

Heart Disease Diagnostic Analysis

Importing Required Libraries

In [11]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

Extracting CSV Dataset From System using Pandas Library

In [285...]

```
df=pd.read_csv('heart_disease_dataset.csv')
df
```

Out[285...]

	age	sex	cp	trestbps	chol	fb	restecg	thalach	exang	oldpeak	slope	ca	thal	num
0	63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
1	67	1	4	160	286	0	2	108	1	1.5	2	3	3	1
2	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
3	37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
4	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
...
298	45	1	1	110	264	0	0	132	0	1.2	2	0	7	1
299	68	1	4	144	193	1	0	141	0	3.4	2	2	7	1
300	57	1	4	130	131	0	0	115	1	1.2	2	1	7	1
301	57	0	2	130	236	0	2	174	0	0.0	2	1	3	1
302	38	1	3	138	175	0	0	173	0	0.0	1	-100000	3	0

303 rows × 14 columns

In [286...]

```
df.head()
```

Out[286...]

	age	sex	cp	trestbps	chol	fb	restecg	thalach	exang	oldpeak	slope	ca	thal	num
0	63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
1	67	1	4	160	286	0	2	108	1	1.5	2	3	3	1
2	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
3	37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
4	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0

In [287...]

```
df.tail()
```

Out[287...]

	age	sex	cp	trestbps	chol	fb	restecg	thalach	exang	oldpeak	slope	ca	thal	num
298	45	1	1	110	264	0	0	132	0	1.2	2	0	7	1
299	68	1	4	144	193	1	0	141	0	3.4	2	2	7	1
300	57	1	4	130	131	0	0	115	1	1.2	2	1	7	1
301	57	0	2	130	236	0	2	174	0	0.0	2	1	3	1
302	38	1	3	138	175	0	0	173	0	0.0	1	-100000	3	0

In [288...]

```
#Checking Not null values
df.info()
```

We can see that majority of the variables are of int64 type and are non-null.

```
RangeIndex: 303 entries, 0 to 302
Data columns (total 14 columns):
 #   Column      Non-Null Count Dtype  
 ---  --          --           --      
 0   age         303 non-null    int64  
 1   sex         303 non-null    int64  
 2   cp          303 non-null    int64  
 3   trestbps   303 non-null    int64  
 4   chol        303 non-null    int64  
 5   fbs         303 non-null    int64  
 6   restecg    303 non-null    int64  
 7   thalach    303 non-null    int64  
 8   exang       303 non-null    int64  
 9   oldpeak    303 non-null    float64 
 10  slope       303 non-null    int64  
 11  ca          303 non-null    int64  
 12  thal        303 non-null    int64
```

```
301 57 0 2 130 236 0 2 174 0 0.0 2 1 3 1
302 38 1 3 138 175 0 0 173 0 0.0 1 -100000 3 0
```

In [288...]

```
#Checking Not null values
df.info()

# We can see that majority of the variables are of int64 type and are non-null.
```

```
RangeIndex: 303 entries, 0 to 302
Data columns (total 14 columns):
 #   Column      Non-Null Count  Dtype  
 --- 
 0   age         303 non-null    int64  
 1   sex          303 non-null    int64  
 2   cp           303 non-null    int64  
 3   trestbps     303 non-null    int64  
 4   chol          303 non-null    int64  
 5   fbs           303 non-null    int64  
 6   restecg       303 non-null    int64  
 7   thalach        303 non-null    int64  
 8   exang          303 non-null    int64  
 9   oldpeak        303 non-null    float64 
 10  slope          303 non-null    int64  
 11  ca             303 non-null    int64  
 12  thal            303 non-null    int64  
 13  num            303 non-null    int64  
dtypes: float64(1), int64(13)
memory usage: 33.3 KB
```

There are 14 features in Dataset

1. age: The person's age in years
2. sex: The person's sex (1 = male, 0 = female)
3. cp: The chest pain experienced (Value 1: typical angina, Value 2: atypical angina, Value 3: non-anginal pain, Value 4: asymptomatic)
4. trestbps: The person's resting blood pressure (mm Hg on admission to the hospital)
5. chol: The person's cholesterol measurement in mg/dl
6. fbs: The person's fasting blood sugar (> 120 mg/dl, 1 = true; 0 = false)
7. restecg: Resting electrocardiographic measurement (0 = normal, 1 = having ST-T wave abnormality, 2 = showing probable or definite left ventricular hypertrophy by Estes' criteria)
8. thalach: The person's maximum heart rate achieved
9. exang: Exercise induced angina (1 = yes; 0 = no)
10. oldpeak: ST depression induced by exercise relative to rest
- 11)slope: the slope of the peak exercise ST segment (Value 1: upsloping, Value 2: flat, Value 3: downsloping)
- 12)ca: The number of major vessels (0-3)
- 13)thal: A blood disorder called thalassemia (3 = normal; 6 = fixed defect; 7 = reversable defect)
- 14)num: Heart disease (0 = no, 1 = yes)

In [289...]

```
# On closer analysis of the dataset it is visible that there are some attributes which are classified
# but they are categorical variables having a specific number of classes. Eg. sex, cp, slope etc.
```

In [290...]

```
df.shape
```

```
# The dataset contains 303 records and 14 different attributes / variables
```

Out[290... (303, 14)

In [291...]

```
df.describe()
```

```
# The describe() function gives the statistical summary of the numerical columns of the dataset.
```

Out[291...]

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak
count	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000
mean	54.438944	0.679868	3.158416	131.689769	246.693069	0.148515	0.990099	149.607261	0.326733	1.039604
std	9.038662	0.467299	0.960126	17.599748	51.776918	0.356198	0.994971	22.875003	0.469794	1.161075
min	29.000000	0.000000	1.000000	94.000000	126.000000	0.000000	0.000000	71.000000	0.000000	0.000000

In [291...]

```
df.describe()

# The describe() function gives the statistical summary of the numerical columns of the dataset.
```

Out[291...]

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak
count	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000
mean	54.438944	0.679868	3.158416	131.689769	246.693069	0.148515	0.990099	149.607261	0.326733	1.039604
std	9.038662	0.467299	0.960126	17.599748	51.776918	0.356198	0.994971	22.875003	0.469794	1.161075
min	29.000000	0.000000	1.000000	94.000000	126.000000	0.000000	0.000000	71.000000	0.000000	0.000000
25%	48.000000	0.000000	3.000000	120.000000	211.000000	0.000000	0.000000	133.500000	0.000000	0.000000
50%	56.000000	1.000000	3.000000	130.000000	241.000000	0.000000	1.000000	153.000000	0.000000	0.800000
75%	61.000000	1.000000	4.000000	140.000000	275.000000	0.000000	2.000000	166.000000	1.000000	1.600000
max	77.000000	1.000000	4.000000	200.000000	564.000000	1.000000	2.000000	202.000000	1.000000	6.200000

Percentage of people having Heart Disease

In [292...]

```
num=df.groupby('num').size()
num
```

Out[292...]

```
num
0    164
1    139
dtype: int64
```

In [293...]

```
def heart_d(r):
    if r==0:
        return 'Absence'
    elif r==1:
        return 'Presence'
```

In [294...]

```
#Applying converted data into our dataset with new column - Heart_Disease
df['Heart_Disease']=df['num'].apply(heart_d)
df.head()
```

Out[294...]

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num	Heart_Disease
0	63	1	1	145	233	1	2	150	0	2.3	3	0	6	0	Absence
1	67	1	4	160	286	0	2	108	1	1.5	2	3	3	1	Presence
2	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1	Presence
3	37	1	3	130	250	0	0	187	0	3.5	3	0	3	0	Absence
4	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0	Absence

In [295...]

```
hd=df.groupby('Heart_Disease')['num'].count()
hd
```

Out[295...]

```
Heart_Disease
Absence    164
Presence   139
Name: num, dtype: int64
```

In [296...]

```
#Converting Numerical Data into Categorical Data
```

```
def gen(r):
    if r==1:
        return 'Male'
    elif r==0:
        return 'Female'
```

In [297...]

```
#Applying converted data into our dataset with new column - sex1
df['sex1']=df['sex'].apply(gen)
df.head()
```

Out[297...]

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num	Heart_Disease	sex1
0	63	1	1	145	233	1	2	150	0	2.3	3	0	6	0	Absence	Male
1	67	1	4	160	286	0	2	108	1	1.5	2	3	3	1	Presence	Male
2	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1	Presence	Male
3	37	1	3	130	250	0	0	187	0	3.5	3	0	3	0	Absence	Male
4	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0	Absence	Female

In [297...]

```
#Applying converted data into our dataset with new column - sex1
```

```
df['sex1']=df['sex'].apply(gen)
df.head()
```

Out[297...]

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num	Heart_Disease	sex1
0	63	1	1	145	233	1	2	150	0	2.3	3	0	6	0	Absence	Male
1	67	1	4	160	286	0	2	108	1	1.5	2	3	3	1	Presence	Male
2	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1	Presence	Male
3	37	1	3	130	250	0	0	187	0	3.5	3	0	3	0	Absence	Male
4	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0	Absence	Female

In [298...]

```
#Converting Numerical Data into Categorical Data
```

```
def age_rng(r):
    if r>=29 and r<40:
        return 'Young Age'
    elif r>=40 and r<55:
        return 'Middle Age'
    elif r>55:
        return 'Elder Age'
```

In [299...]

```
#Applying converted data into our dataset with new column - Age_Range
```

```
df['Age_Range']=df['age'].apply(age_rng)
df.head()
```

Out[299...]

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num	Heart_Disease	sex1	Age_Range
0	63	1	1	145	233	1	2	150	0	2.3	3	0	6	0	Absence	Male	Elder Age
1	67	1	4	160	286	0	2	108	1	1.5	2	3	3	1	Presence	Male	Elder Age
2	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1	Presence	Male	Elder Age
3	37	1	3	130	250	0	0	187	0	3.5	3	0	3	0	Absence	Male	Young Age
4	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0	Absence	Female	Middle Age

Exploratory Data Analysis

In [300...]

```
|pip install pandas-profiling
```

```
Requirement already satisfied: pandas-profiling in c:\users\gigabyte\anaconda3\lib\site-packages (3.1.0)
Requirement already satisfied: tqdm>=4.48.2 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (4.59.0)
Requirement already satisfied: pandas!=1.0.0,!>1.0.1,!>1.0.2,!>1.1.0,>=0.25.3 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (1.2.4)
Requirement already satisfied: htmlmin>=0.1.12 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (0.1.12)
Requirement already satisfied: pydantic>=1.8.1 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (1.9.0)
Requirement already satisfied: visions[type_image_path]==0.7.4 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (0.7.4)
Requirement already satisfied: tangled-up-in-unicode==0.1.0 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (0.1.0)
Requirement already satisfied: requests>=2.24.0 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (2.27.1)
Requirement already satisfied: multimethod>=1.4 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (1.6)
Requirement already satisfied: scipy>=1.4.1 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (1.6.2)
Requirement already satisfied: jinja2>=2.11.1 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (3.0.3)
Requirement already satisfied: markupsafe~=2.0.1 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (2.0.1)
Requirement already satisfied: joblib~=1.0.1 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (1.0.1)
Requirement already satisfied: PyYAML>=5.0.0 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (5.4.1)
Requirement already satisfied: missingno>=0.4.2 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (0.5.0)
Requirement already satisfied: seaborn>=0.10.1 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (0.11.1)
Requirement already satisfied: numpy>=1.16.0 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (1.22.1)
Requirement already satisfied: phik>=0.11.1 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (0.12.0)
Requirement already satisfied: matplotlib>=3.2.0 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (3.3.4)
Requirement already satisfied: attrs>=19.3.0 in c:\users\gigabyte\anaconda3\lib\site-packages (from visions[type_image_path]==0.7.4->pandas-profiling) (20.3.0)
Requirement already satisfied: networkx>=2.4 in c:\users\gigabyte\anaconda3\lib\site-packages (from visions[type_image_path]==0.7.4->pandas-profiling) (2.5)
Requirement already satisfied: Pillow in c:\users\gigabyte\anaconda3\lib\site-packages (from visions[type_image_path]==0.7.4->pandas-profiling) (6.2.1)
Requirement already satisfied: imagehash in c:\users\gigabyte\anaconda3\lib\site-packages (from visions[type_image_path]==0.7.4->pandas-profiling) (4.2.1)
Requirement already satisfied: pyparsing!=2.0.4,!>2.1.2,!>2.1.6,>=2.0.3 in c:\users\gigabyte\anaconda3\lib\site-packages (from matplotlib>=3.2.0->pandas-profiling) (2.4.7)
Requirement already satisfied: python-dateutil>=2.1 in c:\users\gigabyte\anaconda3\lib\site-packages (from matplotlib>=3.2.0->pandas-profiling) (2.8.1)
Requirement already satisfied: cycler>=0.10 in c:\users\gigabyte\anaconda3\lib\site-packages (from matplotlib>=3.2.0->pandas-profiling) (0.10.0)
```

Exploratory Data Analysis

In [300...]

```
| pip install pandas-profiling
```

```
Requirement already satisfied: pandas-profiling in c:\users\gigabyte\anaconda3\lib\site-packages (3.1.0)
Requirement already satisfied: tqdm>=4.48.2 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (4.59.0)
Requirement already satisfied: pandas!=1.0.0,!>=1.0.1,!>=1.0.2,!>=1.1.0,>=0.25.3 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (1.2.4)
Requirement already satisfied: htmlmin>=0.1.12 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (0.1.12)
Requirement already satisfied: pydantic>=1.8.1 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (1.9.0)
Requirement already satisfied: visions[type_image_path]==0.7.4 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (0.7.4)
Requirement already satisfied: tangled-up-in-unicode==0.1.0 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (0.1.0)
Requirement already satisfied: requests>=2.24.0 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (2.27.1)
Requirement already satisfied: multimethod>=1.4 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (1.6)
Requirement already satisfied: scipy>=1.4.1 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (1.6.2)
Requirement already satisfied: jinja2>=2.11.1 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (3.0.3)
Requirement already satisfied: markupsafe~=2.0.1 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (2.0.1)
Requirement already satisfied: joblib~=1.0.1 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (1.0.1)
Requirement already satisfied: PyYAML>=5.0.0 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (5.4.1)
Requirement already satisfied: missingno>=0.4.2 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (0.5.0)
Requirement already satisfied: seaborn>=0.10.1 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (0.11.1)
Requirement already satisfied: numpy>=1.16.0 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (1.22.1)
Requirement already satisfied: phik>=0.11.1 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (0.12.0)
Requirement already satisfied: matplotlib>=3.2.0 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas-profiling) (3.3.4)
Requirement already satisfied: attrs>=19.3.0 in c:\users\gigabyte\anaconda3\lib\site-packages (from visions[type_image_path]==0.7.4->pandas-profiling) (20.3.0)
Requirement already satisfied: networkx>=2.4 in c:\users\gigabyte\anaconda3\lib\site-packages (from visions[type_image_path]==0.7.4->pandas-profiling) (2.5)
Requirement already satisfied: Pillow in c:\users\gigabyte\anaconda3\lib\site-packages (from visions[type_image_path]==0.7.4->pandas-profiling) (6.2.1)
Requirement already satisfied: imagehash in c:\users\gigabyte\anaconda3\lib\site-packages (from visions[type_image_path]==0.7.4->pandas-profiling) (4.2.1)
Requirement already satisfied: pyparsing!=2.0.4,!>=2.1.2,!>=2.1.6,>=2.0.3 in c:\users\gigabyte\anaconda3\lib\site-packages (from matplotlib>=3.2.0->pandas-profiling) (2.4.7)
Requirement already satisfied: python-dateutil>=2.1 in c:\users\gigabyte\anaconda3\lib\site-packages (from matplotlib>=3.2.0->pandas-profiling) (2.8.1)
Requirement already satisfied: cycler>=0.10 in c:\users\gigabyte\anaconda3\lib\site-packages (from matplotlib>=3.2.0->pandas-profiling) (0.10.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\gigabyte\anaconda3\lib\site-packages (from matplotlib>=3.2.0->pandas-profiling) (1.3.1)
Requirement already satisfied: six in c:\users\gigabyte\anaconda3\lib\site-packages (from cycler>=0.10->matplotlib>=3.2.0->pandas-profiling) (1.13.0)
Requirement already satisfied: decorator>=4.3.0 in c:\users\gigabyte\anaconda3\lib\site-packages (from networkx>=2.4->visions[type_image_path]==0.7.4->pandas-profiling) (5.0.6)
Requirement already satisfied: pytz>=2017.3 in c:\users\gigabyte\anaconda3\lib\site-packages (from pandas!=1.0.0,!>=1.0.1,!>=1.0.2,!>=1.1.0,>=0.25.3->pandas-profiling) (2021.1)
Requirement already satisfied: typing-extensions>=3.7.4.3 in c:\users\gigabyte\anaconda3\lib\site-packages (from pydantic>=1.8.1->pandas-profiling) (3.7.4.3)
Requirement already satisfied: idna<4,>=2.5 in c:\users\gigabyte\anaconda3\lib\site-packages (from requests>=2.24.0->pandas-profiling) (2.8)
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\users\gigabyte\anaconda3\lib\site-packages (from requests>=2.24.0->pandas-profiling) (2.0.9)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\gigabyte\anaconda3\lib\site-packages (from requests>=2.24.0->pandas-profiling) (1.24.3)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\gigabyte\anaconda3\lib\site-packages (from requests>=2.24.0->pandas-profiling) (2019.9.11)
Requirement already satisfied: PyWavelets in c:\users\gigabyte\anaconda3\lib\site-packages (from imagehash->visions[type_image_path]==0.7.4->pandas-profiling) (1.1.1)
```

In [301...]

```
from pandas_profiling import ProfileReport
```

In [302...]

```
prof=ProfileReport(df,title="Heart_Dataset_Profile_Report_Before_Cleanup.html", html={'style':{'fullwidth': True}}
```

In [303...]

```
prof
```

```
Summarize dataset: 0% | 0/5 [00:00, ?it/s]
Generate report structure: 0% | 0/1 [00:00, ?it/s]
Render HTML: 0% | 0/1 [00:00, ?it/s]
Heart_Dataset_Profile_Report_Before_Cleanup.html
Heart_Dataset_Profile_Report_Before_Cleanup.html
```

- Overview
- Variables
- Interactions
- Correlations
- Missing values
- Sample

In [301...]

```
from pandas_profiling import ProfileReport
```

In [302...]

```
prof=ProfileReport(df,title="Heart_Dataset_Profile_Report_Before_Cleanup.html", html={'style':{'fullwidth':True}})
```

In [303...]

```
prof
```

```
Summarize dataset: 0% | 0/5 [00:00, ?it/s]
Generate report structure: 0% | 0/1 [00:00, ?it/s]
Render HTML: 0% | 0/1 [00:00, ?it/s]
Heart_Dataset_Profile_Report_Before_Cleanup.html
Heart_Dataset_Profile_Report_Before_Cleanup.html
```

- Overview
- Variables
- Interactions
- Correlations
- Missing values
- Sample

Overview

- Overview
- Alerts 29
- Reproduction

Dataset statistics

Number of variables	17
Number of observations	303
Missing cells	8
Missing cells (%)	0.2%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	40.4 KiB
Average record size in memory	136.4 B

Variable types

Numeric	5
Categorical	12

Alerts

oldpeak	is highly correlated with	slope	High correlation	
slope	is highly correlated with	oldpeak	High correlation	
thal	is highly correlated with	num	High correlation	
num	is highly correlated with	thal	High correlation	
oldpeak	is highly correlated with	slope	High correlation	
slope	is highly correlated with	oldpeak	High correlation	
oldpeak	is highly correlated with	slope	High correlation	
slope	is highly correlated with	oldpeak	High correlation	
thal	is highly correlated with	Heart_Disease	and 1 other fields	High correlation
Heart_Disease	is highly correlated with	thal	and 2 other fields	High correlation
cp	is highly correlated with	Heart_Disease	and 1 other fields	High correlation
sex	is highly correlated with	sex1	High correlation	
num	is highly correlated with	thal	and 2 other fields	High correlation
sex1	is highly correlated with	sex	High correlation	
age	is highly correlated with	thalach	and 1 other fields	High correlation
sex	is highly correlated with	thal	and 1 other fields	High correlation
cp	is highly correlated with	exang	and 2 other fields	High correlation
restecg	is highly correlated with	oldpeak	High correlation	
thalach	is highly correlated with	age	and 3 other fields	High correlation
exang	is highly correlated with	cp	and 3 other fields	High correlation
oldpeak	is highly correlated with	restecg	and 1 other fields	High correlation
slope	is highly correlated with	oldpeak	High correlation	
thal	is highly correlated with	sex	and 3 other fields	High correlation
num	is highly correlated with	cp	and 4 other fields	High correlation
Heart_Disease	is highly correlated with	cp	and 4 other fields	High correlation
sex1	is highly correlated with	sex	and 1 other fields	High correlation
Age_Range	is highly correlated with	age	High correlation	
Age_Range	has 8 (2.6%) missing values		Missing	
oldpeak	has 99 (32.7%) zeros		Zeros	

oldpeak	is highly correlated with	slope	High correlation	
slope	is highly correlated with	oldpeak	High correlation	
thal	is highly correlated with	num	High correlation	
num	is highly correlated with	thal	High correlation	
oldpeak	is highly correlated with	slope	High correlation	
slope	is highly correlated with	oldpeak	High correlation	
oldpeak	is highly correlated with	slope	High correlation	
slope	is highly correlated with	oldpeak	High correlation	
thal	is highly correlated with	Heart_Disease	and 1 other fields	High correlation
Heart_Disease	is highly correlated with	thal	and 2 other fields	High correlation
cp	is highly correlated with	Heart_Disease	and 1 other fields	High correlation
sex	is highly correlated with	sex1	High correlation	
num	is highly correlated with	thal	and 2 other fields	High correlation
sex1	is highly correlated with	sex	High correlation	
age	is highly correlated with	thalach	and 1 other fields	High correlation
sex	is highly correlated with	thal	and 1 other fields	High correlation
cp	is highly correlated with	exang	and 2 other fields	High correlation
restecg	is highly correlated with	oldpeak	High correlation	
thalach	is highly correlated with	age	and 3 other fields	High correlation
exang	is highly correlated with	cp	and 3 other fields	High correlation
oldpeak	is highly correlated with	restecg	and 1 other fields	High correlation
slope	is highly correlated with	oldpeak	High correlation	
thal	is highly correlated with	sex	and 3 other fields	High correlation
num	is highly correlated with	cp	and 4 other fields	High correlation
Heart_Disease	is highly correlated with	cp	and 4 other fields	High correlation
sex1	is highly correlated with	sex	and 1 other fields	High correlation
Age_Range	is highly correlated with	age	High correlation	
Age_Range	has 8 (2.6%) missing values		Missing	
oldpeak	has 99 (32.7%) zeros		Zeros	

Reproduction

Analysis started	2022-01-24 10:15:51.588304
Analysis finished	2022-01-24 10:16:00.326195
Duration	8.74 seconds
Software version	pandas-profiling v3.1.0
Download configuration	config.json

Variables

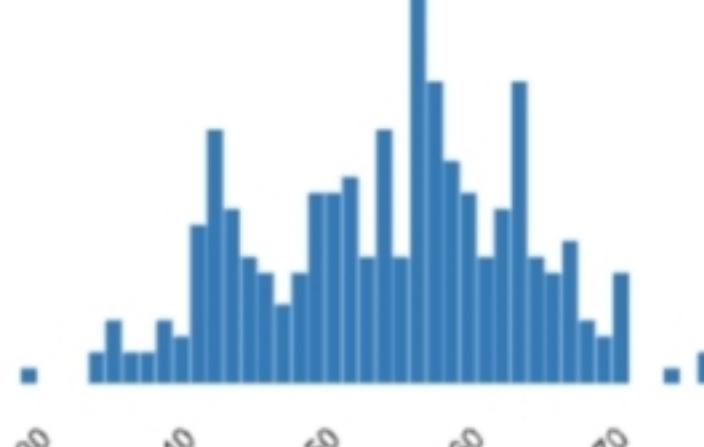
age

Real number (ℝ ≥ 0)

HIGH CORRELATION

Distinct	41
Distinct (%)	13.5%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	54.43894389

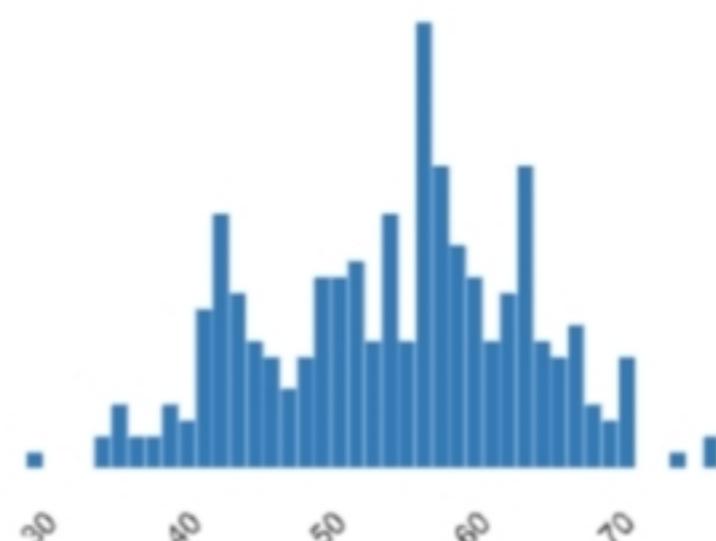
Minimum	29
Maximum	77
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	2.5 Kib



- Statistics
- Histogram
- Common values
- Extreme values

Quantile statistics

Minimum	29
---------	----



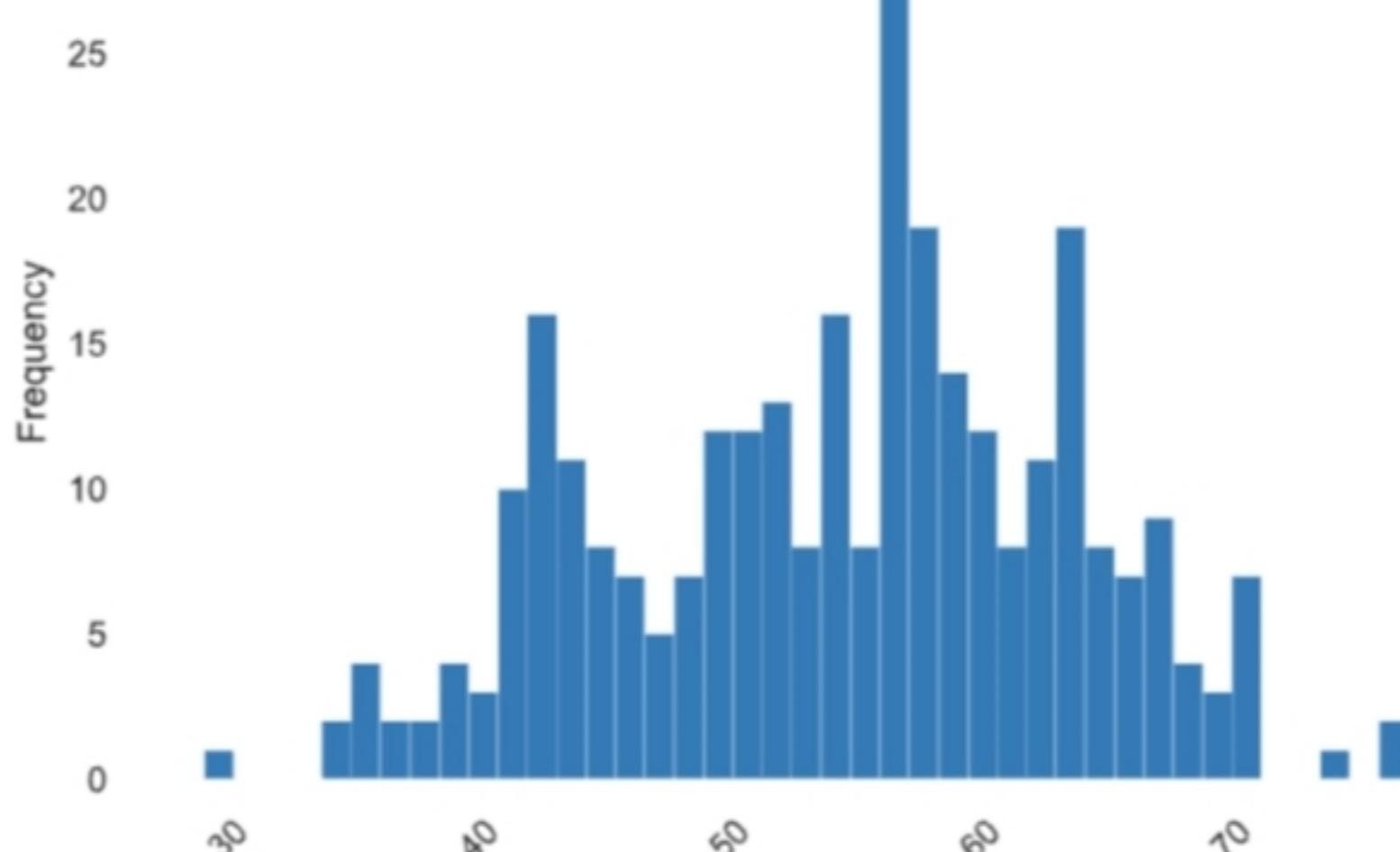
- Statistics
- Histogram
- Common values
- Extreme values

Quantile statistics

Minimum	29
5-th percentile	40
Q1	48
median	56
Q3	61
95-th percentile	68
Maximum	77
Range	48
Interquartile range (IQR)	13

Descriptive statistics

Standard deviation	9.038662442
Coefficient of variation (CV)	0.1660330233
Kurtosis	-0.5233827452
Mean	54.43894389
Median Absolute Deviation (MAD)	6
Skewness	-0.2090604688
Sum	16495
Variance	81.69741875
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=41)

Value	Count	Frequency (%)
58	19	6.3%
57	17	5.6%
54	16	5.3%
59	14	4.6%
52	13	4.3%
51	12	4.0%
60	12	4.0%
62	11	3.6%
44	11	3.6%
56	11	3.6%
Other values (31)	167	55.1%

- Minimum 5 values
- Maximum 5 values

Value	Count	Frequency (%)
29	1	0.3%
34	2	0.7%

Histogram with fixed size bins (bins=41)

Value	Count	Frequency (%)
58	19	6.3%
57	17	5.6%
54	16	5.3%
59	14	4.6%
52	13	4.3%
51	12	4.0%
60	12	4.0%
62	11	3.6%
44	11	3.6%
56	11	3.6%
Other values (31)	167	55.1%

- Minimum 5 values
- Maximum 5 values

Value Count Frequency (%)

29	1	0.3%
34	2	0.7%
35	4	1.3%
37	2	0.7%
38	2	0.7%
39	4	1.3%
40	3	1.0%
41	10	3.3%
42	8	2.6%
43	8	2.6%

Value Count Frequency (%)

77	1	0.3%
76	1	0.3%
74	1	0.3%
71	3	1.0%
70	4	1.3%
69	3	1.0%
68	4	1.3%
67	9	3.0%
66	7	2.3%
65	8	2.6%

sex

Categorical

HIGH CORRELATION

HIGH CORRELATION

Distinct 2

Distinct (%) 0.7%

Missing 0

Missing (%) 0.0%

Memory size 2.5 KiB

1 206

0 97

- Overview
- Categories
- Words
- Characters

Length

Max length 1

Median length 1

Mean length 1

Min length 1

Characters and Unicode

Total characters 0

sex

Categorical

HIGH CORRELATION

HIGH CORRELATION

Distinct 2

Distinct (%) 0.7%

Missing 0

Missing (%) 0.0%

Memory size 2.5 KIB

1 206

0 97

- Overview
- Categories
- Words
- Characters

Length

Max length 1

Median length 1

Mean length 1

Min length 1

Characters and Unicode

Total characters 0

Distinct characters 0

Distinct categories 0

Distinct scripts 0

Distinct blocks 0

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique 0

Unique (%) 0.0%

Sample

1st row 1

2nd row 1

3rd row 1

4th row 1

5th row 0

Common Values

Value	Count	Frequency (%)
-------	-------	---------------

1	206	68.0%
---	-----	-------

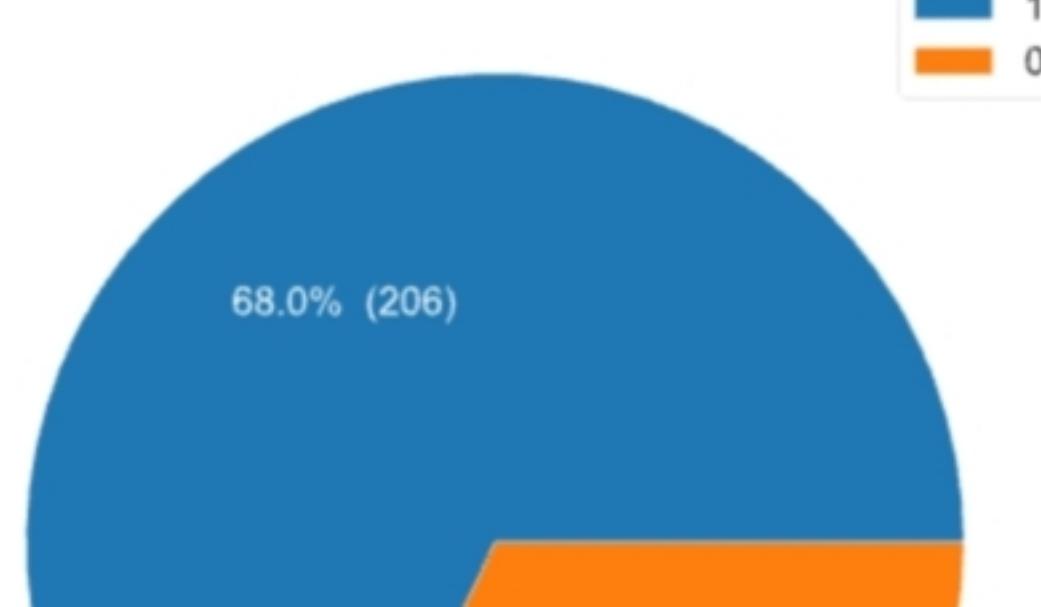
0	97	32.0%
---	----	-------

Length



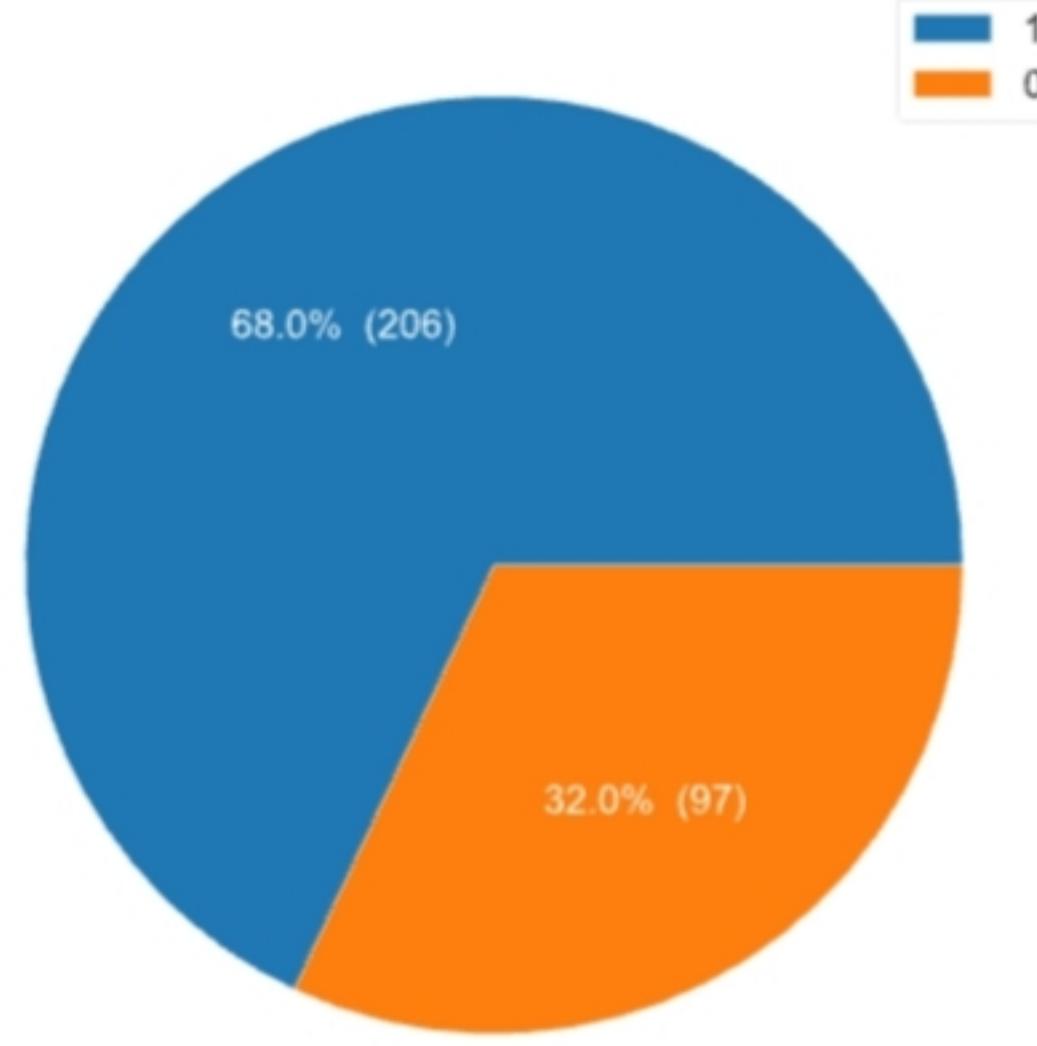
Histogram of lengths of the category

Pie chart



Histogram of lengths of the category

Pie chart



Value Count Frequency (%)

1	206	68.0%
0	97	32.0%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value Count Frequency (%)

No values found.

