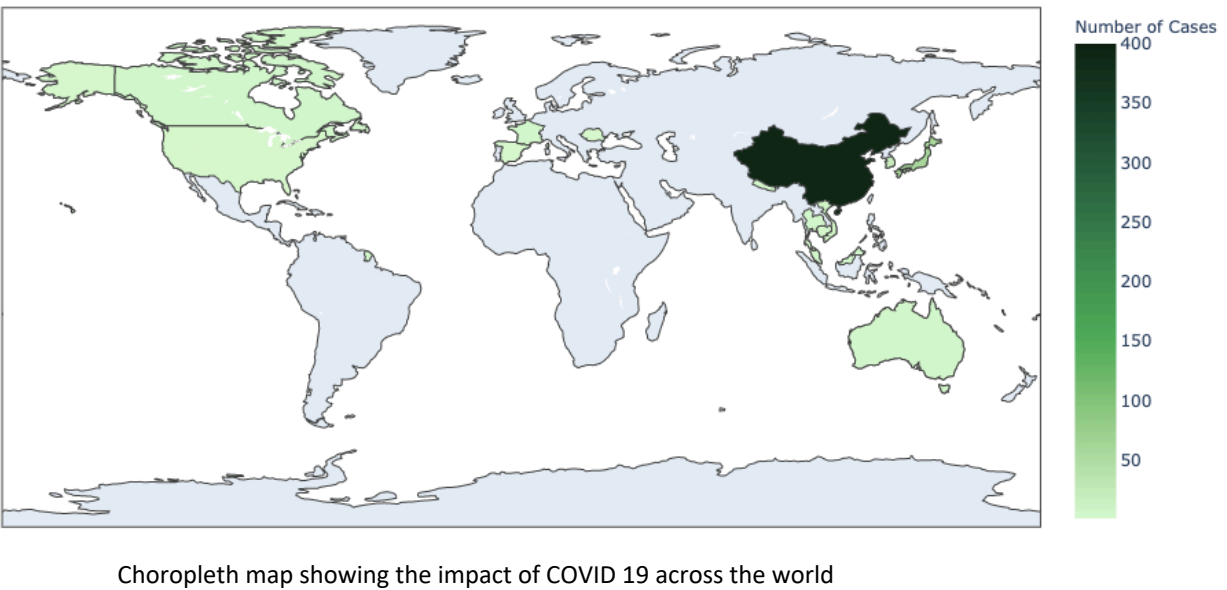
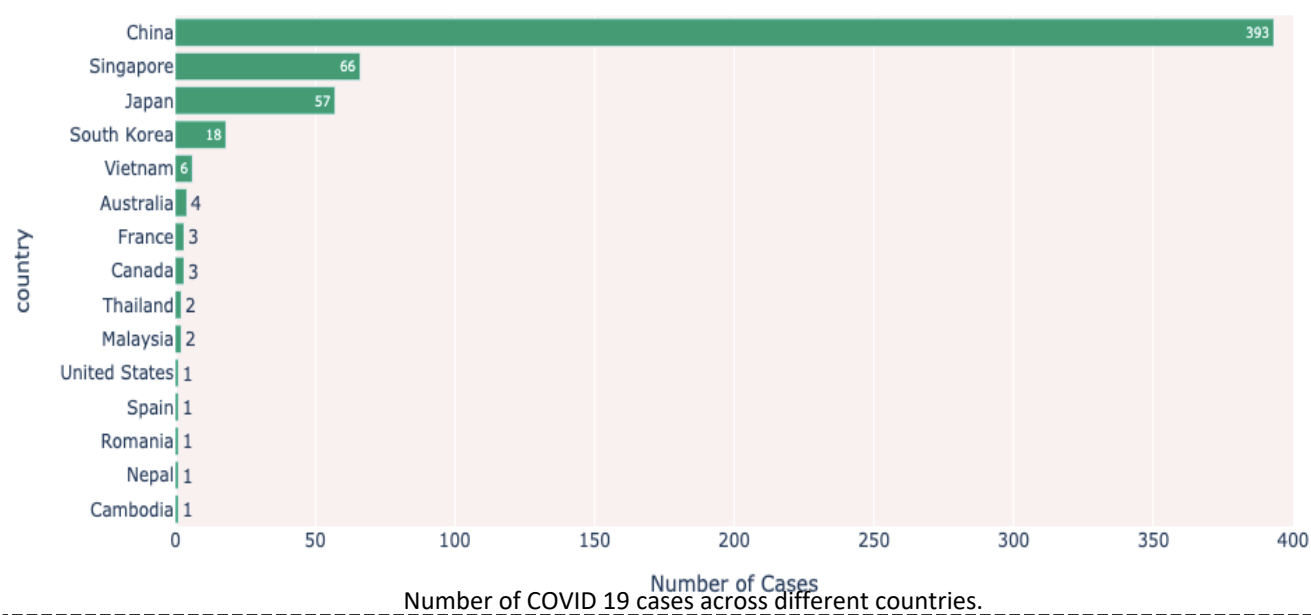


