



CS5346: Information Visualization

Assign1: Covid_19 Viz

Haroon Basheer	A0118085e
----------------	-----------

1. Background

In this viz challenge we will be visualising the covid 19 datasets across different countries for the period of January-March. Covid-19 is declared as a global pandemic by the World Health Organization, the attempt to visualise the containment effort by individual countries is futile given the limited amount of dataset depth for every country. As the spread has reached beyond Asia into Europe and Americas. In this visualisation challenge we have attempted to visualise spread of the viral strain across the country, age group, gender and how the growth is accelerating after china containment efforts.

2. Sources of Data

The data source for the reported cases about the covid-19 is extracted from:

<https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset>.

The data from following csv files is selected for visualisation:

covid_19_data.csv	Data about the list of countries, reported cases across month and breakdown of patient status
COVID19_open_line_list.csv	Contain data about gender, age group and symptoms about the patient.

The dataset was preprocessed using pandas and following cleaning was performed to remove dirty record from the dataset. We have dropped all the null values in the dataset and merged the two tables using country as the key.

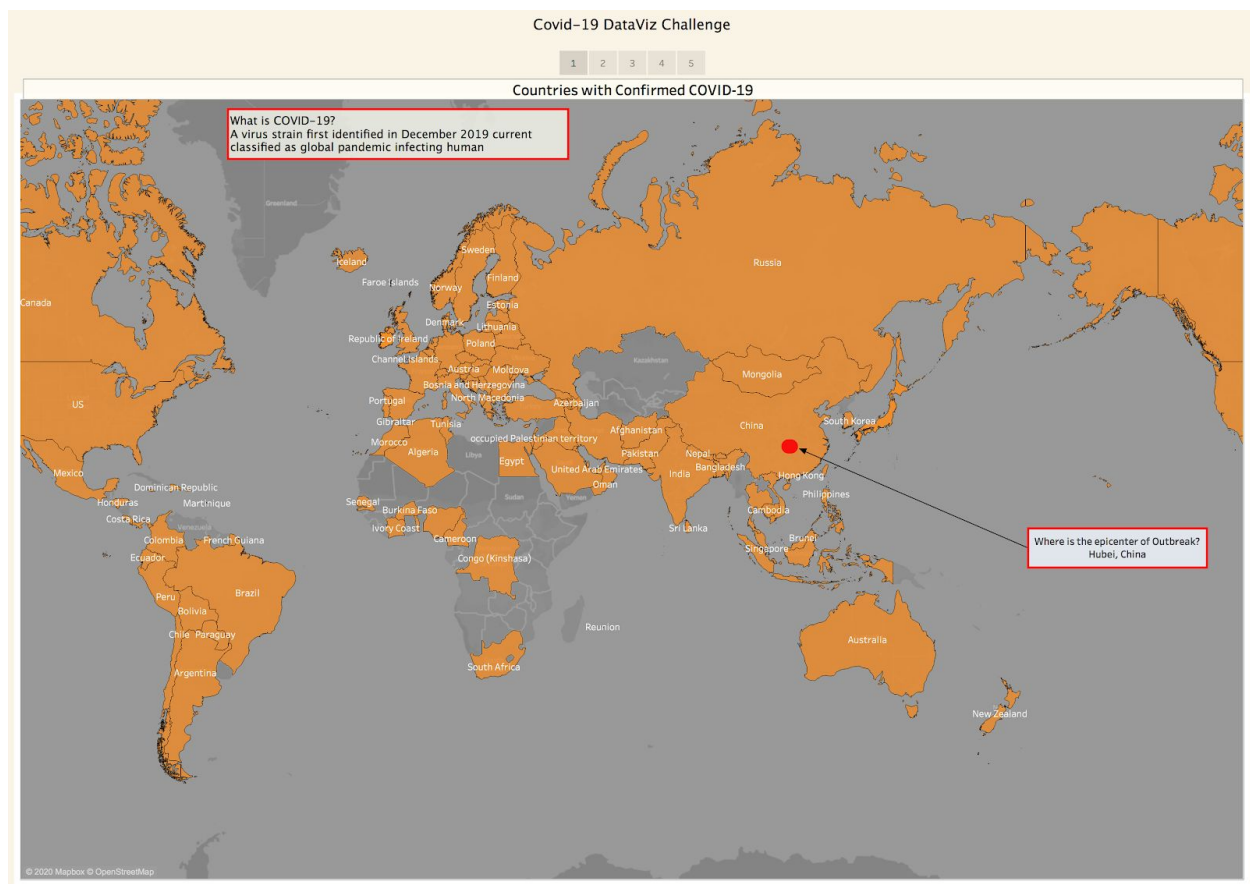
Country	List of countries with reported cases	Geo Location
Observed month	Daily updates collated monthly	Calendar
Numbers	Patient count	integer, whole
Status	Confirmed, Death, Recovered	String
Gender	Male, Female	string
Age	Age group of patient	integer, whole

