

# healdata assignment

May 20, 2023

```
[6]: from matplotlib import pyplot as plot
```

```
[7]: import numpy as np
```

```
[8]: import pandas as pd
```

```
[15]: document = pd.read_csv(r'C:\Users\bonin\Downloads\healdata2.csv')
```

```
[19]: average_age = np.mean(document.age)
```

```
[20]: print(average_age)
```

42.51428571428571

```
[21]: df = pd.DataFrame(document)
```

```
[23]: df.head()
```

```
[23]:
```

	age	gender	bmi	smoker	region	exercise	alcohol	cholesterol	\
0	49	Male	28.7	Yes	Regular	High	Very High	135/85	
1	67	Female	26.8	No	Occasional	Moderate	High	140/90	
2	41	Male	24.9	No	Occasional	Moderate	Normal	125/80	
3	31	Female	29.2	Yes	Regular	Low	High	130/85	
4	42	Male	23.5	No	Never	Moderate	Normal	120/80	

	systolic_bp	diastolic_bp
0	92	71
1	110	60
2	88	84
3	105	63
4	95	80

```
[43]: num_smokers = len(df[df['smoker'] == 'Yes'])
```

```
[44]: print(num_smokers)
```

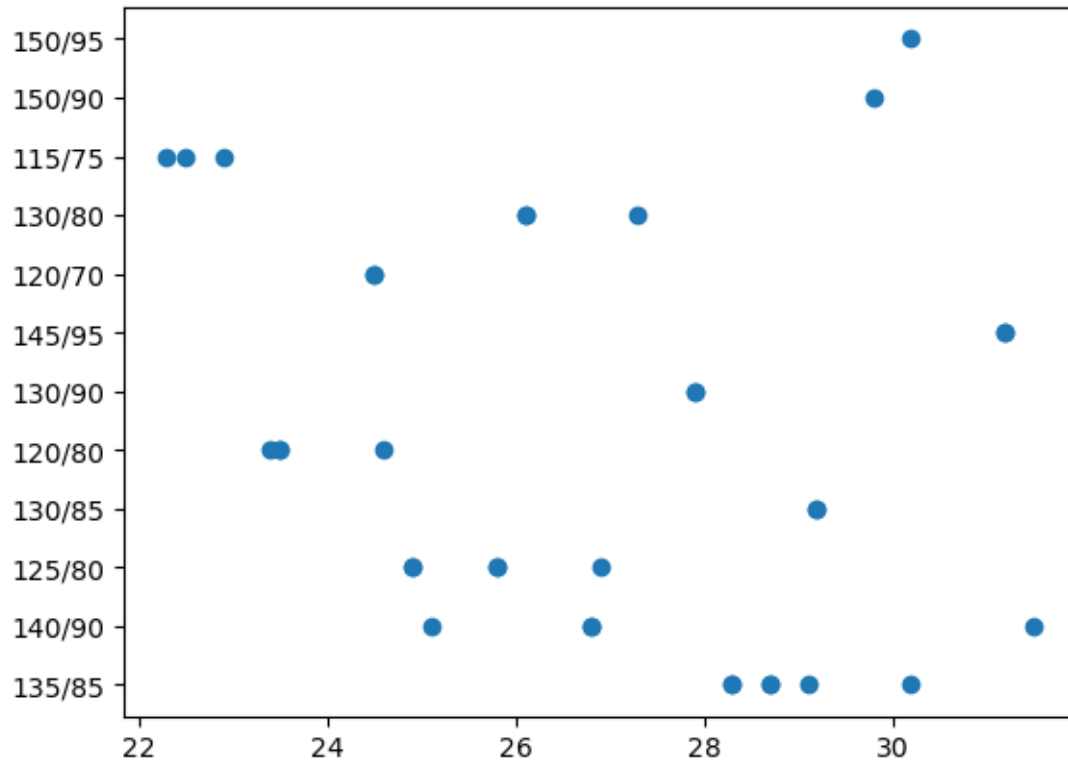
17

```
[31]: bmi = df.bmi
```

```
[32]: cholesterol = df.cholesterol
```

```
[33]: plot.scatter(bmi, cholesterol)
```

```
[33]: <matplotlib.collections.PathCollection at 0x2a3741a01c0>
```



