

omicron-vs-delta

February 7, 2022

```
[1]: import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

from matplotlib import pyplot as plt
import plotly.express as px
import plotly.graph_objects as go
from plotly.subplots import make_subplots
```

Omicron vs Delta

1 Introduction

The goal of this analysis is compare and contrast the data of the omicron and delta variant to gain insight into differences by country.

2 Data reading and preprocessing

```
[2]: data=pd.read_csv("covid-variants.csv")
```

```
[3]: data.head()
```

```
[3]:
```

	location	date	variant	num_sequences	perc_sequences	\
0	Angola	2020-07-06	Alpha	0	0.0	
1	Angola	2020-07-06	B.1.1.277	0	0.0	
2	Angola	2020-07-06	B.1.1.302	0	0.0	
3	Angola	2020-07-06	B.1.1.519	0	0.0	
4	Angola	2020-07-06	B.1.160	0	0.0	

	num_sequences_total
0	3
1	3
2	3
3	3
4	3

```
[4]: data["variant"].unique()
```

```
[4]: array(['Alpha', 'B.1.1.277', 'B.1.1.302', 'B.1.1.519', 'B.1.160',
        'B.1.177', 'B.1.221', 'B.1.258', 'B.1.367', 'B.1.620', 'Beta',
        'Delta', 'Epsilon', 'Eta', 'Gamma', 'Iota', 'Kappa', 'Lambda',
        'Mu', 'Omicron', 'S:677H.Robin1', 'S:677P.Pelican', 'others',
        'non_who'], dtype=object)
```

It's hard to understand the meaning of the variant "non_who" so we will ignore it.

```
[5]: data=data[data["variant"]!="non_who"]
```

Now we add a new column that will serve as an unique key for each row: this will be useful for filtering the data.

```
[6]: selection_zip=list(zip(data['location'],data['date']))
c=-1
old=""
selection_col=[]
for x in selection_zip:
    if (x!=old):
        old=x
        c+=1
    selection_col.append(c)
data["selection_col"]=selection_col
```

The last preprocessing step is to obtain the filtered dataset for the variants omicron and delta, while the rest of the variants will be contained in a third dataset.

```
[7]: omicron=data[data["variant"]=="Omicron"]
delta=data[data["variant"]=="Delta"]
others=data[(data["variant"]!="Delta") & (data["variant"]!="Omicron")]
```

3 How many omicron and delta variants have been sequenced for each date globally?

```
[8]: omicron_global=omicron.groupby("date").sum()
omicron_global["date"]=omicron_global.index
omicron_global=omicron_global[omicron_global["num_sequences"]>0]

delta_global=delta.groupby("date").sum()
delta_global["date"]=delta_global.index
delta_global=delta_global[delta_global["perc_sequences"]>0]

fig = go.Figure()
fig.add_trace(go.
    ↳Scatter(x=omicron_global["date"],y=omicron_global["num_sequences"],mode='lines',line_width=
    ↳num_sequences"))
```

```
fig.add_trace(go.Scatter(x=delta_global["date"],y=delta_global["num_sequences"],mode='lines',line_width=5,
    name="num_sequences"))
fig.update_layout(title='Global log num_sequences for each date')
fig.show()
```

From the plot we can notice that: * The delta variant was being found in the last days of October, then it began slowly to spread. * The omicron variant was being found one year later, then it began, faster than the delta variant, to spread.

At the end of the plot (end of 2021/ begin of 2022) it's possible to observe that for both variants we have very low num_sequences: we will investigate this behaviour later.

Let's plot the logarithmic version of the previous plot.

