|------------------|------------|-------------------|
| | 2020-01-03 | AF | Afghanistan | EMRO | 0 | 0 | 0 | 0 |
| | 2020-01-03 | ME | Montenegro | EURO | 0 | 0 | 0 | 0 |
| | 2020-01-03 | TG | Togo | AFRO | 0 | 0 | 0 | 0 |
| | 2020-01-03 | BO | Bolivia (Plurinational State of) | AMRO | 0 | 0 | 0 | 0 |
| | 2020-01-03 | IQ | Iraq | EMRO | 0 | 0 | 0 | 0 |

با توجه به خروجی فوق می‌بینیم داده‌ها بر اساس تاریخ ثبت مرتب شده‌اند اما برخی از ردیف‌ها تمام مقادیرشان برابر صفر می‌باشد. به جهت کاهش حجم داده‌ها و در پی آن سادگی کار، ردیف‌هایی را که شامل هیچ مقدار عددی نمی‌باشند حذف می‌کنیم:

```
data3 = data3.loc[(data3["New_cases"]!= 0) | (data3["New_deaths"]!= 0) |
(data3["Cumulative_cases"]!= 0) | (data3["Cumulative_deaths"]!= 0) ]
```

```
data3.head()
```

Out[405]:

| | Date_reported | Country_code | Country | WHO_region | New_cases | Cumulative_cases | New_deaths | Cumulative_deaths |
|-------|---------------|--------------|---------|------------|-----------|------------------|------------|-------------------|
| 99295 | 2020-01-04 | DE | Germany | EURO | 1 | 1 | 0 | 0 |
| 53467 | 2020-01-04 | CN | China | WPRO | 1 | 1 | 0 | 0 |
| 90384 | 2020-01-04 | FI | Finland | EURO | 1 | 1 | 0 | 0 |
| 53468 | 2020-01-05 | CN | China | WPRO | 0 | 1 | 0 | 0 |
| 99296 | 2020-01-05 | DE | Germany | EURO | 0 | 1 | 3 | 3 |
| 90385 | 2020-01-05 | FI | Finland | EURO | 0 | 1 | 0 | 0 |
| 53469 | 2020-01-06 | CN | China | WPRO | 3 | 4 | 0 | 0 |
| 90386 | 2020-01-06 | FI | Finland | EURO | 0 | 1 | 0 | 0 |
| 99297 | 2020-01-06 | DE | Germany | EURO | 0 | 1 | 0 | 3 |
| 99298 | 2020-01-07 | DE | Germany | EURO | 0 | 1 | 0 | 3 |

همانطور که در جدول فوق می‌بینیم دیگر هیچ ردیفی که شامل اطلاعات مفیدی نباشد وجود ندارد. و همه‌ی ردیف‌ها حداقل یکی از ستون‌های متناظرشان دارای مقدار عددی است. اما نکته‌ی قابل توجه این است که

تاریخ های ثبت شده بسیار جزئی هستند بنابراین حجم داده ها بسیار بالاست برای راحتی کار می توانیم تاریخ ها را در مقیاس کلی تر مانند مقیاس ماهانه بررسی کنیم.

برای تبدیل تاریخ از مقیاس روزانه به ماهانه از دستورات زیر کمک میگیریم :

```
data4 = data3.resample('M').sum()
```

```
data4.head(10)
```

Out[407]:

| | New_cases | Cumulative_cases | New_deaths | Cumulative_deaths |
|---------------|-----------|------------------|------------|-------------------|
| Date_reported | | | | |
| 2020-01-31 | 9846 | 37542 | 219 | 1010 |
| 2020-02-29 | 75362 | 1633808 | 2717 | 47134 |
| 2020-03-31 | 696282 | 8087962 | 40577 | 374902 |
| 2020-04-30 | 2223610 | 57137169 | 193612 | 4269307 |
| 2020-05-31 | 2887315 | 136685188 | 164398 | 10082180 |
| 2020-06-30 | 4268490 | 237586653 | 153671 | 14389302 |
| 2020-07-31 | 6973772 | 419917707 | 184438 | 20040235 |
| 2020-08-31 | 8287664 | 662789587 | 193832 | 26086537 |
| 2020-09-30 | 8623887 | 893625296 | 168305 | 30679343 |
| 2020-10-31 | 11771167 | 1230304436 | 177318 | 36907448 |

پس از اجرای دستورات، جدول فوق را به عنوان خروجی مشاهده می کنیم که تاریخ ثبت اطلاعات به صورت ماهانه برای تمام کشورها در نظر گرفته شده است. به عنوان مثال در ردیف اول می بینیم که از تاریخ ۲۰۲۰.۱.۳۱ تا ۲۰۲۰.۱۰.۳۱ در سراسر جهان چه تعداد بیمار و فوتی ناشی از کرونا ثبت شده است.

