

Covid 19, India Analysis

Name: Gurmeet Singh

Roll no : 2017CSC1030

GitHub

lnk : <https://github.com/gurmeeetsingh0111/Data-Analysis-project/>

Agenda / Topics

Introduction

Datasets

Data Visualization/Analysis

Research problem and conclusions

Future Scope of work

- The aim of this kernel is to understand COVID-19 India better, by implementing/viewing the data graphically
- The COVID-19 pandemic is the defining global health crisis of our time and the greatest global humanitarian challenge the world has faced since World War II.
- The virus has spread widely, and the number of cases is rising daily as governments work to slow its spread. India has moved quickly, implementing a proactive, nationwide, 21-day lockdown, with the goal of flattening the curve and using the time to plan and resource responses adequately.
- The lockdown is also extended to 3rd May as per recent orders

Introduction

Datasets

Source for the data

sets :Covid19India <https://docs.google.com/spreadsheets/d/e/2PACX-1vSc_Zy5N0I67wDU38DjDh35lZSlS30rQf7_NYZhtYYGU1jJYT6_kDx4YpF-qw0LSIGsBYP8pqM_a1Pd/pubhtml>

We Extracted "*IndividualDetails.csv*" from the given the site and also computed a small file "*AgeGroupDetails.csv*" using the first file

We will use these csv files in order to analys the data.

NOTE : After cleaning the data by removing the `NA` values and unnecessary coloumns , the data remained is just 12.35% of the original data

This dataset has daily level information on the number of affected cases, deaths and recovery from 2019-20 novel coronavirus. Please note that this is a time series data and so the number of cases on any given day is the cumulative number.

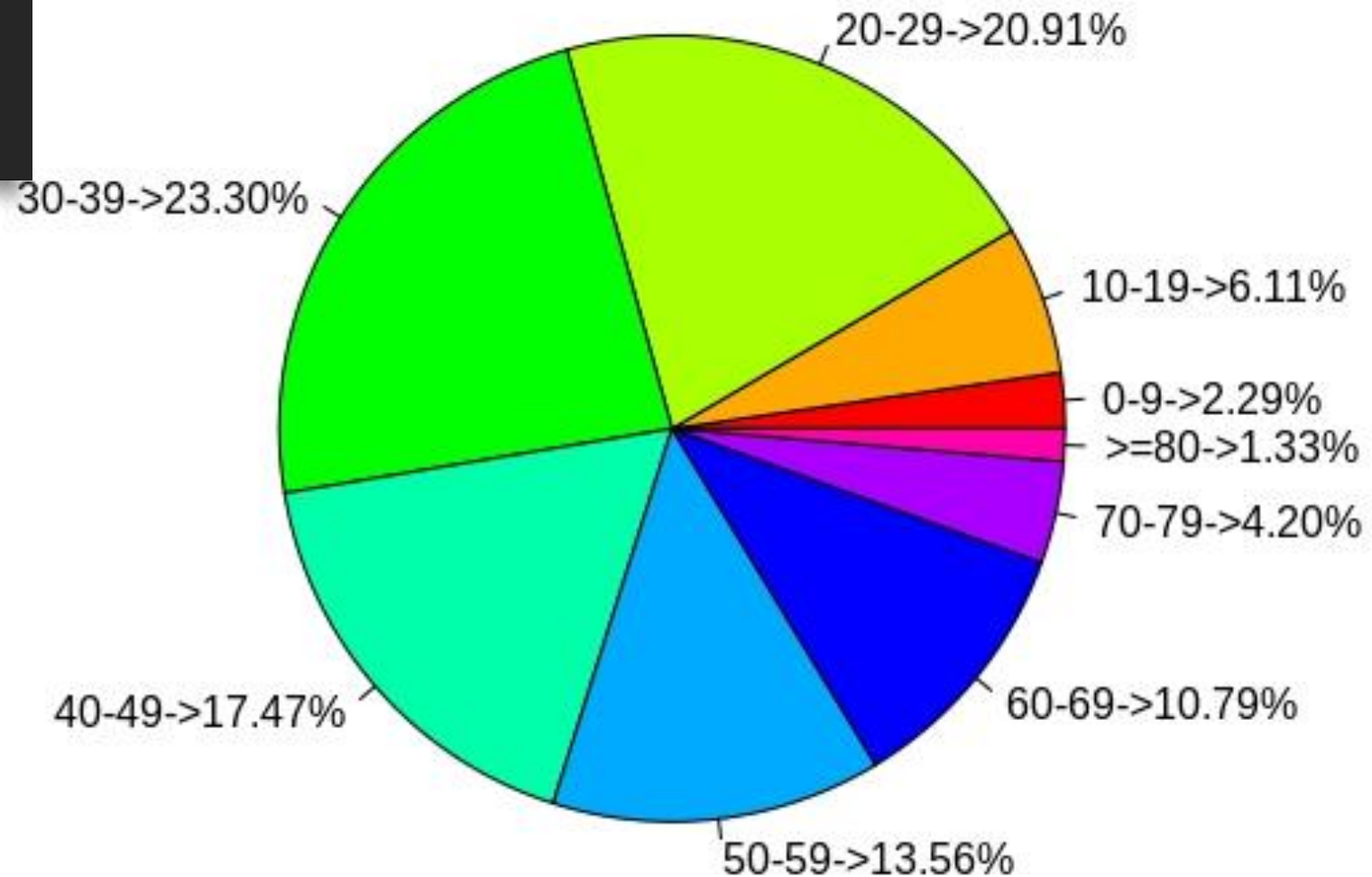
The data encourage the analysis of time series, which will allow us to look at the studied phenomenon not only in a point-like manner.

