SI 618 Project Report

Omicron or Policy:

What caused 2022 China Covid Outbreak

Charles Song

songsc@umich.edu

December 6, 2022

Motivation

China has been stuck to Zero-COVID policy since the COVID-19 pandemic. It worked relatively well at first. However, people find that the number of covid cases has been rising for months recently. Some people claim that it is due to the decrease in entry quarantine time according to "Guidance for COVID-19 Control (9th Edition)"; while other people believe it is because the SARS-CoV-2 Omicron variant spread more quickly. I want to conduct a statistical analysis, trying to find out the pattern and reason of recent COVID case increasing.

Research Questions

1. How does the pattern of new COVID cases look like in China?

2. Is there a day of the week effect in China COVID cases?

3. Did omicron or policy change cause a significant difference to China COVID cases?

Data Sources

1. I Crawled covid case data from Oct 2020 to Dec 2022 from China National Health Commission. (http://www.nhc.gov.cn/xcs/yqtb/list_gzbd_36.shtml). The data is in csv format and includes these columns: date, total new cases, local new cases, imported new cases. 709 entries in total are included.

**Note: In this report, China refers to mainland China. Case number do not include non-symptom cases.*

2. Get COVID new cases data in 2022 of some other countries from Coronavirus (COVID-19) Cases - Our World in Data. After manipulating the data using *pandas*, it is now in csv format and concludes three columns: month, location, new cases. 95 entries in total.

Methods

1. How does the pattern of new COVID cases look like in China?

For this question, I convert the “date” column into pd.Timestamp type by using pd.to_datetime function.

New cases by day

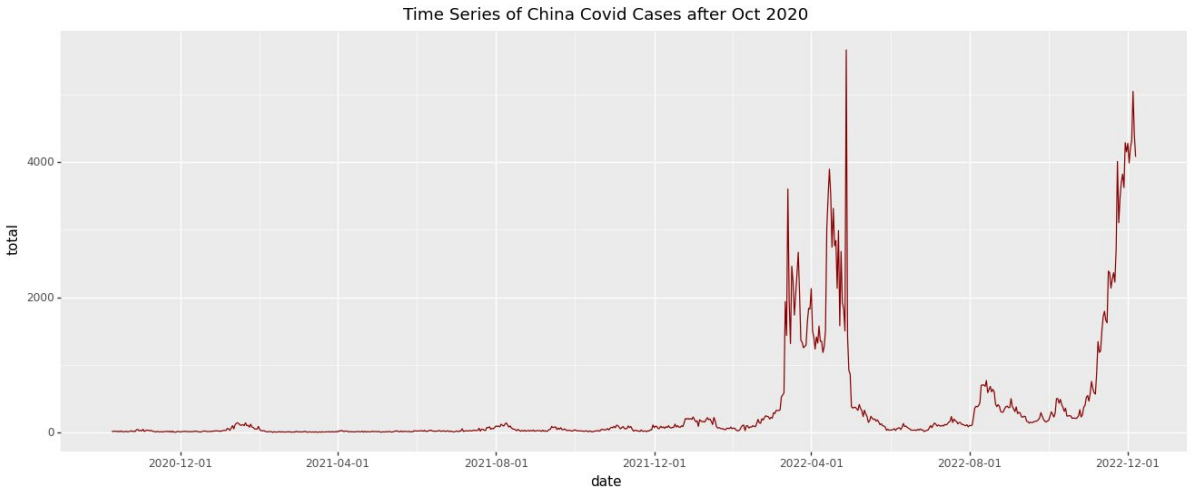

Figure 1 New cases by day.

According to the time series plot, we can roughly divide the data into three parts to analysis:

1. zero. Before 2022. This was when China's zero-covid policy took effect.
2. spring. Spring 2022 (Feb to May). This was a wave of covid in Shanghai.
3. winter. Winter 2022 (late Oct to now). This is a nation-wide wave of covid happening now and China is turning from zero-covid policy to "live with covid" policy.

