Visualization with Bar Plots

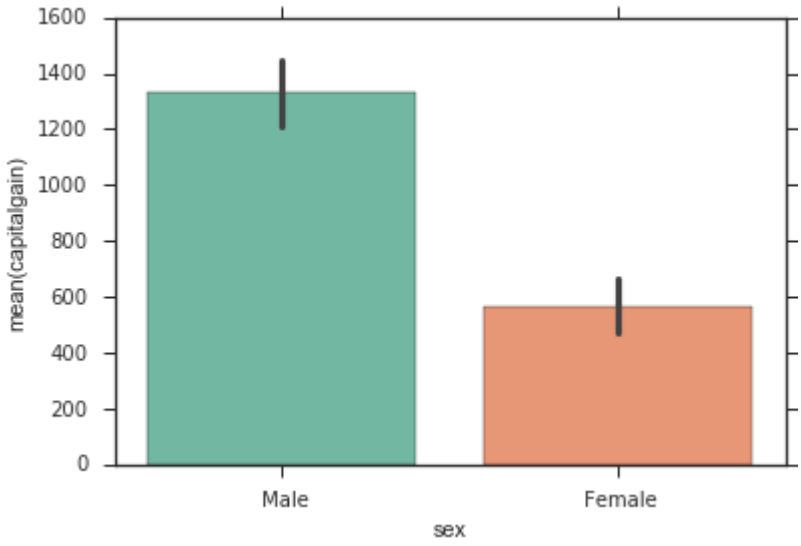
This lesson explains what a bar plot is, why it is used, and how to visualize data with bar plot using Python libraries.

WE'LL COVER THE FOLLOWING

Introduction to bar plots

Bar plots can be useful when comparing categories. For example, the bar plot below compares the capital gain values across gender.

The bar plot allows us to easily see the difference in mean capital gain tax across genders for this dataset.



Bar plots in Python

We can create this bar plot using the `barplot()` command in `Seaborn`.

The first argument to the command is the column to be used for the groups. In the plot above, **gender**.

The second argument is the column to be used for comparison. In the plot above, **capital gain**.

Let's take a look at an example with our Boston data:

```
from sklearn.datasets import load_boston
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
# Set the palette and style to be more minimal
sns.set(style='ticks', palette='Set2')

# Load data as explained in introductory lesson
boston_data = load_boston()
boston_df = pd.DataFrame(boston_data.data, columns=boston_data.feature_names)

# Only keep for ages 96 and 98.2
boston_df = boston_df[boston_df["AGE"].isin([96, 98.2])]

# Create the bar plot
sns.barplot(boston_df['AGE'], boston_df['NOX'])
# Remove excess chart lines and ticks for a nicer looking plot
sns.despine()
```

In the code, we limit our data to only include homes with `age` of 96 or 98.2 to simplify our example.

The first argument to the `barplot()` command is the `AGE` column and the second argument is the `NOX` column. Thus, the plot shows the average `NOX` value by `AGE`.

Seaborn also plots the error bars (the black lines). The error bars are calculated via bootstrapping which randomly resamples our data with replacement. It then draws 95% error bars which are the 95th and 5th percentiles.

You can see all the options for this plot here.

That's how your data can be visualized through a bar plot. Next, we'll look at a few different ways to visualize distributions.