In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [6]:
```
#Read the CSV File
df=pd.read_csv("diabetes.csv")
```

In [7]:
```
df.head()
```

Out[7]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction
0	6	148	72	35	0	33.6	0.627
1	1	85	66	29	0	26.6	0.351
2	8	183	64	0	0	23.3	0.672
3	1	89	66	23	94	28.1	0.167
4	0	137	40	35	168	43.1	2.288

In [8]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   Pregnancies               768 non-null    int64  
 1   Glucose                   768 non-null    int64  
 2   BloodPressure             768 non-null    int64  
 3   SkinThickness             768 non-null    int64  
 4   Insulin                   768 non-null    int64  
 5   BMI                       768 non-null    float64
 6   DiabetesPedigreeFunction  768 non-null    float64
 7   Age                       768 non-null    int64  
 8   Outcome                   768 non-null    int64  
dtypes: float64(2), int64(7)
memory usage: 54.1 KB
```

In [9]:
```
#lets find the average age of  diabetic patient
mean_Age=df['Age'].mean()
print(f"Mean age is:{mean_Age}")
```

```
Mean age is:33.240885416666664
```

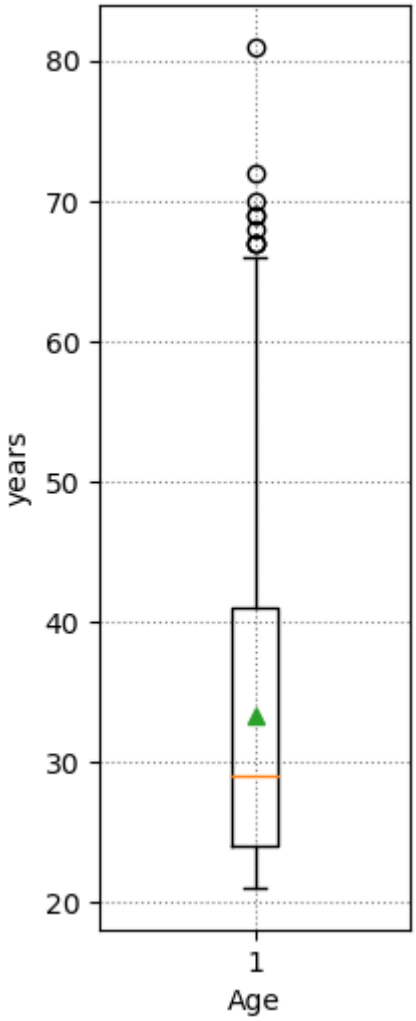
In [10]:
```
#variance
var_age=df['Age'].var()
