COVID-19 CASE SURVEILLANCE

PART 1:

The dataset "COVID-19 Case Surveillance Public Use Data" contains 12 columns, including:

"cdc_case_earliest_dt": This date column provides the best estimate of when a COVID-19 case likely started, based on either the Clinical Date or the Date Received by CDC. If both dates are available, the earlier one is used. If no valid date is available, the field is left blank.

"cdc_report_dt": This column records the date when the case information was first reported to the CDC. It is deprecated, and it is recommended to use "cdc_case_earliest_dt" for analyses.

"pos_spec_dt": This column represents the date of the first positive specimen collection.

"onset_dt": If the patient is symptomatic, this column records the date of symptom onset.

"current_status": It indicates the case status as either "laboratory confirmed" or "probable."

"Sex": This column records the gender of the patient.

"age_group": It categorizes the patient's age into specific age groups.

"race_ethnicity_combined": This column captures the patient's race and ethnicity combined.

"hosp_yn": It indicates whether the patient was hospitalized ("Yes" or "No").

"icu_yn": This column indicates whether the patient was admitted to an intensive care unit ("Yes" or "No").

"death_yn": It records whether the patient died as a result of the illness ("Yes" or "No").

"medcond_yn": This column indicates the presence of underlying comorbidity or pre-existing medical conditions ("Yes" or "No").

Overall, the dataset contains essential information about COVID-19 cases, including dates, demographics, hospitalization status, and outcome, enabling researchers and analysts to study and understand the trends and characteristics of the COVID-19 cases.

Covid patients trend:

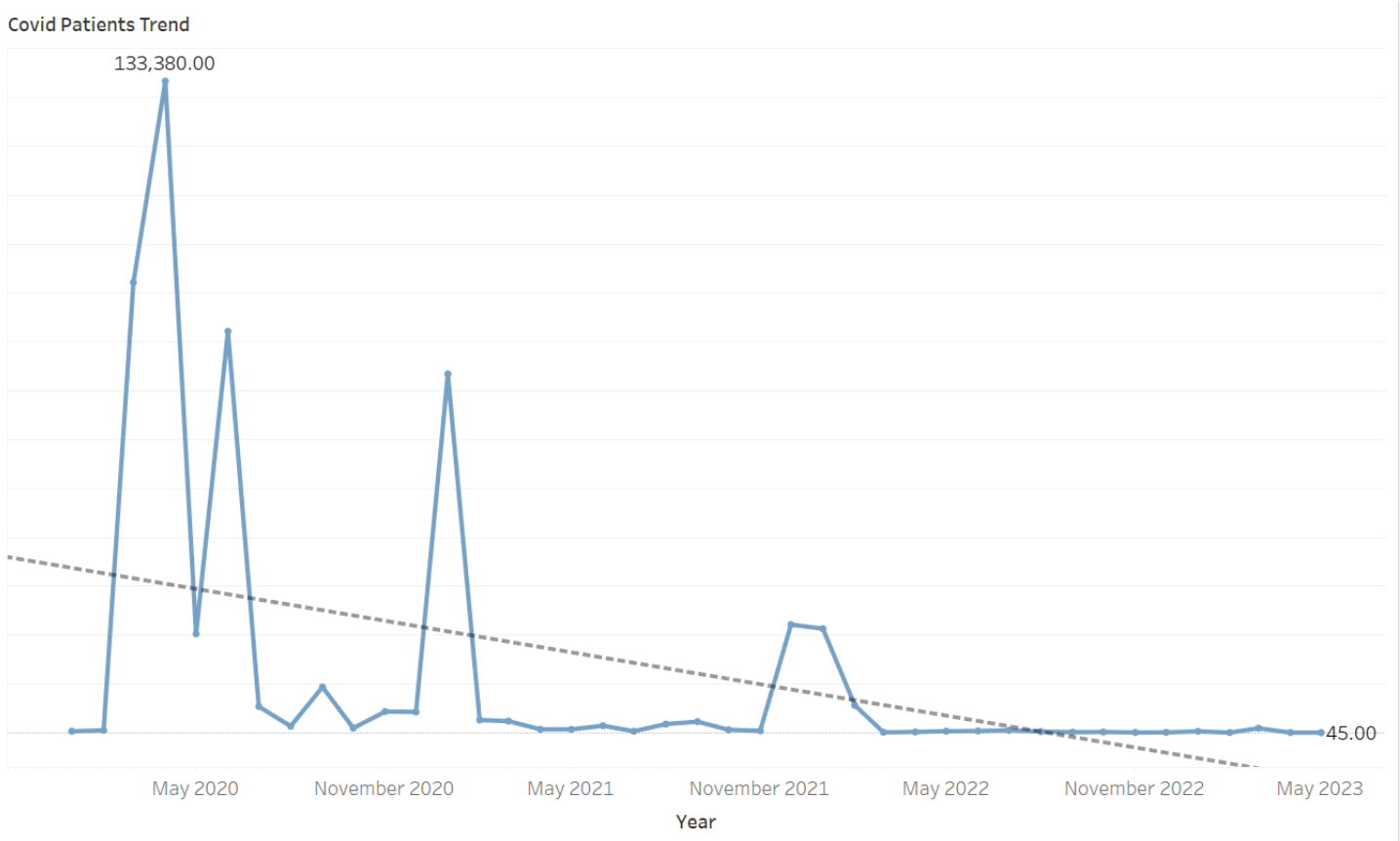
A line chart is a graphical representation of data points connected by straight lines. It is used to visualize the trend and changes in values over time or continuous intervals.

