**Exercise 2.5**

**Histogram**

https://public.tableau.com/app/profile/aysha.samsudeen/viz/DataImmersionAchievement2Ex2\_5StatisticalVisualisationsHistogramsandBoxPlots\_17291462669180/Histogram-InfluenzaDeathsbyagegroup

**A screenshot of a graph

Description automatically generated**

**The histogram above highlights key issues,**

* **Vulnerable Age Groups**: The young (under 5) and elderly (65+) populations show the highest mortality rates. The older age groups would be more vulnerable than younger age groups. We can visually view this as the left bins on the graph are predominatly occupied by 65 years and older age groups. This showcases once again that influenza rate is higher in older population than younger one. We cannot measure if there were zero deaths as the histogram uses bins to measure number of deaths with left most bins contains values from 0-19 but we cannot determine what the age groups contain without evaluating the raw data first.
* **Death Counts**: The under-5 age group and the 65+ age group have the highest percentages and counts of deaths.
* **Trend**: As age increases, particularly from 65+, the death count becomes more pronounced.
* **Overall Findings**: No age group recorded zero deaths; all groups experienced mortality, highlighting a widespread public health issue.

**Histogram (visual checklist factors)**

* Title: Clearly conveys frequency of influenza deaths by age group.
* Text Labels: Percentages displayed for each age category.
* Redundancy: No redundant information present.
* Color Scheme:
  + Legend provided for age groups.
  + Each age group has a distinct color.
  + More than 5 colors used, as there are multiple age groups.
  + Darker colors highlight vulnerable groups (<5 and 65+).
  + Color differentiation is clear and effective.
* Bar Sizes: Vary based on frequency count; height indicates greater frequency.
* Whitespace: Ample whitespace is present.
* Accessibility: Color-blind friendly; colors are easily distinguishable and labels are large.
* Informational Value: Effectively shows frequency of influenza deaths across age groups.
* Labels: Only necessary labels used (axes, title, legend, percentages).
* Title Size: Title is larger than other text elements.
* Percentage Placement: Percentages are clearly located within corresponding bar portions.
* Labeling: No excessive labeling.

**Box plot**

A box plot is a great way to analyze data and make sense of the information listed. However, there are several differences between a box plot and histogram.

A screenshot of a computer screen

Description automatically generated

**Box Plot vs. Histogram Insights**

1. **Summary Statistics**:
   * **Box Plot**: Displays the median, quartiles, and potential outliers. This helps identify the central tendency and variability within each age group.
   * **Histogram**: Shows the frequency distribution but does not convey specific summary statistics like median or quartiles.
2. **Distribution Shape**:
   * **Box Plot**: Provides a clear visual of the spread and symmetry of the data (e.g., whether the data is skewed or has outliers).
   * **Histogram**: While it shows the shape of the distribution, it may be less clear about exact central values and spread.
3. **Comparison Across Groups**:
   * **Box Plot**: Facilitates easy comparison of medians and spread between age groups, making it straightforward to identify trends and differences.
   * **Histogram**: Effective for understanding overall distributions but less effective for direct comparison between multiple groups.

**Comaprisons of Visualization Checklist Feedback**

* **Titles and Text**: Both visualizations are clear and descriptive.
* **Text Labels**: The histogram’s labels enhance clarity, while the box plot could benefit from more descriptive annotations.
* **Color Use**: The histogram’s color scheme is well-thought-out and intuitive, aiding interpretation.
* **Sizes**: In both visualizations, size variations effectively represent frequency and spread, respectively.
* **Whitespace**: Both visualizations maintain good spacing, aiding readability.
* **Accessibility**: The histogram may pose challenges for colorblind viewers, while the box plot remains accessible.
* **Educational Value**: Both visualizations effectively convey important information about influenza-related deaths across age groups.

Overall, both visualizations have their strengths, and together they provide a comprehensive understanding of the data. The box plot is particularly useful for summarizing and comparing distributions, while the histogram excels at illustrating the frequency of occurrences within specific ranges.

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 blue palette where the older legends were assigned the darkest shade possible.

A screenshot of a graph

Description automatically generated