Most occurring categories

Value Count Frequency (%)

No values found.

Most frequent character per category

Most occurring scripts

Value Count Frequency (%)

No values found.

Most frequent character per script

Most occurring blocks

Value Count Frequency (%)

No values found.

Most frequent character per block

cp

Categorical

HIGH CORRELATION
HIGH CORRELATION

Distinct 4

Distinct (%) 1.3%

Missing 0

Missing (%) 0.0%

Memory size 2.5 KIB

4 144

3 86

2 50

1 23

- Overview
- Categories
- Words
- Characters

Length

Max length 1

Median length 1

Mean length 1

Min length 1

Characters and Unicode

Total characters 0

Length

Max length	1
Median length	1
Mean length	1
Min length	1

Characters and Unicode

Total characters	0
Distinct characters	0
Distinct categories	0
Distinct scripts	0
Distinct blocks	0

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0
Unique (%)	0.0%

Sample

1st row	1
2nd row	4
3rd row	4
4th row	3
5th row	2

Common Values

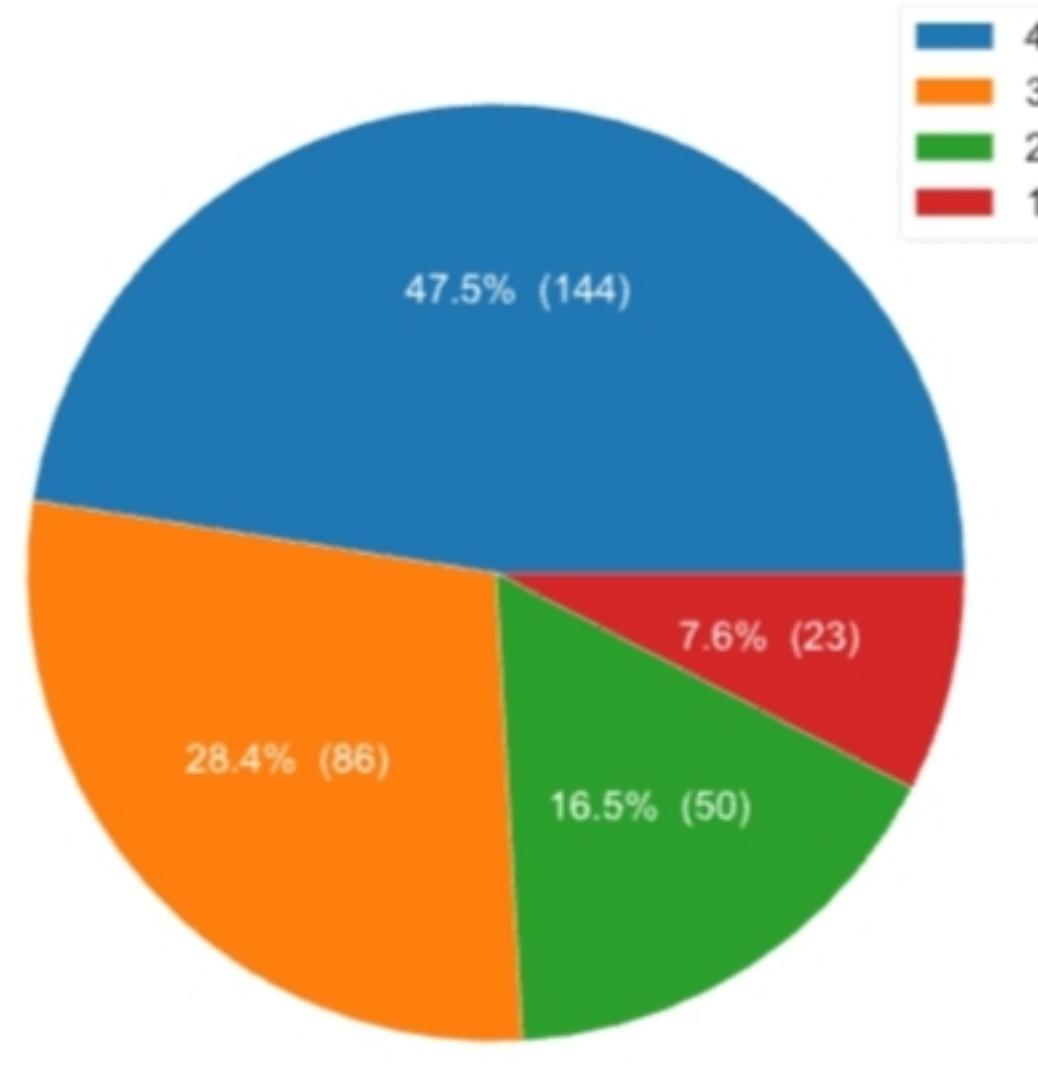
Value	Count	Frequency (%)
4	144	47.5%
3	86	28.4%
2	50	16.5%
1	23	7.6%

Length



Histogram of lengths of the category

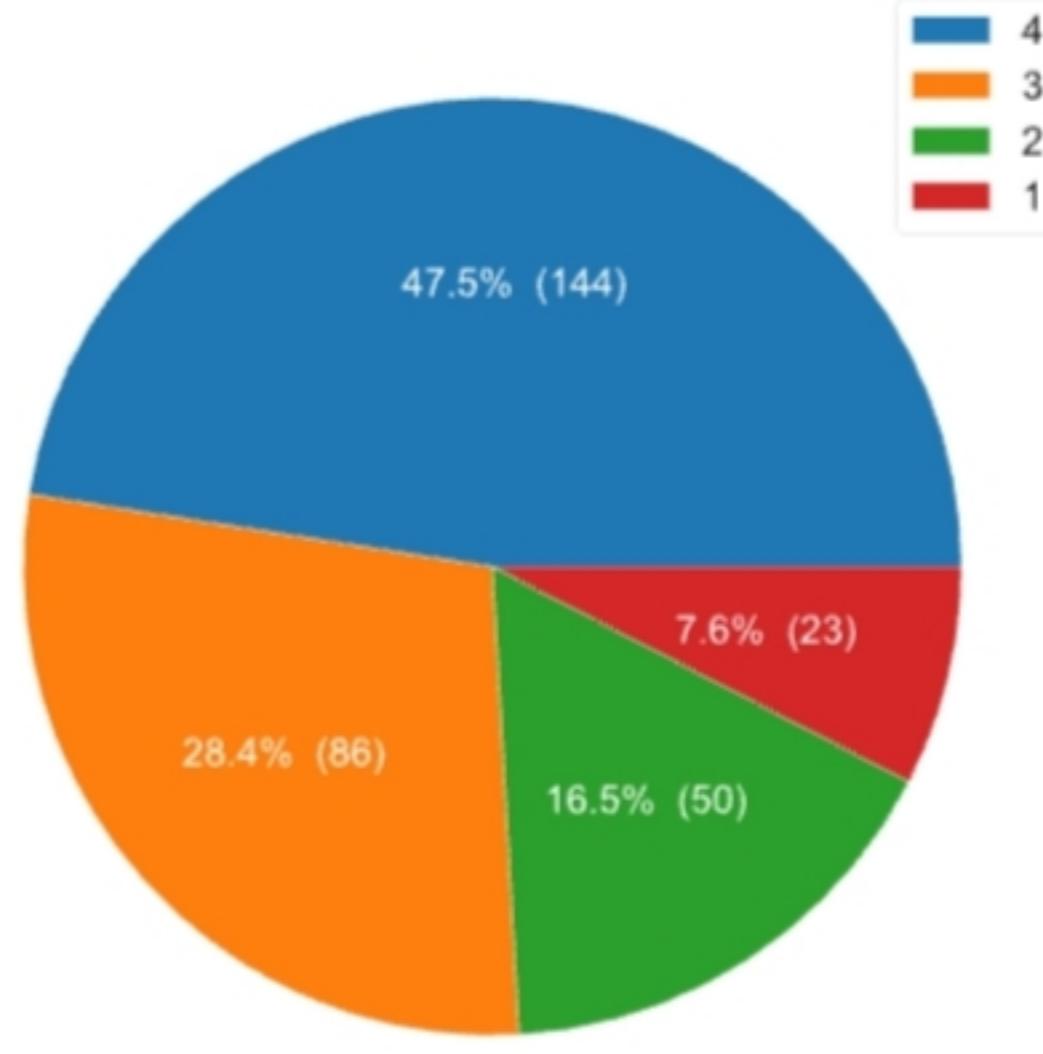
Pie chart



Value	Count	Frequency (%)
4	144	47.5%
3	86	28.4%
2	50	16.5%

Histogram of lengths of the category

Pie chart



Value Count Frequency (%)

4	144	47.5%
3	86	28.4%
2	50	16.5%
1	23	7.6%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value Count Frequency (%)

No values found.

Most occurring categories

Value Count Frequency (%)

No values found.

Most frequent character per category

Most occurring scripts

Value Count Frequency (%)

No values found.

Most frequent character per script

Most occurring blocks

Value Count Frequency (%)

No values found.

Most frequent character per block

trestbps

Real number (ℝ ≥ 0)

Distinct 50

Distinct (%) 16.5%

Missing 0

Missing (%) 0.0%

Infinite 0

Infinite (%) 0.0%

Mean 131.689769

Minimum 94

Maximum 200

Zeros 0

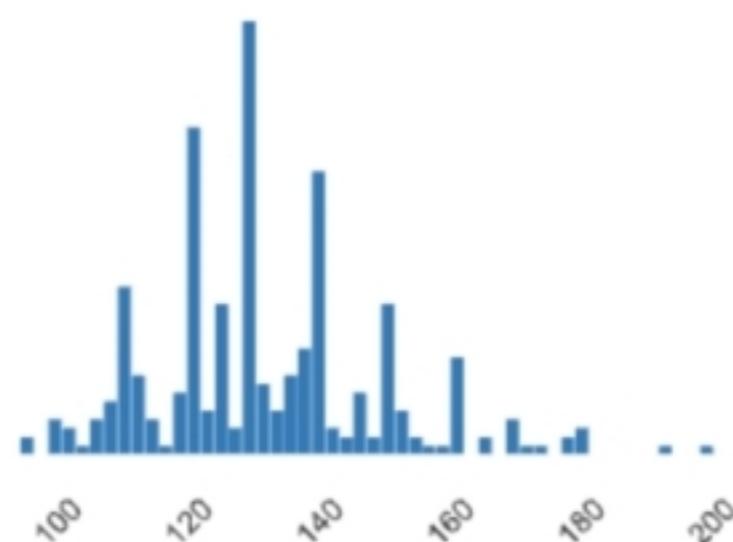
Zeros (%) 0.0%

Negative 0

Negative (%) 0.0%

Memory size 2.5 KiB





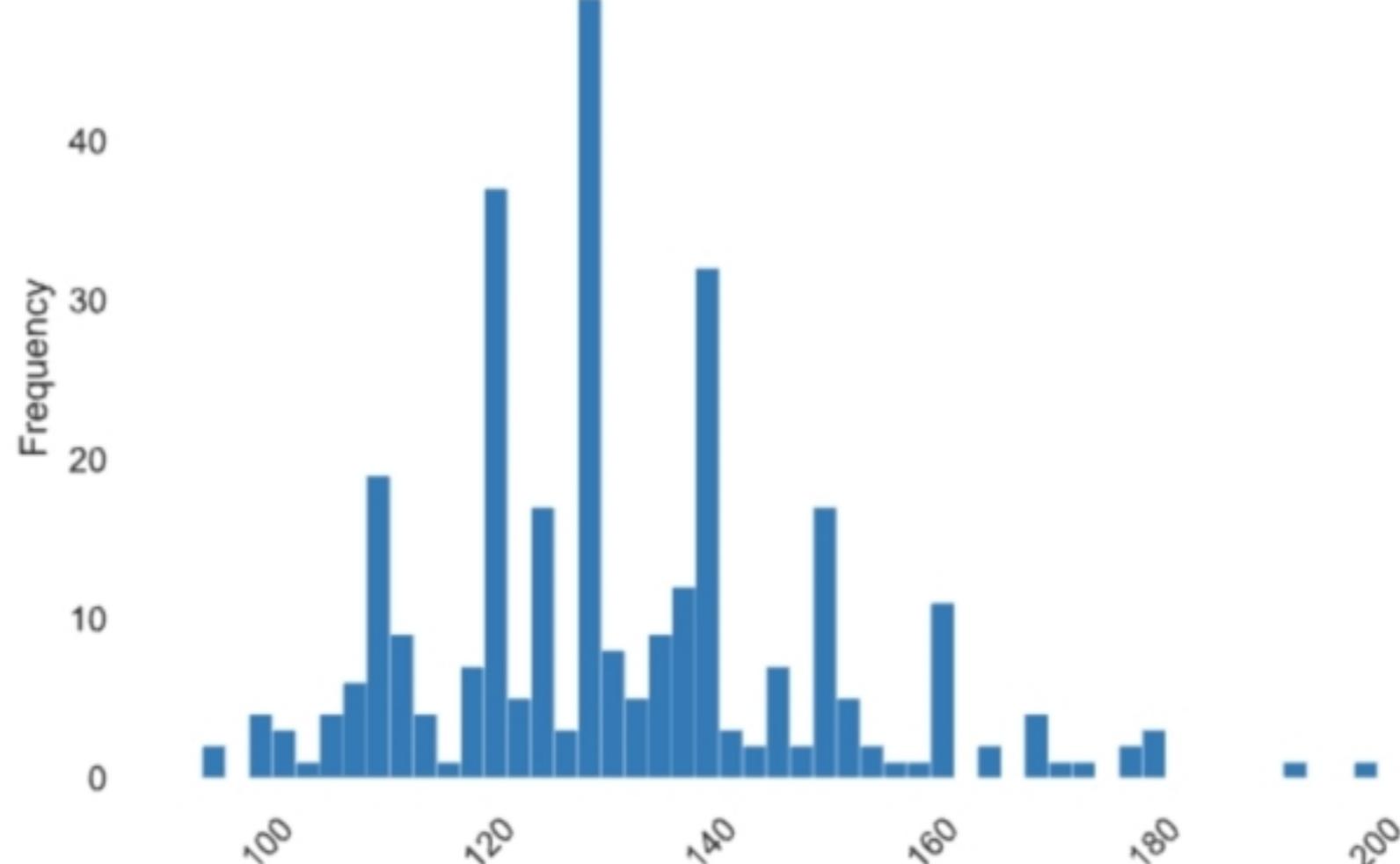
- Statistics
- Histogram
- Common values
- Extreme values

Quantile statistics

Minimum	94
5-th percentile	108
Q1	120
median	130
Q3	140
95-th percentile	160
Maximum	200
Range	106
Interquartile range (IQR)	20

Descriptive statistics

Standard deviation	17.59974773
Coefficient of variation (CV)	0.1336455206
Kurtosis	0.8800738686
Mean	131.689769
Median Absolute Deviation (MAD)	10
Skewness	0.7060346498
Sum	39902
Variance	309.7511201
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
120	37	12.2%
130	36	11.9%
140	32	10.6%
110	19	6.3%
150	17	5.6%
128	12	4.0%
138	12	4.0%
125	11	3.6%
160	11	3.6%
112	9	3.0%
Other values (40)	107	35.3%

- Minimum 5 values
- Maximum 5 values

Value	Count	Frequency (%)
94	2	0.7%
100	4	1.3%

- Minimum 5 values
- Maximum 5 values

Value	Count	Frequency (%)
94	2	0.7%
100	4	1.3%
101	1	0.3%
102	2	0.7%
104	1	0.3%
105	3	1.0%
106	1	0.3%
108	6	2.0%
110	19	6.3%
112	9	3.0%

Value	Count	Frequency (%)
200	1	0.3%
192	1	0.3%
180	3	1.0%
178	2	0.7%
174	1	0.3%
172	1	0.3%
170	4	1.3%
165	1	0.3%
164	1	0.3%
160	11	3.6%

chol

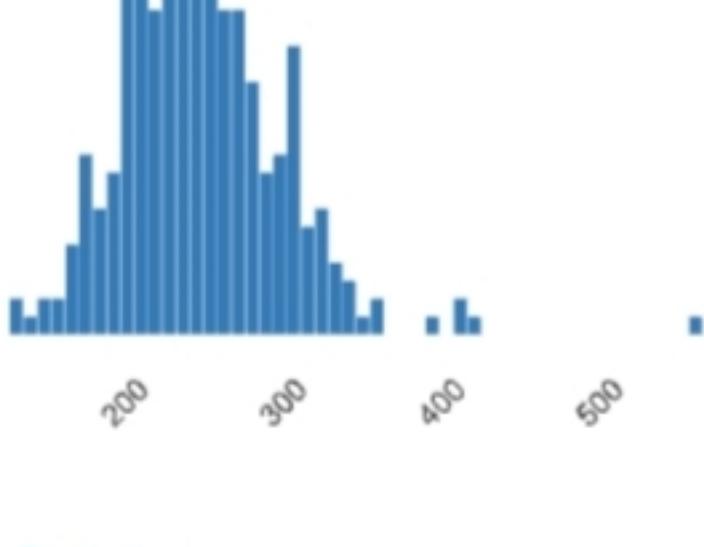
Real number (ℝ ≥ 0)

Distinct	152
Distinct (%)	50.2%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%

Mean 246.6930693

Minimum	126
Maximum	564
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%

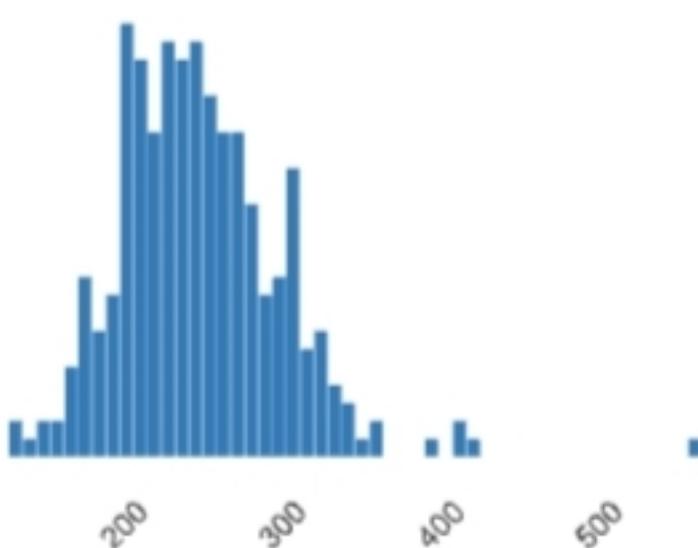
Memory size 2.5 KiB



- Statistics
- Histogram
- Common values
- Extreme values

Quantile statistics

Minimum	126
5-th percentile	175.1
Q1	211
median	241
Q3	275
95-th percentile	326.9
Maximum	564
Range	438
Interquartile range (IQR)	64



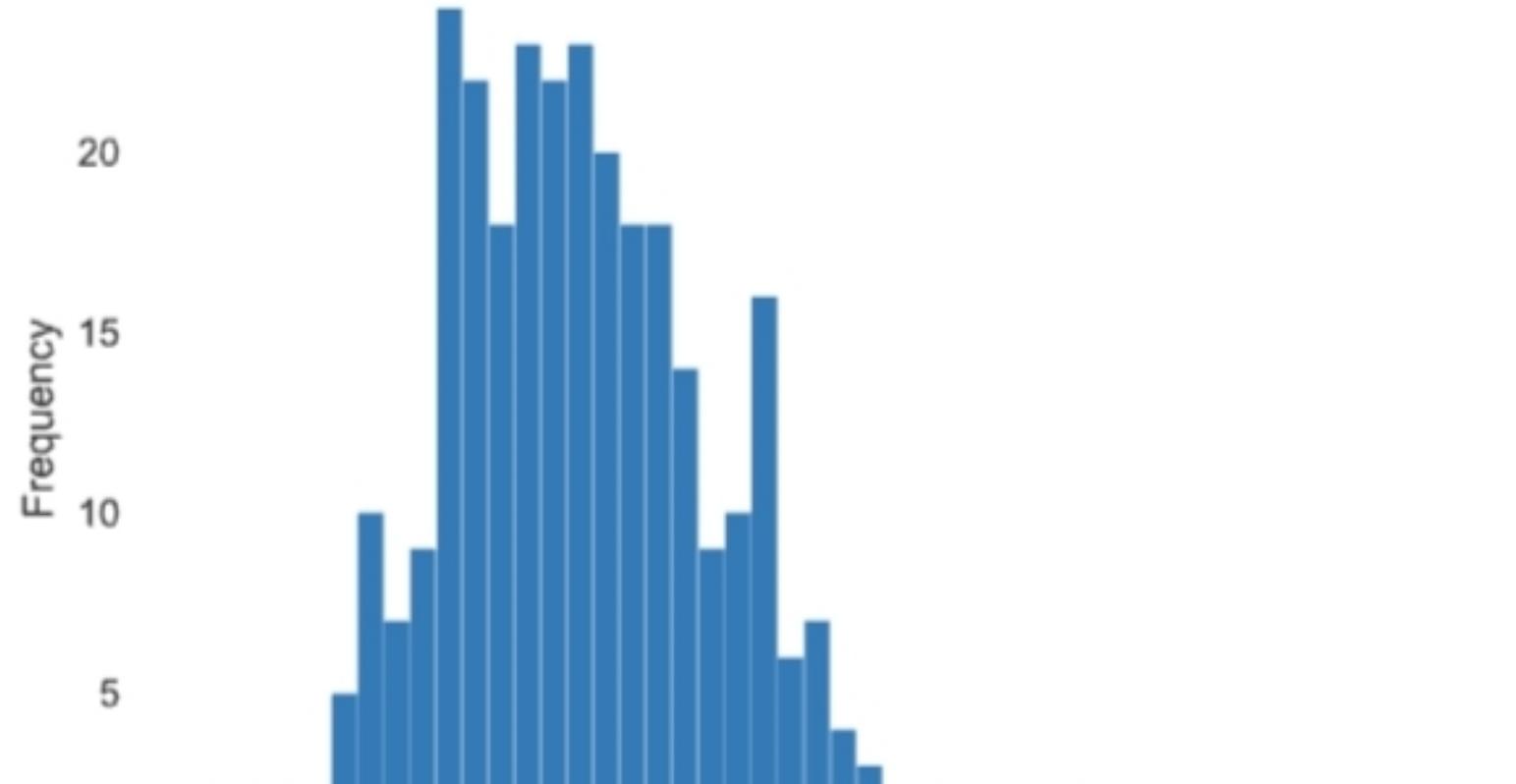
- Statistics
- Histogram
- Common values
- Extreme values

Quantile statistics

Minimum	126
5-th percentile	175.1
Q1	211
median	241
Q3	275
95-th percentile	326.9
Maximum	564
Range	438
Interquartile range (IQR)	64

Descriptive statistics

Standard deviation	51.77691754
Coefficient of variation (CV)	0.209883957
Kurtosis	4.491724287
Mean	246.6930693
Median Absolute Deviation (MAD)	32
Skewness	1.135503153
Sum	74748
Variance	2680.84919
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
204	6	2.0%
197	6	2.0%
234	6	2.0%
269	5	1.7%
212	5	1.7%
254	5	1.7%
240	4	1.3%
226	4	1.3%
239	4	1.3%
177	4	1.3%
Other values (142)	254	83.8%

- Minimum 5 values
- Maximum 5 values

Value	Count	Frequency (%)
126	1	0.3%
131	1	0.3%
141	1	0.3%
149	2	0.7%

Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
204	6	2.0%
197	6	2.0%
234	6	2.0%
269	5	1.7%
212	5	1.7%
254	5	1.7%
240	4	1.3%
226	4	1.3%
239	4	1.3%
177	4	1.3%
Other values (142)	254	83.8%

- Minimum 5 values
- Maximum 5 values

Value

Value	Count	Frequency (%)
126	1	0.3%
131	1	0.3%
141	1	0.3%
149	2	0.7%
157	1	0.3%
160	1	0.3%
164	1	0.3%
166	1	0.3%
167	1	0.3%
168	1	0.3%

Value

Value	Count	Frequency (%)
564	1	0.3%
417	1	0.3%
409	1	0.3%
407	1	0.3%
394	1	0.3%
360	1	0.3%
354	1	0.3%
353	1	0.3%
342	1	0.3%
341	1	0.3%

fbs

Categorical

Distinct 2

Distinct (%) 0.7%

Missing 0

Missing (%) 0.0%

Memory size 2.5 KiB

0 258

1 45

- Overview
- Categories
- Words
- Characters

Length

Max length 1

Median length 1

Mean length 1

Min length 1

Characters and Unicode

Total characters 0

Distinct characters 0

Distinct categories 0 ⓘ

Distinct scripts 0 ⓘ

Distinct blocks 0 ⓘ

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique 0 ⓘ

Length

Max length	1
Median length	1
Mean length	1
Min length	1

Characters and Unicode

Total characters	0
Distinct characters	0
Distinct categories	0 
Distinct scripts	0 
Distinct blocks	0 

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0 
Unique (%)	0.0%

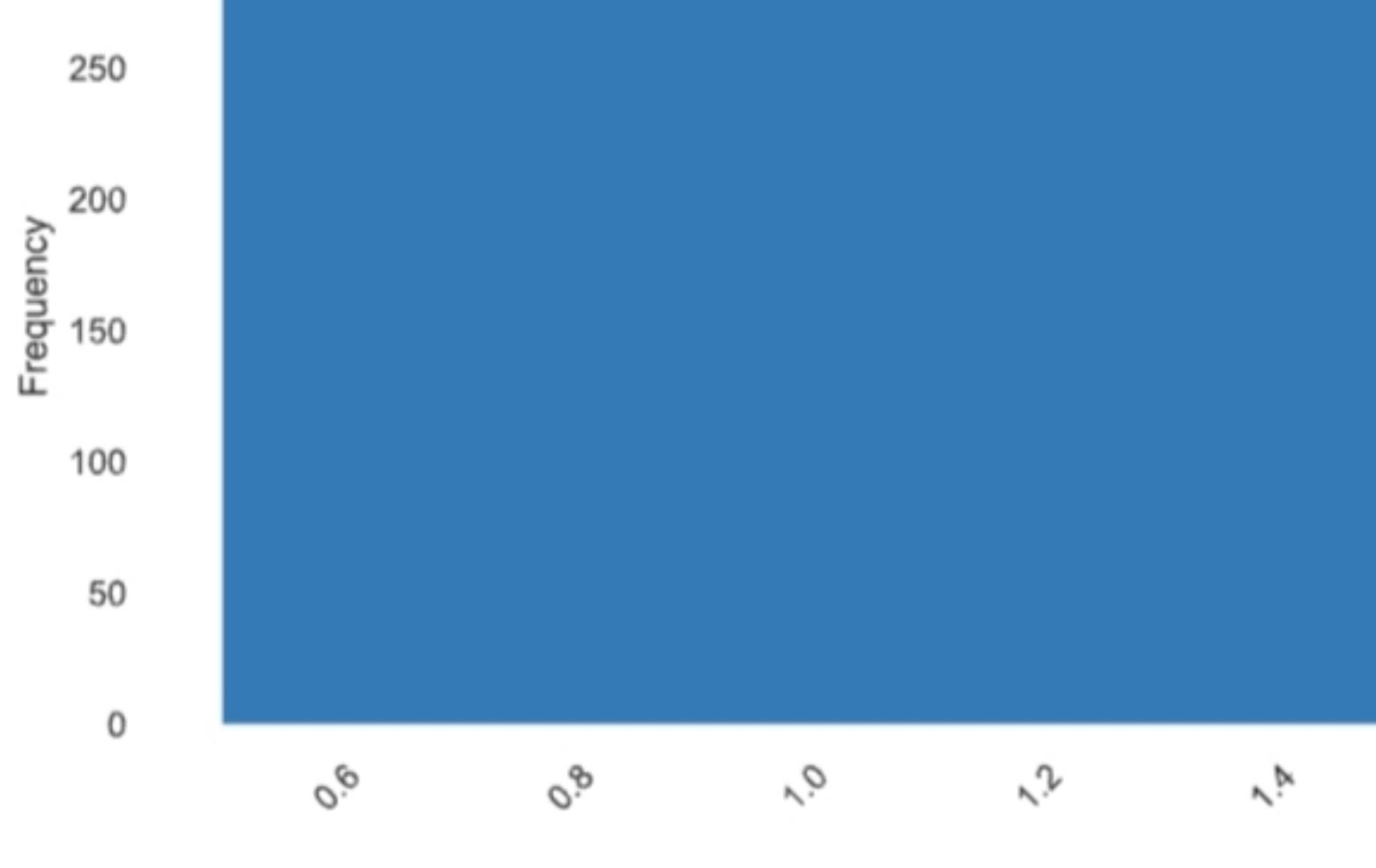
Sample

1st row	1
2nd row	0
3rd row	0
4th row	0
5th row	0

Common Values

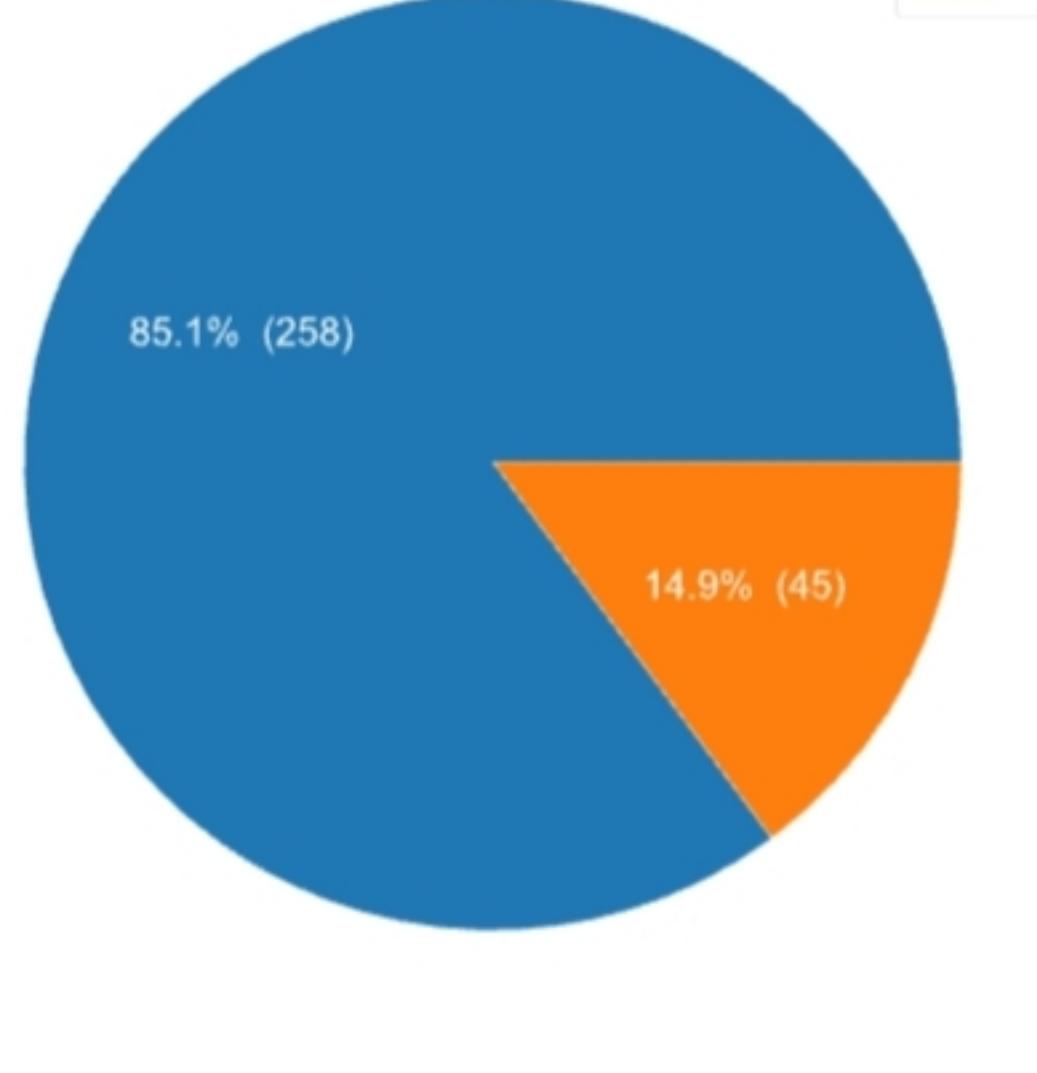
Value	Count	Frequency (%)
0	258	85.1%
1	45	14.9%

Length



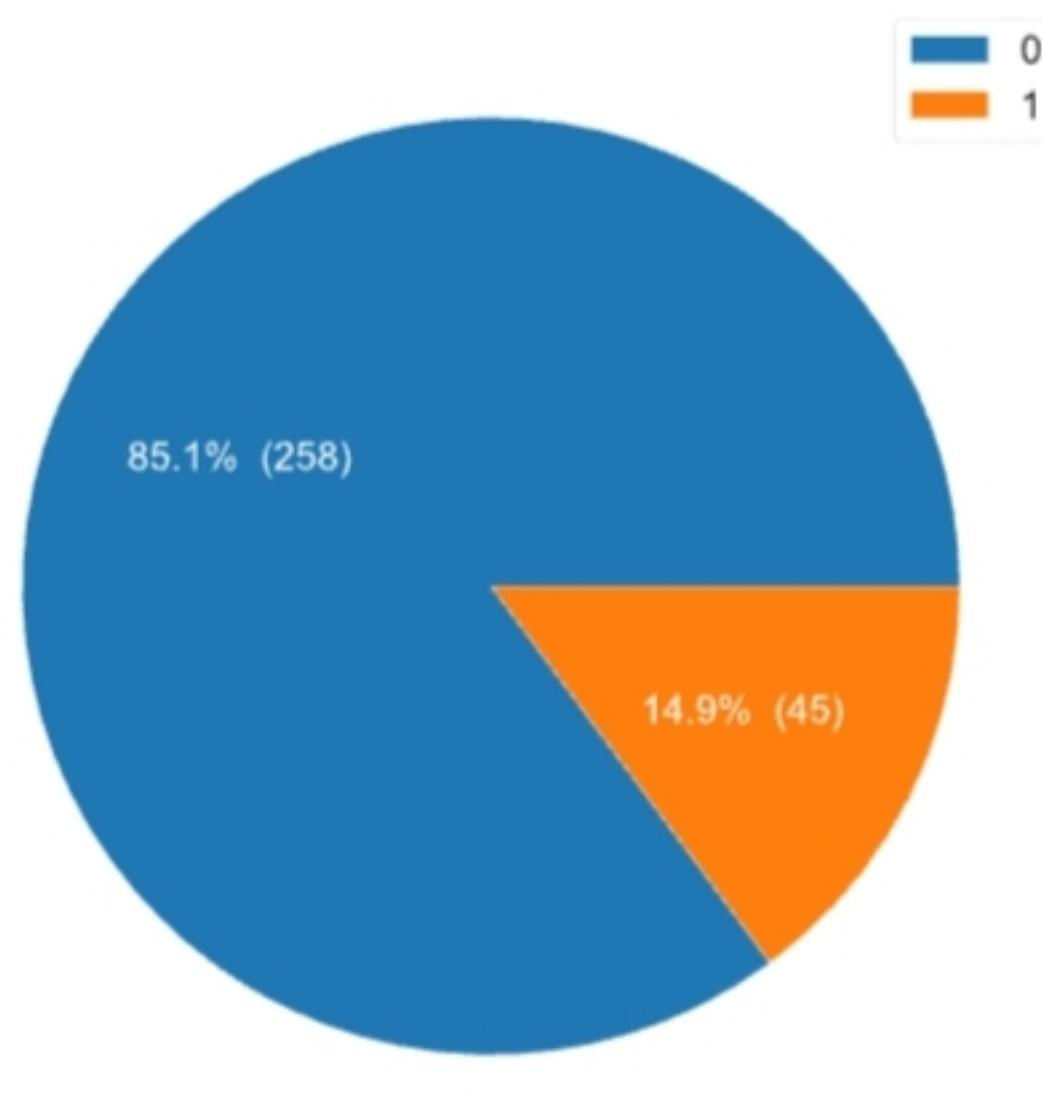
Histogram of lengths of the category

Pie chart



Value	Count	Frequency (%)
0	258	85.1%
1	45	14.9%

- Characters
- Categories
- Scripts



Value Count Frequency (%)

0	258	85.1%
1	45	14.9%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value Count Frequency (%)

No values found.