print(f"variance in age :{var_age}")
```

```
variance in age :138.30304589037365
```

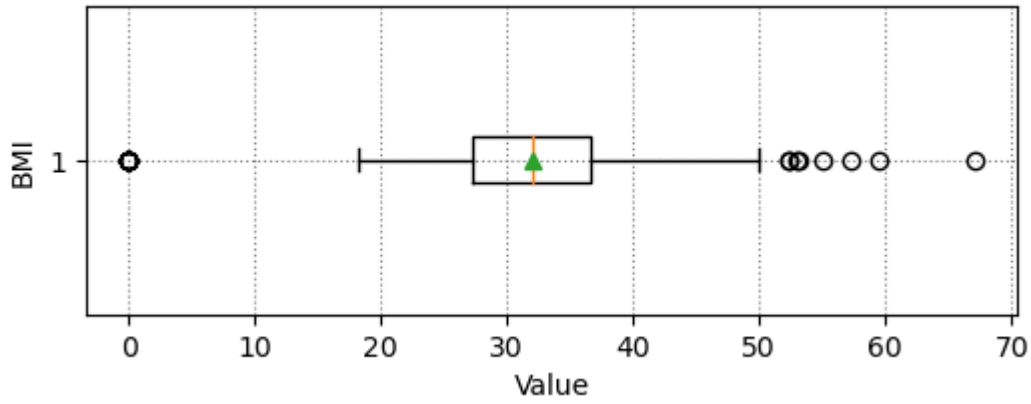
In [11]:
```
#standard deviation
std_Age=df['Age'].std()
print(f"standard deviation:{std_Age}")
```

```
standard deviation:11.76023154067868
```

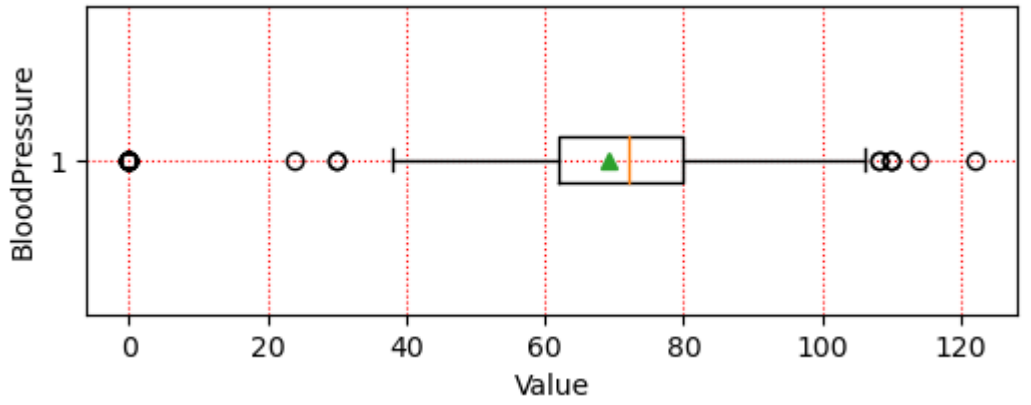
In [16]:
```
#boxplotting for the agew field
plt.figure(figsize=(2,6))
plt.boxplot(df['Age'],showmeans=True)
plt.grid(color='gray',linestyle='dotted')
plt.xlabel("Age")
plt.ylabel("years")
plt.show()
```



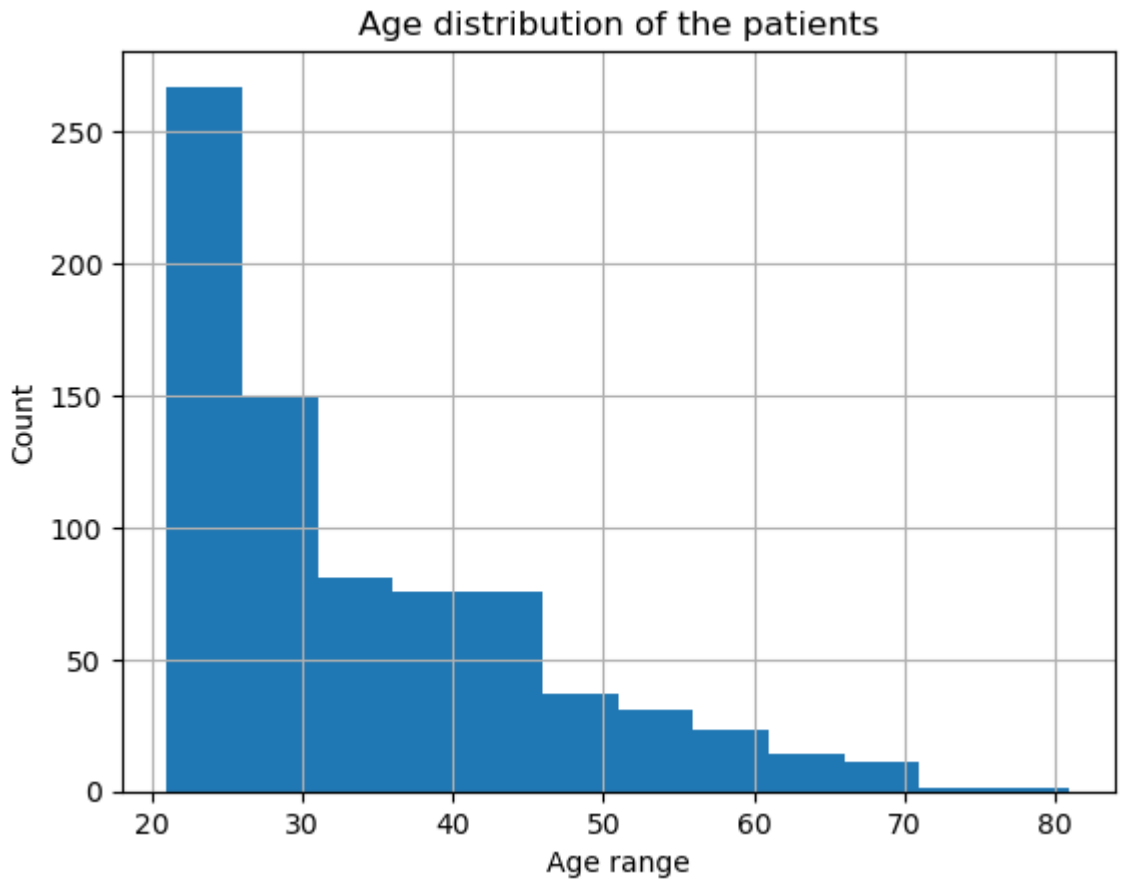
In [17]:
```
#boxplotting for BMI