The line chart illustrates the growth trend of COVID-19 cases over a four-year period, from 2020 to 2023. The data is based on the date when each individual's COVID-19 case started. Initially, in April 2020, the number of cases peaked at 133,380, indicating a surge in infections. As time progressed, the number of cases showed fluctuations but generally decreased, reaching a minimum of 45 cases in May 2023.

The chart clearly shows a decline in the total number of COVID-19 cases over the given years, indicating an overall positive trend in controlling the spread of the virus. The presentation is designed with the general public in mind, making it easy for individuals without specific domain knowledge to understand the data and observe the progression of the pandemic from its peak to a significant reduction in cases by 2023.

The visualization is tailored for an audience categorized as "The Public." This line chart adheres to several Gestalt principles, making it visually cohesive and easy to comprehend. It demonstrates continuity by connecting data points

with smooth lines, allowing viewers to perceive the trend over time. The use of proximity groups related data points closely, enabling viewers to associate them as part of the same trend. Additionally, the similarity in color and style creates consistency across the chart, making it visually unified.



Division of patients by gender:

In Tableau, a treemap is a type of data visualization that uses nested rectangles to display hierarchical data. Each rectangle in the treemap represents a category or dimension within the data, and the size of the rectangle corresponds to a chosen measure or metric associated with that category. This visualization allows users to explore the composition and distribution of data across different categories, making it particularly useful for understanding hierarchical relationships and identifying patterns.

Now, let's explain the above case without using it as an example:

