

Assignment: Milestone 2-3

Group: Outbreak

Members:

James Valles

Kriti Srivastava

Dufang Qu

Frank DeRango

Hima Spandana Barla

Shadhana Palaniswami

Milestone 2

Part A:

a. First meeting was held on February 19. Our team is currently using Slack for communication: csc465group.slack.com.

Group Liaison responsible for communication: James Valles

Part B:

Chosen Datasets

Our team has identified a preliminary list of datasets for the final project outlined below. We would like to visualize different disease outbreaks, including the world's most recent Wuhan Coronavirus. Not only are we interested in seeing how these diseases involve, we think it would be a good idea to see their impact while comparing them to each other. We feel this would be an outstanding opportunity to raise awareness

- **Ebola data from 2014 to 2016**

<https://data.world/brianray/ebola-cases>

- **Mers data in June, 4, 2014**

https://figshare.com/articles/MERS_data_June_4_2014/1243671

- **Updated Daily - Coronavirus Dataset** (contains a map, with link source to each update, notes serious/critical, death, suspected cases, recovery, including international case totals. Updated daily).

<https://bnonews.com/index.php/2020/01/the-latest-coronavirus-cases/>

- **Outbreaks: a compilation of disease outbreak data**

<https://www.repidemicsconsortium.org/outbreaks/>

- **Interesting report on the Impacts of Ebola on Tourism**

<https://www.wttc.org/-/media/files/reports/2018/impact-of-the-ebola-epidemic-on-travel-and-tourism.pdf>

- **2018/W25: U.S. Influenza Surveillance Report**

<https://data.world/makeovermonday/us-influenza-surveillance-report>

- **Influenza Laboratory-Confirmed Cases By County: Beginn... 09**

<https://data.world/healthdatany/jr8b-6gh6>

- **Number of ebola cases and death in affected countries**

<https://data.world/hdx/0d089fa0-3567-4b01-9c03-39d340ff34e3>

- **Ebola Cases and Deaths in the Equateur region Ebola outbreak**

<https://data.world/hdx/2a1b9737-0197-4b46-9ef9-31dda5fa53a6>

- **Number of health-care workers infected with Ebola**

<https://data.world/hdx/2f564400-943b-4c46-9c65-a447355c8587>

- **Ebola outbreaks before 2014** <https://data.world/hdx/8bd9d9f8-6259-4b29-9acb-7e7d479da3e2>

- **Ebola Cases and Deaths in the North Kivu Ebola Outbreak in the Democratic Republic of the Congo (DRC)** <https://data.world/hdx/383945db-762c-46e2-bec6-07adc41fbd16>

- **Tracking the Wuhan Coronavirus reported cases.** WHO: Tracking of Coronavirus and its situational report Link:

<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>

- **US.CDC: Novel Coronavirus Spread in USA (Has confirmed, suspected, recovered, deaths)**

Link: <https://www.cdc.gov/coronavirus/2019-ncov/index.html>

<https://www.cdc.gov/coronavirus/2019-ncov/cases-in-us.html>

- **Tracking Mainland china cases we can use DXY data set** Link:

<https://ncov.dxy.cn/ncovh5/view/pneumonia?scene=2&clicktime=1579582238&enterid=1579582238&from=singlemessage&isappinstalled=0>

- **DXY is a Chinese website that aggregates NHC and local CCDC situation reports in near real-time, providing more current regional case estimates than the national level reporting organizations** NHC Link: http://www.nhc.gov.cn/yjb/s3578/new_list.shtml CCDC Link:

<http://www.chinacdc.cn/en/>

- **Dataset (Contains Province/State, Country/Region, Last Update, Confirmed, Deaths, Recovered, Time and Date) Novel Coronavirus (2019-nCoV) Cases, provided by above websites**
- Google sheet:**

<https://docs.google.com/spreadsheets/d/1wQVypefm946ch4XDp37uZ-wartW4V7ILdg-qYiDXUHM/htmlview?usp=sharing&sle=true>

Time Series:

<https://docs.google.com/spreadsheets/d/1UF2pSkFTURko2OvfHWWIFpDFAr1UxCBA4JLwISP6KFo/edit#gid=0>

Feature layer: <https://www.arcgis.com/home/item.html?id=c0b356e20b30490c8b8b4c7bb9554e7c>

Dataset Details on Influenza

Datasets Source: <https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html>

A. ILINet_National.csv : This file contains the information of the percentage of visits for INFLUENZA-LIKE-ILLNESS REPORTED BY SENTINEL PROVIDERS.

Number of Rows: 1198

Number of Columns: 15

- **REGION TYPE (Text):** This tells how the data is captured region wise. Since I downloaded the Nation wise data therefore, this column only contain “National”.
- **REGION (Text):** : Non applicable.
- **YEAR (Date):** The year when the reports from public health and clinical laboratories are presented.
- **WEEK(Number):** The week when the reports from public health and clinical laboratories are presented.
- **% WEIGHTED ILI (Decimal):** Weighted Percent of positive Influenza-Like-Illness.
- **%UNWEIGHTED ILI (Decimal):** Unweighted Percent of positive Influenza-Like-Illness.
- **AGE 0-4:** Number of positive cases from Age 0 to 4.
- **AGE 25-49:** Number of positive cases from Age 25 to 49.
- **AGE 25-64 :** Number of positive cases from Age 25 to 65.
- **AGE 5-24 :** Number of positive cases from Age 5 to 24.
- **AGE 50-64:** Number of positive cases from Age 50 to 64.
- **AGE 65:** Number of positive cases from Age 65 and above.
- **ILITOTAL:** Total number of cases.
- **NUM. OF PROVIDERS:** Number of providers.
- **TOTAL PATIENTS:** Total number of patients.

B. WHO_NREVSS_Clinical_Labs.csv: This file contains beginning for the 2015-16 season, reports from public health and clinical laboratories are presented separately in the weekly influenza update, FluView. Data from clinical laboratories include the weekly total. It contains following columns:

Number of Rows: 12311

Number of Columns: 10

- REGION TYPE (Text): This tells how the data is captured region wise. Since I downloaded the state wise data therefore, this column only contains “States”.
- REGION (Text): : USA states name.
- YEAR (Date): The year when the reports from public health and clinical laboratories are presented.
- WEEK(Number): The week when the reports from public health and clinical laboratories are presented.
- TOTAL SPECIMENS (Number): Total number of specimens per state taken every week.
- TOTAL A (Number): Number of type A flue.
- TOTAL B(Number): Total number of type B flue.
- PERCENT POSITIVE(Decimal): Percent of positive test for both type A and B.
- PERCENT A(Decimal): Percent of positive test for both type A .
- PERCENT B(Decimal): Percent of positive test for both type B.

C. WHO_NREVSS_Combined_prior_to_2015_16.csv: This data file includes only data prior to the 2015-16.

Number of Rows: 14098

Number of Columns: 13

- REGION TYPE (Text): This tells how the data is captured region wise. Since I downloaded the state wise data therefore, this column only contain “States”.
- REGION (Text): : USA states name.
- YEAR (Date): The year when the reports from public health and clinical laboratories are presented.
- WEEK(Number): The week when the reports from public health and clinical laboratories are presented.
- TOTAL SPECIMENS (Number): Total number of specimens per state taken every week.
- PERCENT POSITIVE(Decimal): Percent of positive test for both type A and B.
- A(2009H1N1) (Number): This column is to record number of positive cases of H1N1 Swine flue.
- A(H1) (Number): This column is to record number of positive cases of H1 flue.
- A(H3) (Number): This column is to record number of positive cases of H3 flue.
- A(Subtyping Not Performed) (Number): This column is to record number of positive cases of Type A.
- A(Unable to Subtype) (Number): This column is to record number of positive cases of Type A.
- B(Number): Total number of type B flue.
- H4N2v(Number): Total number of type B, H4N2v flue.

D. WHO_NREVSS_Public_Health_Labs.csv: Beginning for the 2015-16 season, reports from public health and clinical laboratories are presented separately in the weekly influenza update, FluView. Data presented from public health laboratories include the weekly total number of specimens tested, the number of positive influenza tests, and the number by influenza virus type, subtype, and influenza B lineage.

Number of Rows: 270

Number of Columns: 11

- REGION TYPE (Text): This tells how the data is captured region wise. Since I downloaded the state wise data therefore, this column only contain "States".
- REGION (Text): : USA states name.
- SEASON_DESCRIPTION (Text): The year when the reports from public health and clinical laboratories are presented.
- TOTAL SPECIMENS (Number): Total number of specimens per state taken every week.
- PERCENT POSITIVE(Decimal): Percent of positive test for both type A and B.
- A(2009H1N1) (Number): This column is to record number of positive cases of H1N1 Swine flue.
- A(H3) (Number): This column is to record number of positive cases of H3 flu.
- A(Subtyping Not Performed) (Number): This column is to record number of positive cases of Type A.
- B(Number): Total number of type B flue.
- BVic(Number): Total number of type B, H4N2v flue.
- BYam(Number): Total number of type B, H4N2v flue.
- H4N2v(Number): Total number of type B, H4N2v flue.

D:

Datasets Source: <https://www.cdc.gov/flu/weekly/#S5>

NCHS.csv: This file contains mortality surveillance data available on February 20, 2020 based on National Center for Health Statistics (NCHS).

Number of Rows: 332

Number of Columns: 8

Columns:

- Year (Time series)
- Week (Time series)
- Percent of Death due to Pneumonia and Influenza
- Expected
- Threshold
- All Deaths
- Pneumonia Deaths
- Influenza Deaths

Dataset Details on Coronavirus (COVID-19)

Dataset Details on Coronavirus (COVID-19) numbers overall including international numbers

<https://bnonews.com/index.php/2020/02/the-latest-coronavirus-cases/>

This file contains the latest stats on COVID-19 including the number of suspected, confirmed cases. Including the number of patients in serious condition, those who have died and recovered.

Number of Rows: 332

Number of Columns: 8

Columns:

- Location
- Cases
- Deaths
- Notes
- Source Link

time_series_19-covid-Recovered.csv: This time series dataset provides a daily count for all patients recovered by location. It includes province/state, country/region, lat, long.

Dataset Source:

https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series

Number of Rows: 94

Number of Columns: *39 This value increased by 1 daily.

Columns:

- Province/State
- Country/Region
- Lat
- Long
- Dates with daily recovered count

time_series_19-covid-Deaths.csv: This time series dataset provides a daily count for all patients who have died by location. It includes province/state, country/region, lat, long.

Dataset Source:

https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_19-covid-Deaths.csv

Number of Rows: 94

Number of Columns: 8

Columns: *39 This value increased by 1 daily.

- Province/State
- Country/Region
- Lat
- Long
- Dates with daily death count

time_series_19-covid-Confirmed.csv: This time series dataset provides a daily count for all patients recovered by location. It includes province/state, country/region, lat, long.

Dataset Source:

https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_19-covid-Confirmed.csv

Number of Rows: 94

Number of Columns: *39 This value increased by 1 daily.