3. Analytical Question pursued

a. Distribution across the World

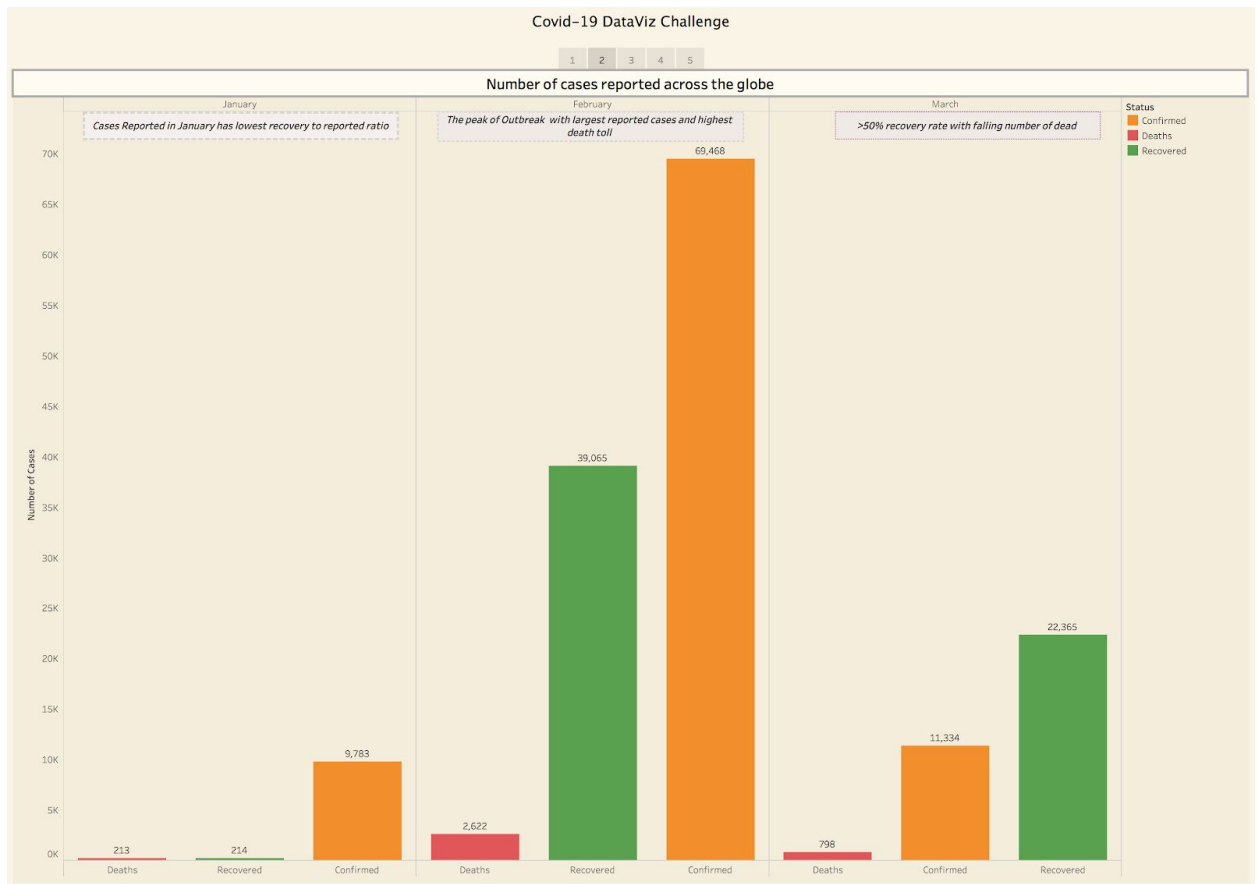
Using the confirmed columns and list of countries reported, we have visualised the overall spread of the infections across the globe