We will analyse these three data separately or together to gain some insights.

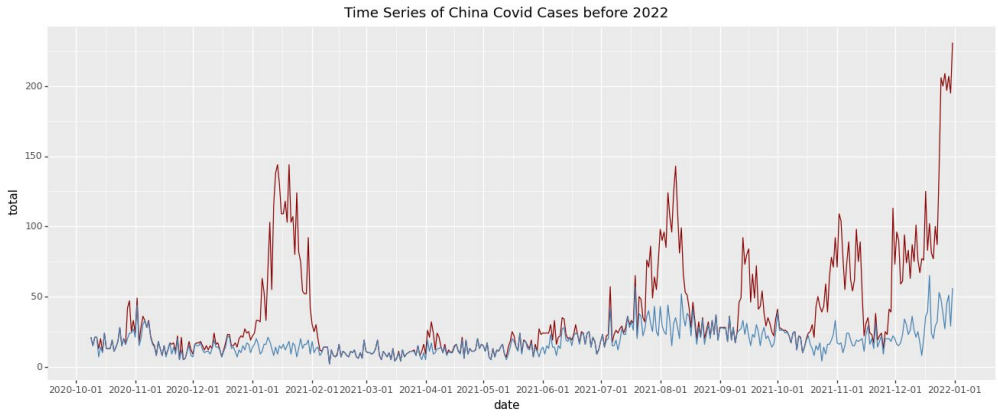

Figure 2 Local and Imported COVID Cases by day before 2022. Red line: local cases, blue line: imported cases.

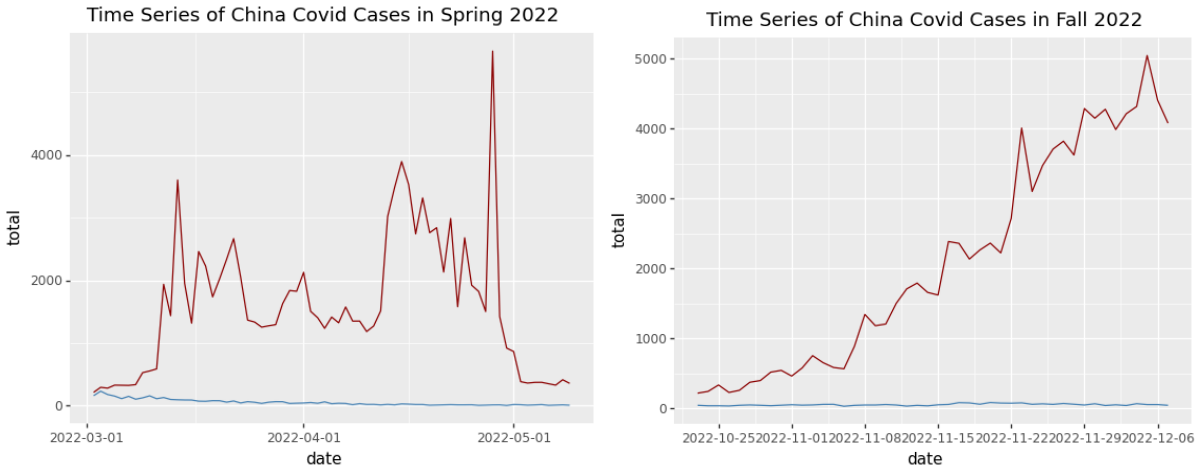

Figure 3 Left: Local and Imported COVID Cases by day in Spring 2022. Right: Local and Imported COVID Cases by day in Winter 2022. Red line: local cases, blue line: imported cases.

We also care about the "imported cases". Because the local cases are sometimes random events. But imported cases are the number of people entering China, China's policy of entering and COVID condition outside China. We can feel the influence of policy and COVID variant by analysing imported cases.