Columns:

- Province/State
- Country/Region
- Lat
- Long
- Dates with daily confirmed case count

Infectious Disease Outbreak from 2002 to 2020

Outbreak.csv : Cleaned from multiple sources mentioned above

Dataset source:

https://drive.google.com/file/d/14G_2eDZ_IADBLEPtVDI9XiUI47upGg-t/view?usp=sharing

Number of Rows: 252

Number of Columns: 4

Columns:

- Disease Type
- Year
- Country.Region
- Confirmed

Number of Ebola Cases and Deaths in Affected Countries

Dataset source;

<https://data.world/hdx/0d089fa0-3567-4b01-9c03-39d340ff34e3>

Number of rows: 17586

Number of Columns: 4

Columns:

Indicator

Country

Date

Values

Ebola outbreaks before 2014

Data source:

<https://data.world/hdx/8bd9d9f8-6259-4b29-9acb-7e7d479da3e2>

Number of rows :33

Number of columns: 6

Columns:

Years

Country

Ebola subtype

Number human cases reported

Number deaths among the cases reported

Percentage of death from the reported

Middle East respiratory syndrome (MERS-CoV)

Dataset source:

<https://docs.google.com/spreadsheets/d/1nd-tNGRhDaT6PCvleDWWVdhPQIXdcLb8/edit#gid=647819139>

Number of Rows: 682

Number of Columns: 14

Columns:

- number
- gender
- age
- country
- city
- reported
- death
- discharge
- comorbidity
- severity
- outcome
- clinical
- animal_contact
- camel contact

Description:

These datasets correspond to the initial information collected by the Epidemic Intelligence group at European Centre for Disease Prevention and Control (ECDC) during the first weeks of the outbreak of Middle East respiratory syndrome (MERS-CoV) outbreak

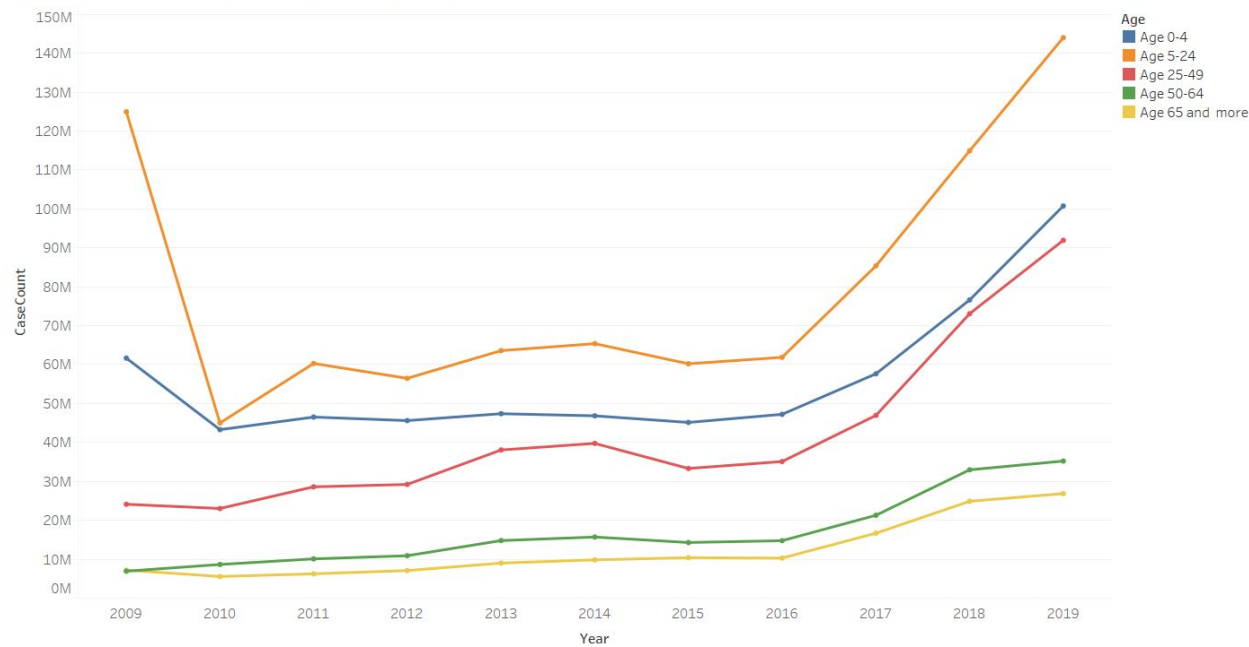
Part C

Influenza Exploratory Graphs

All these dataset has a time series which is year and weak and the Region columns geographical data types.

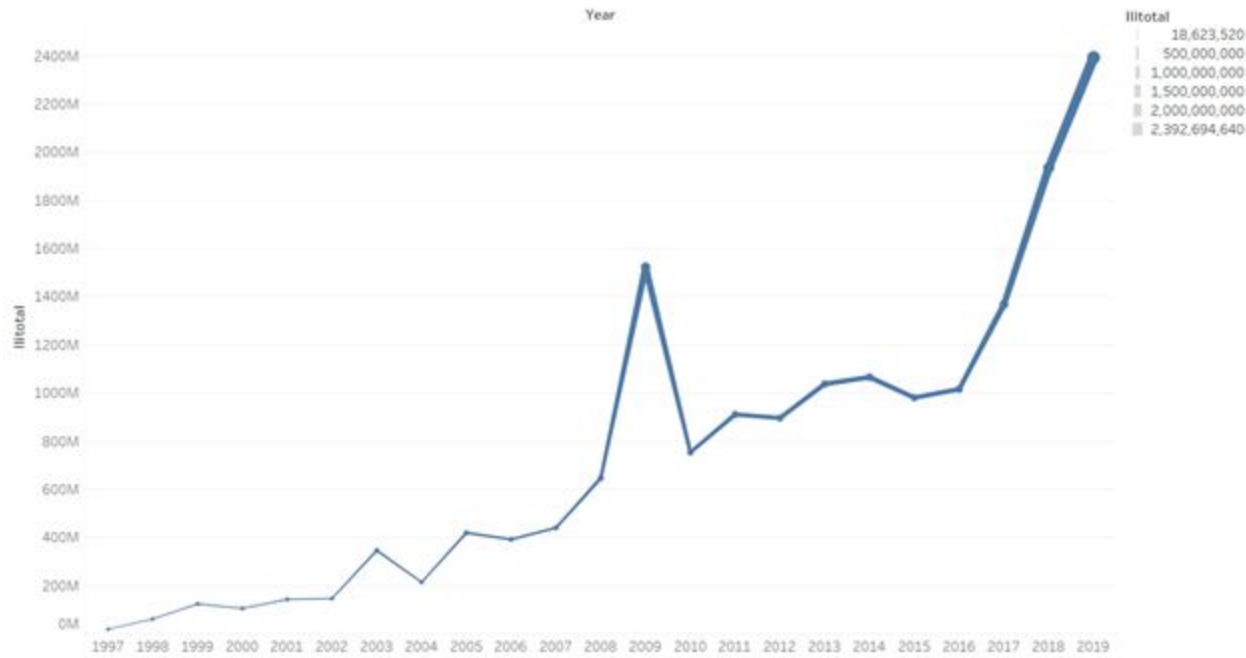
The spike in 2009 is due the (H1N1) cases.

Agewise Infuenza case count from 2009 to 2019



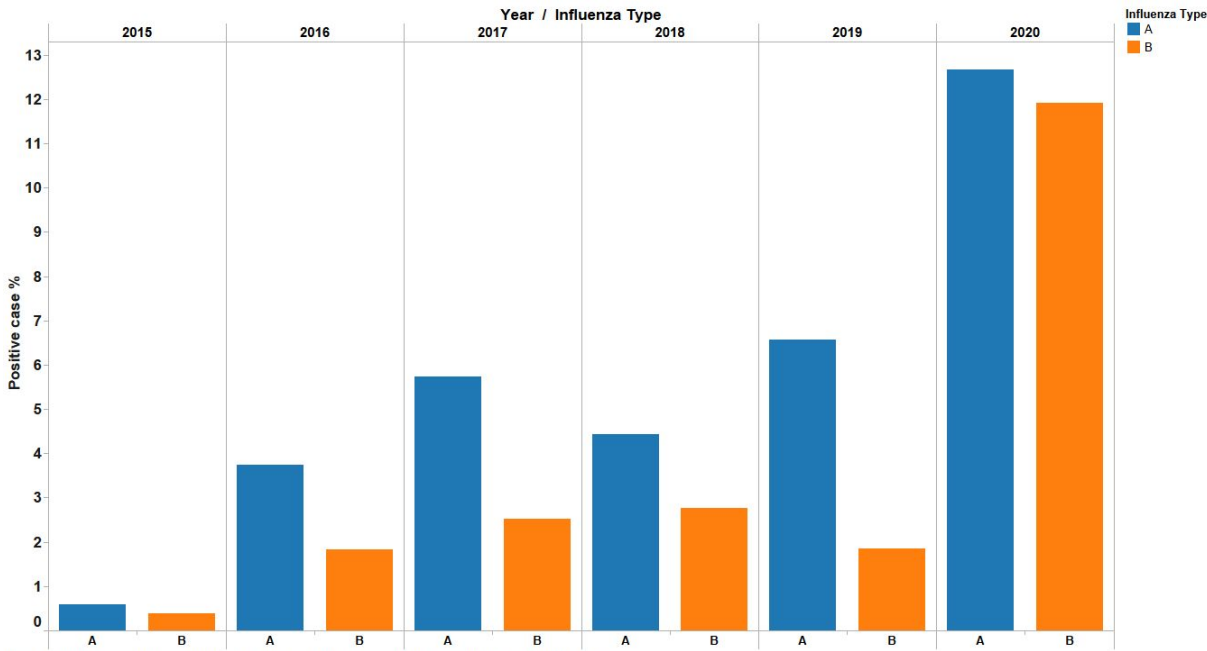
The trend of sum of CaseCount for Year Year. Color shows details about Age. The data is filtered on Year Year, which keeps 11 members. The view is filtered on Age, which has multiple members selected.

Infuenza like confirm cases in USA from 1997 to 2019



The trend of sum of liltotal for Year Year. Size shows sum of liltotal. The view is filtered on Year Year, which excludes 2020.

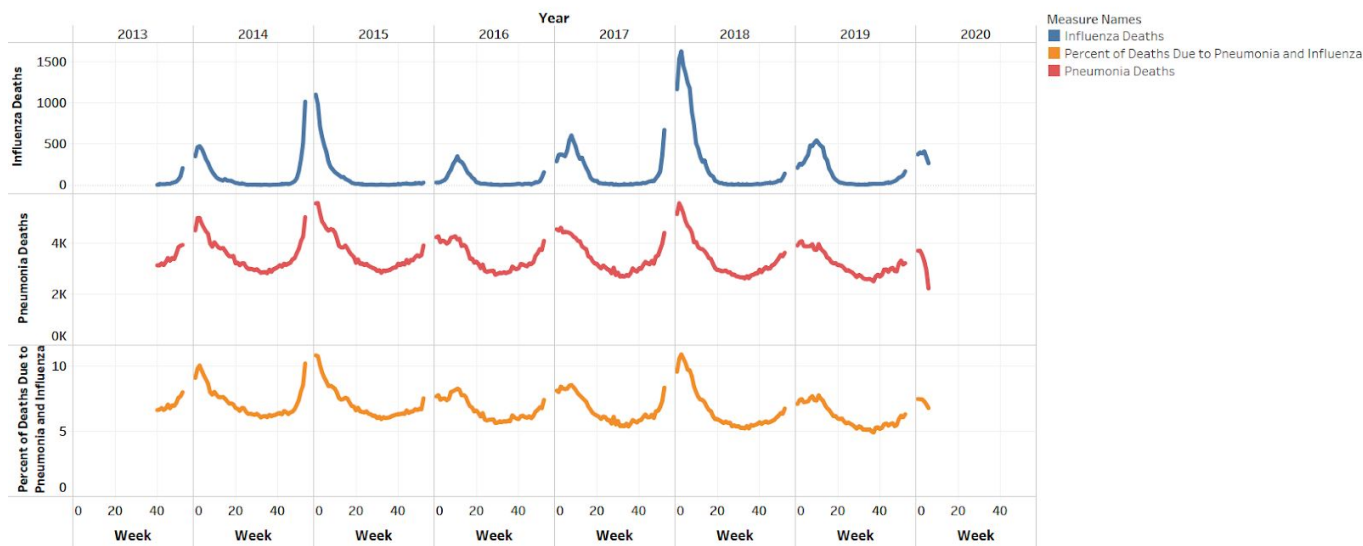
Cases of Infuenza from 40th week of 2015 to 8th week of 2020



Average of Positive case % for each Influenza Type broken down by Year. Color shows details about Influenza Type.