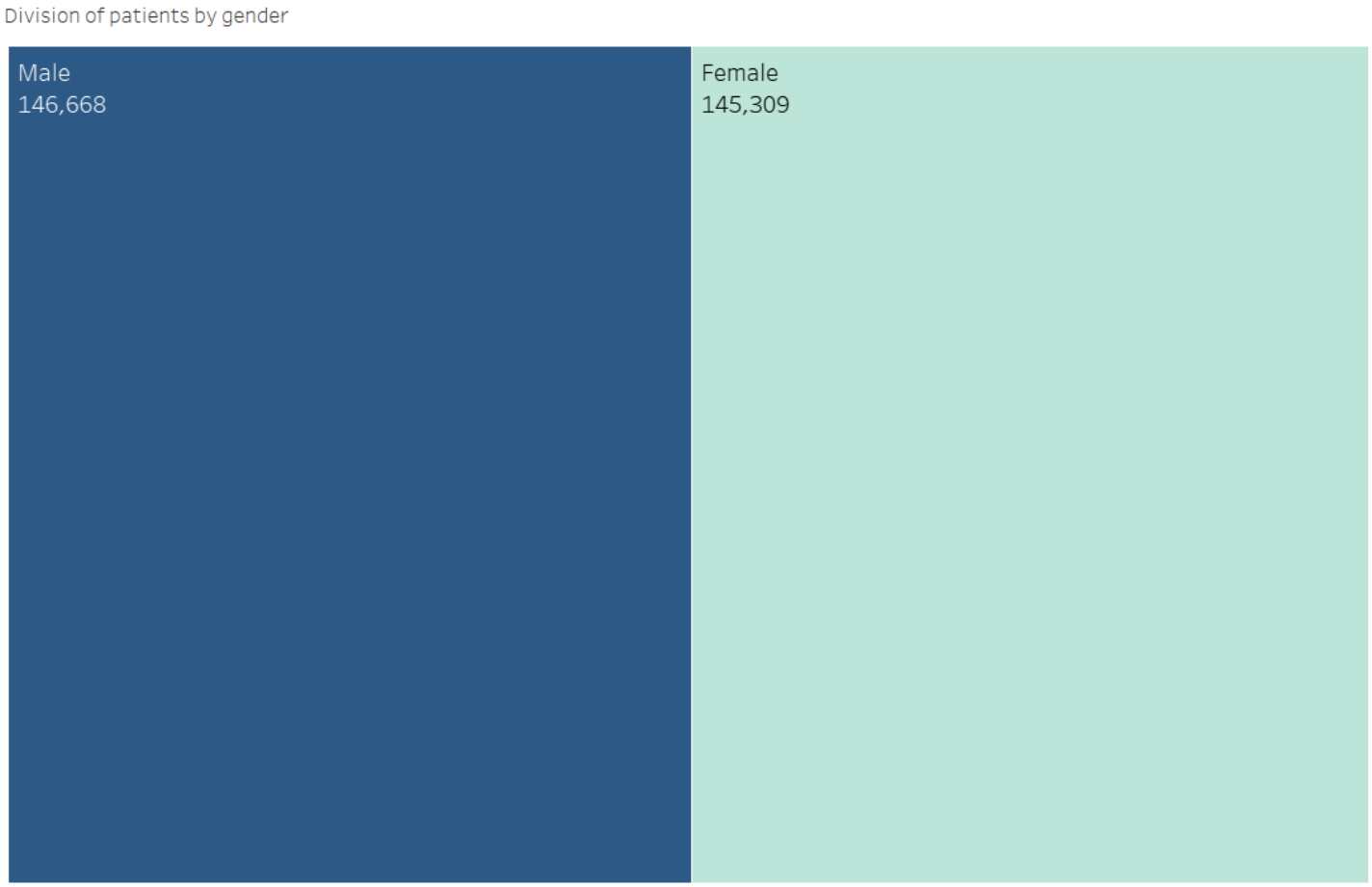
In the given case, a treemap is used in Tableau to represent data from the period 2020 to 2023, specifically showing the number of individuals affected by the coronavirus, categorized by gender. The treemap consists of two main rectangles: one representing males and the other representing females.

The size of each rectangle in the treemap corresponds to the total number of COVID-19 cases for that particular gender during the specified time frame. By comparing the sizes of these rectangles, it becomes evident that the number of male individuals affected by the pandemic is slightly greater than the number of female individuals affected.

This treemap visualization provides a clear and concise way to compare the impact of the pandemic on males and females, allowing for quick and easy understanding of the relative vulnerability of each gender group over the four-year period.

The design of this chart is specifically intended for "The Public" to understand. This treemap adheres to Gestalt principles, ensuring a clear and straightforward representation. The principle of proximity is applied by placing the two

rectangles, representing male and female COVID patients, in close proximity, allowing viewers to associate them as part of the same chart. The principle of similarity is observed as both rectangles share the same visual style and color, creating consistency and visual unity. This treemap effectively uses these Gestalt principles to present the data in a concise and understandable manner, making it accessible to "The Public" audience.



Most Vulnerable Age Group:

A vertical bar chart, also known as a column chart, is a type of data visualization that uses vertical bars to display the relative frequencies or quantities of different categories or data points. Each bar represents a specific category, and the height of the bar corresponds to the value or count associated with that category.

