Capstone-1

Anish.V.Karanth

anish.karanth@gmail.com

Import all Libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import zscore, pearsonr, spearmanr
```

Download and read dataset

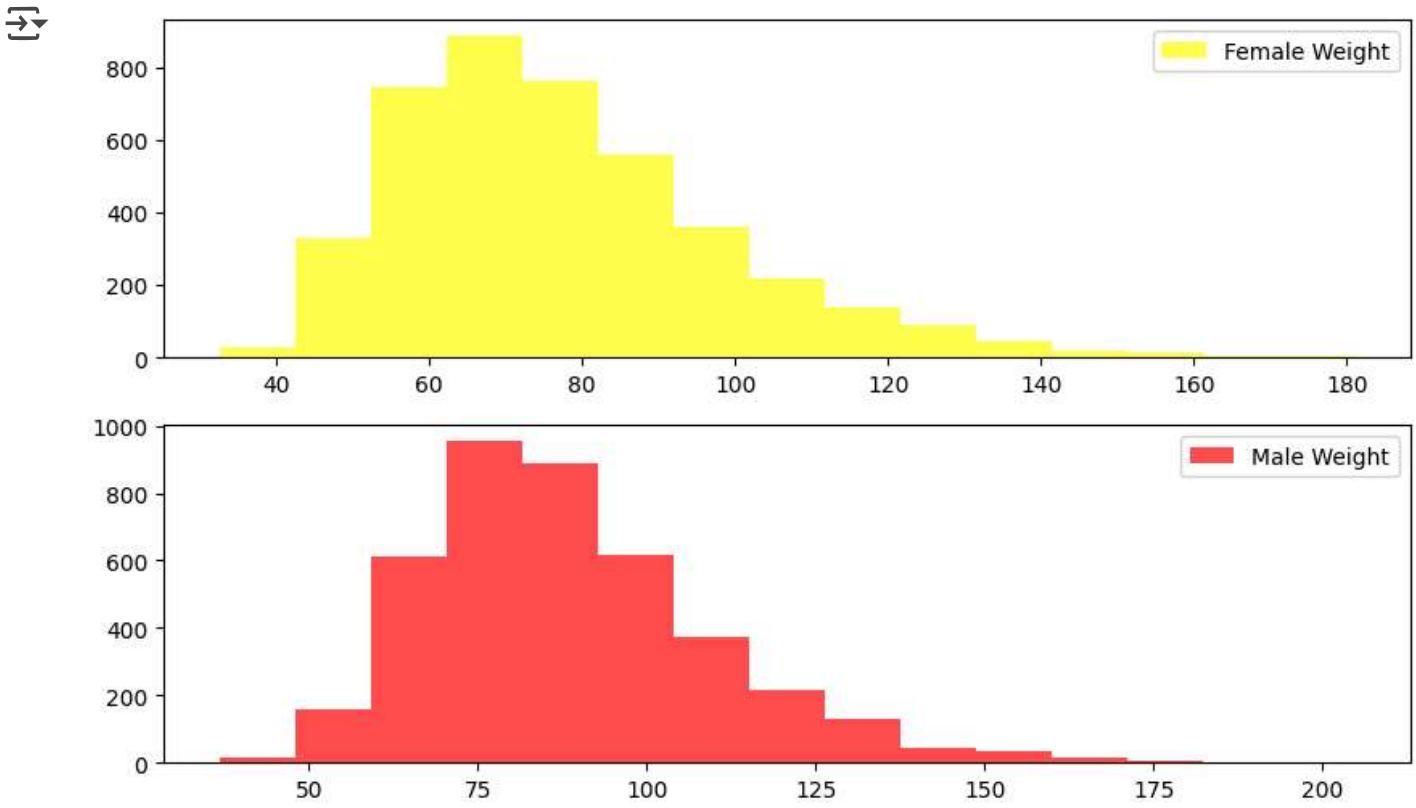
```
male_url = "https://raw.githubusercontent.com/gagolews/teaching-data/master/marek/nhanes_adu
female_url = "https://raw.githubusercontent.com/gagolews/teaching-data/master/marek/nhanes_a

male = np.genfromtxt(male_url, delimiter=",", skip_header=1, dtype=float)
female = np.genfromtxt(female_url, delimiter=",", skip_header=1, dtype=float)


print(f"Male data shape: {male.shape}")
print(f"Female data shape: {female.shape}")
```

