Introduction

* Choice of two most deadly diseases, heart disease and cancer
  + Source of data – Medicare Disparities Data
* Related to income
  + Taken from IRS tax return data for each county
* Metric of annual cost to income ratio
* Initial Application of the data
  + Initial description for both cancer data and heart attack data
    - Use choropleth and bar charts
      * Counties with highest cost-to-income ratio
* Sought additional data to model what factors influence the cost-to-income ratio.
  + Population data, including changes in population
  + Demographic (age, gender)
  + Percentage of people without insurance
* Describe the model (method slides)
  + Random forest models for both cancer and heart disease
  + Predictive power of the models
* Predicting variables, with choropleths
  + Counties with higher population density are lower in cost-income ratio
  + Counties with higher migration rates (in or out) are lower in cost-income ratio
  + Counties with higher medical expense deductions per capita are lower in cost-income ratio