# Analysis of a COVID 19 Dataset

Santosh Nanduru

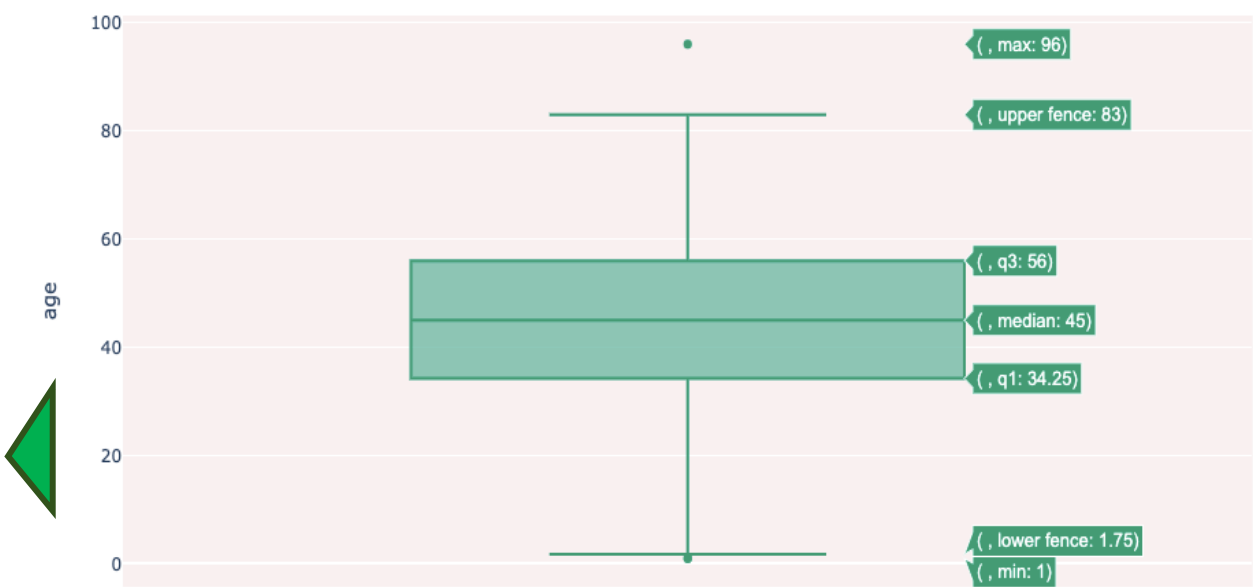
China has the greatest number of COVID 19 cases followed by Singapore and Japan. The COVID 19 seems to be highly concentrated in the Asian region. The percentage of males affected by the virus is 16% more than that of females. The median age of COVID 19 victim is 45.



Gender	Number of Cases	Percentage of Cases
female	237	42
male	322	58

Total Percentage of Female and Male COVID 19 patients

- A person of age 96 is the oldest person to contact COVID 19 virus.
- A child of age 1 is the youngest person to contact COVID 19 virus.



Box plot depicting outliers and median age of COVID 19 patients.

Around 200 people travelled to/from Wuhan during this period and 24 individuals experienced both fever and respiratory symptoms\*. Seoul and Nagoya city from Japan are the cities with maximum patients outside the Chinese border\*\*. No country has more proportion of female patients.

country	female	male	Total People
Australia	1	3	4
Cambodia	0	1	1
Canada	1	2	3
France	1	2	3
Japan	19	38	57
Malaysia	1	1	2
Nepal	0	1	1
Romania	1	0	1
Singapore	29	37	66
South Korea	7	11	18
Spain	0	1	1
Thailand	1	1	2
United States	0	1	1
Vietnam	3	3	6
Total People	64	102	166



country	femaleproportion	maleproportion
Australia	0.25	0.75
Cambodia	0	1
Canada	0.33	0.67
France	0.33	0.67
Japan	0.33	0.67
Malaysia	0.5	0.5
Nepal	0	1
Romania	1	0
Singapore	0.44	0.56
South Korea	0.39	0.61
Spain	0	1
Thailand	0.5	0.5
United States	0	1
Vietnam	0.5	0.5
Total People	0.39	0.61

