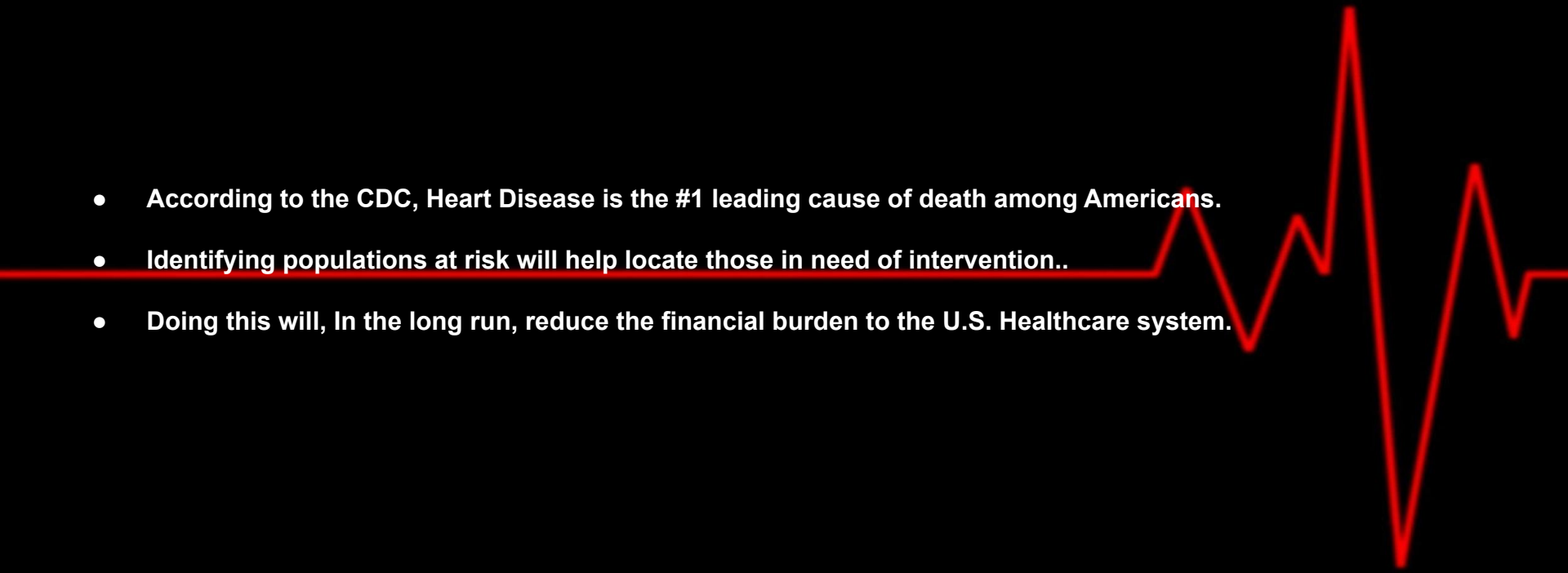
Populations at Risk for CHD

By: Zachary Greenberg & Jonathan Silverman

Business Problem



- **According to the CDC, Heart Disease is the #1 leading cause of death among Americans.**
- **Identifying populations at risk will help locate those in need of intervention..**
- **Doing this will, In the long run, reduce the financial burden to the U.S. Healthcare system.**