Most occurring categories

Value Count Frequency (%)

No values found.

Most frequent character per category

Most occurring scripts

Value Count Frequency (%)

No values found.

Most frequent character per script

Most occurring blocks

Value Count Frequency (%)

No values found.

Most frequent character per block

restecg

Categorical

HIGH CORRELATION

Distinct 3

Distinct (%) 1.0%

Missing 0

Missing (%) 0.0%

Memory size 2.5 KiB

0 151

2 148

1 4

- Overview
- Categories
- Words
- Characters

Length

Max length 1

Median length 1

Mean length 1

Min length 1

Characters and Unicode

Total characters 0

Distinct characters 0

Distinct categories 0

Distinct scripts 0

Distinct blocks 0

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Length

Max length 1

Median length 1

Mean length 1

Min length 1

Characters and Unicode

Total characters 0

Distinct characters 0

Distinct categories 0

Distinct scripts 0

Distinct blocks 0

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique 0

Unique (%) 0.0%

Sample

1st row 2

2nd row 2

3rd row 2

4th row 0

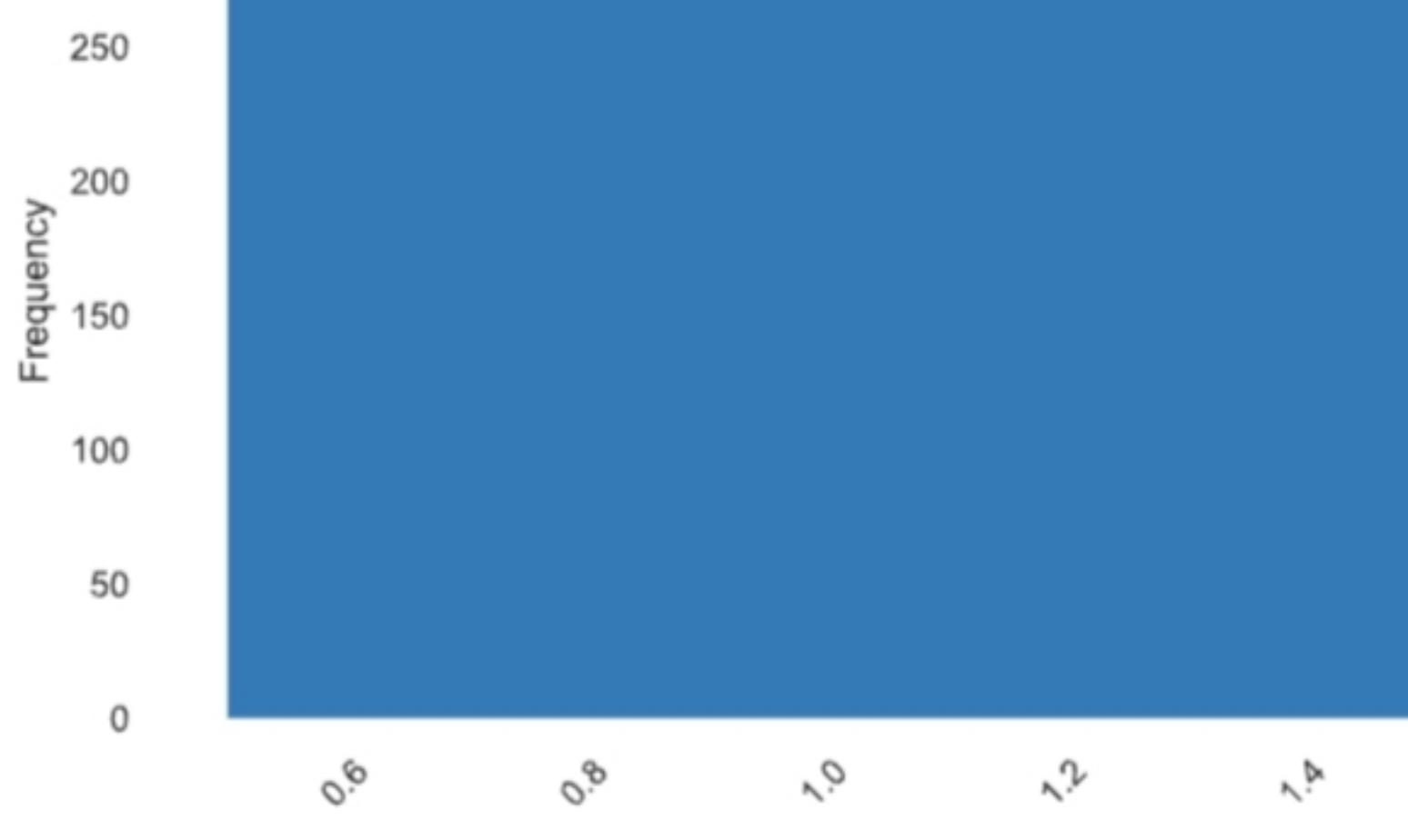
5th row 2

Common Values

Value	Count	Frequency (%)
0	151	49.8%
2	148	48.8%

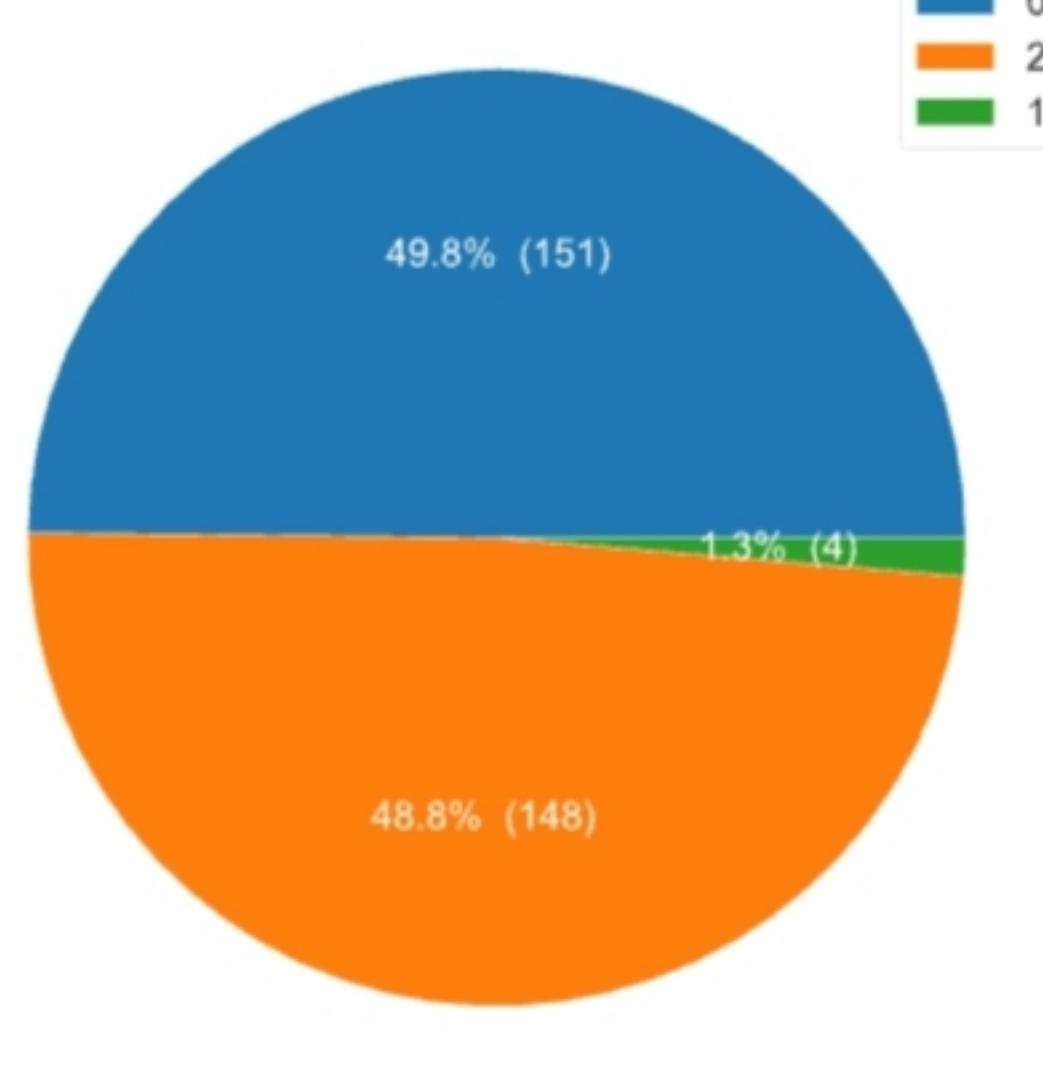
1 4 1.3%

Length



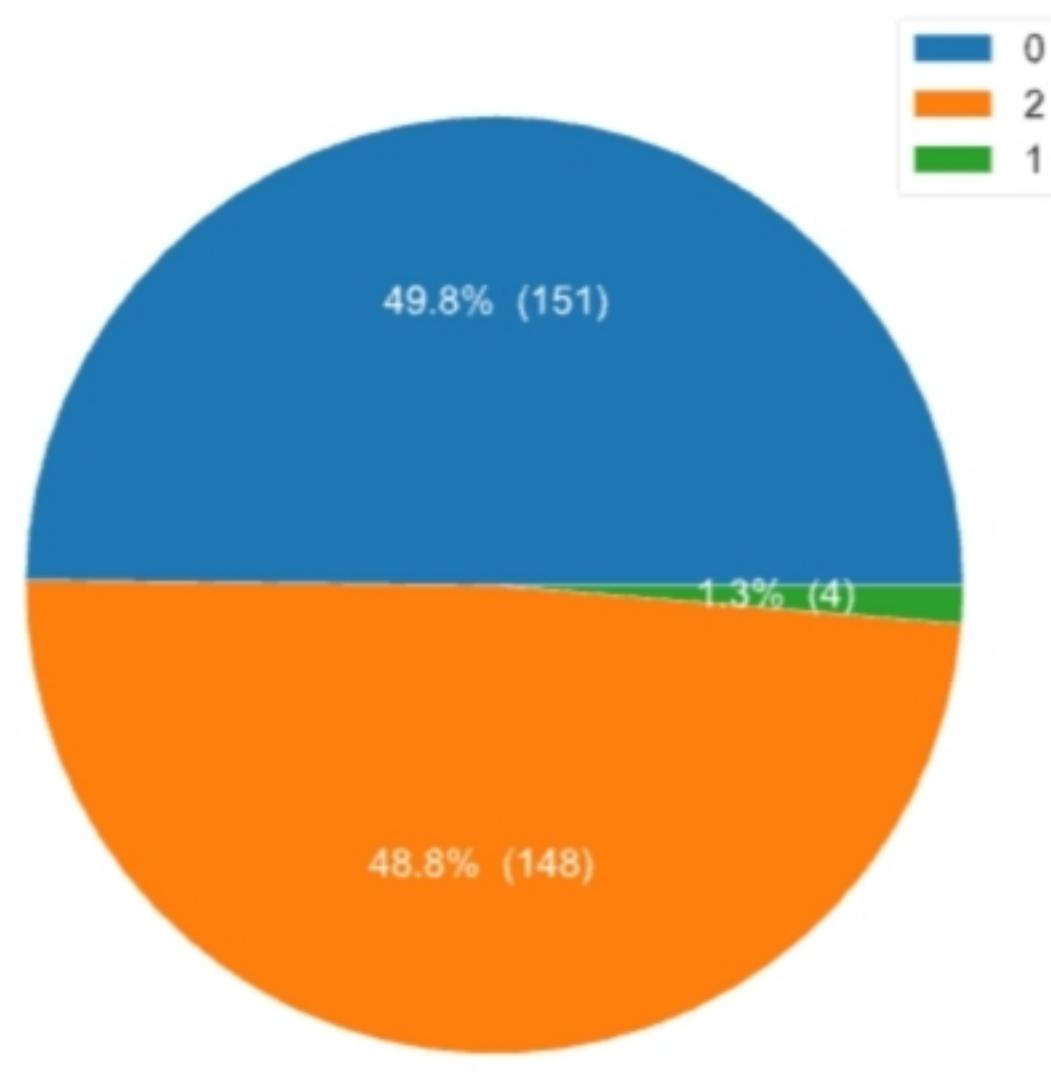
Histogram of lengths of the category

Pie chart



Value	Count	Frequency (%)
0	151	49.8%
2	148	48.8%

1 4 1.3%



Value Count Frequency (%)

Value	Count	Frequency (%)
0	151	49.8%
2	148	48.8%
1	4	1.3%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value Count Frequency (%)

No values found.

Most occurring categories

Value Count Frequency (%)

No values found.

Most frequent character per category

Most occurring scripts

Value Count Frequency (%)

No values found.

Most frequent character per script

Most occurring blocks

Value Count Frequency (%)

No values found.

Most frequent character per block

thalach

Real number (ℝ ≥ 0)

HIGH CORRELATION

Distinct 91

Distinct (%) 30.0%

Missing 0

Missing (%) 0.0%

Infinite 0

Infinite (%) 0.0%

Mean 149.6072607

Minimum 71

Maximum 202

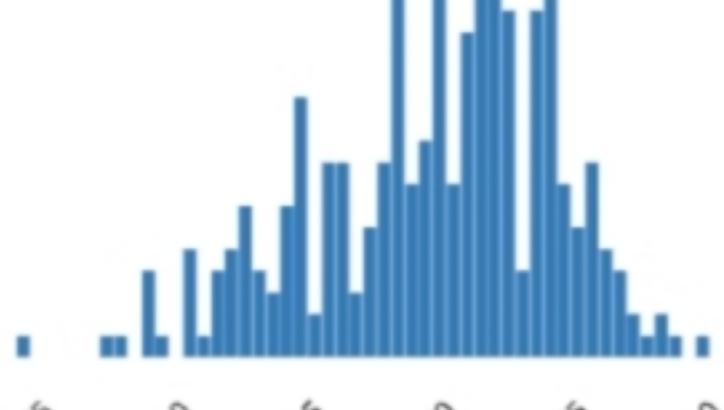
Zeros 0

Zeros (%) 0.0%

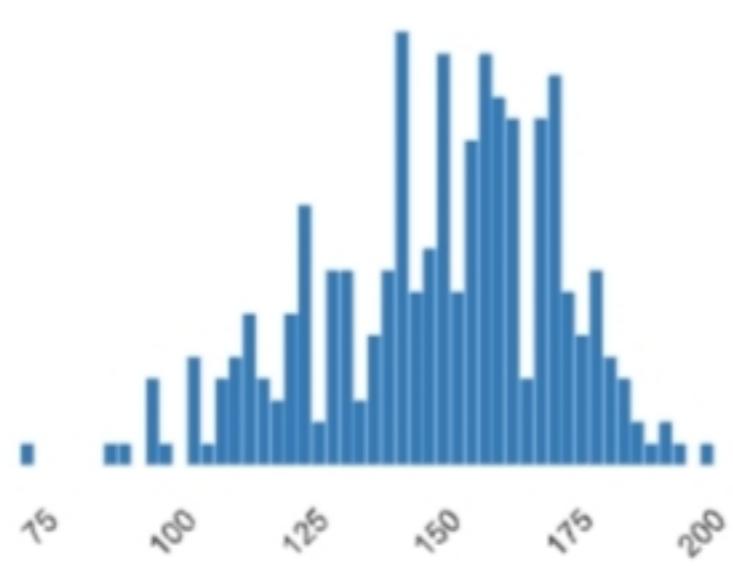
Negative 0

Negative (%) 0.0%

Memory size 2.5 KiB



- Statistics
- Histogram
- Common values



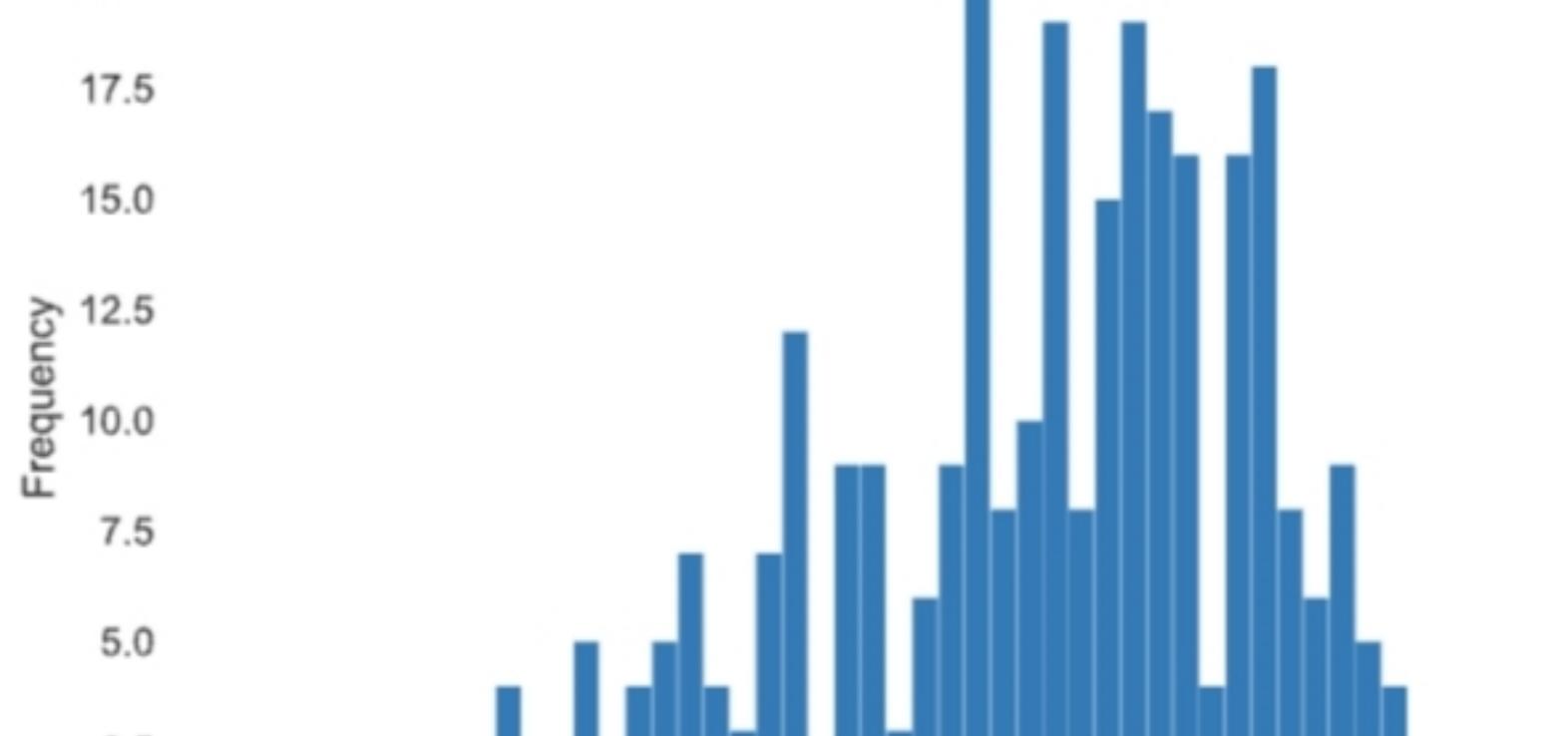
- Statistics
- Histogram
- Common values
- Extreme values

Quantile statistics

Minimum	71
5-th percentile	108.1
Q1	133.5
median	153
Q3	166
95-th percentile	181.9
Maximum	202
Range	131
Interquartile range (IQR)	32.5

Descriptive statistics

Standard deviation	22.87500328
Coefficient of variation (CV)	0.152900355
Kurtosis	-0.05354095895
Mean	149.6072607
Median Absolute Deviation (MAD)	15
Skewness	-0.5374486699
Sum	45331
Variance	523.2657749
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
162	11	3.6%
163	9	3.0%
160	9	3.0%
152	8	2.6%
172	7	2.3%
173	7	2.3%
144	7	2.3%
132	7	2.3%
143	7	2.3%
150	7	2.3%
Other values (81)	224	73.9%

- Minimum 5 values
- Maximum 5 values

Value	Count	Frequency (%)
71	1	0.3%
88	1	0.3%

- Minimum 5 values
- Maximum 5 values

Value	Count	Frequency (%)
71	1	0.3%
88	1	0.3%
90	1	0.3%
95	1	0.3%
96	2	0.7%
97	1	0.3%
99	1	0.3%
103	2	0.7%
105	3	1.0%
106	1	0.3%

Value	Count	Frequency (%)
202	1	0.3%
195	1	0.3%
194	1	0.3%
192	1	0.3%
190	1	0.3%
188	1	0.3%
187	1	0.3%
186	2	0.7%
185	1	0.3%
184	1	0.3%

exang

Categorical

HIGH CORRELATION

Distinct	2
Distinct (%)	0.7%
Missing	0
Missing (%)	0.0%

Memory size 2.5 KiB

0	204
1	99

- Overview
- Categories
- Words
- Characters

Length

Max length	1
Median length	1
Mean length	1
Min length	1

Characters and Unicode

Total characters	0
Distinct characters	0
Distinct categories	0 
Distinct scripts	0 
Distinct blocks	0 

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0 
Unique (%)	0.0%

Sample

1st row	0
2nd row	1
3rd row	1
4th row	0
5th row	0

Common Values

Value	Count	Frequency (%)
0	204	67.3%
1	99	32.7%

Length

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique 0

Unique (%) 0.0%

Sample

1st row 0

2nd row 1

3rd row 1

4th row 0

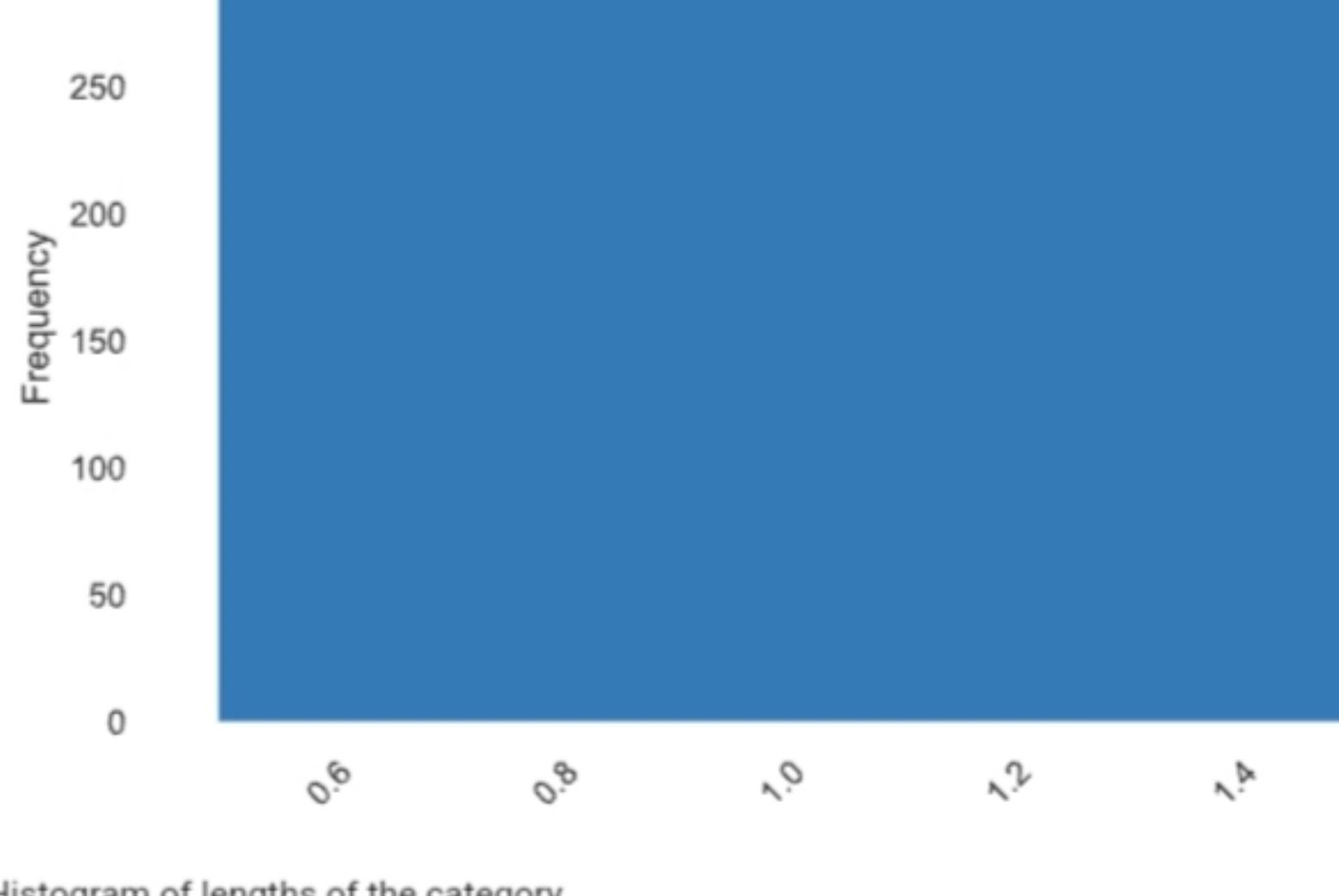
5th row 0

Common Values

Value Count Frequency (%)

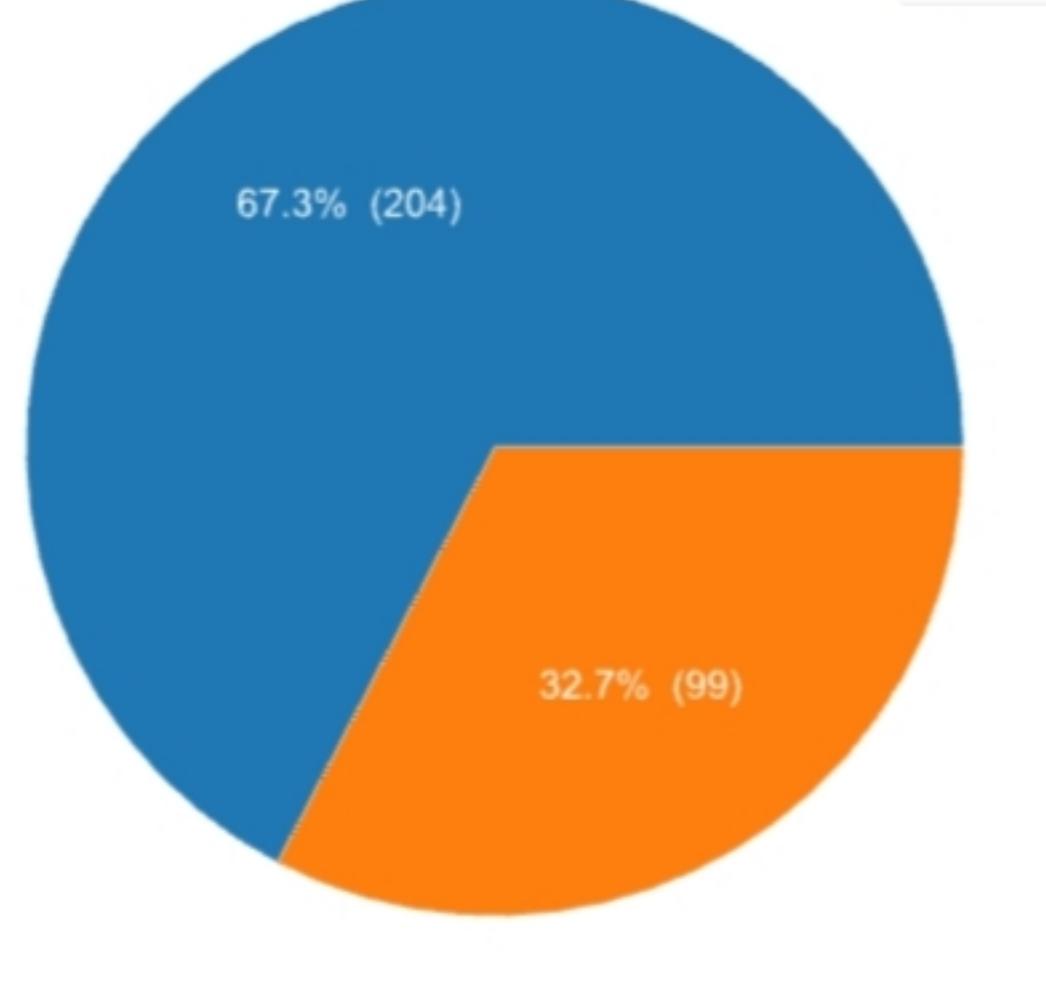
0	204	67.3%
1	99	32.7%

Length



Histogram of lengths of the category

Pie chart



Value Count Frequency (%)

0	204	67.3%
1	99	32.7%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value Count Frequency (%)

No values found.

Most occurring categories

Value Count Frequency (%)

No values found.

Most frequent character per category

Most occurring scripts

Value Count Frequency (%)

No values found.

Most frequent character per script

Value	Count	Frequency (%)
0	204	67.3%
1	99	32.7%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value	Count	Frequency (%)
No values found.		

