

IBM Data Science Capstone

COVID 19 in Orange County, California

----- Yuyan Shi

Why This Problem?

- Almost 40 millions people lost their jobs due to the pandemic.
- California is the state with second most cases.
- Learning the cases and deaths of different demography.
- It helped us to identify the vulnerable groups and pay more attention to them.
- We can learn experience from the data, such as the outcomes of reopening.

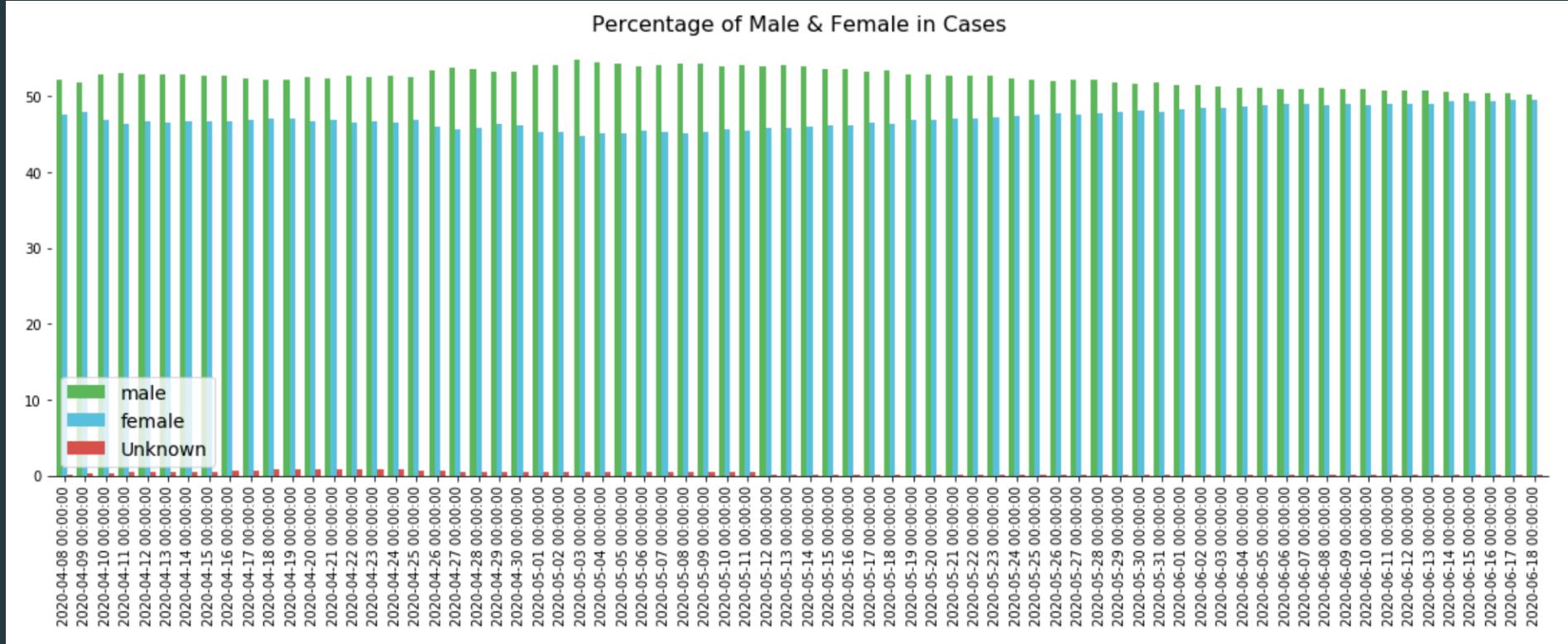


Data Sources

- Kaggle: <https://www.kaggle.com/shubhamkulkarni01/orange-county-covid19-data>
- Author: Shubham Kulkarni
- Foursquare: <https://developer.foursquare.com>