CHD Data

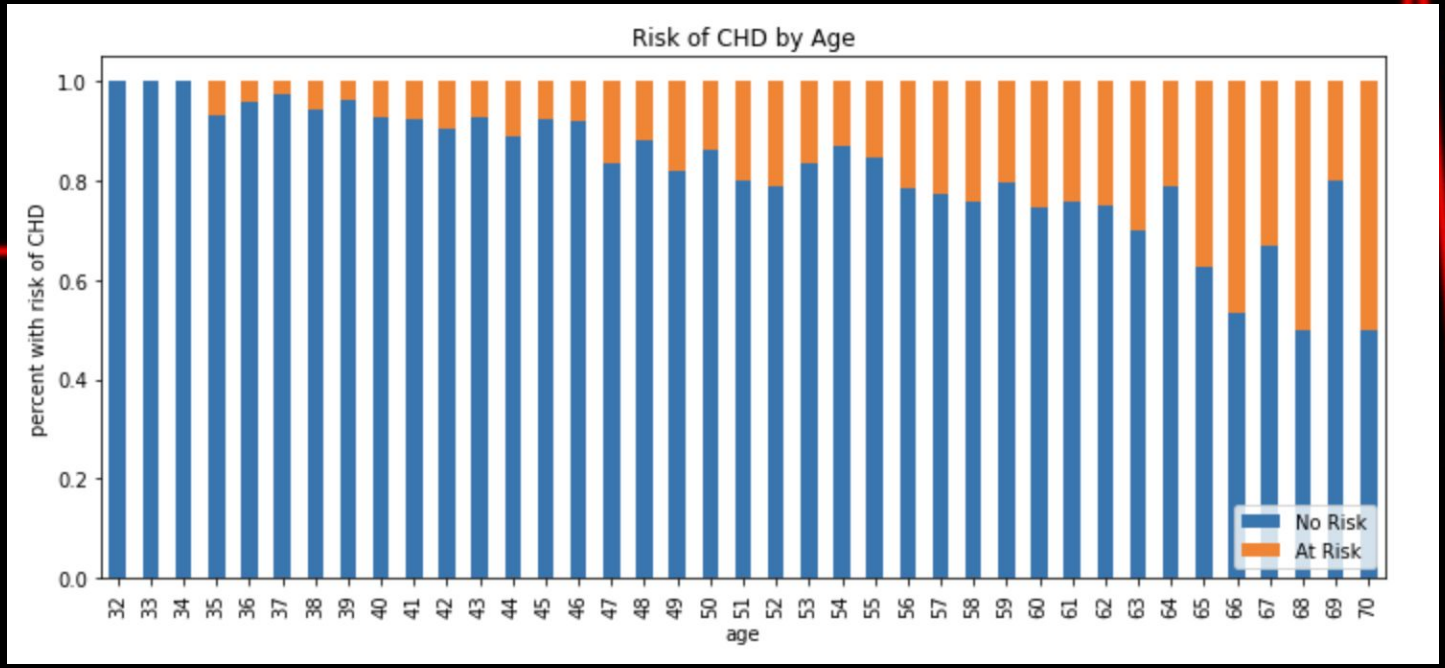
- Public Dataset from Kaggle.
- Data on 4000 individuals
- Fifteen Measures of Health

Information provided:

Sex	Stroke	Diastolic Blood Pressure
Age	Hypertension	Body Mass Index
Smoking	Diabetes	Heart Rate
Cigarettes Per Day	Total Cholesterol	Glucose
BP Meds	Systolic Blood Pressure	**Ten Year Risk for CHD**

Data Exploration

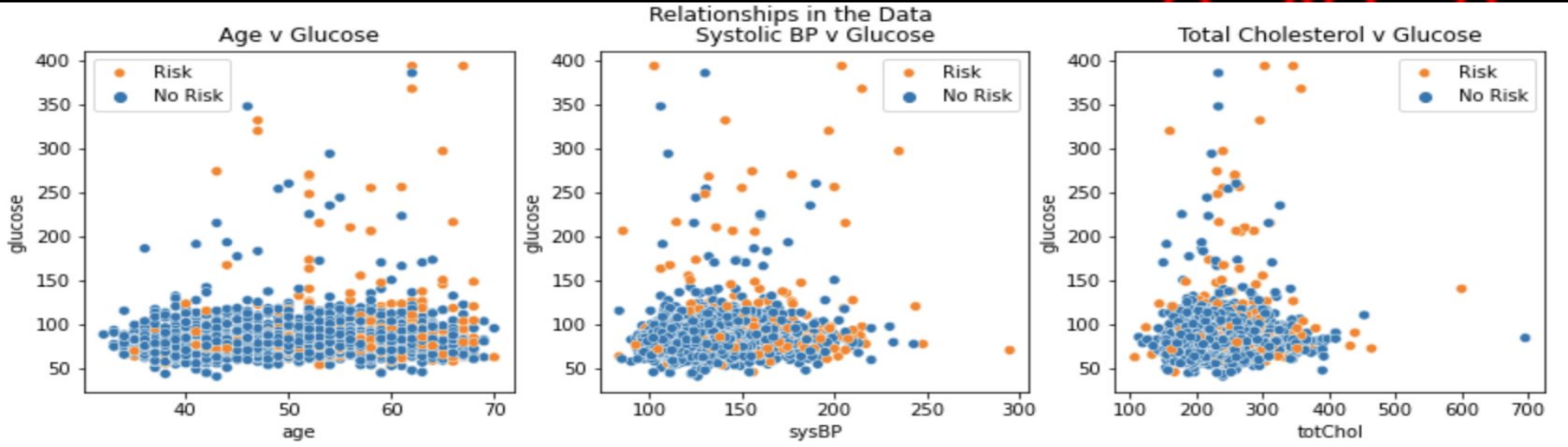
The risk of CHD seems to exponentially increase with age.



