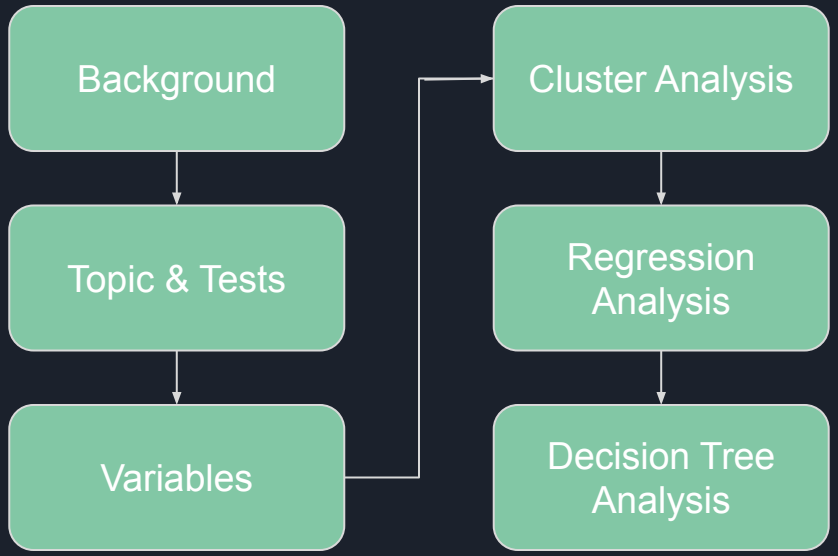
Key Indicators of Heart Disease

Andrey Burns, Daniel Gillinger, Colin Rehnert, Cody Turpin

Overview of Our Presentation



Selected Topic and Tests

- Dataset: Personal Key Indicators of Heart Disease
  - Survey conducted by the Center for Disease Control (CDC) to help determine what risk factor causes heart disease
- Heart disease is one of the leading cause of deaths in the United States
  - About half (47%) of Americans have at least 1 of 3 key risk factors for heart disease which are:
    - High blood pressure, high cholesterol, and diabetes
- 300,000+ respondents to CDC survey
  - 8.6% of respondents reported having heart disease
- Conducted logistic regression, cluster analysis, and a decision tree

Variables from the Dataset

- Heart Disease - Yes, No (Target Variable)
- Race
- Sex
- Age Category
- BMI Category
- How many hours on average do you sleep?
- How can you define your general health?
- For how many days during the past 30 days was your physical health not good?
- For how many days during the past 30 days was your mental health not good?
- Have you played any sports (running, biking, etc.) in the past month?
- Have you smoked at least 100 cigarettes in your entire life (approx. 5 packs)?
- Do you have more than 14 drinks of alcohol (men) or more than 7 (women) in a week?
- Did you have a stroke?
- Do you have serious difficulty walking or climbing stairs?
- Have you ever had diabetes?
- Do you have asthma?
- Do you have kidney disease?
- Do you have skin cancer?

Regression Analysis

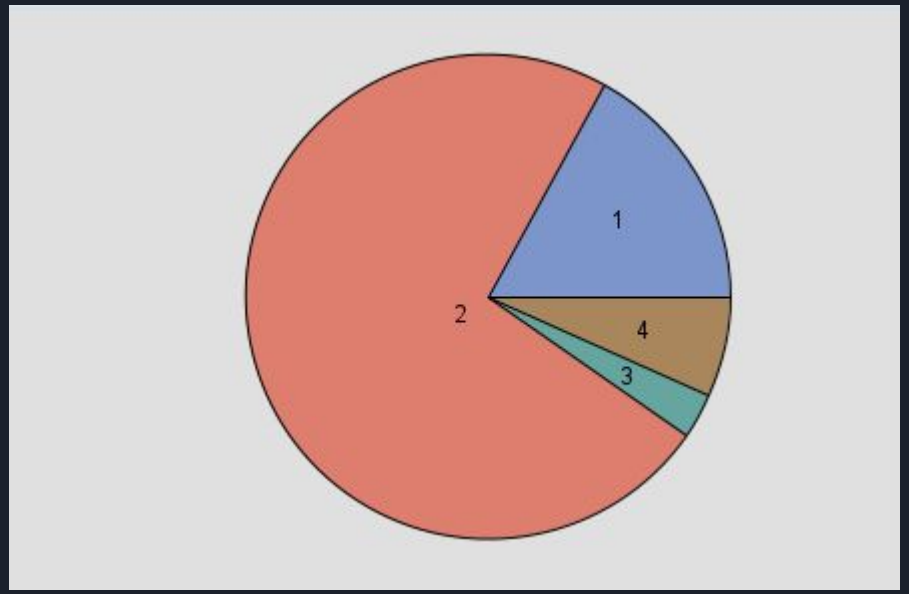
Effect	DF	Wald Chi-Square	Pr > ChiSq
AgeCategory	12	7595.5875	<.0001
AlcoholDrinking	1	51.4398	<.0001
Asthma	1	208.5260	<.0001
BMI	1	54.9867	<.0001
Diabetic	3	814.1896	<.0001
Difficulty_Walking_Up_Stairs	1	137.7417	<.0001
GenHealth	4	3476.9581	<.0001
KidneyDisease	1	542.5172	<.0001
MentalHealth____of_Days_of_Poor	1	28.5515	<.0001
PhysicalActivity	1	1.5308	0.2160
PhysicalHealth____of_Days_Exerci	1	13.5465	0.0002
Race	5	155.9119	<.0001
Sex	1	2361.6469	<.0001
SkinCancer	1	34.4265	<.0001
SleepTime__Hours_per_Night_	1	33.3417	<.0001
Smoking	1	614.6773	<.0001
Stroke	1	2140.1148	<.0001