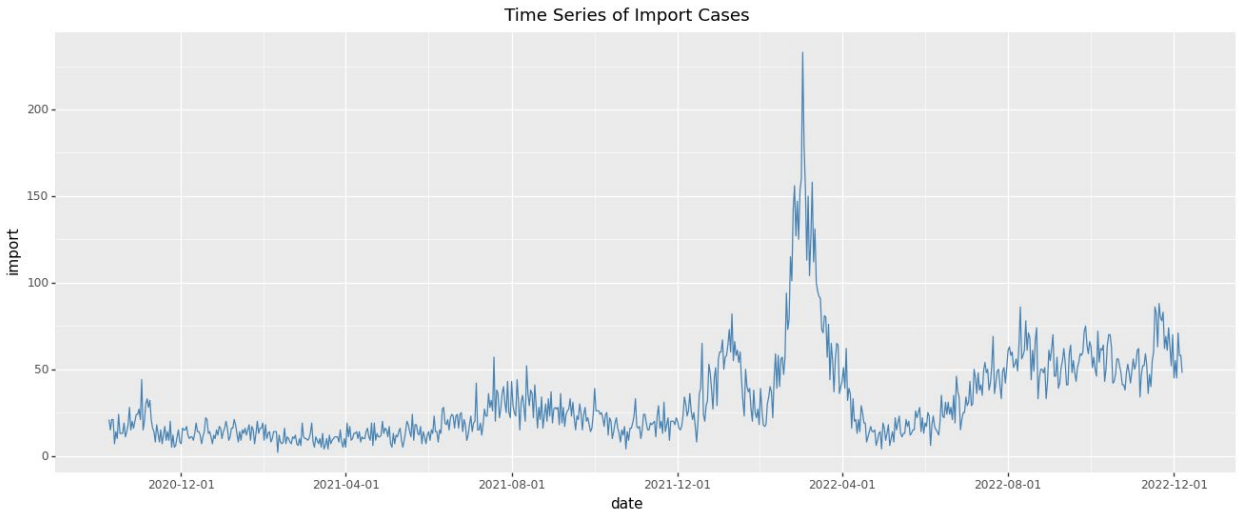

Figure 4 Imported COVID cases of China. Oct 2020 - Dec 2022

New cases by month

By adding a month column with `case['month'] = case['date'].dt.month`, we can get time series of COVID cases by month:

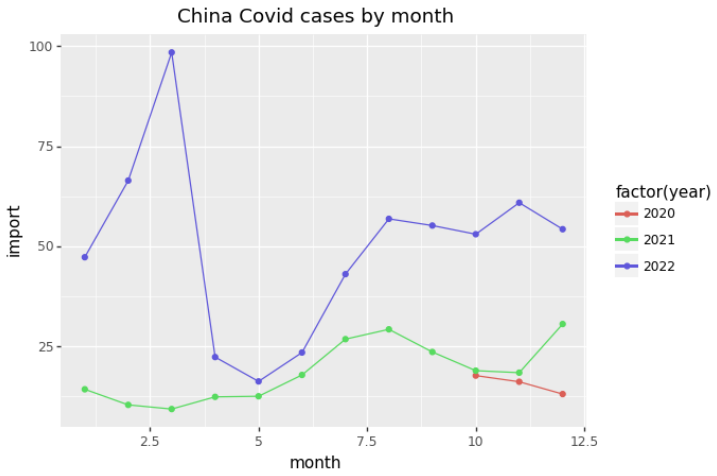

Figure 5 New COVID cases in China by month, Oct 2020 - Dec 2022.

Linear Regression Model: Seasonal Effect of COVID cases.

The spread ability of COVID is affected by temperature. Thus, to inspect the reason for the increase in covid cases, we need to eliminate the effect of season.

I use the COVID case data of some other countries and locations to build a linear regression model of new cases – month in 2022.

