Heart disease risk factors the Panda way

...

Spaar, Andrew, Ivan, Michel, Danielle

Motivation & Summary

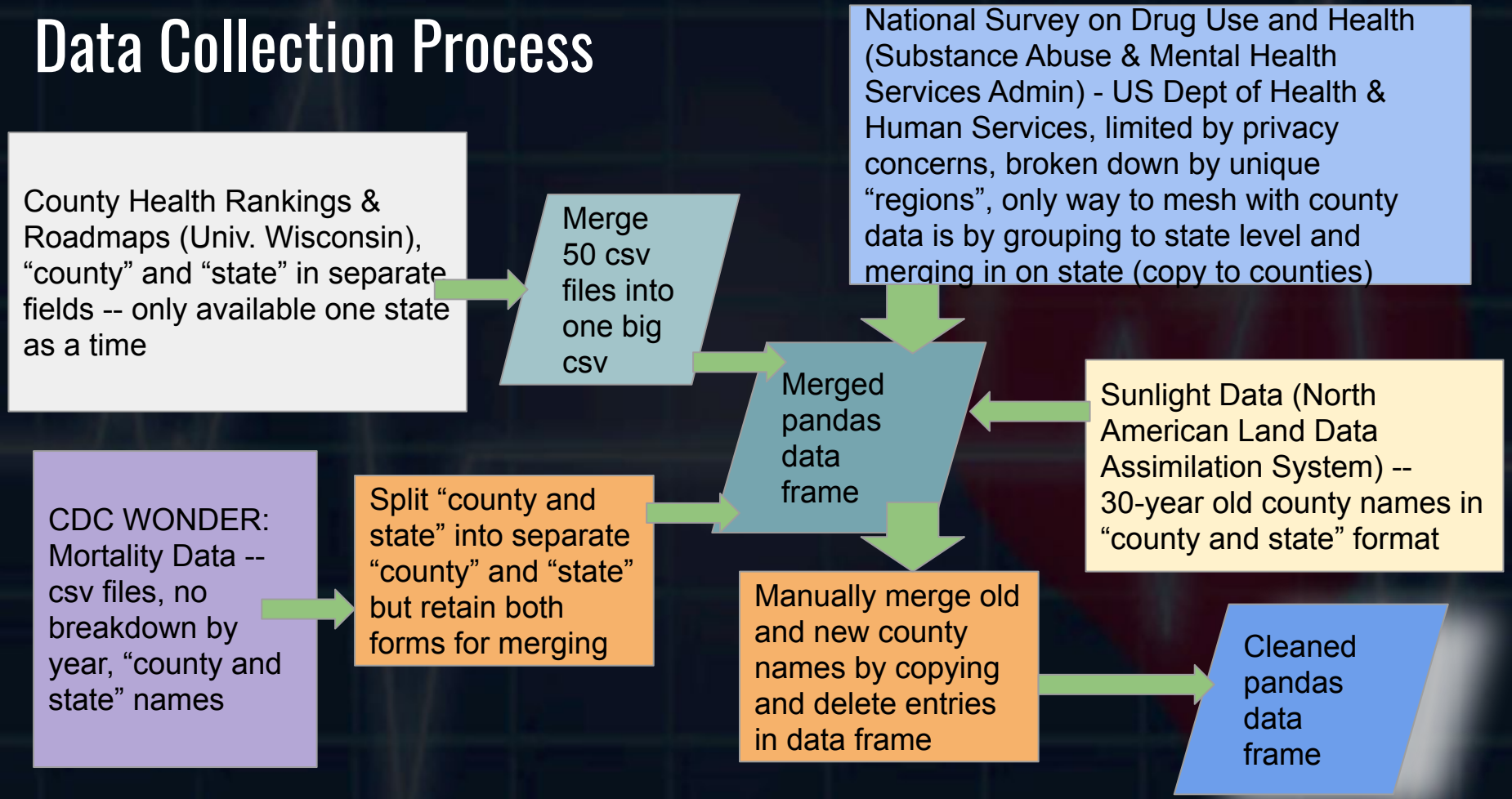
Thesis:

Freely available county-level data will identify impact of different heart disease risk factors.

