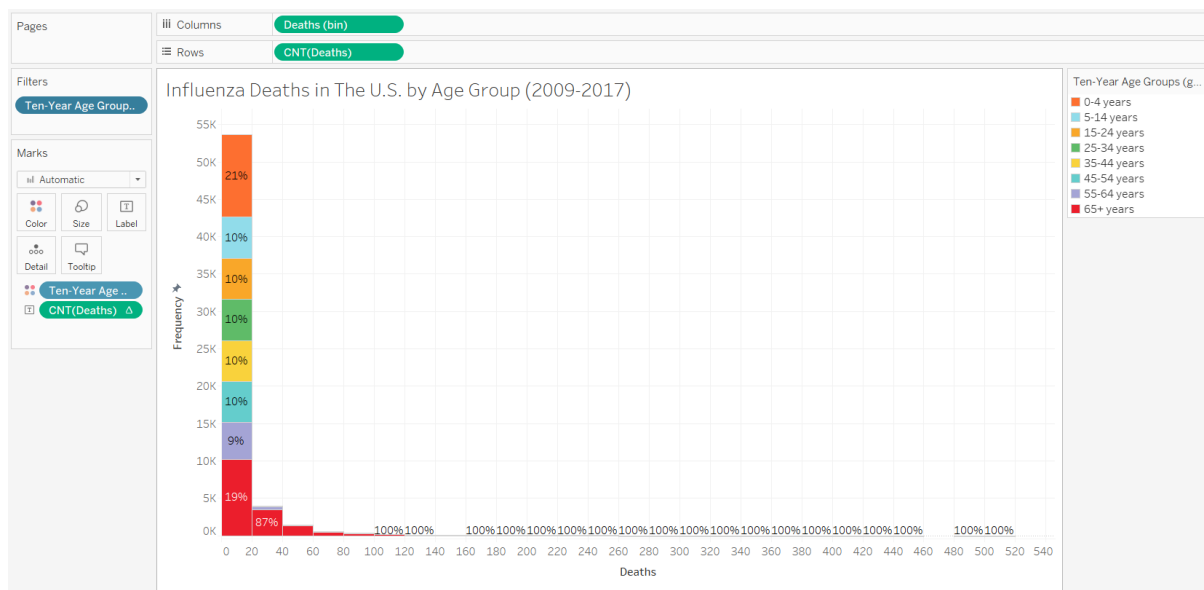


## EXERCISE 2.5: STATISTICAL VISUALIZATIONS: HISTOGRAMS & BOX PLOTS

BY ANUSHMA SHARMA

1. Create a histogram of influenza deaths by age. The histogram should be a frequency of deaths. Examine age by adding age categories as colours.

[https://public.tableau.com/app/profile/anushma.sharma/viz/InfluenzaDeathsbyAge-Histogram\\_17185926486600/Histogram\\_InfluenzaDeathsbyAge?publish=yes](https://public.tableau.com/app/profile/anushma.sharma/viz/InfluenzaDeathsbyAge-Histogram_17185926486600/Histogram_InfluenzaDeathsbyAge?publish=yes)



- **Are young and old populations more vulnerable (because they have more deaths)?**
  - Young and old populations are more vulnerable, as shown by the histogram. The <5 and 65+ age groups have the highest death counts. The 65+ group especially dominates with a larger death count as we move right on the chart.
- **Are there any age groups that have no deaths?**
  - It's impossible to determine from the histogram alone if any age groups have zero deaths, as the leftmost bin covers deaths from 0 to 19. We need to reference the raw data for exact values.

### Style Guide Checklist (Histogram):

#### Text:

- **Title and Text:** Clear and descriptive. The title conveys the purpose effectively.
- **Labels:** Axis labels are present and clear. No additional annotations.
- **Redundancy:** No redundant text.
- **Legends:** Legends present for colours.