Data Visualization/Analysis

Age Group Analysis

- We can clearly see that the age group <50 is the most affected which is against the trend which says elderly people are more at risk of being affected. Only 16% of people >60 are affected.
- One of the main reason could be that, because young people meet and travel much more often than older people so they were under the threat of community transmission more than the others

Age group wise distribution



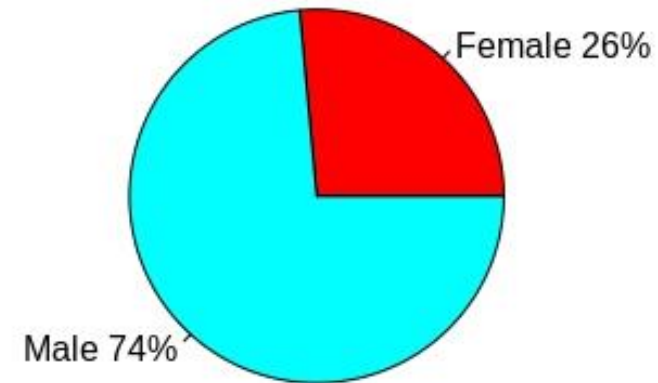
Results of the analysis done may vary than the original results as earlier mentioned in slide 4 under "note" point that after cleaning , 87.65% of the data was lost due to incomplete rows or columns in the original data

Data Visualization/Analysis , continued...

Percentage distribution Gender Wise

- Here we could see the relationship between the two genders which depicts no. Of cases in percenatge manner.
- This Tells that Female patients were less affected by the virus than the male patients.
- Men were the most affected accounting to 74%