Weekly Influenza and Pneumonia Mortality

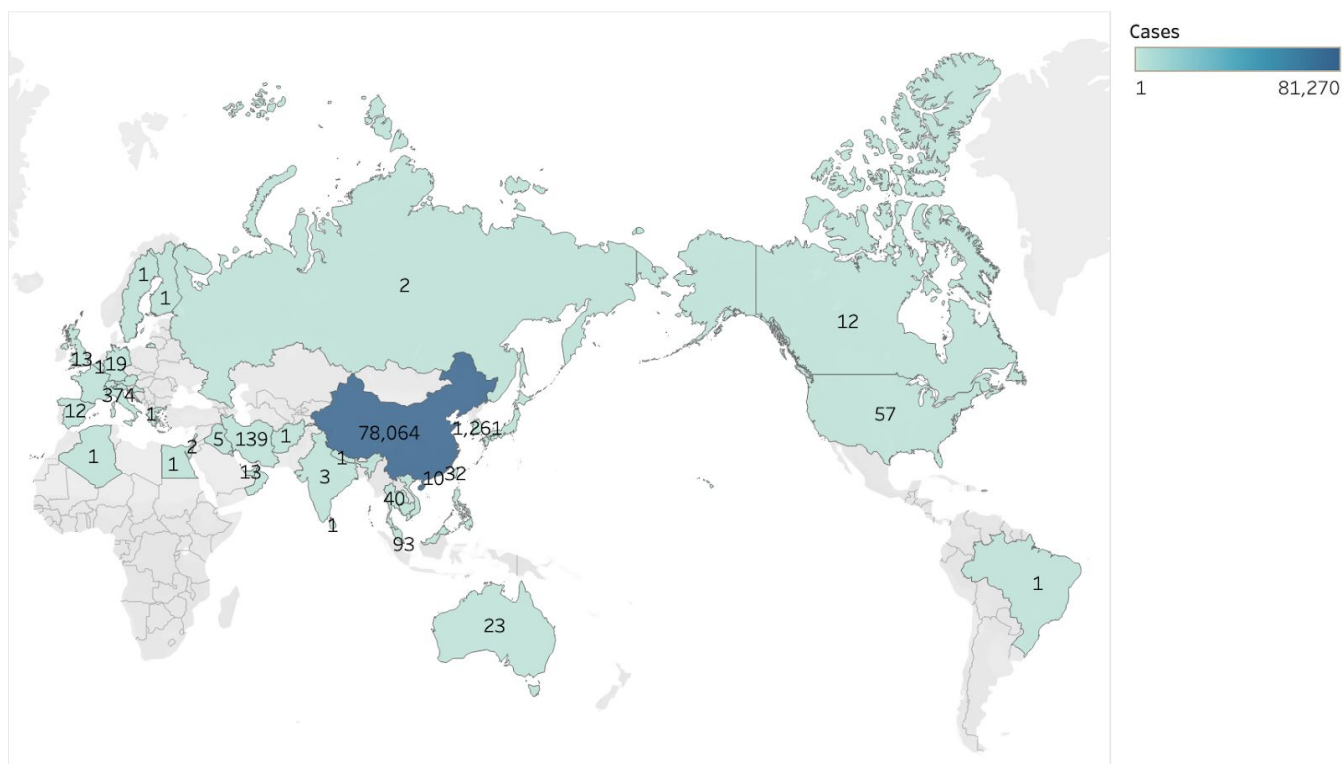
From 2015 to 20-Feb-2020



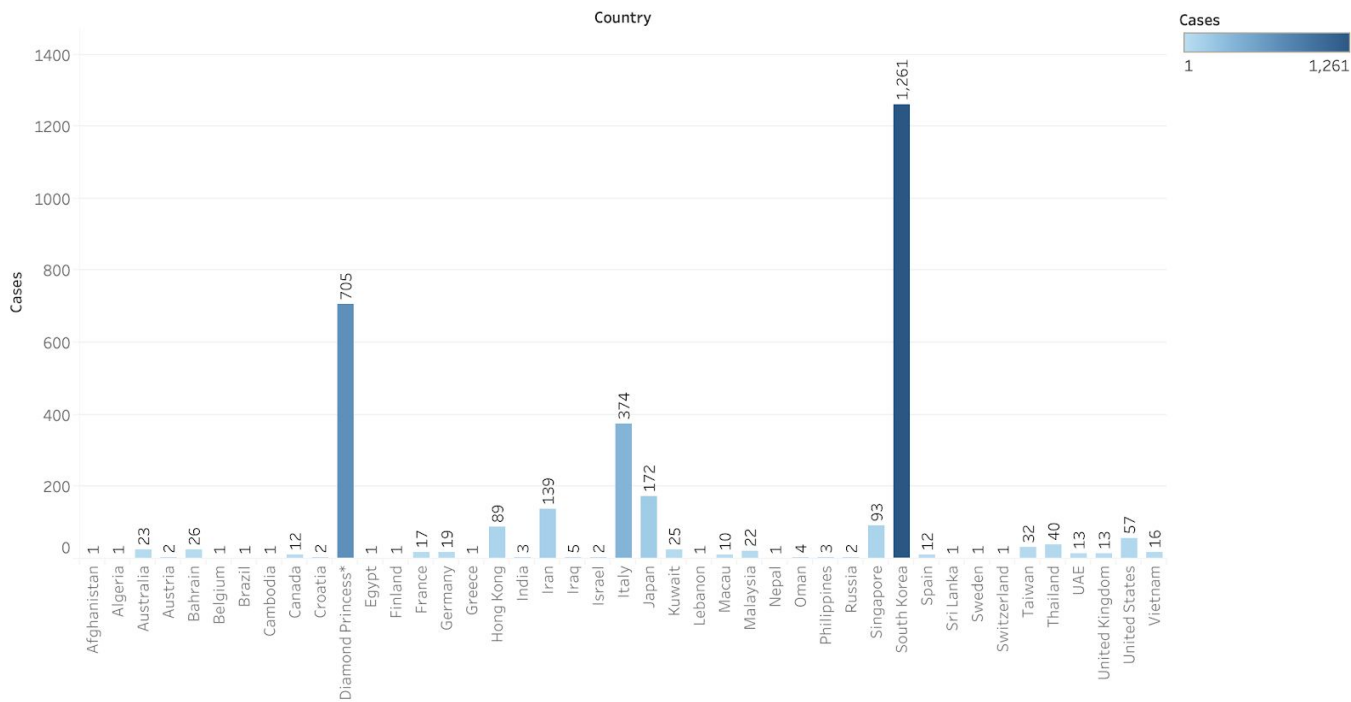
The trends of Influenza Deaths, Pneumonia Deaths and Percent of Deaths Due to Pneumonia and Influenza for Week broken down by Year. Color shows details about Influenza Deaths, Pneumonia Deaths and Percent of Deaths Due to Pneumonia and Influenza.

Coronavirus (COVID-19) Graphs

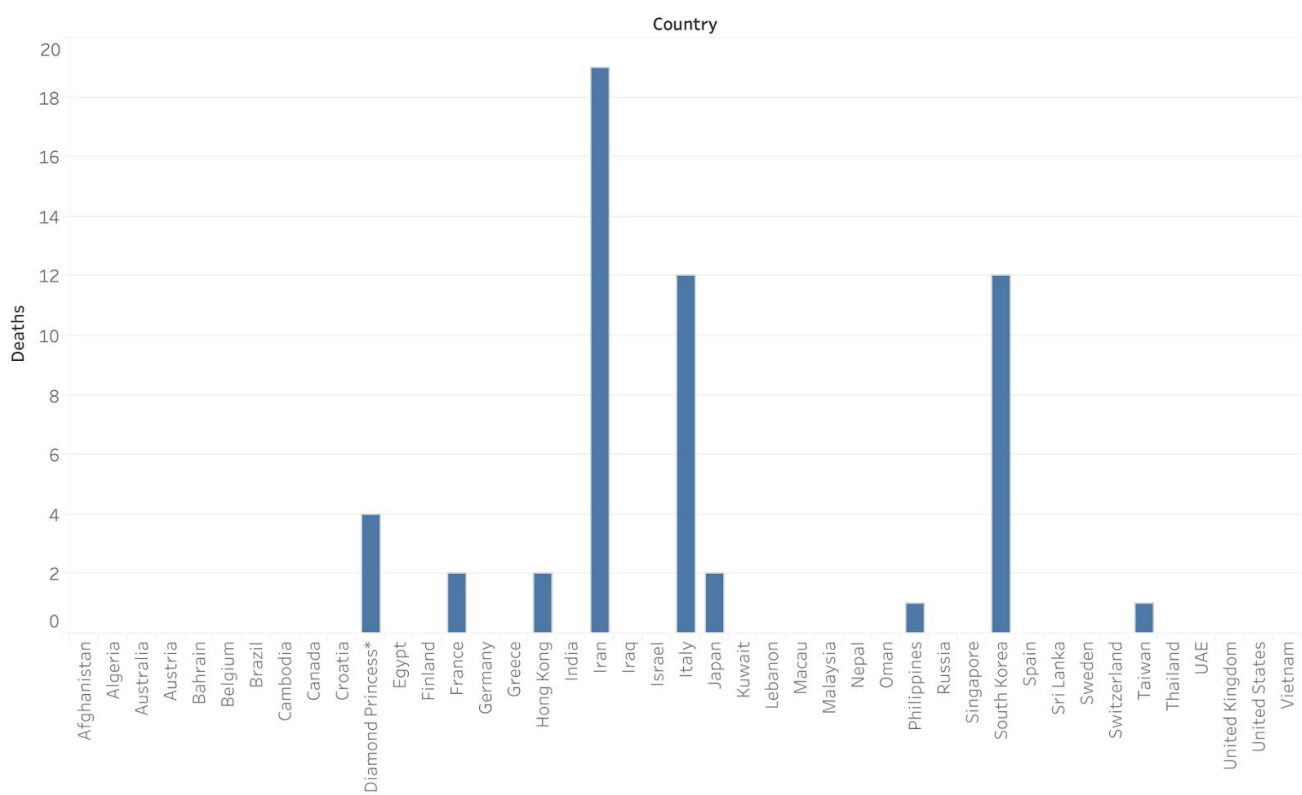
Worldwide View of Suspected Coronavirus Cases



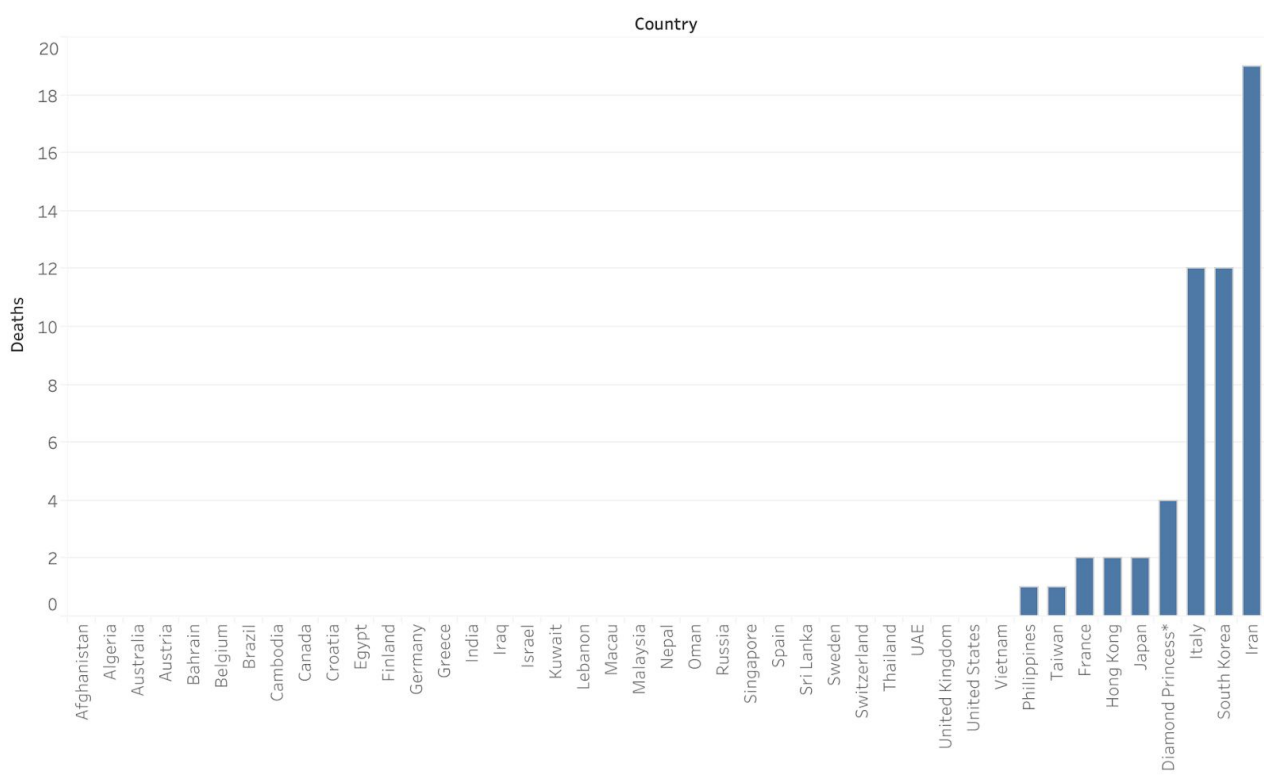
Worldwide View of Suspected Coronavirus Cases (excluding China)



