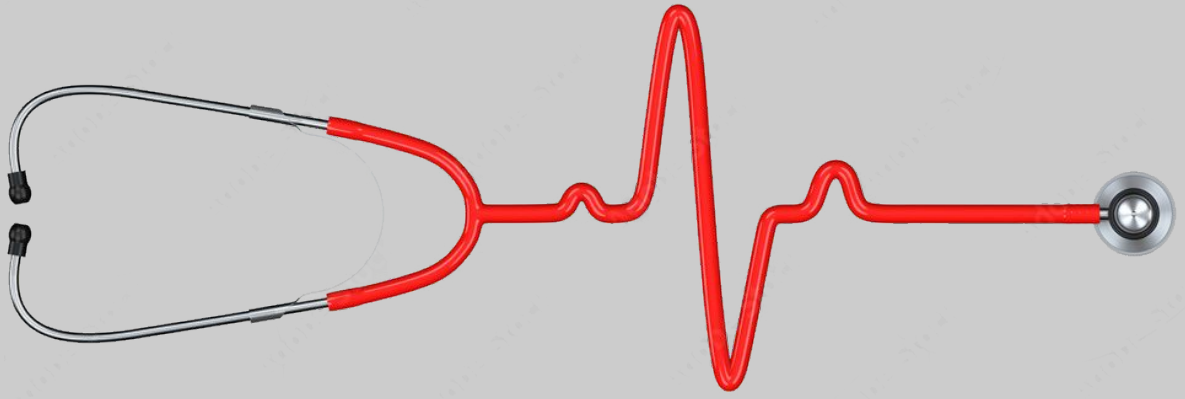


# Stroke Mortality Predictor

University of Minnesota Data Visualization and Analytics Boot Camp

Team 2 — Janice Courtois, Alex Norgren, Tom Pankratz, Rachel Rautenberg

June 9, 2022

# Team 2

Team 2 members all work at Mayo Clinic.

**Janice Courtois**

- Works in Healthcare Technology Management
- Lives on horse ranch in AZ
- Travels often to visit kids & grandson

**Alex Norgren**

- Works in a Lab
- Enjoys to golf

**Tom Pankratz**

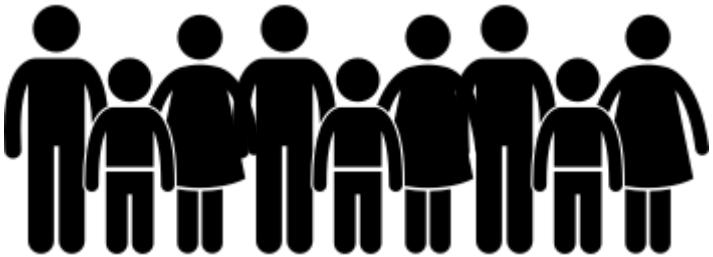
- 19 years at Mayo Clinic
- Manages a digital experimentation team
- Dad of 4