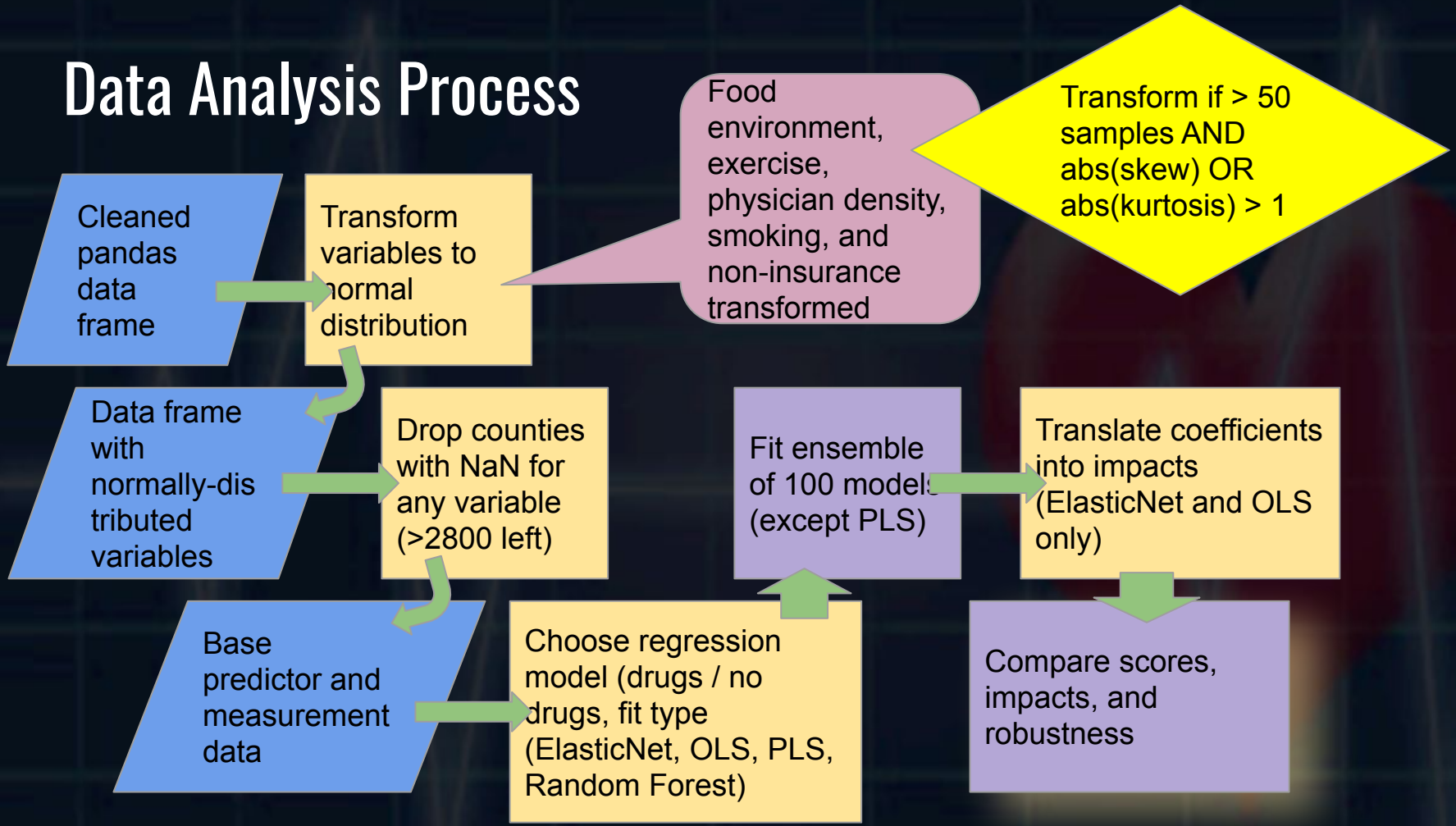
Why This Is Important:

- Heart disease results in 600,000 deaths in the US each year (#1 cause)
- The costs of heart disease exceed $200 billion annually in the US
- The results of traditional medical studies have not yet led to hoped-for reductions in risk factors.

