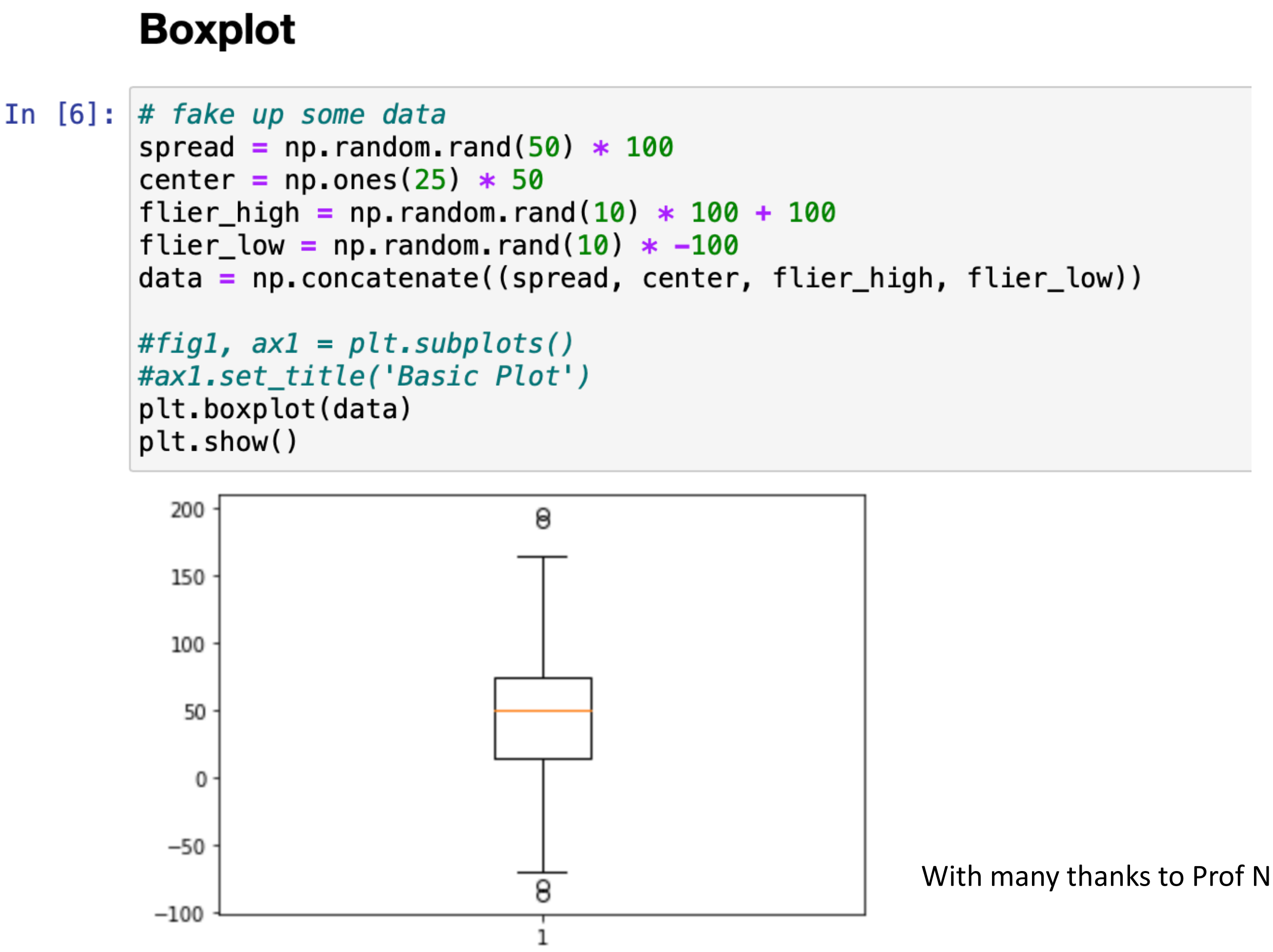
Structure

1. Boxplot
2. Histogram
3. Heatmap

CM3015

Visualisations in Python

Boxplot



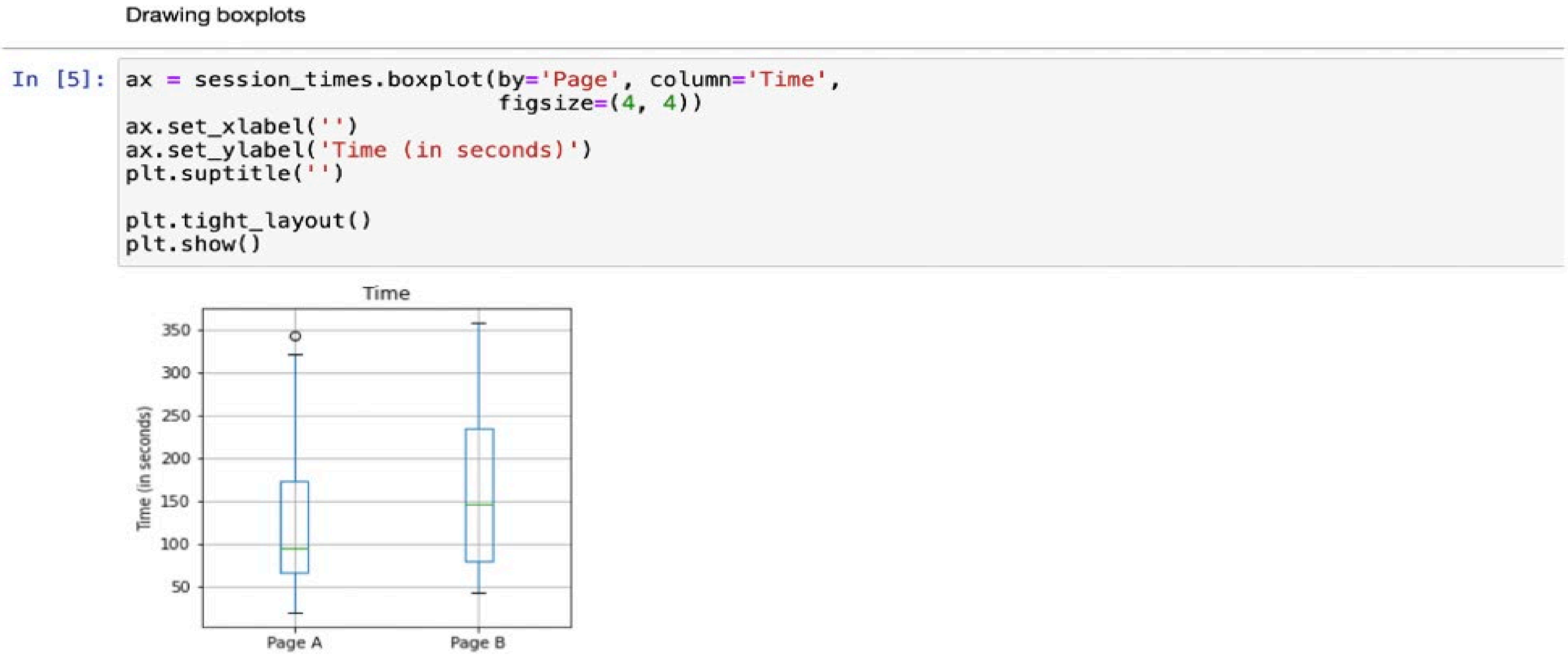
Boxplot

In [6]:

```
# fake up some data
spread = np.random.rand(50) * 100
center = np.ones(25) * 50
flier_high = np.random.rand(10) * 100 + 100
flier_low = np.random.rand(10) * -100
data = np.concatenate((spread, center, flier_high, flier_low))

#fig1, ax1 = plt.subplots()
#ax1.set_title('Basic Plot')
plt.boxplot(data)
plt.show()
```