**Rachel Rautenberg**

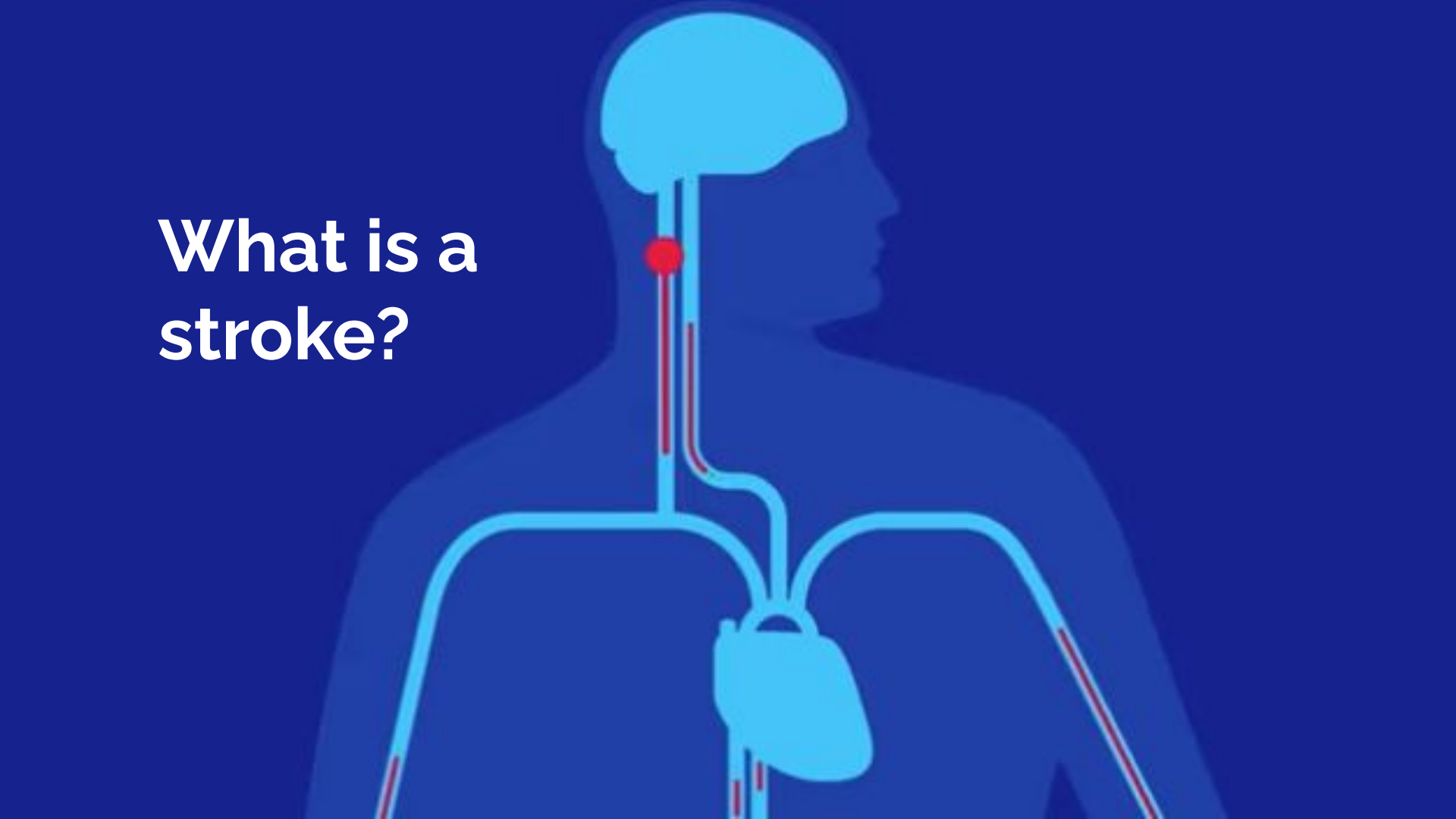
- Holds MHA
- 14 years at Mayo
- Mom of 4
- Enjoys the chaos