Table 1 OLS Linear Regression Result of new COVID cases - months in 2022

OLS Regression Results

Dep. Variable:	new_cases	**R-squared:**	0.085
Model:	OLS	**Adj. R-squared:**	-0.036
Method:	Least Squares	**F-statistic:**	0.7045
Date:	Thu, 08 Dec 2022	**Prob (F-statistic):**	0.731
Time:	11:12:47	**Log-Likelihood:**	-1488.2
No. Observations:	95	**AIC:**	3000.
Df Residuals:	83	**BIC:**	3031.
Df Model:	11		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	1.133e+06	6.22e+05	1.821	0.072	-1.04e+05	2.37e+06
C(month)[T.2]	2.071e+05	8.52e+05	0.243	0.808	-1.49e+06	1.9e+06
C(month)[T.3]	6.361e+05	8.52e+05	0.747	0.457	-1.06e+06	2.33e+06
C(month)[T.4]	-2.157e+05	8.52e+05	-0.253	0.801	-1.91e+06	1.48e+06
C(month)[T.5]	-2.745e+05	8.52e+05	-0.322	0.748	-1.97e+06	1.42e+06
C(month)[T.6]	-3.455e+05	8.52e+05	-0.406	0.686	-2.04e+06	1.35e+06
C(month)[T.7]	1.633e+05	8.52e+05	0.192	0.848	-1.53e+06	1.86e+06
C(month)[T.8]	6.605e+05	8.52e+05	0.776	0.440	-1.03e+06	2.35e+06
C(month)[T.9]	-2.246e+05	8.52e+05	-0.264	0.793	-1.92e+06	1.47e+06
C(month)[T.10]	-5.476e+05	8.52e+05	-0.643	0.522	-2.24e+06	1.15e+06
C(month)[T.11]	-2.844e+05	8.52e+05	-0.334	0.739	-1.98e+06	1.41e+06
C(month)[T.12]	-1.089e+06	8.52e+05	-1.279	0.205	-2.78e+06	6.05e+05

Then, I made a deseasonified time series analysis.