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

oldpeak

Real number (ℝ ≥ 0)

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
ZEROS

Distinct 40

Distinct (%) 13.2%

Missing 0

Missing (%) 0.0%

Infinite 0

Infinite (%) 0.0%

Mean 1.03960396

Minimum 0

Maximum 6.2

Zeros 99

Zeros (%) 32.7%

Negative 0

Negative (%) 0.0%

Memory size 2.5 KiB



- Statistics
- Histogram
- Common values
- Extreme values

Quantile statistics

Minimum 0

5-th percentile 0

Q1 0

median 0.8

Q3 1.6

95-th percentile 3.4

Maximum 6.2

Range 6.2

Interquartile range (IQR) 1.6

Descriptive statistics

Standard deviation 1.161075022

Coefficient of variation (CV) 1.116012502

- Statistics
- Histogram
- Common values
- Extreme values

Quantile statistics

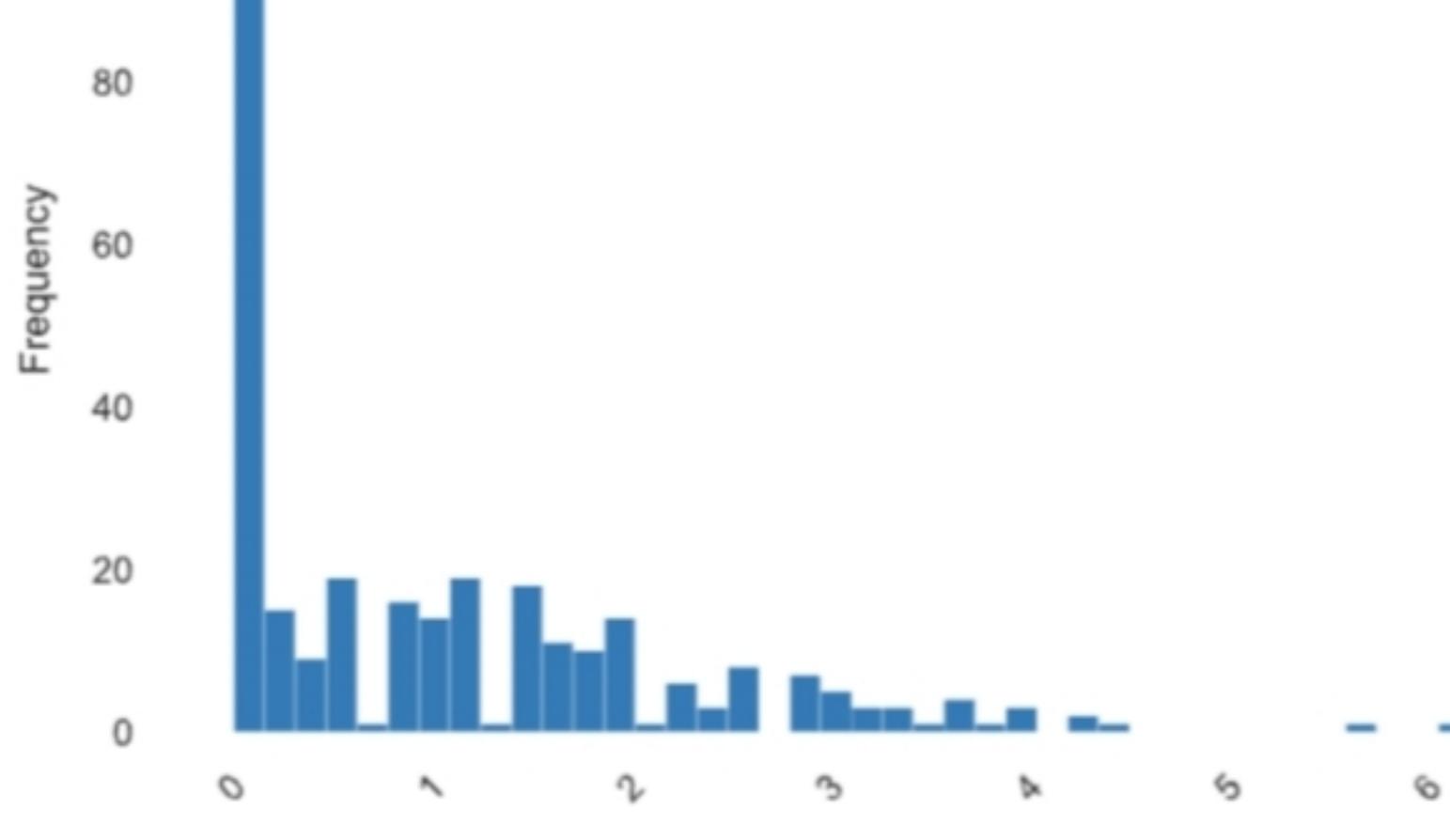
Minimum	0
5-th percentile	0
Q1	0
median	0.8
Q3	1.6
95-th percentile	3.4
Maximum	6.2
Range	6.2

Interquartile range (IQR) 1.6

Descriptive statistics

Standard deviation	1.161075022
Coefficient of variation (CV)	1.116843593
Kurtosis	1.575813073
Mean	1.03960396
Median Absolute Deviation (MAD)	0.8
Skewness	1.269719931
Sum	315
Variance	1.348095207

Monotonicity Not monotonic



Histogram with fixed size bins (bins=40)

Value	Count	Frequency (%)
0	99	32.7%
1.2	17	5.6%
1	14	4.6%
0.6	14	4.6%
0.8	13	4.3%
1.4	13	4.3%
0.2	12	4.0%
1.6	11	3.6%
1.8	10	3.3%
0.4	9	3.0%
Other values (30)	91	30.0%

- Minimum 5 values
- Maximum 5 values

Value Count Frequency (%)

0	99	32.7%
0.1	7	2.3%
0.2	12	4.0%
0.3	3	1.0%
0.4	9	3.0%
0.5	5	1.7%
0.6	14	4.6%
0.7	1	0.3%

- Minimum 5 values
- Maximum 5 values

Value	Count	Frequency (%)
0	99	32.7%
0.1	7	2.3%
0.2	12	4.0%
0.3	3	1.0%
0.4	9	3.0%
0.5	5	1.7%
0.6	14	4.6%
0.7	1	0.3%
0.8	13	4.3%
0.9	3	1.0%

Value	Count	Frequency (%)
6.2	1	0.3%
5.6	1	0.3%
4.4	1	0.3%
4.2	2	0.7%
4	3	1.0%
3.8	1	0.3%
3.6	4	1.3%
3.5	1	0.3%
3.4	3	1.0%
3.2	2	0.7%

slope

Categorical

HIGH CORRELATION
HIGH CORRELATION

HIGH CORRELATION
HIGH CORRELATION

HIGH CORRELATION
HIGH CORRELATION

Distinct	3
Distinct (%)	1.0%
Missing	0
Missing (%)	0.0%

Memory size 2.5 KIB

1 142
2 140
3 21

- Overview
- Categories
- Words
- Characters

Length

Max length 1
Median length 1
Mean length 1
Min length 1

Characters and Unicode

Total characters 0
Distinct characters 0
Distinct categories 0 ⓘ
Distinct scripts 0 ⓘ
Distinct blocks 0 ⓘ

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.
Junique

Unique 0 ⓘ

Unique (%) 0.0%

Sample

1st row 3
2nd row 2
3rd row 2
4th row 3
5th row 1

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique 0 

Unique (%) 0.0%

Sample

1st row 3

2nd row 2

3rd row 2

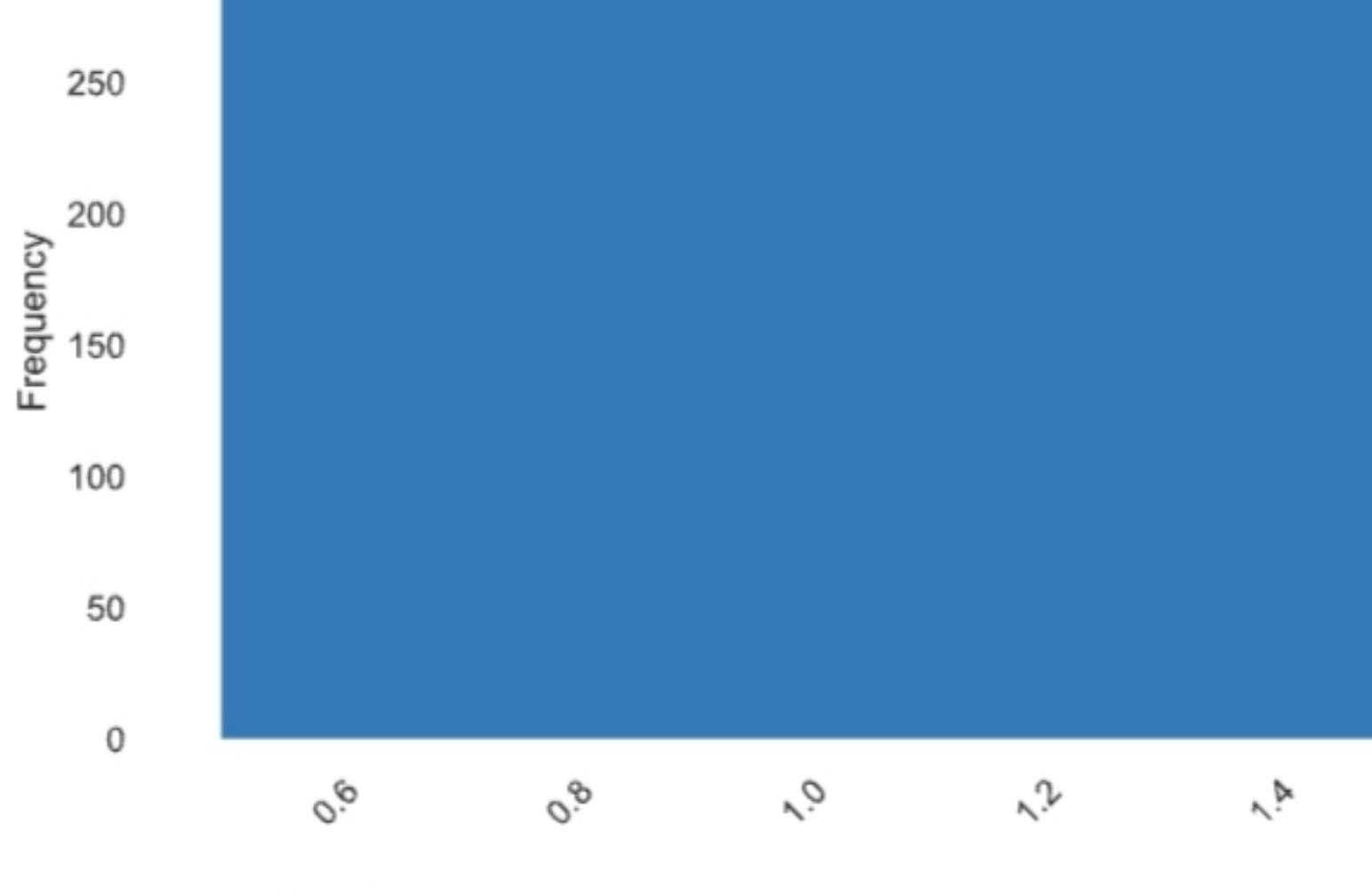
4th row 3

5th row 1

Common Values

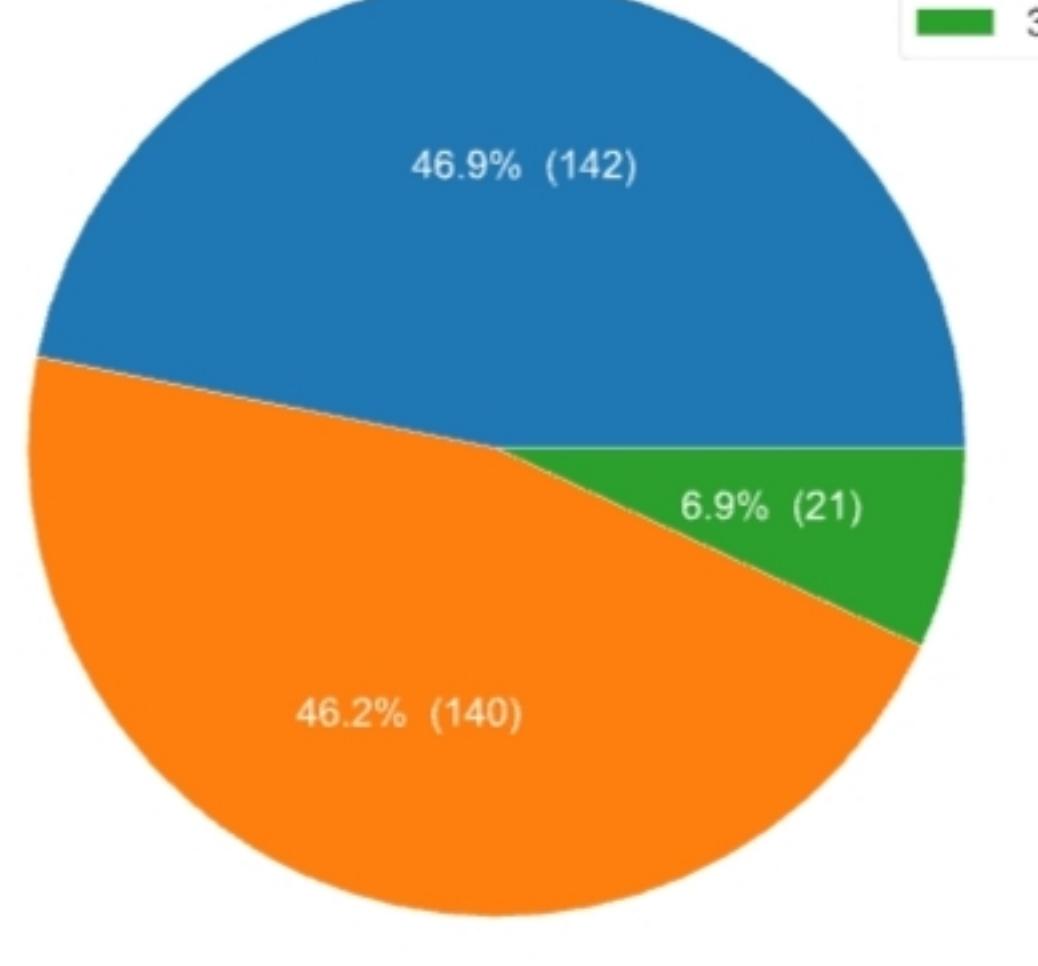
Value	Count	Frequency (%)
1	142	46.9%
2	140	46.2%
3	21	6.9%

Length



Histogram of lengths of the category

Pie chart



Value	Count	Frequency (%)
1	142	46.9%
2	140	46.2%
3	21	6.9%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value	Count	Frequency (%)
No values found.		

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
1	142	46.9%
2	140	46.2%
3	21	6.9%

- [Characters](#)
- [Categories](#)
- [Scripts](#)
- [Blocks](#)

Most occurring characters

Value	Count	Frequency (%)
No values found.		

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

ca

Categorical

Distinct	5
Distinct (%)	1.7%
Missing	0
Missing (%)	0.0%

Memory size 2.5 KiB

0	176
1	65
2	38
3	20
-100000	4

- [Overview](#)
- [Categories](#)
- [Words](#)
- [Characters](#)

Length

Max length	7
Median length	1
Mean length	1.079207921
Min length	1

Characters and Unicode

Total characters	0
Distinct characters	0
Distinct categories	0 
Distinct scripts	0 
Distinct blocks	0 

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0 
Unique (%)	0.0%

Sample

1st row	0
2nd row	3
3rd row	2
4th row	0
5th row	0

Common Values

Value	Count	Frequency (%)
0	176	58.1%
1	65	21.5%
2	38	12.5%

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0
Unique (%)	0.0%

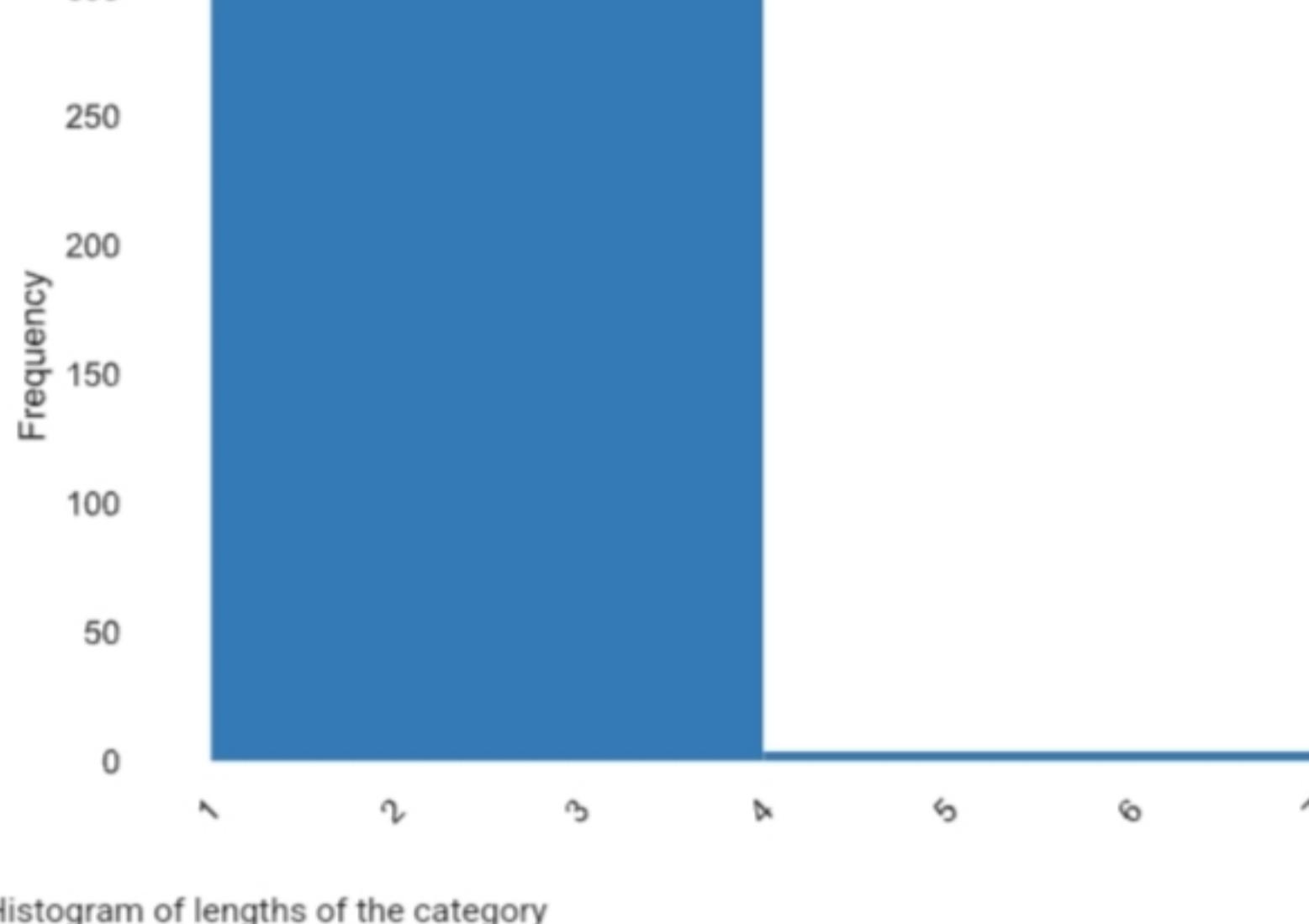
Sample

1st row	0
2nd row	3
3rd row	2
4th row	0
5th row	0

Common Values

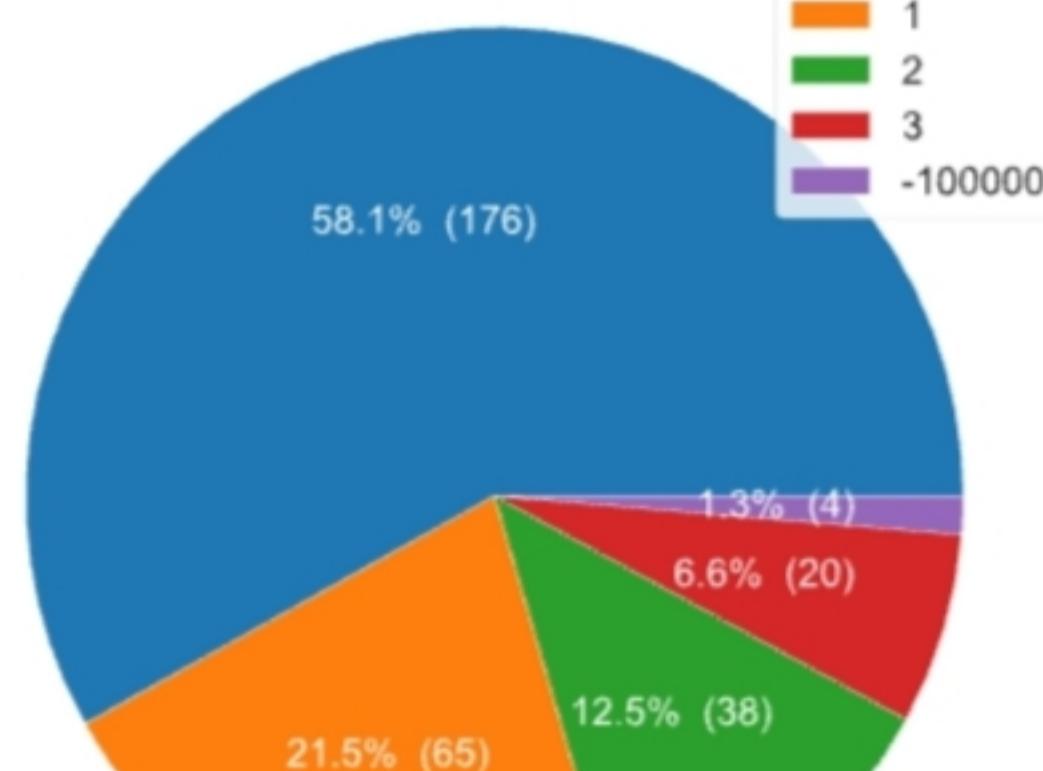
Value	Count	Frequency (%)
0	176	58.1%
1	65	21.5%
2	38	12.5%
3	20	6.6%
-100000	4	1.3%

Length



Histogram of lengths of the category

Pie chart



Most occurring characters

Value	Count	Frequency (%)
0	176	58.1%
1	65	21.5%
2	38	12.5%
3	20	6.6%
100000	4	1.3%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value	Count	Frequency (%)
0	176	58.1%
1	65	21.5%
2	38	12.5%
3	20	6.6%
100000	4	1.3%

Value	Count	Frequency (%)
0	176	58.1%
1	65	21.5%
2	38	12.5%
3	20	6.6%
100000	4	1.3%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value	Count	Frequency (%)
No values found.		

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

:hal

Categorical

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION

Distinct	4
Distinct (%)	1.3%
Missing	0
Missing (%)	0.0%

Memory size 2.5 KIB

3	166
7	117
6	18

-100000	2
---------	---

- Overview
- Categories
- Words
- Characters

Length

Max length	7
Median length	1
Mean length	1.03960396
Min length	1

Characters and Unicode

Total characters	0
Distinct characters	0
Distinct categories	0
Distinct scripts	0
Distinct blocks	0

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0
Unique (%)	0.0%

Sample

1st row	6
2nd row	3
3rd row	7
4th row	3
5th row	3

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0
Unique (%)	0.0%

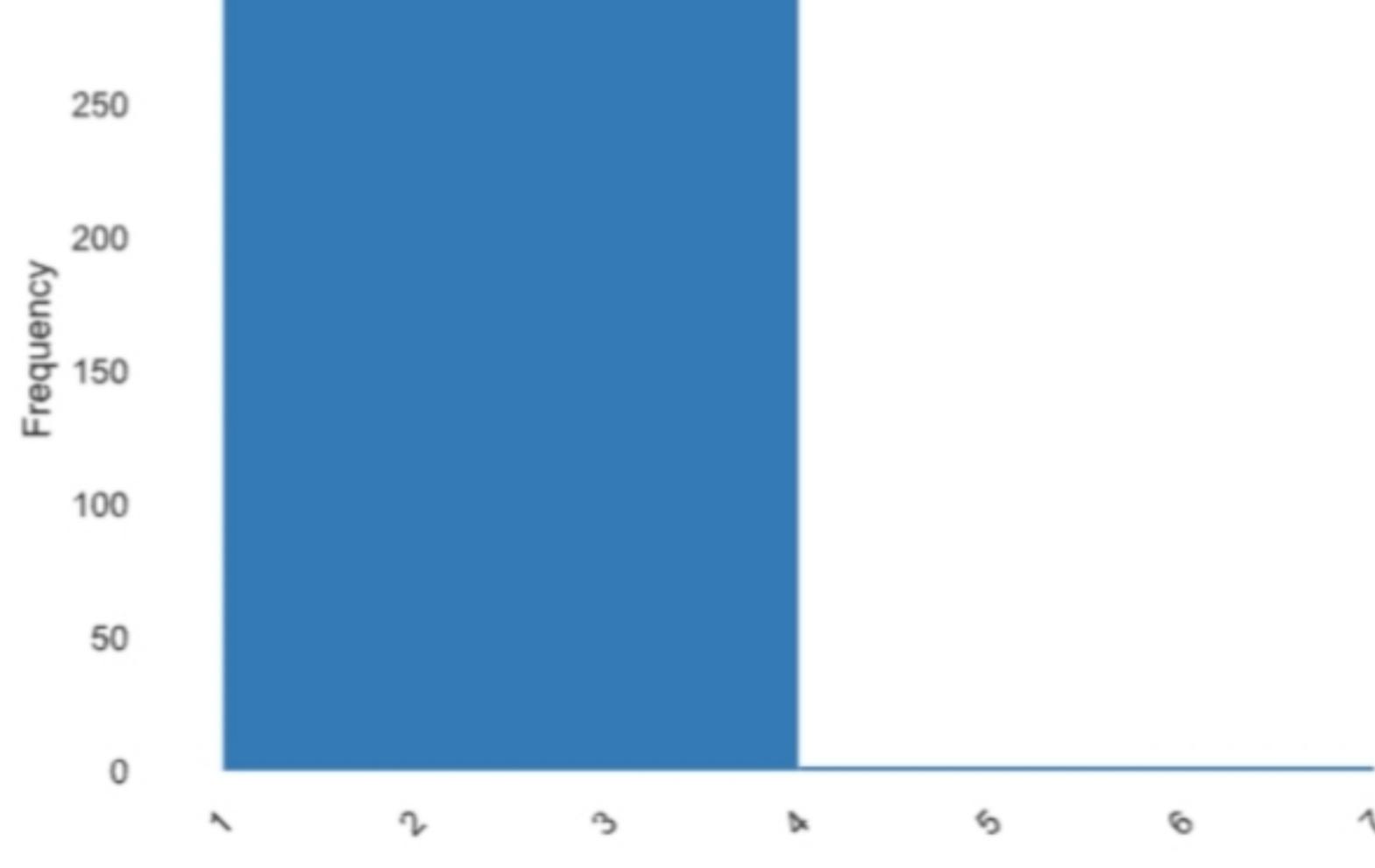
Sample

1st row	6
2nd row	3
3rd row	7
4th row	3
5th row	3

Common Values

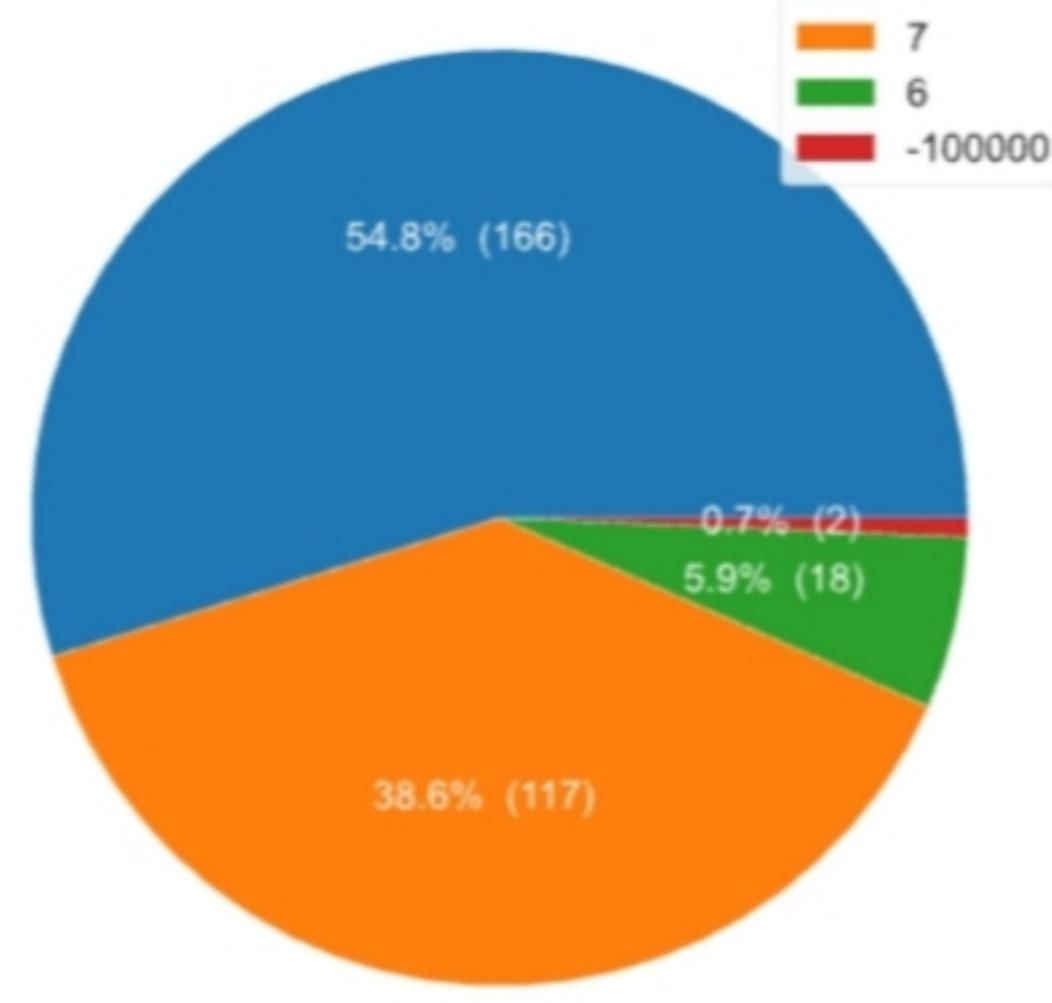
Value	Count	Frequency (%)
3	166	54.8%
7	117	38.6%
6	18	5.9%
-100000	2	0.7%

Length



Histogram of lengths of the category

Pie chart



Value Count Frequency (%)

3	166	54.8%
7	117	38.6%
6	18	5.9%
100000	2	0.7%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value	Count	Frequency (%)
No values found.		

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Value	Count	Frequency (%)
3	166	54.8%
7	117	38.6%
6	18	5.9%
100000	2	0.7%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value	Count	Frequency (%)
No values found.		

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

num

Categorical

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION

Distinct	2
Distinct (%)	0.7%
Missing	0
Missing (%)	0.0%

Memory size 2.5 KiB

0	164
1	139

- Overview
- Categories
- Words
- Characters

Length

Max length	1
Median length	1
Mean length	1
Min length	1

Characters and Unicode

Total characters	0
Distinct characters	0
Distinct categories	0
Distinct scripts	0
Distinct blocks	0

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0
Unique (%)	0.0%

Sample

1st row	0
2nd row	1
3rd row	1
4th row	0
5th row	0

Common Values

Value	Count	Frequency (%)
0	164	54.1%
1	139	45.9%

Length

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique 0

Unique (%) 0.0%

Sample

1st row 0

2nd row 1

3rd row 1

4th row 0

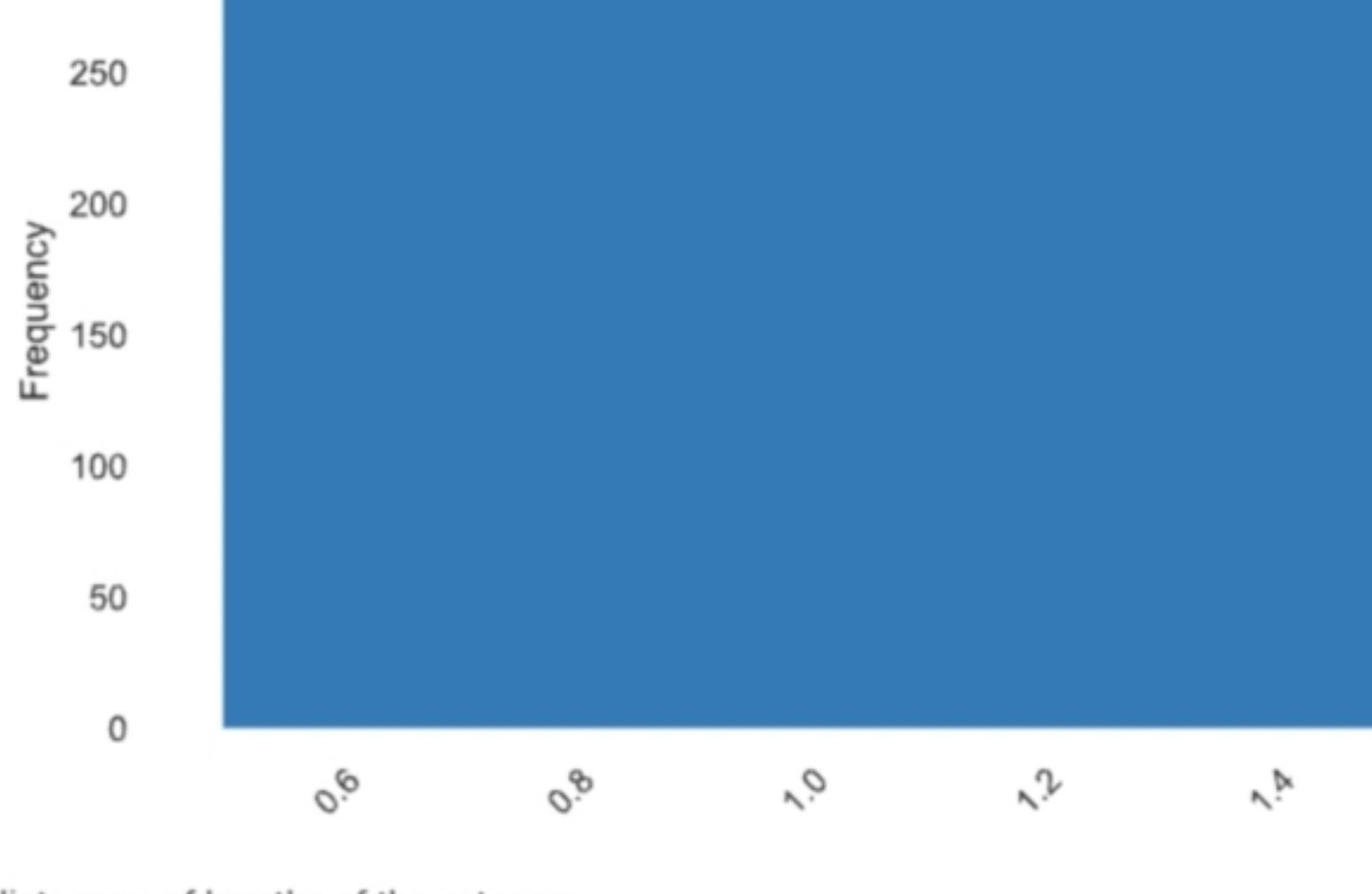
5th row 0

Common Values

Value Count Frequency (%)

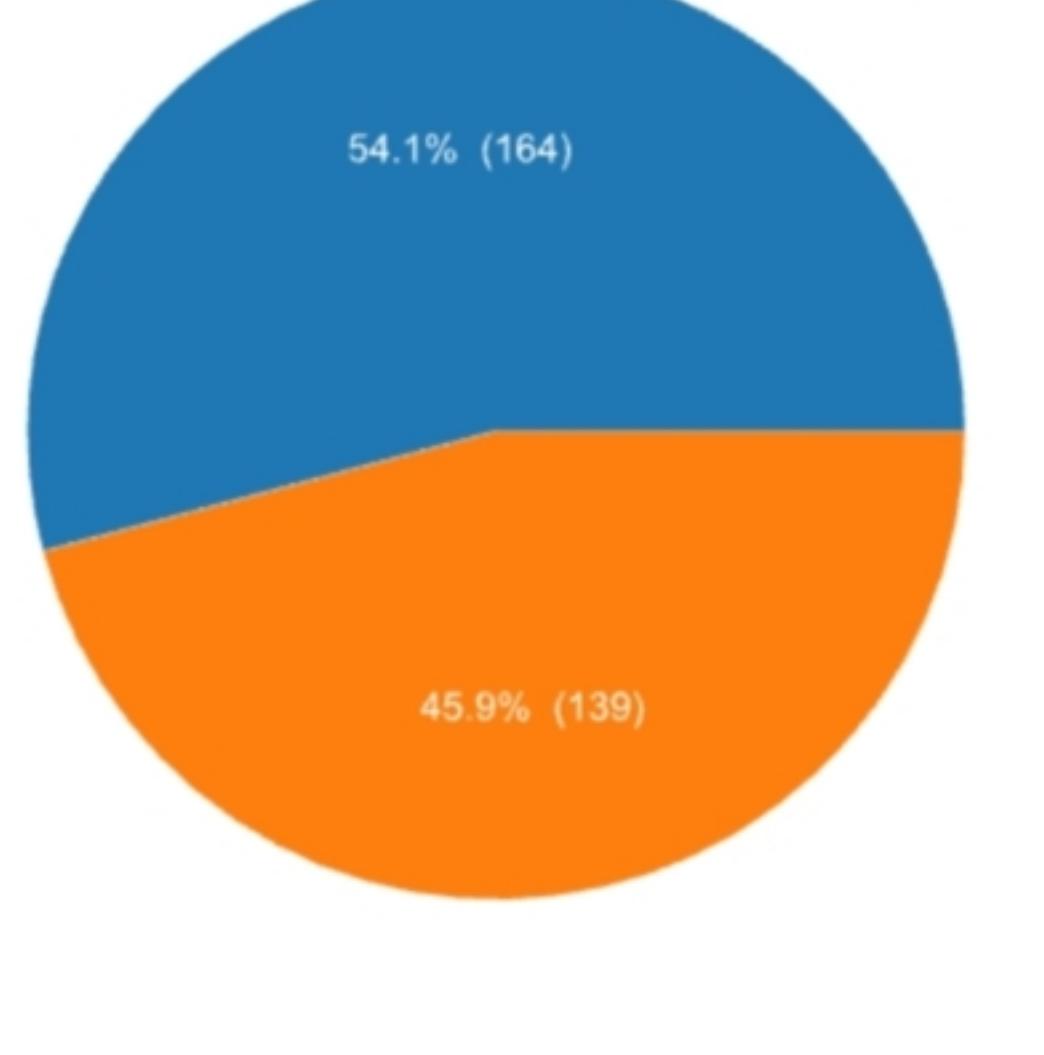
0	164	54.1%
1	139	45.9%

Length



Histogram of lengths of the category

Pie chart



Value Count Frequency (%)

0	164	54.1%
1	139	45.9%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value Count Frequency (%)

No values found.