# Topic: Stroke mortality

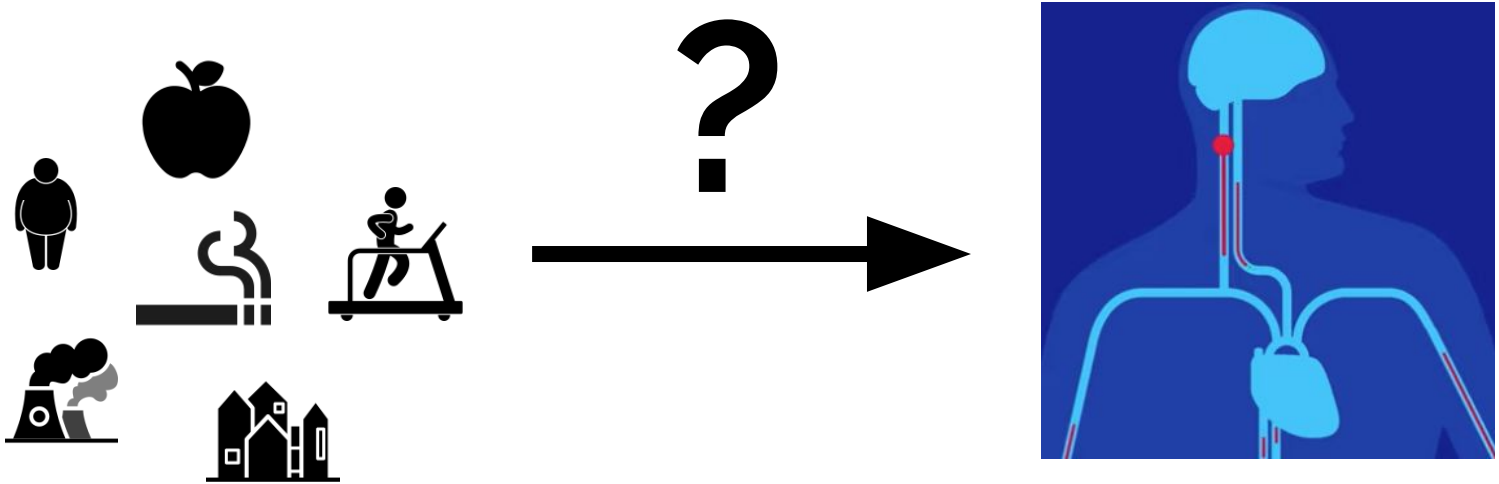
MAYO CLINIC



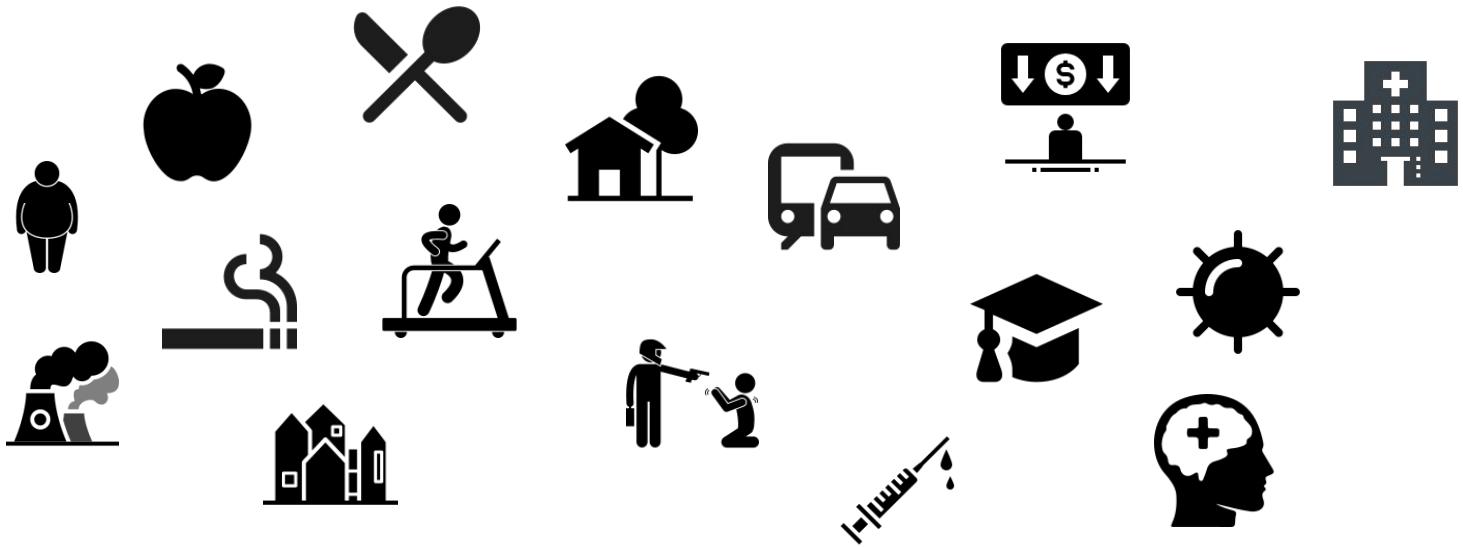
# What is a stroke?