Visual Encoding: Geographic encoding with latitude and longitude data to represent Country. Text encoding for name of the country in the map plus color encoding to visualize the county affected



B. The number of case spike every month and status of each case

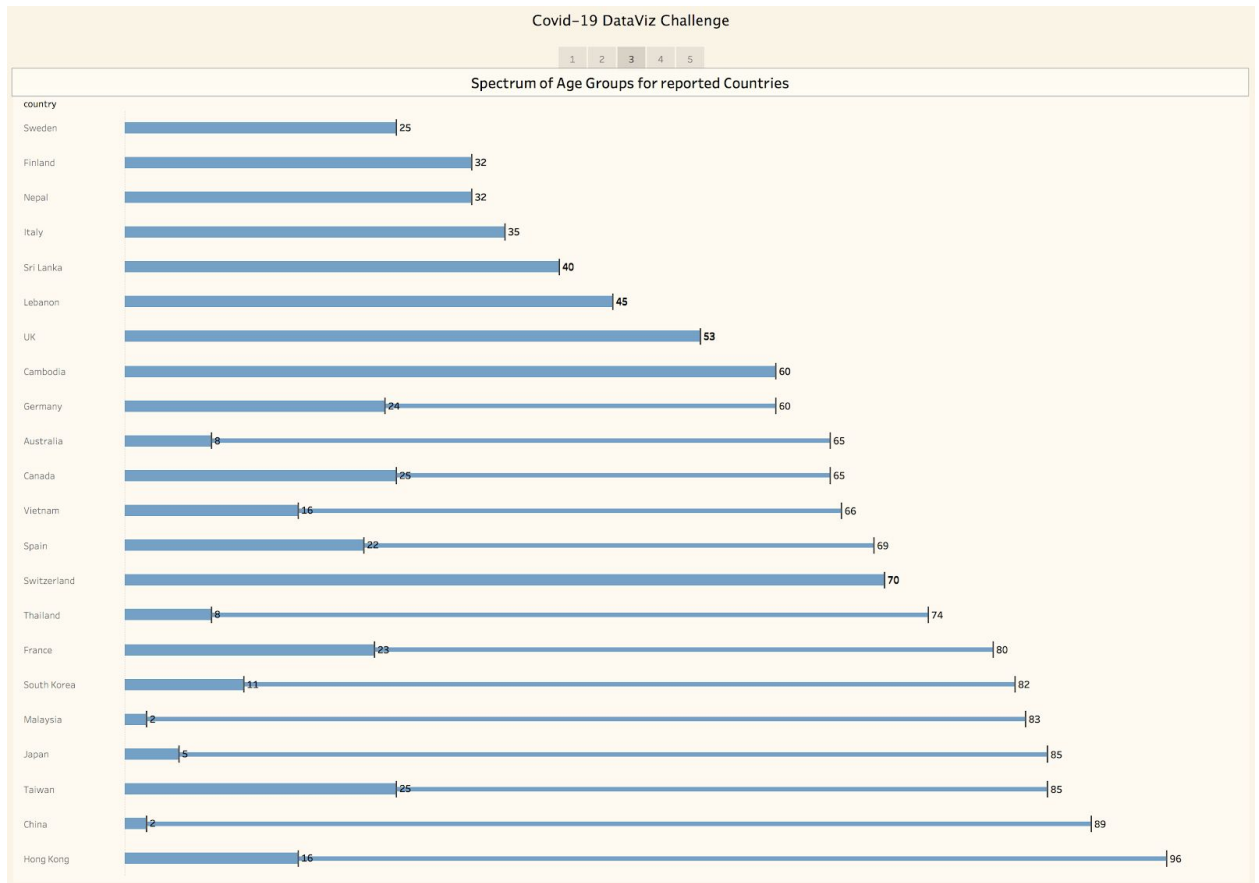
Using the status, country and numbers reported we have visualised the cases across the globe in a bar chart. We have observed that in January due to limited containment effort the number of death and recovery is less than the reported cases. However the number of recovery increases significantly over the month



We have used color encoding to represent status for the case the barchart with text encoding to represent case number for each status