With many thanks to Prof Noureddin Sadawi

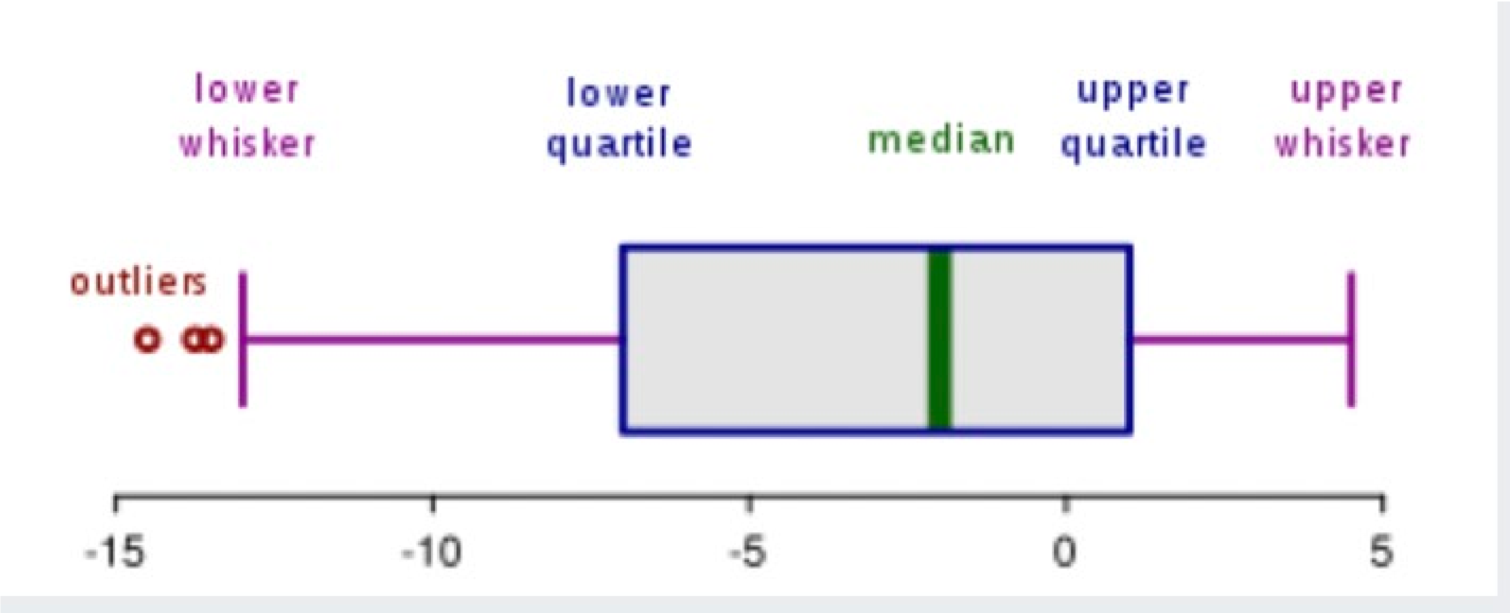
From Bruce P., Bruce A. & Gedeck, P. “Practical Statistics for Data Scientists” 2nd ed.



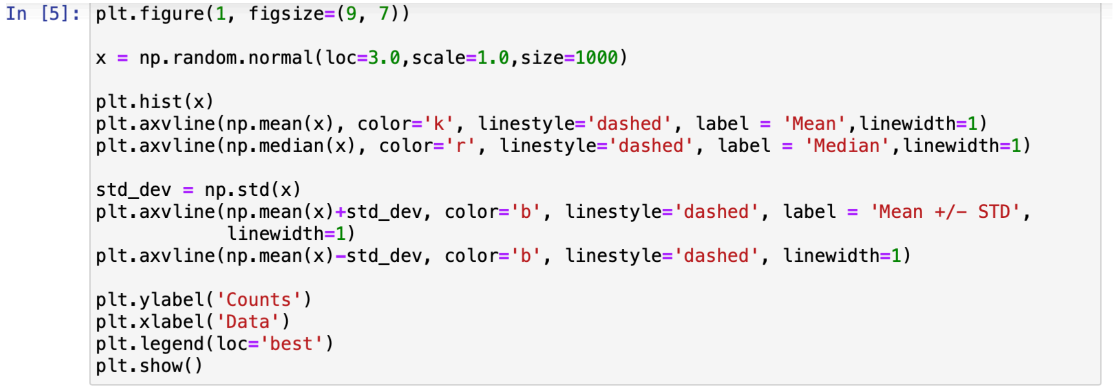
Drawing boxplots

```
In [5]: ax = session_times.boxplot(by='Page', column='Time',
                                   figsize=(4, 4))
        ax.set_xlabel('')
        ax.set_ylabel('Time (in seconds)')
        plt.suptitle('')

        plt.tight_layout()
        plt.show()
```

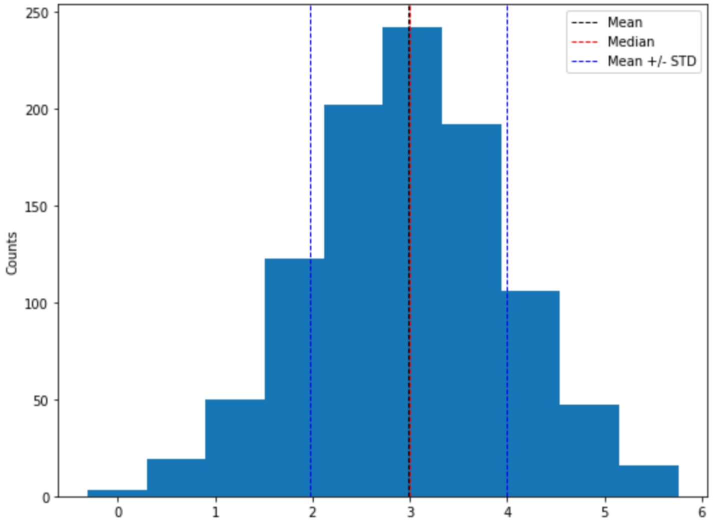
What does Boxplot represent?