مصور سازی

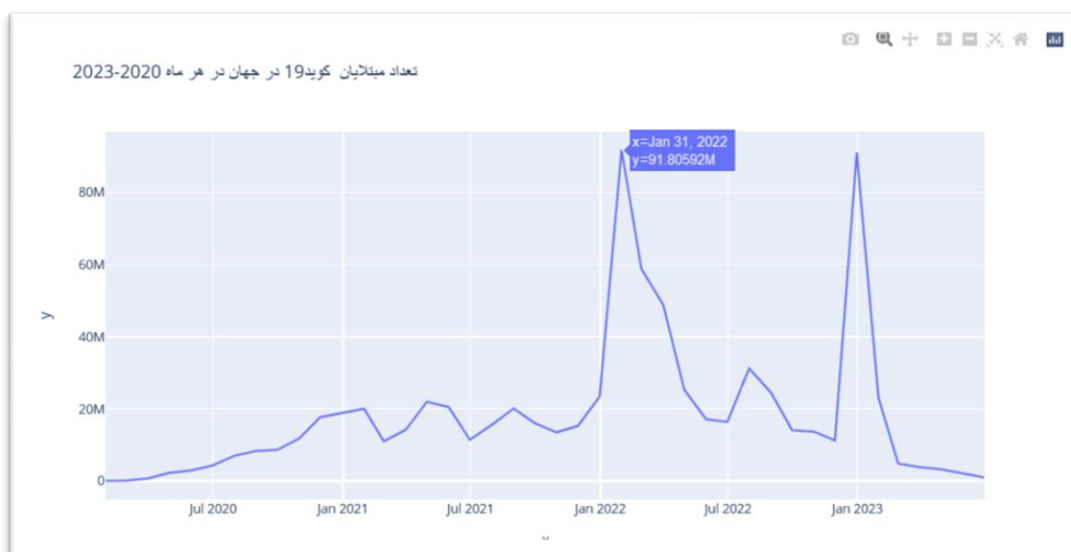
ابتدا می‌خواهیم نموداری را جهت بررسی روند ابتلای کرونا در جهان از سال ۲۰۲۰ لغایت ۲۰۲۳ رسم کنیم:

```
x= data4.index
```

```
y= data4["New_cases"]
```

```
fig = px.line( x = x , y = y, title = 'تعداد مبتلایان کوید ۱۹ در جهان در هر ماه ۲۰۲۰-۲۰۲۳')
```

```
fig.show()
```



همانطور که در تصویر بالا مشاهده می‌کنیم تعداد مبتلایان جهان در طول زمان ترسیم شده‌اند نمودار رسم شده دارای ۶ موج است که بیشترین ابتلا در این ۳ سال مربوط به ۳۱ jan ۲۰۲۲ و پس از آن اوایل سال ۲۰۲۳ می‌باشد. برای بررسی بهتر می‌توانیم این نمودار را برای تعداد فوتی‌های ناشی از کرونا نیز رسم کنیم. تا به نتایج مفیدتری برسیم:

```
x= data4.index
```

```
y= data4["New_deaths"]
```

```
fig2 = px.line( x = x , y = y, title = 'تعداد فوتی‌های کوید ۱۹ در جهان در هر ماه ۲۰۲۰-۲۰۲۳')
```

```
fig2.show()
```



با توجه به نمودار بالا بیشترین فوتی های ناشی از کرونا در تاریخ ۳۱ jan ۲۰۲۱ ثبت شده است در واقع یعنی در زمانی که تازه موج اول کرونا جهان را برگرفته بوده است. در تاریخ jan 2023 که تعداد مبتلایان بسیار بیشتر از jan 2021 است اما تعداد فوتی ها بسیار کمتر می باشد. کاهش چشگیر فوتی های ناشی از کرونا نسبت به مبتلایان میتواند ناشی از آشنایی مردم با شیوه های مراقبتی در طول بیماری و همچنین استفاده از واکسن کرونا باشد.

