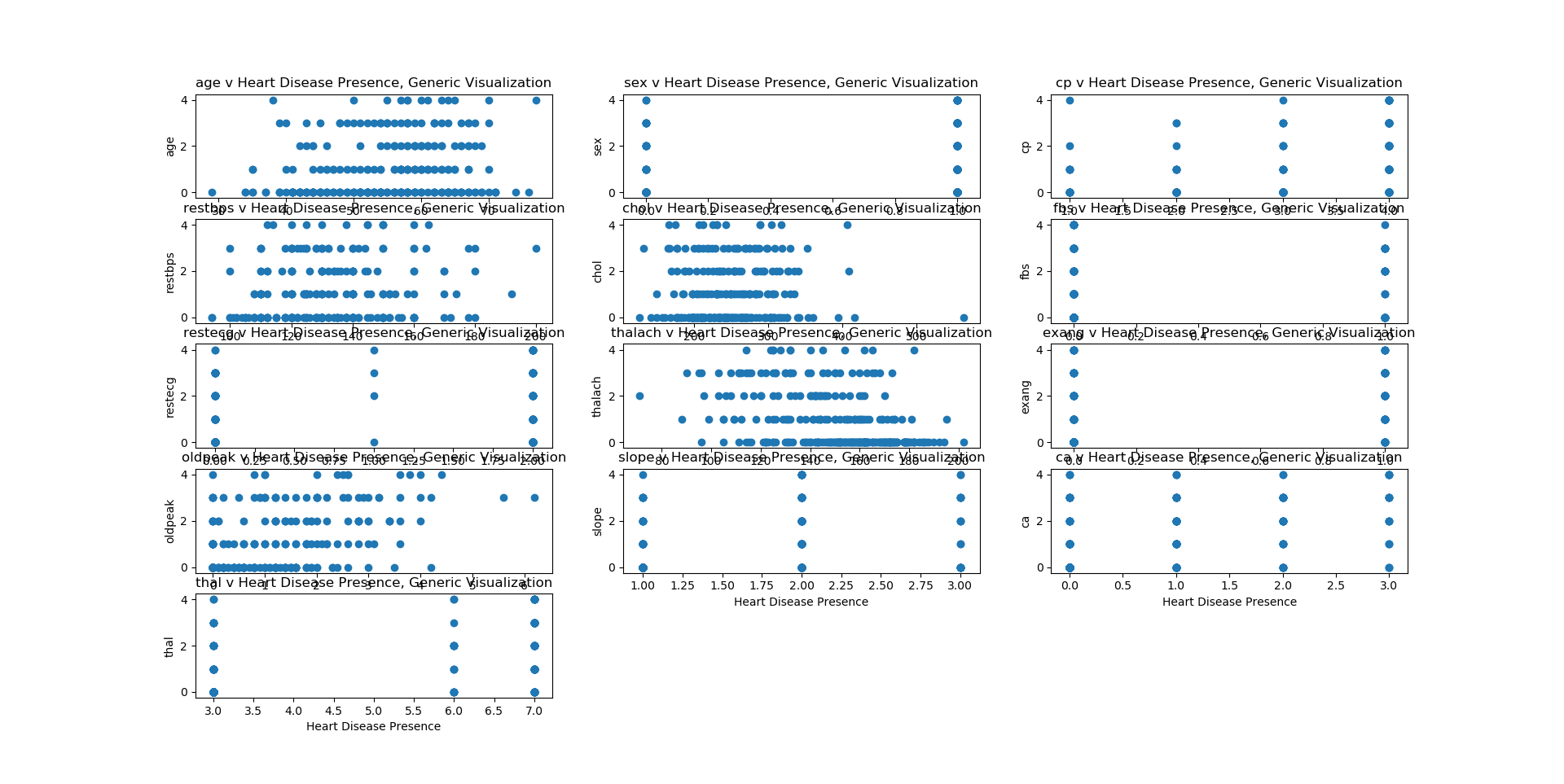
Distribution of data points:

Chart, bar chart

Description automatically generated

Sample points are more heavily weighted towards no presence of heart disease (0). Similar to normal distribution, 53.87% being 0, 72.05% between 0 and 1

Data points separated by category:



Nothing notable from this visualization, need to weight points for discrete categories, similarly something different for continuous values.

Weighted points, discrete values

Calendar

Description automatically generated

Sex(1): heart disease more prevalent in one sex compared to other. If female, highly likely to have no presence. If male, higher chance for any bit of presence

Cp(2): cp = 1,2, or 3, appears to follow normal distribution. When cp = 4, equal chance for all 5 values. I.e. for cp=4, there is chance you have some HD presence.

Fbs(3): non-normal distribution for fbs = 1

restECG(4): normal distr when ecg = 0, non-normal for restECG = 1, 2

exercise angina(5): nonnormal distr when exang = 1

slope(6): non-normal for slope not 1 (slope = 2, 3)

ca(7): normal distr when ca = 0, still similar distr but less 0-value heavy when ca = 1, trend breaks at ca = 2, 3

thal(8): normal distr for thal = 3, when all other values (6,7) no-normal

for all categories, need to test for statistical significance as each point size was scaled for the total count of each individual discrete category. For example, may not be enough points with restecg = 1, also lack of values for thal = 4, 5.

TO-DO, test for statistical significance

Continuous categories

Chart, box and whisker chart

Description automatically generated

For the continuous data, only plots 4 and 5 seem to have any visual trend. One solution to better visualize would be to bin values and perform plot similar to discrete plot. In this case, we can only say for heart disease, average is so and so. Want to work the other way around. (ie. Predicting HD presence with the continuous values.