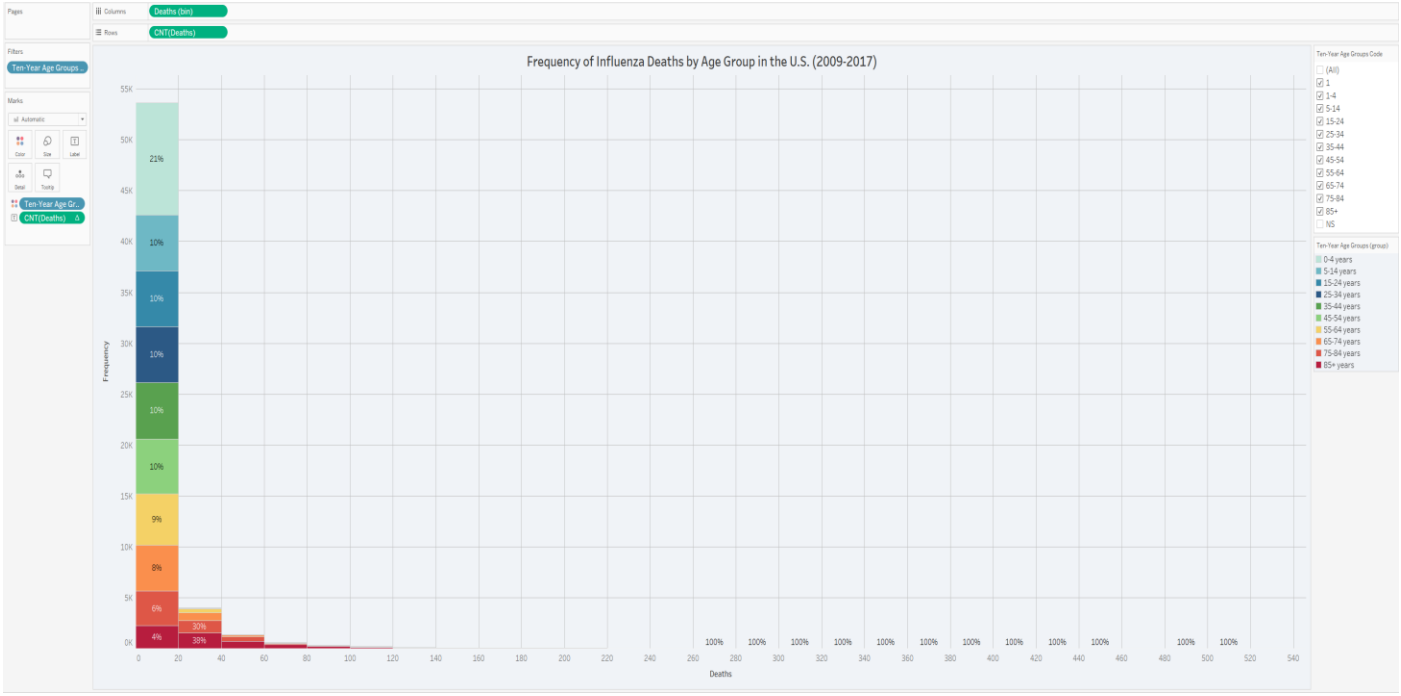
Kyle Stanford
Data Analytics Immersion – Exercise 2.5

Statistical Visualisations: Histograms and Box Plots

https://public.tableau.com/views/DataImmersionAchievement2Ex2_5StatisticalVisualisationsHistogramsandBoxPlots/Story1?:language=en-US&publish=yes&:sid=&:display_count=n&:origin=viz_share_link

Histogram:



Are young and old populations more vulnerable (because they have more deaths)?