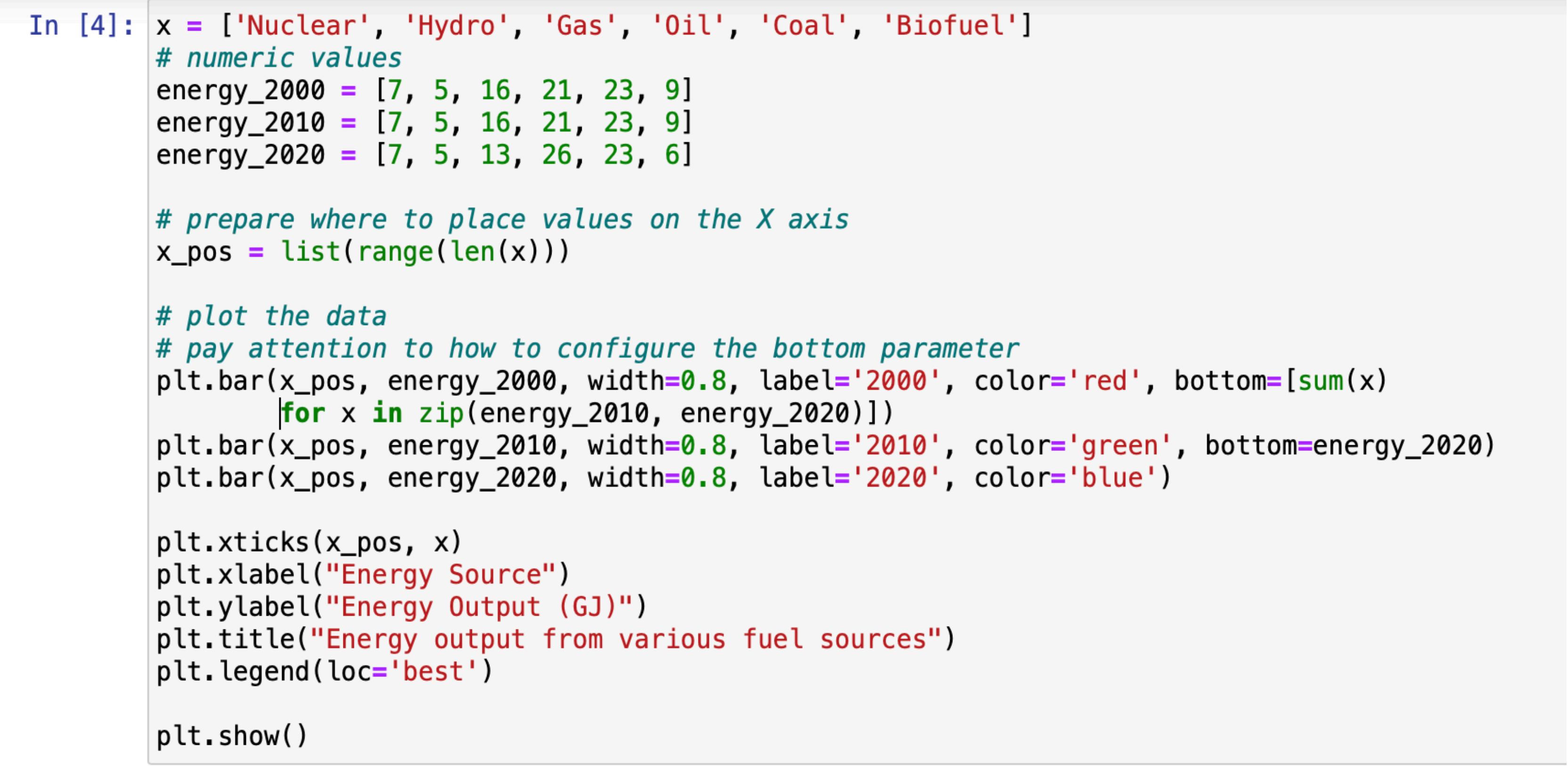
Histogram



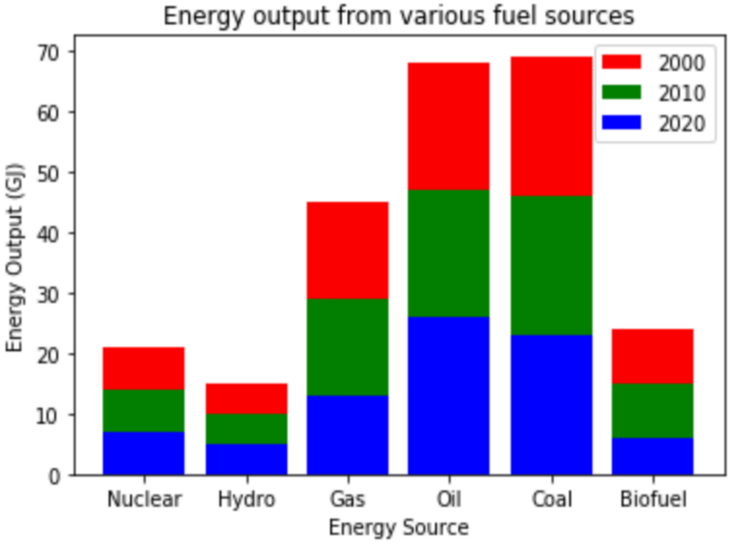
```
In [5]: plt.figure(1, figsize=(9, 7))

        x = np.random.normal(loc=3.0,scale=1.0,size=1000)

        plt.hist(x)
        plt.axvline(np.mean(x), color='k', linestyle='dashed', label = 'Mean',linewidth=1)
        plt.axvline(np.median(x), color='r', linestyle='dashed', label = 'Median',linewidth=1)

        std_dev = np.std(x)
        plt.axvline(np.mean(x)+std_dev, color='b', linestyle='dashed', label = 'Mean +/- STD',
                    linewidth=1)
        plt.axvline(np.mean(x)-std_dev, color='b', linestyle='dashed', linewidth=1)

        plt.ylabel('Counts')
        plt.xlabel('Data')
        plt.legend(loc='best')
        plt.show()
```



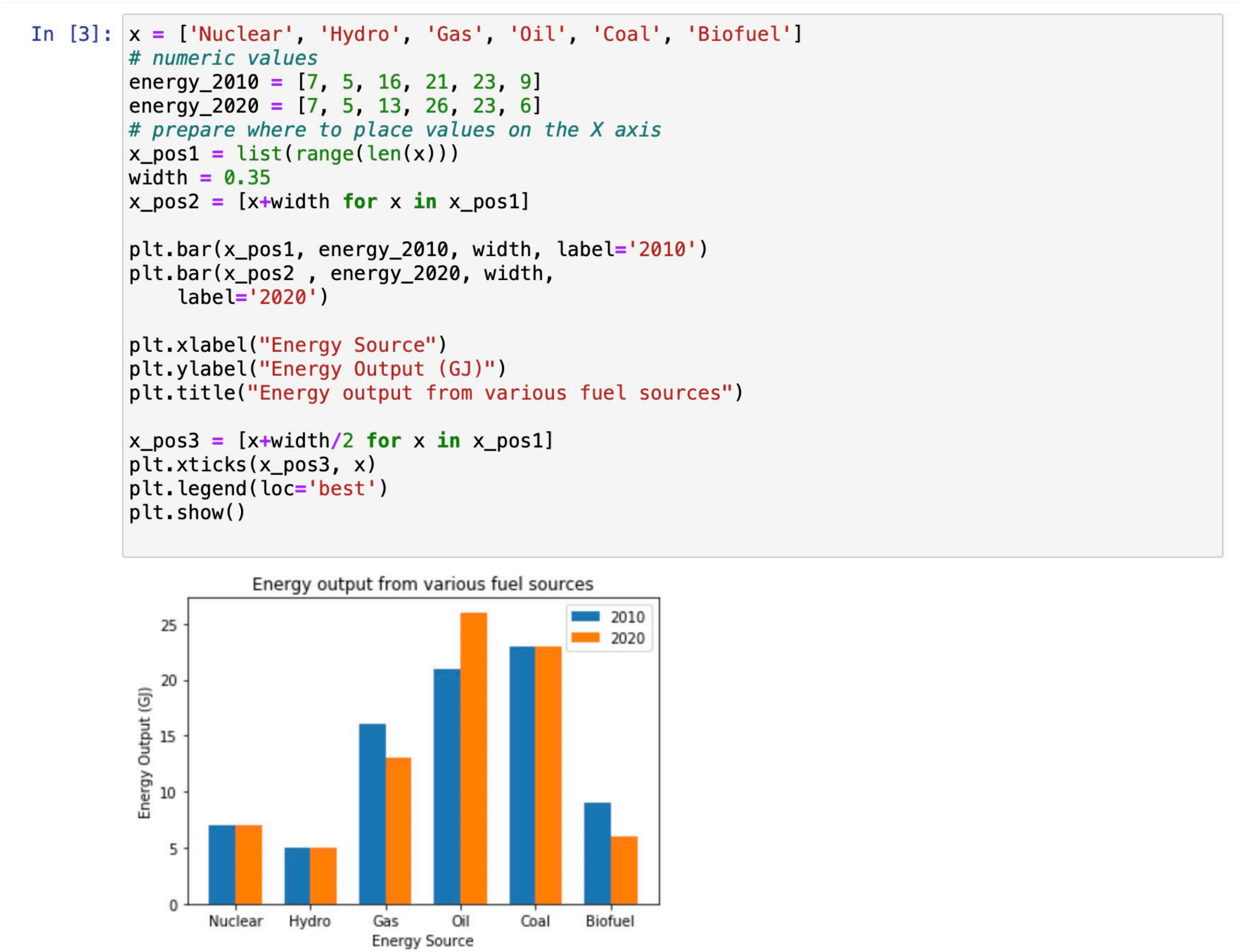
With many thanks to Prof Noureddin Sadawi



```
In [4]: x = ['Nuclear', 'Hydro', 'Gas', 'Oil', 'Coal', 'Biofuel']
        # numeric values
        energy_2000 = [7, 5, 16, 21, 23, 9]
        energy_2010 = [7, 5, 16, 21, 23, 9]
        energy_2020 = [7, 5, 13, 26, 23, 6]

        # prepare where to place values on the X axis
        x_pos = list(range(len(x)))

        # plot the data
        # pay attention to how to configure the bottom parameter
        plt.bar(x_pos, energy_2000, width=0.8, label='2000', color='red', bottom=[sum(x)
                for x in zip(energy_2010, energy_2020)])
        plt.bar(x_pos, energy_2010, width=0.8, label='2010', color='green', bottom=energy_2020)
        plt.bar(x_pos, energy_2020, width=0.8, label='2020', color='blue')

        plt.xticks(x_pos, x)
        plt.xlabel("Energy Source")
        plt.ylabel("Energy Output (GJ)")
        plt.title("Energy output from various fuel sources")
        plt.legend(loc='best')

        plt.show()
```



