

HW2 Selected problem solutions

Part I

Import Packages

```
import statistics
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

Read in the CSV file

```
df = pd.read_csv("C:/Users/seh00004.UCONN/Desktop/CSE5520/HW/Data/nba.csv")
```

Step2

```
height = df["Height"]
mu = height.mean()
sigma = height.std()
```

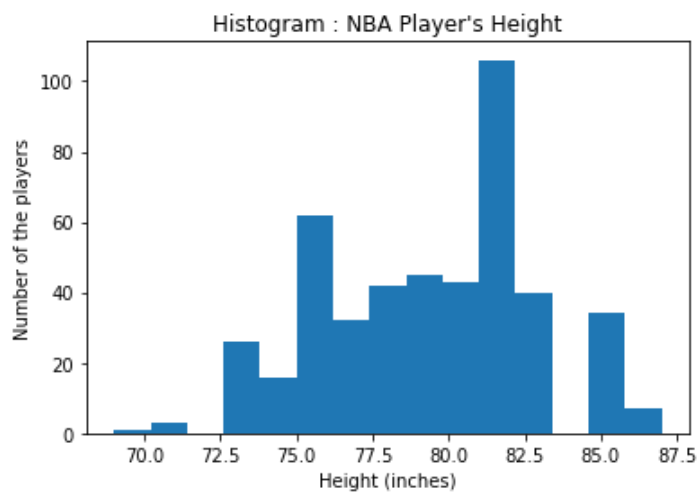
```
print("The average height : %4.2f (inches)" % mu)
print("The standard deviation of the height : %4.2f (inches)" % sigma)
```

The average height : 79.27 (inches)

The standard deviation of the height : 3.48 (inches)

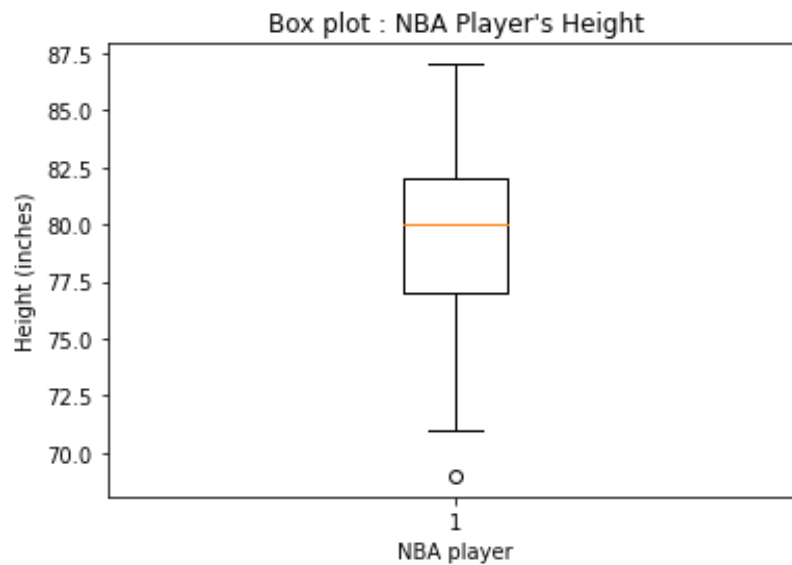
Step3

```
hist_heights = plt.hist(height, 15)
plt.xlabel("Height (inches)")
plt.ylabel("Number of the players")
plt.title("Histogram : NBA Player's Height")
plt.show()
```



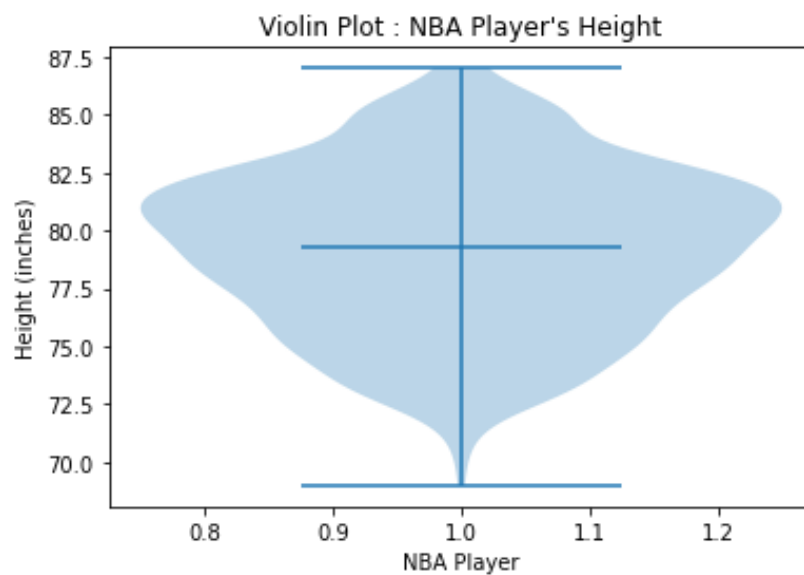
Step4

