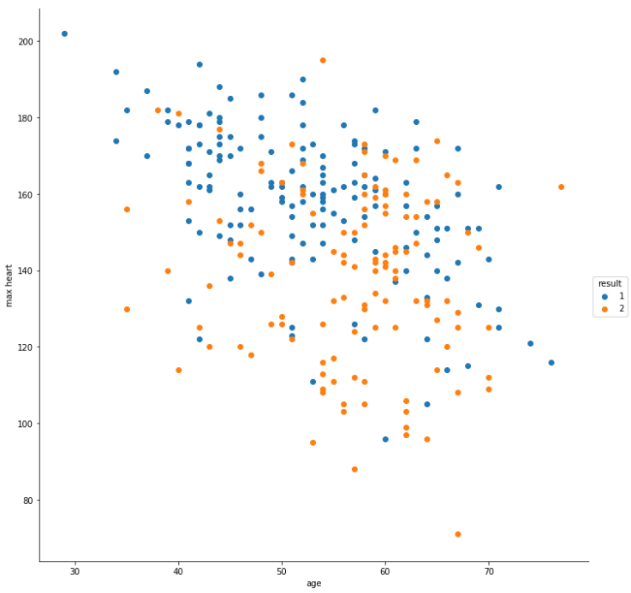
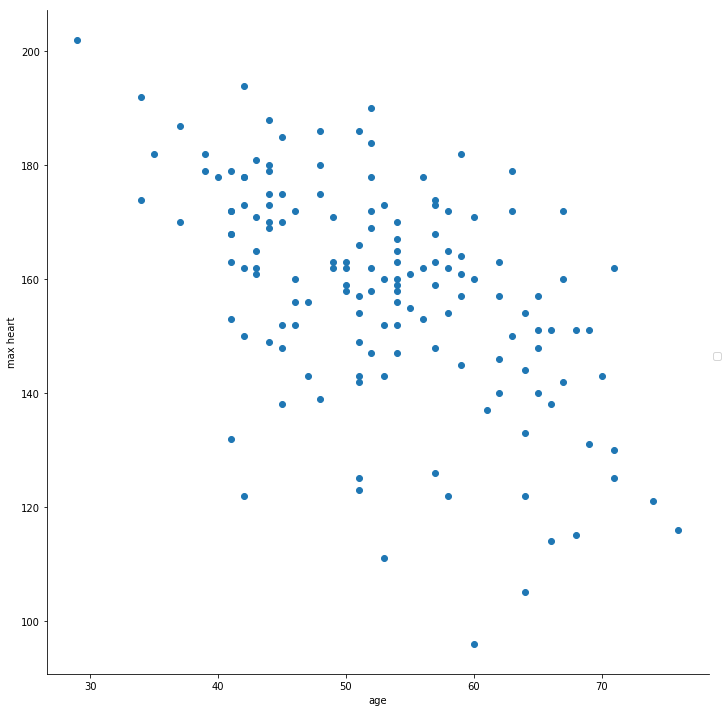
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Y  X | Age | Blood Pressure | Serum | Max Heart | Old Peak |
| Age |  | No | No | Correlation | No |
| Blood Pressure |  |  | No | No | No |
| Serum |  |  |  | No | No |
| Max Heart |  |  |  |  | No |
| Old Peak |  |  |  |  |  |