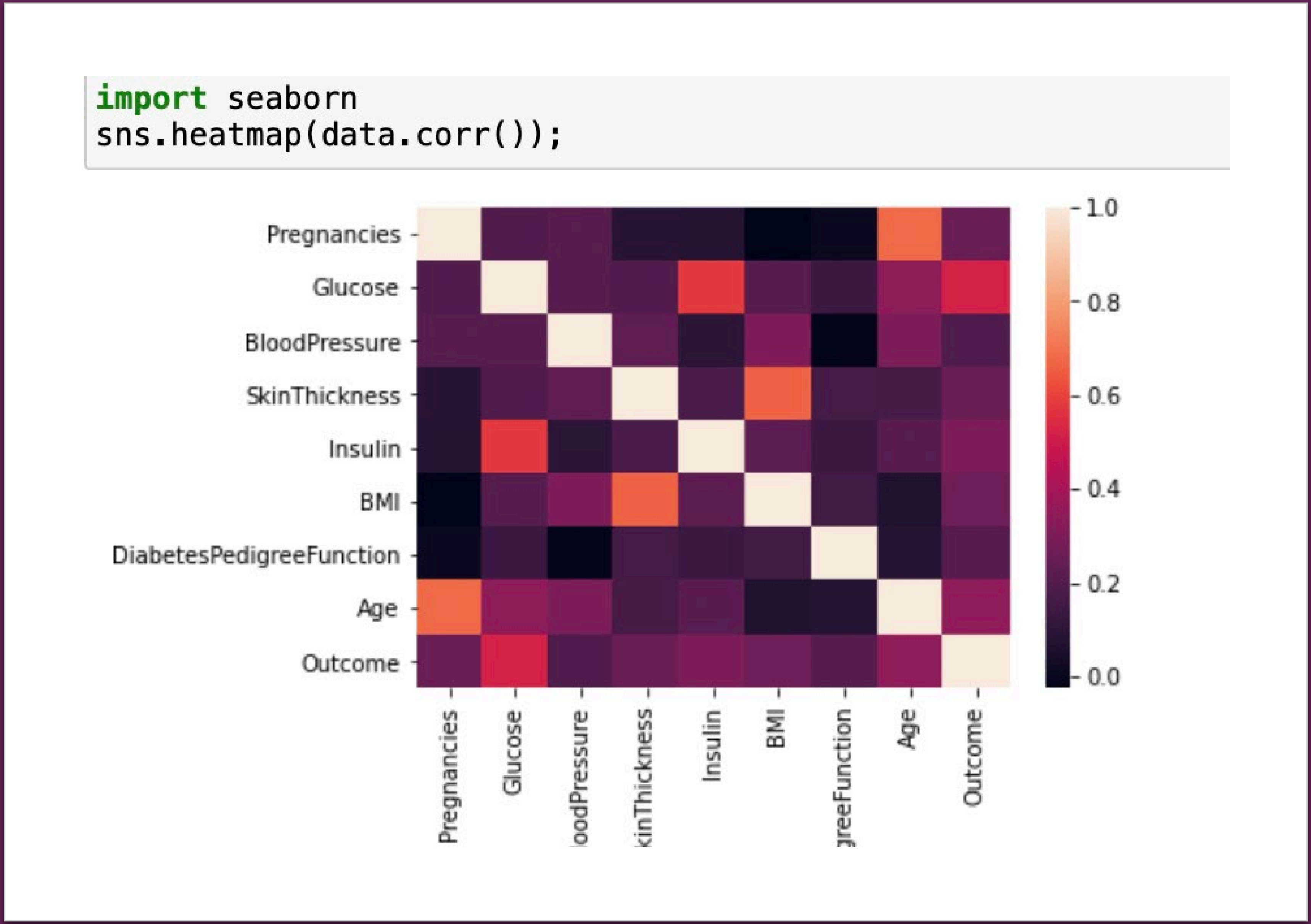
With many thanks to Prof Noureddin Sadawi