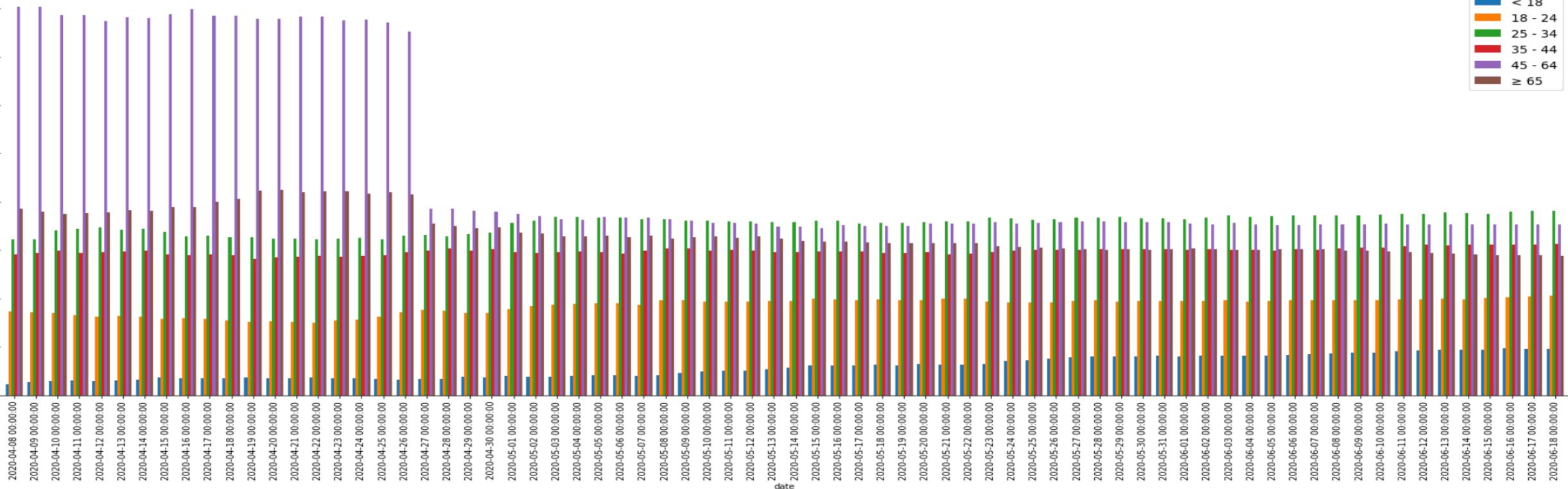


COVID 19 Cases by Sex



- In the first two months, the difference between the percentage of female cases and the percentage of male cases is significant.
- Recently, the difference becomes less.

