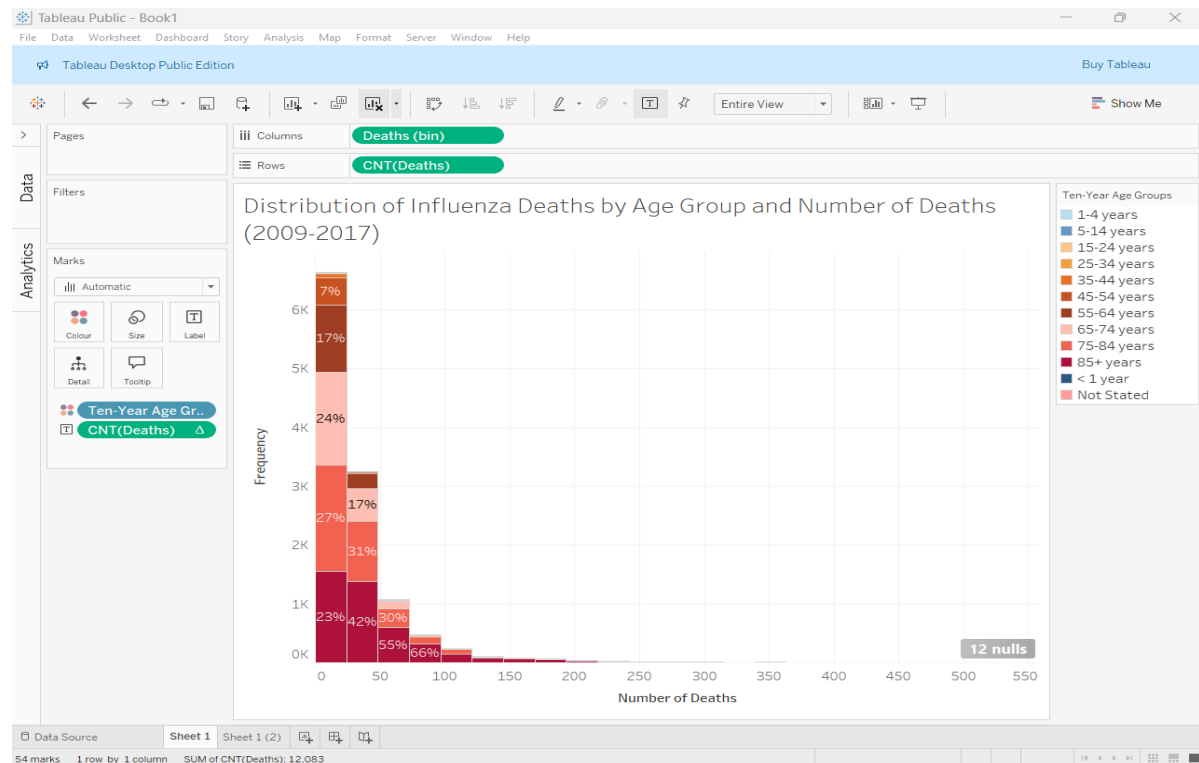


Statistical Visualizations: Histograms & Box Plots.