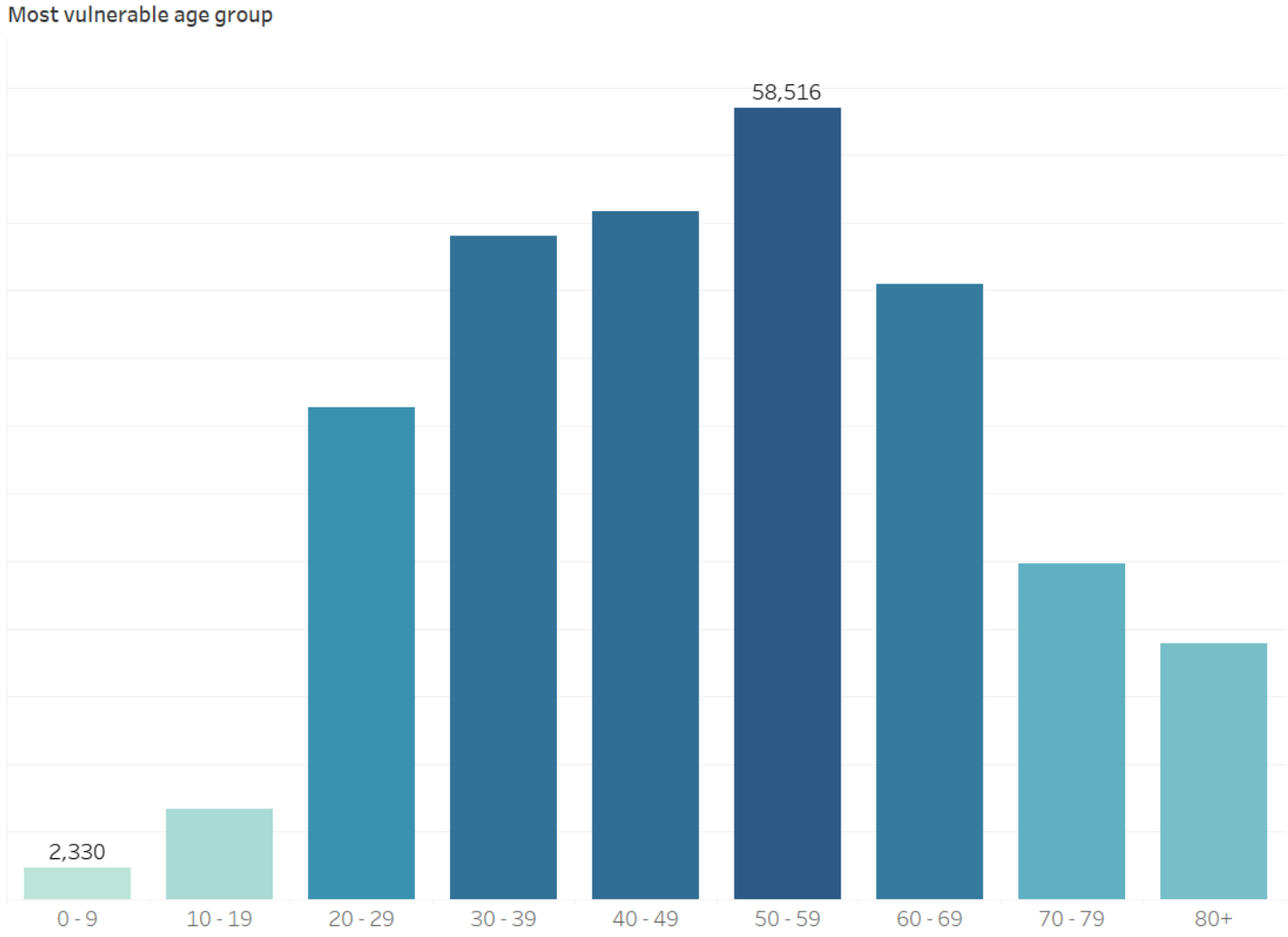
In this context, the vertical bar chart is used to present data on the number of patients falling within various age groups over a four-year period from 2020 to 2023. The chart displays multiple age groups on the horizontal axis, and the vertical axis represents the count or number of patients in each group.

The chart allows us to quickly identify trends and patterns in the data. From the chart, it becomes evident that the age group with the highest number of patients is 50-59, with a total count of 58,516 patients. On the other hand, the age group with the fewest patients is 0-9, with only 2,330 patients.

By comparing the heights of the bars, we can observe the relative vulnerability of different age groups to the disease. The data suggests that people in their 50s are more susceptible to the disease compared to other age groups. Additionally, the age group of 0-9 seems to be the least affected.

This visual representation makes it easy to comprehend the distribution of patients across age groups and to identify which groups are more at risk and which ones are relatively less affected. Vertical bar charts are widely used in various fields to present categorical data and help in making data-driven decisions based on visual insights.

The intended viewers for this visualization are people from "The Public" category. The bar chart effectively incorporates Gestalt principles to facilitate clear data representation. The principle of proximity is demonstrated as each bar is positioned close to each other, grouping them according to age groups. The principle of similarity is evident as each bar shares a consistent width and style, creating visual unity. The tallest bar being the highest number of patients showcases the principle of closure, making it easy for viewers to identify the most affected age group. Through these principles, the bar chart offers an intuitive and accessible visualization of patient data for "The Public" audience.