```
boxplot_heights = plt.boxplot(height)
plt.ylabel("Height (inches)")
plt.xlabel("NBA player")
plt.title("Box plot : NBA Player's Height")
plt.show()
```



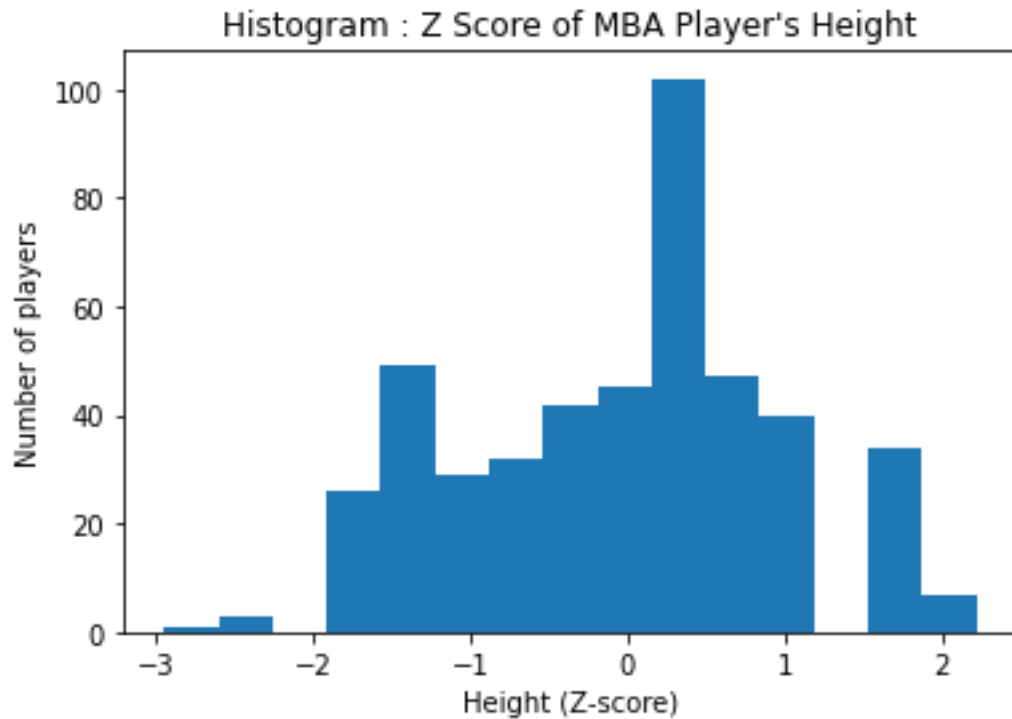
Step5

```
violinplot_heights = plt.violinplot(height, showmeans=True)
plt.ylabel("Height (inches)")
plt.xlabel("NBA Player")
plt.title("Violin Plot : NBA Player's Height")
plt.show()
```



Step6

```
zscore_height = (height - mu) / sigma
hist_heights_zscore = plt.hist(zscore_height, 15)
plt.xlabel("Height (Z-score)")
plt.ylabel("Number of players")
plt.title("Histogram : Z Score of MBA Player's Height")
plt.show()
```



Step7

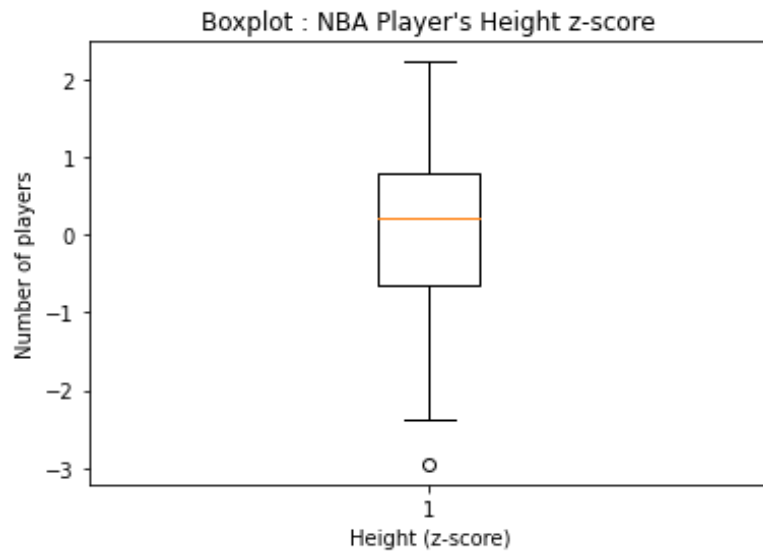
```
mu_z = zscore_height.mean()
sigma_z = zscore_height.std()