Data Collection Process



Data Analysis Process

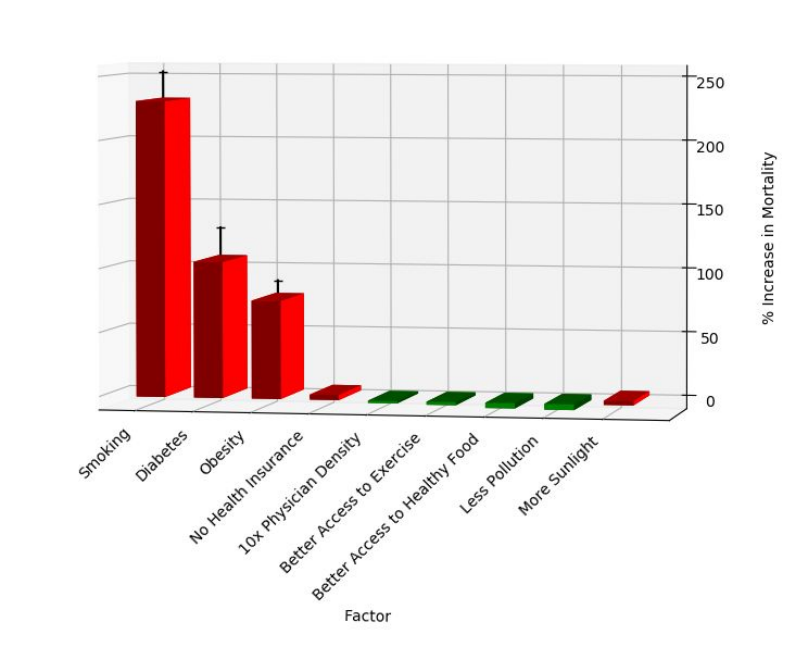


Statistical Results

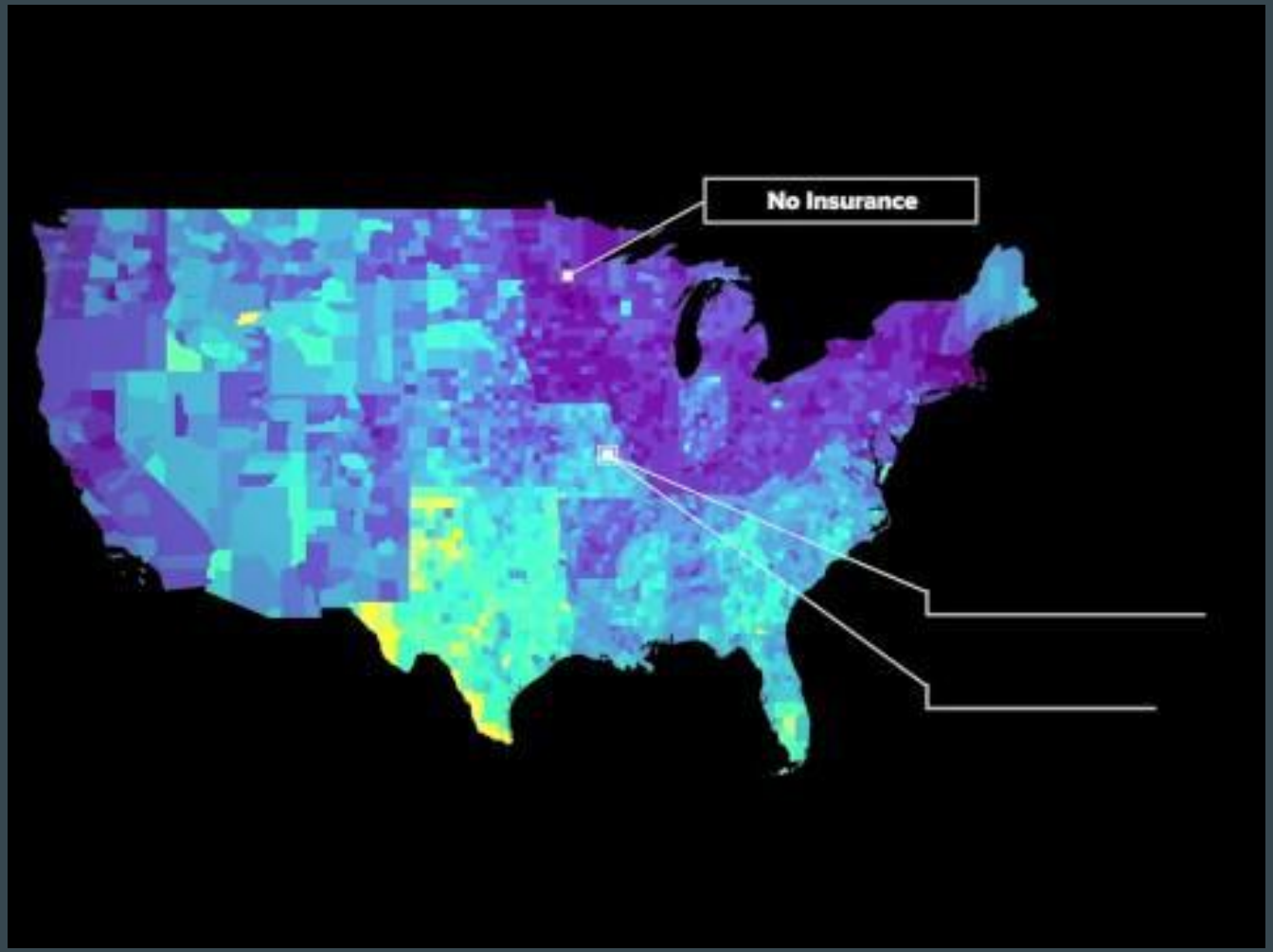
Factors	P-Values	Findings
Smoking	4 x 10 ^ -35	SIGNIFICANT
Diabetes	6 x 10 ^ -10	SIGNIFICANT
Obesity	3 x 10 ^ -17	SIGNIFICANT
Lack of Health Insurance	2 x 10 ^ -5	SIGNIFICANT
Access to Physicians	2 x 10 ^ -3	SIGNIFICANT
Sunlight Exposure	5 x 10 ^ -10	SIGNIFICANT
Pollution	2 x 10 ^ -20	SIGNIFICANT
Food Environment Index	5 x 10 ^ -13	SIGNIFICANT
Exercise	6 x 10 ^ -5	SIGNIFICANT

Assumes n > 2000 independent, random samples. Does not take systematic effects into account.