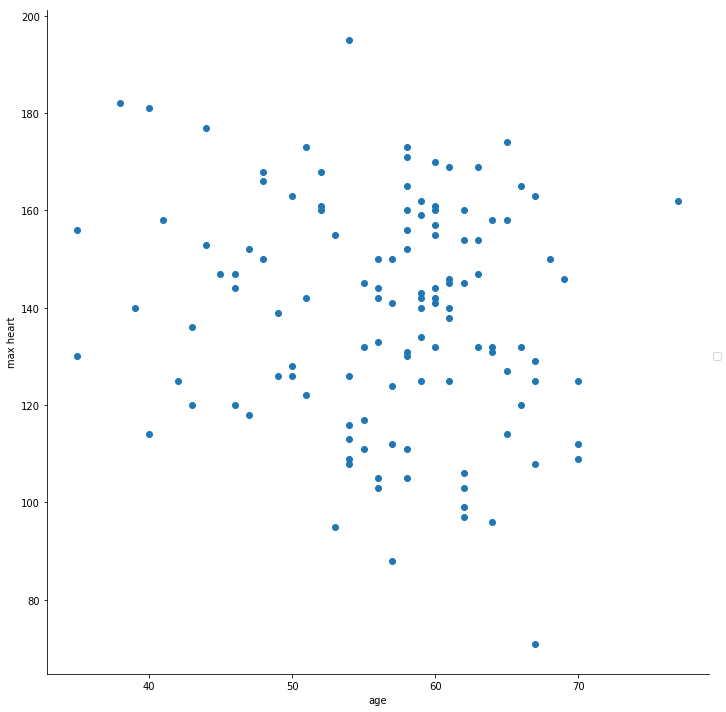
It’s hard to see correlation on oldpeak

Age and max heart correlation

No heart disease

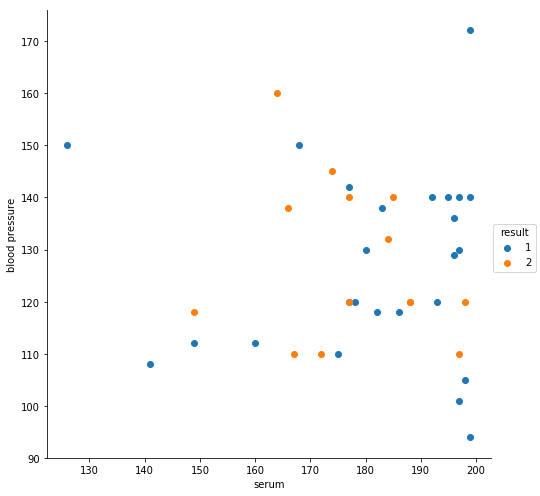


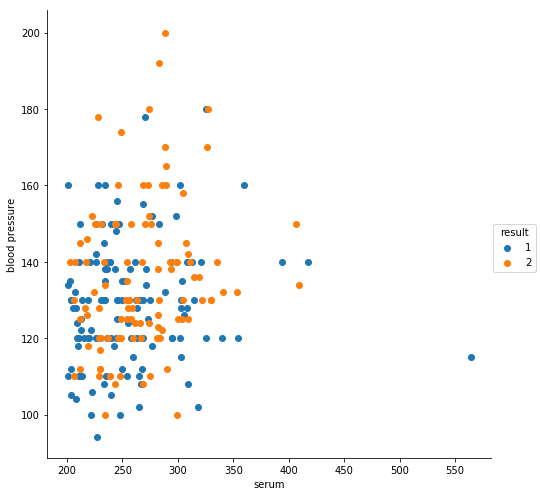
With heart disease



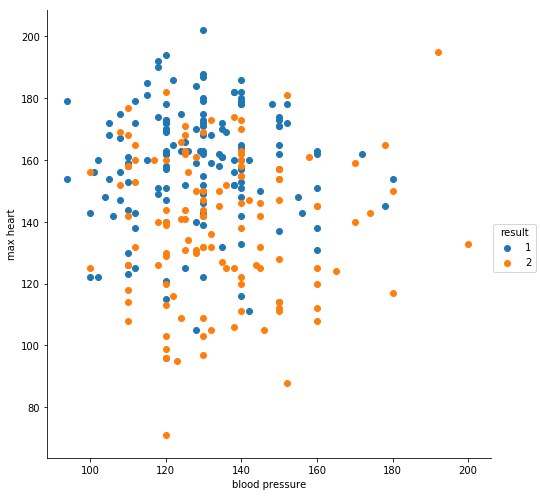
People who have more than 200 dl/mg on cholesterol serum have higher chance to get heart

Diseases



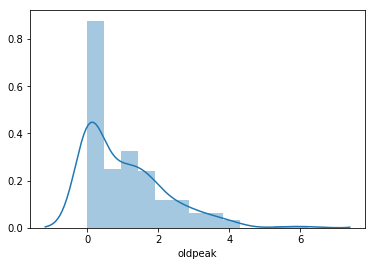


Blood pressure vs max heart has no correlation but has interesting outliers for those who have heart diseases. See below



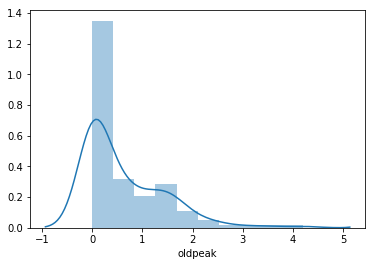
In a [cardiac stress test](https://en.wikipedia.org/wiki/Cardiac_stress_test), an ST depression of at least 1 mm after [adenosine](https://en.wikipedia.org/wiki/Adenosine) administration indicates a reversible ischaemia, while an exercise stress test requires an ST depression of at least 2 mm to significantly indicate reversible ischaemia. Source: [wikipedia](https://en.wikipedia.org/wiki/ST_depression)

sns.distplot(data['oldpeak'])