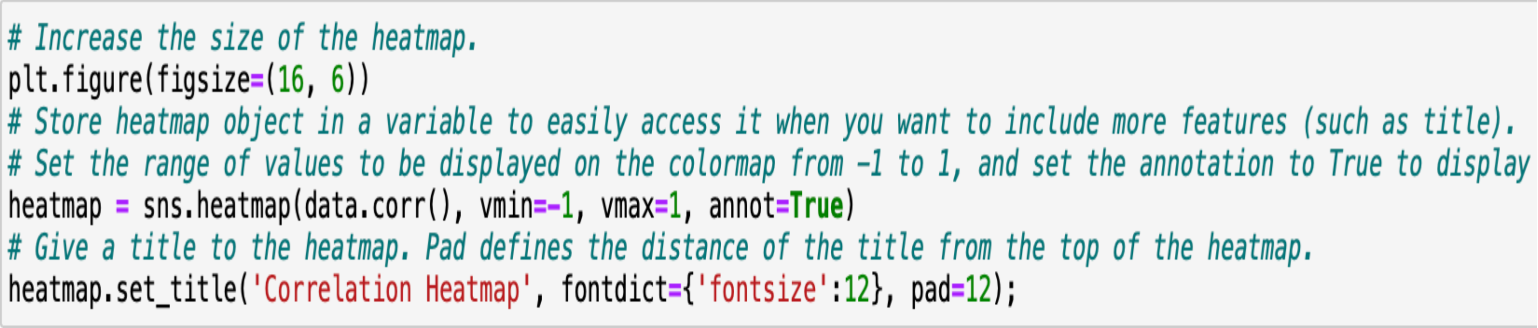
In [3]:

```
x = ['Nuclear', 'Hydro', 'Gas', 'Oil', 'Coal', 'Biofuel']
# numeric values
energy_2010 = [7, 5, 16, 21, 23, 9]
energy_2020 = [7, 5, 13, 26, 23, 6]
# prepare where to place values on the X axis
x_pos1 = list(range(len(x)))
width = 0.35
x_pos2 = [x+width for x in x_pos1]

plt.bar(x_pos1, energy_2010, width, label='2010')
plt.bar(x_pos2 , energy_2020, width,
    label='2020')

plt.xlabel("Energy Source")
plt.ylabel("Energy Output (GJ)")
plt.title("Energy output from various fuel sources")

x_pos3 = [x+width/2 for x in x_pos1]
plt.xticks(x_pos3, x)
plt.legend(loc='best')
plt.show()
```