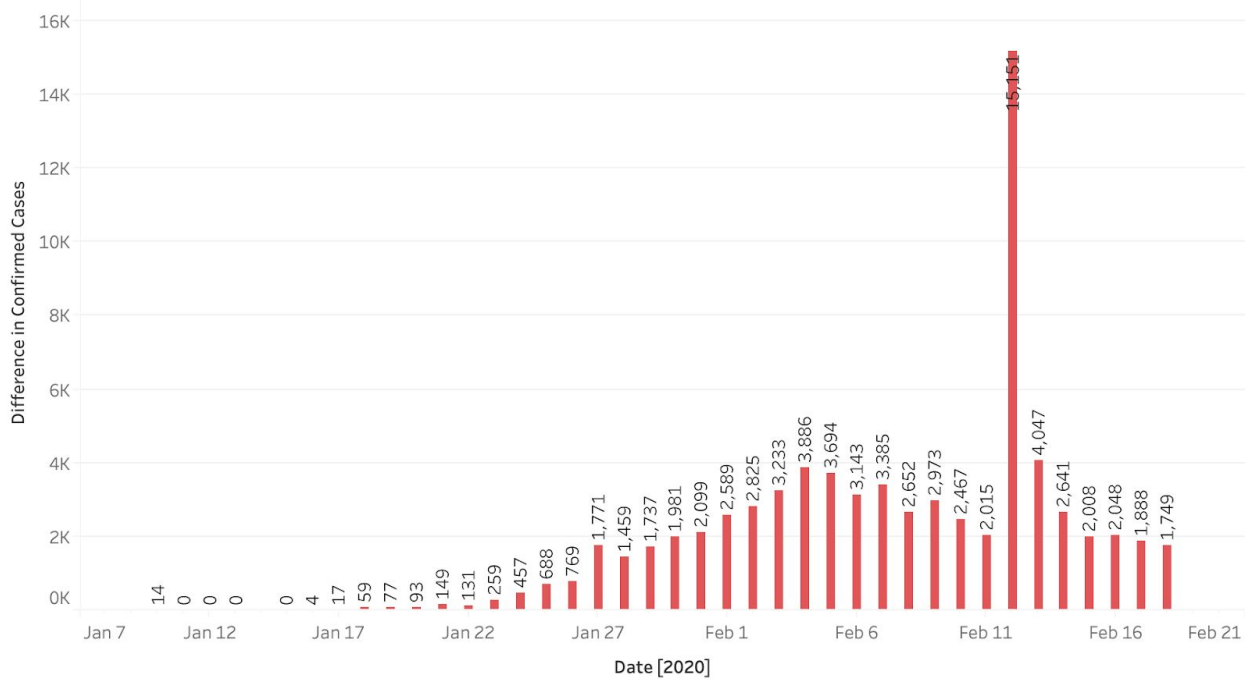
Worldwide View of Coronavirus Deaths (excluding China)



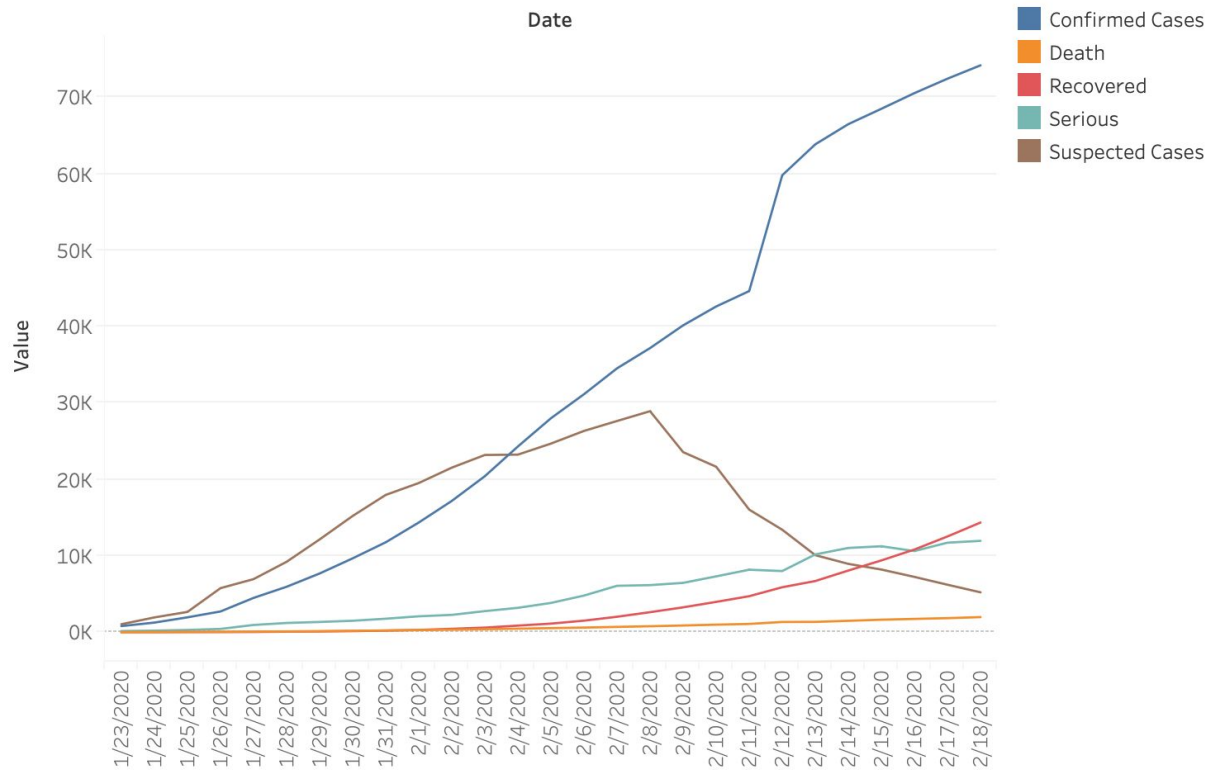
Worldwide View of Coronavirus Deaths (excluding China)



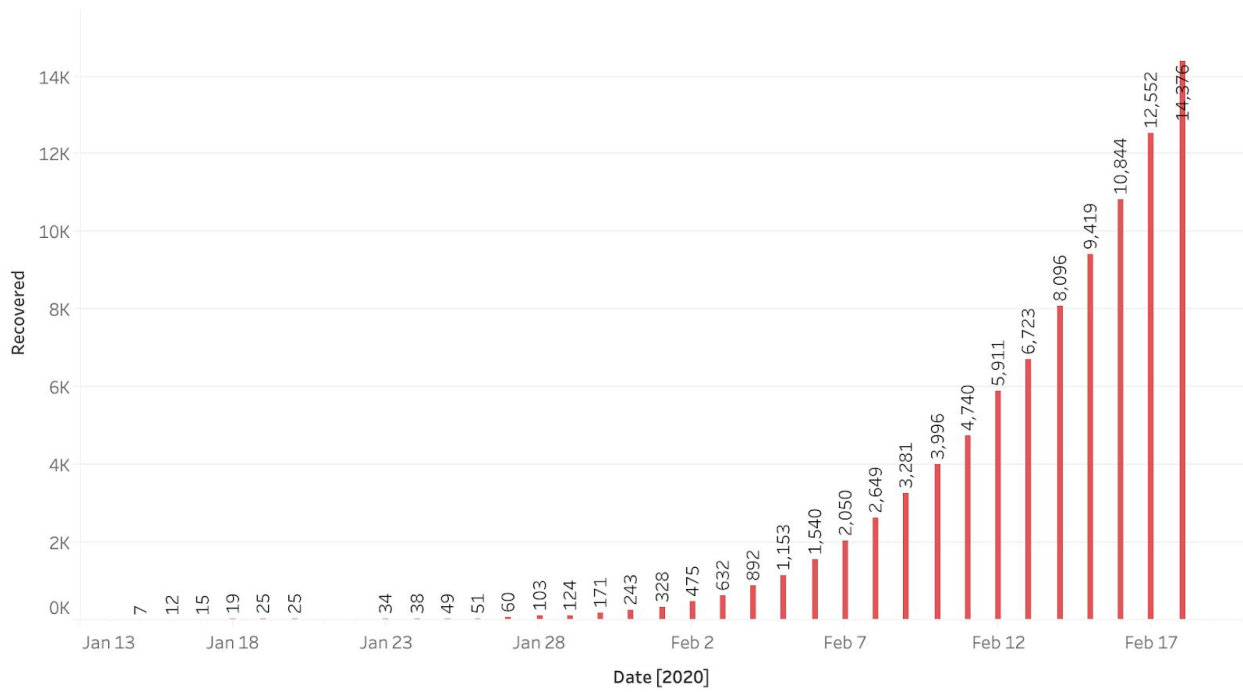
Daily change in total number of Wuhan coronavirus cases in China



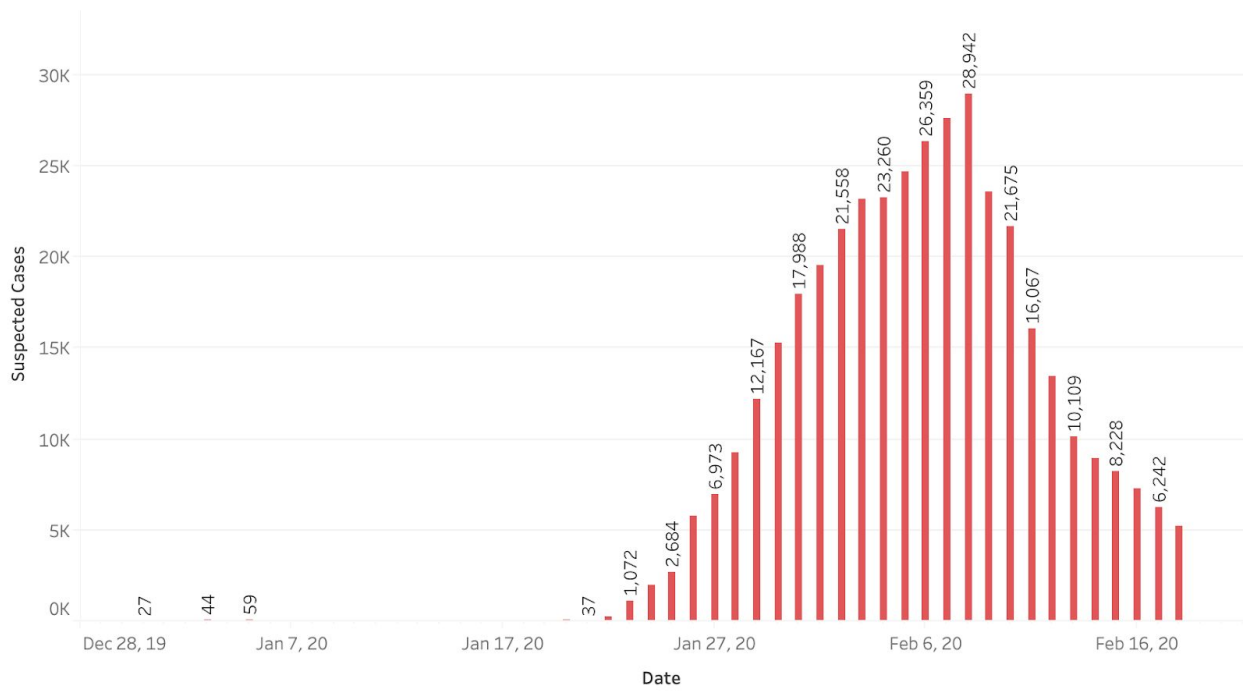
Wuhan coronavirus outbreak in China



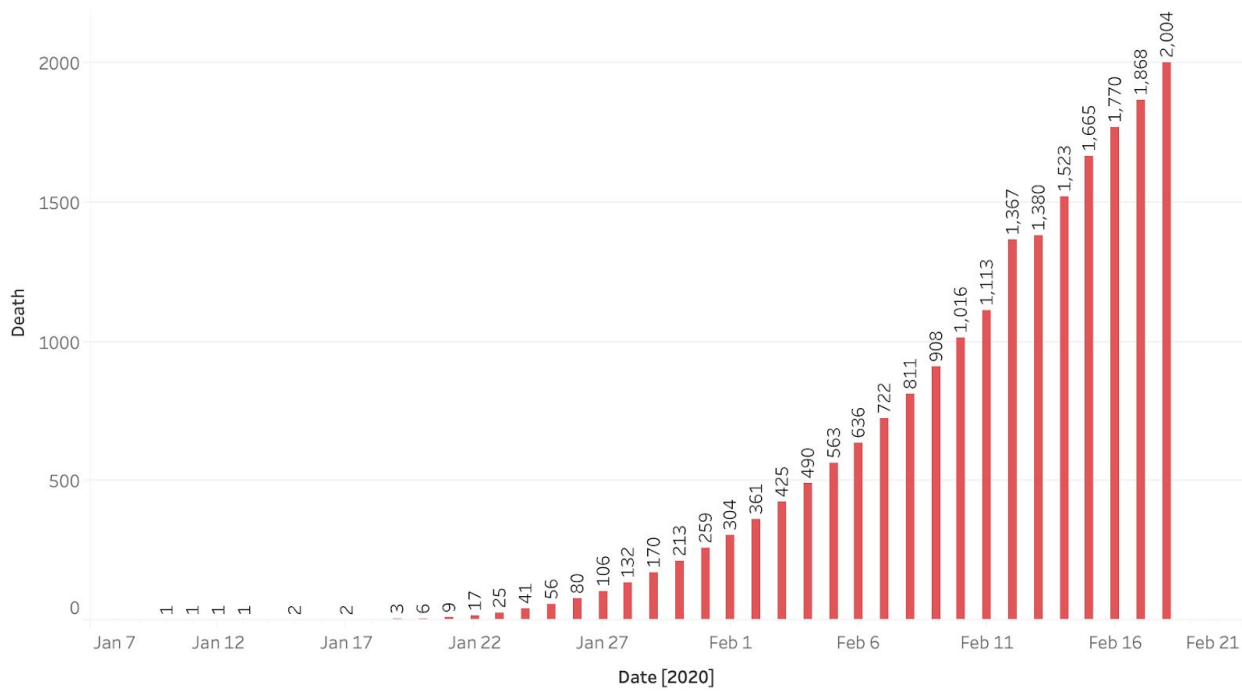
Number of recovered coronavirus patients in China (cumulative)