```
Male data shape: (4082, 7)
Female data shape: (4222, 7)
```

Compare the weight distributions of male and female participants

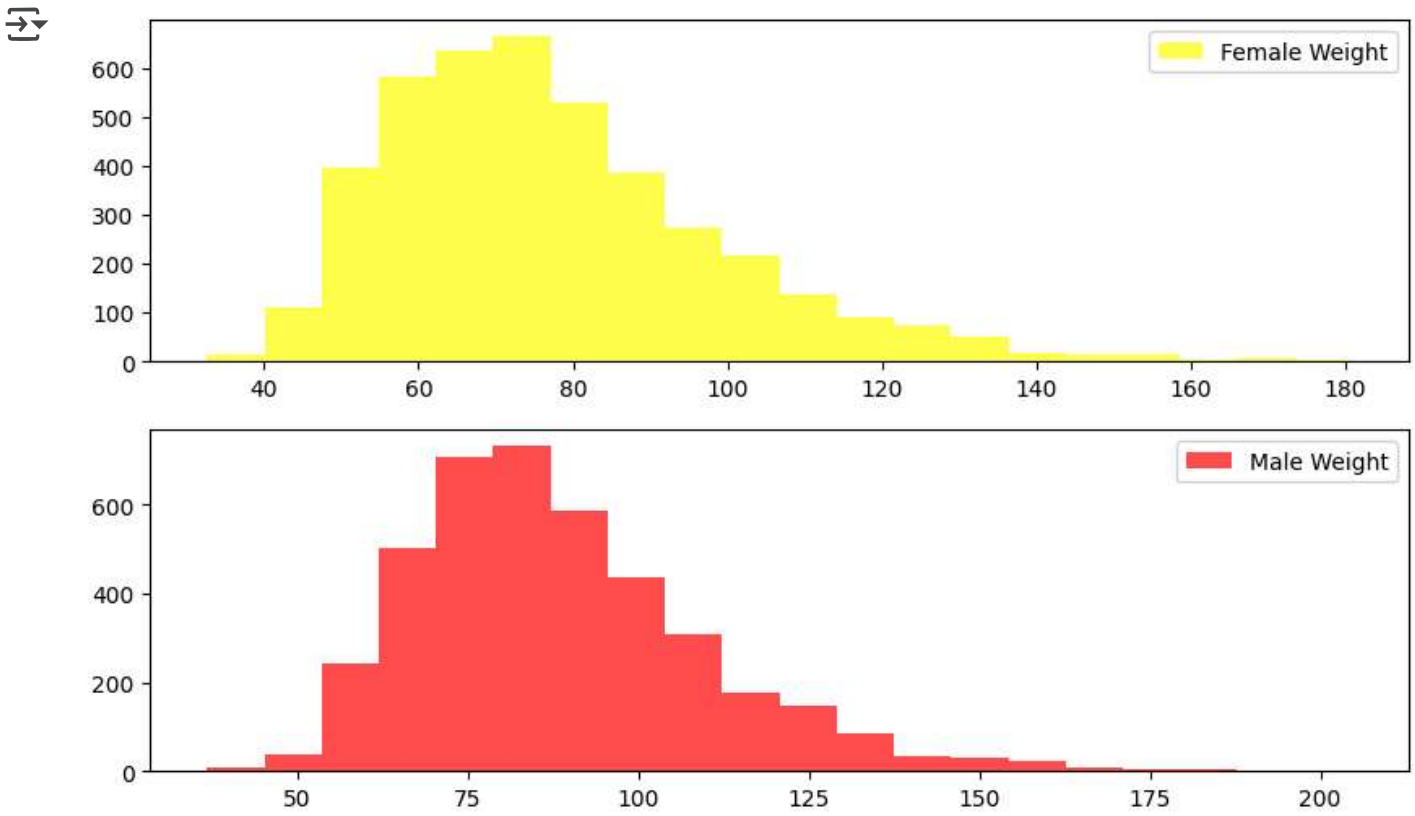
```
plt.figure(figsize=(10, 6))
plt.subplot(2, 1, 1)
plt.hist(female[:, 0], bins=15, alpha=0.7, color='yellow', label='Female Weight')
plt.legend()

plt.subplot(2, 1, 2)
```

```
plt.hist(male[:, 0], bins=15, alpha=0.7, color='red', label='Male Weight')
plt.legend()
plt.show()
```



Plot Histograms of Weights