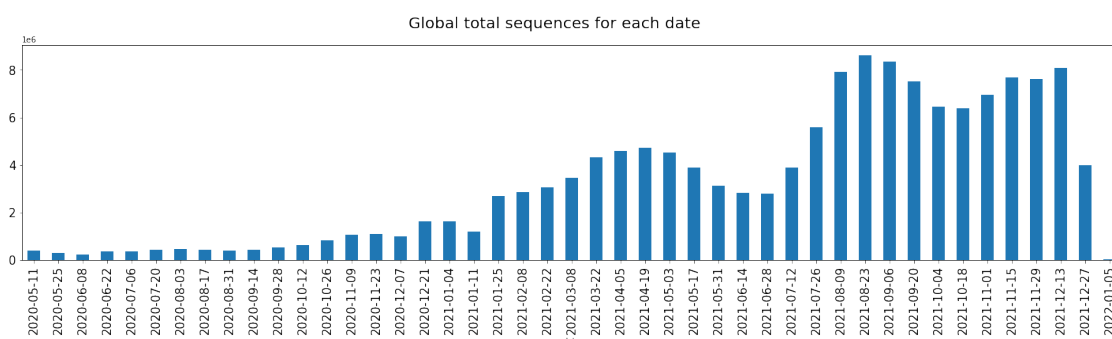
```
[9]: fig = go.Figure()
fig.add_trace(go.Scatter(x=omicron_global['date'],y=np.
    log1p(omicron_global["num_sequences"]),mode='lines+markers',line_width=5,marker_size=10,
    name="num_sequences logarithm"))
fig.add_trace(go.Scatter(x=delta_global['date'],y=np.
    log1p(delta_global["num_sequences"]),mode='lines+markers',line_width=5,marker_size=10,
    name="num_sequences logarithm"))
fig.update_layout(title='Global num_sequences logarithm for each data')
fig.show()
```

From this plot it's clear that the omicron variant spreads faster than the delta variant: as we can see, the omicron curve is very likely to be exponential.

At the end of the plot we observe the same behaviour we noticed before.

4 How many samples have been sequenced every date globally?

```
[10]: ax=data.groupby("date").sum().plot.bar(y="num_sequences_total",figsize=(25,5),
    fontsize=15,legend=False)
title=plt.suptitle("Global total sequences for each date",fontsize=20)
```



For each year, we can assert that: * In the year 2020 the countries started to analyze the viruses (as the pandemic started in this year), so for each date we have very low sequences. * In the year 2021 the countries put more efforts, in fact in the second semester they more than doubled the sequenced for each date they did in respect to the first semester.

However, in the date 2021/12/27 we can notice that the total number of sequence is half than the usual number the countries were able to reach in the last months, meanwhile the lowest peak, in this plot, is reached at the date 2022/01/05.

My interpretation for this behaviour is that the countries did less analysis because of Christmas holidays and/or the data is incomplete (for example, most of the countries didn't provide their latest data).