## Colour:

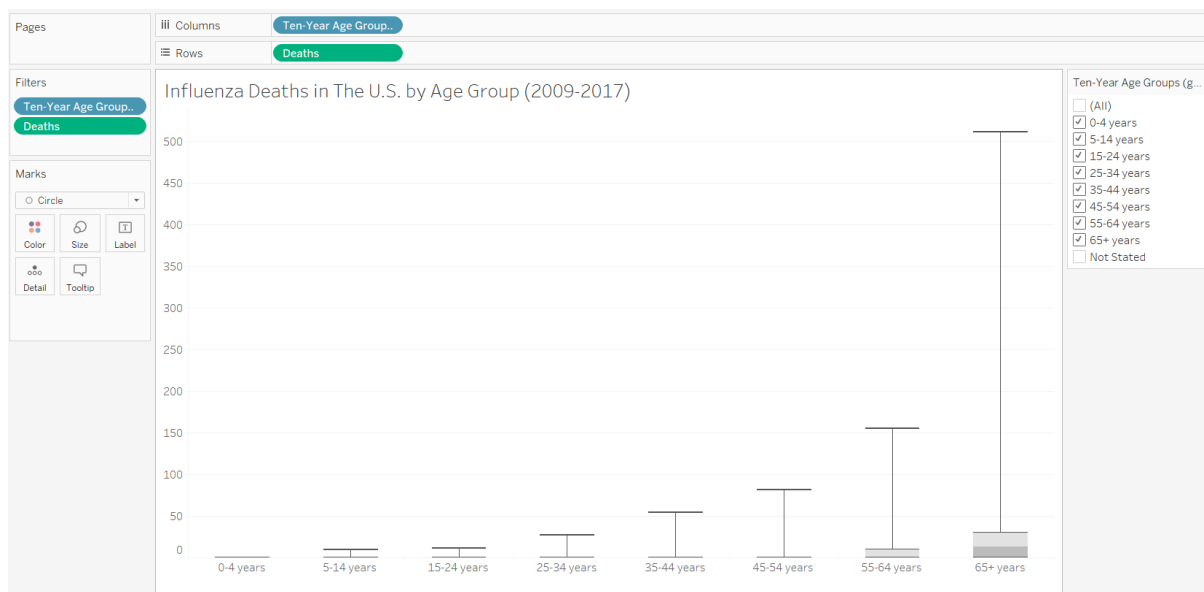
- **Significance:** Differentiates age groups.
- **Number of Colors:** Use more than five colours, as there are more than 5 age groups.
- **Logic:** Colour scheme logic is clear. The darkest colour highlights key data.

## Other

- **Sizes:** The bar sizes vary according to the frequency of deaths.
- **Groupings:** Higher bars indicate greater frequency. Darker colours highlight more vulnerable populations. Size reflects frequency count.
- **Whitespace:** Adequate whitespace, making it readable.
- **Accessibility:** Generally accessible and uses colourblind-friendly palettes.
- **Educational Value:** The histogram illustrates the frequency of influenza deaths across different age groups.

2. Create a box and whisker plot of this same information. Each age category should be its own plot.

[https://public.tableau.com/app/profile/anushma.sharma/viz/InfluenzaDeathsbyAge-BoxPlot/BoxPlot InfluenzaDeathsbyAge?publish=yes](https://public.tableau.com/app/profile/anushma.sharma/viz/InfluenzaDeathsbyAge-BoxPlot/BoxPlot%20InfluenzaDeathsbyAge?publish=yes)



The box plot provides detailed statistical distribution for each age category, revealing median, quartiles, and outliers.

## Style Guide Checklist (Box Plot):

### Text

- **Title and Text:** The title clearly indicates that the visualization shows influenza deaths by age group.
- **Labels:** Only the x-axis has group labels; no additional text labels are provided.
- **Redundancy:** No redundant information is present.
- **Legends:** There is no legend, as a box plot typically does not use multiple colours.

### Colour

- **Colour Scheme:** There is no colour scheme used.
- **Number of Colours:** Only one colour is used.
- **Colour Scheme Logic:** No colour scheme is applied.
- **Highlighting Important Information:** Since there is no colour scheme, this is not applicable.

### Other

- **Sizes:** The size of the box plot reflects the presence of outliers.
- **Groupings:** No groupings are portrayed through colour, size, or position.
- **Whitespace:** Adequate whitespace is present.
- **Accessibility:** The visualization is colour-blind friendly, using only one colour, with large enough labels.
- **Educational Value:** The line chart illustrates the number of influenza deaths across age groups and provides descriptive statistics.

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

- The box plot shows the median, quartiles, interquartile range (IQR), and outliers, highlighting central tendency, variability, and unusual values within each age group. It also indicates data skewness. Unlike histograms, box plots provide detailed summary statistics and identify outliers.