Gender % of affected people

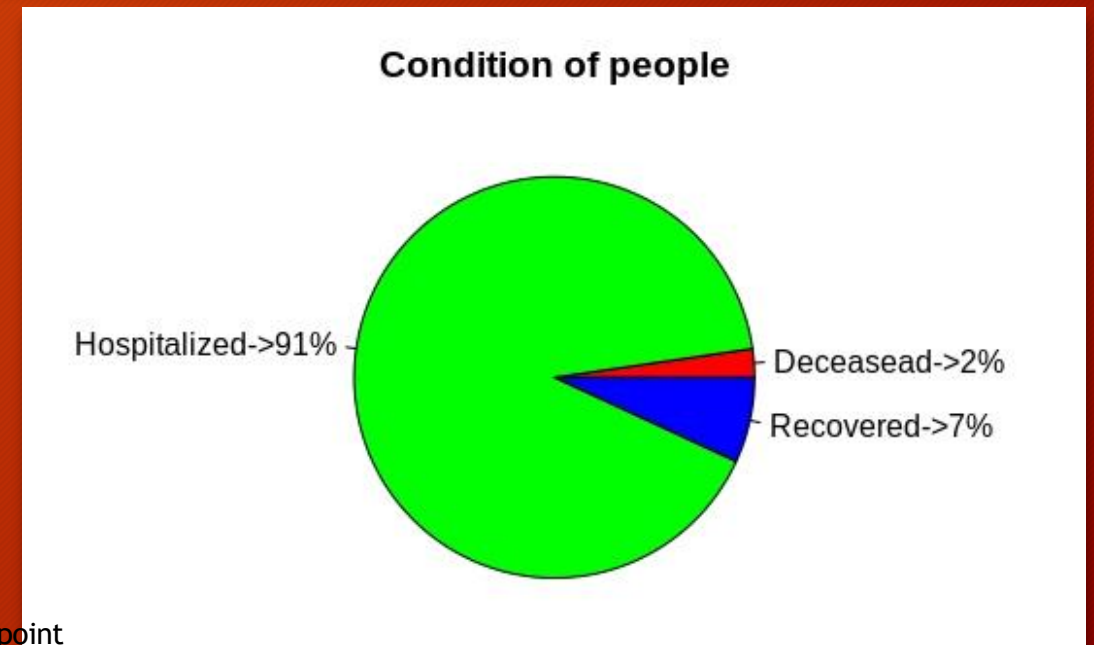
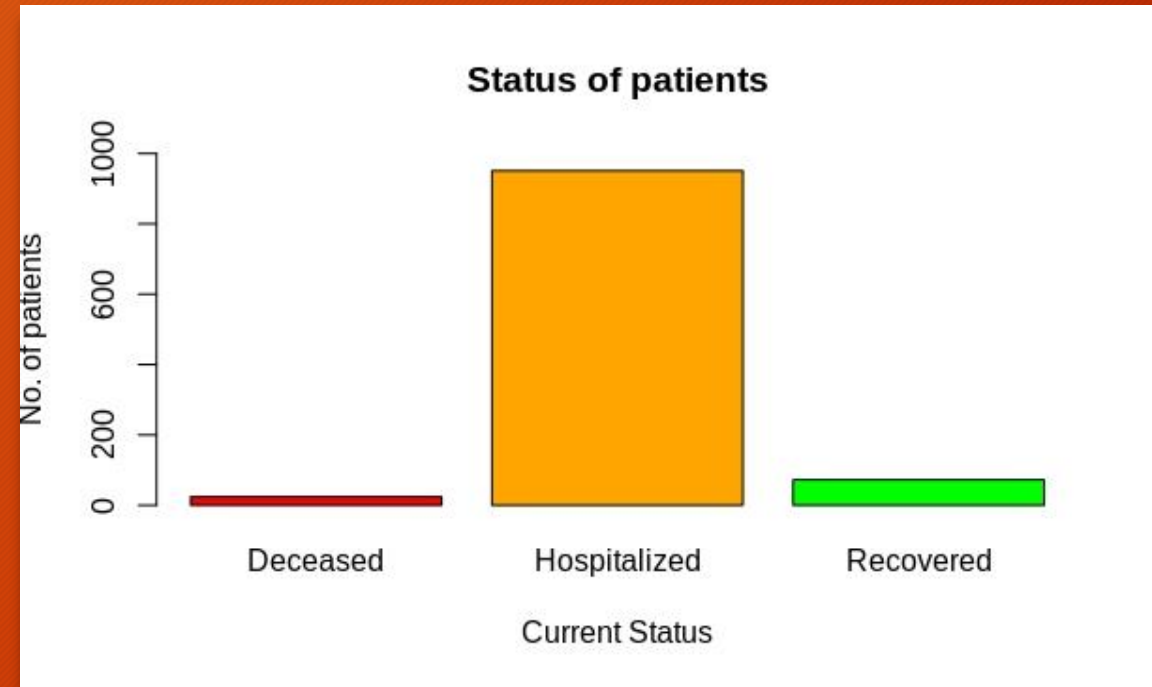


Results of the analysis done may vary than the original results as earlier mentioned in slide 4 under "note" point that after cleaning , 87.65% of the data was lost due to incomplete rows or columns in the original data

Data Visualization/Analysis , continued...