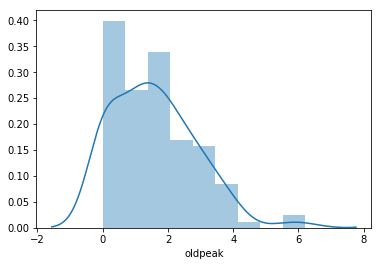
filteredData = data.loc[data['result'] == 1]

sns.distplot(filteredData['oldpeak'])



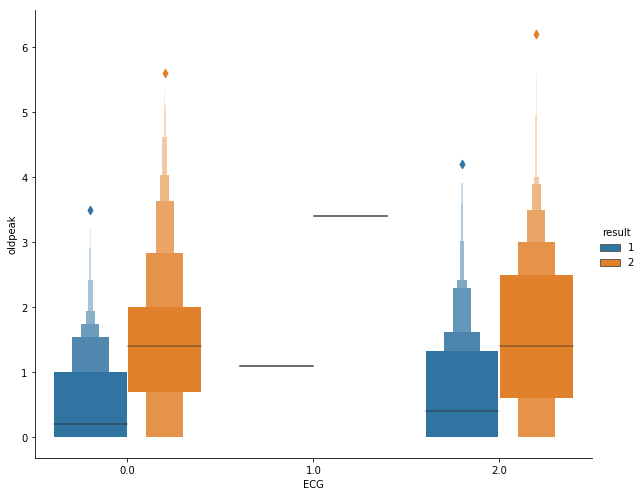
filteredData = data.loc[data['result'] == 2]

sns.distplot(filteredData['oldpeak'])



Normal person can have either 0 or 2, but a person with have higher than normal oldpeak will most likely have disease if the person has 1.

sns.catplot(x="ECG", y="oldpeak", data=data, hue='result', aspect=1.2, kind='boxen', height=7)



print ("{:<20}{:<8}{:<8}{:<8}{:<8}{:<8}".format('column','min','max','std','mean','median'))

for x in range(14):

if data.columns.values[x] == 'age' or data.columns.values[x] == 'blood pressure' or \

data.columns.values[x] == 'serum' or data.columns.values[x] == 'max heart' or\

data.columns.values[x] == 'oldpeak':

print ("{:<20}{:<8}{:<8}{:<8}{:<8}{:<8}".format(data.columns.values[x], data.iloc[:, x].min(),

data.iloc[:, x].max(),round(data.iloc[:, x].std(),2),

round(data.iloc[:, x].mean(),2),round(data.iloc[:, x].median(),2)))

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Column | Min | Max | Std | Mean | Median |
| Age | 29 | 77 | 9.11 | 54.43 | 55 |
| Blood Pressure | 94 | 200 | 17.86 | 131.34 | 130 |
| Serum | 126 | 564 | 51.69 | 249.66 | 245 |
| Max Heartrate | 71 | 202 | 23.17 | 149.68 | 153.5 |
| oldpeak | 0 | 6.2 | 1.15 | 1.05 | 0.8 |

#Printing the

for x in range(14):

if data.columns.values[x] != 'age' and data.columns.values[x] != 'blood pressure' and \

data.columns.values[x] != 'serum' and data.columns.values[x] != 'max heart' and\

data.columns.values[x] != 'oldpeak':

display(data.groupby(data.columns.values[x]).size().reset\_index(name='Count').rename(columns={'Col1':'Col\_value'}))

|  | **sex** | **Count** |
| --- | --- | --- |
| **0** | 0.0 | 87 |
| **1** | 1.0 | 183 |

|  | **chest pain** | **Count** |
| --- | --- | --- |
| **0** | 1.0 | 20 |
| **1** | 2.0 | 42 |
| **2** | 3.0 | 79 |
| **3** | 4.0 | 129 |

|  | **fasting** | **Count** |
| --- | --- | --- |
| **0** | 0.0 | 230 |
| **1** | 1.0 | 40 |

|  | **ECG** | **Count** |
| --- | --- | --- |
| **0** | 0.0 | 131 |
| **1** | 1.0 | 2 |
| **2** | 2.0 | 137 |