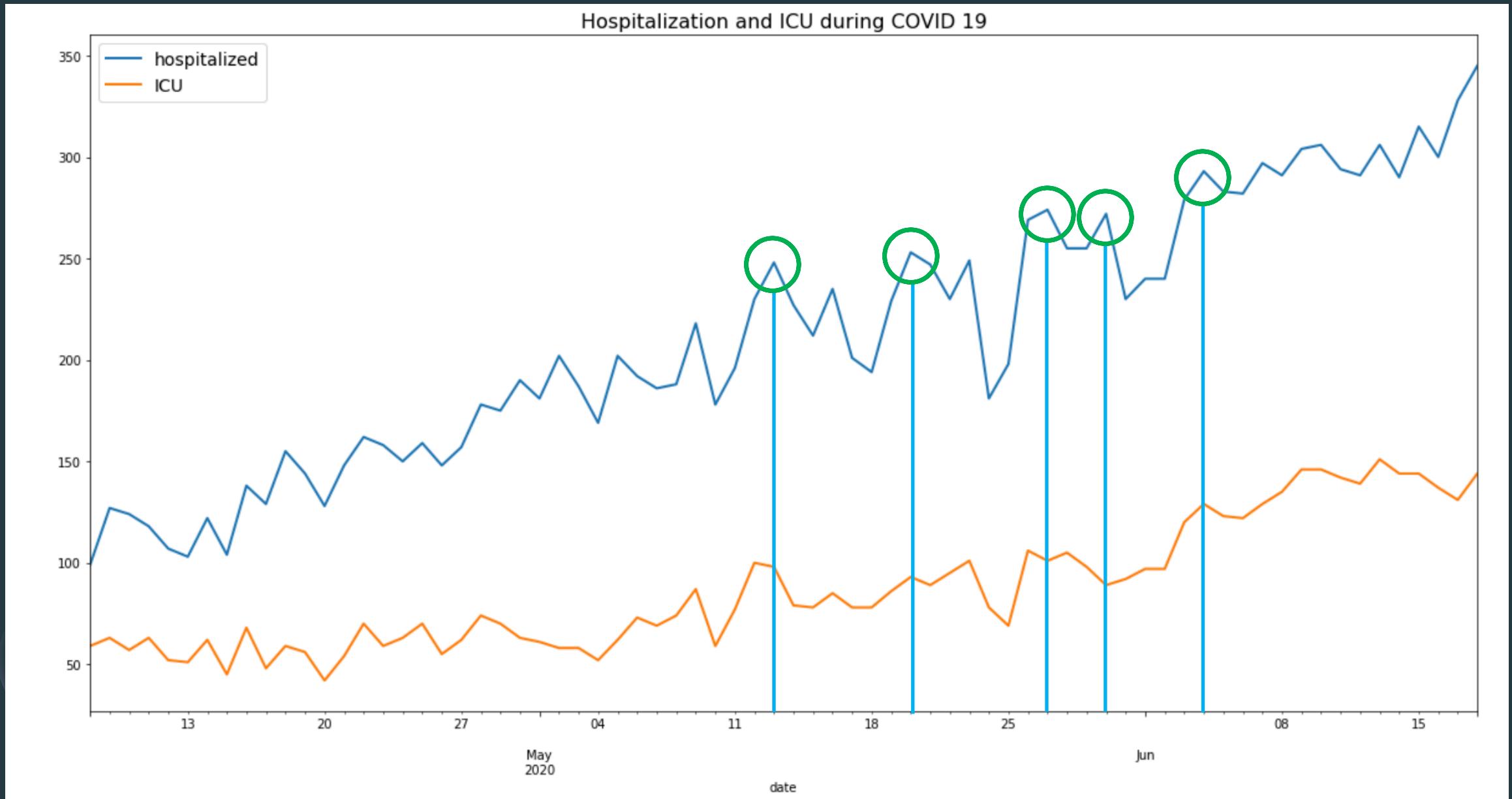
Percentage of Different Ages in Cases



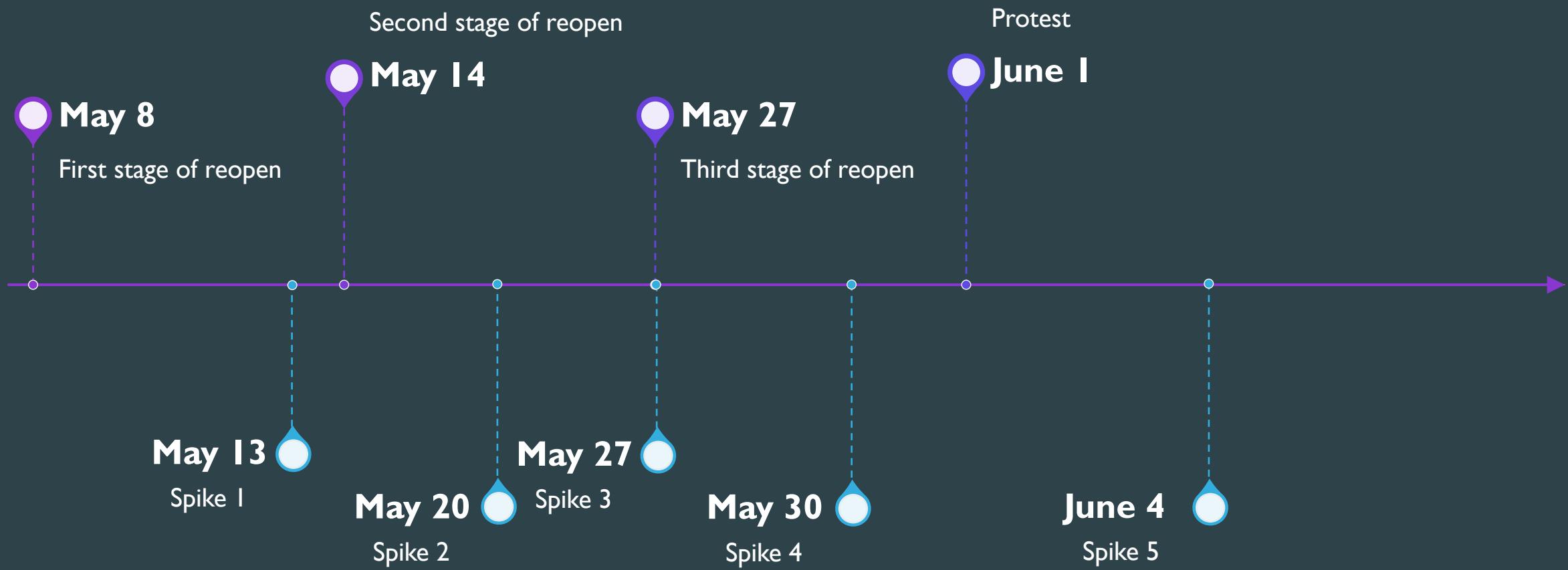
COVID 19 Cases by Age

- The percentage of people whose ages are above 45 takes up most.
- The percentage of people whose ages are below 18 gradually increases, because of the incautious protection.