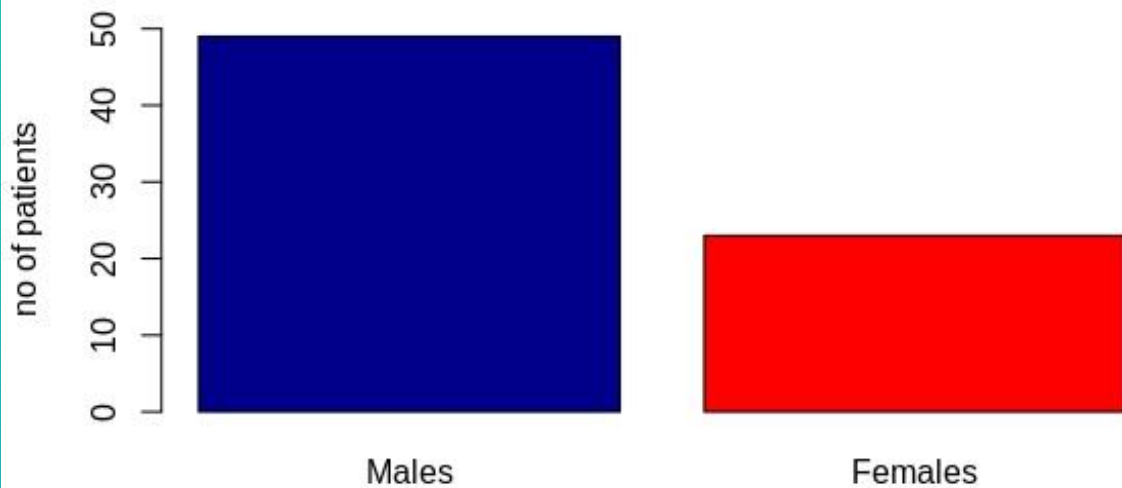
Current status of patients

- Most of them are still held in quarantine
- No. Of people Deceased = 24 , 2%
- No. Of people Hospitalized = 951 , 91%
- No. Of people Recovered = 72 , 7%

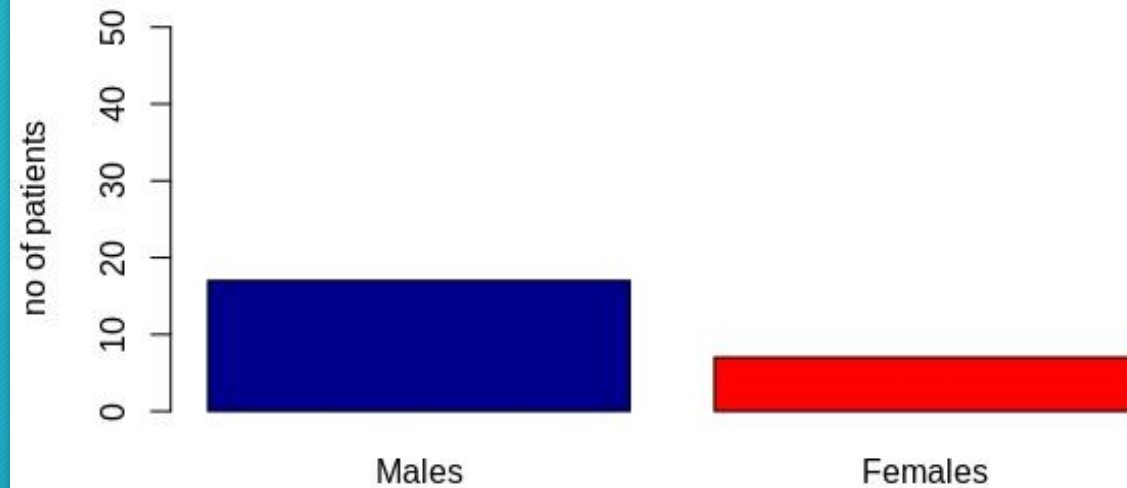


Results of the analysis done may vary than the original results as earlier mentioned in slide 4 under "note" point that after cleaning , 87.65% of the data was lost due to incomplete rows or columns in the original data

People Recovered



People Deceased



Gender Wise Patient Status

Data Visualization/Analysis , continued...