Coutry/Province	Number of Travelers
Wuhan	199
Hubei	15
Thailand	8
Malaysia	5
Guangzhou	5
no	4
Japan	4
Beijing	4
Jiangxi	3

Table showing the number of people traveling through various ports

City	Number of cases
Nagoya City	7
Seoul	7

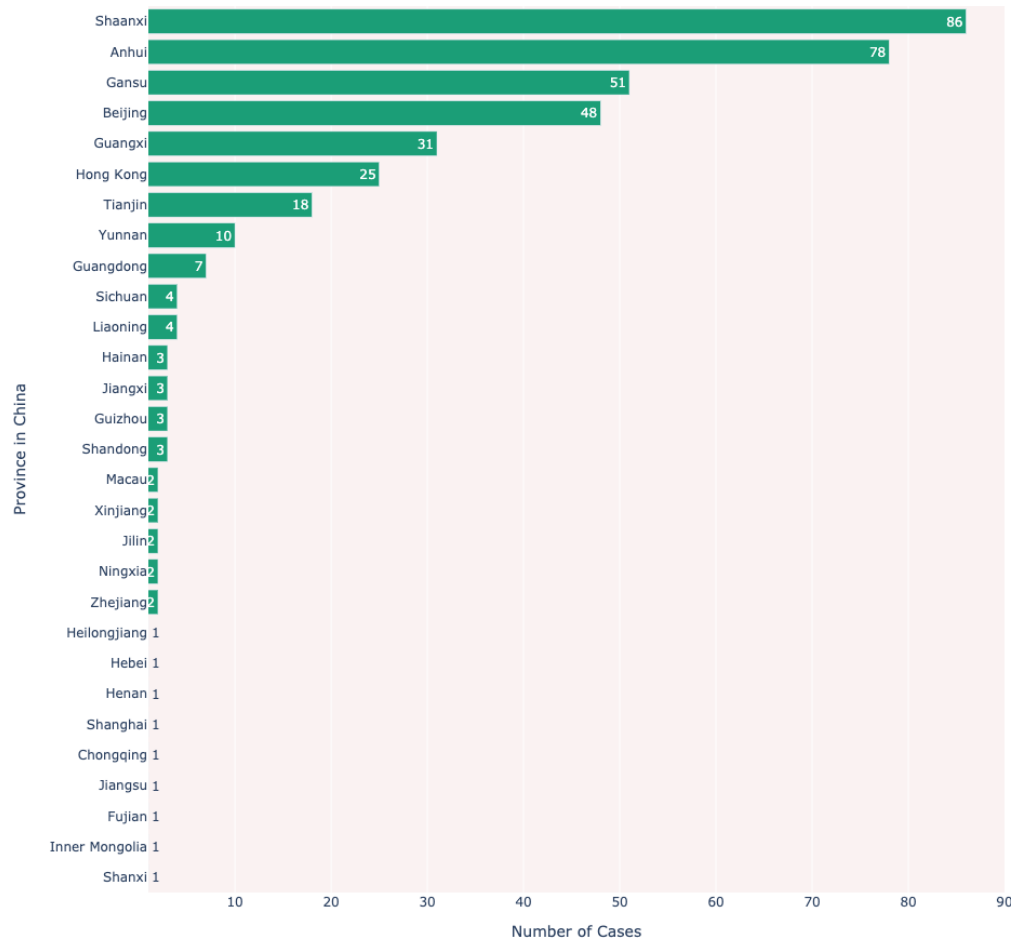
Cities with maximum cases outside China

Tables depicting the number of Females and Males affected across different countries

\*Respiratory symptoms include Rhinorrhea, Expectoration, Pharyngeal Discomfort, Pneumonia, Pharyngalgia, Sputum, Shortness of breath, Dyspnea, Pleuritic chest pain, Pleural effusion, Pneumonitis.

\*\* Singapore is a city-country. Even though it has the highest number of cases, it has been considered as a country for this statistic.