```
plt.figure(figsize=(10, 6))
plt.subplot(2, 1, 1)
plt.hist(female[:, 0], bins=20, alpha=0.7, color='yellow', label='Female Weight')
plt.legend()
plt.subplot(2, 1, 2)
plt.hist(male[:, 0], bins=20, alpha=0.7, color='red', label='Male Weight')
plt.legend()
plt.show()
```



Check for NaN or invalid values

```
print("Male dataset preview:\n", male[:5])
print("Female dataset preview:\n", female[:5])
print("Any NaNs in Male Data:", np.isnan(male).any())
print("Any NaNs in Female Data:", np.isnan(female).any())
```

```
Male dataset preview:
 [[  nan   nan   nan   nan   nan   nan   nan]
 [ 98.8 182.3  42.   40.1  38.2 108.2 120.4]
 [ 74.3 184.2  41.1  41.   30.2  94.5  86.8]
 [103.7 185.3  47.   44.   32.  107.8 109.6]
 [ 86.  167.8  39.5  38.4  29.  106.4 108.3]]
Female dataset preview:
 [[  nan   nan   nan   nan   nan   nan   nan]
 [ 97.1 160.2  34.7  40.8  35.8 126.1 117.9]
 [ 91.1 152.7  33.5  33.   38.5 125.5 103.1]
 [ 73.  161.2  37.4  38.   31.8 106.2  92. ]
 [ 61.7 157.4  38.   34.7  29.  101.   90.5]]
Any NaNs in Male Data: True
Any NaNs in Female Data: True
```

Remove NaN data

```
male = male[~np.isnan(male).any(axis=1)]
female = female[~np.isnan(female).any(axis=1)]
```

Checking Data post clearing

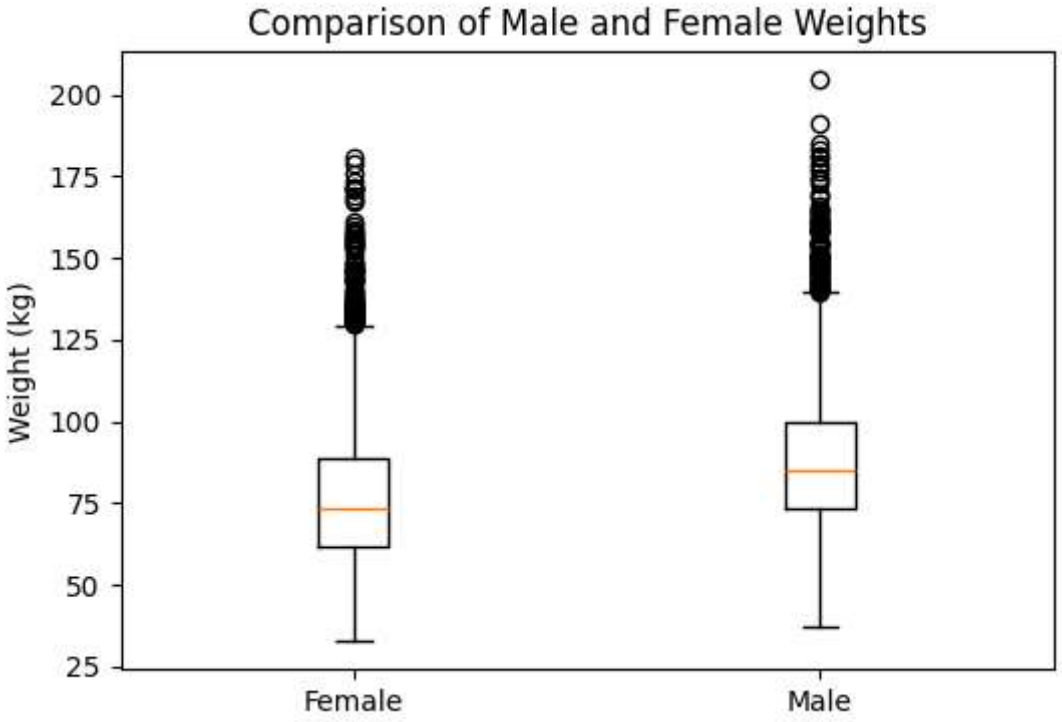
```
print("Male dataset preview:\n", male[:5])
print("Female dataset preview:\n", female[:5])

print("Any NaNs in Male Data:", np.isnan(male).any())
print("Any NaNs in Female Data:", np.isnan(female).any())
```