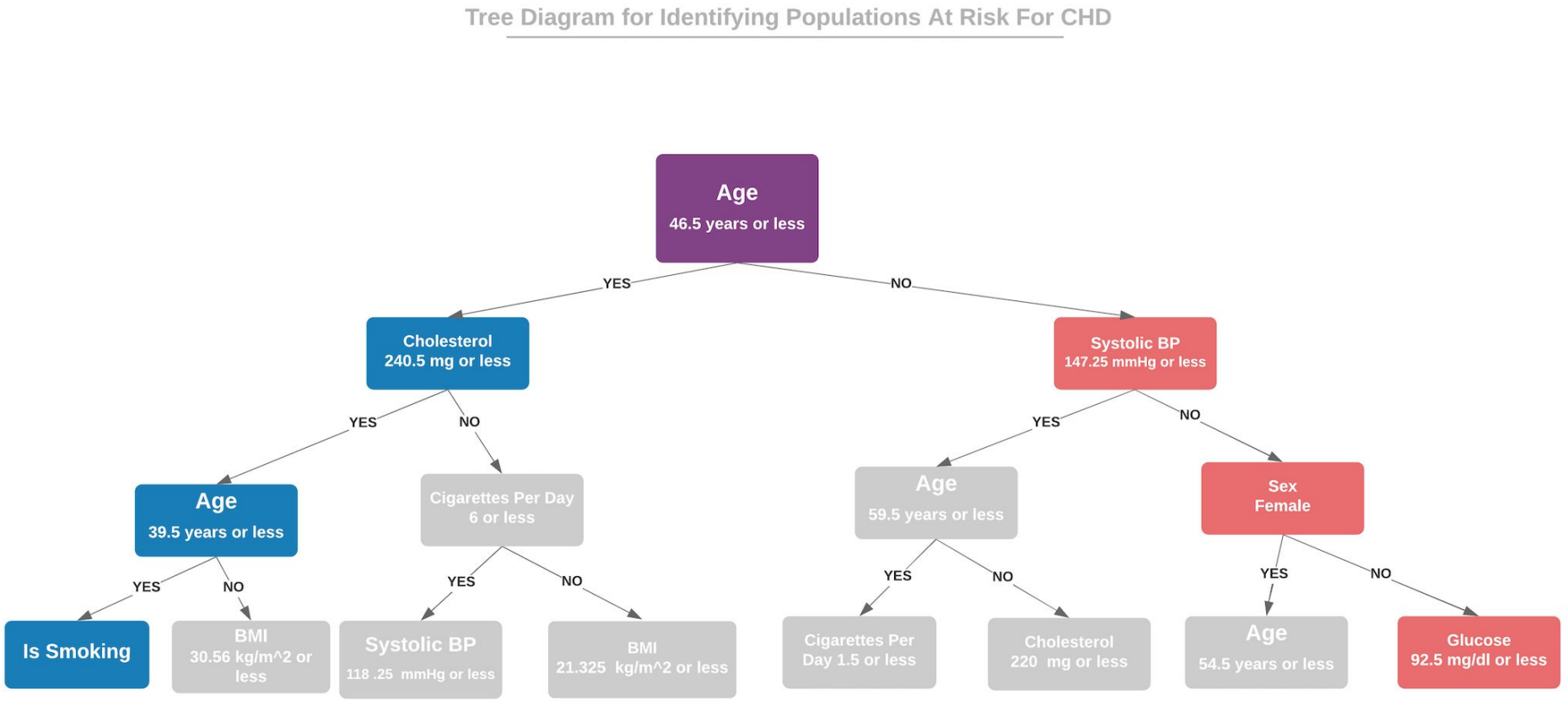
Feature Engineering
from further exploration

Given 2 conditions, the risk of CHD is made clear.