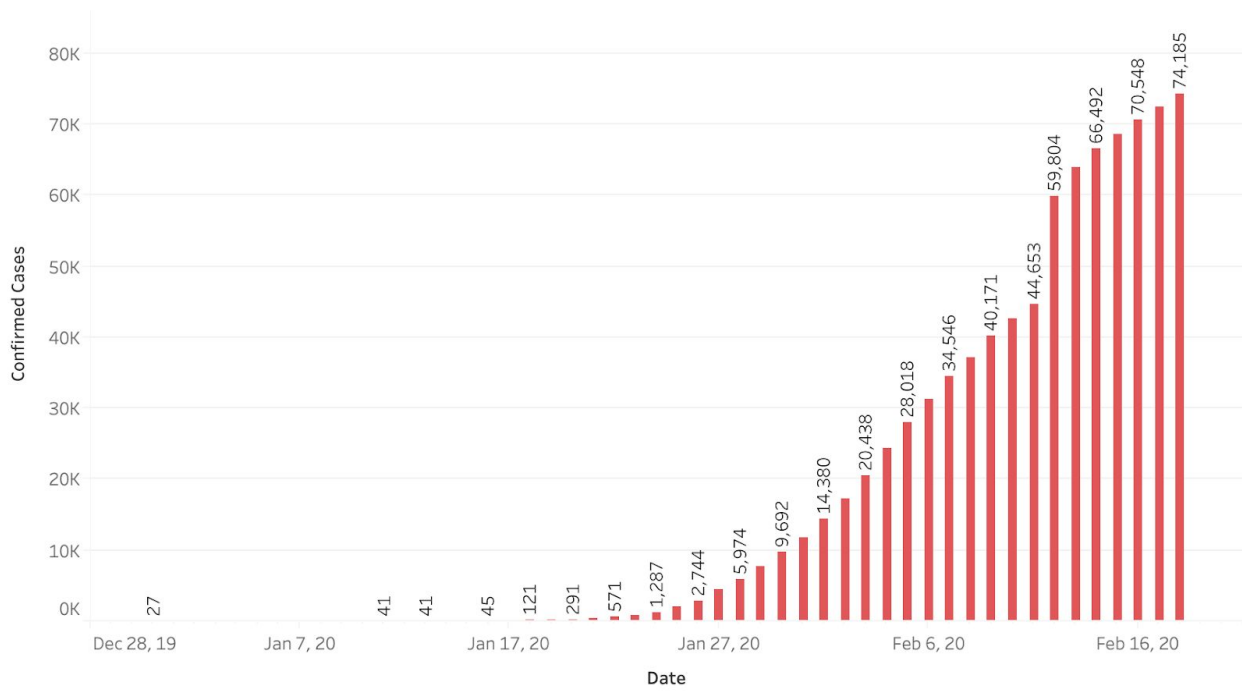
Number of suspected Wuhan coronavirus cases in China (cumulative)



Number of Wuhan coronavirus deaths in China (cumulative)



Number of confirmed Wuhan coronavirus cases in China (cumulative)



Milestone 3

A.

Here is a rough plan for the final visualizations we intend to create. Our audience is comprised of students who have taken at least an elementary statistics class and shown an interest in visualizing data.

There are several messages that we can tell with the data. The overall theme is the Impact of Outbreaks on the world within a 30 day period. Some of the areas of interest we believe are worth exploring include: impact of these outbreaks on world economies (we will have to create our datasets as none exist), the number of deaths (morality), confirmed cases and geographic spread, peak (an potential estimate of when corona might peak?), recovery rate, and airline disruption.

The outbreaks that we are studying include: Mers, Sars, Coronavirus, Ebola, and H1N1 (Swine Flu).

B.

Below you will find multiple exploratory visualizations and more than two draft explanatory visualizations, including drafts for each of the diseases, as well as an animated world map showing an overview of those outbreaks in the past 18 years.

Tentative Plan for Final Visualizations

Influenza H3N2

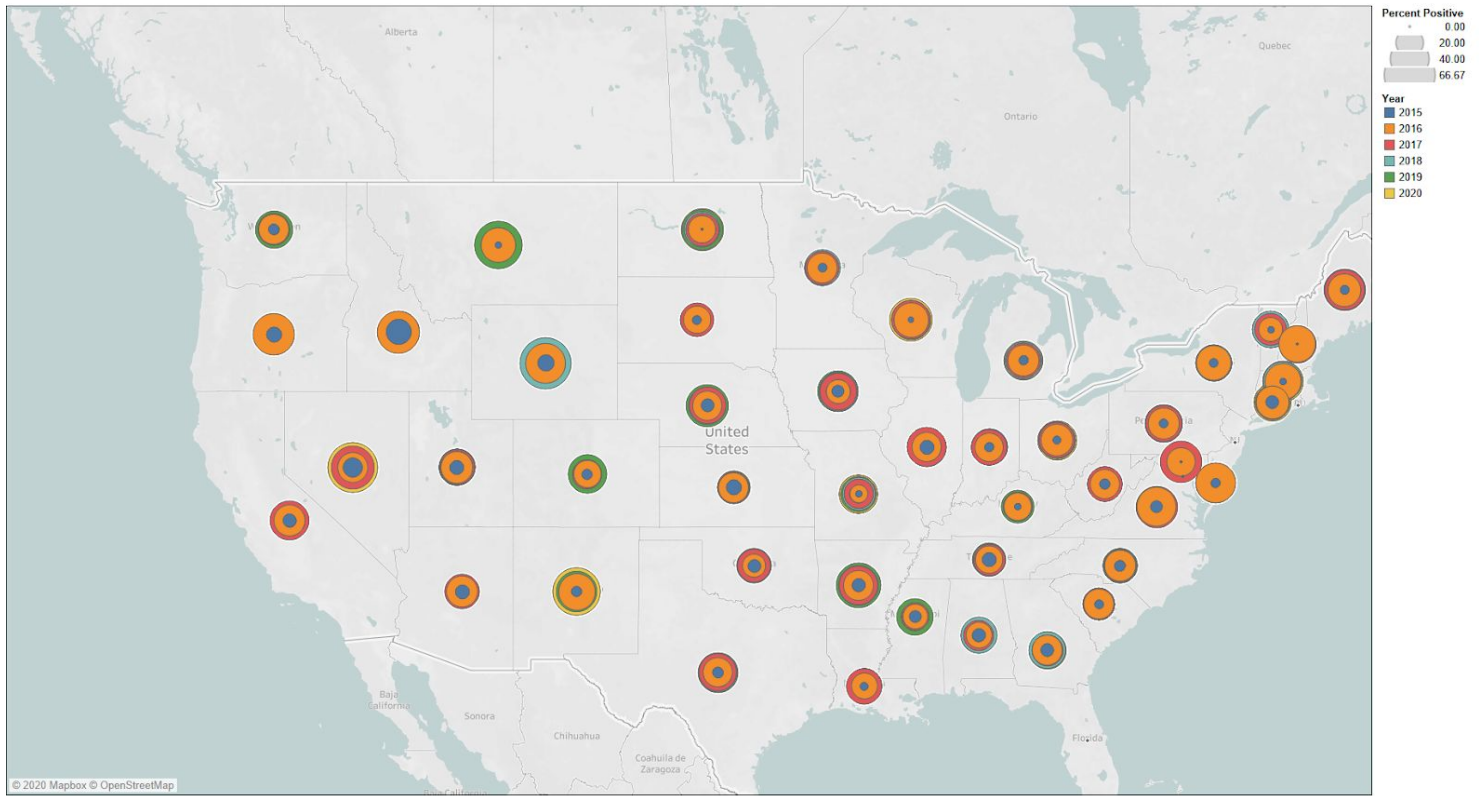
We can start with the less stressful and general situation of the viruses as Influenza by giving the data and virtualization of the USA. Then we can move our focus on the more serious subject of the outbreaks world wide like SARS and Ebola and then move to the current situation of the Coronavirus outbreak.

Infuenza cases in USA



Sum of Positive cases (color) broken down by Year vs. Region.

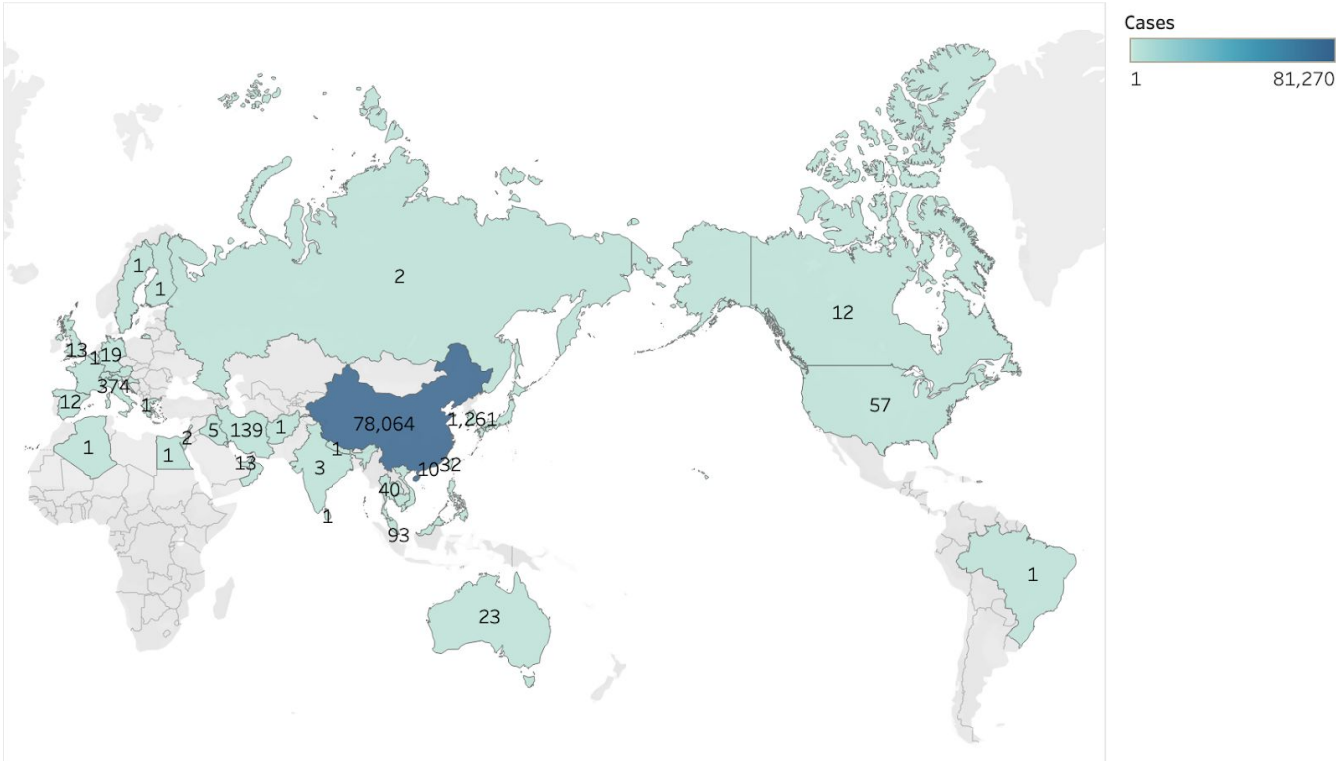
Percent of Influenza cases in USA from 2015 to 8th week of 2020