Most occurring categories

Value Count Frequency (%)

No values found.

Most frequent character per category

Most occurring scripts

Value Count Frequency (%)

No values found.

Most frequent character per script

Value Count Frequency (%)

0	164	54.1%
1	139	45.9%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value Count Frequency (%)

No values found.

Most occurring categories

Value Count Frequency (%)

No values found.

Most frequent character per category

Most occurring scripts

Value Count Frequency (%)

No values found.

Most frequent character per script

Most occurring blocks

Value Count Frequency (%)

No values found.

Most frequent character per block

Heart_Disease

Categorical

HIGH CORRELATION

HIGH CORRELATION

Distinct 2

Distinct (%) 0.7%

Missing 0

Missing (%) 0.0%

Memory size 2.5 KiB

Absence 164

Presence 139

- Overview
- Categories
- Words
- Characters

Length

Max length 8

Median length 7

Mean length 7.458745875

Min length 7

Characters and Unicode

Total characters 0

Distinct characters 0

Distinct categories 0 ⓘ

Distinct scripts 0 ⓘ

Distinct blocks 0 ⓘ

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique 0 ⓘ

Unique (%) 0.0%

Sample

1st row Absence

2nd row Presence

3rd row Presence

4th row Absence

5th row Absence

Common Values

Value Count Frequency (%)

Absence 164 54.1%

Presence 139 45.9%

Length

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique 0 

Unique (%) 0.0%

Sample

1st row Absence

2nd row Presence

3rd row Presence

4th row Absence

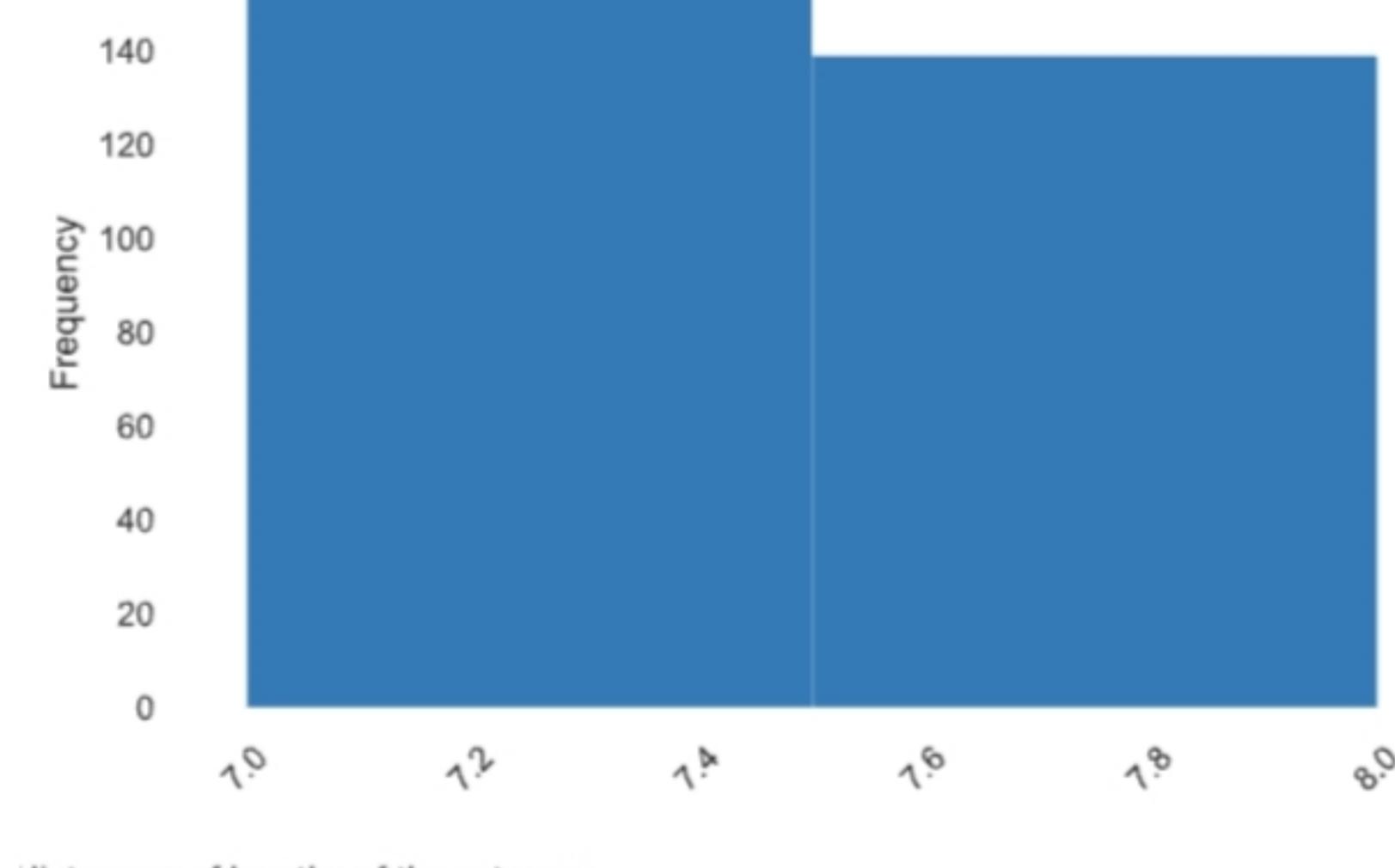
5th row Absence

Common Values

Value	Count	Frequency (%)
-------	-------	---------------

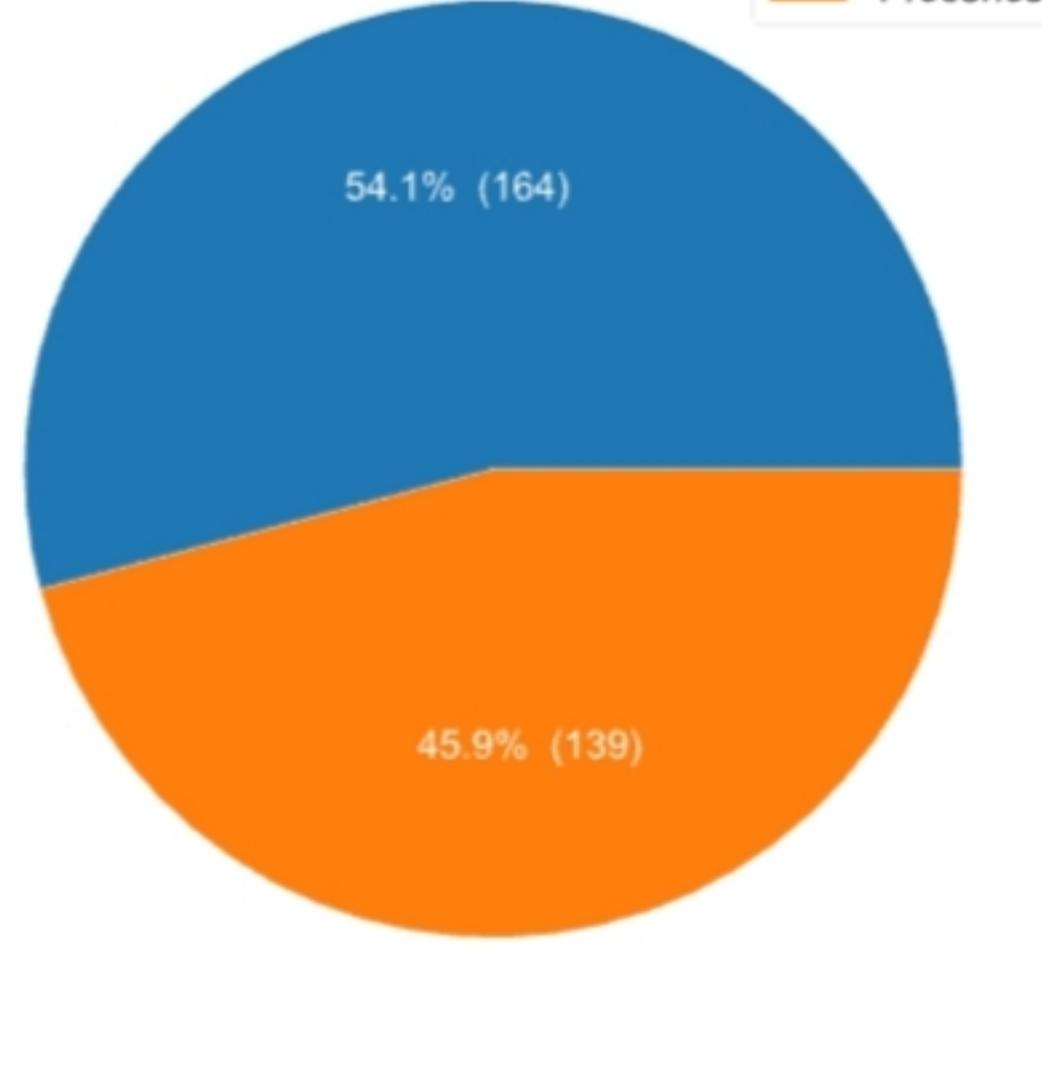
Absence	164	54.1%
Presence	139	45.9%

Length



Histogram of lengths of the category

Pie chart



Value	Count	Frequency (%)
-------	-------	---------------

Absence	164	54.1%
presence	139	45.9%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value	Count	Frequency (%)
-------	-------	---------------

No values found.

Most occurring categories

Value	Count	Frequency (%)
-------	-------	---------------

No values found.

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
-------	-------	---------------

No values found.

Value	Count	Frequency (%)
absence	164	54.1%
presence	139	45.9%

- [Characters](#)
- [Categories](#)
- [Scripts](#)
- [Blocks](#)

Most occurring characters

Value	Count	Frequency (%)
No values found.		

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

[sex1](#)

Categorical

HIGH CORRELATION
HIGH CORRELATION

Distinct	2
Distinct (%)	0.7%
Missing	0
Missing (%)	0.0%

Memory size 2.5 KiB

Male 206

Female 97

- [Overview](#)
- [Categories](#)
- [Words](#)
- [Characters](#)

Length

Max length	6
Median length	4
Mean length	4.640264026
Min length	4

Characters and Unicode

Total characters	0
Distinct characters	0
Distinct categories	0
Distinct scripts	0
Distinct blocks	0

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Junique

Unique 0
Unique (%) 0.0%

Sample

1st row Male
2nd row Male
3rd row Male
4th row Male
5th row Female

Common Values

Value	Count	Frequency (%)
Male	206	68.0%
Female	97	32.0%

Length

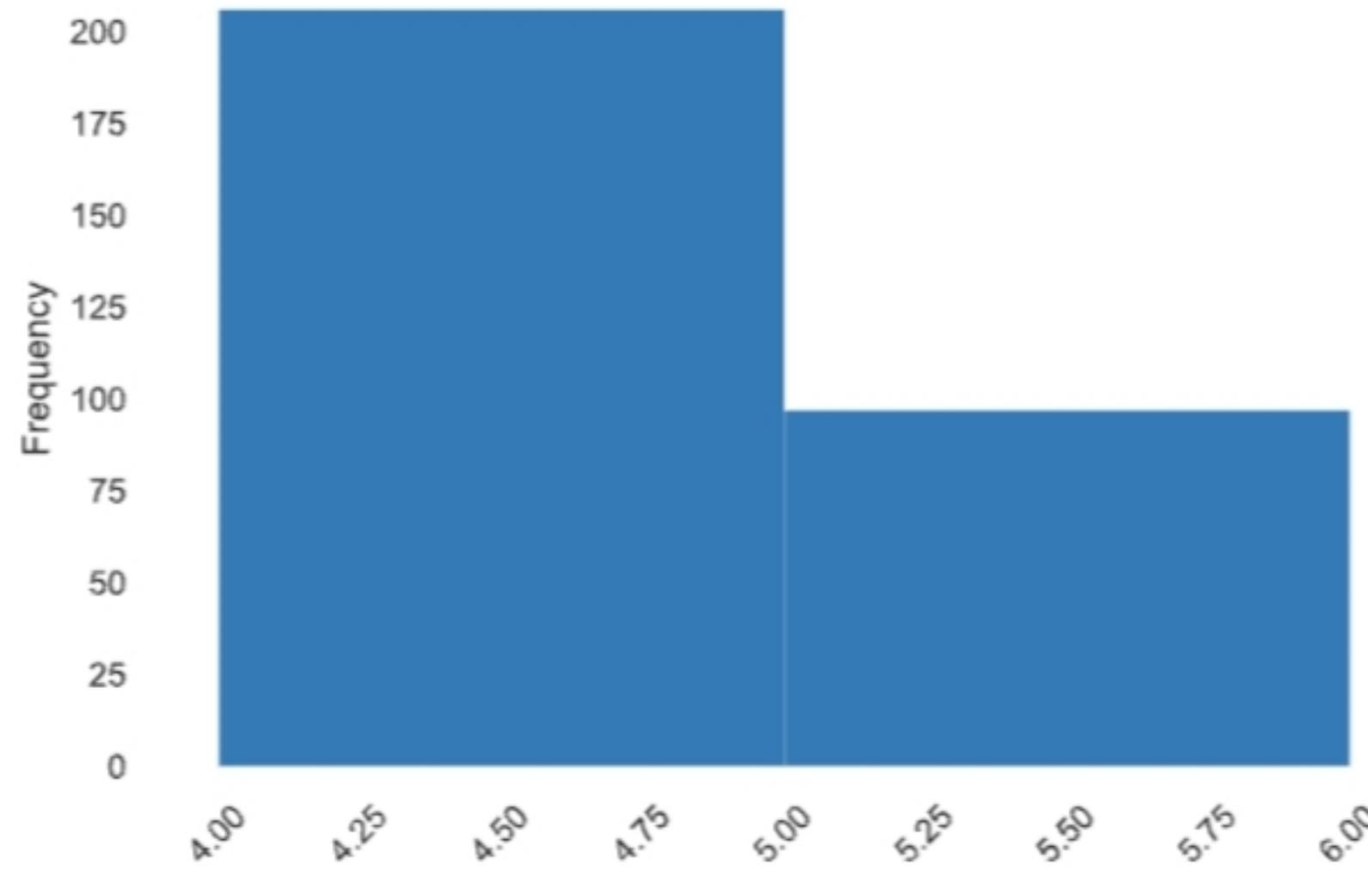
200

175

Common Values

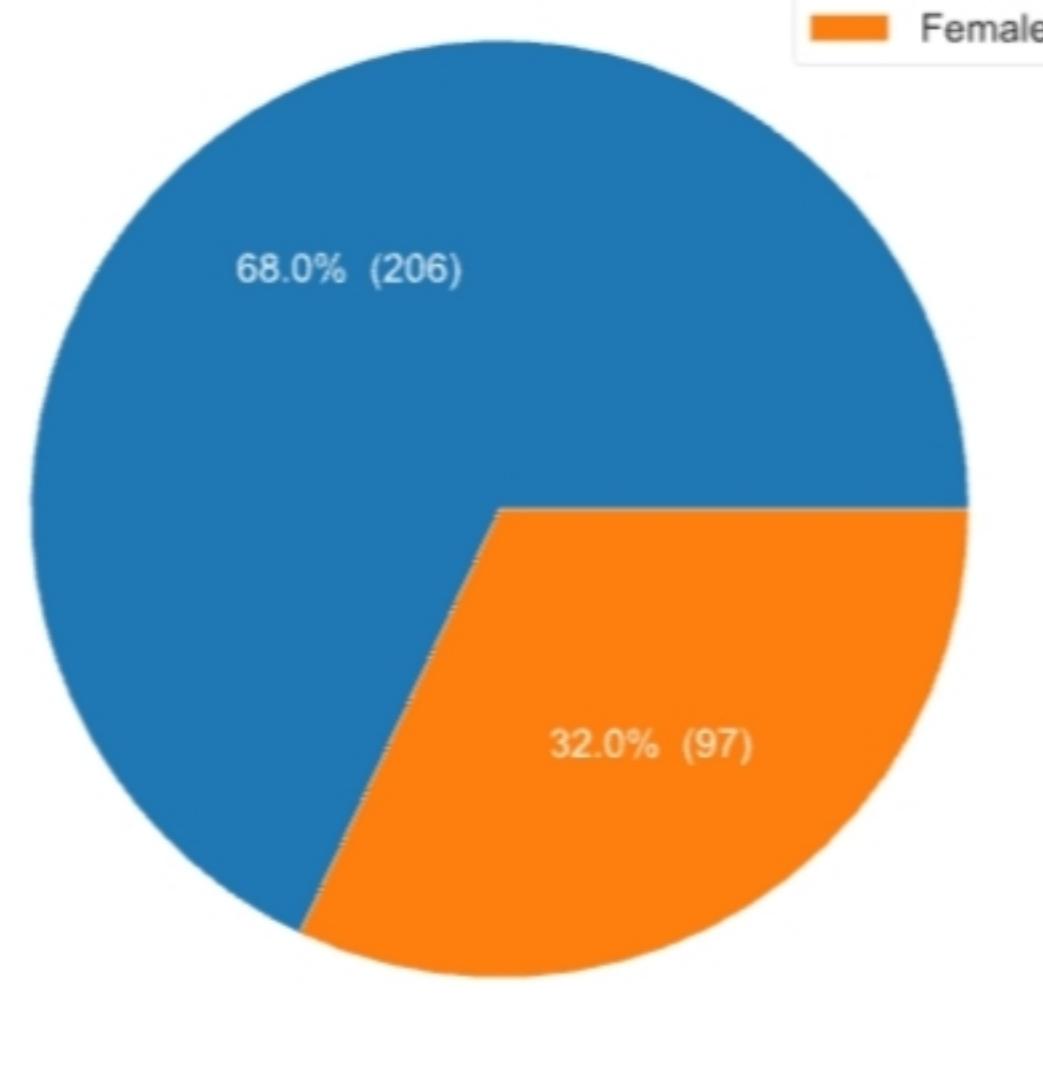
Value	Count	Frequency (%)
Male	206	68.0%
Female	97	32.0%

Length



Histogram of lengths of the category

Pie chart



Value	Count	Frequency (%)
male	206	68.0%
female	97	32.0%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value	Count	Frequency (%)
No values found.		

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

Age_Range

Categorical

HIGH CORRELATION

MISSING

Distinct	3
Distinct (%)	1.0%

Most frequent character per block

Age_Range

Categorical

HIGH CORRELATION

MISSING

Distinct 3

Distinct (%) 1.0%

Missing 8

Missing (%) 2.6%

Memory size 2.5 KiB

Elder Age 152

Middle Age 128

Young Age 15

- Overview

- Categories

- Words

- Characters

Length

Max length 10

Median length 9

Mean length 9.433898305

Min length 9

Characters and Unicode

Total characters 0

Distinct characters 0

Distinct categories 0

Distinct scripts 0

Distinct blocks 0

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique 0

Unique (%) 0.0%

Sample

1st row Elder Age

2nd row Elder Age

3rd row Elder Age

4th row Young Age

5th row Middle Age

Common Values

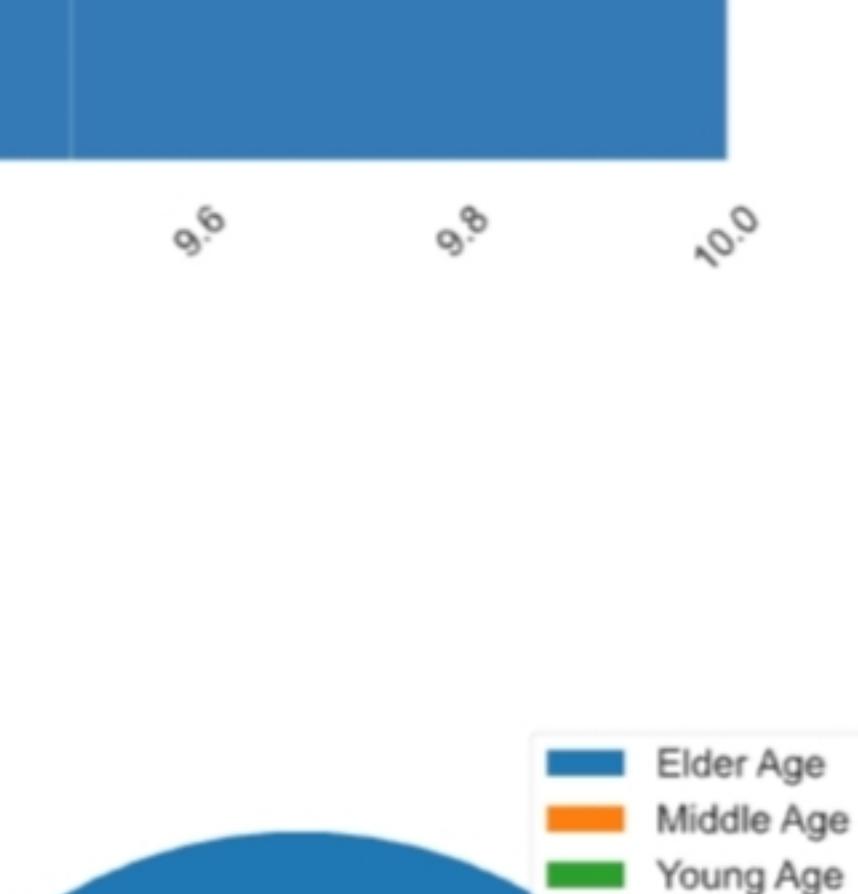
Value	Count	Frequency (%)
Elder Age	152	50.2%
Middle Age	128	42.2%
Young Age	15	5.0%
(Missing)	8	2.6%

Length



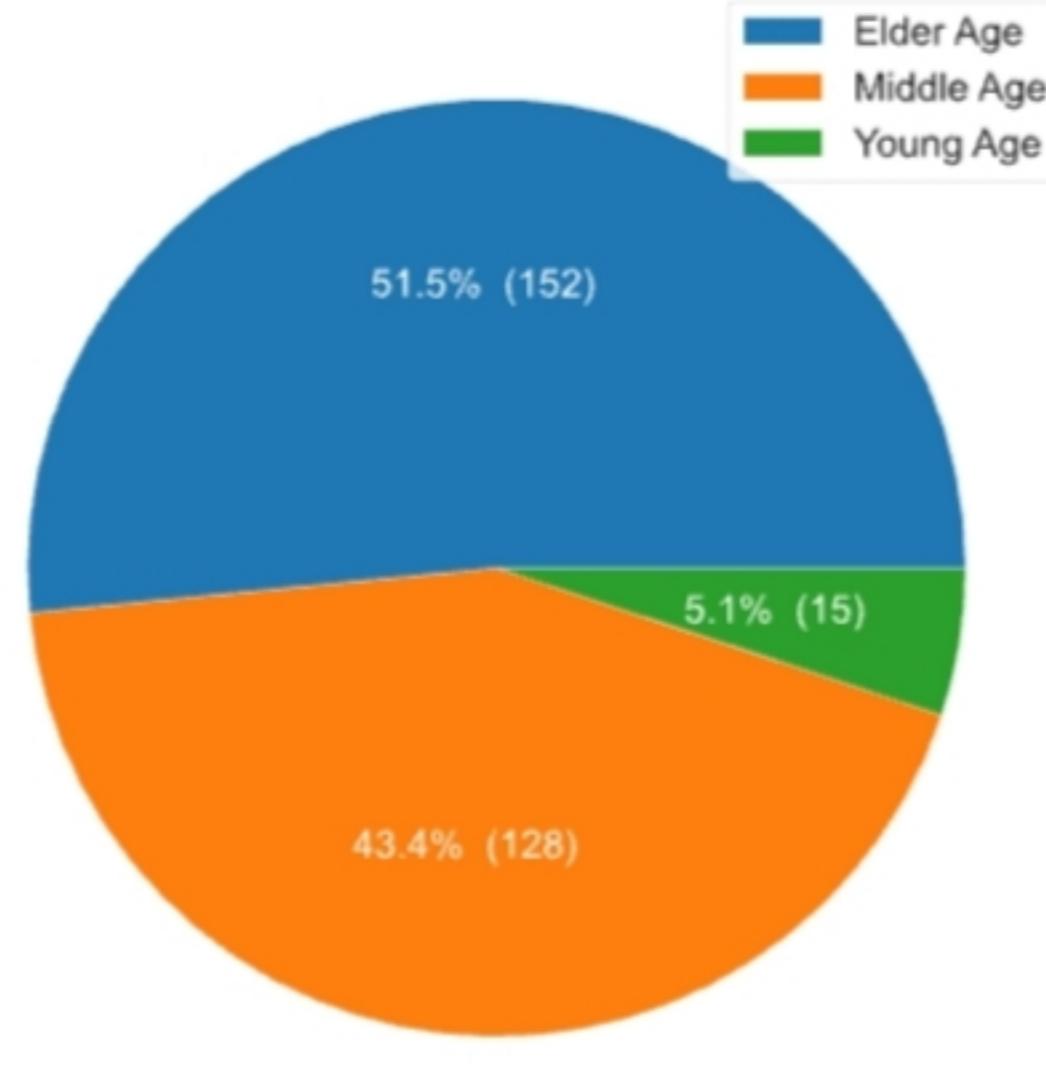
Histogram of lengths of the category

Pie chart



Histogram of lengths of the category

Pie chart



Value Count Frequency (%)

age	295	50.0%
elder	152	25.8%
middle	128	21.7%
young	15	2.5%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value Count Frequency (%)

No values found.

Most occurring categories

Value Count Frequency (%)

No values found.

Most frequent character per category

Most occurring scripts

Value Count Frequency (%)

No values found.

Most frequent character per script

Most occurring blocks

Value Count Frequency (%)

No values found.

Most frequent character per block

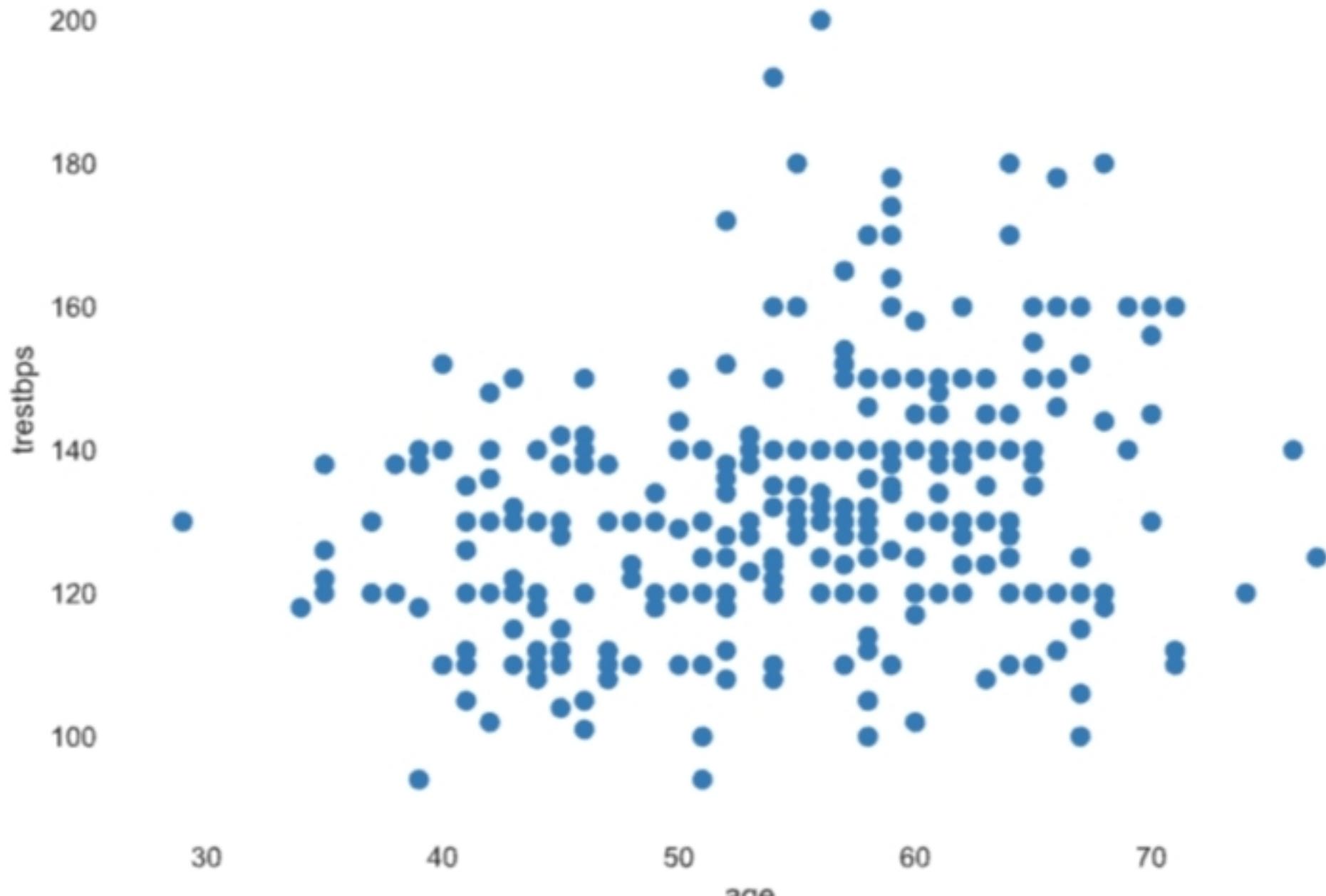
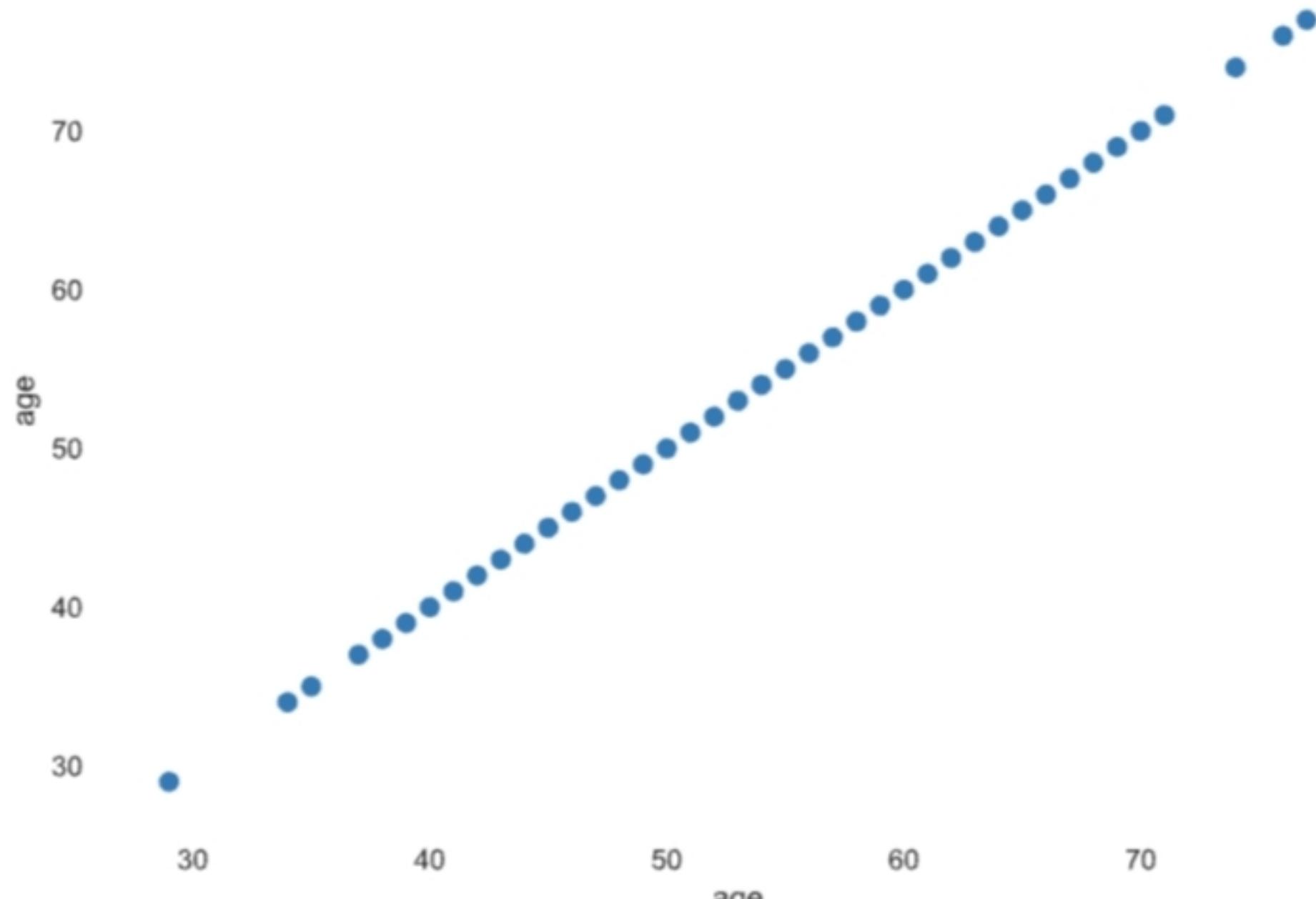
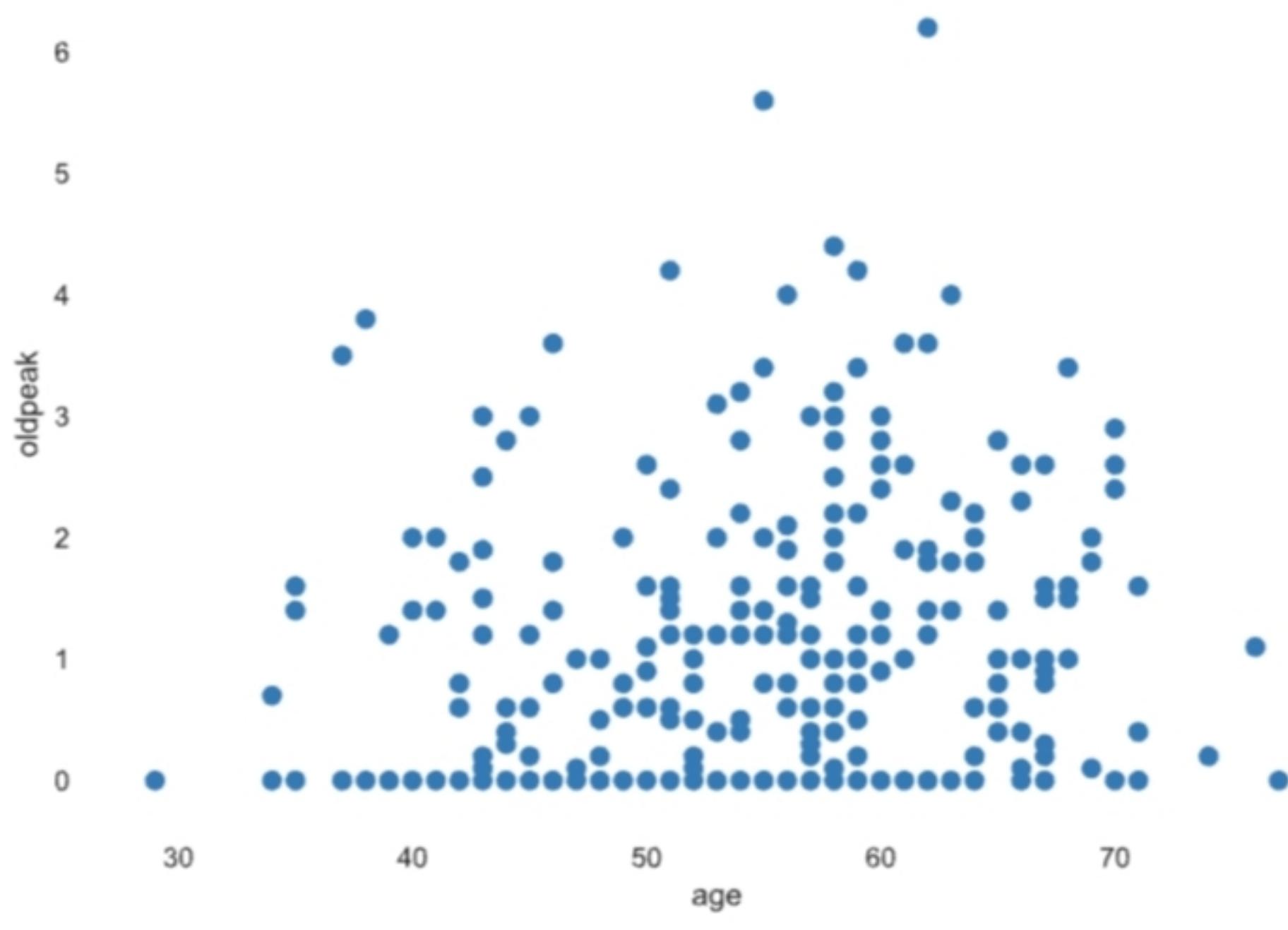
Interactions