plt.figure(figsize=(6,2))
plt.boxplot(df['BMI'],vert=False,showmeans=True)
plt.grid(color='gray',linestyle='dotted')
plt.xlabel("Value")
plt.ylabel("BMI")
plt.show()
```



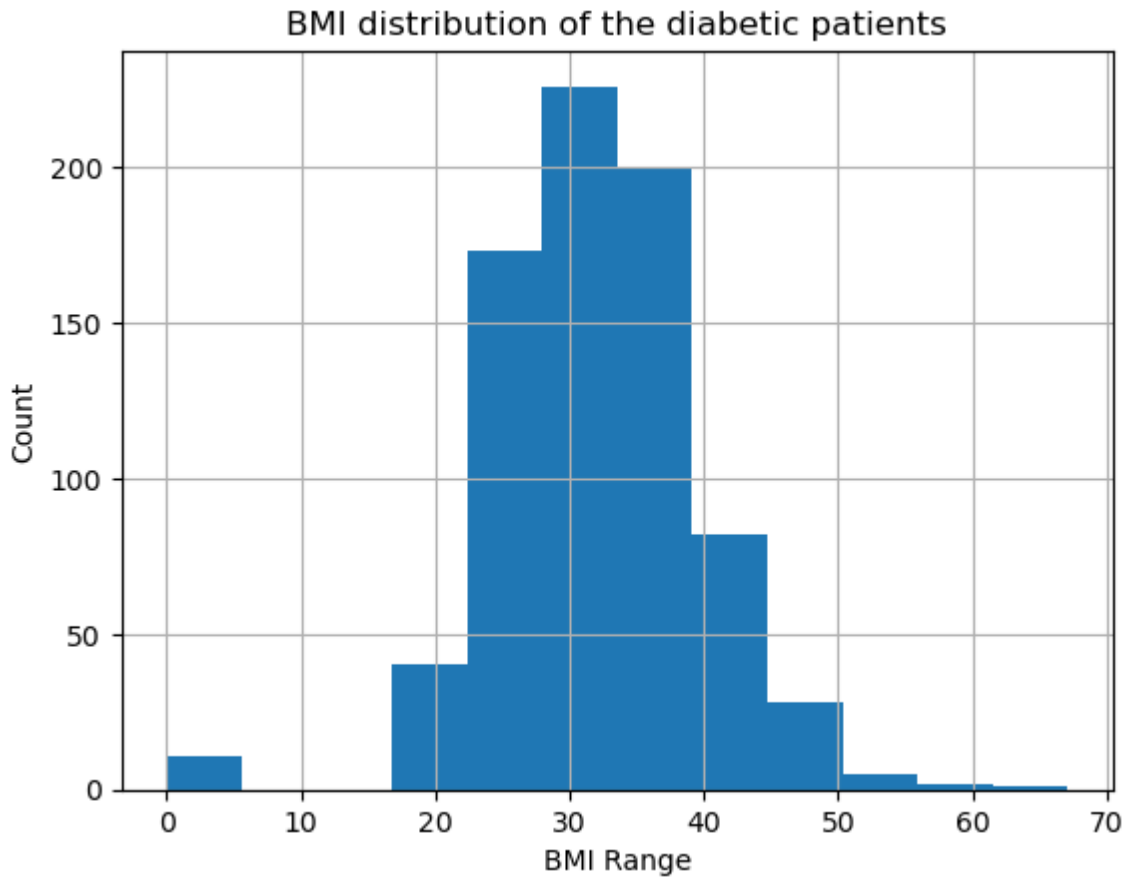
In [18]:
```
#Boxplotting for BloodPressure
plt.figure(figsize=(6,2))
plt.boxplot(df['BloodPressure'],vert=False,showmeans=True)
plt.grid(color='red',linestyle='dotted')
plt.xlabel("Value")
plt.ylabel("BloodPressure")
plt.show()
```



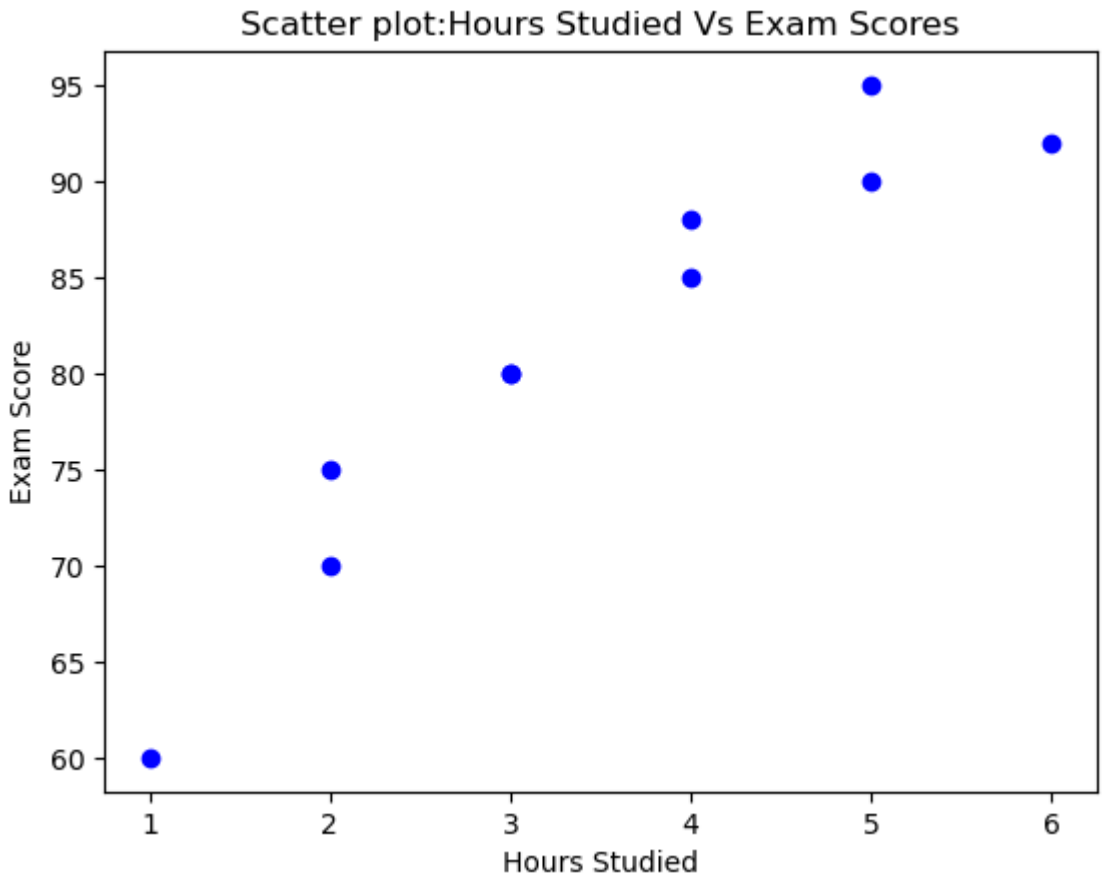
In [22]:
```
#I want to understand what age of the patients are mostly diabetic
#We can create the histograms for the same