```
Male dataset preview:
 [[ 98.8 182.3  42.   40.1  38.2 108.2 120.4]
 [ 74.3 184.2  41.1  41.   30.2  94.5  86.8]
 [103.7 185.3  47.   44.   32.  107.8 109.6]
 [ 86.  167.8  39.5  38.4  29.  106.4 108.3]
 [ 99.4 181.6  40.4  39.9  36.  120.2 107. ]]
Female dataset preview:
 [[ 97.1 160.2  34.7  40.8  35.8 126.1 117.9]
 [ 91.1 152.7  33.5  33.   38.5 125.5 103.1]
 [ 73.  161.2  37.4  38.   31.8 106.2  92. ]
 [ 61.7 157.4  38.   34.7  29.  101.   90.5]
 [ 55.4 154.6  34.6  34.   28.3  92.5  73.2]]
Any NaNs in Male Data: False
Any NaNs in Female Data: False
```

Boxplot for Weight Comparison

```
plt.figure(figsize=(6, 4))
plt.boxplot([female[:, 0], male[:, 0]], labels=['Female', 'Male'])
plt.ylabel("Weight (kg)")
plt.title("Comparison of Male and Female Weights")
plt.show()
```

```
<ipython-input-8-5fd59ffc8dc8>:2: MatplotlibDeprecationWarning: The 'labels' parameter c
  plt.boxplot([female[:, 0], male[:, 0]], labels=['Female', 'Male'])
```



Compute Basic Numerical Aggregates

```
def describe_data(data, label):
    print(f"Statistics for {label}:")
    print(f"Mean: {np.mean(data):.2f}, Median: {np.median(data):.2f}")
    print(f"Standard Deviation: {np.std(data):.2f}")
    print(f"Min: {np.min(data)}, Max: {np.max(data)}")
    print("-----------")

describe_data(female[:, 0], "Female Weights")
describe_data(male[:, 0], "Male Weights")
```

```
Statistics for Female Weights:
Mean: 77.40, Median: 73.60
Standard Deviation: 21.54
Min: 32.6, Max: 180.9
-----------
Statistics for Male Weights:
Mean: 88.36, Median: 85.00
Standard Deviation: 21.42
Min: 36.8, Max: 204.6
-----------
```

Compute and Append BMI

```
def compute_bmi(data):
    weight = data[:, 0]
    height_m = data[:, 1] / 100
    bmi = weight / (height_m ** 2)
    return np.column_stack((data, bmi))