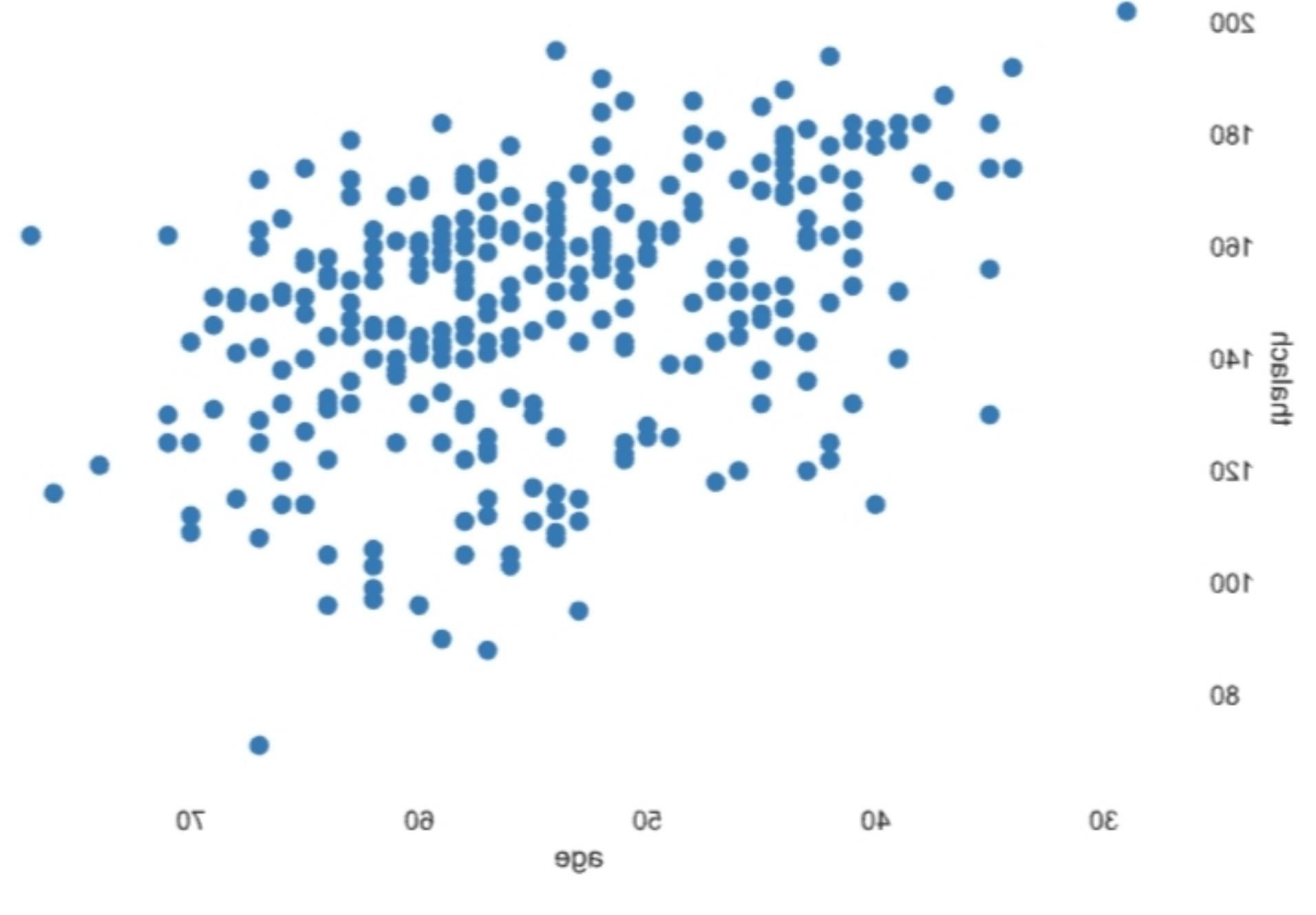
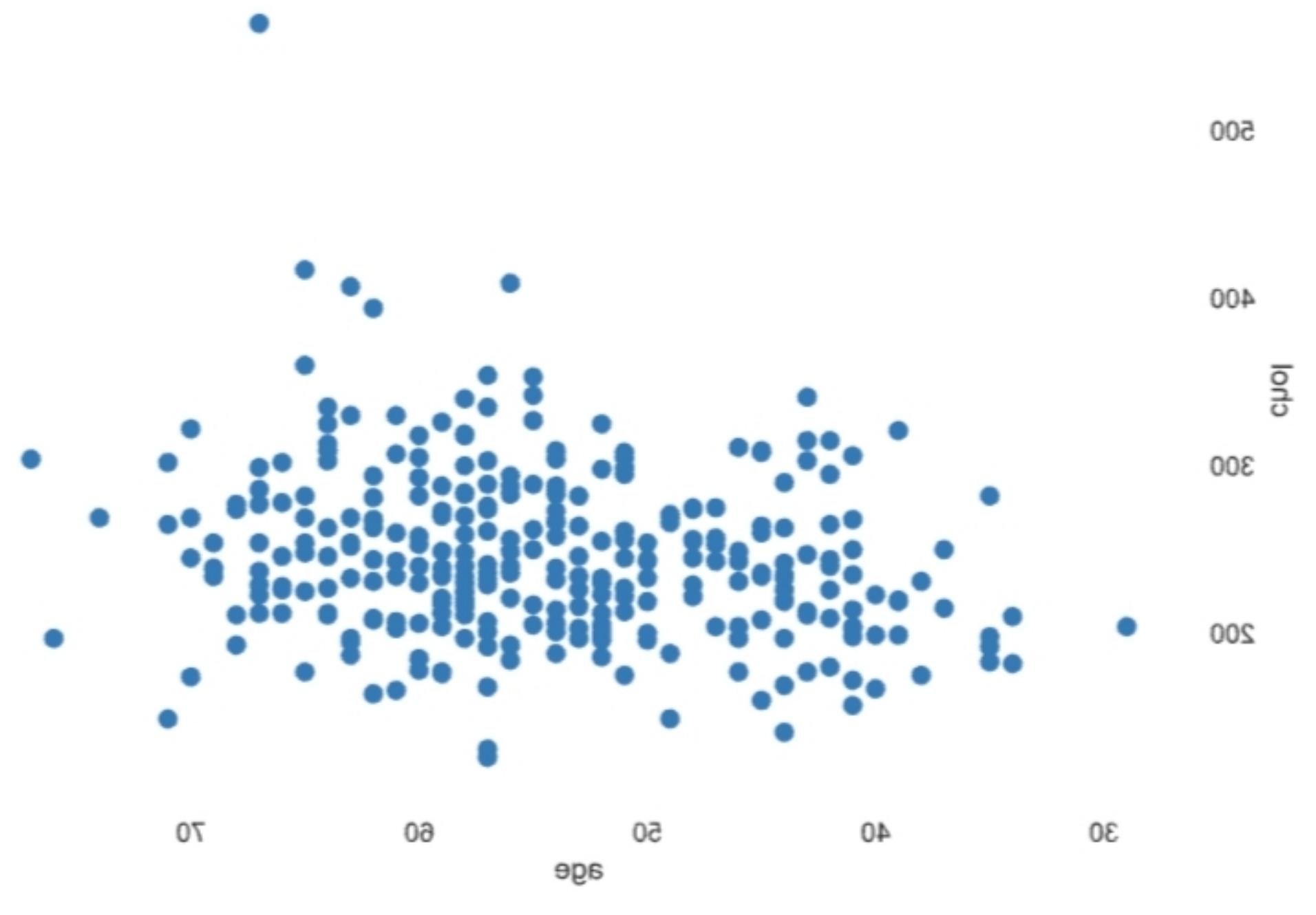
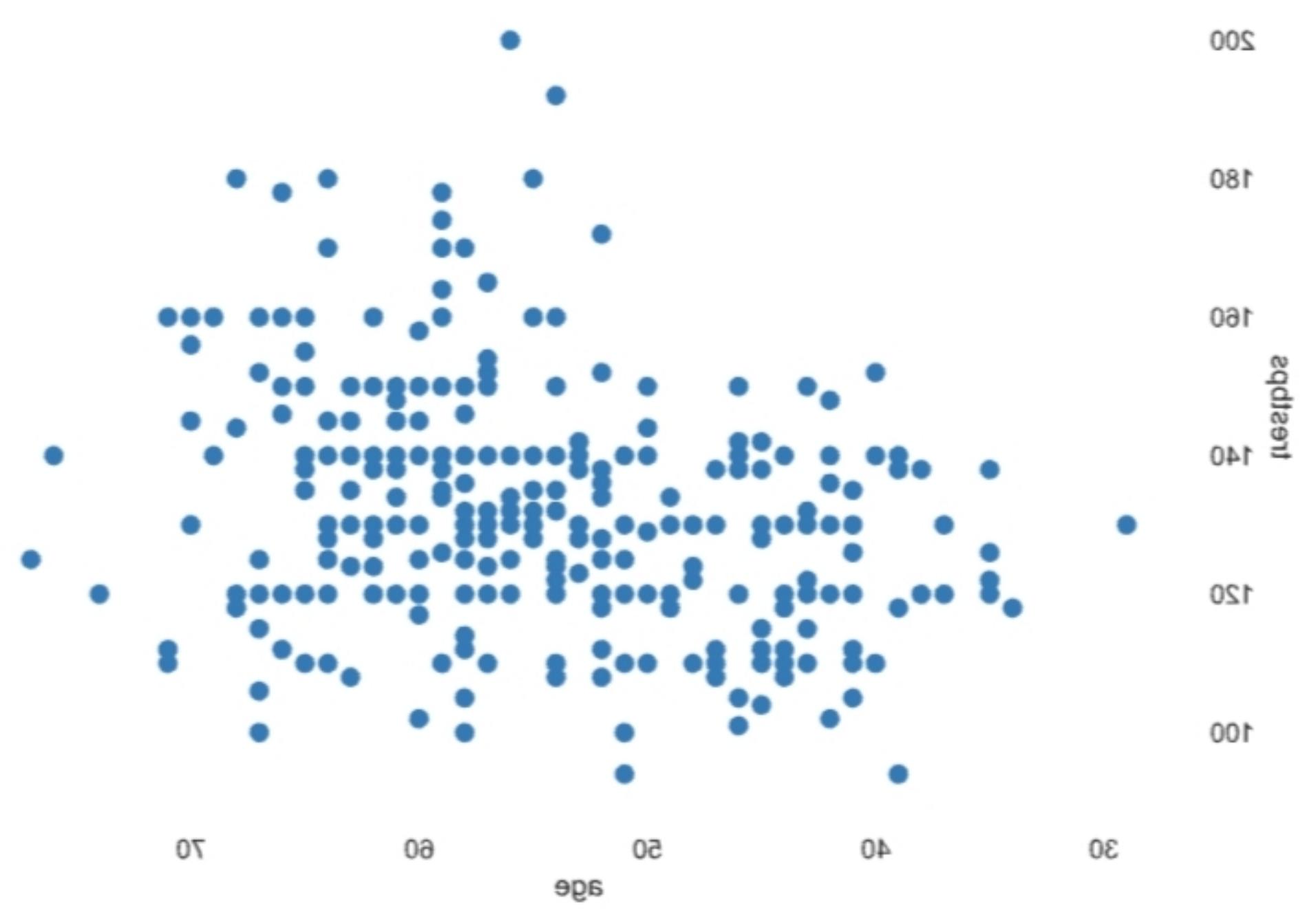
- age
- trestbps
- chol
- thalach
- oldpeak
- oldpeak
- age
- trestbps
- chol
- thalach



Interactions

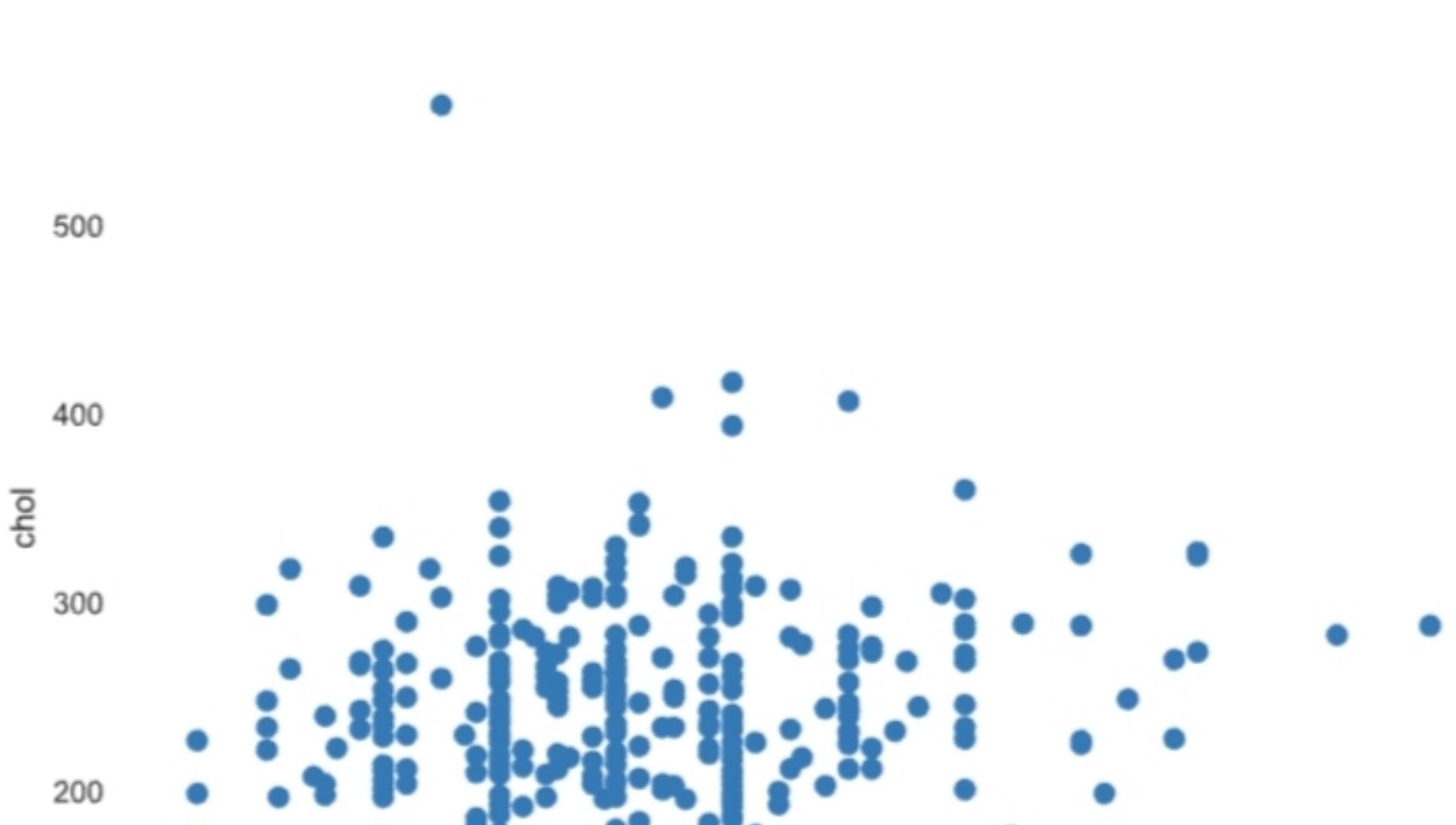
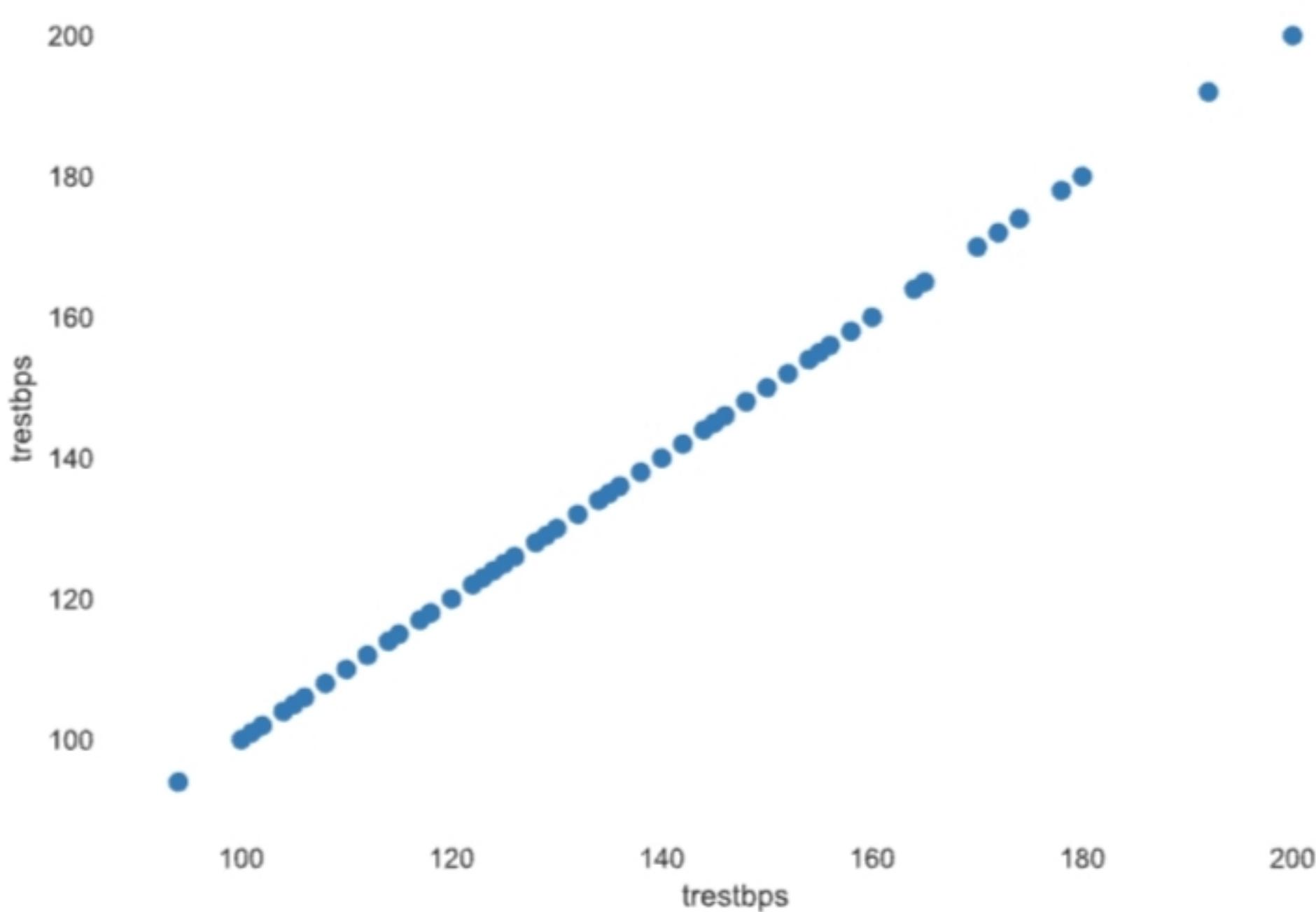
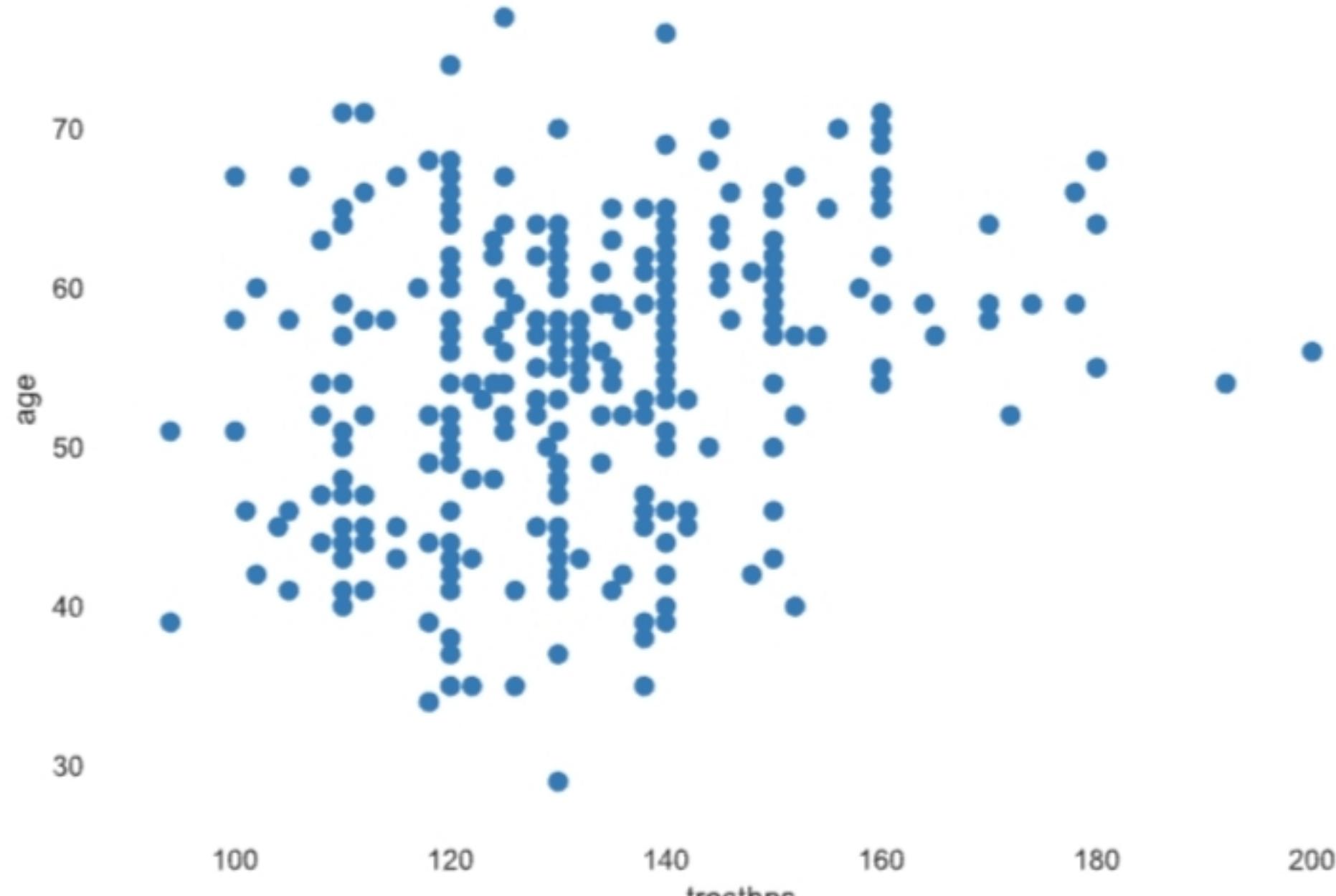
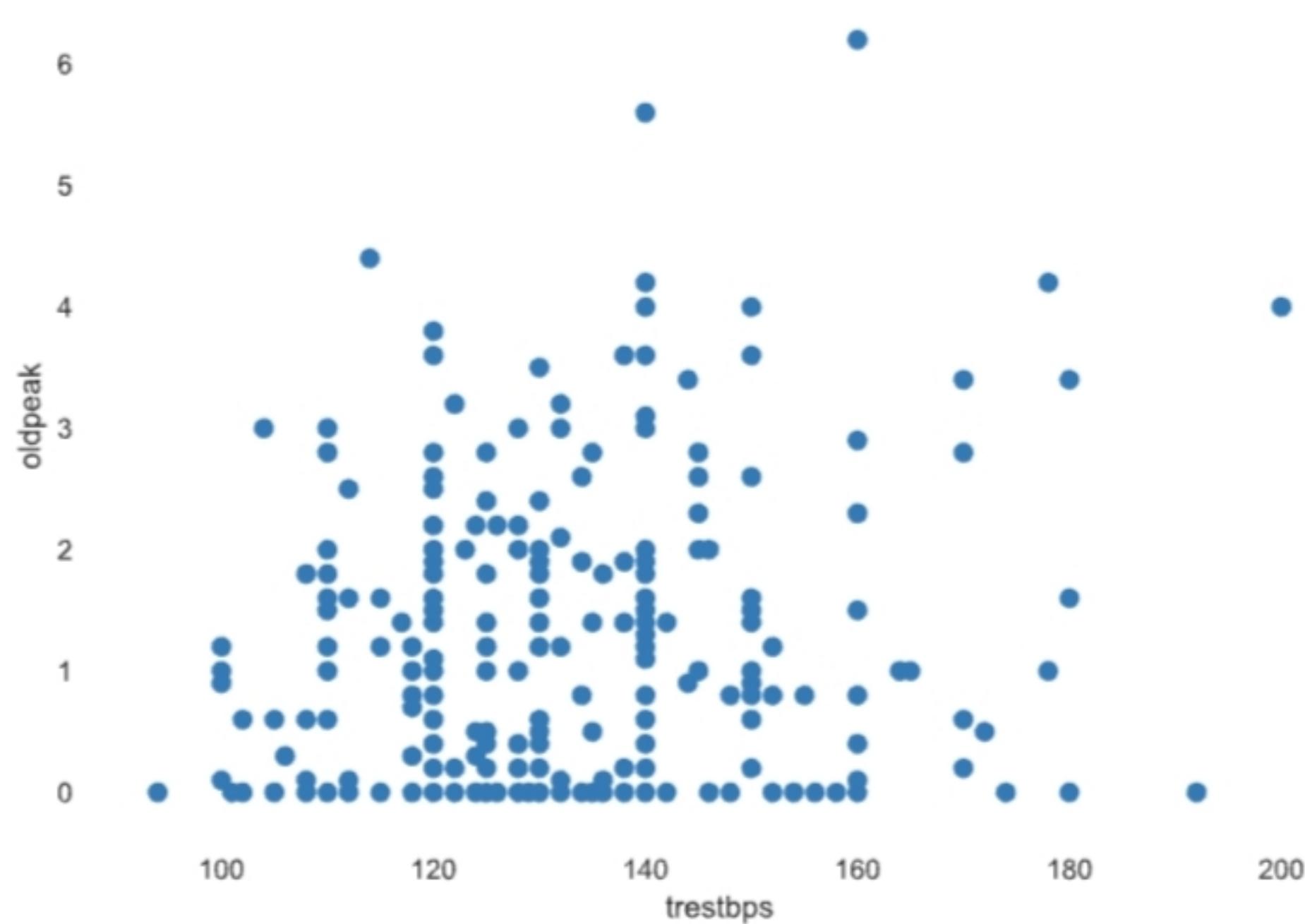
- age
- trestbps
- chol
- thalach
- oldpeak
- oldpeak
- age
- trestbps
- chol
- thalach

