

Plotting Distributions with Seaborn

Seaborn's strength is in visualizing statistical calculations. Seaborn includes several plots that allow you to graph univariate distribution, including KDE plots, box plots, and violin plots. Explore the Jupyter notebook below to get an understanding of how each plot works.

```
In [161]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

First, we'll read in three datasets. In order to plot them in Seaborn, we'll combine them using NumPy's `.concatenate()` function into a Pandas DataFrame.

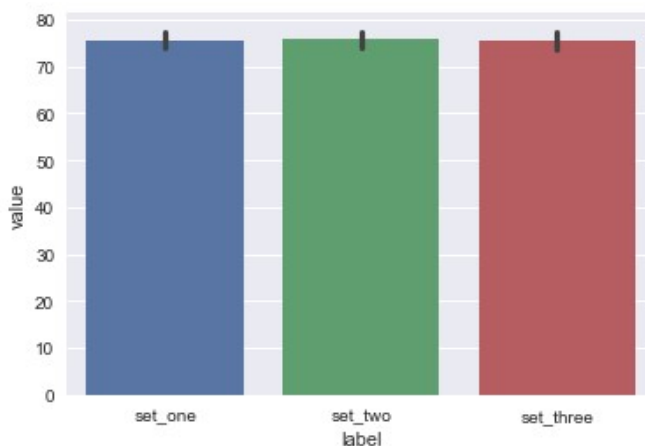
```
In [162]: n = 500
dataset1 = np.genfromtxt("dataset1.csv", delimiter=",")
dataset2 = np.genfromtxt("dataset2.csv", delimiter=",")
dataset3 = np.genfromtxt("dataset3.csv", delimiter=",")

df = pd.DataFrame({
    "label": ["set_one"] * n + ["set_two"] * n + ["set_three"] * n,
    "value": np.concatenate([dataset1, dataset2, dataset3])
})

sns.set()
```

First, let's plot each dataset as bar charts.

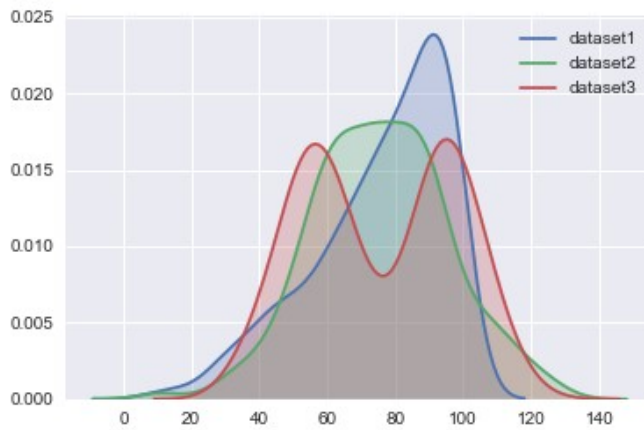
```
In [163]: sns.barplot(data=df, x='label', y='value')
plt.show()
```



We can use barplots to find out information about the mean - but it doesn't give us a sense of how spread out the data is in each set. To find out more about the distribution, we can use a KDE plot.

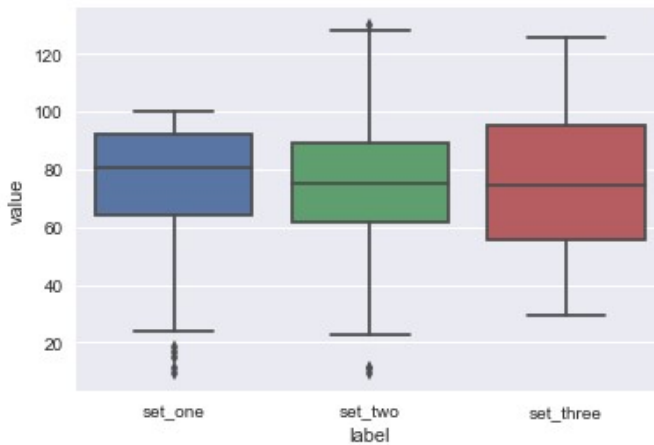
```
In [164]: sns.kdeplot(dataset1, shade=True, label="dataset1")
sns.kdeplot(dataset2, shade=True, label="dataset2")
sns.kdeplot(dataset3, shade=True, label="dataset3")

plt.legend()
plt.show()
```



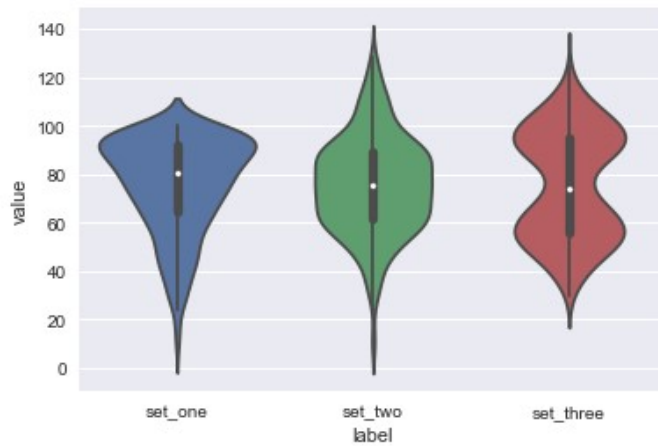
A KDE plot will give us more information, but it's pretty difficult to read this plot.

```
In [165]: sns.boxplot(data=df, x='label', y='value')
plt.show()
```



A box plot, on the other hand, makes it easier for us to compare distributions. It also gives us other information, like the interquartile range and any outliers. However, we lose the ability to determine the shape of the data.

```
In [166]: sns.violinplot(data=df, x="label", y="value")  
plt.show()
```



A violin plot brings together shape of the KDE plot with additional information that a box plot provides. It's understandable why many people like this plot!