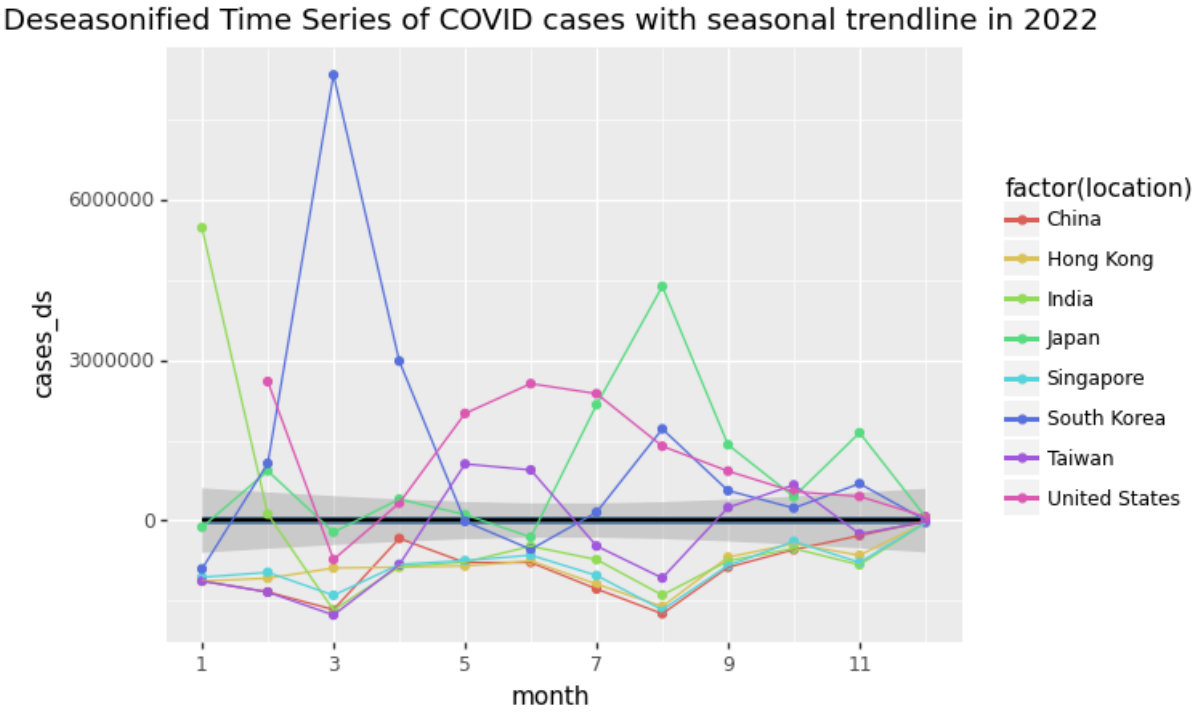

Figure 6 Deseasonified Time Series of COVID cases with seasonal trendline in 2022

After removing the seasonal pattern, we can see there's a rise in China's residual new cases starting from August.

2. Is there a day of the week effect in China COVID cases?

We are going to check the day of the week effect Hypotheses for the day of the week effect.

1. There are more cases in the weekends. The covid case increase is related to day of the week. For example, in weekend, people might tend to hang out together.

2. There are more cases in the weekdays. There's a statistics error pattern related to day of the week. For example, there are fewer testing staff during weekend, resulting in fewer cases detected.

I added a “dayofweek” column, and group the data by it.

Table 2 New COVID Cases by day of the week

dayofweek	total	import	local
5	321.097345	32.247788	288.849558
6	324.345133	32.539823	291.805310
4	335.008850	32.973451	302.026549
1	339.964602	29.238938	310.725664
0	344.619469	32.300885	312.318584
2	354.955752	29.964602	324.991150
3	359.303571	32.535714	326.767857

The local cases on Saturday and Sunday do appear fewer than other days. We can use a T-test to see if the difference is statistically significant.

```
stats.ttest_ind(case[case['day']=='weekdays']['local'], case[case['day']=='weekends']['local'])
```

```
Ttest_indResult(statistic=0.40716330762369735, pvalue=0.683998625000753)
```

P value is not less than 0.05, we cannot reject the null hypothesis of "there is no difference in weekdays and weekends". The local cases on the weekend are the same as those on weekdays.

3. Did omicron or policy change cause a significant difference to China COVID cases?

First, let's see if imported cases are related to local cases.

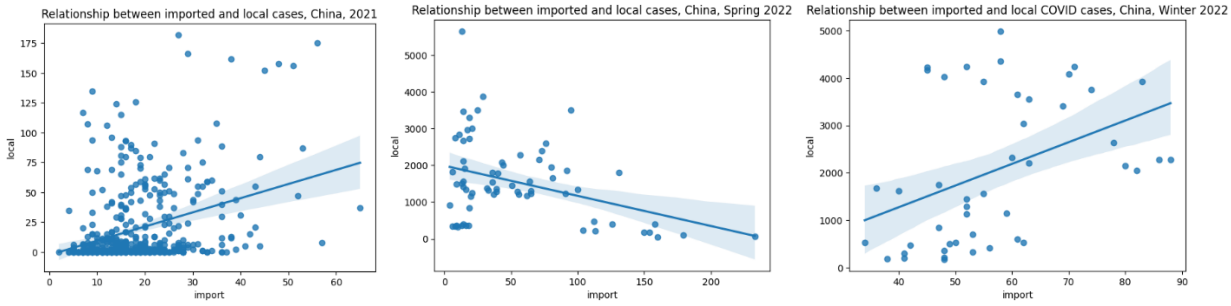

Figure 7 Relationships between imported and local cases in China. Left: before 2022; Middle: Spring 2022; Right: Winer 2022

From the plots above, we can see that there is a positive relationship between import and local cases before 2022 and in winter 2022, indicating the possibility that more import cases will result in more local cases. However, in the wave of spring 2022, there is a negative relationship between import and local cases. I don't know the reason why.

However, correlation does not imply causation. There is a positive relationship between imported and local cases does not mean the increase in imported cases results in the increase in local cases for sure. We still need more analysis.

Then, the t-tests for the influence of Omicron vs the influence of "9th Edition".