Results of the analysis done may vary than the original results as earlier mentioned in slide 4 under "note" point that after cleaning , 87.65% of the data was lost due to incomplete rows or columns in the original data

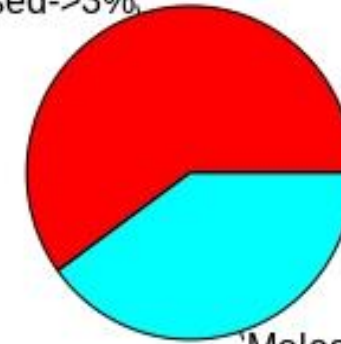
Data Visualization/Analysis , continued...

Gender Wise Patient Status , Contd...

- 3% of the total Females affected were Deceased i.e. '7' but as you can see only 2% of the total men affected were deceased i.e. 17
- And 8% of the total Females affected were Recovered i.e. '23' and only 6% of the total men affected were recovered i.e. '49'

Deceased percenatge of females vs males

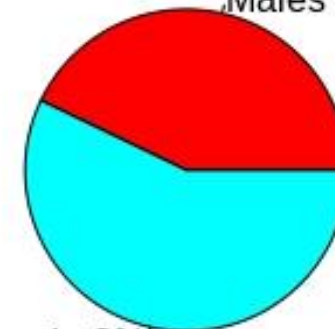
Females Deceased->3%



Males Deceased->2%

Recovery percenatge of females vs males

Males Recovered->6%

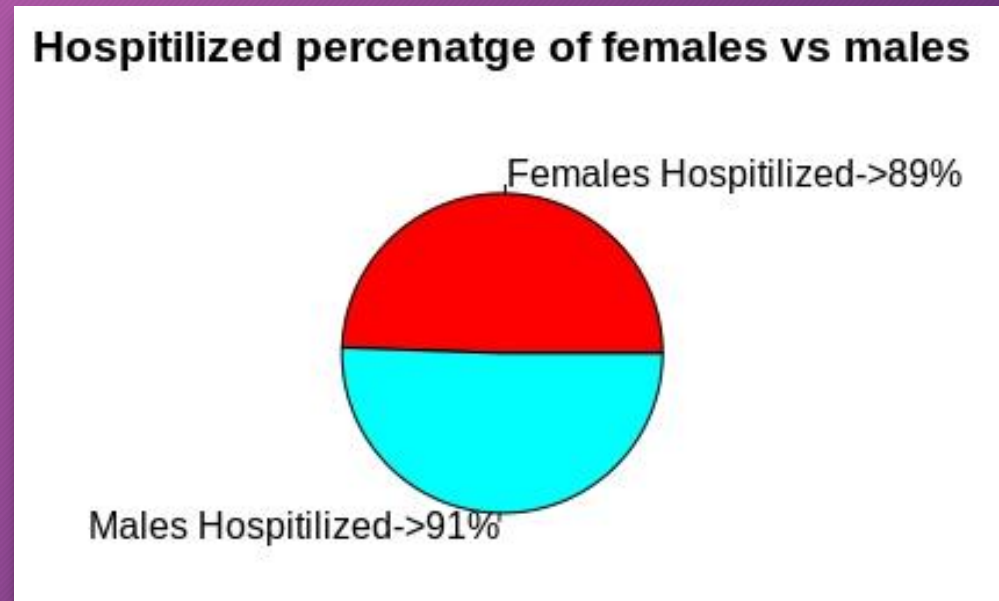


Females Recovered->8%

Data Visualization/Analysis , continued...

Quarantine Status

- A quarantine is a restriction on the movement of people and goods which is intended to prevent the spread of disease or pests. It is often used in connection to disease and illness, preventing the movement of those who may have been exposed to a communicable disease, but do not have a confirmed medical diagnosis
- 91% of the total men affected and 89% of the total Females affected are still kept in quarantine to slow down the community transmission as much as possible

