Visualizing Frequency Distributions: Takeaways

Syntax

- Vertical bar chart:

```
ggplot(data = df,
       aes(x = col)) +
  geom_bar()
```

- Horizontal bar chart

```
ggplot(data = df,
       aes(x = col)) +
  geom_bar() +
  coord_flip()
```

- Proportions with bar charts

```
ggplot(data = df,
       aes(x = col,
             y = ..prop..,
             group = 1)) +
  geom_bar()
```

- Percentages with bar charts

```
ggplot(data = df,
       aes(x = col,
             y = ..prop.. * 100,
             group = 1)) +
  geom_bar()
```

- Horizontal stacked bar chart from summary table of proportions:

```
proportions <- df %>%
  group_by(col) %>%
  summarize(Prop = n() / nrow(df))
ggplot(data = proportions,
              aes(x = "", y = Prop, fill = col)) +
  geom_bar(stat = "identity", width = 0.25) +
  coord_flip()
```

- Simple pie chart for exploratory data analysis:

```
ggplot(data = proportions,

              aes(x = "", y = Prop, fill = col)) +

  geom_bar(stat = "identity") +

  coord_polar(theta = "y")
```

- Histogram - default number of bins (30):

```
ggplot(data = df,

       aes(x = col)) +

  geom_histogram()
```

- 10 bins

```
ggplot(data = df,

       aes(x = col)) +

  geom_histogram(bins = 10)
```

- Aligning 10 bins with range of values present for a single variable

```
custom_binwidth <- (max(df$col) - min(wdf$col)) / 10

ggplot(data = df,

       aes(x = col)) +

  geom_histogram(boundary = min(df$col),

                 binwidth = custom_binwidth)
```

Concepts

- To visualize frequency distributions for *nominal* and *ordinal* variables, we can use:
    - **Bar chart**.
    - **Pie charts**.

- To visualize frequency distributions for variables measured on an interval or ratio scale, we can use a **histogram**.
- Depending on the shape of the histogram, we can have:
    - **Skewed** distributions:
        - Left skewed (negatively skewed) — the tail of the histogram points to the left.
        - Right skewed (positively skewed) — the tail of the histogram points to the right.
    - **Symmetrical** distributions:
        - **Normal** distributions — the values pile up in the middle and gradually decrease in frequency toward both ends of the histogram.
        - **Uniform** distributions — the values are distributed uniformly across the entire range of the distribution.

Resources

- An introduction to bar chart.
- An introduction to pie charts.
- An introduction to histograms.
- An introduction to skewed distributions.
- More details on the normal distribution.