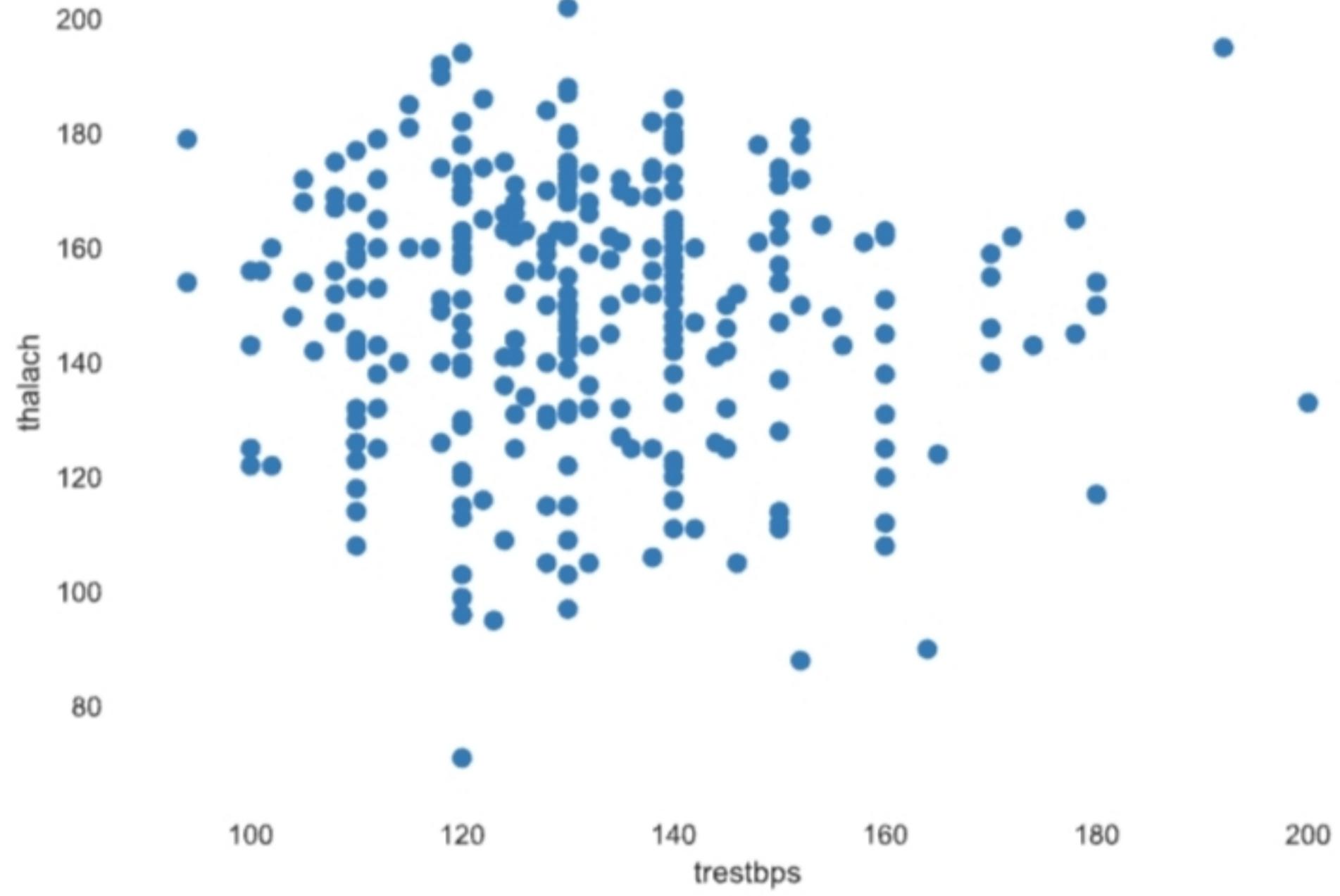
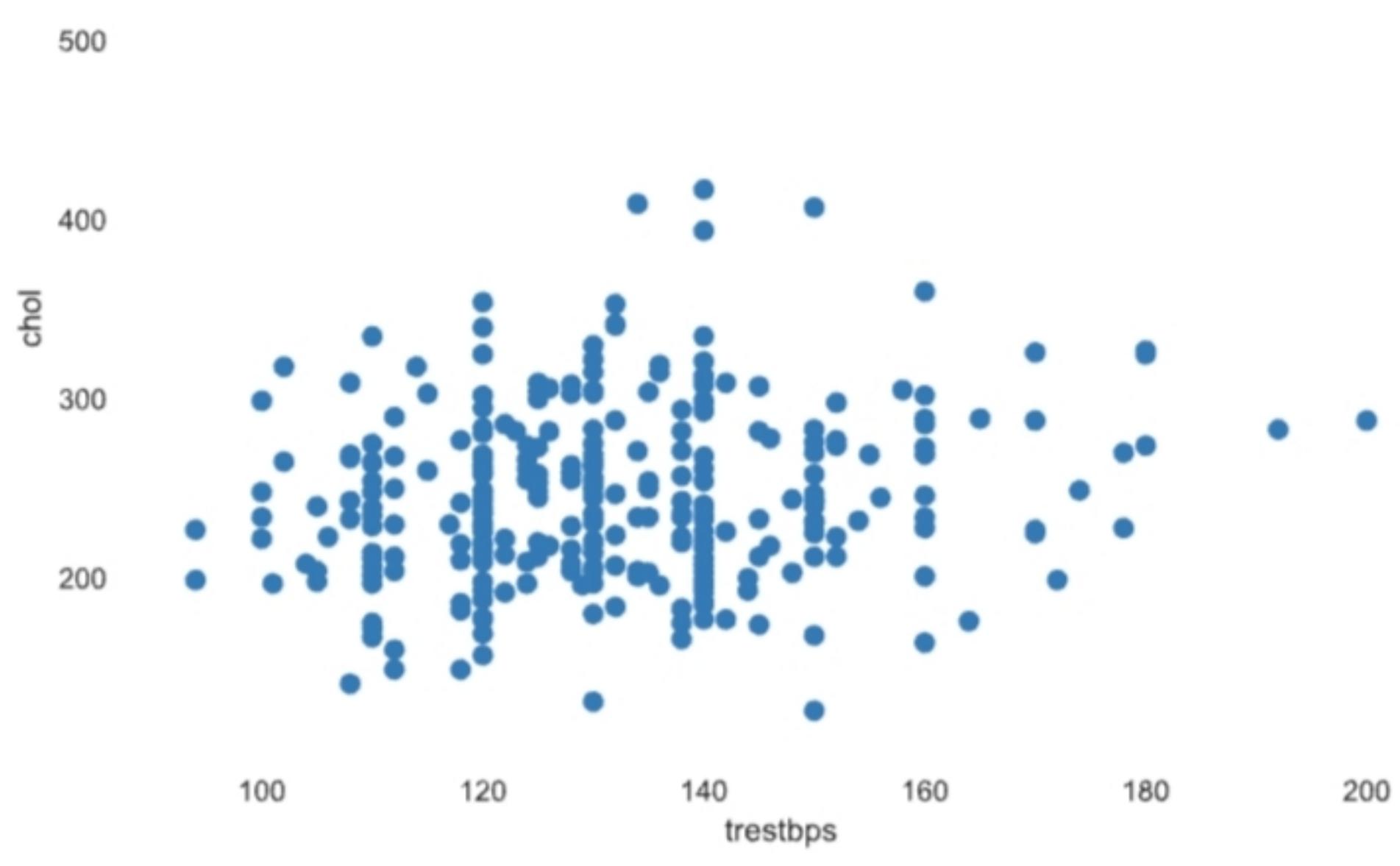




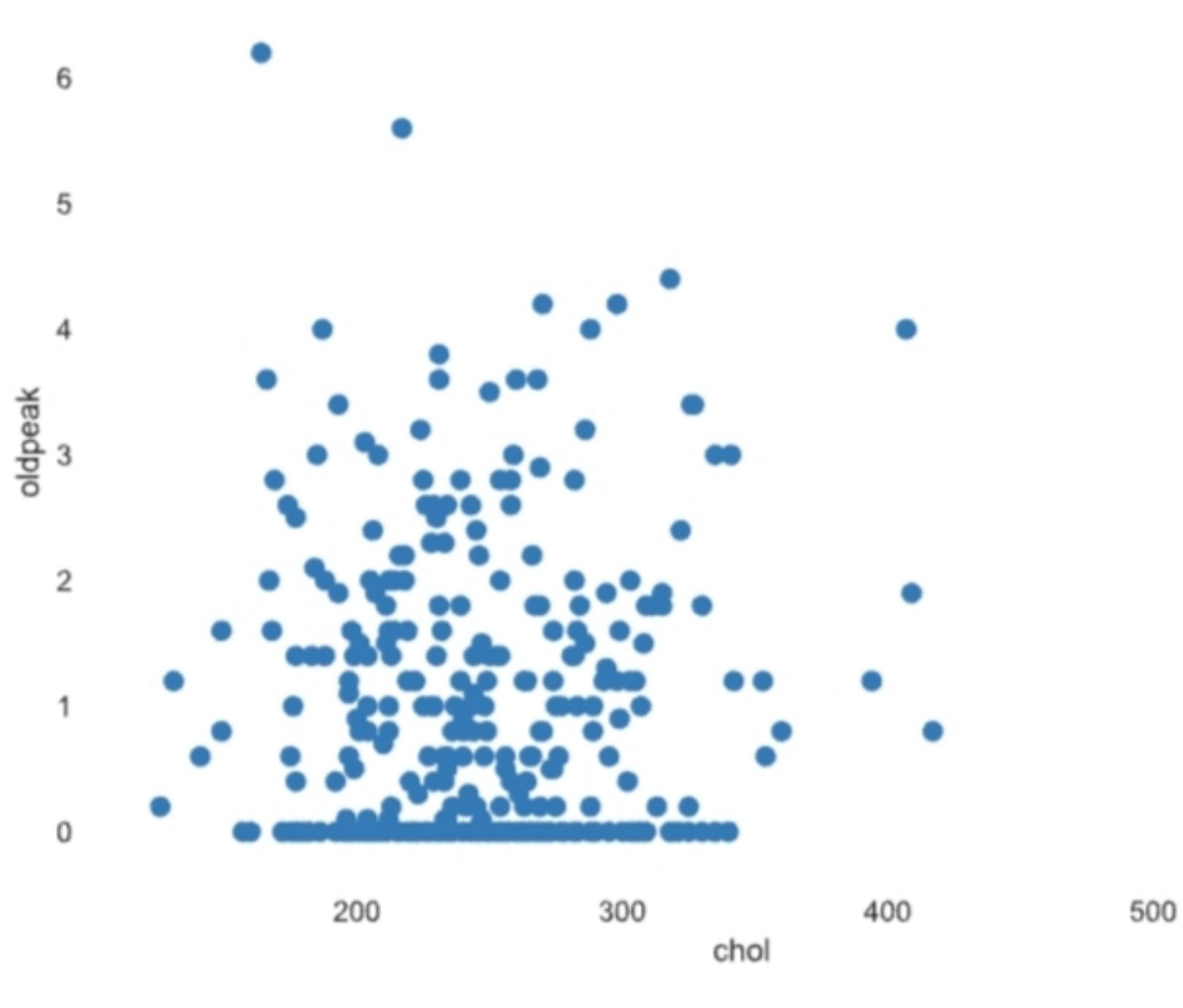
- oldbeak
- age
- testsbps
- chori
- thiasach

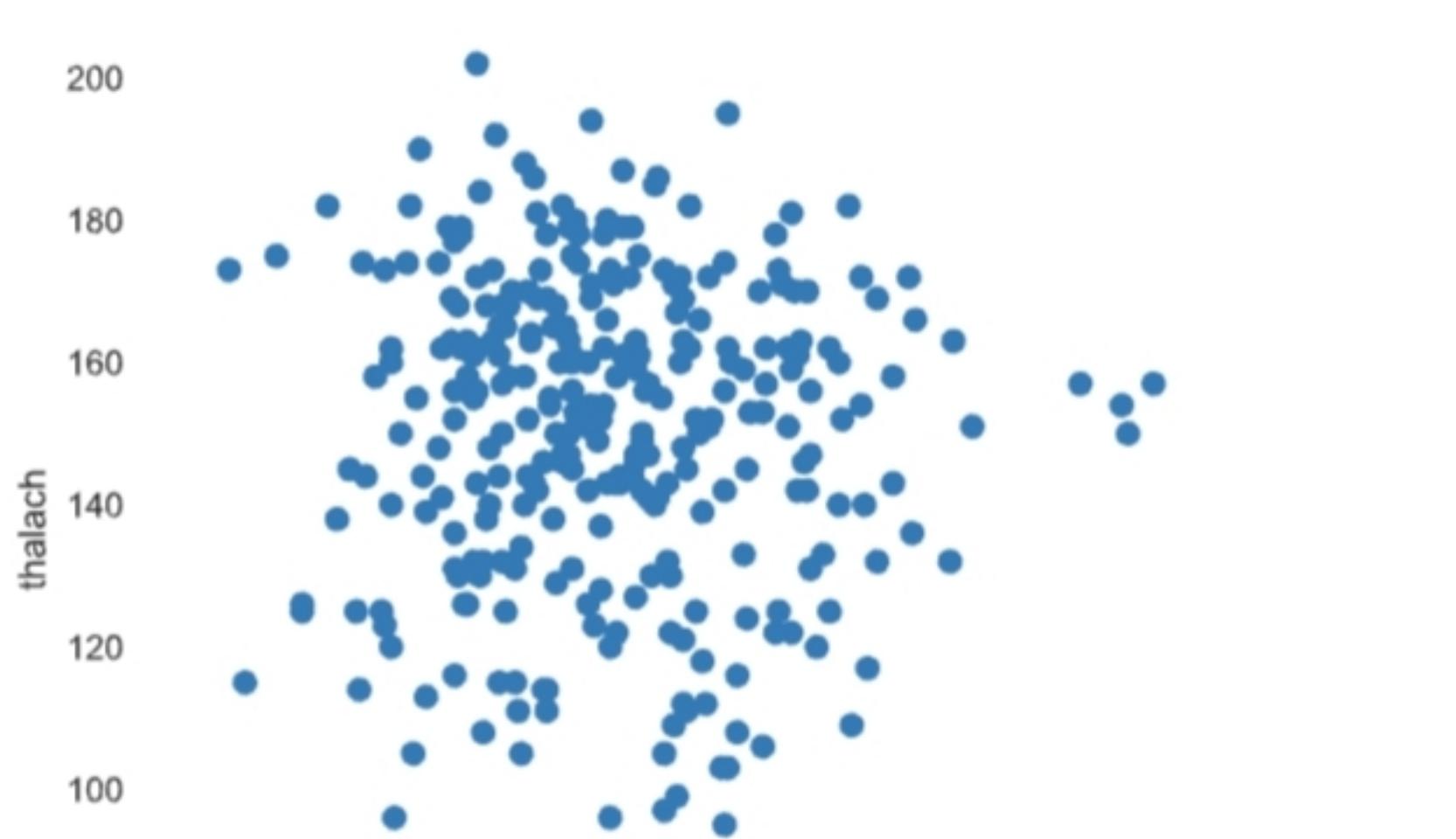
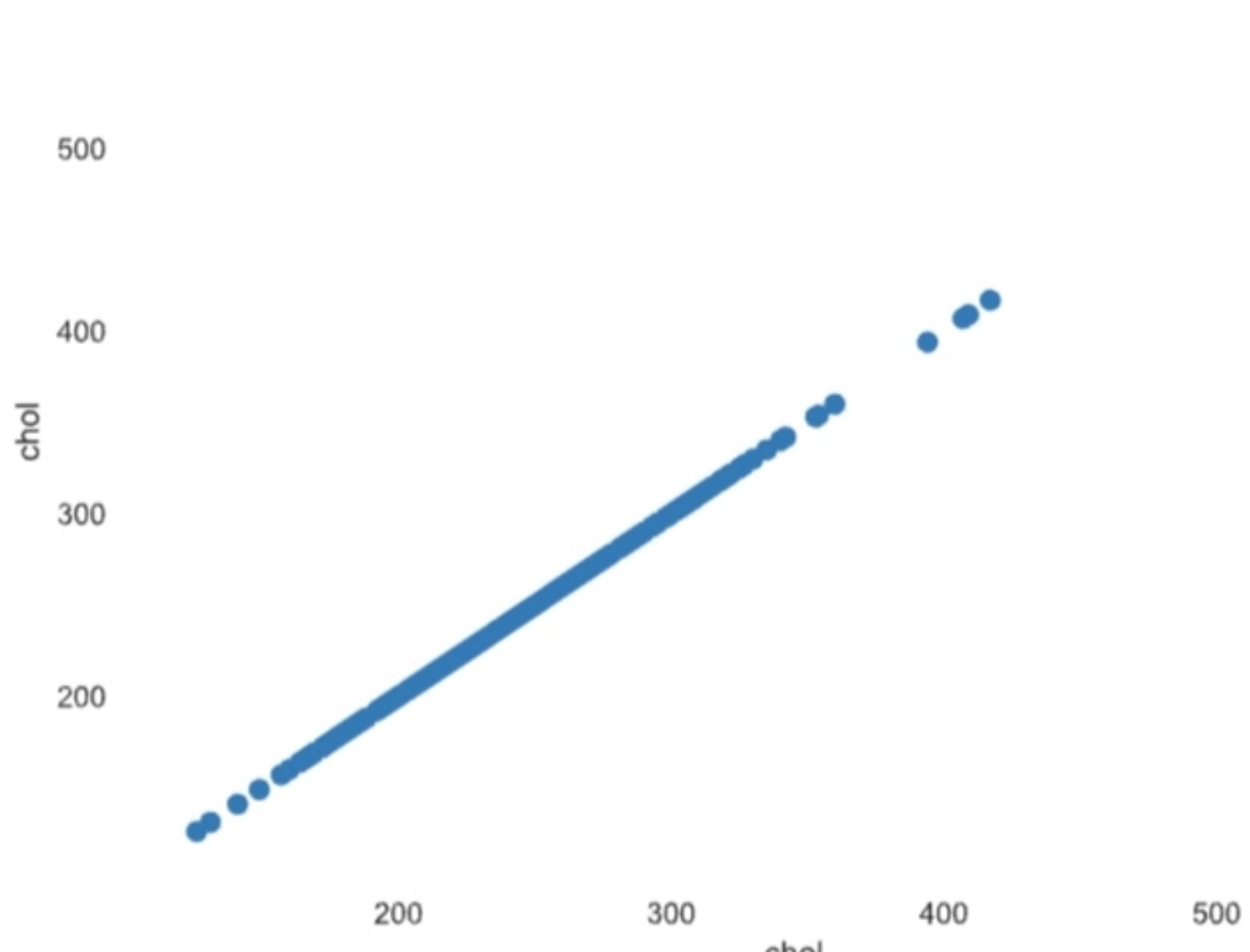
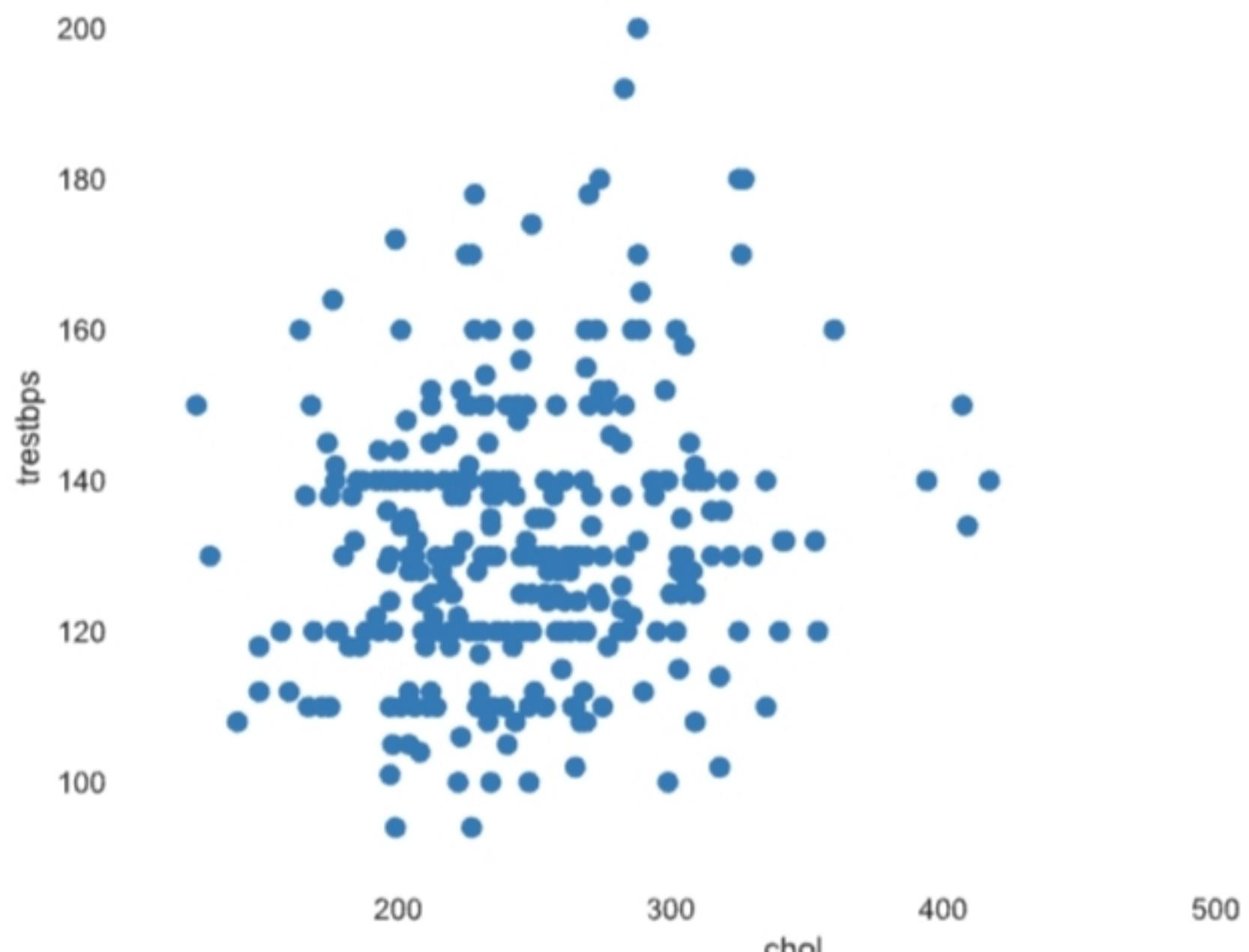
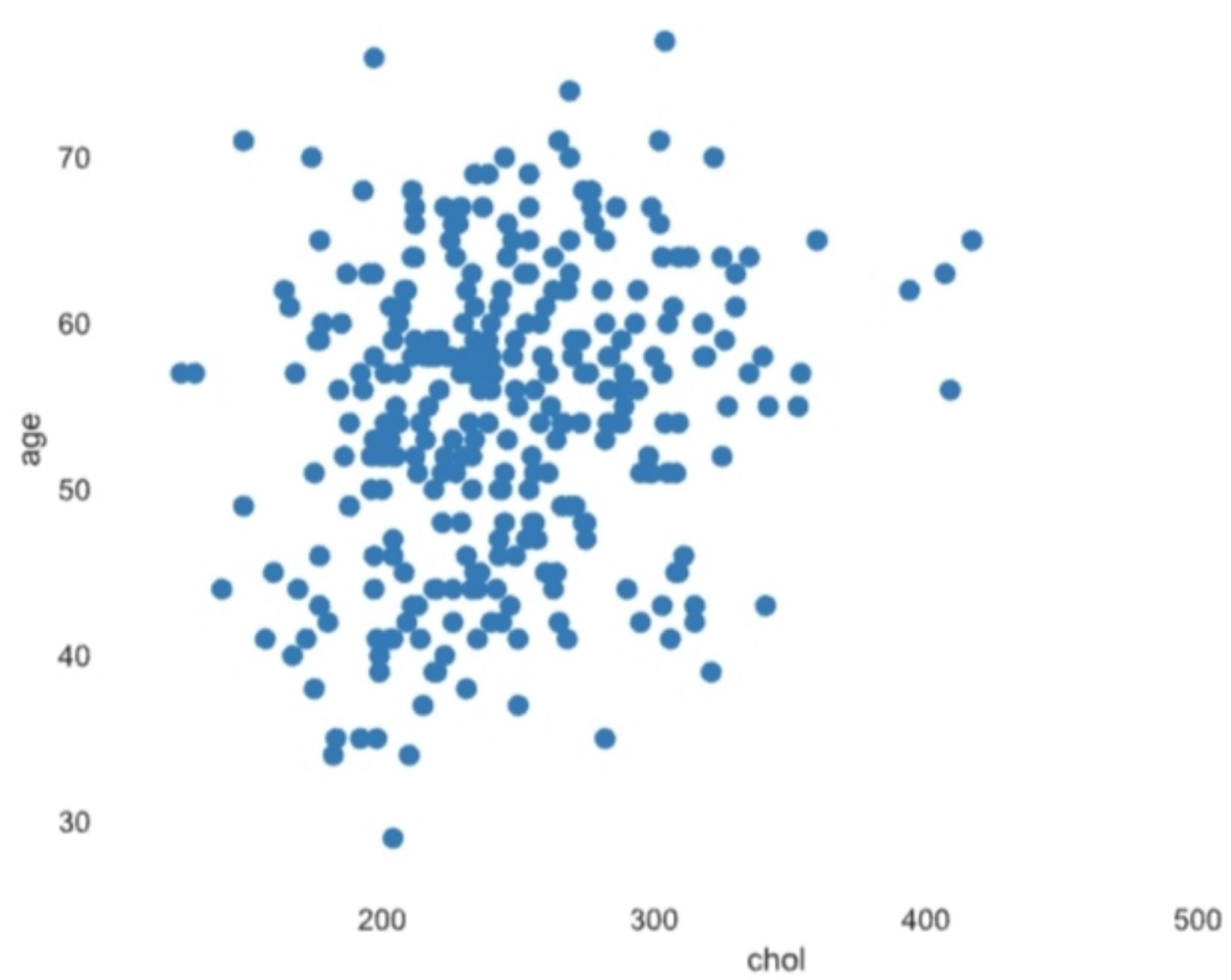


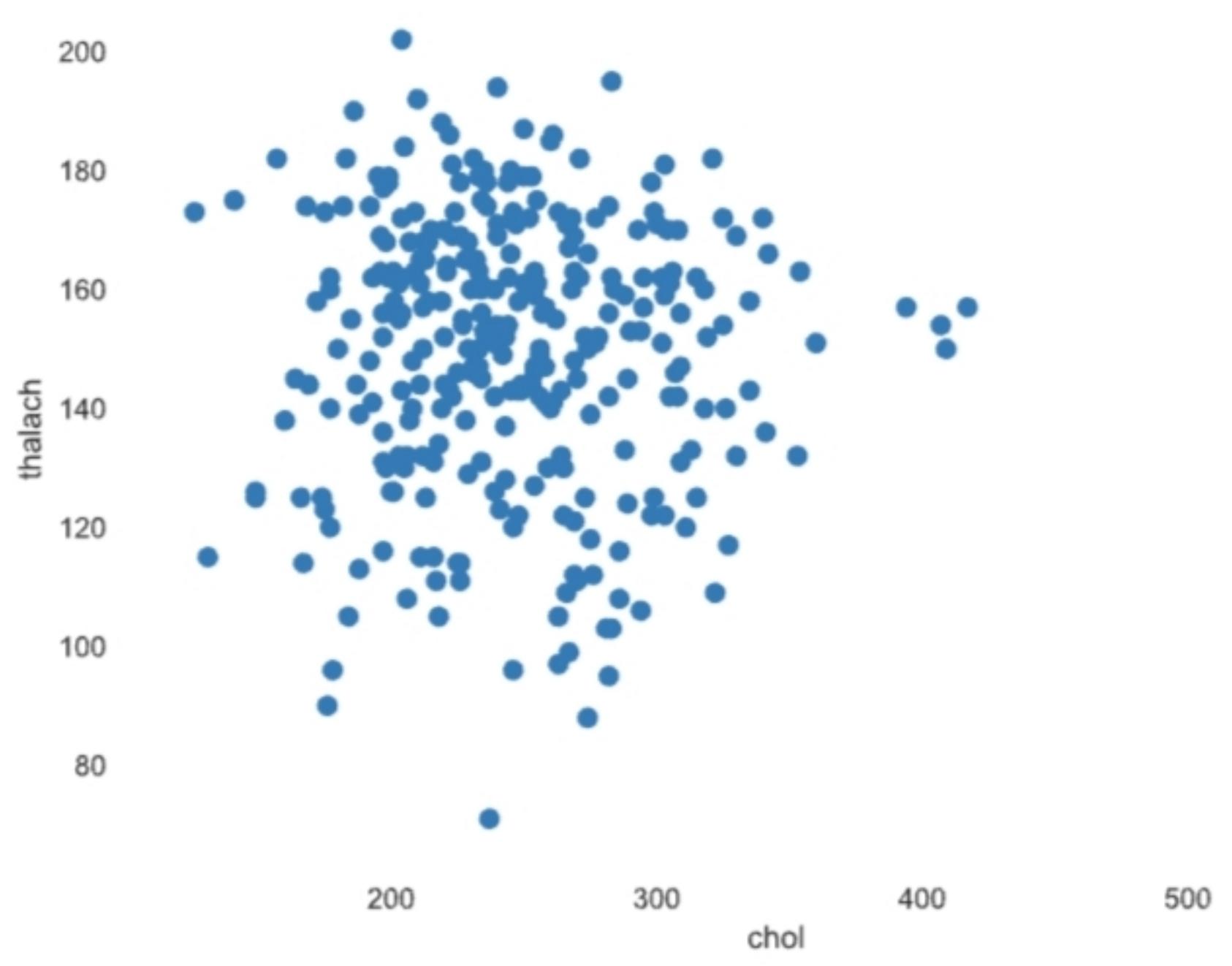




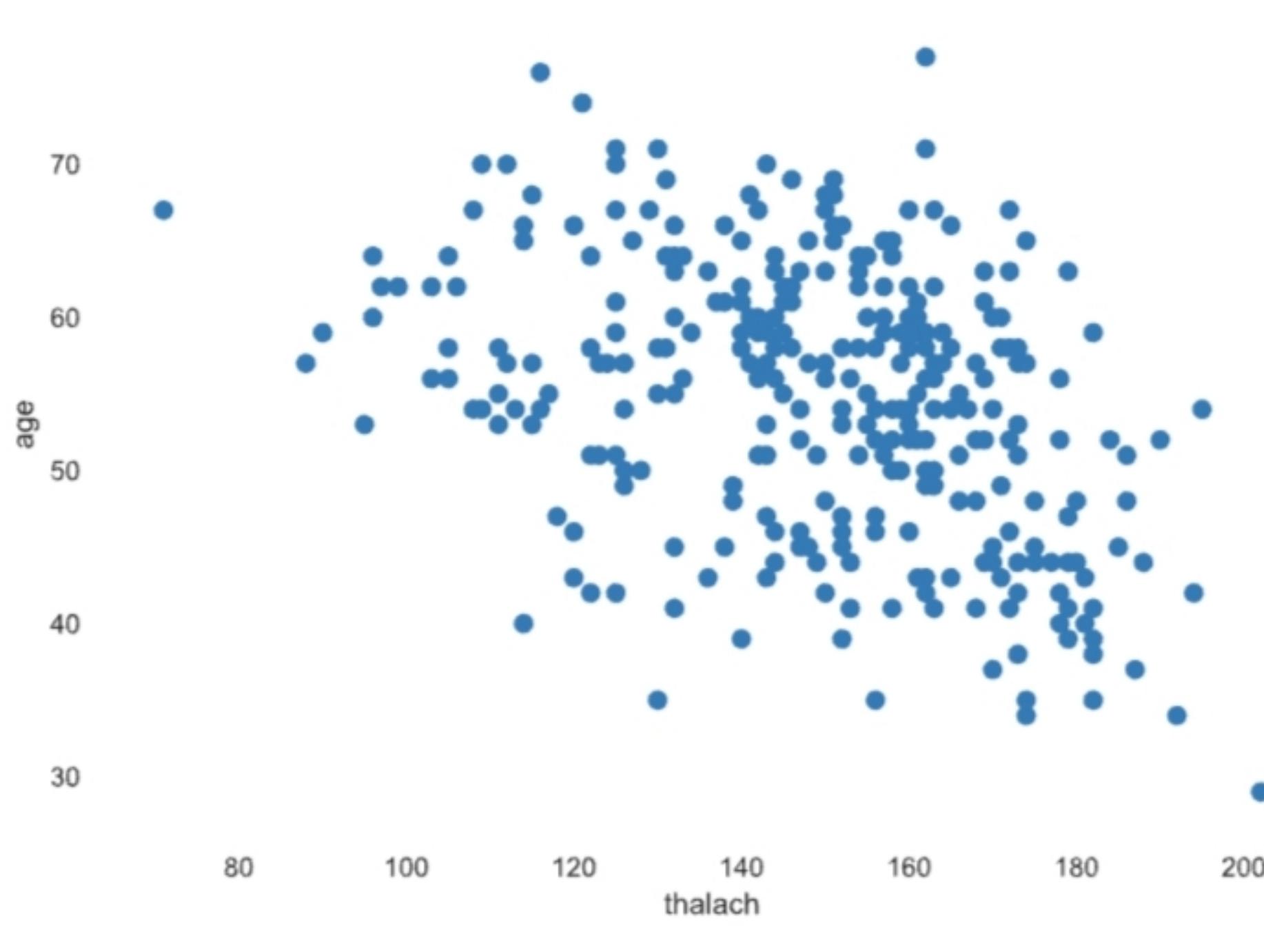
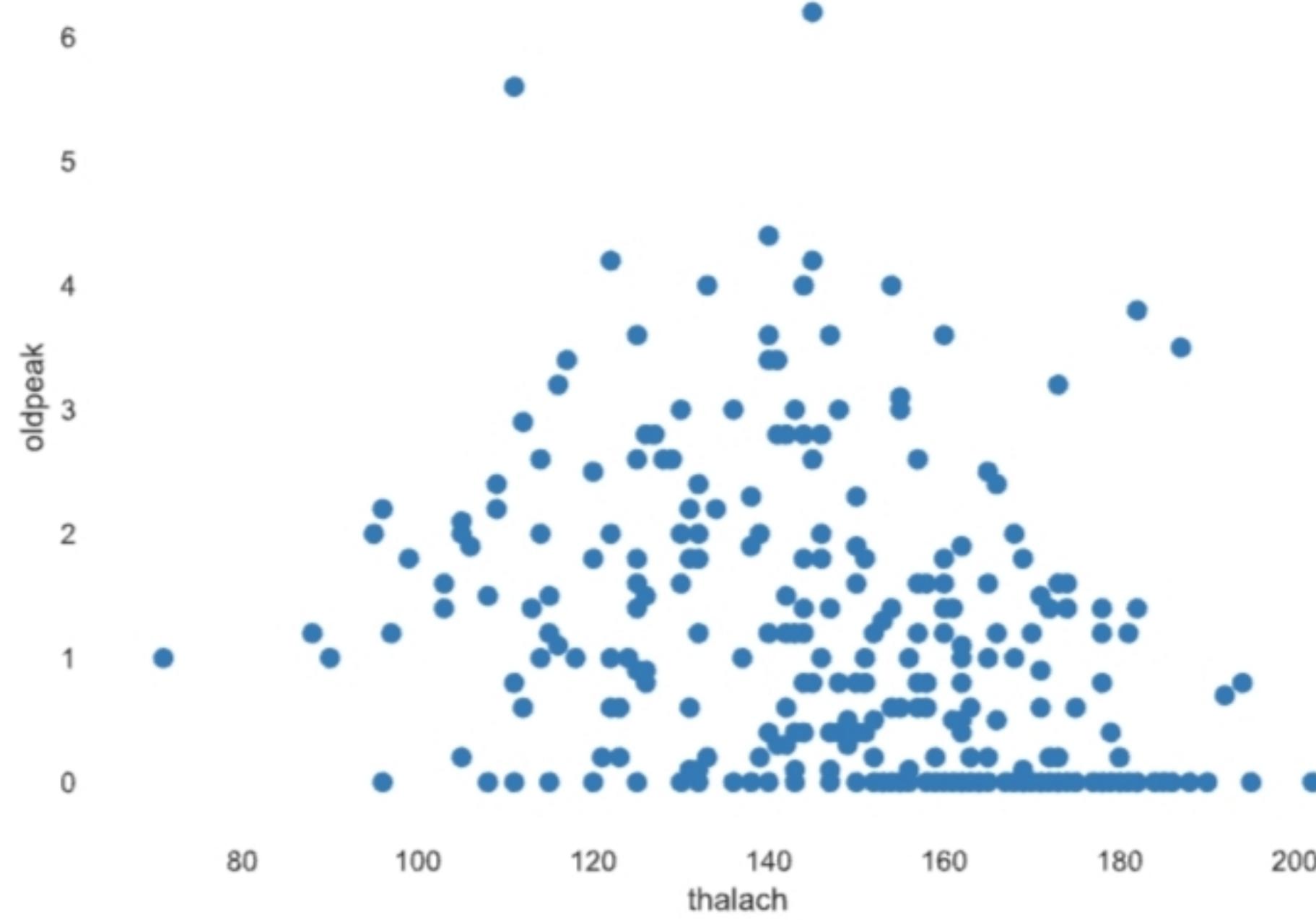
- oldpeak
- age
- trestbps
- chol
- thalach