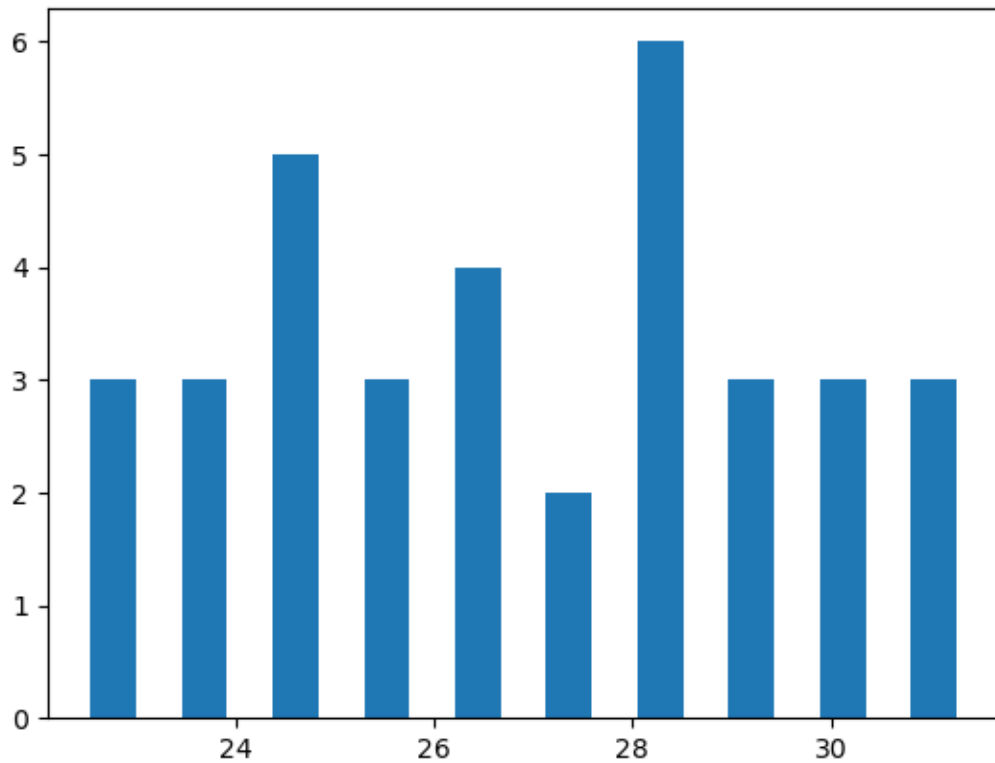
```
[34]: percentage = num_smokers / len(df.smoker) * 100
```

```
[36]: print(percentage)
```

```
48.57142857142857
```

```
[38]: plot.hist(df.bmi, rwidth=0.5)
```

```
[38]: (array([3., 3., 5., 3., 4., 2., 6., 3., 3., 3.]),  
      array([22.3 , 23.22, 24.14, 25.06, 25.98, 26.9 , 27.82, 28.74, 29.66,  
            30.58, 31.5 ]),  
      <BarContainer object of 10 artists>)
```



```
[47]: region_array = np.array(document.region)
```

```
[50]: region, count = np.unique(region_array, return_counts=True)
```

```
[51]: region[count == count.max()]
```

```
[51]: array(['Occasional', 'Regular'], dtype=object)
```

```
[52]: print(region)
```

```
['Never' 'Occasional' 'Regular']
```

```
[57]: print(count)
```

```
[ 7 14 14]
```

```
[59]: sbp_over_140 = len(df[df['systolic_bp'] > 140])
```

```
[60]: print(sbp_over_140)
```

```
0
```

```
[ ]:
```