# Overview

The scope of this analysis is to estimate the how the reported perceived levels of chest pain correlate with a patient’s odds of having hearts disease.

# Research questions

Is chest pain an accurate signal of heart disease? What if it is exercise induced?

Regression with chest pain alone

Is the persistence of chest pain more significant if a patient has other confounding factors?

Regression with chest pain and other variables

How do these effects compare for both men and women?

Interaction between chest pain and sex

WHAT WILL THE ANSWER LOOK LIKE?

Interpreted coefficients for chest pain variables in each of the different cases

# Data

You should describe the data in this section, that is, the source of the data, the variables included, the size of the dataset (for example, how many observations are in your dataset), potential problems, and so on.

This entire database contains data 76 attributes rom 4 different sources, but all published experiments refer to using a subset of 14 of them. The "target" field refers to the presence of heart disease in the patient (angiographic disease status); Value 0: < 50% diameter narrowing – less chance of heart attack; Value 1: > 50% diameter narrowing – more chance of heart attack. The processed dataset from the Cleveland Clinic Foundation of the 14 key attributes consists of 303 rows with no missing values. These other datasets from the Hungarian Institute of Cardiology, the Long Beach, CA, V.A. Medical Center, and University Hospital in Zurich are also available in their unprocessed form.

**Attribute Information**

1) age

2) sex

3) chest pain type (4 values)

4) resting blood pressure

5) serum cholesterol in mg/dl

6) fasting blood sugar > 120 mg/dl

7) resting electrocardiographic results (values 0,1,2)

8) maximum heart rate achieved

9) exercise induced angina

10) oldpeak = ST depression induced by exercise relative to rest

11) the slope of the peak exercise ST segment

12) number of major vessels (0-3) colored by fluoroscopy

13) thal: 0 = normal; 1 = fixed defect; 2 = reversable defect

14) target: 0= less chance of heart attack 1= more chance of heart attack

# Project plan

Finally, outline your analysis plan: which models do you plan to explore, what milestones have you set for ensuring that you complete the project in time, and so on.

I plan to explore a logistic regression model, potentially hierarchical to compare different study locations. If I do plan to include the other unprocessed the datasets my first milestone would be to clean those datasets. Next, would be to explore the data and fit models. Then would be to create a comprehensive summary report of my findings that address the research questions posed.