- As time passed, the percentage of people whose ages are above 45 decreases since people realize that the vulnerability of old people.

Hospitalization and ICU during COVID 19

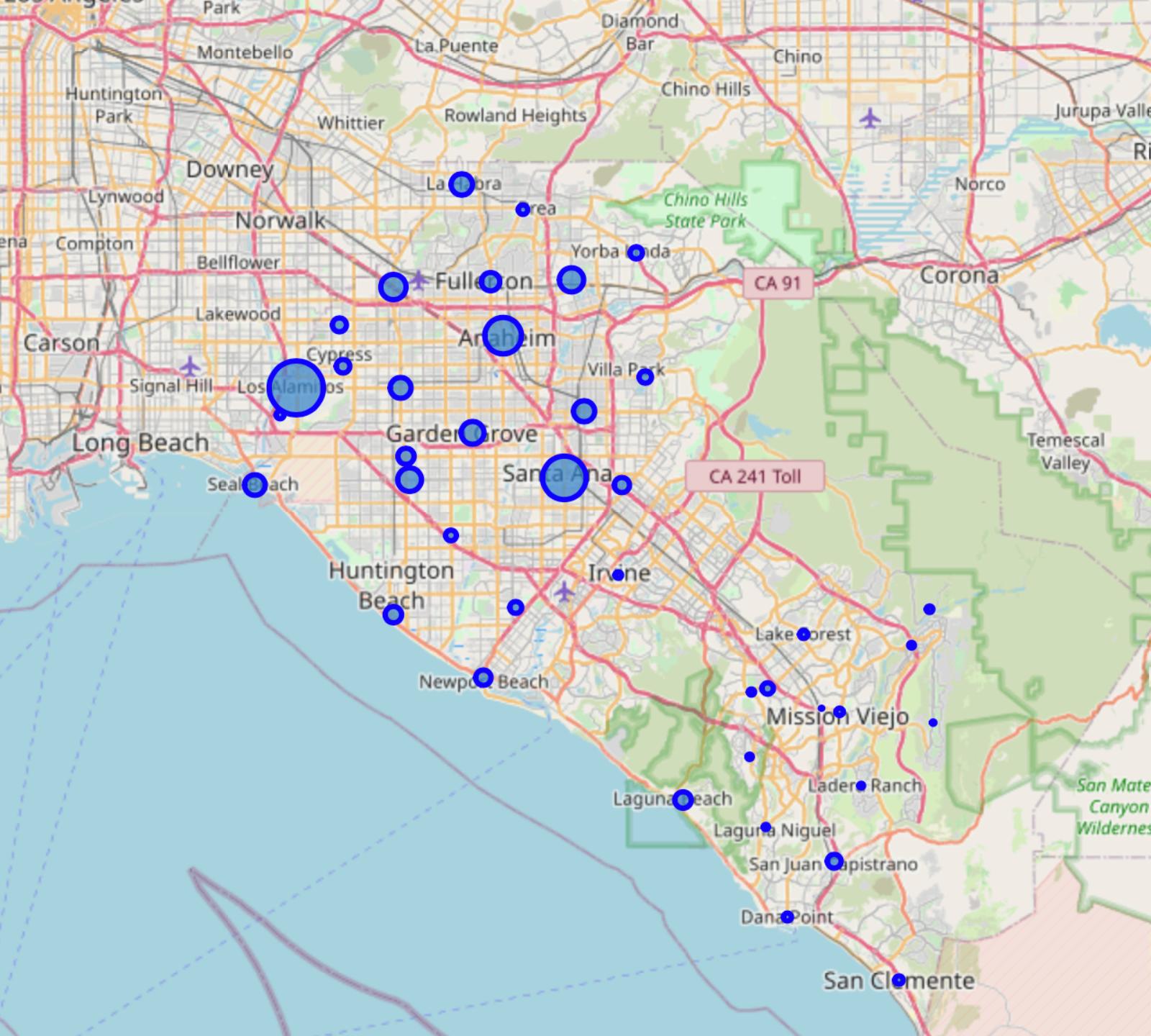


Hospitalization and ICU during COVID 19 (Cont.)