T test for the influence of Omicron

```
stats.ttest_ind(case[case['date'] > '2021-12-13']['total'], case[case['period']=='before omicron']['total'])
```

```
Ttest_indResult(statistic=13.251917056061068, pvalue=2.4797634017006744e-36)
```

```
stats.ttest_ind(case[case['date'] > '2021-12-13']['import'], case[case['period']=='before omicron']['import'])
```

```
Ttest_indResult(statistic=20.50117036200962, pvalue=3.440140753370568e-75)
```

Both p value << 0.05, both Null Hypotheses are rejected.

There is a significant difference in the total covid case between before and after China's first omicron case.

There is a significant difference in the imported covid case between before and after China's first omicron case.

T test for the influence of "9th Edition"

```
stats.ttest_ind(case[case['period']=='omicron, before 9th edition']['total'], case[(case['period']=='omicron, after 9th edition')]['total'])
```

```
Ttest_indResult(statistic=-1.3789576760655198, pvalue=0.1687711350950561)
```

```
stats.ttest_ind(case[case['period']=='omicron, before 9th edition']['import'], case[(case['period']=='omicron, after 9th edition')]['import'])
```

```
Ttest_indResult(statistic=-2.6778173091489252, pvalue=0.007752735351322857)
```

The p value of total case is greater than 0.05; while the p value of imported case is smaller than 0.05.

There **is not** a statistically significant difference in the **total** covid case between before and after "Guidance for COVID-19 Control (9th Edition)" was published.

However, there **is** a statistically significant difference in the **imported** covid case between before and after "Guidance for COVID-19 Control (9th Edition)" was published.

**Note: To control the variable, we only use the covid case data after China's first omicron case to analysis the influence of "9th edition"*

Joint plot of the imported and local cases.

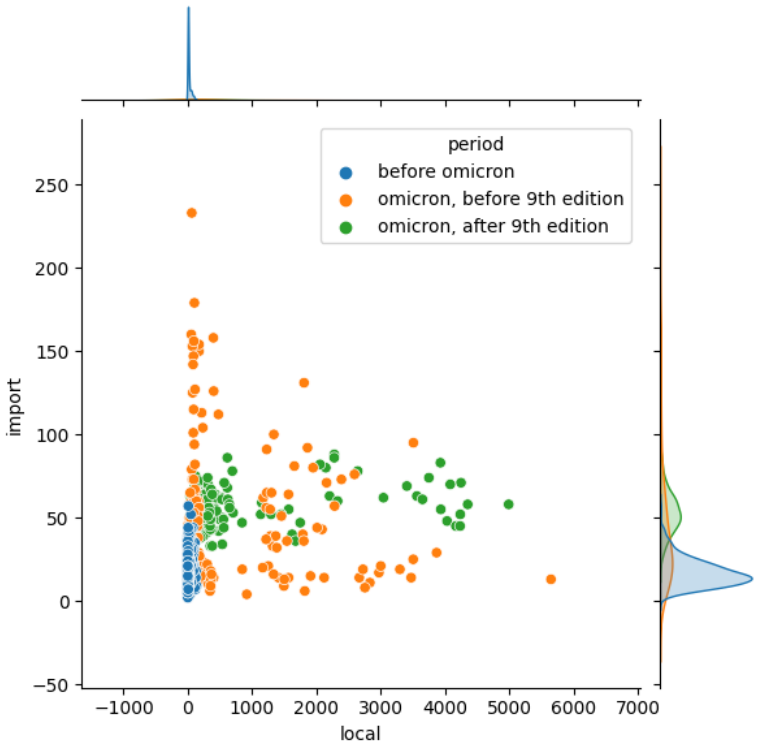

Figure 8 Joint plot of the imported and local cases, classified by omicron and "9th Edition".

Conclusion: Omicron causes the China COVID outbreak in 2022. "9th Edition" does increase imported cases but not total. I guess this is because imported cases are test positive during quarantine, so they are not the reason of the local case outbreak – That's just the omicron.