# Goal of project & questions?



# Brainstorming possible factors



# Factors we landed on

Health-related:

- Smoking
- Obesity
- Access to healthy foods
- Access to exercise opportunities
- Primary care availability
- Availability of mental health providers

Social-related:

- College education
- Unemployment
- Income
- Violent crime rate
- Air pollution
- Length and type of commute to work
- Urban vs. rural

# Source data



**Stroke Mortality Data Among US Adults (35+) by State/Territory and County (2018)**

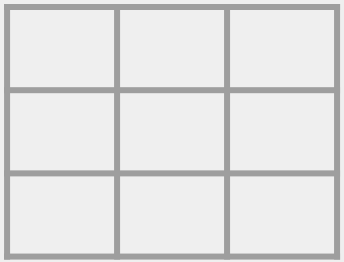
**County Health Rankings (2018)**

# Data exploration and integration


# Target: Stroke mortality dataset

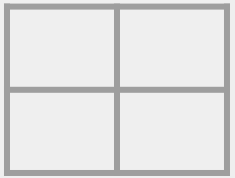
20 columns

59,094 rows



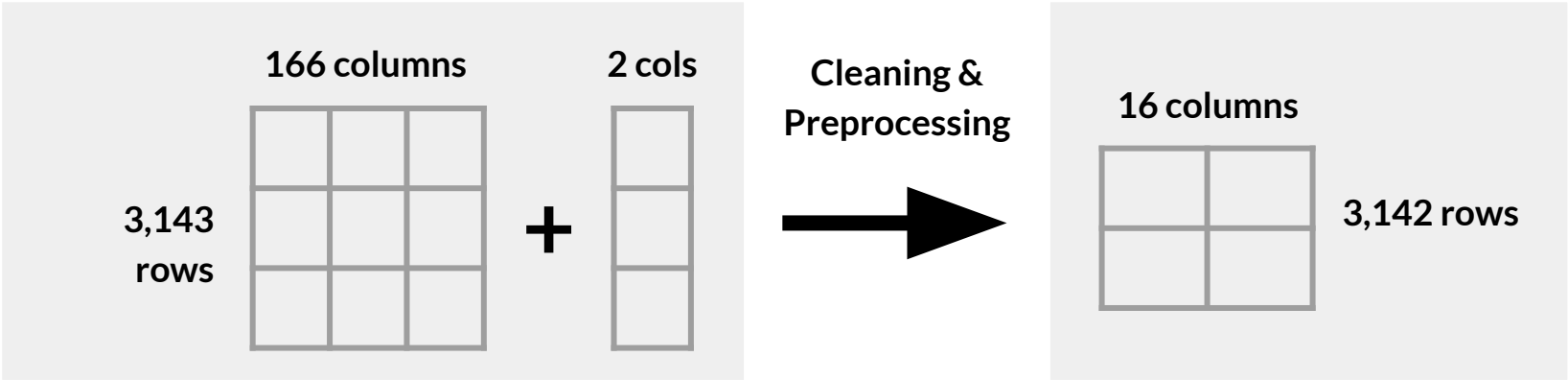
Cleaning &
Preprocessing

2 columns

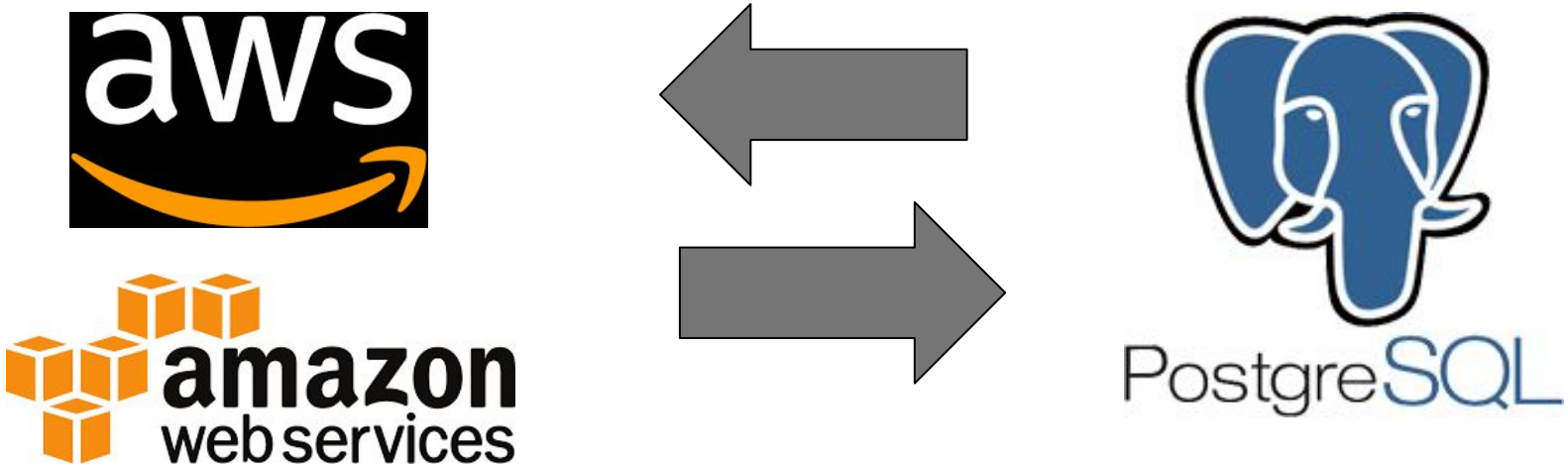


3,142 rows

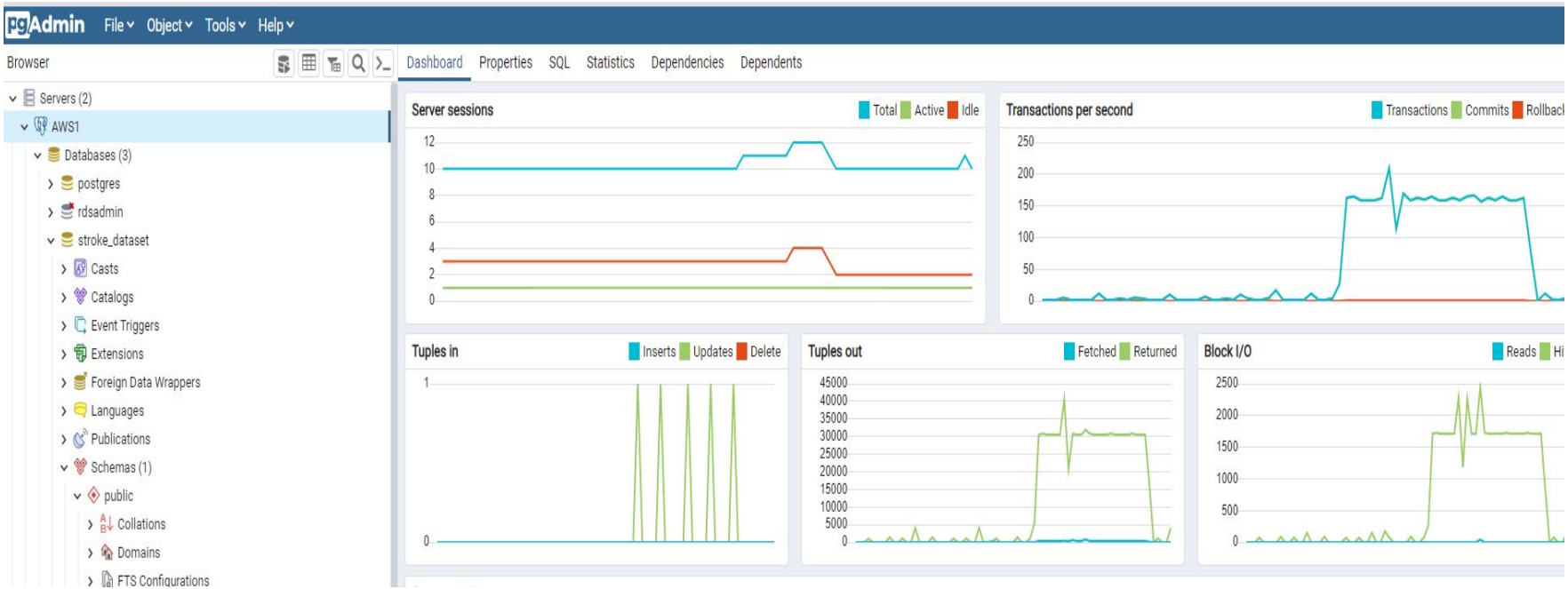
# Features: Health rankings datasets



# Establish Connection to Database

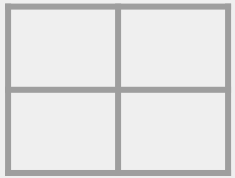


