# calculate correlation matrix

corr = data.corr()

plt.subplots(figsize=(15,10))

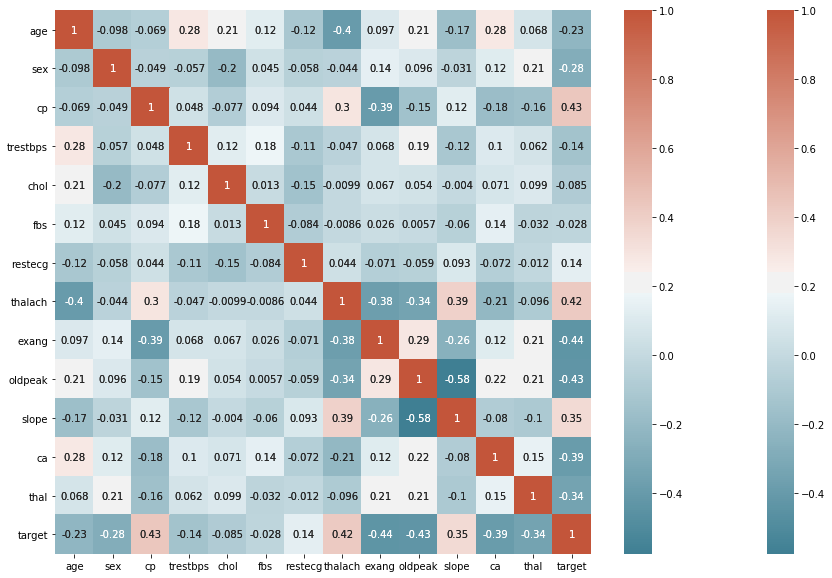
sns.heatmap(corr, xticklabels=corr.columns, yticklabels=corr.columns, annot=True, cmap=sns.diverging\_palette(220, 20, as\_cmap=True))

sns.heatmap(corr, xticklabels=corr.columns,

yticklabels=corr.columns,

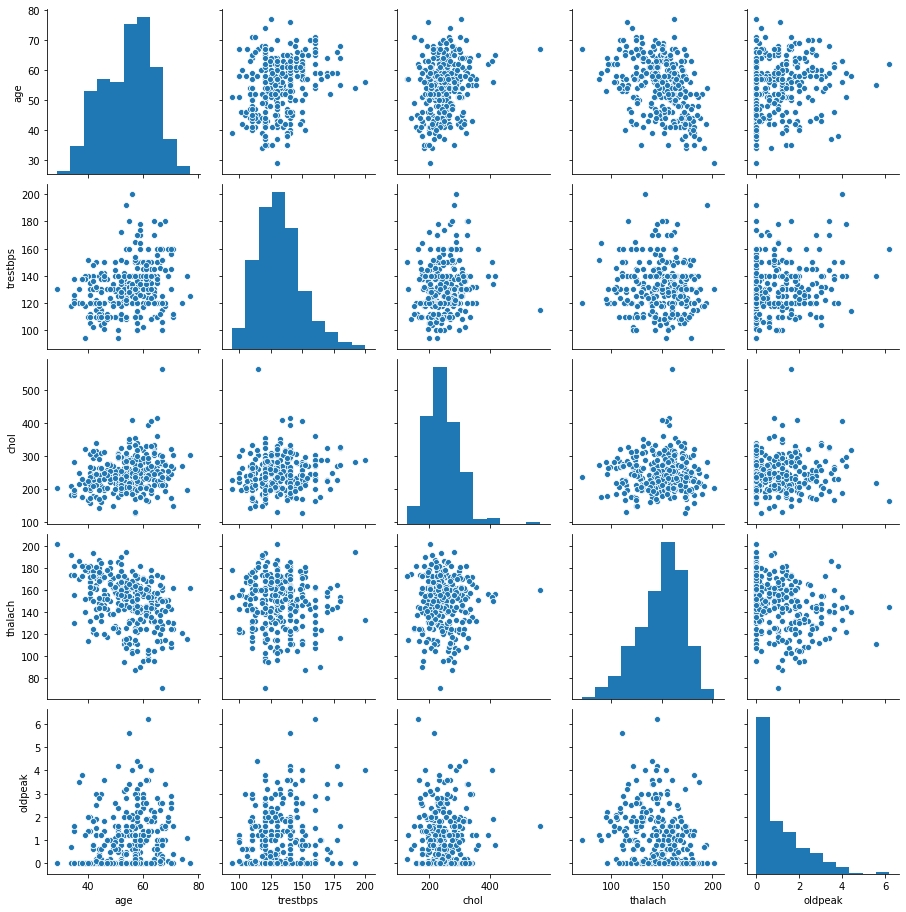
annot=True,

cmap=sns.diverging\_palette(220, 20, as\_cmap=True))



subData = data[['age','trestbps','chol','thalach','oldpeak']]

sns.pairplot(subData)



Chose to make a smaller pairplot with only the continus variables, to dive deeper into the relationships. Also a great way to see if theirs a positve or negative correlation!