5 According to the last trustable data, how much are widely diffused the delta and the omicron variants?

```
[11]: #we filter only the countries that did a reasonable number of sequences
omicron_top=omicron[omicron["num_sequences_total"]>=300]
for country in np.unique(omicron_top["location"]):
    tmp=omicron_top[omicron_top["location"]==country]
    latest=tmp.iloc[-1]
    if (latest["date"]>="2021-12-27"):
        omicron_top=omicron_top.drop((tmp[tmp["date"]<latest["date"]].index))
        continue
    omicron_top=omicron_top.drop((tmp.index))
omicron_top=omicron_top.sort_values("perc_sequences",ascending=True)

#now we filter the delta dataset using the columns we inserted before.
indexes=[delta[delta["selection_col"]==x] for x in omicron_top["selection_col"]]
delta_top=pd.DataFrame(np.array(indexes).reshape(omicron_top.
    ↳shape[0],7),columns=delta.columns)

y=np.arange(omicron_top.shape[0])

fig, axes = plt.subplots(ncols=2, sharey=True,figsize=(25, 15))
axes[0].barh(y, omicron_top["perc_sequences"], align='center',
    ↳color='mediumturquoise', zorder=10)
axes[0].set_title('Omicron sequences (%)',fontsize=30)
axes[1].barh(y, delta_top["perc_sequences"], align='center', color='orange',
    ↳zorder=10)
axes[1].set_title('Delta sequences (%)',fontsize=30)

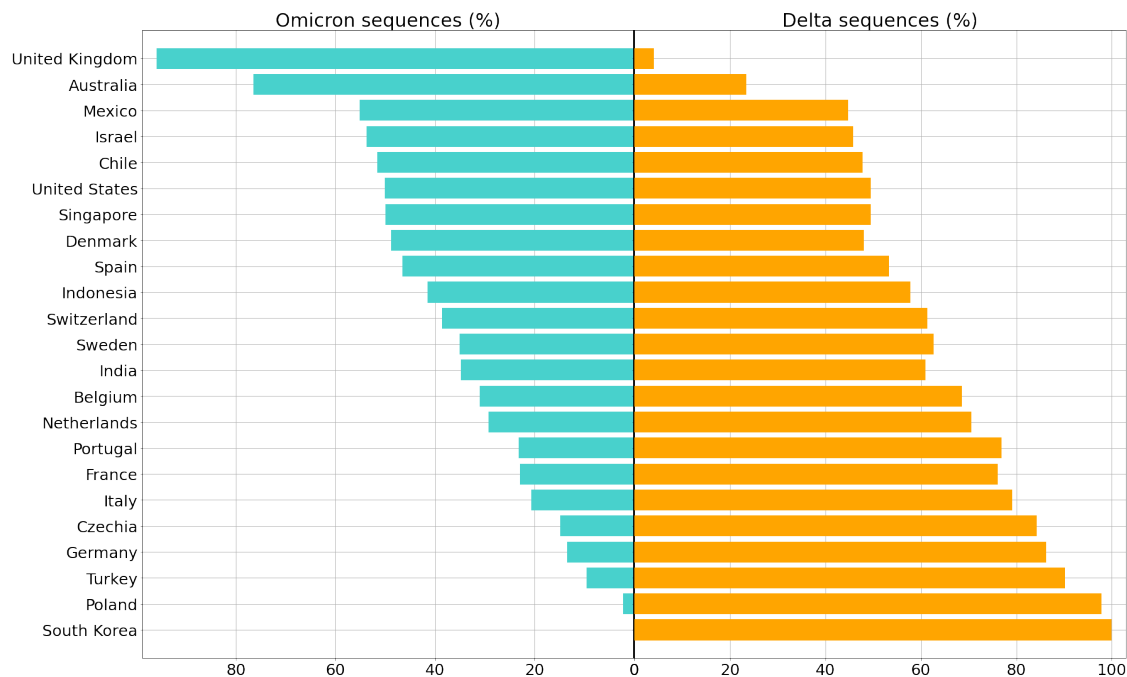
axes[0].invert_xaxis()
axes[0].tick_params(axis="both", labelsize=25)
```

```

axes[0].set(yticks=y, yticklabels=omicron_top["location"])
axes[1].tick_params(axis="both", labelsize=25)
axes[0].axvline(0,zorder=10,linewidth=5,color='black')
for ax in axes.flat:
    ax.margins(0.03)
    ax.grid(True)

fig.tight_layout()
fig.subplots_adjust(wspace=0)
plt.show()

```



It's clear that, if we analyze a random sample it's very likely that we find either the delta or the omicron variant in those countries: however, it's still possible to find other variants (as we will see soon).

6 Are there any differences in the diffusion of the variants in South Africa?

The omicron variant was first found in South Africa. Let's see some plots.

```

[12]: omicron_sa=omicron[omicron["location"]=="South Africa"]
omicron_sa=omicron_sa[omicron_sa['perc_sequences']>0]

delta_sa=delta[delta["location"]=="South Africa"]
filter_dates=np.where(delta_sa['perc_sequences']==0)[0]

