**Marathon Results - 2017**

**Introduction:**

In this report, our focus lies on a comprehensive analysis and visualization of Boston marathon finishers data from the 2017 event. Through meticulous exploration, we aim to unearth the meaningful trends, demographic patterns, and performance metrics embedded within the dataset. Our investigation will shed light on crucial aspects such as the distribution of finishers ages, offering insights into the age demographics prevalent in marathon participation. Additionally, we will delve into participation rates across countries, unraveling the global appeal and geographic diversity of marathon events. Furthermore, we will scrutinize the age distribution by gender, examining potential variations and disparities in the participation and performance based on gender identity.

**Visualizations:**

Initially, the plan was to create five key visualizations:

1. A bar plot showcasing the top 5 countries with the fastest runners, based on average finishing times.
2. A lollipop chart depicting the top 10 countries with the highest number of participants who completed the marathon.
3. A histogram illustrating the distribution of finishers ages.
4. A boxplot displaying the distribution of finishers ages categorized by gender.
5. A pie chart illustrates the distribution of finishers by gender, showcasing the relative proportions of male and female participants in the marathon.
6. **Distribution of Finishers by Gender:**

A blue and pink pie chart

Description automatically generated

The above pie chart offers a visual breakdown of the gender distribution among marathon finishers. The size of each segment corresponds to the proportion of finishers identified as male or female. The use of blue and pink hues distinguishes between genders, facilitating easy interpretation. Through this visualization, one can readily observe the relative representation of males and females in the marathon, providing insights into the gender dynamics of participation in endurance sports.

1. **Top 10 Countries with Maximum Number of Finishers:**

A graph with numbers and dots

Description automatically generated

The lollipop chart above visualizes the number of finishers by country in the marathon event. Each lollipop represents a country, with its length corresponding to the number of finishers from that country. The chart showcases the top 10 countries with the highest number of finishers, providing a clear comparison of participation levels across different nations. By arranging the countries in descending order based on the starting alphabet of the country name, the visualization effectively highlights the countries with the most significant representation in the marathon and we can easily look through each country based on the name. We can notice that USA has the highest number of finishers compared to all other nations. This intuitive representation allows viewers to easily identify the required countries in terms of participation, offering valuable insights into the global reach and popularity of the marathon event.

1. **Distribution of Finishers Ages in the Marathon:**

A graph of a number of years

Description automatically generated with medium confidence

The histogram presented above offers a comprehensive visualization of the distribution of finishers’ ages participating in the marathon event. Each bar in the histogram represents a range of ages, with the height of the bar indicating the frequency or number of finishers falling within that age range. By employing a bin width of 5 years, the histogram provides a granular view of the age distribution, allowing viewers to discern patterns and trends more effectively. The bars are filled in blue with black outlines, enhancing contrast and visibility, while an alpha value of 0.7 adds a subtle transparency effect to the bars, reducing visual clutter. Overall, this histogram serves as a valuable tool for understanding the age demographics of participants in the marathon, facilitating insights into the age composition and diversity of the finishers.

1. **Age Distribution based on Gender in the Marathon:**

A diagram of a gender

Description automatically generated with medium confidence

The boxplot depicted above offers a comprehensive visualization of the distribution of finishers’ ages stratified by gender in the marathon event. Each boxplot represents the age distribution for a specific gender category, with the median age marked by a horizontal line inside the box. The boxes themselves indicate the interquartile range (IQR), encompassing the middle 50% of the data, while the whiskers extend to the minimum and maximum ages within 1.5 times the IQR from the first and third quartiles, respectively. By employing a fill color corresponding to gender (blue for males and pink for females), the plot effectively distinguishes between the two groups. An alpha value of 0.3 adds a subtle transparency effect to the boxplots, aiding in visual interpretation without overwhelming the viewer. This boxplot provides valuable insights into the age distributions for male and female finishers, facilitating comparisons and uncovering potential differences or similarities in age demographics based on gender identity.

1. **Top 5 Countries with Fastest Runners:**

A graph of blue rectangular bars

Description automatically generated with medium confidence

The bar plot showcased above illustrates the top 5 countries with the fastest average finishing time of the finishers from the marathon event. Countries are sorted in ascending order based on their average finishing times, with the top 5 countries with the shortest average times selected for visualization. Each bar in the plot represents a country, with the length of the bar corresponding to its average finishing time. The x-axis displays the countries, while the y-axis represents the average finishing time. The bars are filled in a steel blue color for visual appeal and ease of interpretation. Additionally, the rotation of country labels on the x-axis enhances readability. Overall, this visualization offers valuable insights into the performance of different countries runners in the marathon, highlighting those with the quickest average finishing times.

1. **Age vs. Official Finishing Time**

A graph showing a number of blue and pink dots

Description automatically generated

This above shown scatter plot enables us to explore the relationship between the official finishing times and the ages of the marathon participants. This can provide insights into whether certain age groups tend to finish faster or slower on average. Also, it will allow us to visualize how the finishing times vary with age and whether there are noticeable differences between genders in this aspect.

**Application of Data Visualization Principles:**

Several principles of data visualization and design were applied throughout the assignment:

1. **Clarity and Simplicity:** Each plot is designed and displayed in a way to be easily interpretable, with clear labels, appropriate titles, and minimal clutter, to enhance the readability.
2. **Color and Contrast:** Color choices are deliberate, with contrasting hues usage to have a clear differentiation between the categories and these highlight key insights.
3. **Consistency:** A consistent theme and style are maintained across all visualizations, ensuring coherence and facilitating comparison.
4. **Engagement:** The visualizations are engaging and visually appealing, encouraging viewers to explore and derive insights from the data.

**Conclusion:**

In conclusion, this analysis of Boston marathon data from 2017 has provided valuable insights into the demographics and performance trends of participants. By exploring factors such as age distribution, gender disparities, and the top-performing countries, we have gained a deeper understanding of the dynamics within the marathon community. These insights not only offer valuable guidance for marathon organizers, trainers, and participants in optimizing future events but also contribute to the broader conversation surrounding endurance sports and athletic achievement. Moving forward, further exploration of certain performance factors and regional variations could enrich our understanding and help drive continued improvements in marathon experiences for all involved.