در ادامه قصد داریم همه گیری ویروس کرونا و فوتی های ناشی از آن را به تفکیک مناطق بهداشتی بر روی نمودار مشاهده کنیم:

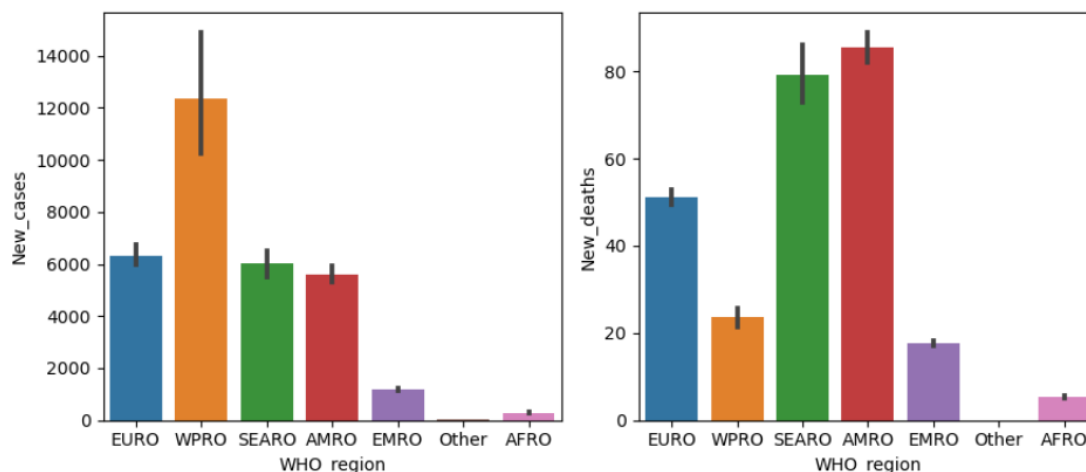
```
plt.rcParams["figure.figsize"] = [9.00, 4.00]
```

```
plt.rcParams["figure.autolayout"] = True
```

```
fig, axs = plt.subplots(1,2)
```

```
sbn.barplot(x="WHO_region", y="New_cases", data=data3, ax=axs[0])
```

```
sbn.barplot(x="WHO_region", y="New_deaths", data=data3, ax=axs[1])
```



در تصویر فوق نمودار سمت چپ نمودار تعداد مبتلایان جدید به تفکیک منطقه بهداشتی و نمودار سمت راست تعداد فوتی‌های جدید به تفکیک مناطق بهداشت است. همانطور که در نمودار سمت چپ می‌بینیم بیشترین ابتلا مربوط به منطقه WPRO یعنی غرب اقیانوس آرام و کمترین ابتلا مربوط به منطقه AFRO یعنی کشورهای آفریقایی می‌باشد. اما آمار فوتی‌ها مطابق آمار مبتلایان نیست. و بیشترین فوتی‌ها مربوط به منطقه AMRO یعنی کشورهای آمریکایی است بنابراین با توجه به تعداد مبتلایان و تعداد فوتی‌های کشورهای آمریکایی می‌توان نتیجه گرفت کشورهای آمریکایی در بهبود حال بیماران کرونایی عملکرد ضعیفی داشته‌اند اما در مقابل کشورهای منطقه WPRO با توجه به این که بیشترین آمار مبتلایان ثبت شده مربوط به این منطقه است اما در بهبود حال بیماران عملکرد خوبی داشته‌اند چرا که آمار فوتی‌های این منطقه نسبت به سایر مناطق کمتر است.

می‌خواهیم برای متغیر تعداد مبتلایان جدید و متغیر تعداد فوتی‌های جدید نمودار جعبه ای رسم کنیم تا ببینیم آیا مشاهده پرت برایشان وجود دارد یا خیر. همچنین ارتباط آنها را نیز روی نمودار ببینیم. اما برای این کار ابتدا لازم است داده های خود را یکبار دیگر آماده‌سازی کنیم چرا که نباید ستون تاریخ به عنوان index تعریف شده باشد :

```
data2 = pd.read_csv("ProjectData/WHO-COVID-19-global-data.csv")
```

```
data2.sort_values(by=['Date_reported'],inplace=True)
```

```
data2 = data2.loc[(data2["New_cases"]!= 0) | (data2["New_deaths"]!= 0) |  
(data2["Cumulative_cases"]!= 0) | (data2["Cumulative_deaths"]!= 0) ]
```