Affected Ethnicities:

A highlighted table in Tableau is a visually enhanced way of presenting tabular data. Users can format or highlight cells based on specific conditions or measures, emphasizing important information and patterns within the table. This visual enhancement helps viewers quickly identify trends and key data points. In this highlighted table, the first column contains different ethnicities affected by the pandemic, while the second column shows the number of people from each ethnicity impacted by the virus. The data reveals that the category with the least number of patients is Native Hawaiian, with 364 individuals affected. On the other hand, the ethnicity with the highest number of people affected is White, Non-Hispanic, with a total of 39,327 cases.

The use of conditional formatting or highlighting in the table could emphasize these numbers further, drawing attention to the disparities in COVID-19 cases across various ethnicities. This presentation makes it easier for users to quickly identify the ethnic groups that have been more or less affected by the pandemic.

This chart is specially created to cater to the understanding of individuals who fall under the category of "The Public." The highlighted table incorporates Gestalt principles to optimize data comprehension. The principle of similarity is applied through the consistent use of a teal blue color palette, making it easy to distinguish related data points. The principle of proximity is observed as the table arranges data rows and columns closely, aiding viewers to associate relevant information. These principles enhance the table's visual organization, allowing the general to quickly grasp the number of patients affected in each ethnicity.

Affected Ethnicities

Race Ethnicity	
American Indian	386
Asian, Non-Hispanic	6,862
Black, Non-Hispanic	37,483
Hispanic/Latino	39,327
Multiple/Other	6,916
Native Hawaiian	364
White, Non-Hispanic	68,797

Symtomatic?:

A table in Tableau is a basic tabular representation of data, displaying rows and columns without any specific formatting or visual enhancements. It presents the data in a straightforward manner, making it easy to view and analyze the values in a structured format.

This table contains data from 2020 to 2023, consisting of five columns and two rows. It compares the number of patients who displayed symptoms with those who did not. In 2020, 2021, and 2023, the number of symptomatic patients was lower than the number of asymptomatic patients. However, in 2021, this relationship was reversed. It is also notable that the number of symptomatic patients was highest in 2020 and gradually decreased in 2023, likely due to a decline in overall COVID-19 cases. Conversely, the maximum number of asymptomatic patients was observed in 2020, and it reduced by 2023.

"The Public" is the focus of this chart's design and presentation. The simple table follows Gestalt principles to create a clear and structured presentation. The principle of similarity is observed as the table uses consistent formatting and styling for symptomatic and asymptomatic patient data. The principle of proximity is applied by arranging the data in rows and columns, grouping related information together. These principles ensure that audience can easily comprehend the number of symptomatic and asymptomatic patients for each year from 2020 to 2023.