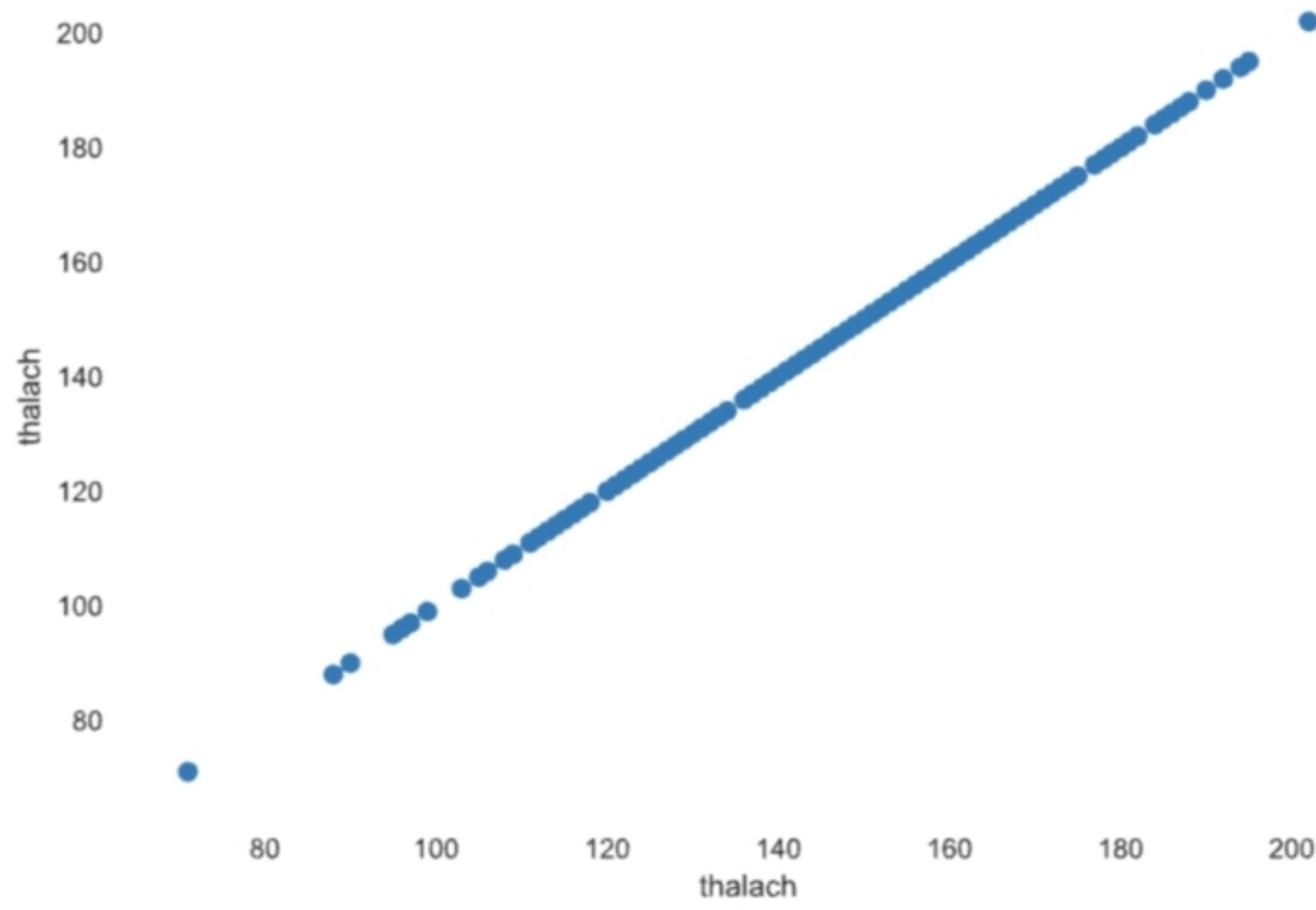
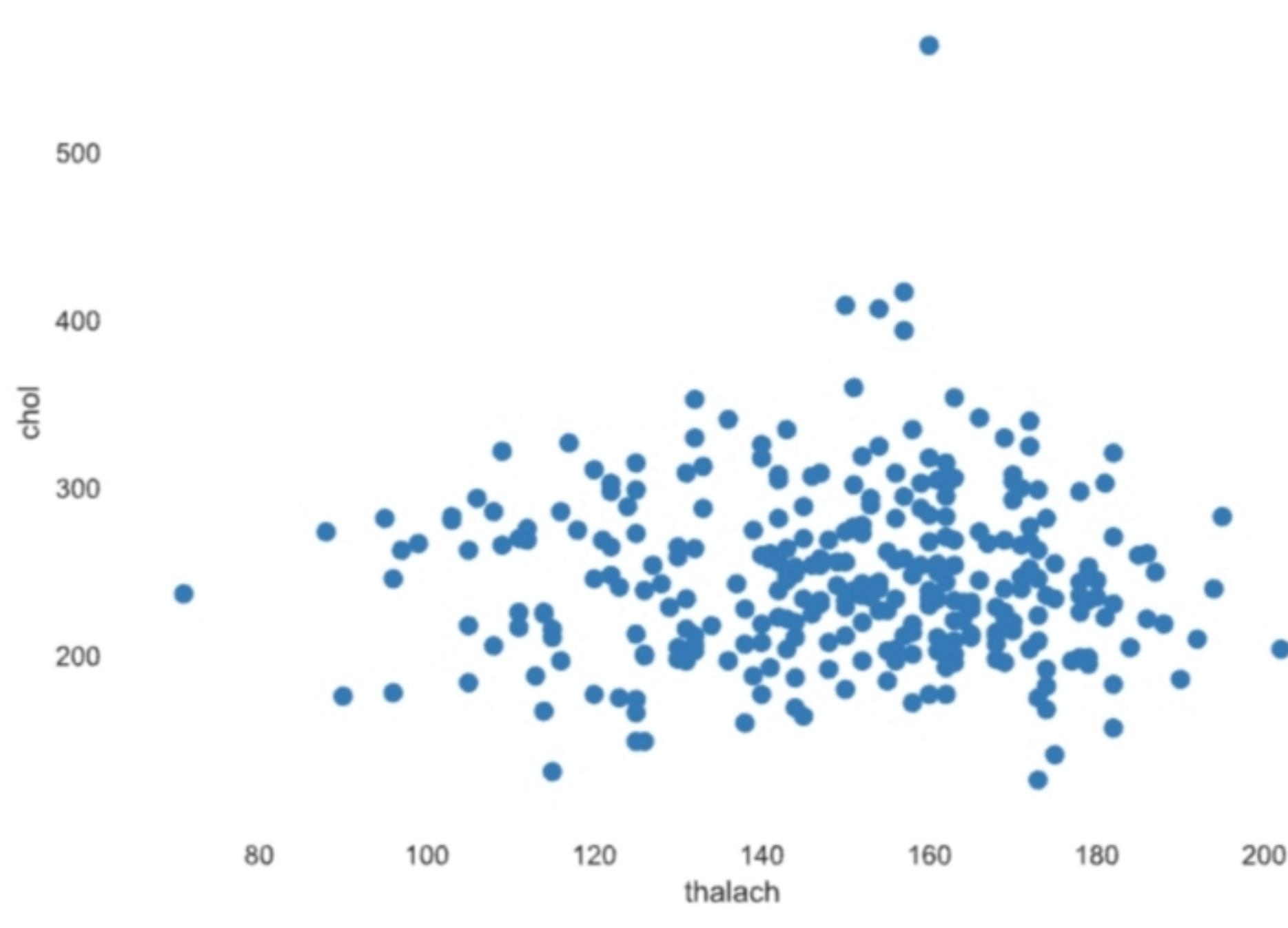
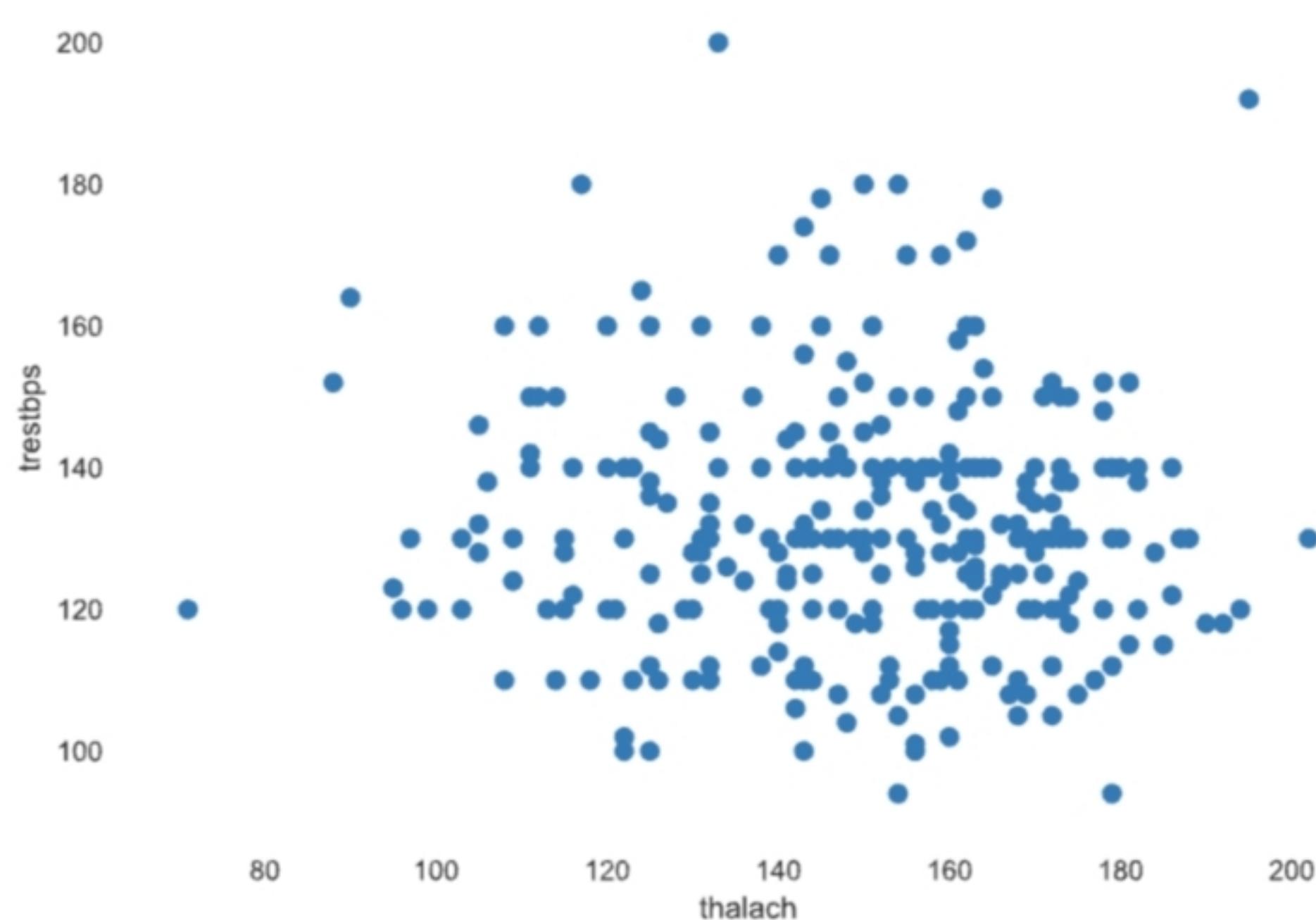






- oldpeak
- age
- trestbps
- chol
- thalach

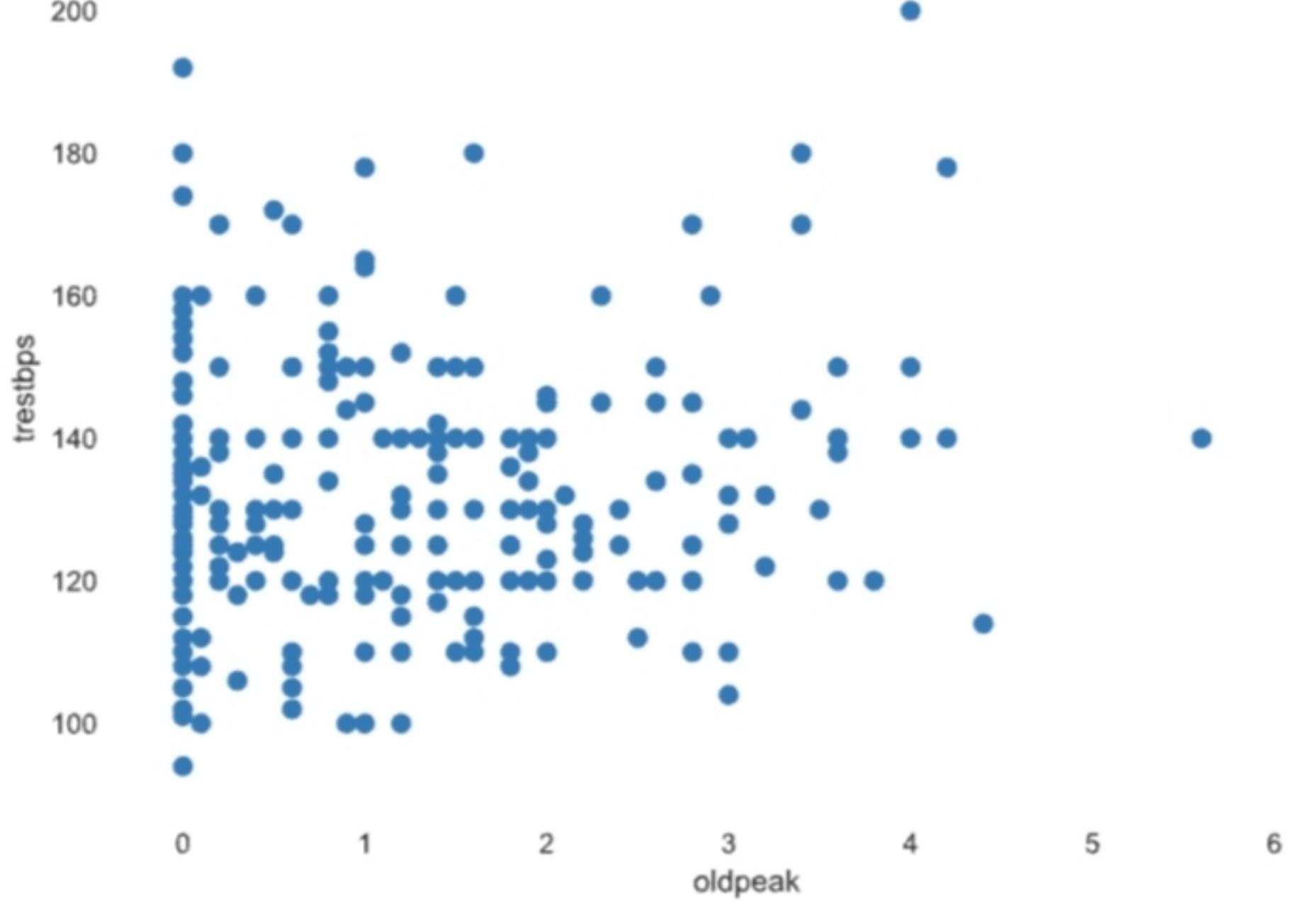
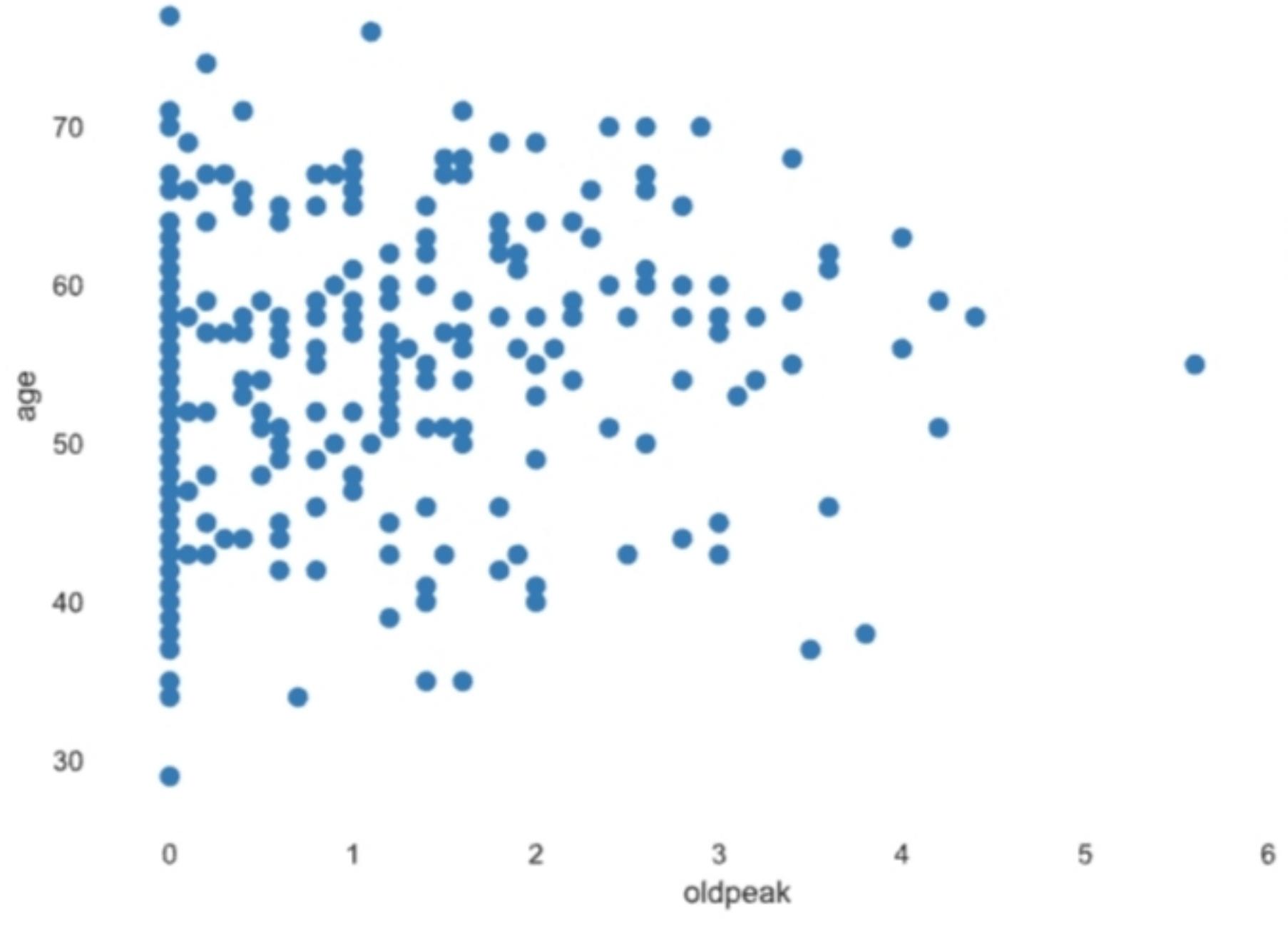
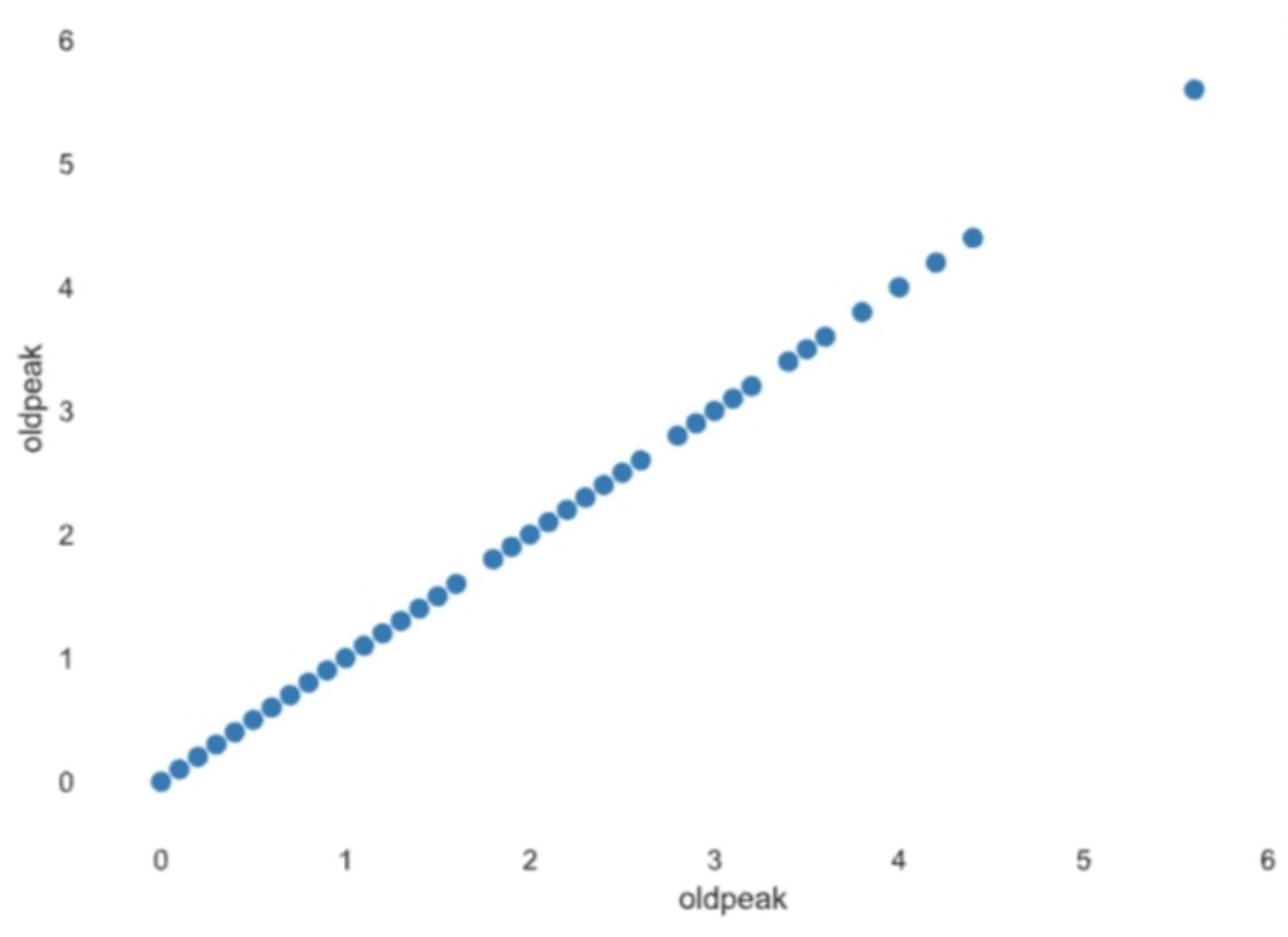


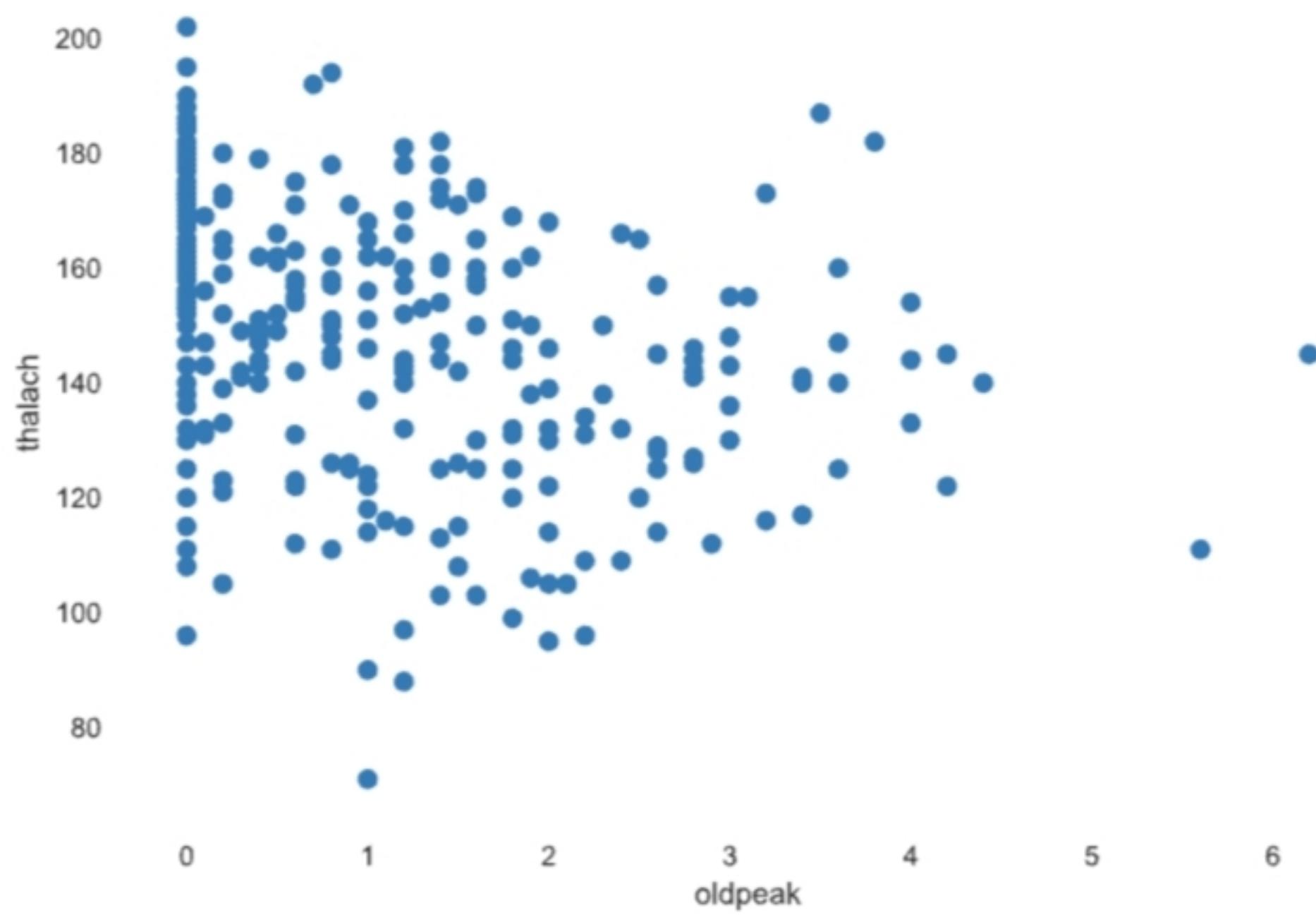


- oldpeak
- age
- trestbps
- chol
- thalach



- oldpeak
- age
- trestbps
- chol
- thalach





Correlations

- Spearman's ρ
- Pearson's r
- Kendall's τ
- Cramér's V (φ_c)
- Phik (φ_k)

Spearman's ρ

The Spearman's rank correlation coefficient (ρ) is a measure of monotonic correlation between two variables, and is therefore better in catching nonlinear monotonic correlations than Pearson's r . Its value lies between -1 and +1, -1 indicating total negative monotonic correlation, 0 indicating no monotonic correlation and 1 indicating total positive monotonic correlation.

To calculate ρ for two variables X and Y , one divides the covariance of the rank variables of X and Y by the product of their standard deviations.

Pearson's r

The Pearson's correlation coefficient (r) is a measure of linear correlation between two variables. Its value lies between -1 and +1, -1 indicating total negative linear correlation, 0 indicating no linear correlation and 1 indicating total positive linear correlation. Furthermore, r is invariant under separate changes in location and scale of the two variables, implying that for a linear function the angle to the x-axis does not affect r .

To calculate r for two variables X and Y , one divides the covariance of X and Y by the product of their standard deviations.

Kendall's τ

Similarly to Spearman's rank correlation coefficient, the Kendall rank correlation coefficient (τ) measures ordinal association between two variables. Its value lies between -1 and +1, -1 indicating total negative correlation, 0 indicating no correlation and 1 indicating total positive correlation.

To calculate τ for two variables X and Y , one determines the number of concordant and discordant pairs of observations. τ is given by the number of concordant pairs minus the discordant pairs divided by the total number of pairs.

Cramér's V (ϕ_c)

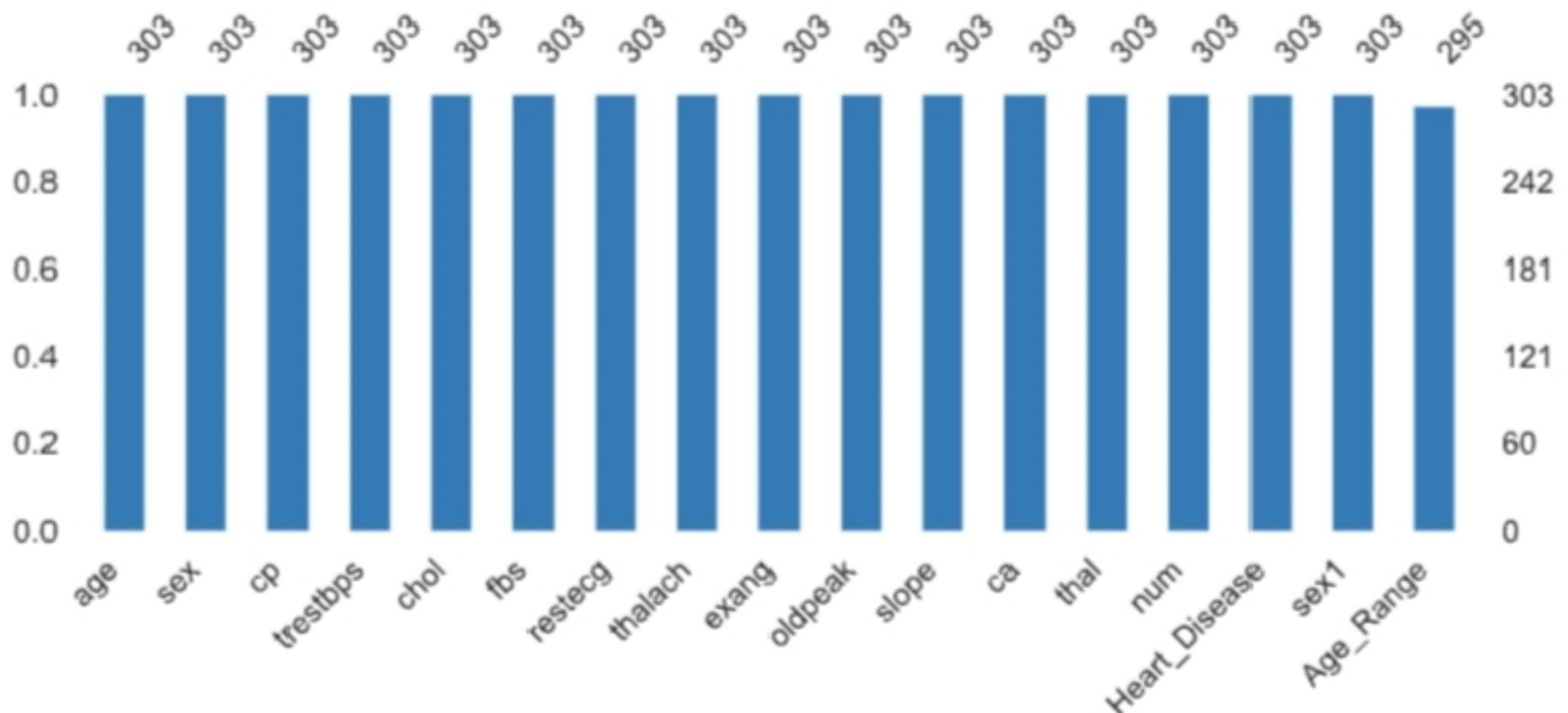
Cramér's V is an association measure for nominal random variables. The coefficient ranges from 0 to 1, with 0 indicating independence and 1 indicating perfect association. The empirical estimators used for Cramér's V have been proved to be biased, even for large samples. We use a bias-corrected measure that has been proposed by Bergsma in 2013 that can be found [here](#).

Phik (φ_k)

Phik (φ_k) is a new and practical correlation coefficient that works consistently between categorical, ordinal and interval variables, captures non-linear dependency and reverts to the Pearson correlation coefficient in case of a bivariate normal input distribution. There is extensive documentation available [here](#).

Missing values

- [Count](#)
- [Matrix](#)
- [Dendrogram](#)



A simple visualization of nullity by column.

Nullity matrix is a data-dense display which lets you quickly visually pick out patterns in data completion.



The dendrogram allows you to more fully correlate variable completion, revealing trends deeper than the pairwise ones visible in the correlation heatmap.

Sample

First rows

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num	Heart_Disease	sex1	Age_Range
0	63	1	1	145	233	1	2	150	0	2.3	3	0	6	0	Absence	Male	Elder Age
1	67	1	4	160	286	0	2	108	1	1.5	2	3	3	1	Presence	Male	Elder Age
2	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1	Presence	Male	Elder Age

Nullity matrix is a data-dense display which lets you quickly visually pick out patterns in data completion.

	sex	age	cp	trestbps	chol	fb	restecg	thalach	exang	oldpeak	slope	ca	thal	num	Heart_Disease	sex1	Age_Range
0																	
1																	
2																	

The dendrogram allows you to more fully correlate variable completion, revealing trends deeper than the pairwise ones visible in the correlation heatmap.

Sample

First rows

	age	sex	cp	trestbps	chol	fb	restecg	thalach	exang	oldpeak	slope	ca	thal	num	Heart_Disease	sex1	Age_Range
0	63	1	1	145	233	1	2	150	0	2.3	3	0	6	0	Absence	Male	Elder Age
1	67	1	4	160	286	0	2	108	1	1.5	2	3	3	1	Presence	Male	Elder Age
2	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1	Presence	Male	Elder Age
3	37	1	3	130	250	0	0	187	0	3.5	3	0	3	0	Absence	Male	Young Age
4	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0	Absence	Female	Middle Age
5	56	1	2	120	236	0	0	178	0	0.8	1	0	3	0	Absence	Male	Elder Age
6	62	0	4	140	268	0	2	160	0	3.6	3	2	3	1	Presence	Female	Elder Age
7	57	0	4	120	354	0	0	163	1	0.6	1	0	3	0	Absence	Female	Elder Age
8	63	1	4	130	254	0	2	147	0	1.4	2	1	7	1	Presence	Male	Elder Age
9	53	1	4	140	203	1	2	155	1	3.1	3	0	7	1	Presence	Male	Middle Age

Last rows

	age	sex	cp	trestbps	chol	fb	restecg	thalach	exang	oldpeak	slope	ca	thal	num	Heart_Disease	sex1	Age_Rang
293	63	1	4	140	187	0	2	144	1	4.0	1	2	7	1	Presence	Male	Elder Ag
294	63	0	4	124	197	0	0	136	1	0.0	2	0	3	1	Presence	Female	Elder Ag
295	41	1	2	120	157	0	0	182	0	0.0	1	0	3	0	Absence	Male	Midd Ag
296	59	1	4	164	176	1	2	90	0	1.0	2	2	6	1	Presence	Male	Elder Ag
297	57	0	4	140	241	0	0	123	1	0.2	2	0	7	1	Presence	Female	Elder Ag
298	45	1	1	110	264	0	0	132	0	1.2	2	0	7	1	Presence	Male	Midd Ag
299	68	1	4	144	193	1	0	141	0	3.4	2	2	7	1	Presence	Male	Elder Ag
300	57	1	4	130	131	0	0	115	1	1.2	2	1	7	1	Presence	Male	Elder Ag
301	57	0	2	130	236	0	2	174	0	0.0	2	1	3	1	Presence	Female	Elder Ag
302	38	1	3	138	175	0	0	173	0	0.0	1	-100000	3	0	Absence	Male	Young Ag