Coronavirus

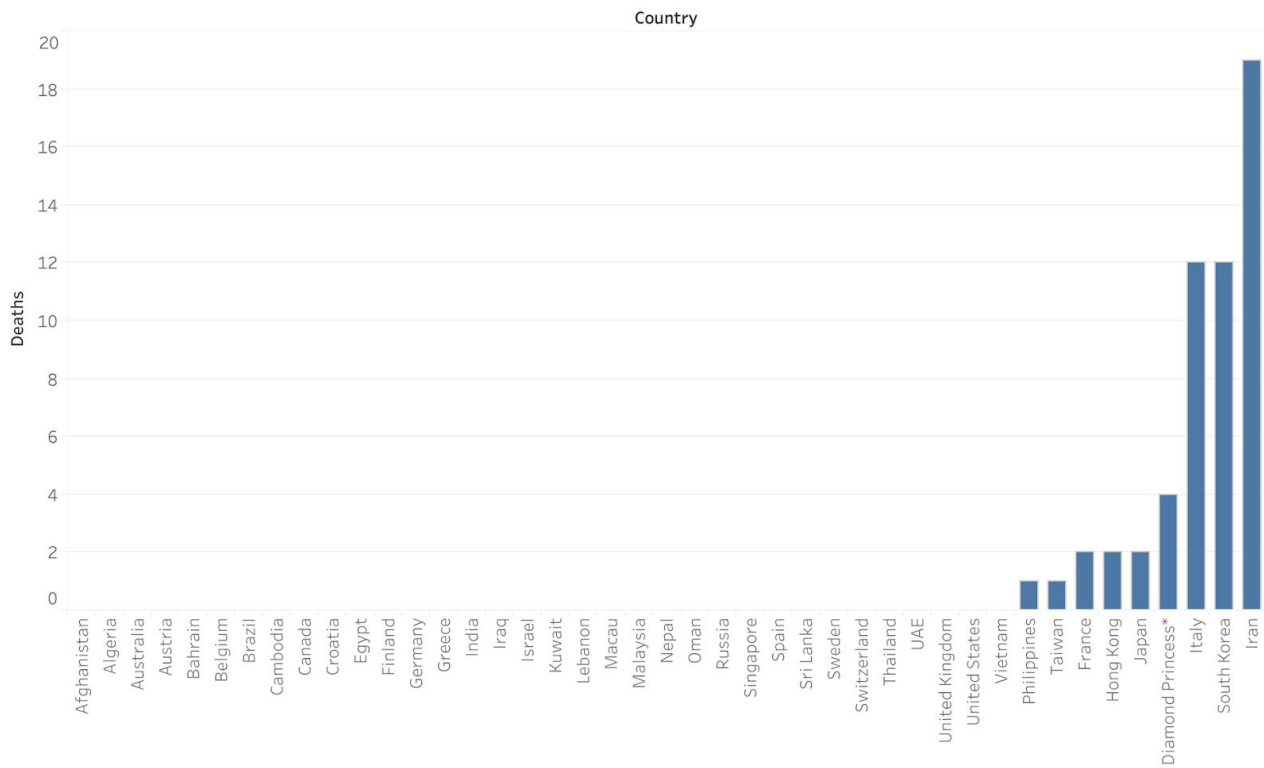
First, we will provide a geo-spatial map highlighting the overall number of confirmed Coronavirus virus cases worldwide. This will enable our audience to see at a glance which areas have the highest number of confirmed cases. We will do a similar map for deaths as well. These types of map will provide a quick visualization to determine how widespread the virus may be. I may also explore using a dot chart if the data is available. Furthermore, we will attempt to create an interactive map.

Worldwide View of Suspected Coronavirus Cases



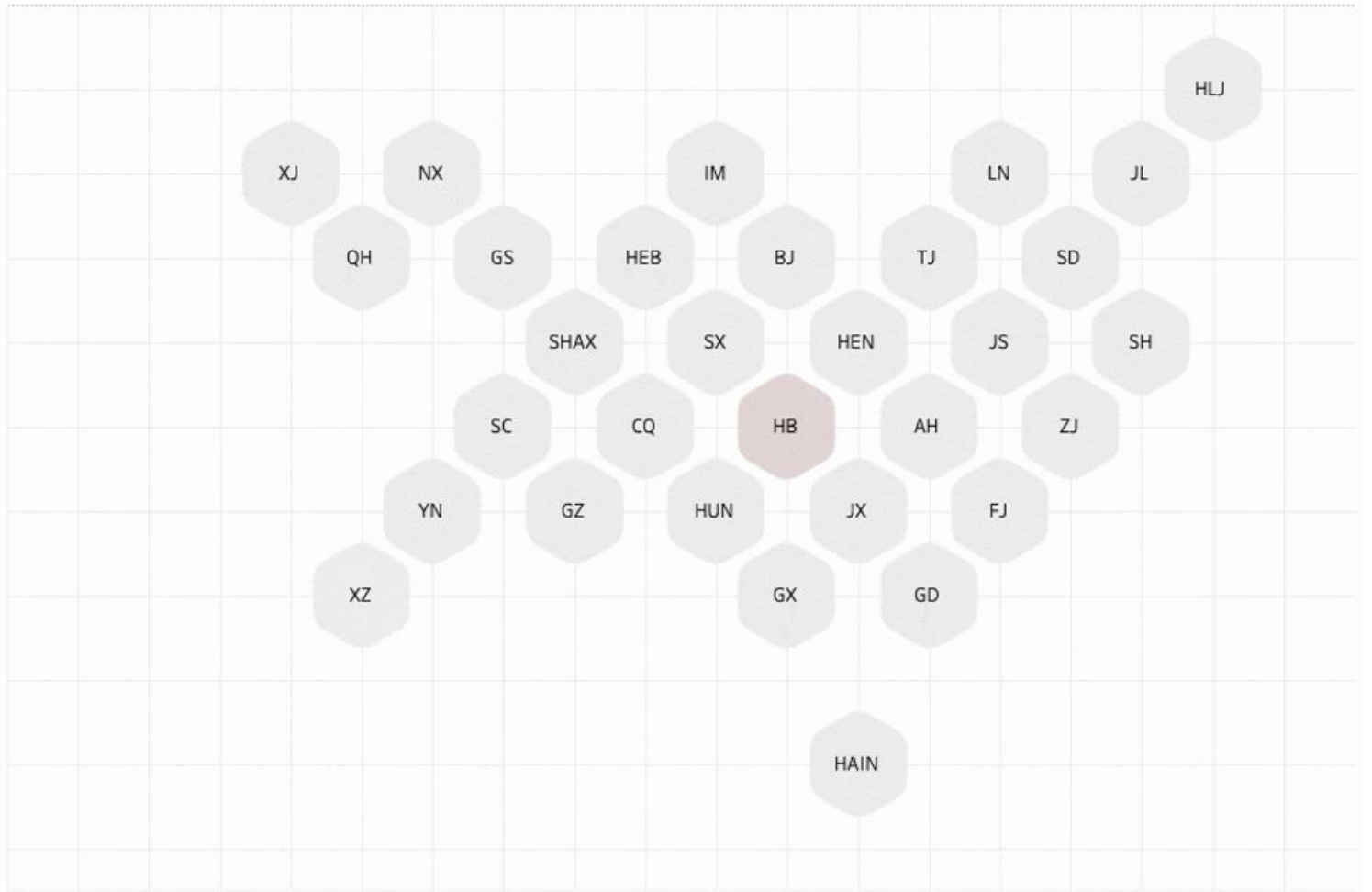
We would also consider creating visualization that show the effect of the outbreak on the economy.

Worldwide View of Coronavirus Deaths (excluding China)



The animation below shows how confirmed cases change in the COVID-19 outbreak in the early epic center, China.

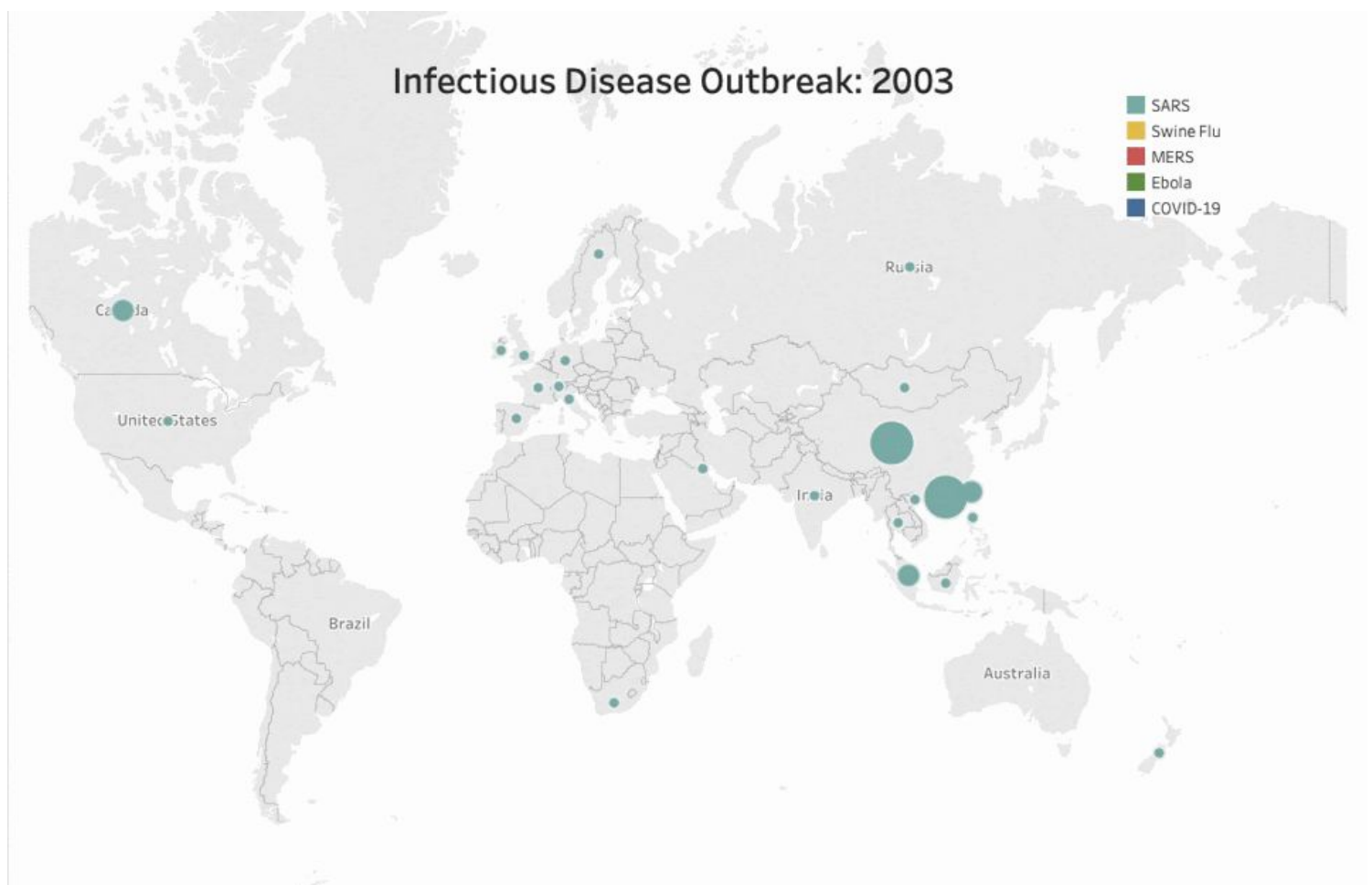
Daily Confirmed Cases In China - January 25, 2020