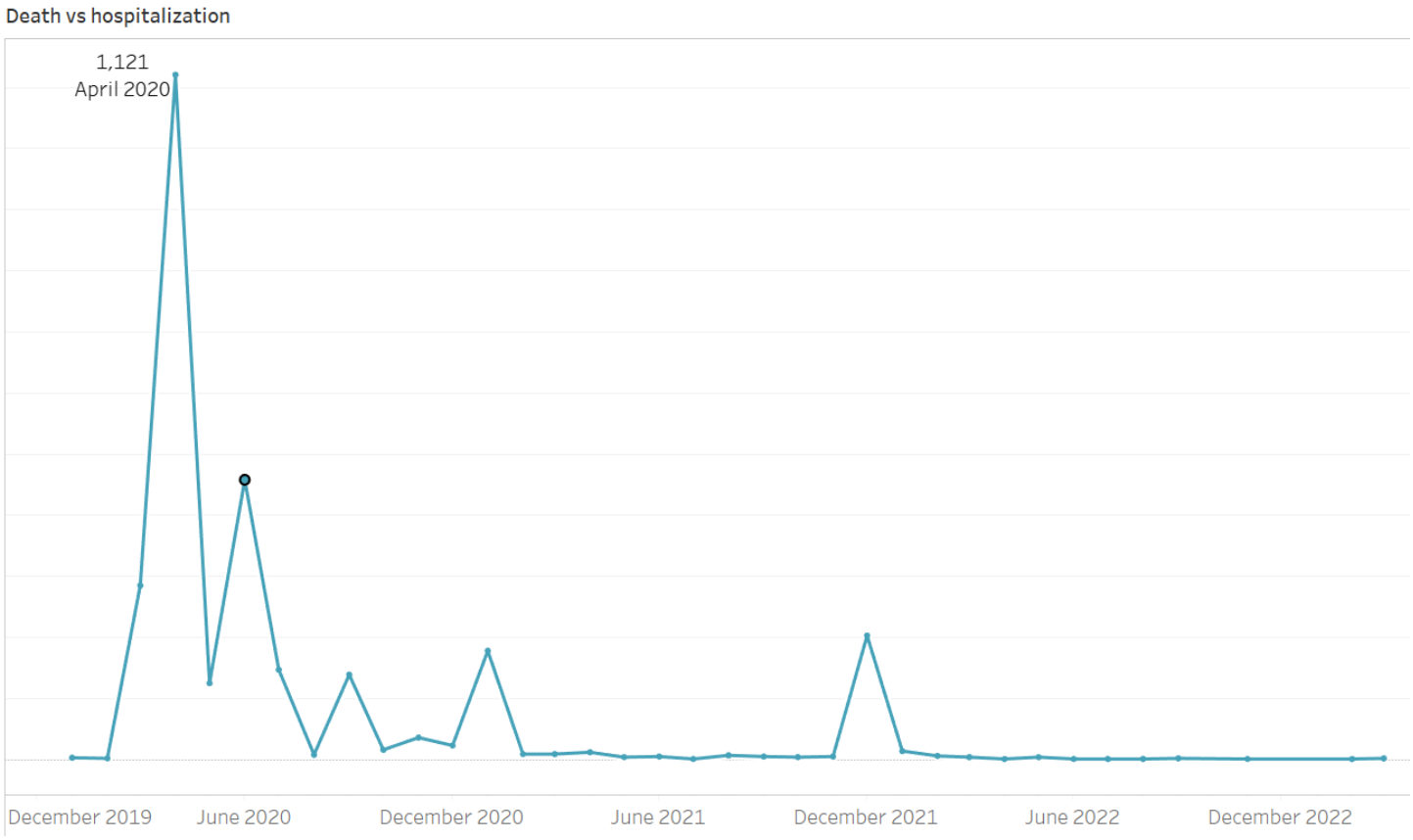
Symptomatic?

	2020	2021	2022	2023
Yes	68,448	50,816	6,002	158
No	101,561	45,669	20,462	228

Death vs Hospitalization:

This line chart presents intriguing insights derived from data spanning 2020 to 2023. It illustrates the number of patients who unfortunately passed away due to illness but were unable to receive hospitalization. This trend indicates a scarcity of healthcare facilities during the early years of the COVID-19 outbreak. However, as the pandemic intensified, countries took measures to strengthen their healthcare infrastructure, leading to an initial upward trend in the chart, followed by a subsequent downward trend.

The target demographic for this visualization is those belonging to "The Public," ensuring it is user-friendly and easy to interpret. The line chart effectively adheres to Gestalt principles for a coherent representation. The principle of continuity is evident as the chart connects data points with smooth lines, enabling viewers to perceive the progression of people who died but were not hospitalized over the years. The principle of proximity is applied by closely aligning the data points along the timeline, illustrating the trend in a cohesive manner. By following these principles, the line chart provides an easily understandable visualization of the data, making it accessible to the audience.



PART 2:

COVID-19 CASE SURVEILLANCE:

The purpose of a Tableau dashboard is to present data in a visually interactive format, enabling users to gain insights and make data-driven decisions effectively. It consolidates key metrics and trends on one screen for easy analysis and communication of information. In our dashboard, designed for "The Public," we've included six easy-to-understand charts: two line charts displaying COVID patient trends and death versus hospitalization, a highlighted table showing affected ethnicities, a bar chart depicting the most vulnerable age group, a simple table indicating symptomatic cases, and a treemap illustrating the division of patients by gender. Our goal was to present the data in a visually appealing and