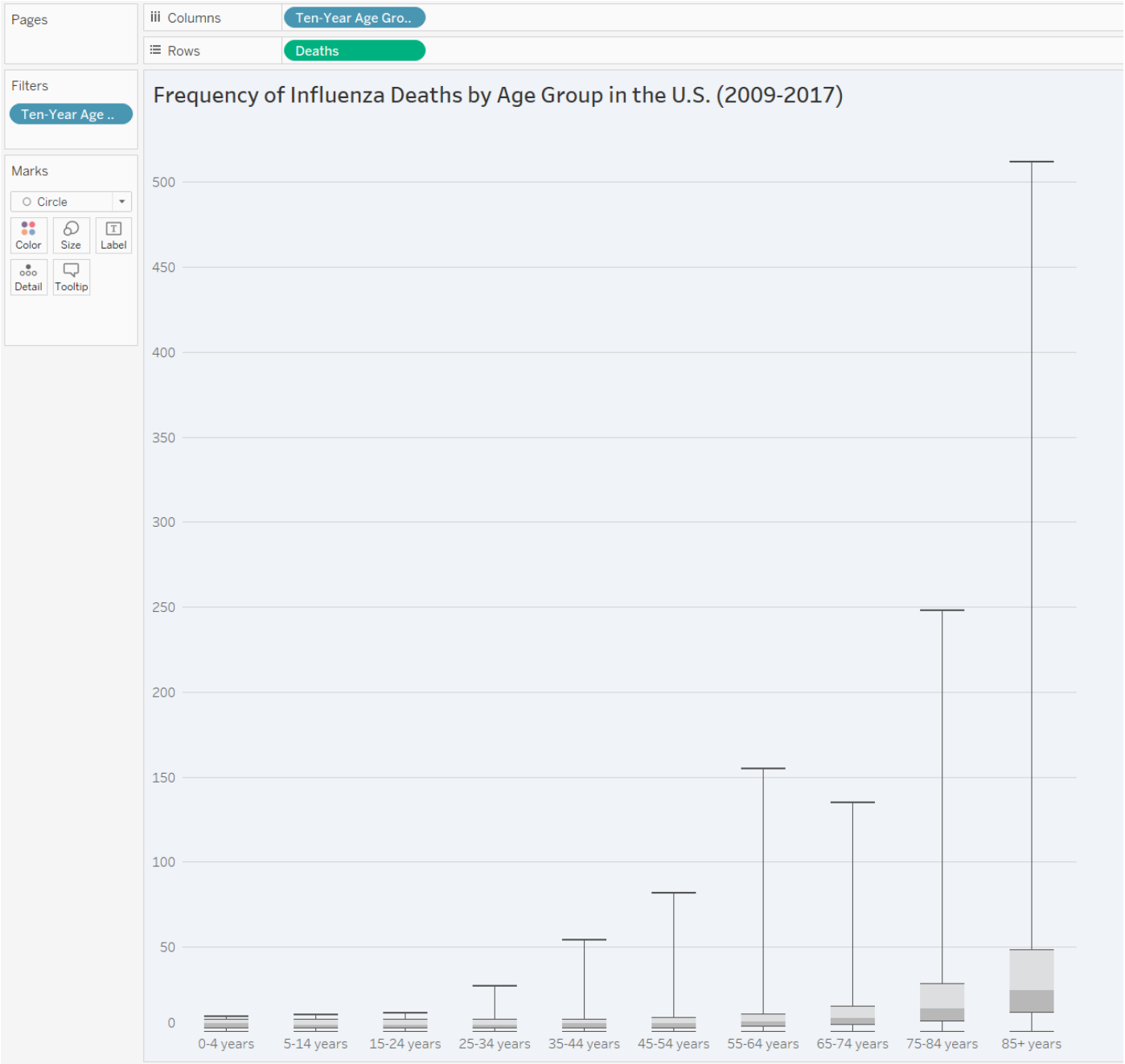
Although the scaling of this histogram makes it difficult to see, older age groupings are far more vulnerable than younger age groups. This is because all bins after the left-most bin are dominated by the 65 years and older groups. This indicates that the influenza-mortality rate of older populations is far higher than that of younger ones.

Are there any age groups that have no deaths?

This is impossible to determine because the histogram uses bins for the number of deaths. Whilst the left-most bin encompasses all death values of 0 to 19, there is no way of knowing if the age groups represented here only have data points of exactly 0 deaths without referencing the raw data.

Note: the raw data uses the code “suppressed” for death values ranging from 0 to 9. This makes it impossible to determine if any age groups have 0 deaths from the raw data as well. However, to ensure Tableau would be able to make a histogram, these cells in the raw data were replaced with a random number from 0 to 9 (using =RANDBETWEEN(0,9)).

Box Plot:



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

Unlike the histogram, the box plot shows median and quartile values for each of the age groups.

Kyle Stanford
Data Analytics Immersion – Exercise 2.5

Visualisation Checklist

(H = Histogram, BP = Box Plot)

Text

- Are the title and text descriptive enough? (i.e., do you understand what the visualization is trying to convey just by looking at the title and text?)
    - H and BP - Yes
- Are there text labels?
    - H – Yes, each age-section shows the percentage that age group occupies of the overall bin (where possible).
    - BP – Only in the x-axis to identify each age groups' box plot
- Does the text portray any redundant information that could be gotten rid of?
    - H and BP – No
- Do colours, shapes, and size scales come with legends?
    - H – Yes, there is a colour legend to the right indicating which age groups are represented with each colour.
    - BP – No since a box plot does not require colours.

Colour

- What does the colour scheme signify?
    - H – The colour scheme signifies the age groupings
    - BP – N/A
- Are there more than five colours?
    - H – Yes, but there are more than 5 groupings for the age categories. We could combine age groupings to get 5 or less, but we also may want to keep the current level of detail.
    - BP – N/A
- Does the colour scheme make sense? Are colours analogous, complementary, monochromatic, or intuitive?
    - H – The colour scheme seems appropriate for the number of groups. Having analogous colours work through the colour spectrum as the age groups increase feels easy to interpret. Also, the hot colours being linked to the older age groups helps intuitively identify them as the more vulnerable population (red = alarm).
    - BP – N/A

- If colour is used to draw attention to important information, is the darkest colour representing the most important information?
    - H – not exactly, although the hottest colours are drawing attention to the most vulnerable population groups.
    - BP – N/A

Other

- Are different sizes used? If so, is there meaning behind the sizes?
    - H – The columns for each bin are different sizes depending on the frequency of the number of deaths.
    - BP – The length of the box and whisker portions of the plot vary depending on the distribution of the influenza deaths in each age group. The sizes indicate how spread or concentrated the data is.
- Are there groupings in the data that can be portrayed through colour, size, or position?
    - H – They are. Age groupings are portrayed through colour.
    - BP – All age groupings are separated and have their own box plot which is located next to the box plots of other age groups.
- Is there (enough) whitespace?
    - H and BP - Yes
- Is the visualization accessible?
    - H – Probably not given that there are many colours, including red and green.
    - BP - Yes
- Does the visualization teach you something?
    - H – Yes, the histogram shows that older age groups have a greater frequency of influenza related deaths.
    - BP – Yes, the box plot shows that older age groups have a greater median frequency of influenza related deaths. There is also a general trend in which the maximum number, median, and quartiles increase exponentially for each older group.

On the next page, a new histogram has been made to address the issues with colour and number of groups. All other elements have been kept the same. The colour scheme was changed to a monochromatic red palette where the older groupings were assigned the darkest shade:

Kyle Stanford
Data Analytics Immersion – Exercise 2.5