female = compute_bmi(female)
male = compute_bmi(male)
```

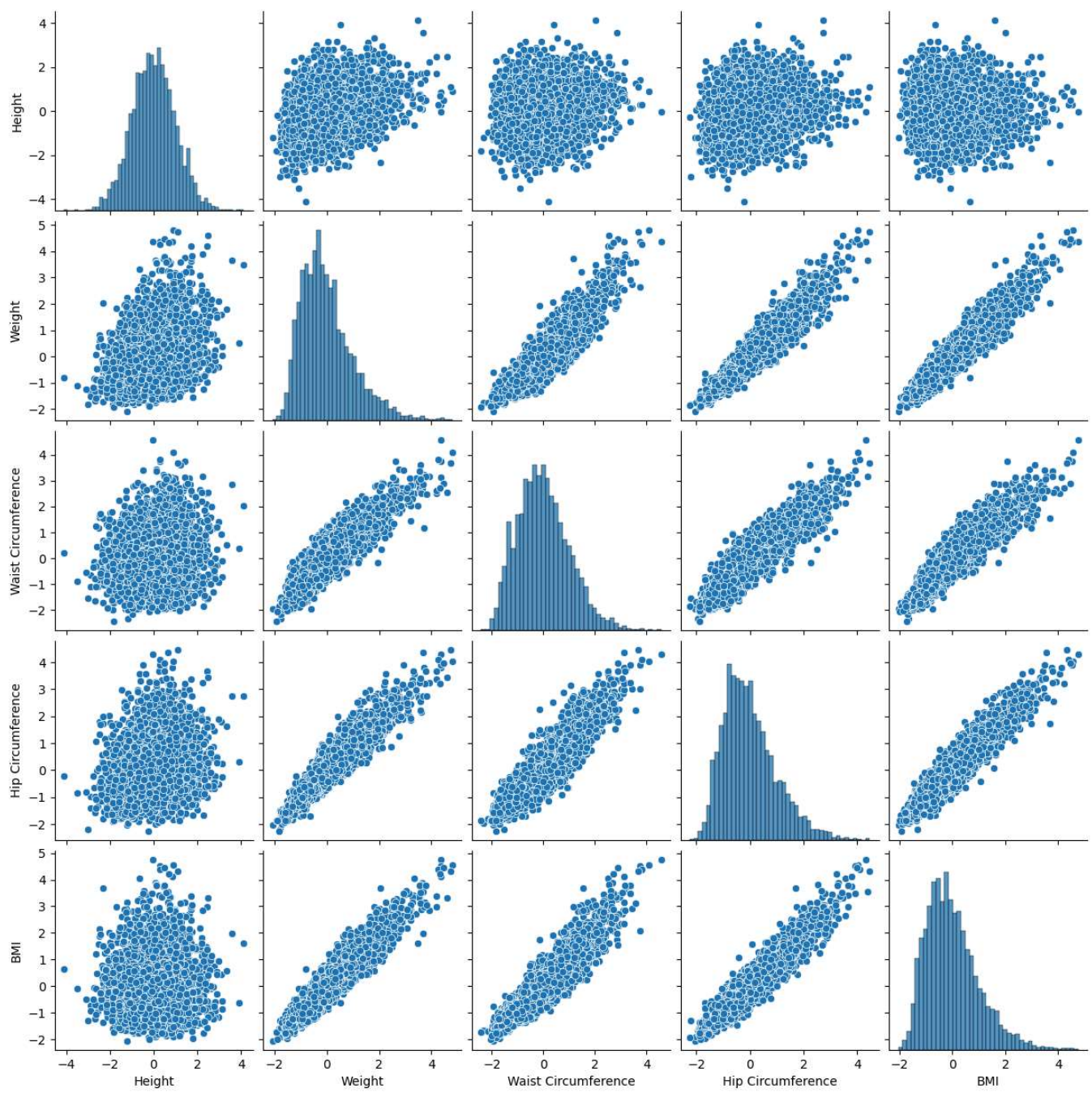
Standardize Female Data

```
zfemale = np.apply_along_axis(zscore, 0, female)
```

Scatterplot Matrix and Correlations

```
df_zfemale = pd.DataFrame(zfemale, columns=["Weight", "Height", "Arm Length", "Leg Length",
sns.pairplot(df_zfemale[["Height", "Weight", "Waist Circumference", "Hip Circumference", "BM
plt.show()
```

```
for col1 in df_zfemale.columns:
    for col2 in df_zfemale.columns:
        if col1 != col2:
            pearson_corr, _ = pearsonr(df_zfemale[col1], df_zfemale[col2])
            spearman_corr, _ = spearmanr(df_zfemale[col1], df_zfemale[col2])
            print(f"Pearson({col1}, {col2}): {pearson_corr:.2f}, Spearman: {spearman_corr:.2
```




```
Pearson(Weight, Height): 0.35, Spearman: 0.34
Pearson(Weight, Arm Length): 0.55, Spearman: 0.54
Pearson(Weight, Leg Length): 0.19, Spearman: 0.20
Pearson(Weight, Arm Circumference): 0.91, Spearman: 0.91
Pearson(Weight, Hip Circumference): 0.95, Spearman: 0.95
Pearson(Weight, Waist Circumference): 0.90, Spearman: 0.90
Pearson(Weight, BMI): 0.95, Spearman: 0.94
Pearson(Height, Weight): 0.35, Spearman: 0.34
Pearson(Height, Arm Length): 0.67, Spearman: 0.67
Pearson(Height, Leg Length): 0.66, Spearman: 0.65
Pearson(Height, Arm Circumference): 0.15, Spearman: 0.14
Pearson(Height, Hip Circumference): 0.20, Spearman: 0.21
Pearson(Height, Waist Circumference): 0.13, Spearman: 0.11
Pearson(Height, BMI): 0.03, Spearman: 0.02
Pearson(Arm Length, Weight): 0.55, Spearman: 0.54
Pearson(Arm Length, Height): 0.67, Spearman: 0.67
Pearson(Arm Length, Leg Length): 0.48, Spearman: 0.46
```