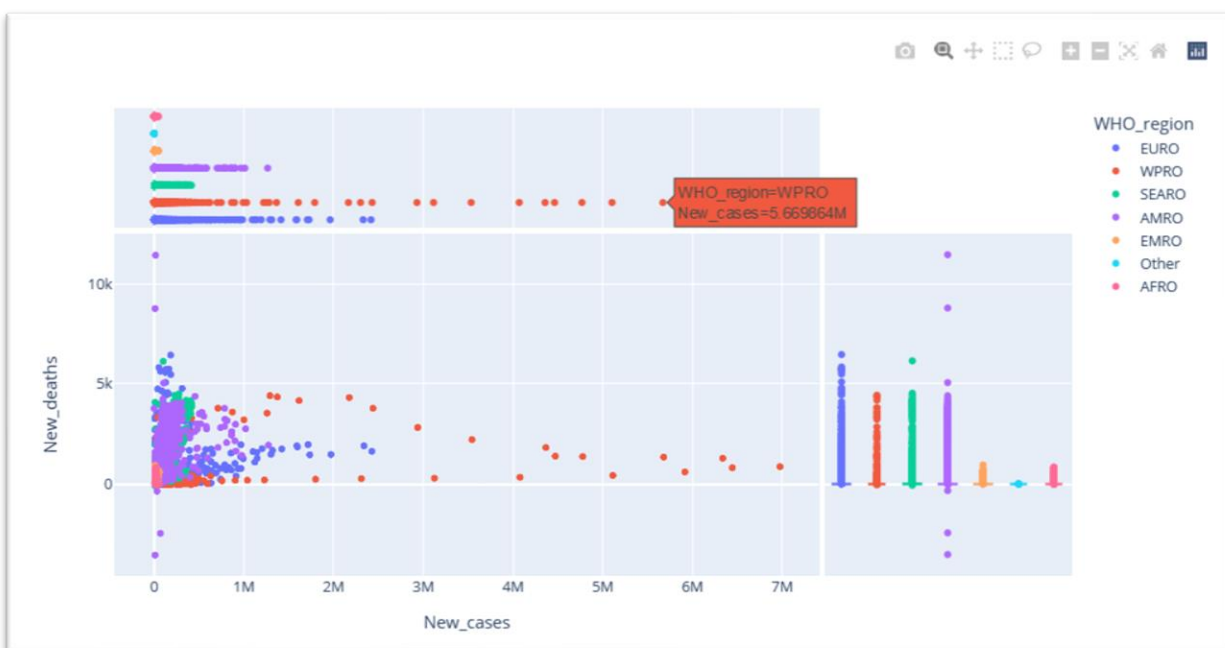
```
data2.head()
```


Out[405]:

| | Date_reported | Country_code | Country | WHO_region | New_cases | Cumulative_cases | New_deaths | Cumulative_deaths |
|-------|---------------|--------------|---------|------------|-----------|------------------|------------|-------------------|
| 99295 | 2020-01-04 | DE | Germany | EURO | 1 | 1 | 0 | 0 |
| 53467 | 2020-01-04 | CN | China | WPRO | 1 | 1 | 0 | 0 |
| 90384 | 2020-01-04 | FI | Finland | EURO | 1 | 1 | 0 | 0 |
| 53468 | 2020-01-05 | CN | China | WPRO | 0 | 1 | 0 | 0 |
| 99296 | 2020-01-05 | DE | Germany | EURO | 0 | 1 | 3 | 3 |
| 90385 | 2020-01-05 | FI | Finland | EURO | 0 | 1 | 0 | 0 |
| 53469 | 2020-01-06 | CN | China | WPRO | 3 | 4 | 0 | 0 |
| 90386 | 2020-01-06 | FI | Finland | EURO | 0 | 1 | 0 | 0 |
| 99297 | 2020-01-06 | DE | Germany | EURO | 0 | 1 | 0 | 3 |
| 99298 | 2020-01-07 | DE | Germany | EURO | 0 | 1 | 0 | 3 |

```
fig2 = px.scatter(data2, x = "New_cases", y = "New_deaths", color="WHO_region",
marginal_y="box",marginal_x="box")
```

```
fig2.show()
```



همانطور که در نمودار فوق می بینیم تعداد مبتلایان برای منطقه WPRO شامل داده های پرت می باشد یعنی در بعضی ماه ها تعداد مبتلایان بسیار بیشتر از ماه های دیگر بوده است که لازم است مورد تحقیق و بررسی قرار بگیرد اما در منطقه WPRO برای تعداد فوتی ها هیچ داده ی پرتی مشاهده نمی شود وقتی به نمودار پراکنش

برای این منطقه نگاه کنیم نتیجه می‌گیریم که درست است که تعداد مبتلایان این منطقه زیاد است اما در مدیریت حال بیماران برای جلوگیری از فوت آنها عملکرد بسیار خوبی داشته‌اند.

