Comparing Frequency Distributions: Takeaways

Syntax

- Generating a grouped bar plot:

```
import seaborn as sns
sns.countplot(x = 'column_name_1', hue = 'column_name_2',
              data = some_dataframe)
```

- Generating only the shape of the histogram for two `Series` objects:

```
Series_1.plot.hist(histtype = 'step')
Series_2.plot.hist(histtype = 'step')
```

- Generating kernel density plots for two `Series` objects:

```
Series_1.plot.kde()
Series_2.plot.kde()
```

- Generating strip plots:

```
import seaborn as sns
sns.stripplot(x = 'column_name_1', y = 'column_name_2',
              data = some_dataframe)
```

- Generating multiple box plots:

```
import seaborn as sns
sns.boxplot(x = 'column_name_1', y = 'column_name_2',
            data = some_dataframe)
```

Concepts

- To compare visually frequency distributions for nominal and ordinal variables we can use **grouped bar plots**.
- To compare visually frequency distributions for variables measured on an interval or ratio scale, we can use:
    - **Step-type histograms**.
    - **Kernel density plots**.
    - **Strip plots**.
    - **Box plots**.
- A value that is much lower or much larger than the rest of the values in a distribution is called an **outlier**. A value is an outlier if:
    - It's larger than the upper quartile by 1.5 times the interquartile range.
    - It's lower than the lower quartile by 1.5 times the interquartile range.

Resources

- A seaborn tutorial on grouped bar plots, strip plots, box plots, and more.
- A seaborn tutorial on kernel density plots, histograms, and more.