```
Pearson(Arm Length, Arm Circumference): 0.45, Spearman: 0.42
Pearson(Arm Length, Hip Circumference): 0.46, Spearman: 0.45
Pearson(Arm Length, Waist Circumference): 0.43, Spearman: 0.40
Pearson(Arm Length, BMI): 0.36, Spearman: 0.34
Pearson(Leg Length, Weight): 0.19, Spearman: 0.20
Pearson(Leg Length, Height): 0.66, Spearman: 0.65
Pearson(Leg Length, Arm Length): 0.48, Spearman: 0.46
Pearson(Leg Length, Arm Circumference): 0.08, Spearman: 0.08
Pearson(Leg Length, Hip Circumference): 0.10, Spearman: 0.12
Pearson(Leg Length, Waist Circumference): -0.03, Spearman: -0.04
Pearson(Leg Length, BMI): -0.01, Spearman: -0.02
Pearson(Arm Circumference, Weight): 0.91, Spearman: 0.91
Pearson(Arm Circumference, Height): 0.15, Spearman: 0.14
Pearson(Arm Circumference, Arm Length): 0.45, Spearman: 0.42
Pearson(Arm Circumference, Leg Length): 0.08, Spearman: 0.08
Pearson(Arm Circumference, Hip Circumference): 0.87, Spearman: 0.87
Pearson(Arm Circumference, Waist Circumference): 0.85, Spearman: 0.84
Pearson(Arm Circumference, BMI): 0.92, Spearman: 0.92
Pearson(Hip Circumference, Weight): 0.95, Spearman: 0.95
Pearson(Hip Circumference, Height): 0.20, Spearman: 0.21
Pearson(Hip Circumference, Arm Length): 0.46, Spearman: 0.45
Pearson(Hip Circumference, Leg Length): 0.10, Spearman: 0.12
Pearson(Hip Circumference, Arm Circumference): 0.87, Spearman: 0.87
Pearson(Hip Circumference, Waist Circumference): 0.90, Spearman: 0.89
Pearson(Hip Circumference, BMI): 0.94, Spearman: 0.93
Pearson(Waist Circumference, Weight): 0.90, Spearman: 0.90
Pearson(Waist Circumference, Height): 0.13, Spearman: 0.11
Pearson(Waist Circumference, Arm Length): 0.43, Spearman: 0.40
Pearson(Waist Circumference, Leg Length): -0.03, Spearman: -0.04
Pearson(Waist Circumference, Arm Circumference): 0.85, Spearman: 0.84
Pearson(Waist Circumference, Hip Circumference): 0.90, Spearman: 0.89
Pearson(Waist Circumference, BMI): 0.92, Spearman: 0.92
Pearson(BMI, Weight): 0.95, Spearman: 0.94
Pearson(BMI, Height): 0.03, Spearman: 0.02
Pearson(BMI, Arm Length): 0.36, Spearman: 0.34
Pearson(BMI, Leg Length): -0.01, Spearman: -0.02
Pearson(BMI, Arm Circumference): 0.92, Spearman: 0.92
Pearson(BMI, Hip Circumference): 0.94, Spearman: 0.93
Pearson(BMI, Waist Circumference): 0.92, Spearman: 0.92
```

Compute Waist-to-Height and Waist-to-Hip Ratios

```
# ## **9. Compute Waist-to-Height and Waist-to-Hip Ratios**
def compute_ratios(data):
    waist_to_height = data[:, 6] / data[:, 1]
    waist_to_hip = data[:, 6] / data[:, 5]
    return np.column_stack((data, waist_to_height, waist_to_hip))

female = compute_ratios(female)
male = compute_ratios(male)
```

Boxplot Comparison of Ratios

```
# ## **10. Boxplot Comparison of Ratios**
plt.figure(figsize=(8, 5))
plt.boxplot([female[:, 8], male[:, 8], female[:, 9], male[:, 9]],
            labels=['Female W/Ht', 'Male W/Ht', 'Female W/Hip', 'Male W/Hip'])
plt.ylabel("Ratio Value")
plt.title("Comparison of Waist Ratios Between Genders")
plt.show()
```