Bar Plots, Histograms, and Distributions: Takeaways

Syntax

- Plotting a vertical bar plot:

```
plt.bar(x=x_coordinates, height=heights)
plt.show()
```

- Plotting a horizontal bar plot:

```
plt.barh(y=y_coordinates, width=widths)
plt.show()
```

- Changing the x-tick labels of a graph (any graph):

```
plt.xticks(ticks=x_coordinates, labels=the_labels_you_want)
```

- Changing the y-tick labels of a graph (any graph):

```
plt.yticks(ticks=y_coordinates, labels=the_labels_you_want)
```

- Plotting a histogram:

```
plt.hist(column)
plt.show()
```

- Generating a frequency table:

```
Series.value_counts(bins=10)
```

- Generating a sorted grouped frequency table:

```
Series.value_counts(bins=10).sort_index()
```

Concepts

- We call the number of times that a unique value occurs the **frequency**. And we call the output of **Series.value_counts()** a **frequency table**.

- To create a readable frequency table for a numerical column, we group the values in intervals — in this case, we call the table **grouped frequency table**.

- **Bar plots** work well for visualizing frequency tables when the frequency tables are generated for categorical columns.

- **Histograms** work well for visualizing frequency tables when the frequency tables are generated for numerical columns.

- A histogram is a modified bar plot — the main visual difference is that there are no gaps between bars. Another equally important difference is that each bar represents an interval, not a single value.

- A histogram shows the distribution of the values, and if its shape is symmetrical, then we say we have a **symmetrical distribution**.

- If we draw a vertical line exactly in the middle of a symmetrical histogram, then we divide the histogram in two halves that are mirror images of one another.

- A common symmetrical distribution is the **normal distribution** — most values pile up in the middle of the range, and value frequencies decrease gradually toward the extremities of the range.

- Another common symmetrical distribution is the **uniform distribution**. The values are uniformly distributed — the unique values have equal frequencies.

- When we plot histograms in practice, we rarely see perfectly symmetrical distributions. However, these ideal cases we learned about serve as a baseline to help us describe and interpret the distributions we see in practice.

- **Skewed distributions** are not symmetrical, and they have the following characteristics:
    - The values pile up toward the end or the starting point of the range, making up the body of the distribution.
    - Then the values decrease in frequency toward the opposite end, forming the tail of the distribution.

- If the tail of a skewed distribution points to the right, then the distribution is right skewed (or positively skewed).

- If the tail of a skewed distribution points to the left, then the distribution is left skewed (or negatively skewed).

Resources

- [A short article on bar plots by The Data Visualization Catalogue](#)
- [A short article on histograms by The Data Visualization Catalogue](#)
- A few fun and useful tutorials from MathIsFun:
    - [Bar Plots](#)
    - [Histograms](#)
    - [Frequency Distribution](#)
    - [Grouped Frequency Distribution](#)
    - [Normal Distribution](#)