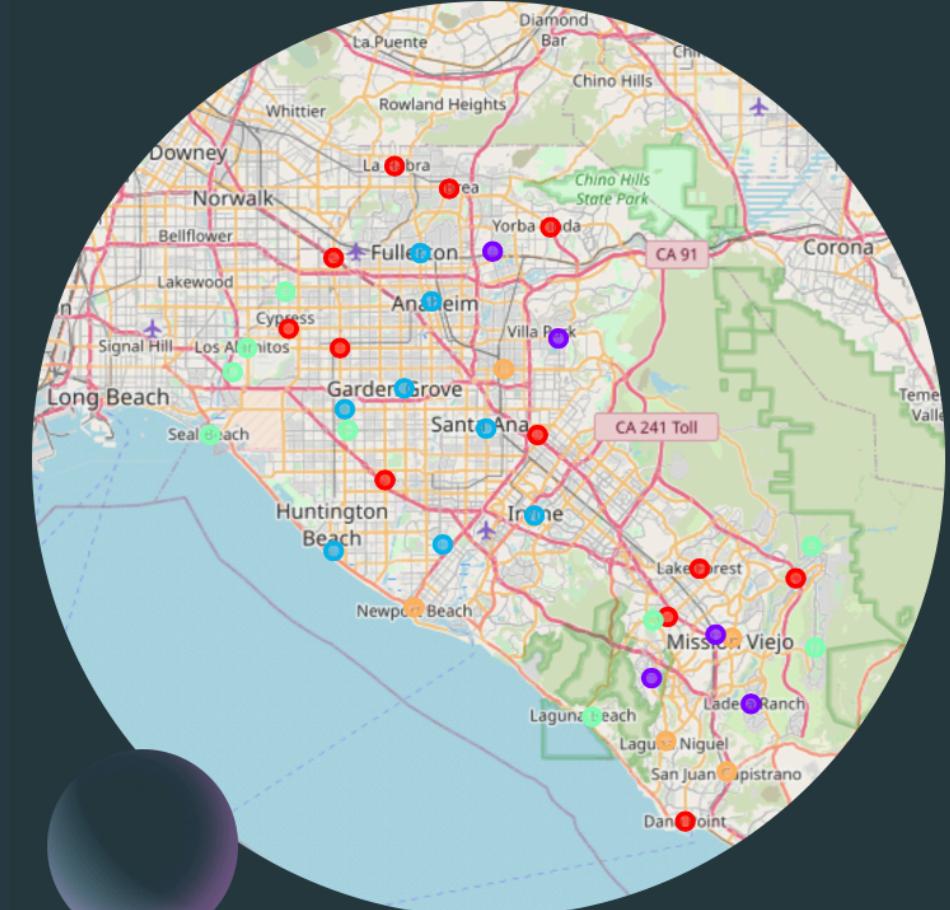
Different Severity in Different Cities

- The bigger circle is , the more case are.
- Cities with more cases are mainly located in the northwest of orange county, where the populations are denser.
- The next step is to divide cities into groups and identify the properties of each cluster.



K-Means Cluster by Severity, Population, and Medical Resources

- I used the number of hospitals in each city as proxy of the medical resources.
- The population and the medical resources have three levels.
- The same color represents the same cluster.
- Cluster 0: cities with medium population and medium medical resources. This cluster is the second serious.
- Cluster 1: cities with medium population and less medical resources. This cluster is the third serious.
- Cluster 2: cities with medium population and less medical resources. This cluster is the first serious.
- Cluster 3: cities with low population and less or medium medical resources. This cluster is the fourth serious.
- Cluster 4: cities with medium or high population and more medical resources. This cluster is the least serious.



Conclusion

- The old people and male are more vulnerable.
- The sudden increase happened several days after the reopen and tests.
- Therefore, more cautious consideration should be brought.
- The severity of the pandemic highly depends on the medical resources.
 - Cities with more medical resources tends to have lower number of cases per ten thousand.
- People whose tests are positive should have more medical resources if they live in cities of Cluster 0, 3, or 4.