Impacts on Mortality

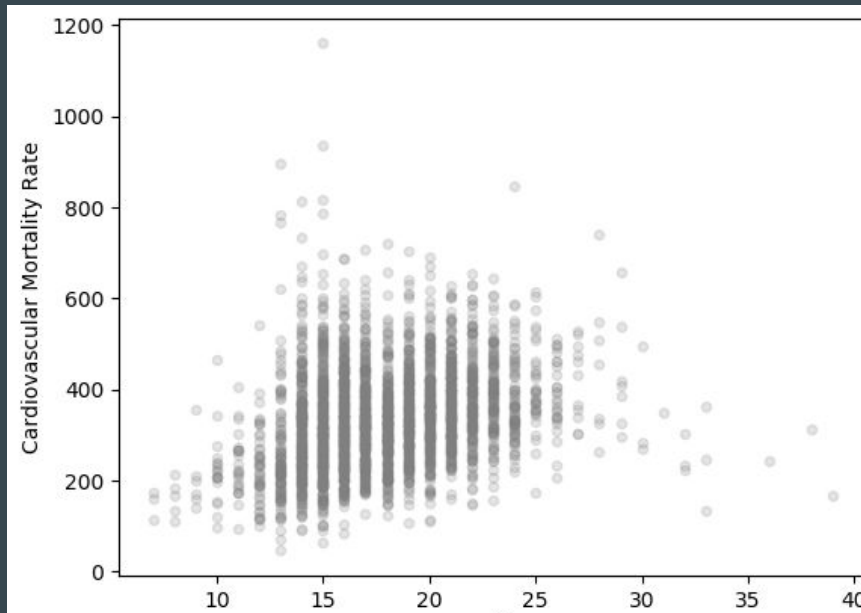


- Impacts generated from average coefficients in ensemble of linear regressions
- Order and magnitude of obesity, diabetes, and smoking impacts computed on county basis qualitatively similar to impacts for individuals reported in the literature..
- Factors such as “food environment” are likely not independent of obesity in a meaningful sense.


No Insurance

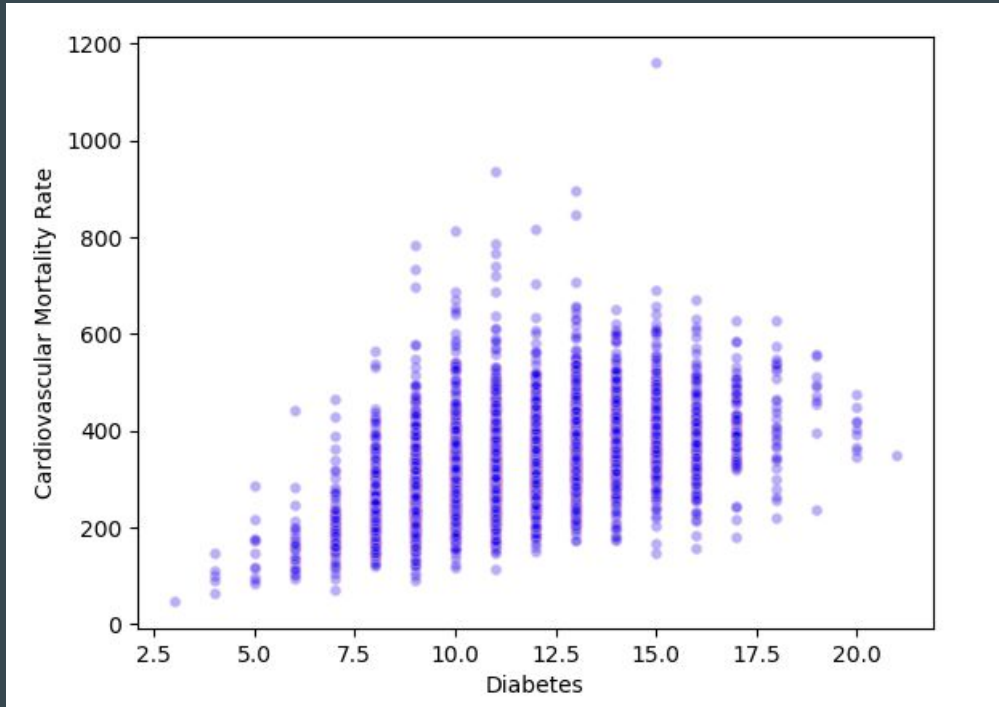
Smoking

- Smoking has the most impact out of any other factor
- Native American reservations have a disproportionate amount of smokers in them