For example if age is greater than or equal to 50, and glucose greater than or equal to 140, 60% of people at in the dataset were at risk for CHD.



Modeling

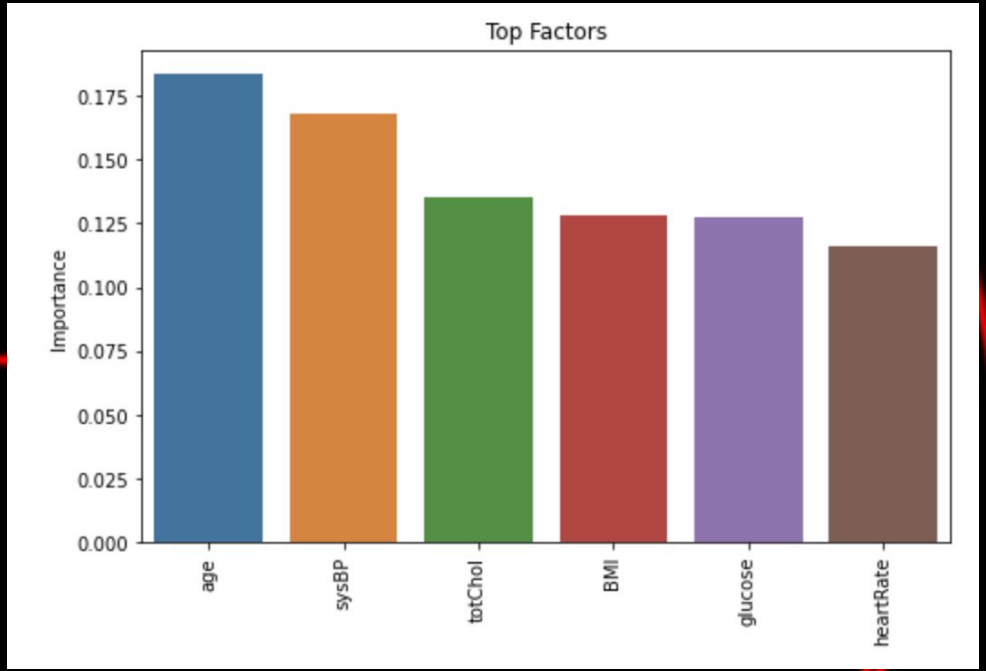

Tree Diagram for Identifying Populations At Risk For CHD
Age
46.5 years or less
YES
NO
Cholesterol
240.5 mg or less
Systolic BP
147.25 mmHg or less
YES
NO
YES
NO
Age
39.5 years or less
Cigarettes Per Day
6 or less
Age
59.5 years or less
Sex
Female
YES
NO
YES
NO
YES
NO
YES
NO
Is Smoking
BMI
30.56 kg/m^2 or less
Systolic BP
118 .25 mmHg or less
BMI
21.325 kg/m^2 or less
Cigarettes Per Day 1.5 or less
Cholesterol
220 mg or less
Age
54.5 years or less
Glucose
92.5 mg/dl or less

Results

The top 6 risk factors are:

Systolic BP, BMI, Cholesterol,

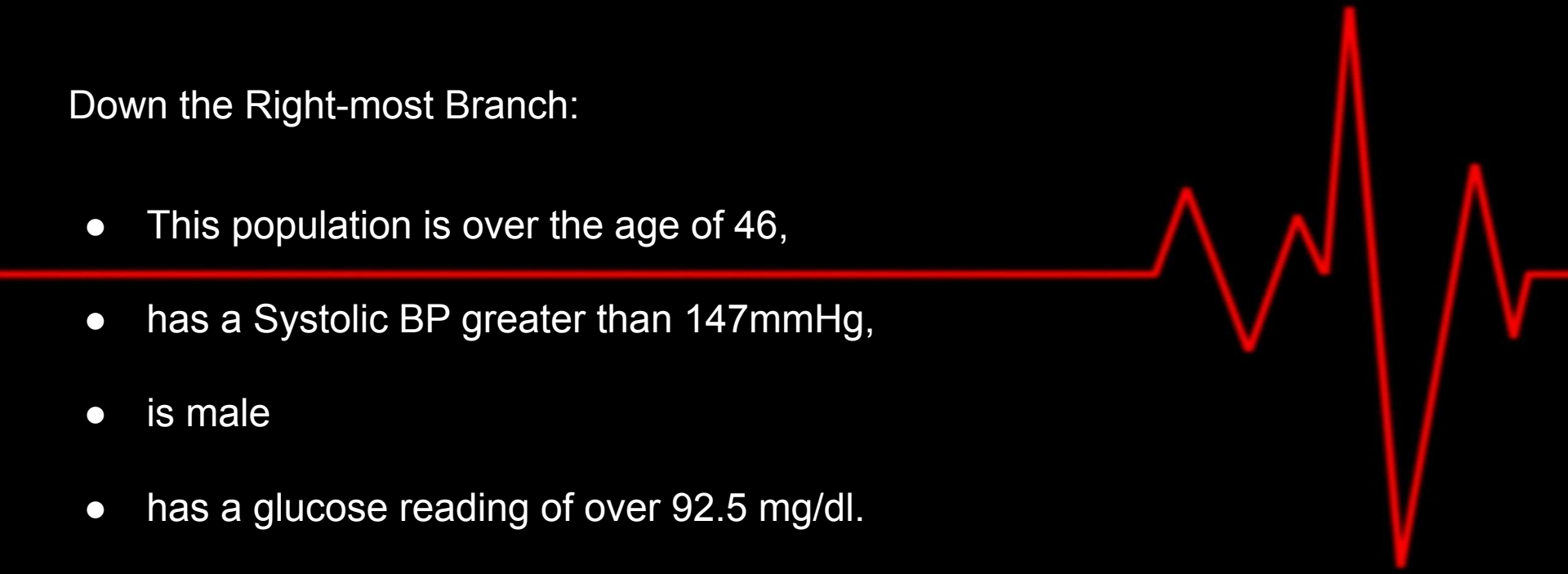
Age, Heart rate, & Glucose Levels.



Population Most At Risk

Down the Right-most Branch:

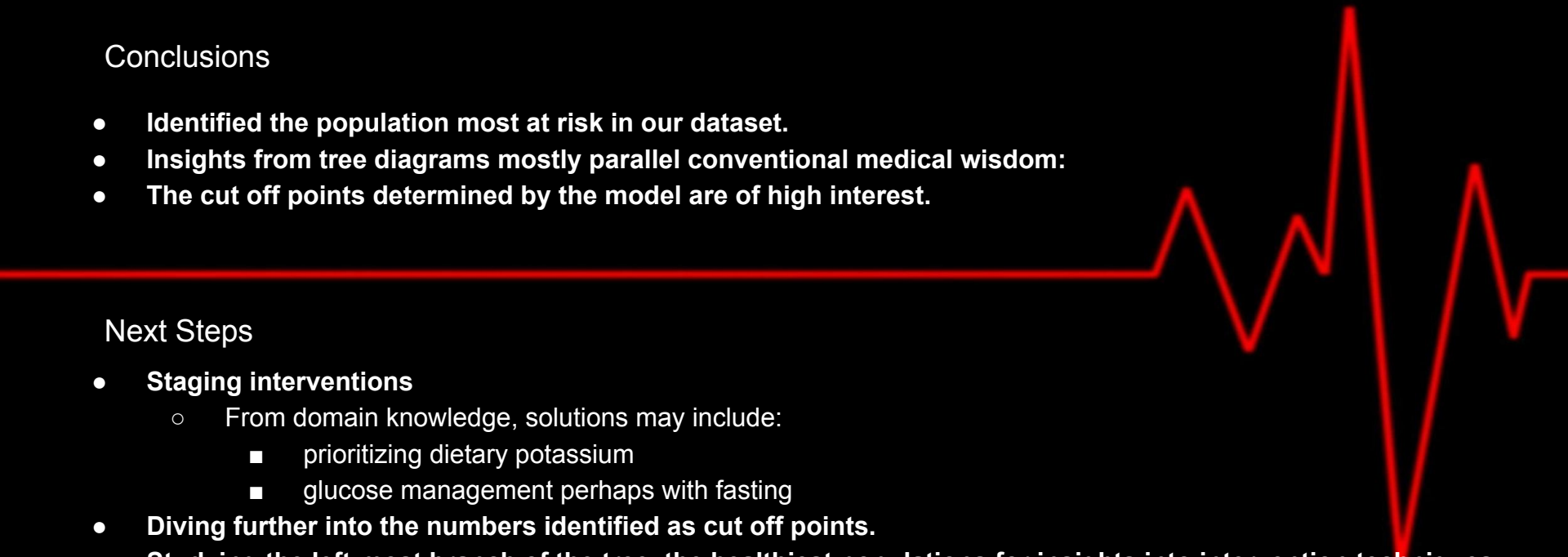
- This population is over the age of 46,
- has a Systolic BP greater than 147mmHg,
- is male
- has a glucose reading of over 92.5 mg/dl.



