

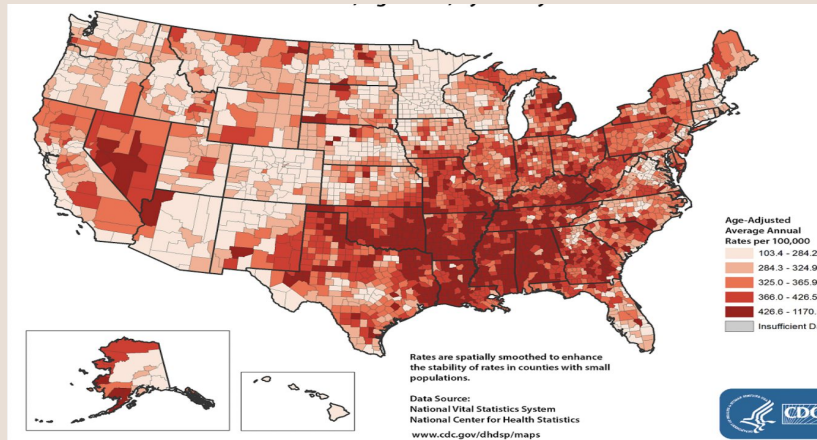


# Heart Disease Detection

Team 2: Angela Zhang, Nava  
Roohi, Jimmy Nguyen

# Introduction

- Heart disease causes the highest fatality rate.
  - About 61,000 people die of heart disease in the United States annually that is 1 in every 4 deaths.



# Heart Disease Prediction

- By early detection, a proper care and treatment could be implemented.
- Difficult to distinguish high risk patients due to the multifactorial nature of contributing risk factors such as high cholesterol or high blood pressure etc.

# Heart Disease Detection Using Data Mining

- Data mining is a novel tool to detect heart disease..
- Classification models from data mining will be introduced to address this problem:
  - ❑ Gaussian Naive Bayesian
  - ❑ Logistic Regression
  - ❑ Random Forest models

# Dataset

- Dataset is provided by UCI repository and was taken from Kaggle.
- 303 rows & 14 columns:
  - ❑ 13 independent variable and 1 dependent variable (Target variable of interest).
  - ❑ Patient's Health related information & Demographic related information

Name	Description
<i>Age</i>	Age in years
<i>Sex</i>	0 = Female 1 = Male
<i>Cp</i>	Chest Pain Type 1 = Typical Angina 2 = Atypical Angina 3 = Non-Angina 4 = Asymptomatic
<i>Trestps</i>	Resting Blood Pressure (in mmHg)
<i>Chol</i>	Serum Cholesterol in mg/dl
<i>Fbs</i>	Fasting Blood Sugar >120 mg/dl: 1 = True 0 = False
<i>Restecg</i>	Resting Cardiographic results: 0=Normal 1=Having ST=T wave abnormality 2=Showing Probable or Define Left Ventricular Hypertrophy by Este's Criteria
<i>Thalach</i>	Maximum Heart Rate Achieved
<i>Exang</i>	Exercise induced angina: 1 = Yes 0 = No
<i>Old Peak</i>	ST Depression Induced by Exercise Relative to Rest
<i>Ca</i>	Number of Major Vessels (0-3) Colored by Fluoroscopy
<i>Thal</i>	Heart beat measurements 3 = Normal 6 = Fixed Defect 7 = Reversible Defect
<i>Target</i>	Diagnosis of Heart Disease (Angiographic Disease Status) 0 = No 1 = Yes