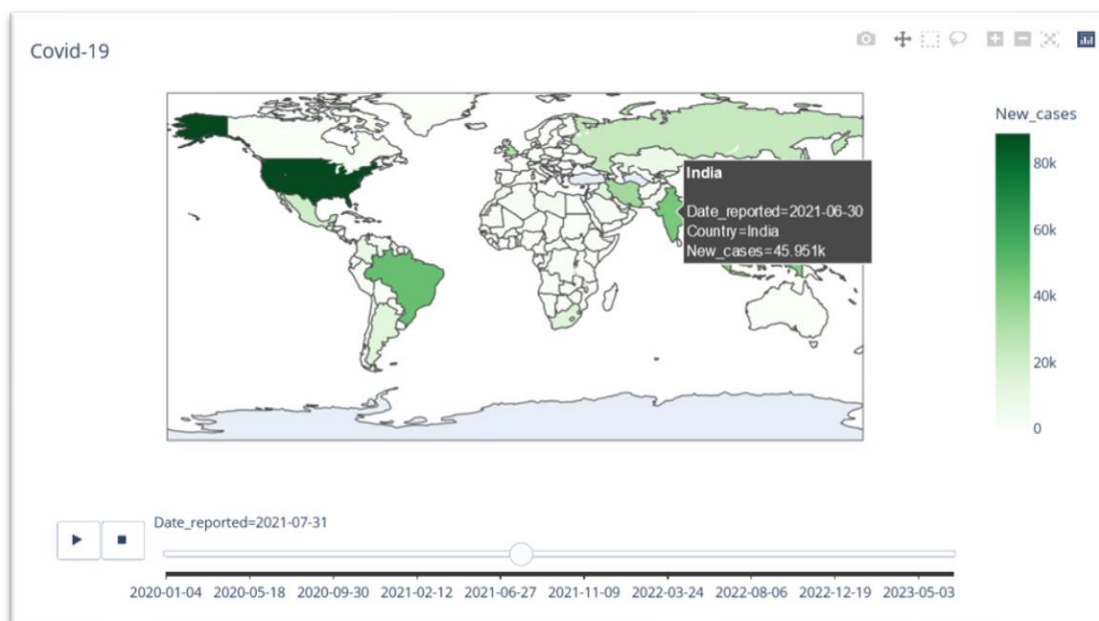
در ادامه می‌خواهیم روند گسترش ویروس کرونا را در جهان بر روی نقشه‌ی جغرافیایی در طول گذر زمان ببینیم:

```
fig = px.choropleth(data2, locations="Country",
locationmode="country names",
color="New_cases",
hover_name="Country",
animation_frame="Date_reported",
animation_group = "Date_reported",
color_continuous_scale='Greens')

fig.update_layout(

title_text = 'Renewable energy production across the world (% of electricity production)',geo =
dict(projection={'type':'equiarectangular'}))

fig.show()
```



همانطور که در تصویر فوق می‌بینیم این نمودار به صورت interactive می‌باشد شما می‌توانید به کمک دکمه‌های موجود برای **play** , **stop** روند گسترش ویروس کرونا را در طول زمان بر روی نقشه مشاهده کنید و یا با **hover** کردن کشورهای روی نقشه می‌توانید آمار مربوط به هر کشور را مشاهده کنید از دیگر قابلیت‌های این نمودار میتوان به قابلیت **zoom in** , **zoom out** اشاره کرد. اگر نمودار فوق را از روز اول داده‌ها **play** کنیم می‌بینیم که شروع ویروس کرونا از کشور چین بوده‌است و سپس در برخی از کشورهای قاره‌ی آمریکا گسترش یافته‌است پس از آنها به تدریج تمام کشورها دچار این بیماری می‌شوند. اما با گذر زمان می‌بینیم کشور چین به سرعت ابتلای ویروس کرونا را در کشور خود مهار میکند اما همچنان سایر کشورها درگیر این بیماری هستند.

منابع

1. <https://covid19.who.int/data>
2. <https://seaborn.pydata.org/generated>
3. <https://www.amazon.com/Introducing-Data-Science-Machine-Learning/dp/1633430030>
4. <https://plotly.com/python/plotly-express/>