In [12]:



sns.catplot(x**=**"target", y**=**"oldpeak", hue**=**"slope", kind**=**"bar", data**=**data);

​

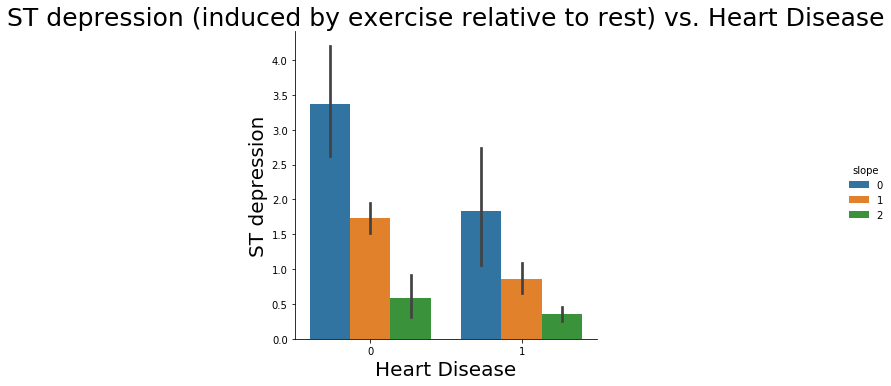
plt.title('ST depression (induced by exercise relative to rest) vs. Heart Disease',size**=**25)

plt.xlabel('Heart Disease',size**=**20)

plt.ylabel('ST depression',size**=**20)

Out[12]:

Text(26.426458333333343, 0.5, 'ST depression')



**Violin & Box Plots**

The advantages of showing the Box & Violin plots is that it showsthe basic statistics of the data, as well as its distribution. These plots are often used to compare the distribution of a given variable across some categories. It shows the median, IQR, & Tukey’s fence. (minimum, first quartile (Q1), median, third quartile (Q3), and maximum). In addition it can provide us with outliers in our data.

In [156]:



plt.figure(figsize**=**(12,8))

sns.violinplot(x**=** 'target', y**=** 'oldpeak',hue**=**"sex", inner**=**'quartile',data**=** data )

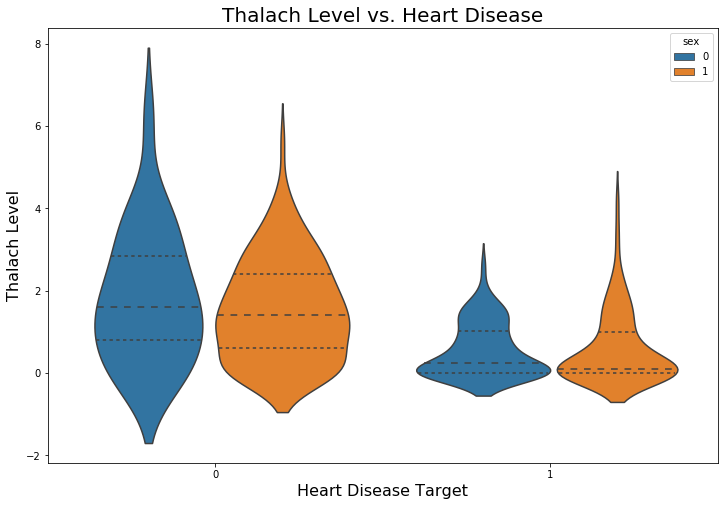
plt.title("Thalach Level vs. Heart Disease",fontsize**=**20)

plt.xlabel("Heart Disease Target", fontsize**=**16)

plt.ylabel("Thalach Level", fontsize**=**16)

Out[156]:

Text(0, 0.5, 'Thalach Level')

****

We can see that the overall shape & distribution for negative & positive patients differ vastly. Positive patients exhibit a lower median for ST depression level & thus a great distribution of their data is between 0 & 2, while negative patients are between 1 & 3. In addition, we dont see many differences between male & female target outcomes.

In [14]:



plt.figure(figsize**=**(12,8))

sns.boxplot(x**=** 'target', y**=** 'thalach',hue**=**"sex", data**=**data )

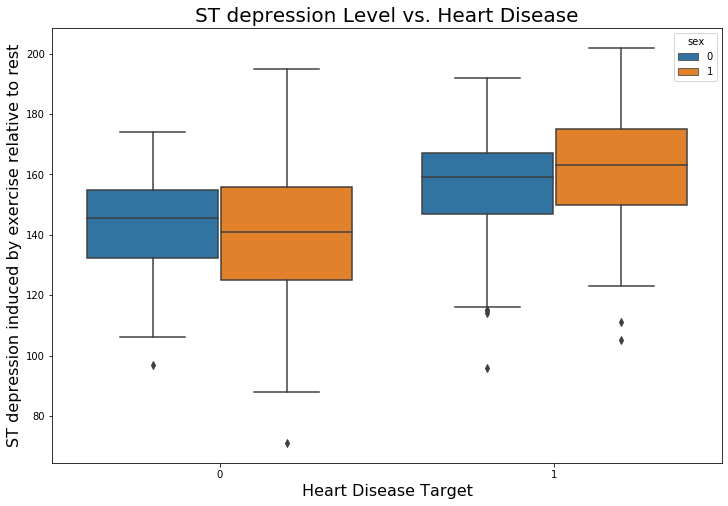
plt.title("ST depression Level vs. Heart Disease", fontsize**=**20)

plt.xlabel("Heart Disease Target",fontsize**=**16)

plt.ylabel("ST depression induced by exercise relative to rest", fontsize**=**16)

Out[14]:

Text(0, 0.5, 'ST depression induced by exercise relative to rest')

****