# View of live connection via AWS



# Datasets merge via PostgreSQL

2 columns

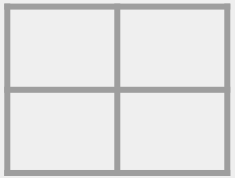


3,142 rows

Stroke Mortality Dataset

+

16 columns

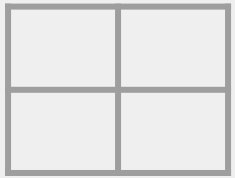


3,142 rows

Health Rankings Dataset

→

17 columns



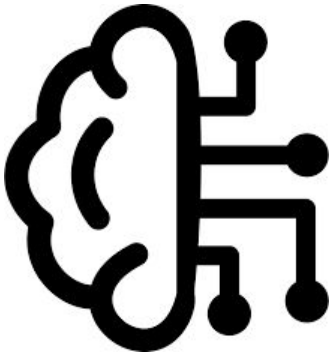
3,142 rows

Clean Dataset


**Ready for Machine Learning!**

# Exploring for Machine Learning

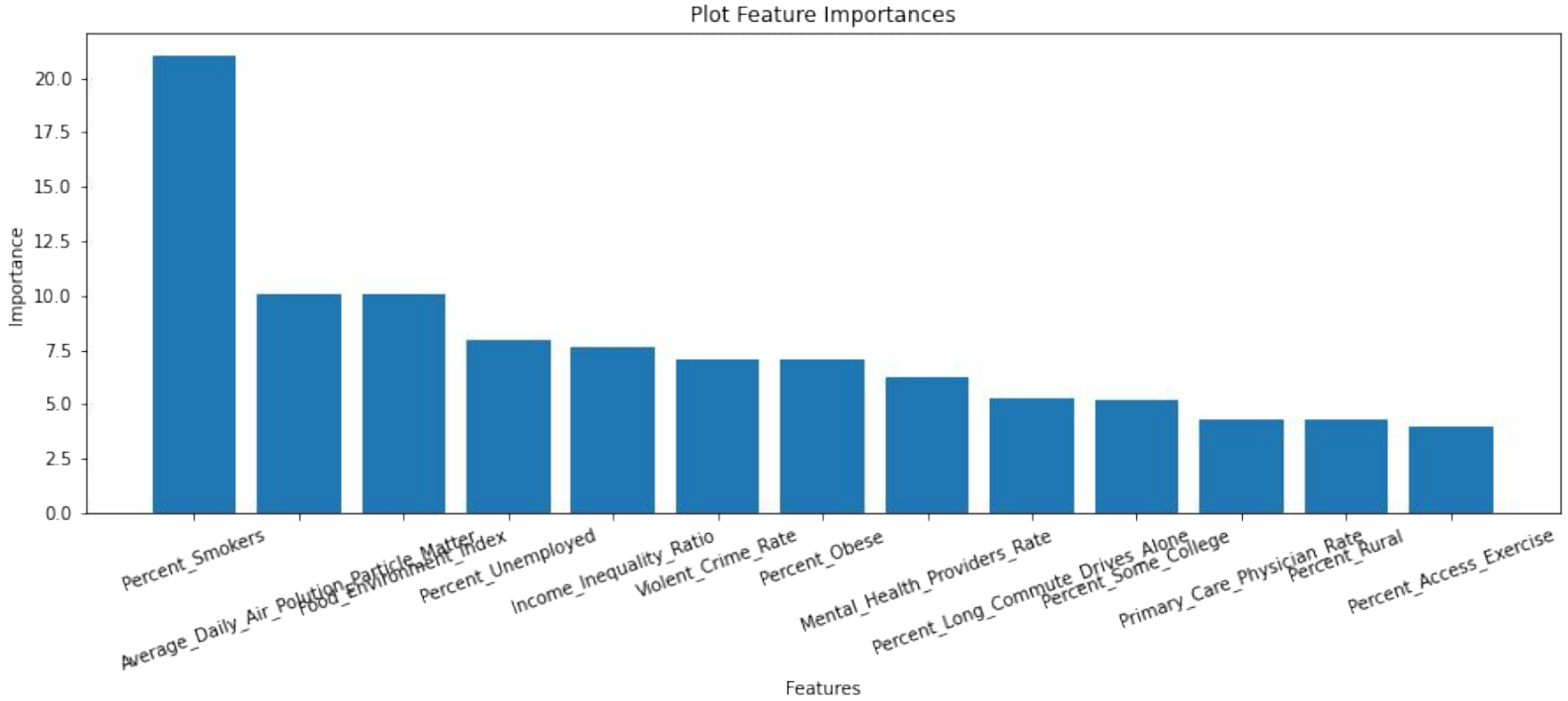
- Decide prediction target
- Research & decide machine learning model
- Build machine learning model