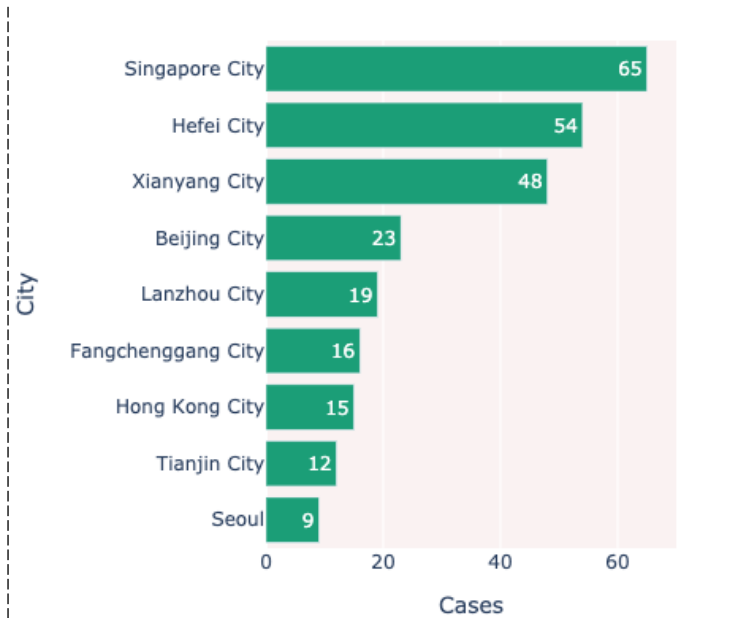
People should avoid travelling to Beijing city and opt to travel to Zhengzhou city instead. Shaanxi province is the worst effected province in China. Singapore city\* has the highest number of COID 19 cases followed by Hefei City and Xianyang City.



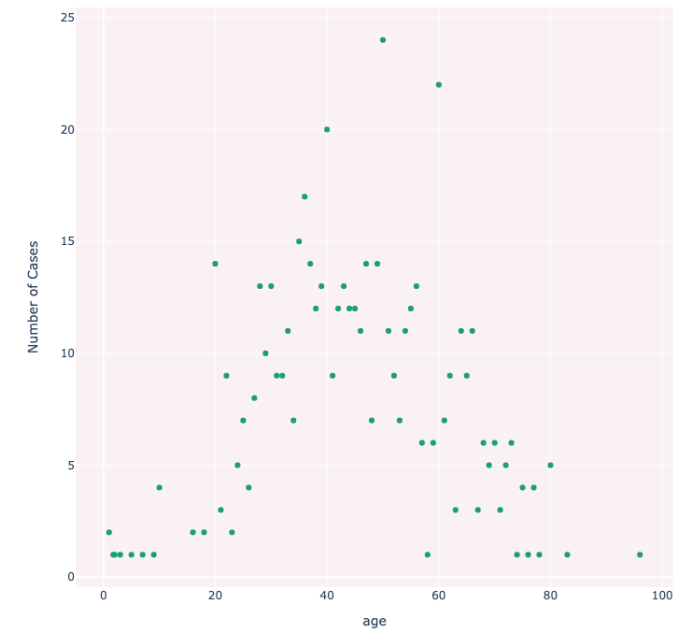
Graph depicting the number of cases across different provinces in China.



Provinces close to Hubei/ Wuhan show maximum number of cases.



Graph depicting the number of cases in Top 10 Cities of the World



Graph depicting the number of cases across different age groups.

