```
import pandas as pd
import numpy as np
import seaborn as sns

data=[1,3,28,27,25,92,30,26,30,39,40,50,26,24,29,94]

df= pd.DataFrame(data)
print(df)

print(sns.scatterplot(data=df))

        0
0       1
1       3
2      28
3      27
4      25
5      92
6      30
7      26
8      30
9      39
10     40
11     50
12     26
13     24
14     29
15     94
Axes(0.125,0.11;0.775x0.77)
```
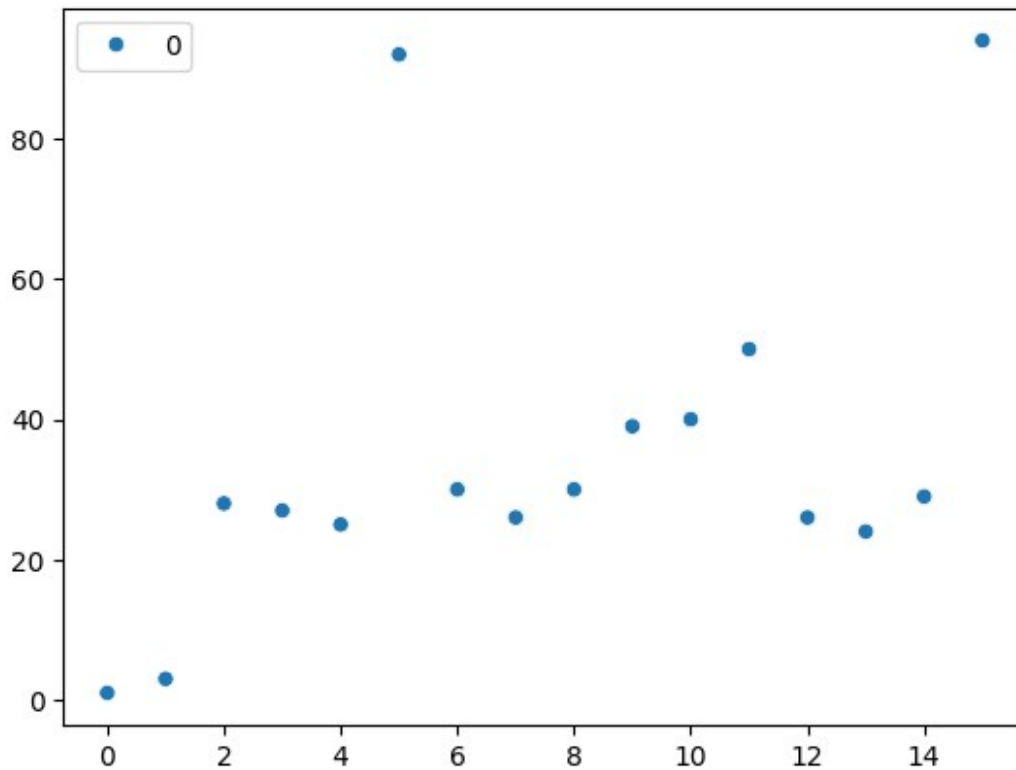
```python
import pandas as pd
import numpy as np
import seaborn as sns

data=[1,3,28,27,25,92,30,26,30,39,40,50,26,24,29,94]

df= pd.DataFrame(data)
print(df)

print(sns.boxplot(data=df))

import pandas as pd
import numpy as np
import seaborn as sns

data=[1,3,28,27,25,92,30,26,30,39,40,50,26,24,29,94]
df= pd.DataFrame(data)
print("data frame:")
print(df,"\n")
Q1=np.percentile(df,25)
Q3=np.percentile(df,75)
IQR=Q3-Q1
lower_bound=  Q1-1.5*IQR
upper_bound=  Q3+1.5*IQR
df= df[((df>=lower_bound)&(df<=upper_bound))]
df.dropna()
```

```
print("Q1:",Q1)
print("Q3:",Q3)
print("IQR:",IQR)
print("lower bound :",lower_bound)
print("upper bound", upper_bound)
print("After removing all the outliers:\n")
print(df)
print(sns.boxplot(df))
```

data frame:
```
      0
0     1
1     3
2    28
3    27
4    25
5    92
6    30
7    26
8    30
9    39
10   40
11   50
12   26
13   24
14   29
15   94
```

Q1: 25.75
Q3: 39.25
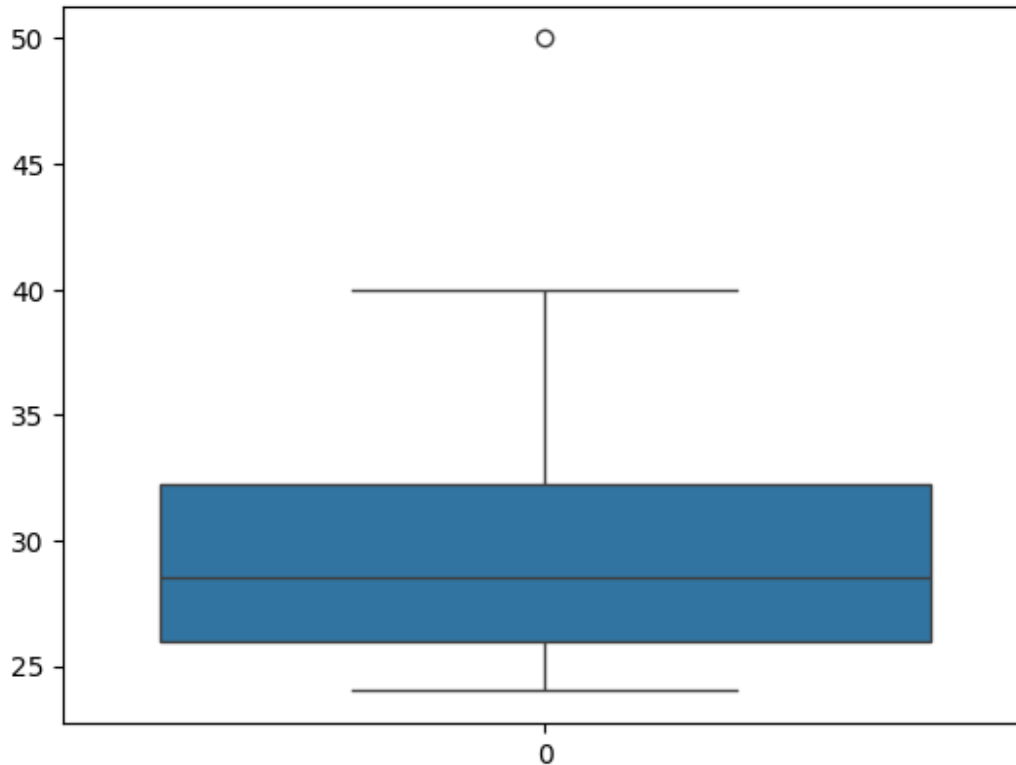IQR: 13.5
lower bound : 5.5
upper bound 59.5
After removing all the outliers:

```
        0
0     NaN
1     NaN
2    28.0
3    27.0
4    25.0
5     NaN
6    30.0
7    26.0
8    30.0
9    39.0
10   40.0
11   50.0
12   26.0
13   24.0
```

```
14   29.0
15    NaN
Axes(0.125,0.11;0.775x0.77)
```



```python
import pandas as pd
import numpy as np
import seaborn as sns

data=[1,3,28,27,25,92,30,26,30,39,40,50,26,24,29,94]
df= pd.DataFrame(data)
print("data frame:")
print(df,"\n")
Q1=np.percentile(df,25)
Q3=np.percentile(df,75)
IQR=Q3-Q1
lower_bound=  Q1-1.5*IQR
upper_bound=  Q3+1.5*IQR
df= df[((df>=lower_bound)&(df<=upper_bound))]
df.dropna()
print("Q1:",Q1)
print("Q3:",Q3)
print("IQR:",IQR)
print("lower bound :",lower_bound)
print("upper bound", upper_bound)
print("After removing all the outliers:\n")
```

```
print(df)
print(sns.scatterplot(df))

data frame:
      0
0     1
1     3
2    28
3    27
4    25
5    92
6    30
7    26
8    30
9    39
10   40
11   50
12   26
13   24
14   29
15   94

Q1: 25.75
Q3: 39.25
IQR: 13.5
lower bound : 5.5
upper bound 59.5
After removing all the outliers:

         0
0      NaN
1      NaN
2     28.0
3     27.0
4     25.0
5      NaN
6     30.0
7     26.0
8     30.0
9     39.0
10    40.0
11    50.0
12    26.0
13    24.0
14    29.0
15     NaN
Axes(0.125,0.11;0.775x0.77)
```
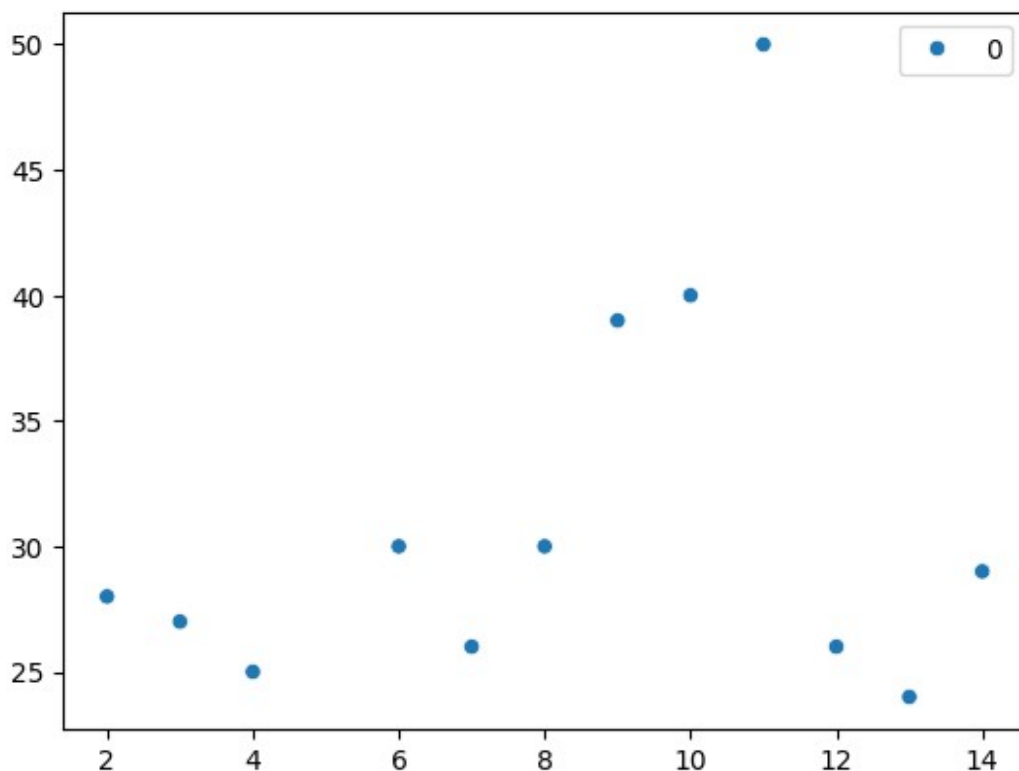
```python
import pandas as pd
import numpy as np
import seaborn as sns

df= pd.read_csv("survey lung cancer.csv")

print(df)
```

```
     GENDER  AGE  SMOKING  YELLOW_FINGERS  ANXIETY  PEER_PRESSURE  \
0         M   69        1               2        2              1
1         M   74        2               1        1              1
2         F   59        1               1        1              2
3         M   63        2               2        2              1
4         F   63        1               2        1              1
..      ...  ...      ...             ...      ...            ...
304       F   56        1               1        1              2
305       M   70        2               1        1              1
306       M   58        2               1        1              1
307       M   67        2               1        2              1
308       M   62        1               1        1              2

     CHRONIC DISEASE  FATIGUE  ALLERGY  WHEEZING  ALCOHOL CONSUMING
\
0                  1        2        1         2                  2

1                  2        2        2         1                  1
```

2	1	2	1	2	1
3	1	1	1	1	2
4	1	1	1	2	1
..	...	...	...	...	...
304	2	2	1	1	2
305	1	2	2	2	2
306	1	1	2	2	2
307	1	2	2	1	2
308	1	2	2	2	2

	COUGHING	SHORTNESS OF BREATH	SWALLOWING DIFFICULTY	CHEST PAIN \
0	2	2	2	2
1	1	2	2	2
2	2	2	1	2
3	1	1	2	2
4	2	2	1	1
..	...	...	...	...
304	2	2	2	1
305	2	2	1	2
306	2	1	1	2
307	2	2	1	2
308	1	1	2	1

	LUNG_CANCER
0	YES
1	YES
2	NO
3	NO
4	NO

```
..          ...
304         YES
305         YES
306         YES
307         YES
308         YES

[309 rows x 16 columns]

import pandas as pd
import numpy as np
import seaborn as sns

df= pd.read_csv("survey lung cancer.csv")

print(sns.boxplot(x='AGE',data=df))

Axes(0.125,0.11;0.775x0.77)
```
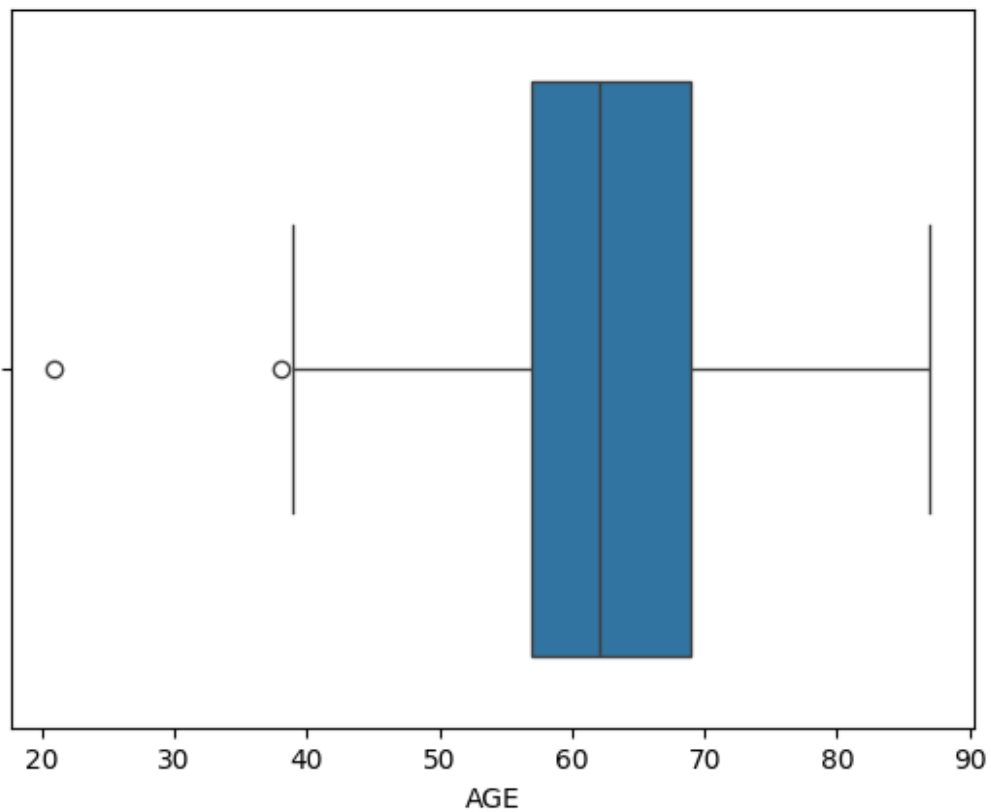


```
import pandas as pd
import numpy as np
import seaborn as sns

df= pd.read_csv("survey lung cancer.csv")
```
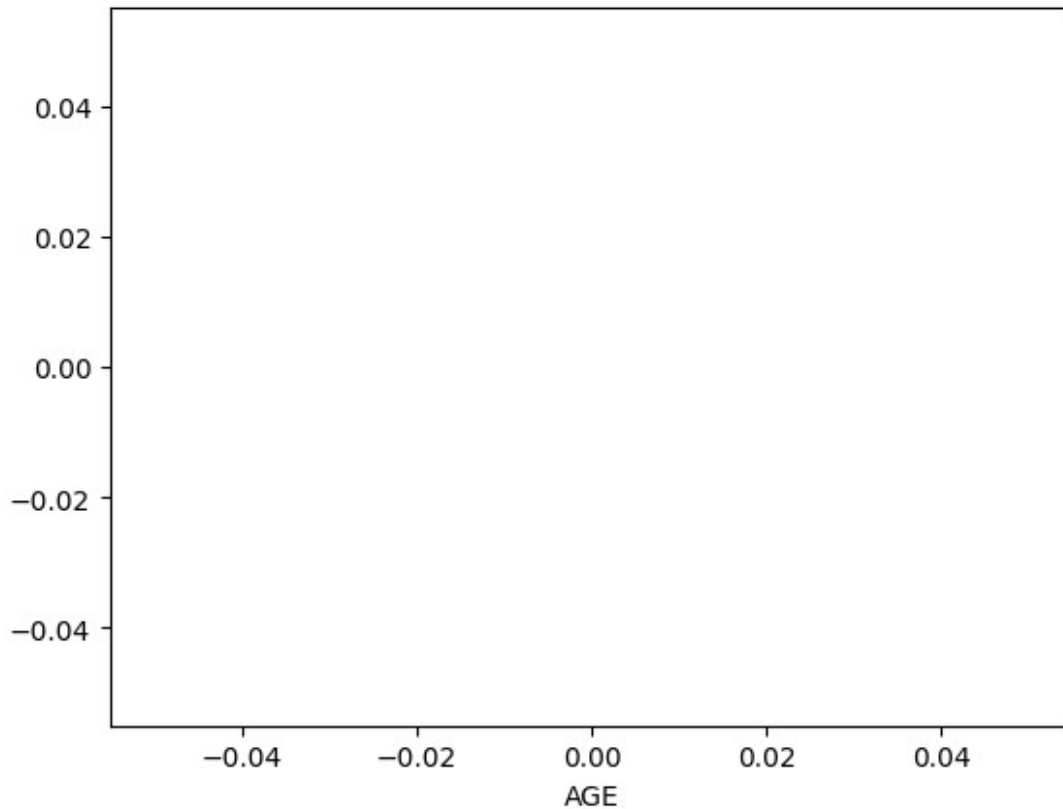
```
print(sns.scatterplot(x='AGE',data=df))
```

```
Axes(0.125,0.11;0.775x0.77)
```



```python
import pandas as pd
import numpy as np
import seaborn as sns

df= pd.read_csv("survey lung cancer.csv")
df=df['AGE']
# Calculate Q1 (25th percentile) and Q3 (75th percentile) for the
"AGE" column
Q1 = df.quantile(0.25)
Q3 = df.quantile(0.75)
IQR = Q3 - Q1

# Print the values for Q1, Q3, and IQR
print("Q1:", Q1)
print("Q3:", Q3)
print("IQR:", IQR)

# Define the lower and upper bounds for outliers
low = Q1 - 1.5 * IQR
```

```python
high = Q3 + 1.5 * IQR
print("Lower bound:", low)
print("Upper bound:", high)

# Filter the dataframe to remove outliers in the "AGE" column
df_filtered = df[(df >= low) & (df <= high)]

# Print the dataframe after removing outliers
print("After removing the outliers:\n")
print(df_filtered)
print(sns.scatterplot(df_filtered))
```

```
Q1: 57.0
Q3: 69.0
IQR: 12.0
Lower bound: 39.0
Upper bound: 87.0
After removing the outliers:

0        69
1        74
2        59
3        63
4        63
        ..
304      56
305      70
306      58
307      67
308      62
Name: AGE, Length: 307, dtype: int64
Axes(0.125,0.11;0.775x0.77)
```
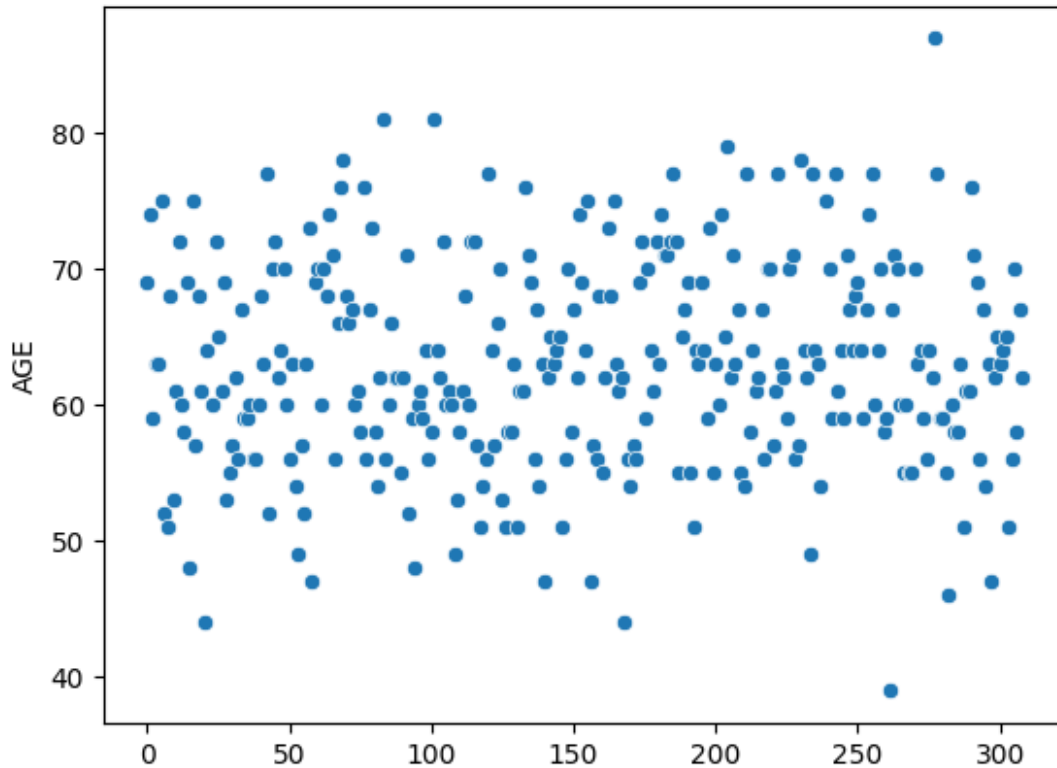
```python
import pandas as pd
import numpy as np
import seaborn as sns

df= pd.read_csv("survey lung cancer.csv")
df=df['AGE']
# Calculate Q1 (25th percentile) and Q3 (75th percentile) for the
"AGE" column
Q1 = df.quantile(0.25)
Q3 = df.quantile(0.75)
IQR = Q3 - Q1

# Print the values for Q1, Q3, and IQR
print("Q1:", Q1)
print("Q3:", Q3)
print("IQR:", IQR)

# Define the lower and upper bounds for outliers
low = Q1 - 1.5 * IQR
high = Q3 + 1.5 * IQR
print("Lower bound:", low)
print("Upper bound:", high)

# Filter the dataframe to remove outliers in the "AGE" column
df_filtered = df[(df >= low) & (df <= high)]
```

```
# Print the dataframe after removing outliers
print("After removing the outliers:\n")
print(df_filtered)
print(sns.boxplot(df_filtered))
```

```
Q1: 57.0
Q3: 69.0
IQR: 12.0
Lower bound: 39.0
Upper bound: 87.0
After removing the outliers:

0      69
1      74
2      59
3      63
4      63
       ..
304    56
305    70
306    58
307    67
308    62
Name: AGE, Length: 307, dtype: int64
Axes(0.125,0.11;0.775x0.77)
```