Conclusions & Next Steps

Conclusions

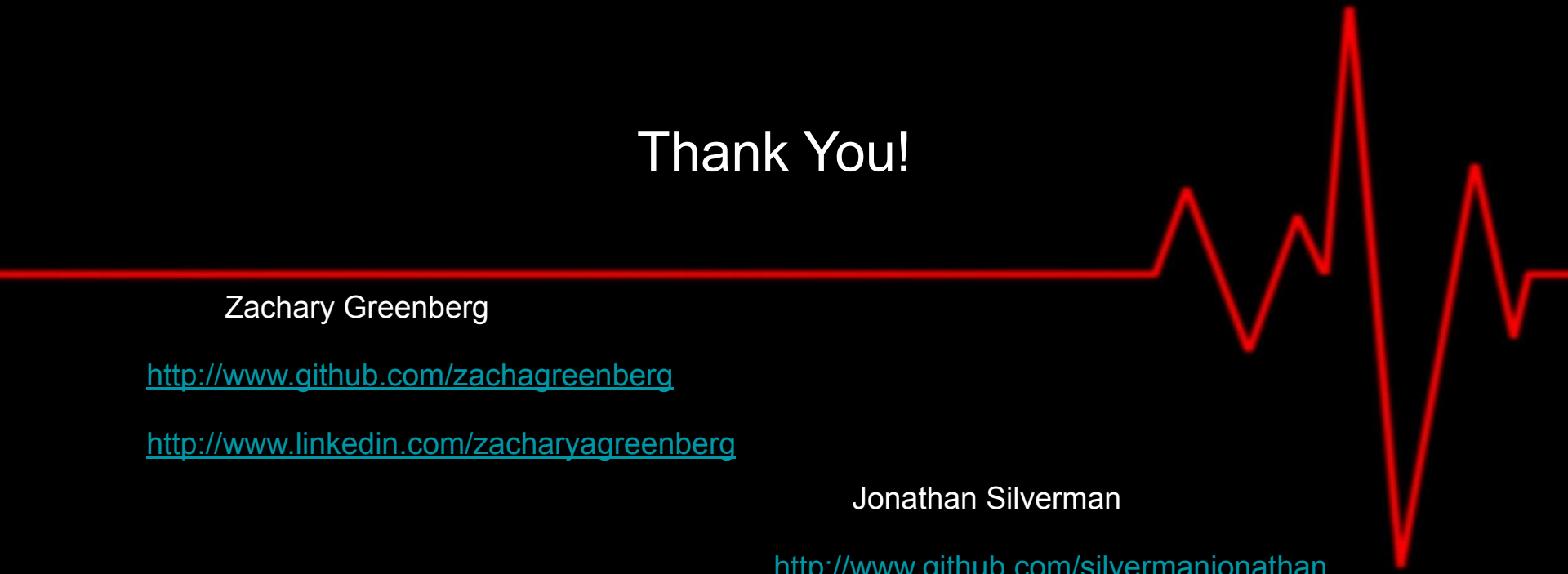
- **Identified the population most at risk in our dataset.**
- **Insights from tree diagrams mostly parallel conventional medical wisdom:**
- **The cut off points determined by the model are of high interest.**



Next Steps

- **Staging interventions**
  - From domain knowledge, solutions may include:
    - prioritizing dietary potassium
    - glucose management perhaps with fasting
- **Diving further into the numbers identified as cut off points.**
- **Studying the left-most branch of the tree, the healthiest populations for insights into intervention techniques**

Thank You!



Zachary Greenberg

http://www.github.com/zachagreenberg

http://www.linkedin.com/zacharyagreenberg

Jonathan Silverman

http://www.github.com/silvermanjonathan

https://www.linkedin.com/in/jonathansilverman007/