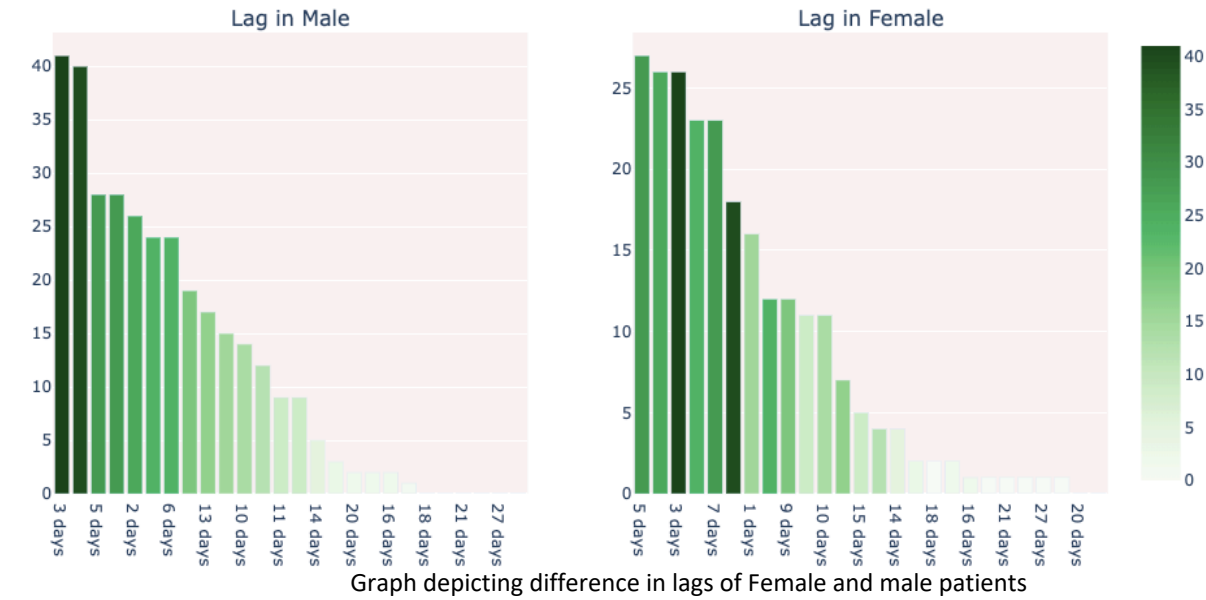


The age group between 40 to 60 is the worst hit age group with 204 cases registered in this age group.

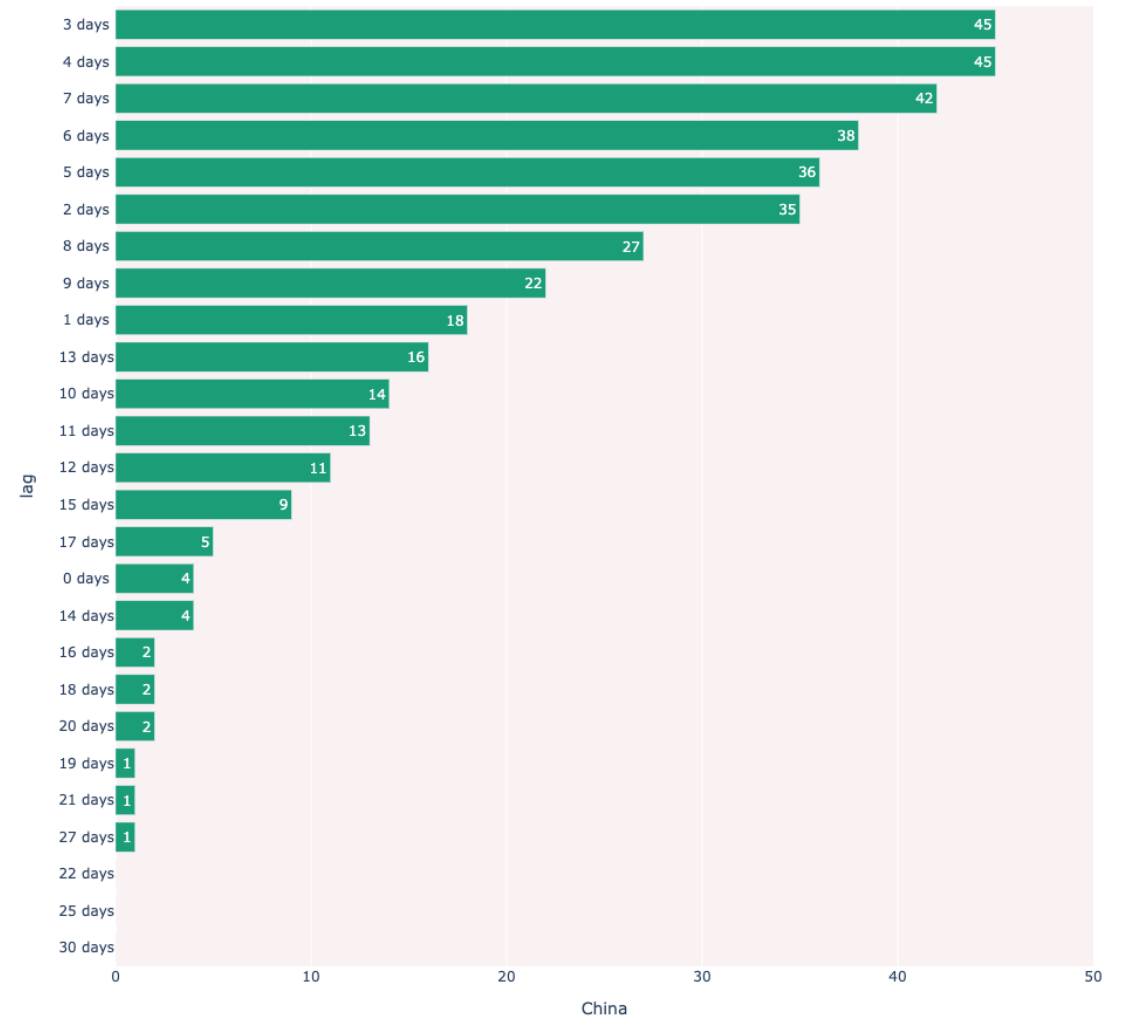
\*Singapore is a city country. It has been accounted into this table considering the size of the country.