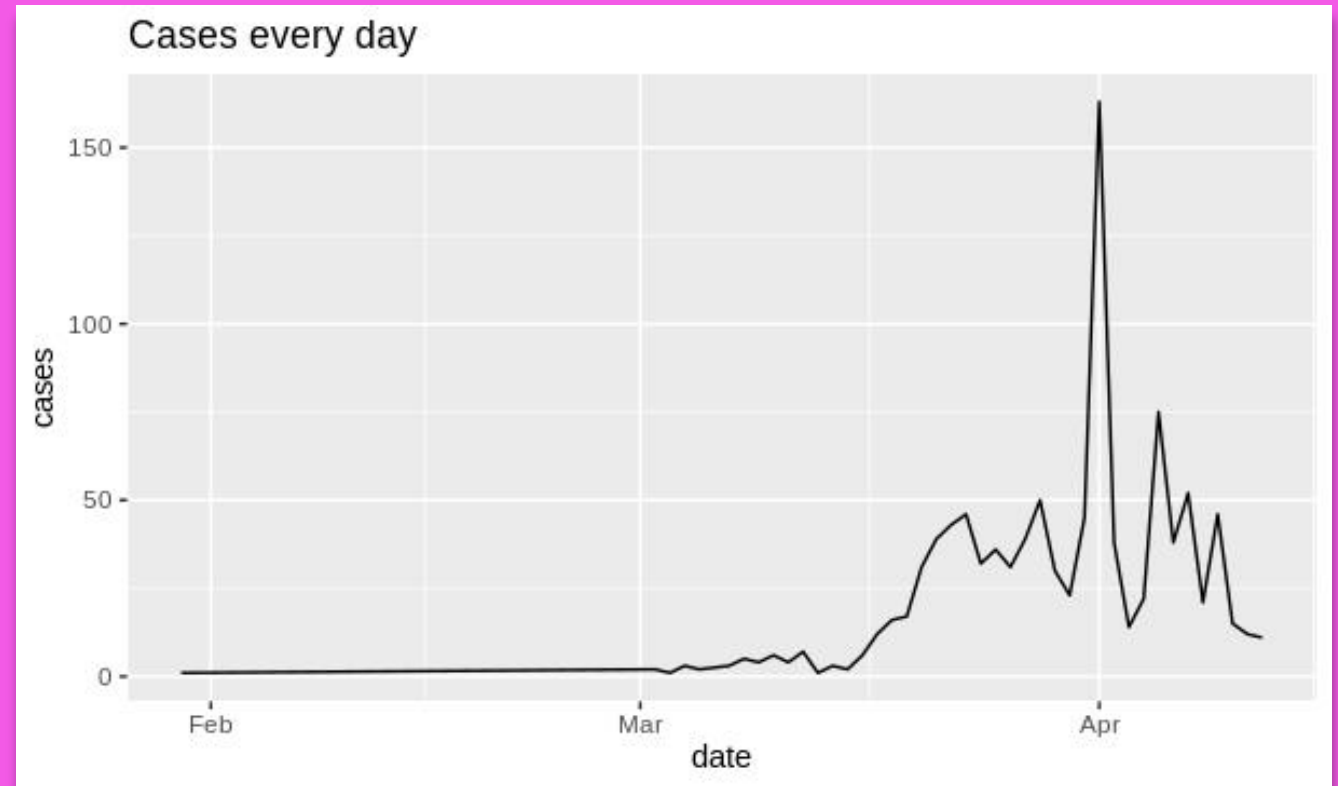


Results of the analysis done may vary than the original results as earlier mentioned in slide 4 under "note" point that after cleaning , 87.65% of the data was lost due to incomplete rows or columns in the original data

Data Visualization/Analysis , continued...

Daily Report

- Here is a graph which depicts the total no. Of cases confirmed in a day
- The maximum no. Of cases in a single day were recorded on April 1st 2020 , 163 cases