```

```

delta_sa=delta_sa[delta_sa['date']>=delta_sa['date'].iloc[filter_dates[-3]]]

others_sa=others[others['location']=='South Africa'].groupby(['date']).sum()
comm=np.unique(np.hstack((delta_sa['date'],omicron_sa['date'])))
others_sa=others_sa[others_sa.apply(lambda x: x.name in comm, axis=1)]
fig = go.Figure()
fig.add_trace(go.
    ↳Scatter(x=omicron_sa['date'],y=omicron_sa["perc_sequences"],mode='lines',line_width=5,name="Del
    ↳perc_sequences"))
fig.add_trace(go.
    ↳Scatter(x=delta_sa['date'],y=delta_sa["perc_sequences"],mode='lines',line_width=5,name="Del
    ↳perc_sequences"))
fig.add_trace(go.Scatter(x=others_sa.
    ↳index,y=others_sa["perc_sequences"],mode='lines',line_width=5,name="Other_
    ↳perc_sequences"))
fig.update_layout(title="South Africa's perc_sequences of variants for each_
    ↳date")
fig.show()

```

The delta variant, from May 2021 and September 2021, became the dominant variant. Instead, the omicron variant was able to be the most diffused variant just more than a month and two weeks!

At last we can also assert that other variants are still diffused: in the 2021/12/27 6.17% of the samples belonged to other variants.

```

[13]: fig = go.Figure()
fig.add_trace(go.Scatter(x=omicron_sa['date'],y=np.
    ↳log1p(omicron_sa["perc_sequences"]),mode='lines+markers',line_width=5,marker_size=10,name="
    ↳perc_sequences logarithm"))
fig.add_trace(go.Scatter(x=delta_sa['date'],y=np.
    ↳log1p(delta_sa["perc_sequences"]),mode='lines+markers',line_width=5,marker_size=10,name="De
    ↳perc_sequences logarithm"))
fig.update_layout(title="South Africa's perc_sequences logarithm of variants_
    ↳for each date")
fig.show()

```

The logarithmic version of the previous plot confirms the fact that the omicron variant is far more contagious than the delta one: the delta variant had a slow and “unsure” grow, meanwhile the omicron variant had a very fast and brave grow.

7 Are there any differences in the diffusion of the variants in India?

India is also an interesting country because it’s where the delta variant was first found.

```

[14]: omicron_in=omicron[omicron["location"]=="India"]
omicron_in=omicron_in[omicron_in['perc_sequences']>0]

```

```

delta_in=delta[delta["location"]=="India"]
delta_in=delta_in[delta_in['perc_sequences']>0]

others_in=others[others['location']=='India'].groupby(['date']).sum()
comm=np.unique(np.hstack((delta_in['date'],omicron_in['date'])))
others_in=others_in[others_in.apply(lambda x: x.name in comm, axis=1)]
fig = go.Figure()
fig.add_trace(go.
    ↳Scatter(x=omicron_in['date'],y=omicron_in["perc_sequences"],mode='lines',line_width=5,name=
    ↳perc_sequences"))
fig.add_trace(go.
    ↳Scatter(x=delta_in['date'],y=delta_in["perc_sequences"],mode='lines',line_width=5,name="Del
    ↳perc_sequences"))
fig.add_trace(go.
    ↳Scatter(x=delta_in['date'],y=others_in["perc_sequences"],mode='lines',line_width=5,name="Ot
    ↳perc_sequences"))
fig.update_layout(title="India's perc_sequences of variants for each date")
fig.show()

```

In India the Omicron variant is not spreaded yet but we can notice how is taking the place of the delta variant in the month of december.

```

[15]: fig = go.Figure()
fig.add_trace(go.Scatter(x=omicron_in['date'],y=np.
    ↳log1p(omicron_in["perc_sequences"]),mode='lines+markers'
    ↳,line_width=5,marker_size=10,name="Omicron_
    ↳perc_sequences logarithm"))
fig.add_trace(go.Scatter(x=delta_in['date'],y=np.
    ↳log1p(delta_in["perc_sequences"]),mode='lines+markers'
    ↳,line_width=5,marker_size=10,name="Delta_
    ↳perc_sequences logarithm"))
fig.update_layout(title="India's perc_sequences logarithm of variants for each_
    ↳date")
fig.show()

```

This plot it's very similar to the one of South Africa and the observations we did before are still possible for India.