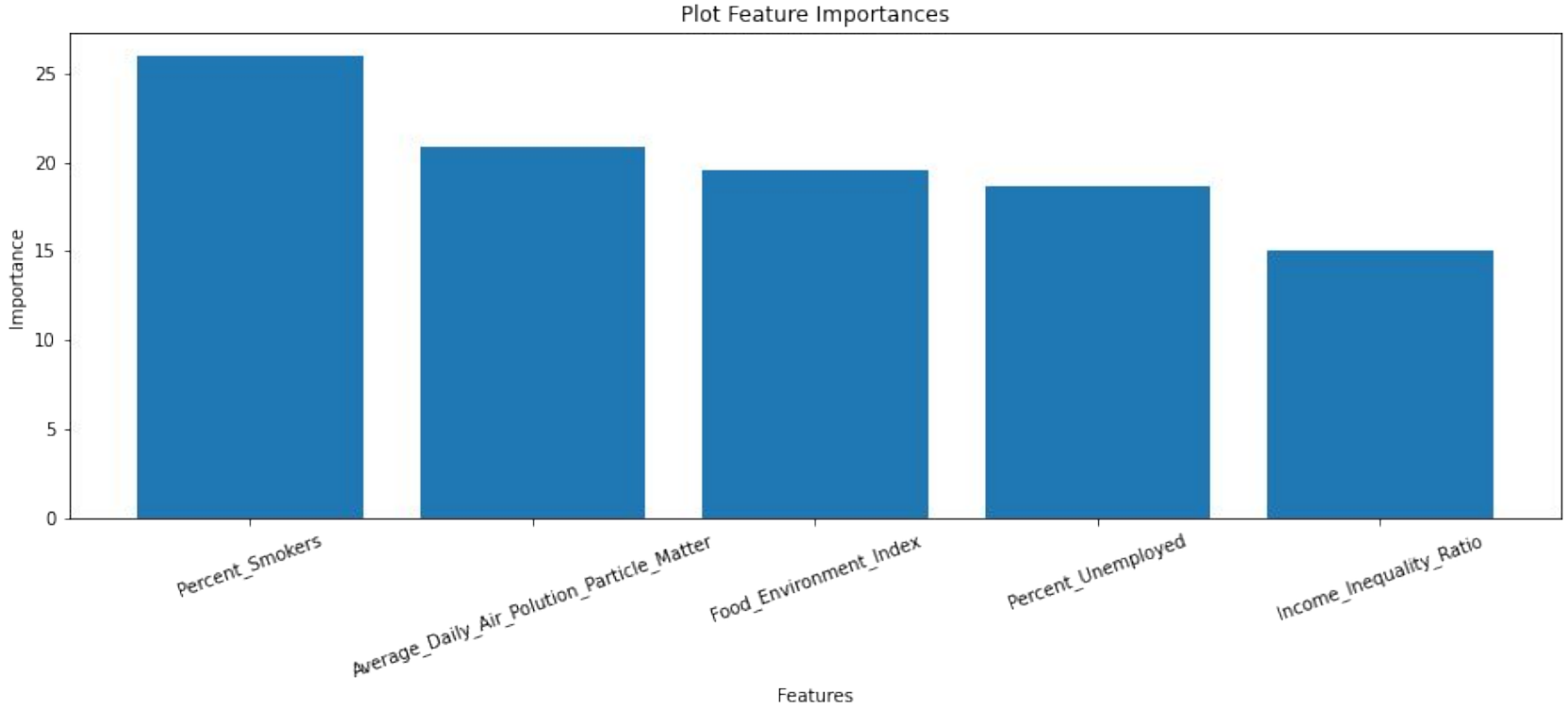
# Machine Learning

- Training & testing split
- Optimization

# Model importances



# Model importances: Top 5



# Results

Did the model work?