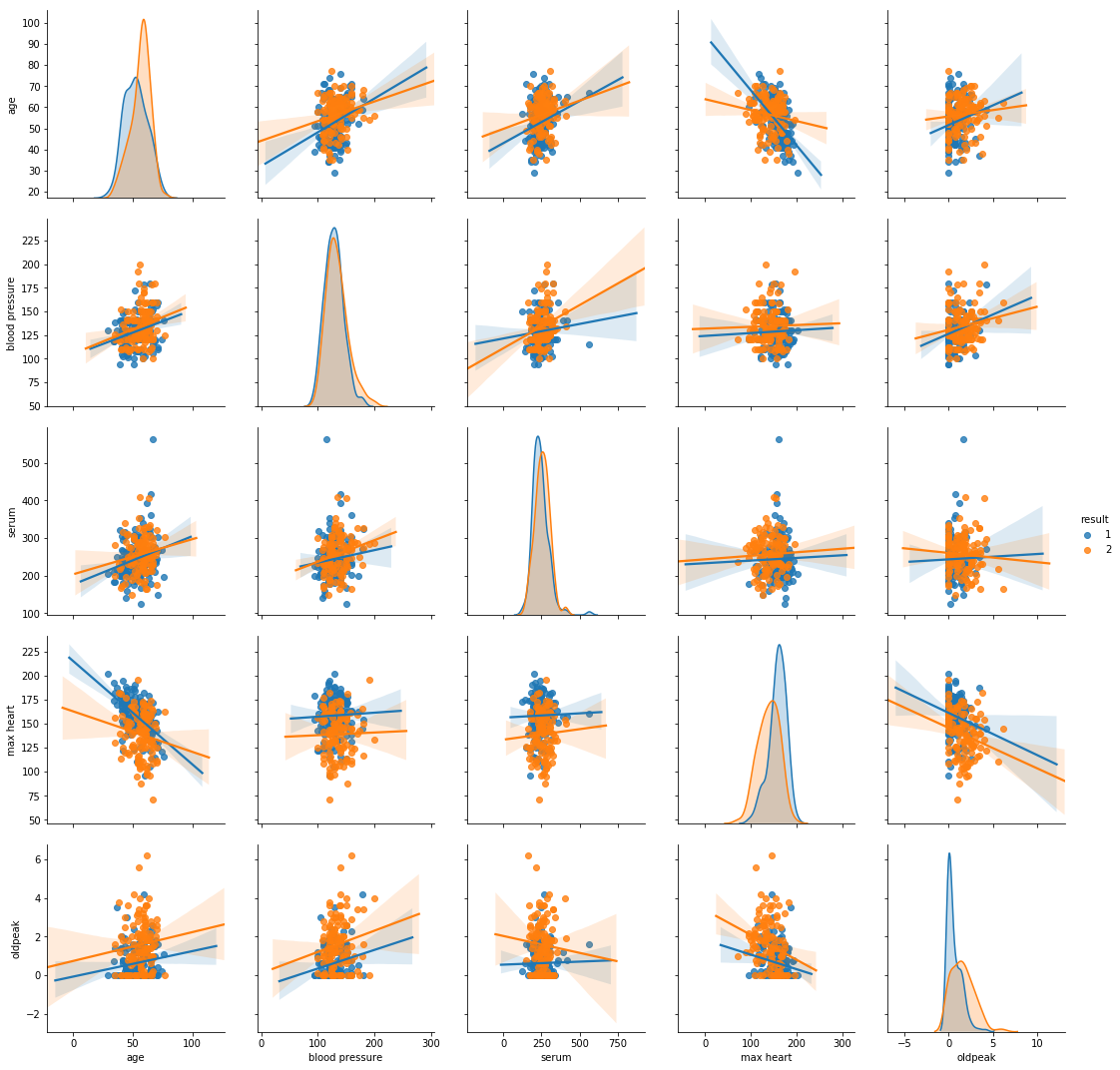
|  | **angina** | **Count** |
| --- | --- | --- |
| **0** | 0.0 | 181 |
| **1** | 1.0 | 89 |

|  | **slope** | **Count** |
| --- | --- | --- |
| **0** | 1.0 | 130 |
| **1** | 2.0 | 122 |
| **2** | 3.0 | 18 |

|  | **major vessel** | **Count** |
| --- | --- | --- |
| **0** | 0.0 | 160 |
| **1** | 1.0 | 58 |
| **2** | 2.0 | 33 |
| **3** | 3.0 | 19 |

|  | **thal** | **Count** |
| --- | --- | --- |
| **0** | 3.0 | 152 |
| **1** | 6.0 | 14 |
| **2** | 7.0 | 104 |

|  | **result** | **Count** |
| --- | --- | --- |
| **0** | 1 | 150 |
| **1** | 2 | 120 |

sns.pairplot(data, vars= ['age','serum','blood pressure','max heart','oldpeak'], height=3,hue='result',diag\_kind="kde",kind="reg") 

Judging from the graph, it seems that blood pressure and serum are not the main factors to determine someone Is having heard disease or not. The best factor is max heart rate, old peak, and age.