C. Spectrum of age:

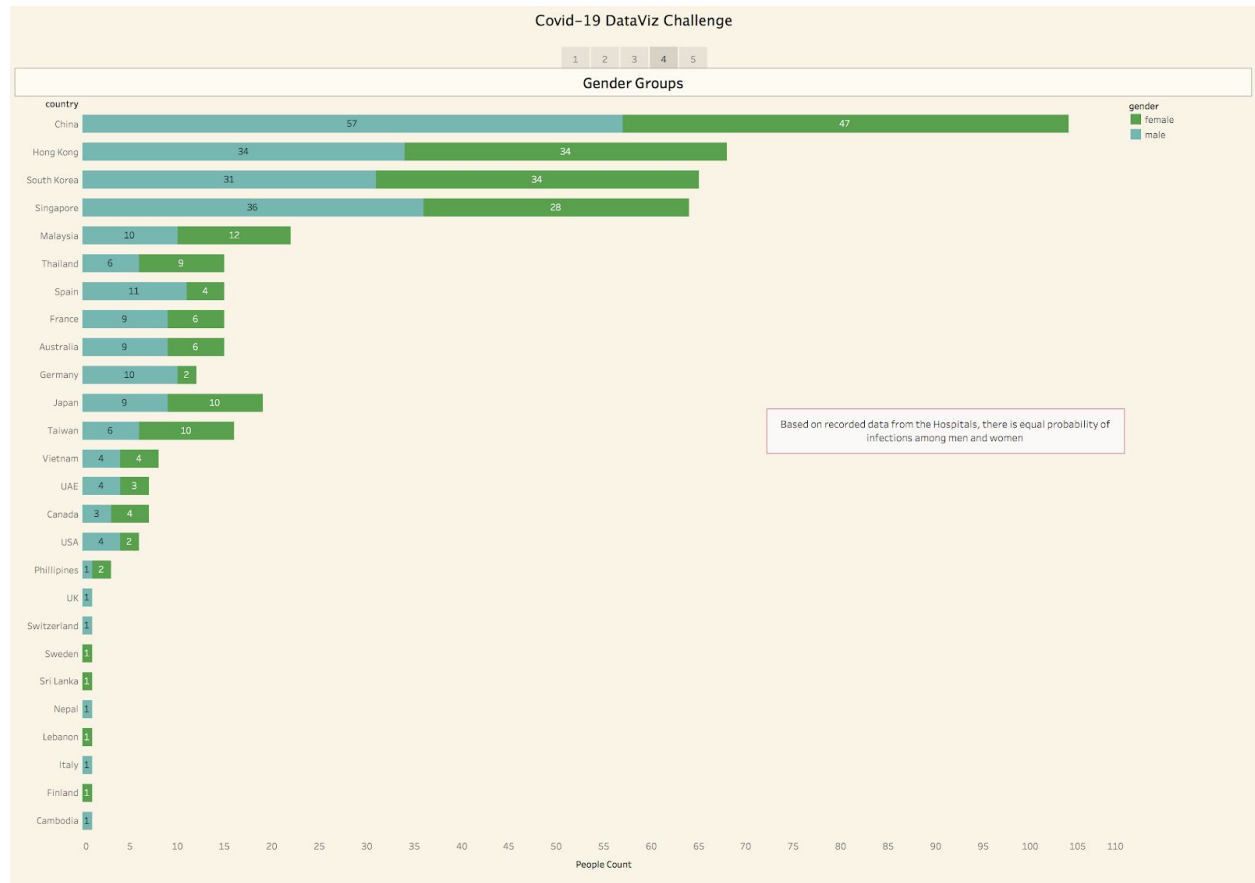
We have found the age spectrum across the reported case with minimum age being 2 and majority reported cases aged above 50



Text encoding is used to represent min and max age across the countries with width of bar to represent min and max age distribution

D. Gender Spectrum

To understand the infection pattern, there was no significant correlation across the countries for gender based infection, In Fact the transmission is not related to gender based



We have used color encoding to represent the gender and text encoding to represent the count of cases for each gender across the country.

E.Lockdown of China and Worldwide Infection

In this we have visualised how the lockdown of china has been effective in decreasing



the local transmission across the epicenter but the missed opportunity for the rest of the world by relative size of bubble across month for other country.