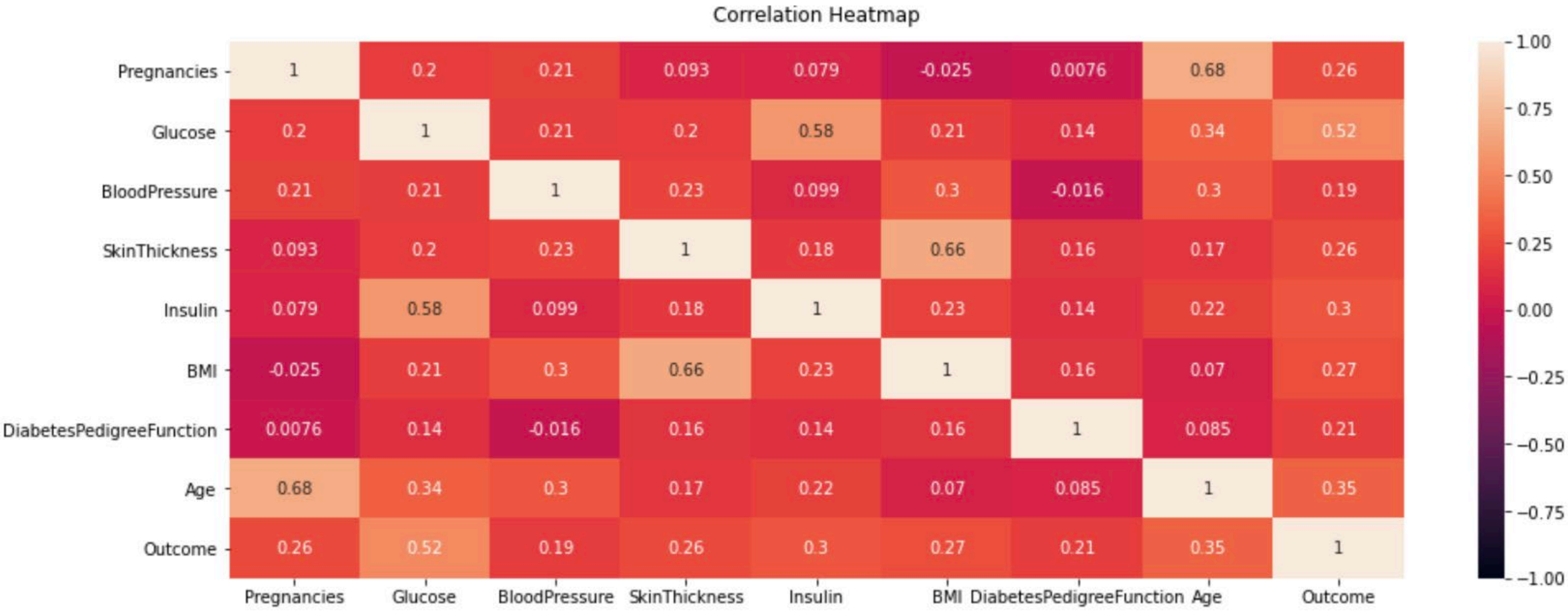
Heatmap



```
import seaborn
sns.heatmap(data.corr());
```



```
# Increase the size of the heatmap.
plt.figure(figsize=(16, 6))
# Store heatmap object in a variable to easily access it when you want to include more features (such as title).
# Set the range of values to be displayed on the colormap from -1 to 1, and set the annotation to True to display
heatmap = sns.heatmap(data.corr(), vmin=-1, vmax=1, annot=True)
# Give a title to the heatmap. Pad defines the distance of the title from the top of the heatmap.
heatmap.set_title('Correlation Heatmap', fontdict={'fontsize':12}, pad=12);
```



Correlation Heatmap

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
Pregnancies	1	0.2	0.21	0.093	0.079	-0.025	0.0076	0.68	0.26
Glucose	0.2	1	0.21	0.2	0.58	0.21	0.14	0.34	0.52
BloodPressure	0.21	0.21	1	0.23	0.099	0.3	-0.016	0.3	0.19
SkinThickness	0.093	0.2	0.23	1	0.18	0.66	0.16	0.17	0.26
Insulin	0.079	0.58	0.099	0.18	1	0.23	0.14	0.22	0.3
BMI	-0.025	0.21	0.3	0.66	0.23	1	0.16	0.07	0.27
DiabetesPedigreeFunction	0.0076	0.14	-0.016	0.16	0.14	0.16	1	0.085	0.21
Age	0.68	0.34	0.3	0.17	0.22	0.07	0.085	1	0.35
Outcome	0.26	0.52	0.19	0.26	0.3	0.27	0.21	0.35	1