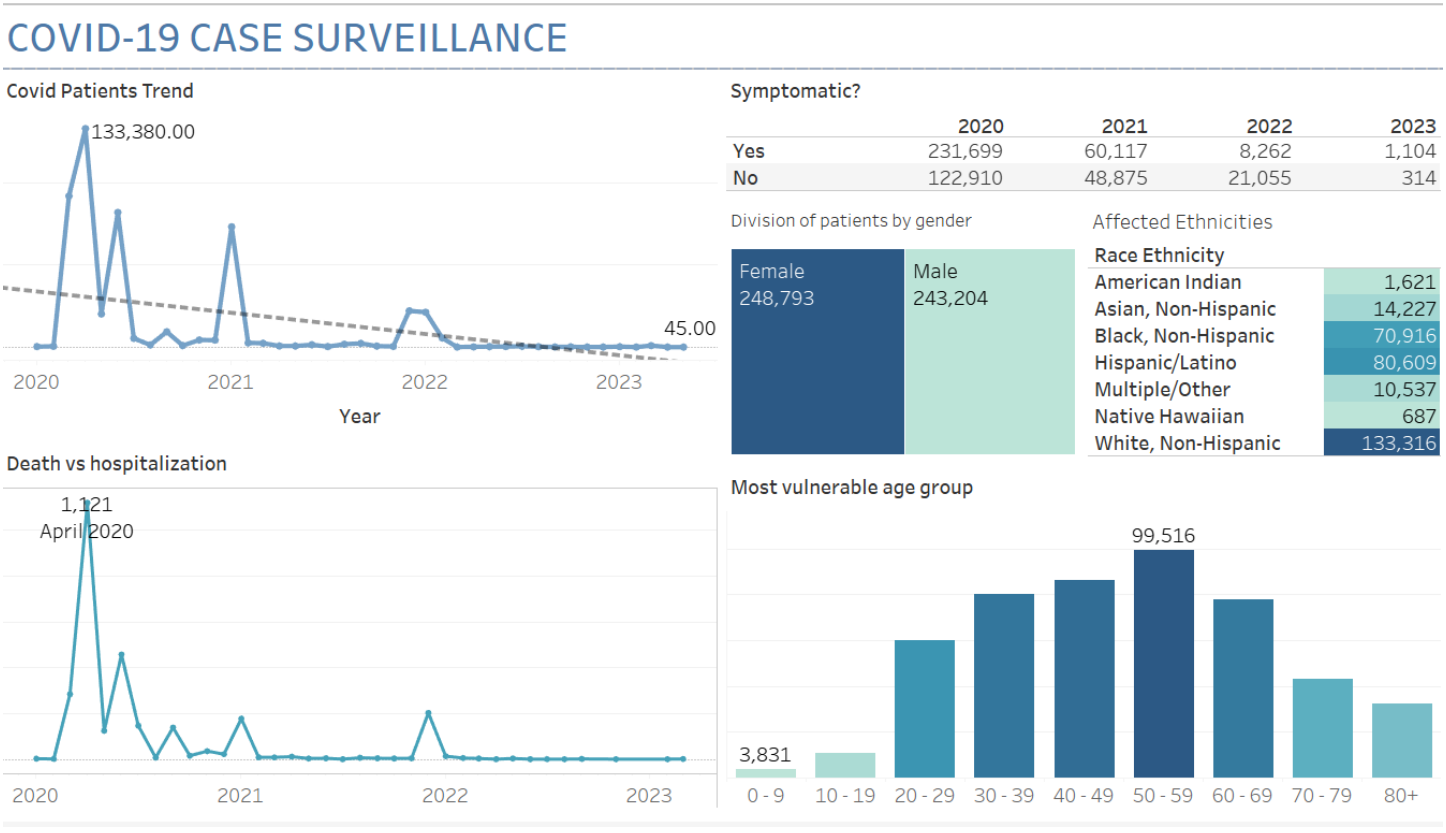
accessible manner, ensuring that even individuals without extensive domain knowledge can easily comprehend the information provided as our chosen audience is **"The Public".** The dashboard sheds light on patient data analysis from 2020 to 2023, providing insights into the overall patient trends, the proportion of symptomatic cases, affected ethnicities, vulnerable ethnicities shown through a bar chart, and the trend of death rate versus hospitalization rate. It offers a comprehensive and detailed analysis, encompassing valuable insights discussed in the overall summary.

This dashboard strategically utilizes Gestalt principles to optimize data visualization and ensure a user-friendly experience for "The Public." Firstly, it employs the principle of Proximity by grouping related charts and tables together, allowing viewers to quickly associate and comprehend data that belongs to the same context. Secondly, Consistency in colors and styles across the dashboard enhances visual harmony, making it easier for users to recognize patterns and maintain focus.

The principle of Continuity is exemplified through smooth lines in the line charts, which aid in perceiving the flow and trends of data over time. Additionally, the use of closed shapes in the treemap, following the principle of Closure, organizes and groups data based on gender, enabling a clear understanding of the distribution.