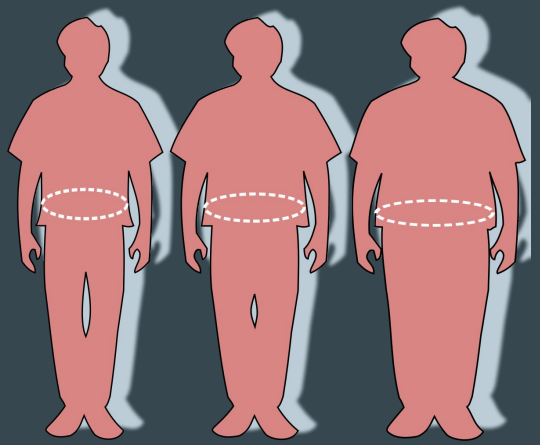
Diabetes

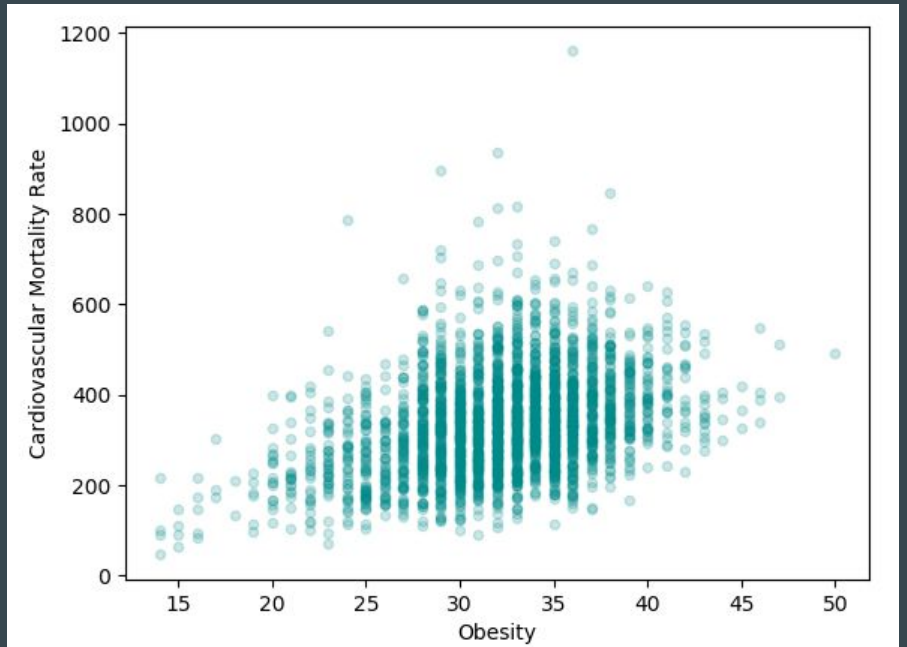
- South western Alabama and Mississippi have some of the highest diabetes counties in the USA.
- Mortality rate seems to peak at 12.5% of diabetic population.



Obesity

- 3rd factor with most impact on cardiovascular mortality rate
- The higher the percentage of obesity(in counties across the U.S.) is, the higher the cardiovascular mortality rate is
- Therefore, there is a correlation between obesity and cardiovascular mortality rate





Conclusions:

- Lifestyle factors have more apparent effects than environmental factors
- County data indicates smoking increases your risk of heart disease at least twice as much as the second leading cause -- similar to individual studies
- The southern and midwest counties tend to have a larger amount of heart disease-related deaths and also engage more in all risk factors

Project Post Mortem

- County-level data is very useful for illustrating the issues and might have value for capturing the effects of factors not otherwise considered
- Drug use data was grouped by regions, when we wanted our data to be grouped by county
  - Group by state for drug use dataset lost a lot of information
  - Most drug user populations too small to use for estimating impacts at county level
- Linear regression provided some useful analysis despite issues ...
  - Assumptions needed to relate county-level and individual impacts
  - Strong correlations and complex causal relationships among factors
  - Counties are not truly random, independent samples


Q&A