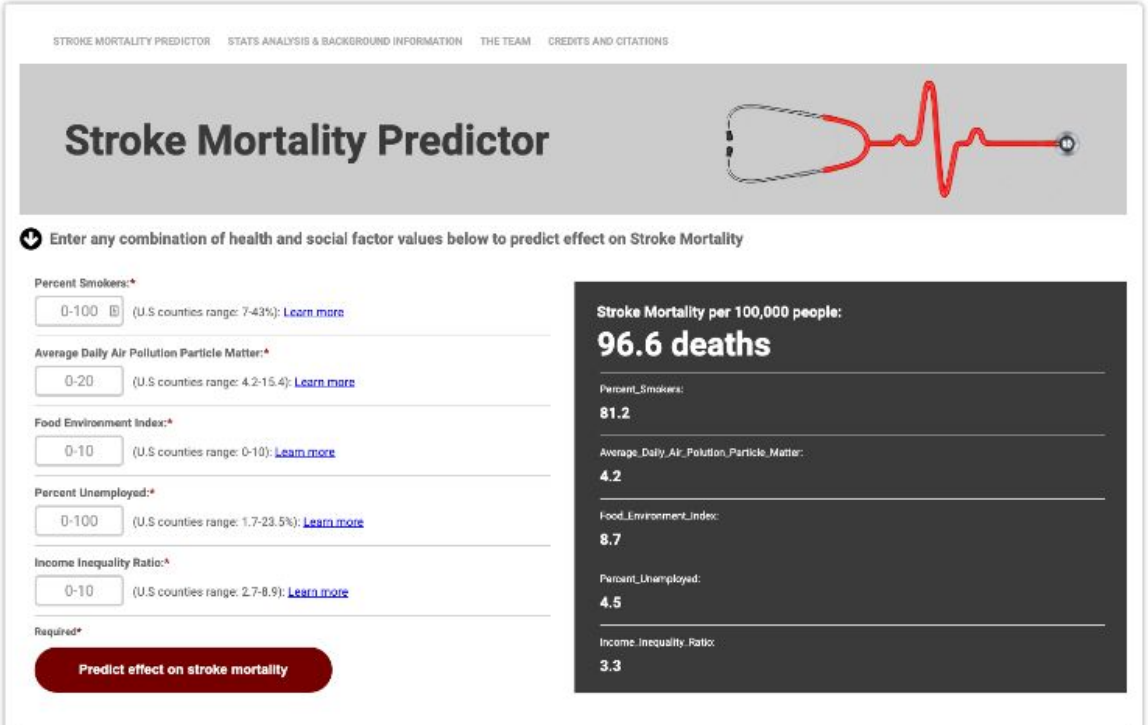
- All data : *12.13*
- Health features: *13.12*
- Economic/Environmental Features: *13.21*

# Model output and usage

We wondered:

Could we input variations of the feature data to determine what sort of effect it would have on stroke mortality?

# The stroke mortality predictor form & dashboard



# Results

**Percent Smokers** appeared to have the largest impact on stroke mortality, but beyond that, it was difficult to determine impacts from the other features. It was likely less a matter of machine learning model choice, and more a matter of the choice of features/factors.

# Recommendations
## (what could we have done differently?)

As a starting point, we could have chosen features/factors that have already been determined by the health care community to have a larger impact on predicting stroke mortality.

According to Mayo Clinic.org, here are several:

- Physical inactivity
- Heavy or binge drinking
- Use of illegal drugs such as cocaine and methamphetamine
- High blood pressure
- Cigarette smoking or secondhand smoke exposure
- High cholesterol
- Diabetes

# Questions?