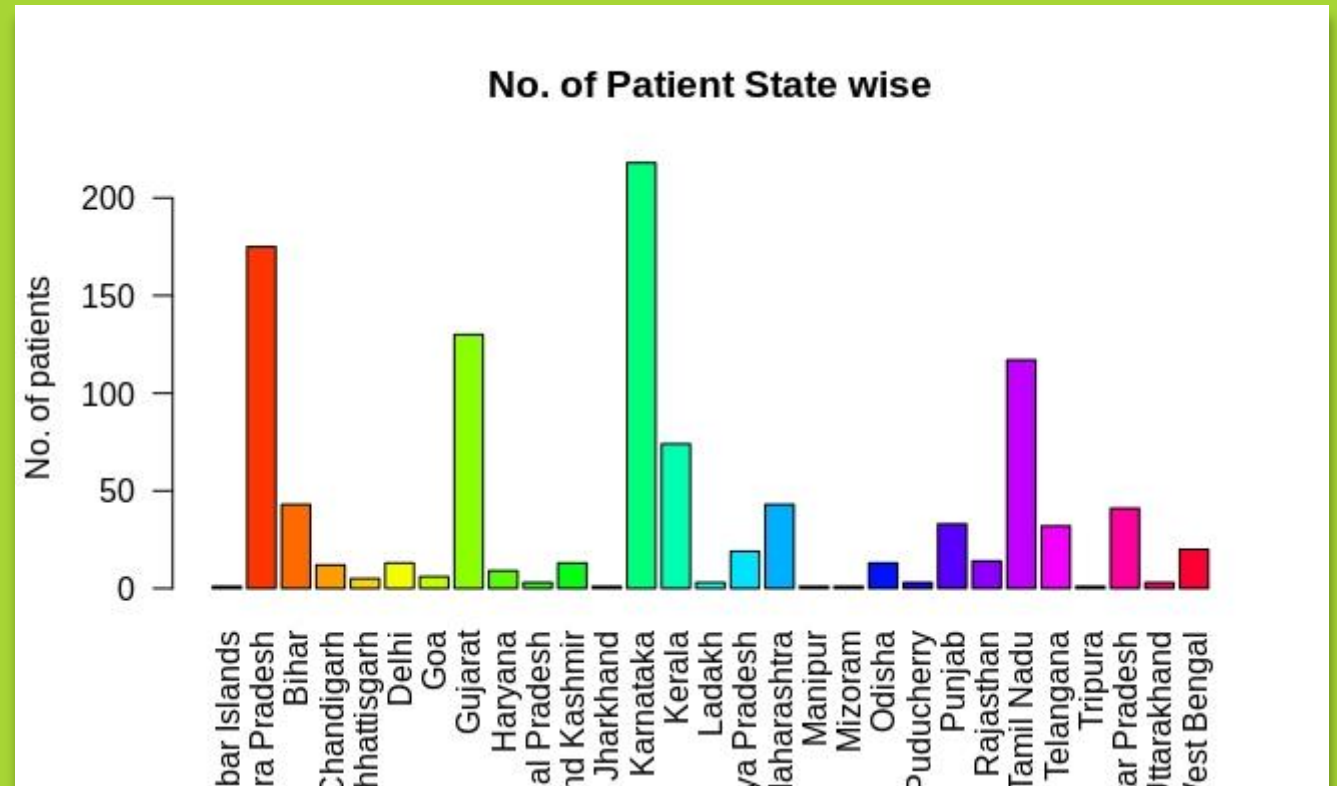


Results of the analysis done may vary than the original results as earlier mentioned in slide 4 under "note" point that after cleaning , 87.65% of the data was lost due to incomplete rows or columns in the original data

Data Visualization/Analysis , continued...

Cases as per State

Karnataka has the maximum no. Of cases(218) followed by Andra Pradesh(175) and Tamil Nadu(177)