df['Age'].hist(bins=12)
plt.title('Age distribution of the patients')
plt.xlabel('Age range')
plt.ylabel('Count')
plt.show()
```



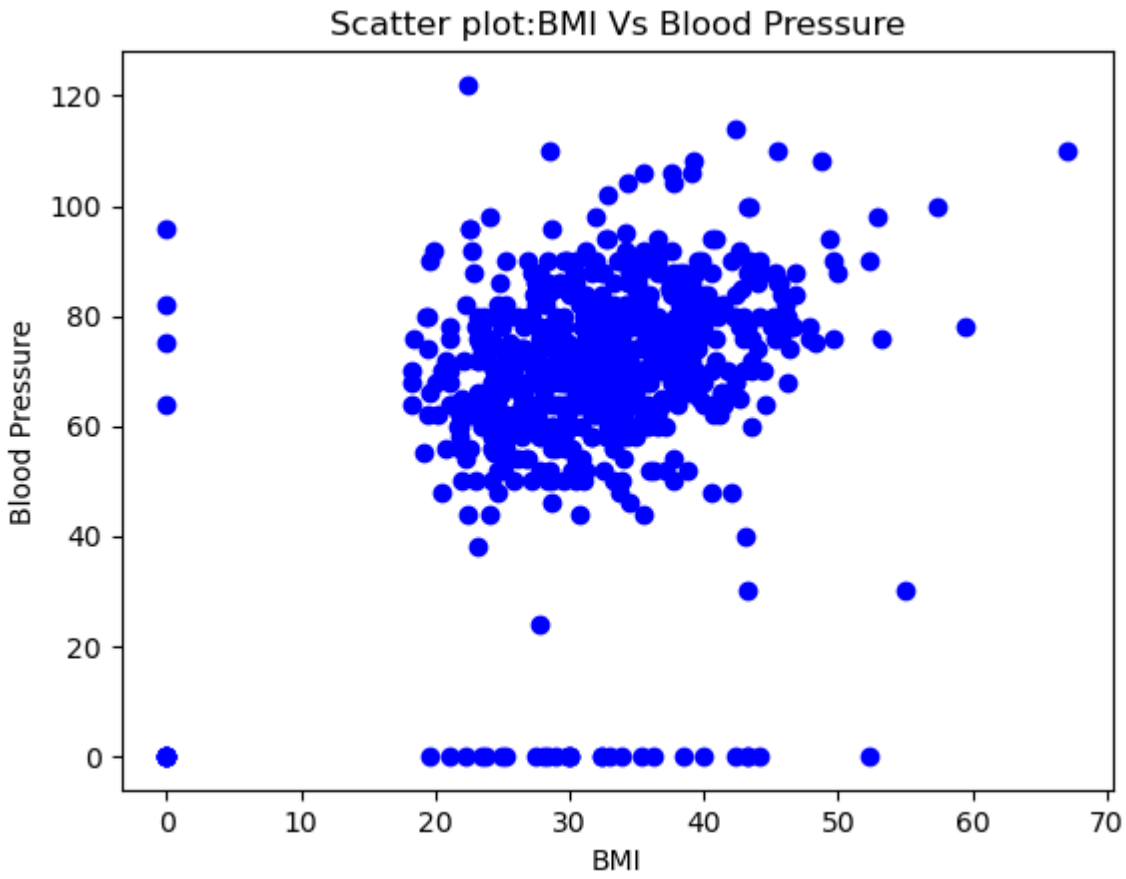
In [23]:
```
df['BMI'].hist(bins=12)
plt.title('BMI distribution of the diabetic patients')
plt.xlabel('BMI Range')
plt.ylabel('Count')
plt.show()
```



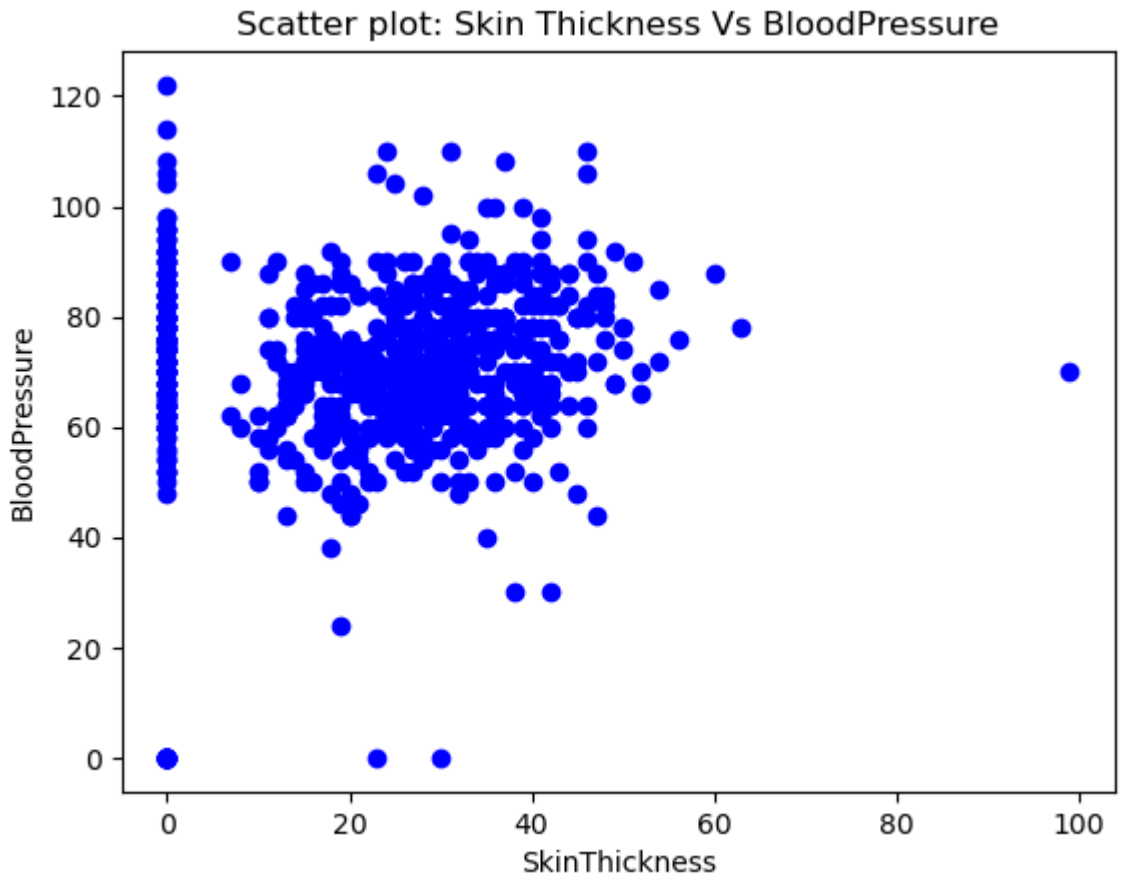
In [24]:
```
hours_studies=[2,3,1,4,2,5,4,6,3,5]
exam_scores=[70,80,60,85,75,90,88,92,80,95]

plt.scatter(hours_studies,exam_scores,color='blue',marker='o')
plt.xlabel("Hours Studied")
plt.ylabel("Exam Score")
plt.title("Scatter plot:Hours Studied Vs Exam Scores")
plt.show()
```



In [25]:
```
plt.scatter(df['BMI'],df['BloodPressure'],color='blue')
plt.xlabel("BMI")
plt.ylabel("Blood Pressure")
plt.title("Scatter plot:BMI Vs Blood Pressure")
plt.show()
```



In [27]:
```
#Inverse relation between Skin thickness and Blood Pressure
plt.scatter(df['SkinThickness'],df['BloodPressure'],color='blue')
plt.xlabel("SkinThickness")
plt.ylabel("BloodPressure")
plt.title("Scatter plot: Skin Thickness Vs BloodPressure")
plt.show()
```


Scatter plot: Skin Thickness Vs BloodPressure
BloodPressure
SkinThickness
0
20
40
60
80
100
120

In [ ]: