import Required library

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
df=pd.read_csv('/content/weight-height.csv')
```

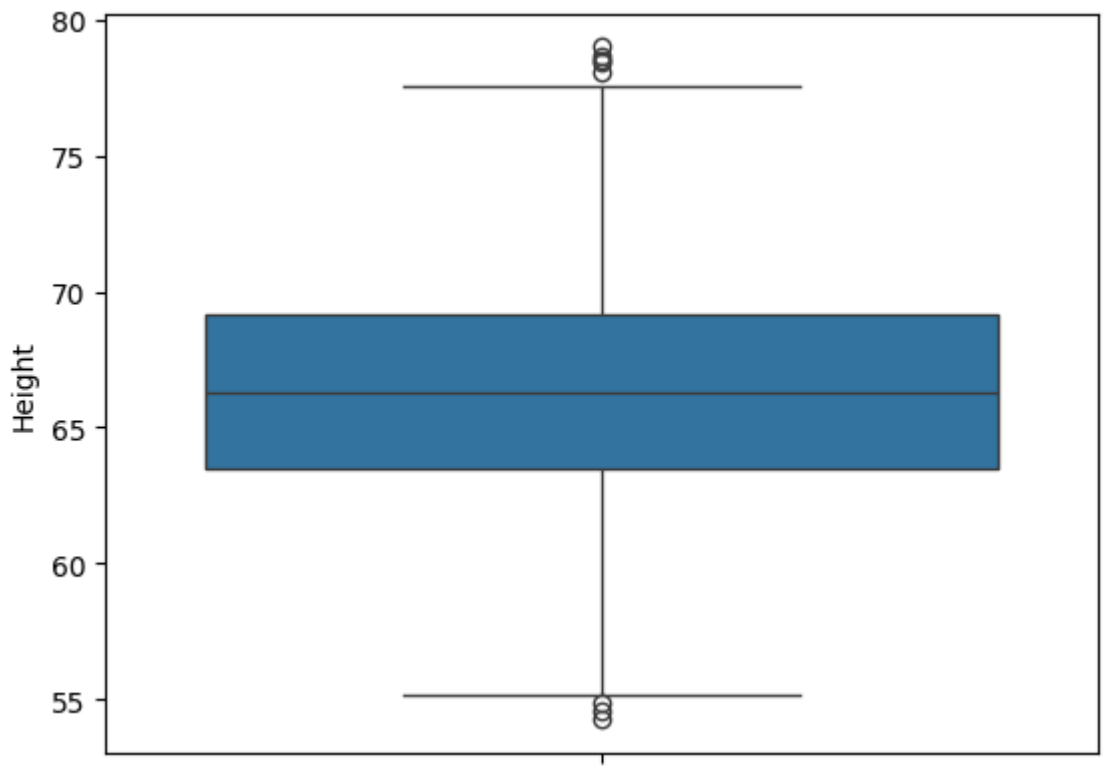
```
df.sample(5)
```

	Gender	Height	Weight
2476	Male	69.587280	194.498356
5446	Female	66.465184	160.083864
9228	Female	63.576101	122.564330
9301	Female	62.144128	130.552298
305	Male	70.185352	197.884507

detect the outlier in box-plot

```
sns.boxplot(df['Height'])
```

```
<Axes: ylabel='Height'>
```



Use percentile method for Outlier Detection

```
upperlimit=df['Height'].quantile(0.99)
lowerlimit=df['Height'].quantile(0.01)
print("UpperLimit",upperlimit)
print("LowerLimit",lowerlimit)
```

```
UpperLimit 74.7857900583366
LowerLimit 58.13441158671655
```

```
df[((df['Height']>=upperlimit) | (df['Height']<=lowerlimit))]
```

	Gender	Height	Weight
23	Male	75.205974	228.761781
190	Male	76.709835	235.035419
197	Male	75.944460	231.924749
202	Male	75.140821	224.124271
215	Male	74.795375	232.635403
...	...	...	...
9761	Female	56.975279	90.341784
9825	Female	55.979198	85.417534
9895	Female	57.740192	93.652957
9904	Female	57.028857	101.202551
9978	Female	57.375759	114.192209

200 rows × 3 columns

Remove Outlier(Trimmimg)

```
new_df=df[((df['Height']<=upperlimit) & (df['Height']>=lowerlimit))]
new_df
```

	Gender	Height	Weight
0	Male	73.847017	241.893563
1	Male	68.781904	162.310473
2	Male	74.110105	212.740856
3	Male	71.730978	220.042470
4	Male	69.881796	206.349801
...	...	...	...
9995	Female	66.172652	136.777454
9996	Female	67.067155	170.867906
9997	Female	63.867992	128.475319
9998	Female	69.034243	163.852461
9999	Female	61.944246	113.649103

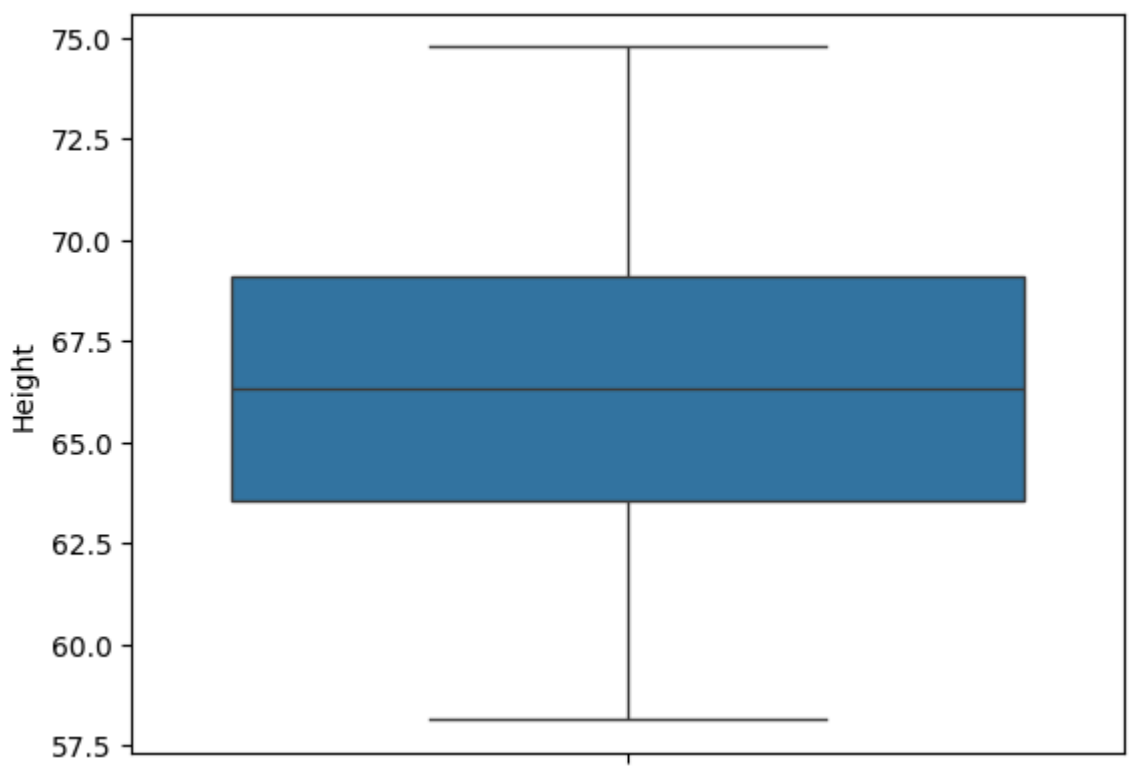
9800 rows × 3 columns

Next steps: Generate code with new_df | New interactive sheet

Box-plot After removing outlier

```
sns.boxplot(new_df['Height'])
```

```
<Axes: ylabel='Height'>
```



percentile Method(Capping/Winsorization)

```
new_df2=df.copy()
new_df2['Height']=np.where(
    new_df2['Height']>upperlimit,
    upperlimit,
    np.where(
        new_df2['Height']<lowerlimit,
        lowerlimit,
        new_df2['Height']
    )
)
```

```
sns.boxplot(new_df2['Height'])
```

```
<Axes: ylabel='Height'>
```