print("The average z-value of the heights : %6.5f" % mu_z)
print("The standard deviation of the z-values of the heights : %4.2f" % sigma_z)
```

The average z-value of the heights : -0.00000
The standard deviation of the z-values of the heights : 1.00

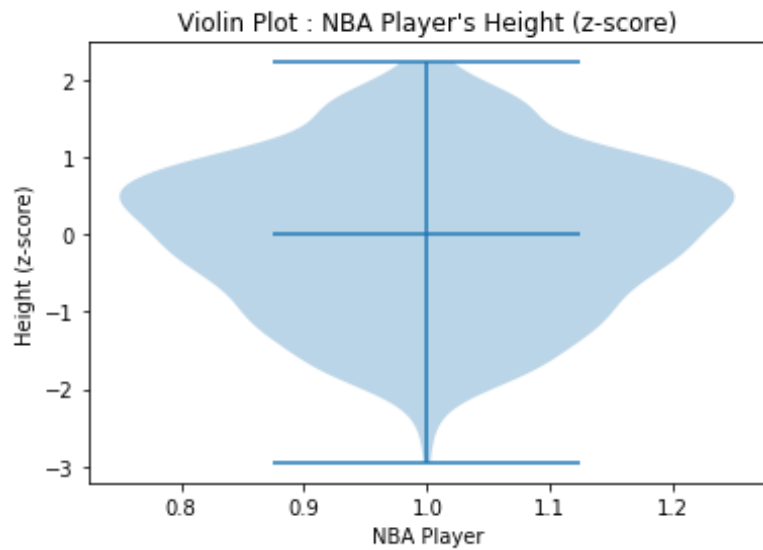
Step8

```
boxplot_zscore_height = plt.boxplot(zscore_height)
plt.xlabel("Height (z-score)")
plt.ylabel("Number of players")
plt.title("Boxplot : NBA Player's Height z-score")
plt.show()
```



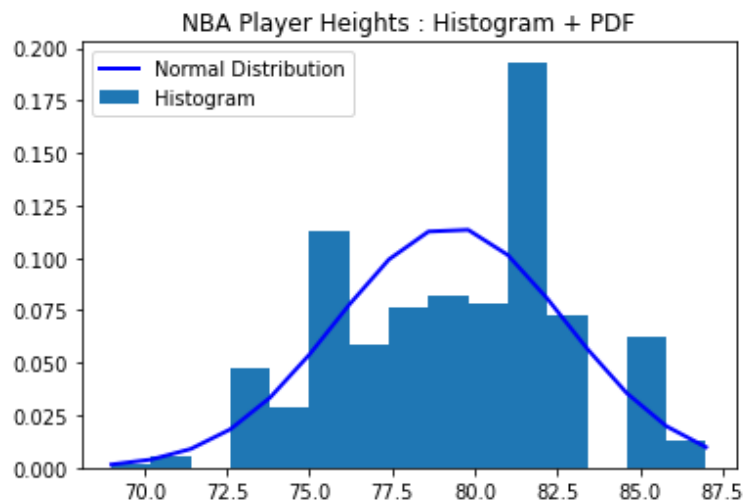
Step9

```
violinplot_xscore_heights = plt.violinplot(zscore_height, showmeans=True)
plt.ylabel("Height (z-score)")
plt.xlabel("NBA Player")
plt.title("Violin Plot : NBA Player's Height (z-score)")
plt.show()
```



Step10

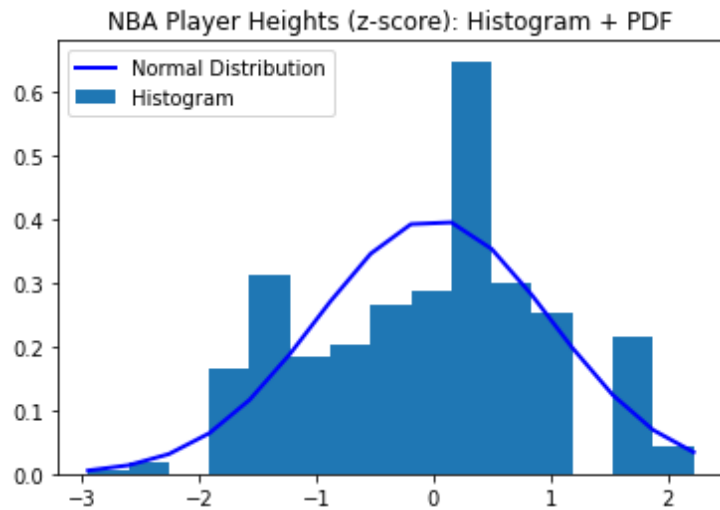
```
count, bins, ignored = plt.hist(height, 15, density=True, label= "Histogram")
plt.plot(bins, 1/(sigma*np.sqrt(2*np.pi))*np.exp(-(bins-mu)**2/(2*sigma**2)),
        linewidth=2, color='b', label="Normal Distribution")
plt.legend()
plt.title("NBA Player Heights : Histogram + PDF")
plt.show()
```



Step11

```
count, bins, ignored = plt.hist(zscore_height, 15, density=True, label= "Histogram")

plt.plot(bins, 1/(sigma_z*np.sqrt(2*np.pi))*np.exp(-(bins-mu_z)**2/(2*sigma_z**2)),
         linewidth=2, color='b', label="Normal Distribution")
plt.legend()
plt.title("NBA Player Heights (z-score): Histogram + PDF")
plt.show()
```



Step13

```
PlyersTallerThan82 = len(height[height>82])/len(height)
print("P(x > 82)= %4.2f" % p)
```

$P(x > 82) = 0.18$