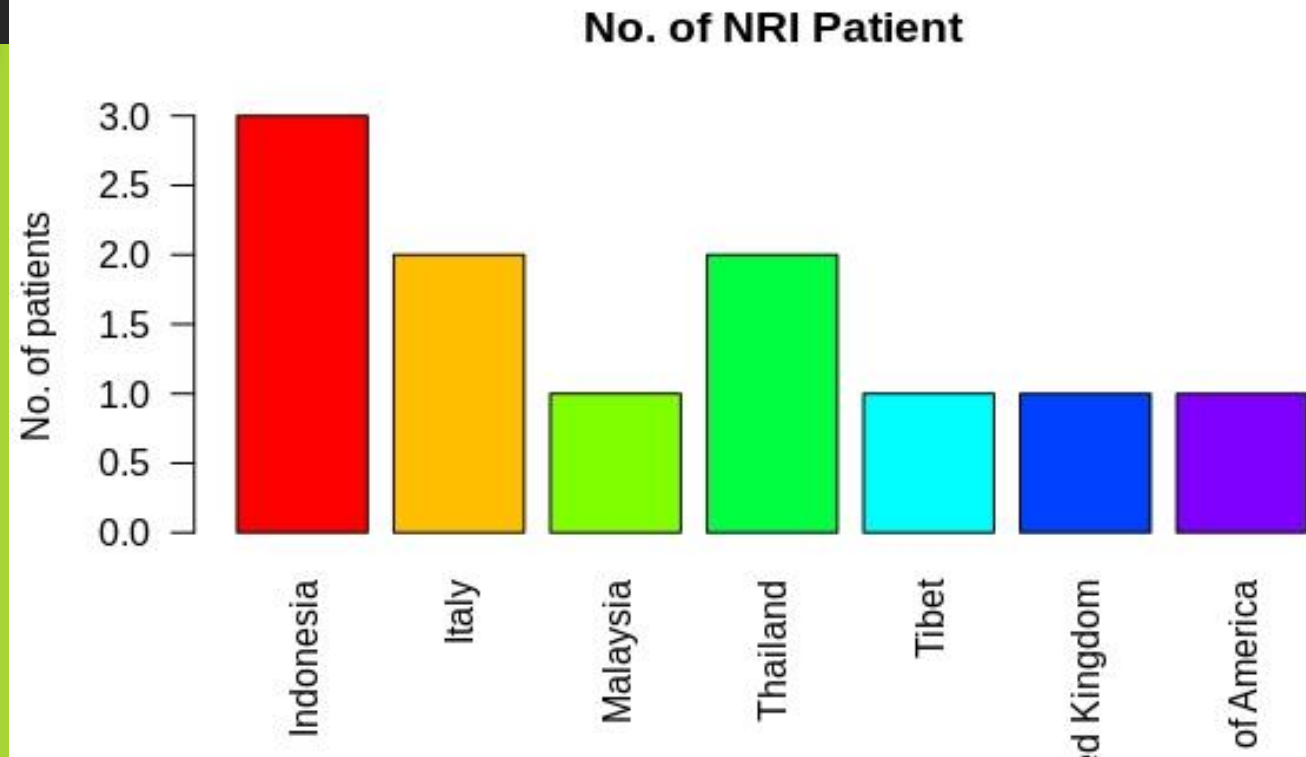


Results of the analysis done may vary than the original results as earlier mentioned in slide 4 under "note" point that after cleaning , 87.65% of the data was lost due to incomplete rows or columns in the original data

Data Visualization/Analysis , continued...

NRI Patients affected

- We might consider that these people carried virus from their respective countries or from any other country to India while travelling around the world .
- They were present in India when tested positive for corona virus , and are held in quarantine facilities near them



Results of the analysis done may vary than the original results as earlier mentioned in slide 4 under "note" point that after cleaning , 87.65% of the data was lost due to incomplete rows or columns in the original data

Research problem and conclusion

The research problem taken here is "measuring the fatality rate of people. "

- In epidemiology, a case fatality rate — sometimes called case fatality risk — is the proportion of deaths from a certain disease compared to the total number of people diagnosed with the disease for a certain period of time

The fatality rate of people calculated from the data is follows :

- Female population has a fatality rate of 2.52%
- Male population has a fatality rate of 2.20%
- Overall fatality rate is 2.29%

Results of the analysis done may vary than the original results as earlier mentioned in slide 4 under "note" point that after cleaning , 87.65% of the data was lost due to incomplete rows or columns in the original data

Research problem and conclusion , continued...

- In Conclusion

We can say that , current evidence for COVID-19 is limited to modelling studies that make parameter assumptions based on the current, fragmented knowledge. Findings consistently indicate that quarantine is important in reducing incidence and mortality during the COVID-19 pandemic. Early implementation of quarantine and combining quarantine with other public health measures is important to ensure effectiveness. In order to maintain the best possible balance of measures, decision makers must constantly monitor the outbreak situation and the impact of the measures implemented. Testing in representative samples in different settings could help assess the true prevalence of infection, and would reduce uncertainty of modelling assumptions.

Future Scope of work

- As we know that this dataset has daily level information on the number of affected cases, deaths and recovery from 2019-20 novel coronavirus. Please note that this is a time series data and so the number of cases on any given day is the cumulative number.
- The data encourage the analysis of time series, which will allow us to look at the studied phenomenon not only in a point-like manner.
- We will continue our analysis till the time this covid-19 is present in the world . And will try to increase the efficiency and correctness of the Analysis.

S T A Y

HOME

S T A Y

SAFE