Moreover, the dashboard applies the principle of Figure-Ground by utilizing contrasting colors to emphasize important data, allowing key insights to stand out from the background. Lastly, the principle of Common Fate is demonstrated in the bar chart, where related data points are grouped together, facilitating effortless comparison and understanding of different vulnerable ethnic groups.

By implementing these Gestalt principles, the dashboard provides a cohesive and organized visual display, enhancing data comprehension for "The Public." The thoughtful application of Proximity, Similarity, Continuity, Closure, Figure-Ground, and Common Fate ensures that viewers can easily perceive patterns, trends, and valuable insights, making data-driven decisions more accessible to a broader audience.



PART 3:

Unraveling COVID-19: Visual Insights and Trends:

Initially, I conducted a thorough analysis of the dataset, identifying the presence of missing and null values. To ensure a focused analysis, I decided to filter the data for the years 2020 to 2023 and removed any records with missing values. With "The public" as my target audience, I opted to present general trends using straightforward and easily comprehensible visuals. Throughout my journey, I maintained consistency by employing a teal-blue color palette, known for its subtlety and visual harmony. I began by creating a line chart to visualize the progression of COVID cases over the years. The chart depicted a declining trend as COVID cases decreased by 2023. The peak occurred in April 2020, with 133,380 patients, while the lowest point was in May 2023, with only 45 patients. This trend illustrates how COVID initially had a devastating impact worldwide, but as vaccinations and immunity spread within the population, its effects significantly diminished.

Subsequently, I employed a visually captivating treemap for a formal representation of the gender division among patients. The treemap was visually engaging and effectively showcased the distribution of genders within the patient population. Next, I crafted a bar chart to illustrate the distribution of patients across different age groups. The bar chart utilized a continuous shading technique in teal blue, where darker shades represented higher patient numbers, gradually transitioning to lighter shades for lower patient counts within each age group. This approach provided a clear visual representation of the patient distribution across various age categories.

Afterward, I aimed to acquaint my audience with the impact on different ethnicities. Initially, I was uncertain about which chart to use and opted for a visually appealing bar chart. However, I later realized the importance of prioritizing informative and communicative visuals over aesthetics. Consequently, I switched to a highlighted table, utilizing continuous shading to emphasize specific details. This approach allowed for better comprehension and comparison of the data related to various ethnicities affected by the pandemic.

Amidst the COVID pandemic, distinguishing certain cases became challenging as they did not exhibit any symptoms. Particularly, asymptomatic cases were considered more perilous as the virus could silently spread and multiply within the body without the individual's awareness, leading to diagnosis at a severe stage. To address this, I designed a straightforward table that allowed the audience to easily differentiate between the number of symptomatic and asymptomatic patients. Creating this table required leveraging my knowledge of creating a calculated field. Since the dataset lacked direct information about symptoms, I used the "Onset Dt" field, which contained the dates when symptoms began to appear, to create a calculated field labeled "Symptoms Y/N." This field was populated with "Yes" if an onset date was present, indicating symptomatic cases, and "No" otherwise, representing asymptomatic cases.

One of my personal favorite charts was the depiction of death and hospitalization rates. I aimed to draw attention to how unprepared the world was for the pandemic. To achieve this, I filtered out patients who had unfortunately died due to COVID and those who did not receive hospital care. The chart showcased the immense strain on hospital facilities, as they struggled to accommodate the high number of patients in the initial stages, particularly in April 2020, when a significant number of people passed away without medical attention. However, the chart also demonstrated a positive trend, with a noticeable improvement in hospital facilities by 2023, leading to a flattening of the graph.

I meticulously organized all these charts in a systematic manner within my dashboard, aiming to help people understand the trends and analyze the impact of the most terrifying pandemic of our lives. This journey of creating the dashboard has been a valuable learning experience for me. The most crucial lesson I've gained is that data visualization is not solely about creating charts using a specific tool, but rather, it begins with analyzing the situation, understanding the audience, and thoroughly examining the data. Crafting an impactful narrative that encompasses valuable insights and effectively communicating that story to the audience is paramount. I am immensely grateful to my professor for imparting this wonderful methodology for data visualization, which has not only enhanced my skills but also transformed my mindset when creating impactful data visualizations. This course has been instrumental in refining my abilities and has been a highly appreciated and enlightening journey, thanks to the guidance of my professor.