Infectious Disease Outbreak from 2002 to 2020

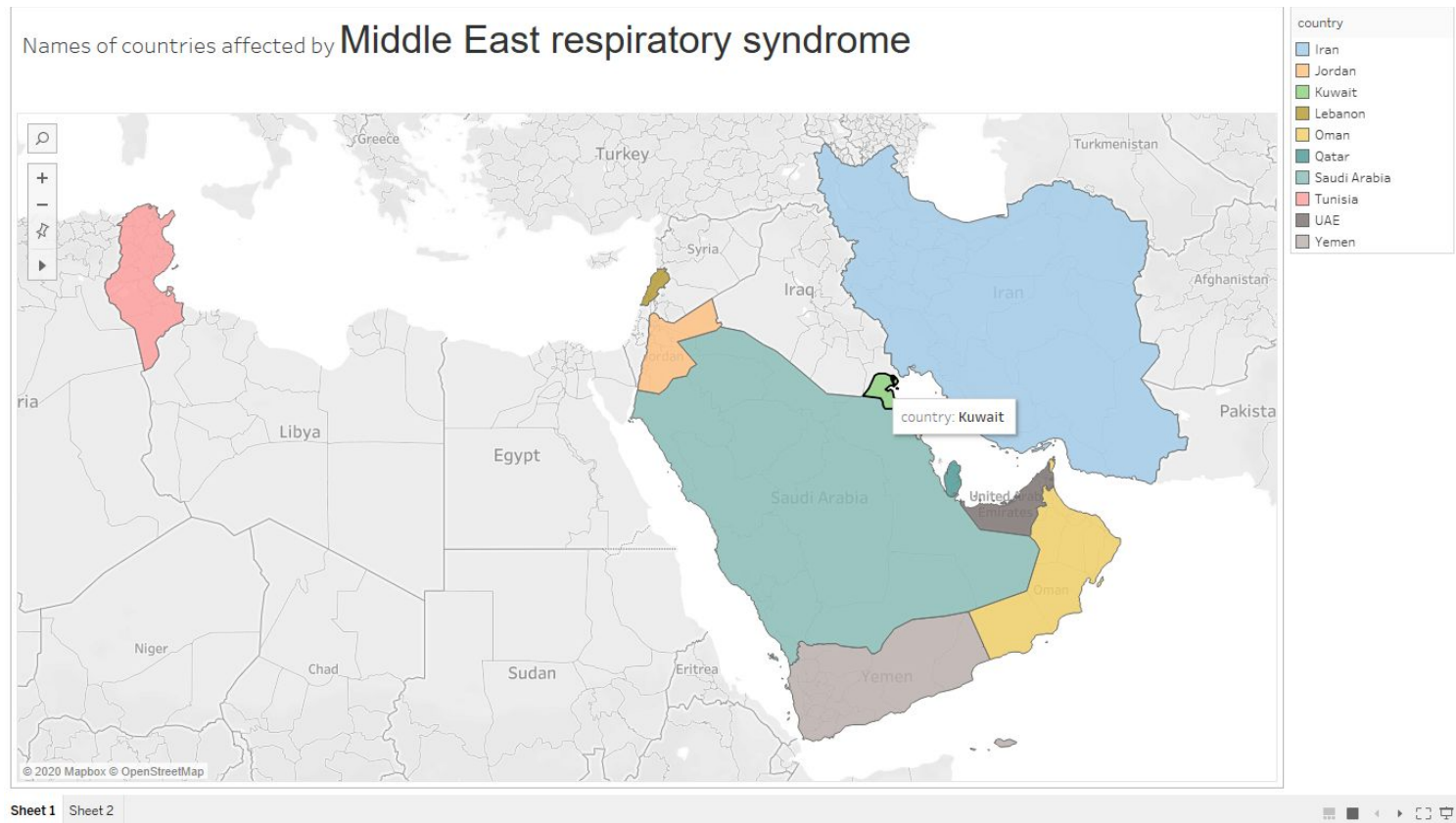
We plan to compare the outbreak speed, death tolls, affected countries and regions of these outbreaks of the five diseases.

So far, we have the graph below to show how the five major diseases outbreaked all over the world in the past 18 years. In the final report, we plan to show not just confirmed cases but also deaths. We are still looking for more related data.



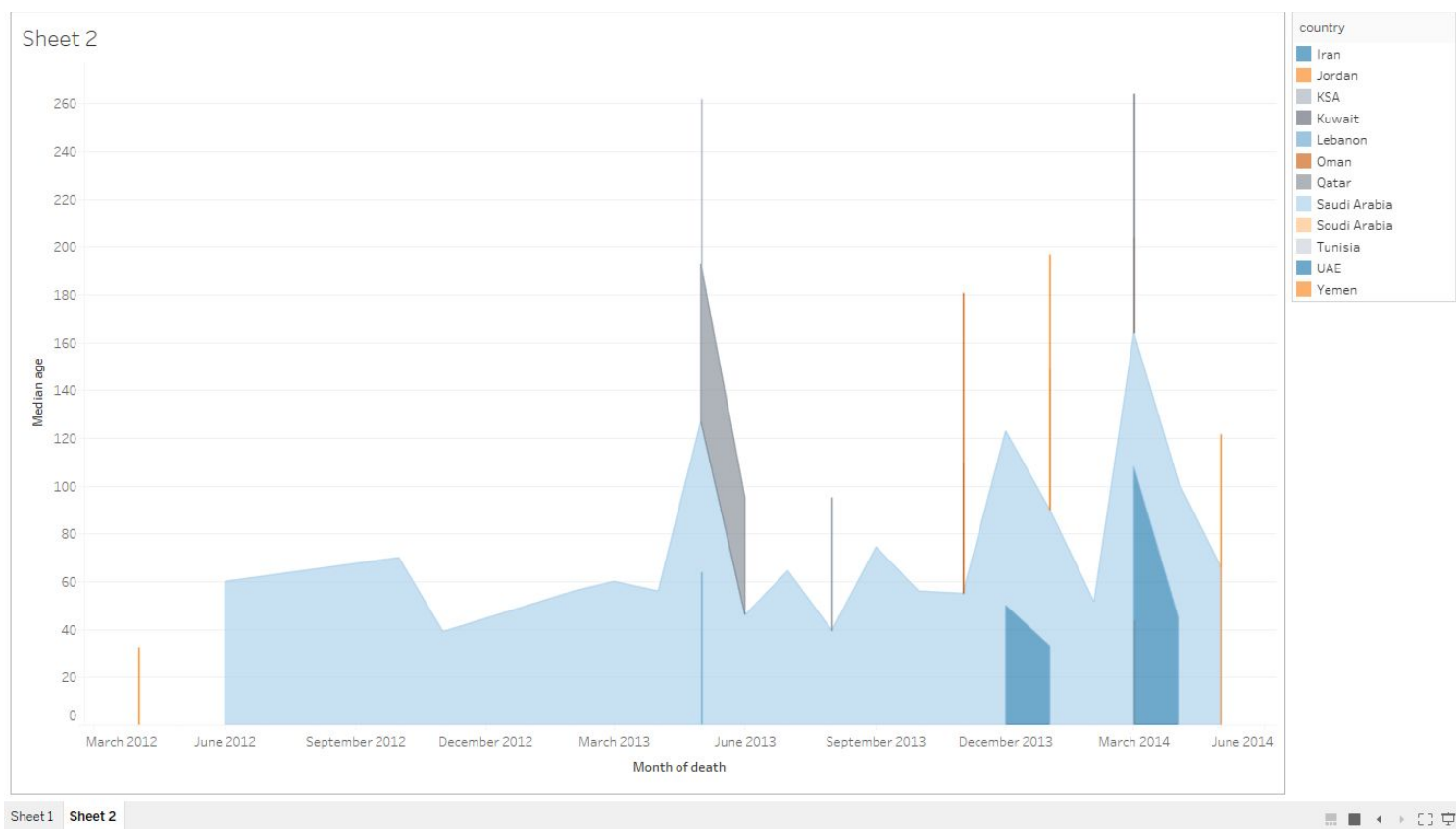
Middle East respiratory syndrome- Exploratory Graphs