The lag\* in Females is higher than that of males implying that Females are more resistant to virus. In China, around 90 people were hospitalized within 4 days from the inception date. The time series data clearly shows the emergence of cases around 2<sup>nd</sup> Feb in countries like Thailand.

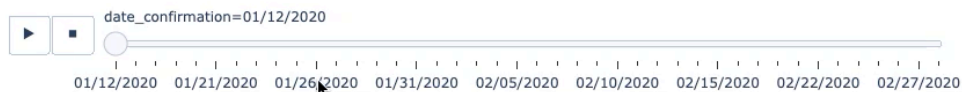
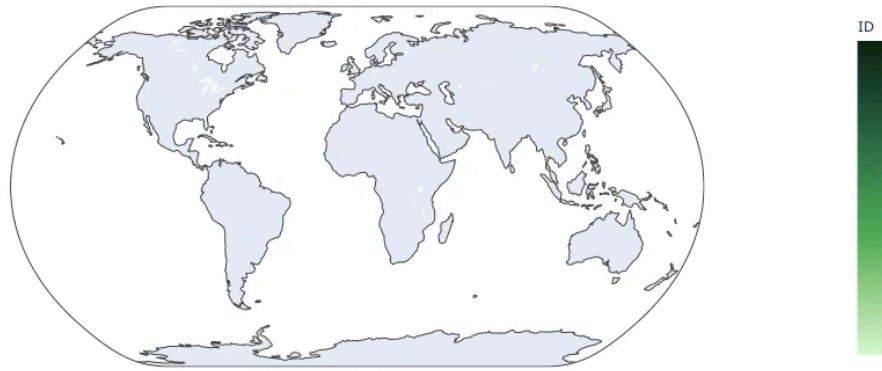
Lag in Females and Males



Lag in China



Graph depicting lag in Chinese patients



Time Series clip showing the first wave and second wave of Coronavirus cases in China and Asia

\* Lag is the difference in the number of days a patient takes to test positive for coronavirus from the day the symptoms begin.

A shutdown 6 days before 23<sup>rd</sup> February\* would have helped Singapore, Japan, United States, Australia and France to avoid their Index cases. Of a total 393 cases in China, 142 cases\*\* are linked to someone who has traveled to Wuhan / Hubei. Young people of Japan traveled to China but the elderly faced the brunt of the virus.

ID	age	sex	city	province	date_onset_symptoms	date_confirmation	symptoms	travel_history_dates	travel_history_location
8904	42.0	female	Longxi County, Dingxi City	Gansu	2020-01-28	2020-02-05	fever, cough	2020-01-21	Thailand
8905	26.0	female	Longxi County, Dingxi City	Gansu	2020-02-01	2020-02-05	fever, cough	2020-01-21	Thailand
8914	54.0	female	Longxi County, Dingxi City	Gansu	2020-01-21	2020-02-07	dry throat; sore throat; fatigue	NaT	Thailand
8915	29.0	female	Longxi County, Dingxi City	Gansu	2020-01-24	2020-02-07	headache; dry mouth	NaT	Thailand

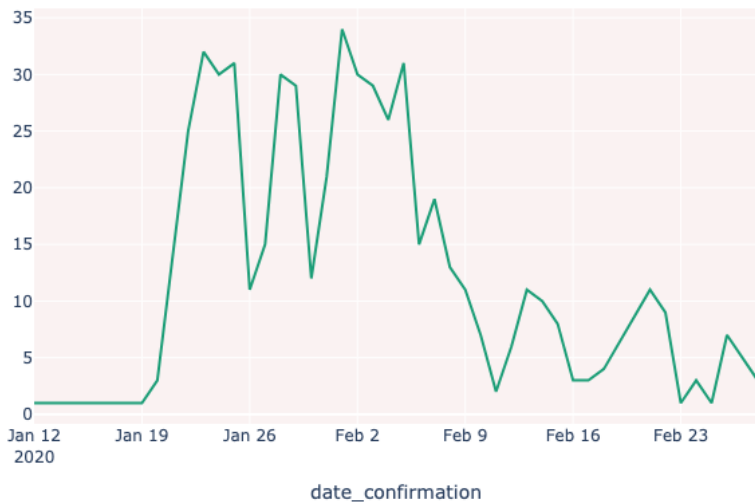
Data showing 4 females traveling out of Thailand and entering China.



A strong reason to hypothesize that an all girls' trip to Thailand from Dingxi City, China proved costly for them.

Country	Date of Travel	Date of Onset of Symptoms
France	19/01/2020	19/01/2020
Australia	06/01/2020	15/01/2020
United States of America	15/01/2020	16/01/2020
Singapore	18/01/2020	21/01/2020

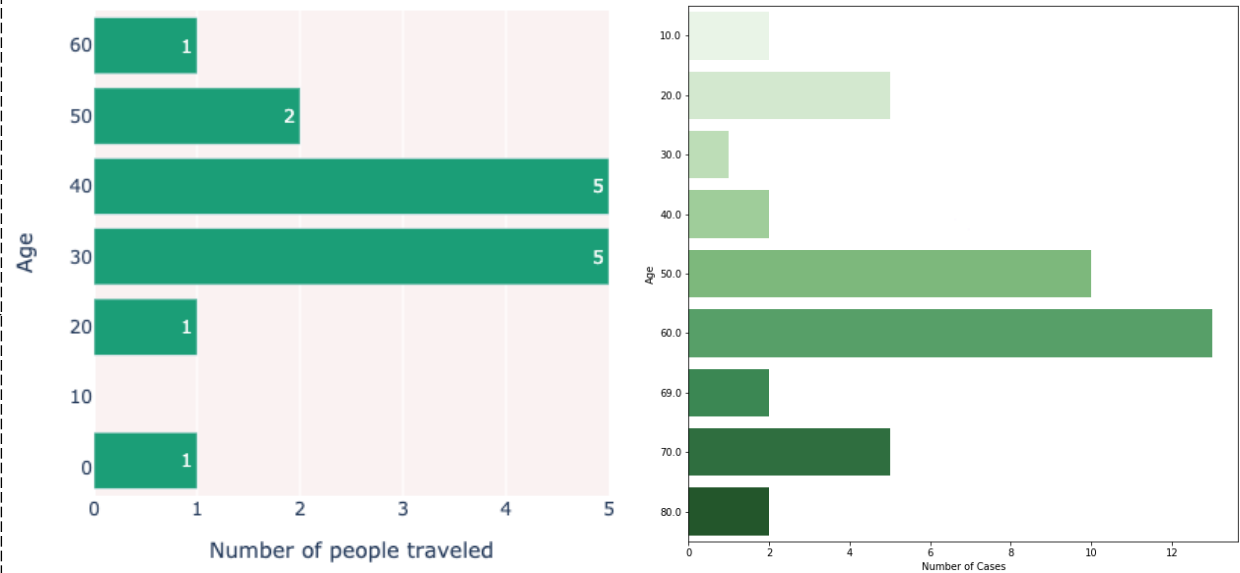
Index cases of different countries.



Graphs depicting the change In number of cases with time



Data shows that the initial clampdown in China worked. However, there was a sudden surge in cases due to rapid testing from January 26 confirming community transmission had already taken place.



Graphs depicting the age group traveling into/out of Japan and the actual impact in the elderly of Japan



Young people can be passive carriers of the virus and the virus strongly impacts the elderly

\* The actual date of shutdown in China.

\*\* Number of cases increase within a province with increase in proximity to Wuhan/ Hubei. The closest province Anhui records the highest number of cases.