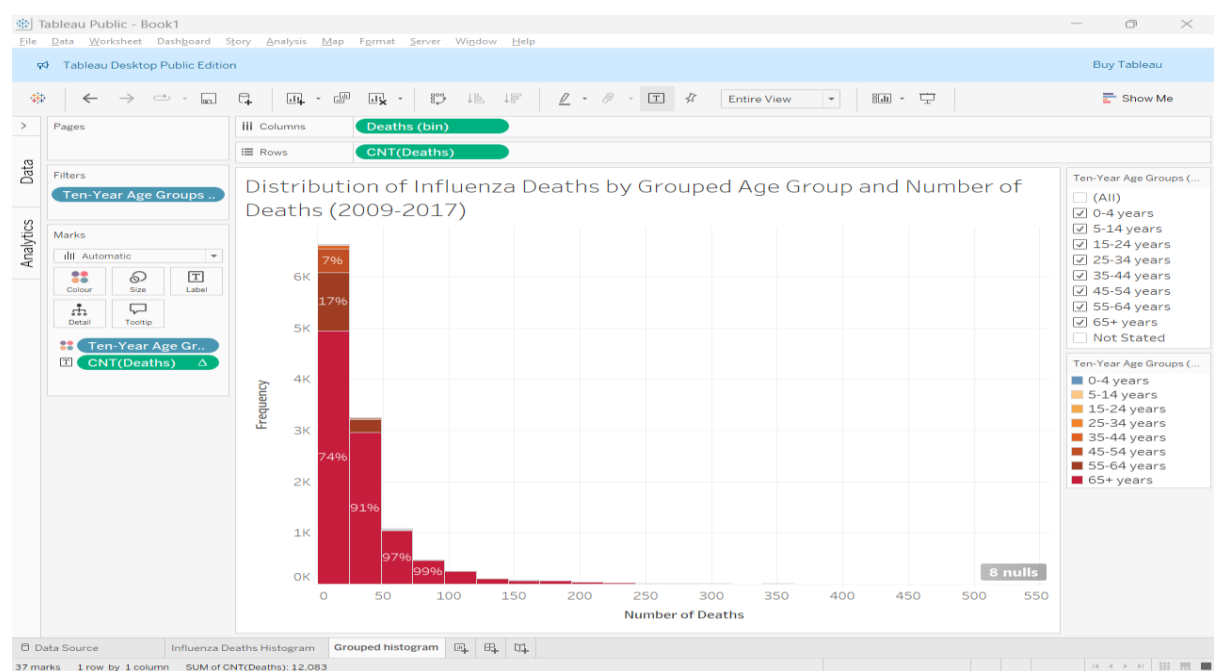
Paul Maden.

Last Update: 12/10/2024

Q1. Create a histogram of influenza deaths by age in Tableau:



When the 'vulnerable age groups', like 65+ years and 0-4 years, are combined to simplify the distribution illustration, this is the histogram created:



This histogram illustrates:

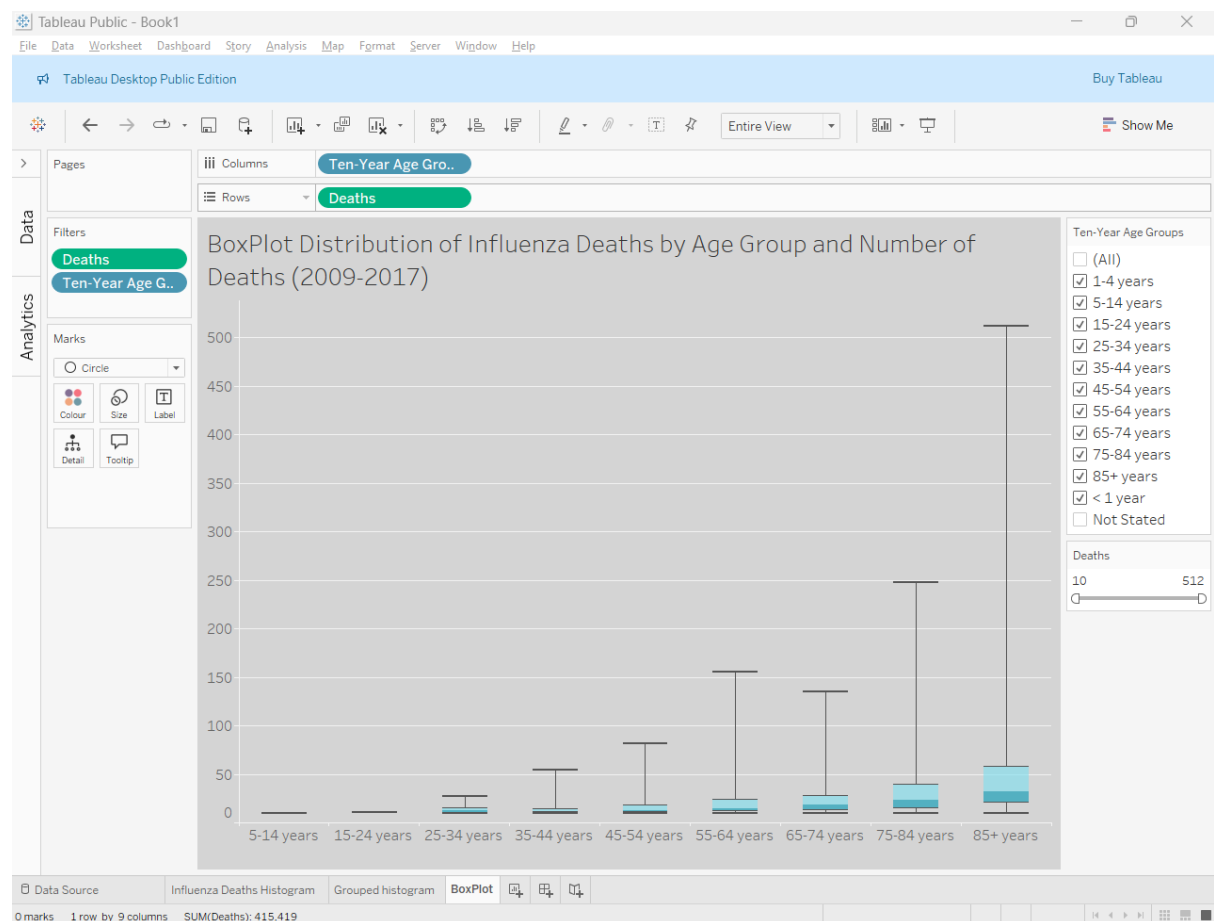
- **Vulnerability of young and old populations:**
 - The colour legend on the right shows different age groups, with the oldest groups bracket (65 plus years) coloured in red dominates the higher parts of the bar, indicating they account for many deaths. This suggests that older populations (65 plus years) are indeed more vulnerable, as they contribute significantly to the higher death counts
- **Age groups with no deaths:**
 - All age groups from the legend are represented in the histogram meaning there are no age groups with 0 deaths.