Middle East countries affected by respiratory syndrome shown using graphical map

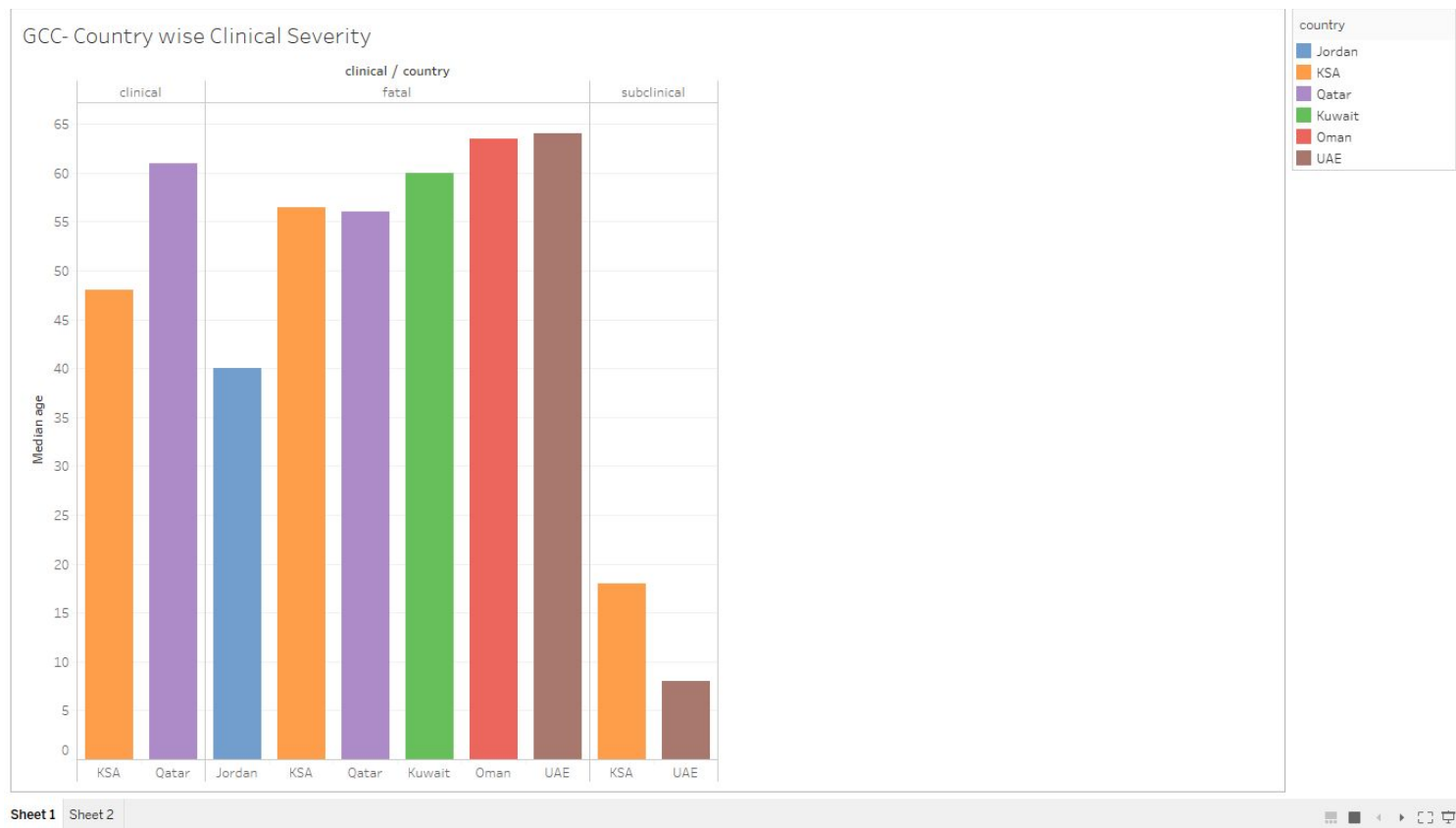


The below graph explains the median age of death by each GCC country affected by

MERS



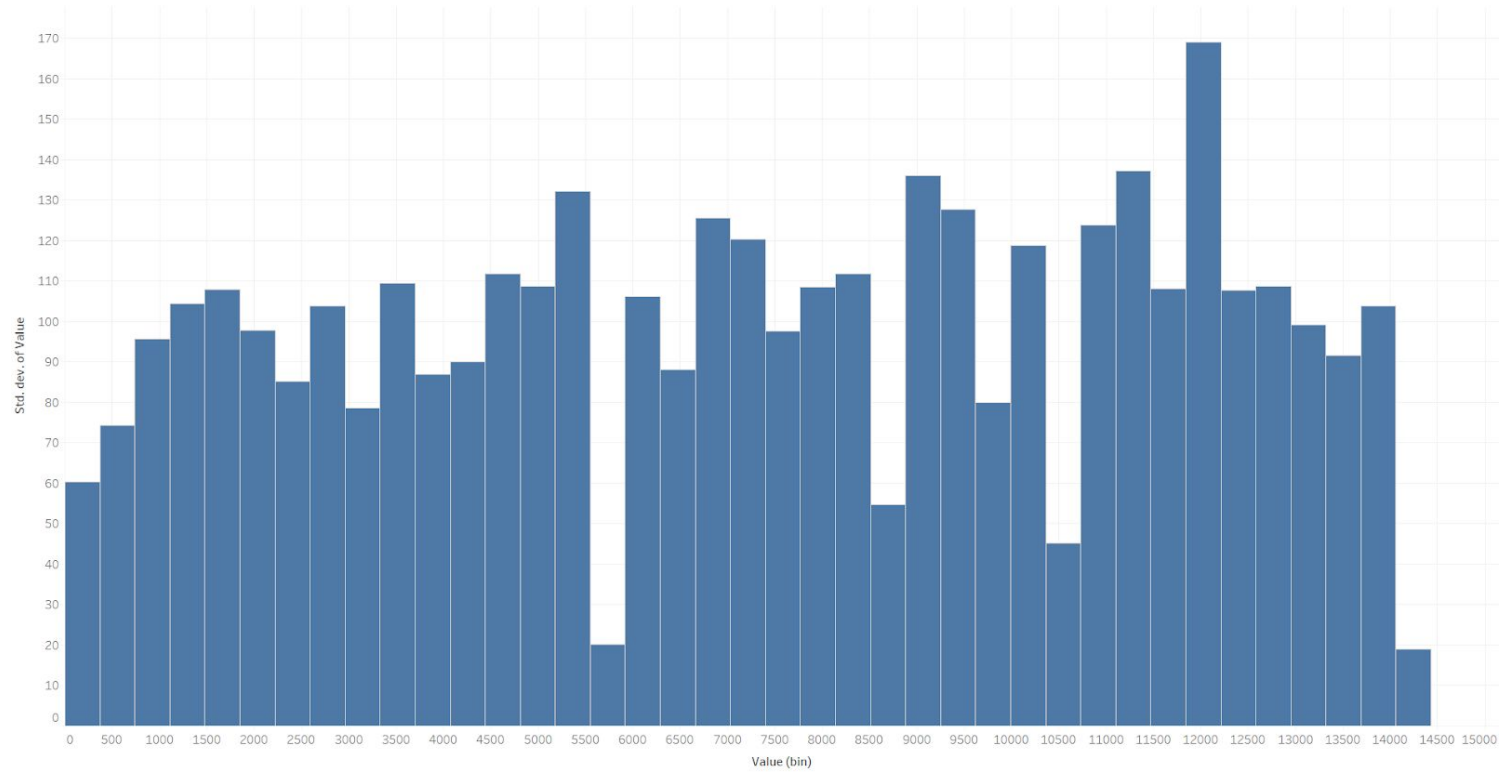
Below graph shows the age and clinical severity by the individual country shown using bar graph



Ebola outbreak

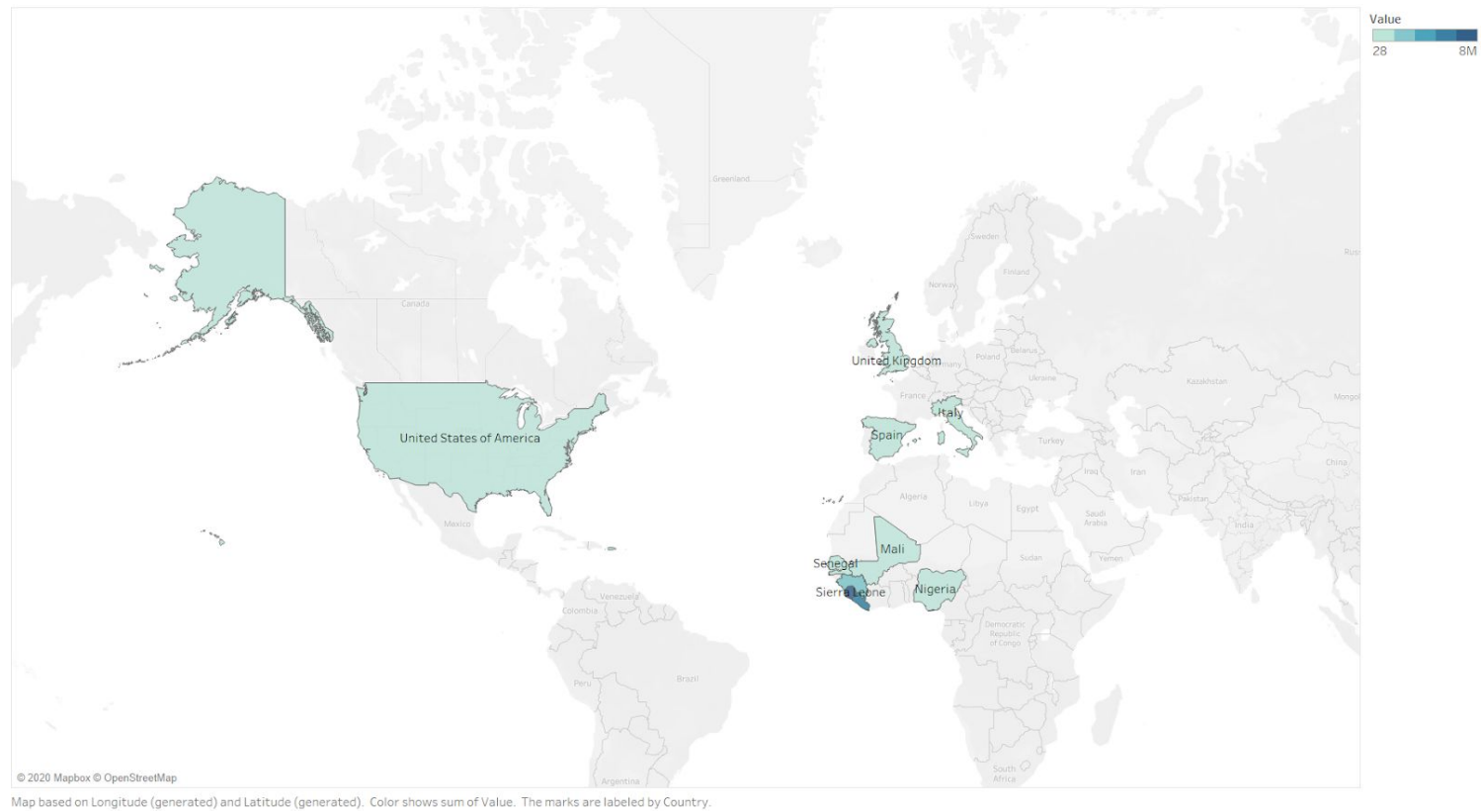
We are trying to visualize the impact of the ebola virus in different countries .We have also included the impact of ebola before 2014

Spread of Disease

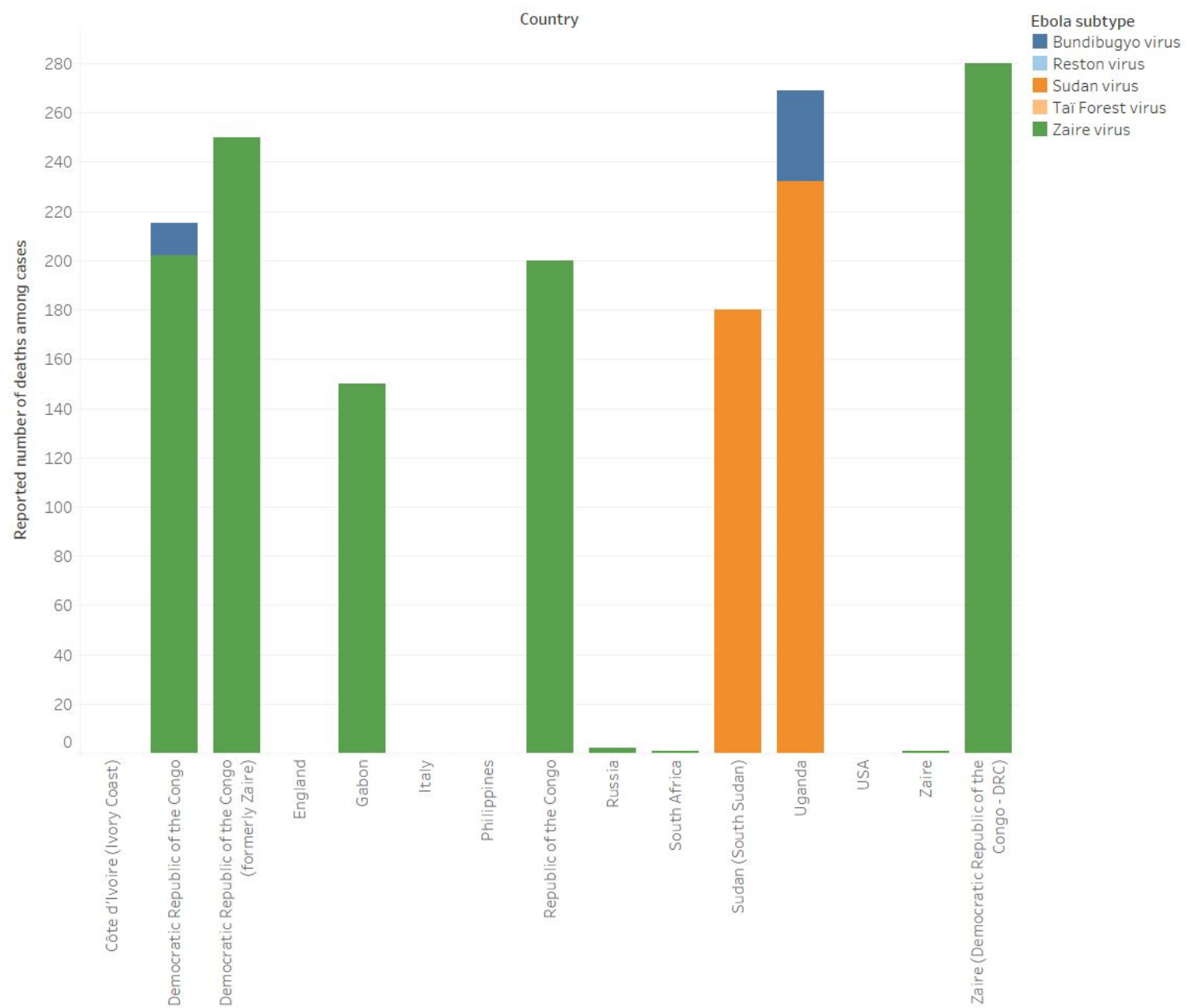


The trend of standard deviation of Value for Value (bin).

THE COUNTRIES THAT WERE AFFECTED

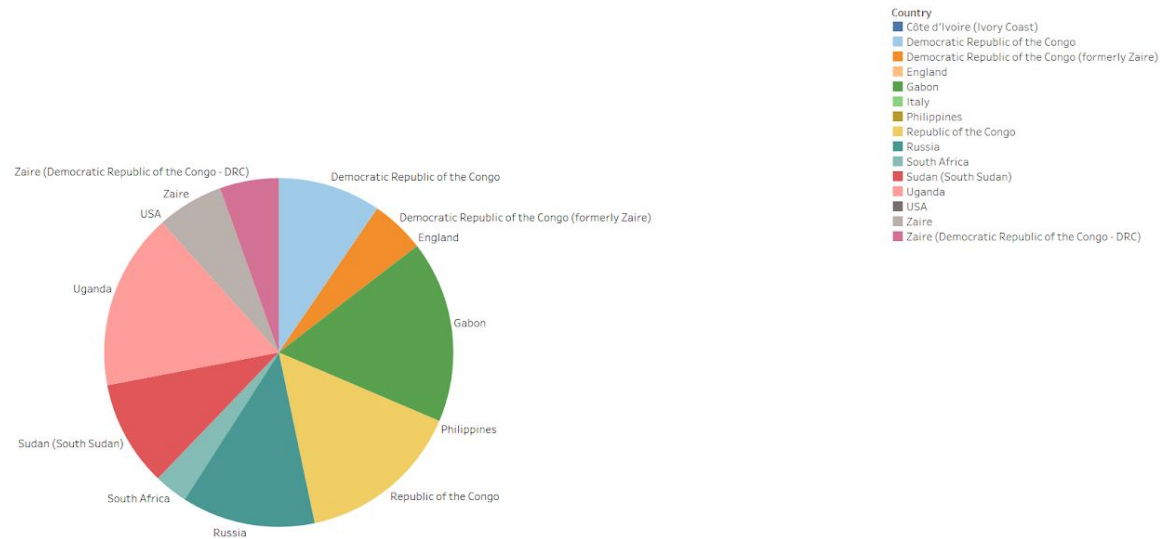


Number of deaths caused by the ebolal subtypes in diff countries



Sum of Reported number of deaths among cases for each Country. Color shows details about Ebola subtype.

Percentage of death Ebola in diff regions



Country. Color shows details about Country. The marks are labeled by Country.