What is also noticeable is that for middle age groups (e.g. 15-24, 25-34 years), the number of deaths is significantly lower, especially in the higher bins (towards the right of the chart). This indicates that middle age populations are less vulnerable compared to the youngest and oldest groups.

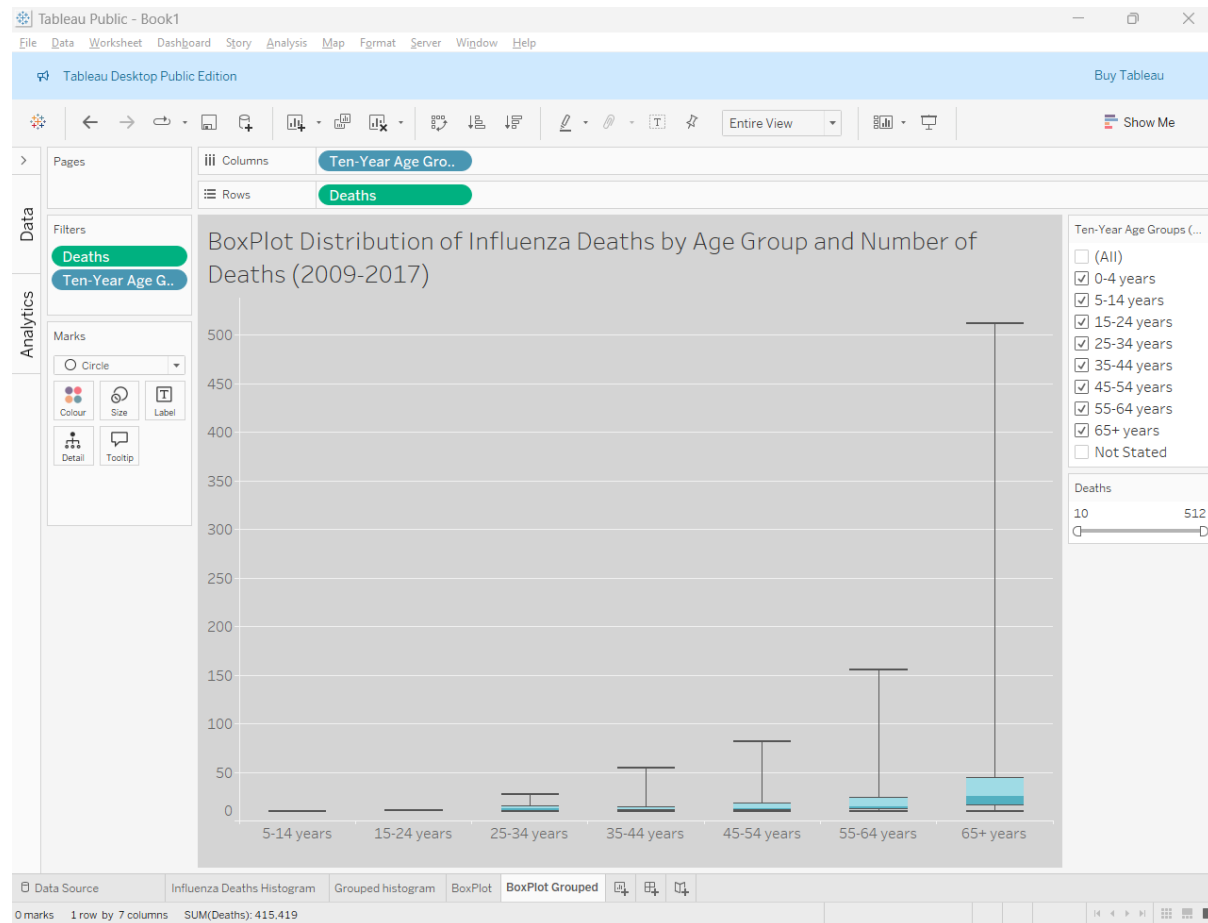
In summary, the vulnerable (like 65+ years) population are more at risk with higher death counts. No age group has 0 deaths, but the middle age groups experienced significantly fewer deaths.

Q2. Create a box and whisker plot of this same information in Tableau:

Box and whisker plot, each age category should be its own plot:



When the ‘vulnerable age groups’, like 65+ years and 0-4 years, are combined to simplify the distribution illustration, this is the boxplot created:



Q3. Update the visualization using the style guide checklist you created in Exercise 2.2:

Using style guide – Exercise 2.2 – in this updated boxplot, we can observe a few changes:

- **Background and grid lines:** the background is now shaded, with grid lines becoming more prominent. This helps in visually distinguishing the range of deaths for each age group.
- **Boxplot markers:** there are now clearer boxplot markers (whiskers, boxes, and lines) for each age group, with the range of death counts better highlighted. The data points for each age group are more distinguishable.

Q4. Explain what the box plot tells you that the histogram can't.

- **Range and Distribution:**
 - The box plot provides more detailed information about the spread of data (range) for each age group. It shows the minimum, maximum, median, and quartiles, which the histogram does not.
 - I.e. the 65+ year group has a very wide range of deaths, from close to zero up to more than 500 deaths. The histogram only shows frequencies of deaths in bins, so it doesn't reveal this variability.

- **Outliers and Variation:**
 - The box plot allows us to see if there are any outliers (extreme values), although none are evident here. It also visually shows the variation within each age group's death counts through the whiskers and box height.
- **Comparing Medians:**
 - The median death count for each age group is indicated by the horizontal line inside each box. You can compare the medians for different age groups, which is not possible with a histogram.
 - E.g. the median deaths for the 65+ years group are clearly much higher than any other group, while groups like 5-14 years have a much smaller range and median.

Summary:

The box plot provides a more nuanced view of the distribution of death counts across age groups, highlighting the range and central tendency (median) of deaths, while the histogram is better for showing frequency distributions across the death count bins. Both are useful, but the box plot gives a clearer picture of the spread and distribution within each age group.

Tableau Link:

https://public.tableau.com/views/UpdatedInfluenzaDeathsDistribution/Groupedhistogram?:language=en-GB&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link