Group Takeaways from the Regression Analysis

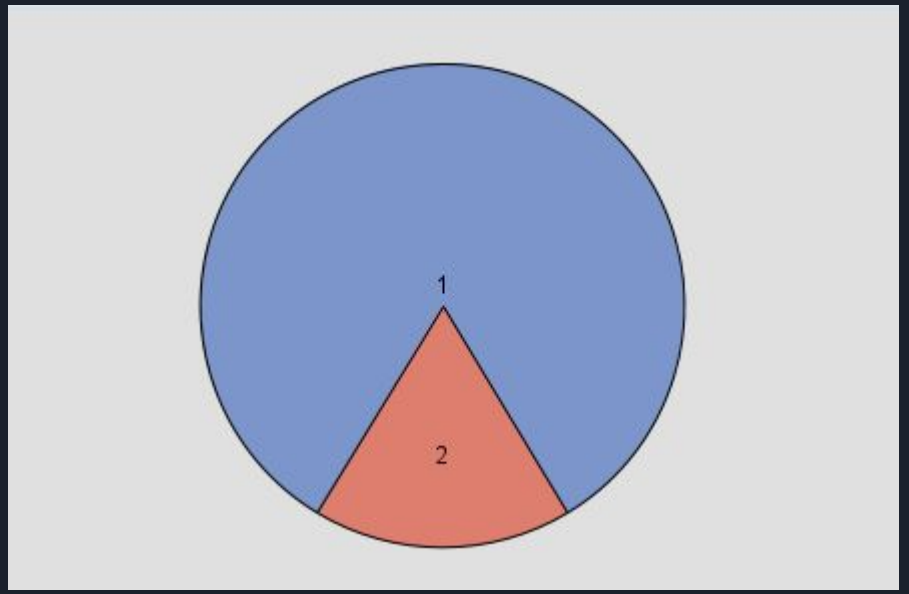
- Physical activity was determined to be insignificant
  - P-value = 21.6%
- Reasons why:
  - Younger people might not be active and still will not develop heart disease
  - Older people with heart disease might start trying to be active to improve their health
  - Likely due to short period of time (30 days)
  - If longer, it would likely be a significant factor
- A closer look to age is necessary

Cluster Analysis: Segment Size

Ward



Centroid



Cluster Analysis (Hierarchical Ward)

- Group 1:
  - Participants: 46,986
  - Physical Activity: 25% No (75% Yes)
  - Diabetic: 77% No (23% Yes)
  - Stroke: 100% No (0% Yes)
  - Heart Disease: 67% No (33% Yes)
- Group 2:
  - Participants: 204,301
  - Physical Activity: 15% No (85% Yes)
  - Diabetic: 91% No (9% Yes)
  - Stroke: 100% No (0% Yes)
  - Heart Disease: 100% No (0% Yes)
- Group 3:
  - Participants: 8,597
  - Physical Activity: 33% No (67% Yes)
  - Diabetic: 69% No (31% Yes)
  - Stroke: 0% No (100% Yes)
  - Heart Disease: 67% No (33% Yes)
- Group 4:
  - Participants: 18,212
  - Physical Activity: 41% No (59% Yes)
  - Diabetic: 74% No (26% Yes)
  - Stroke: 100% No (0% Yes)
  - Heart Disease: 88% No (12% Yes)

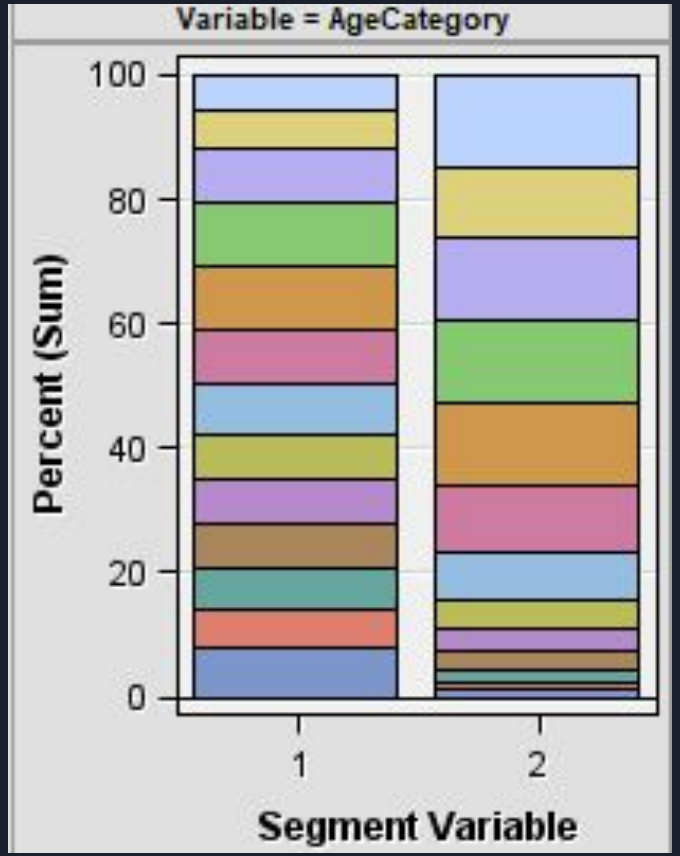
Cluster Analysis (Hierarchical Centroid)

- Group 1:
  - Participants: 265,038
  - Physical Activity: 15% No (85% Yes)
  - Diabetic: 87% No (13% Yes)
  - Stroke: 98% No (2% Yes)
  - Heart Disease: 95% No (5% Yes)
- Group 2:
  - Participants: 54,757
  - Physical Activity: 57% No (43% Yes)
  - Diabetic: 62% No (38% Yes)
  - Stroke: 83% No (17% Yes)
  - Heart Disease: 71% No (29% Yes)

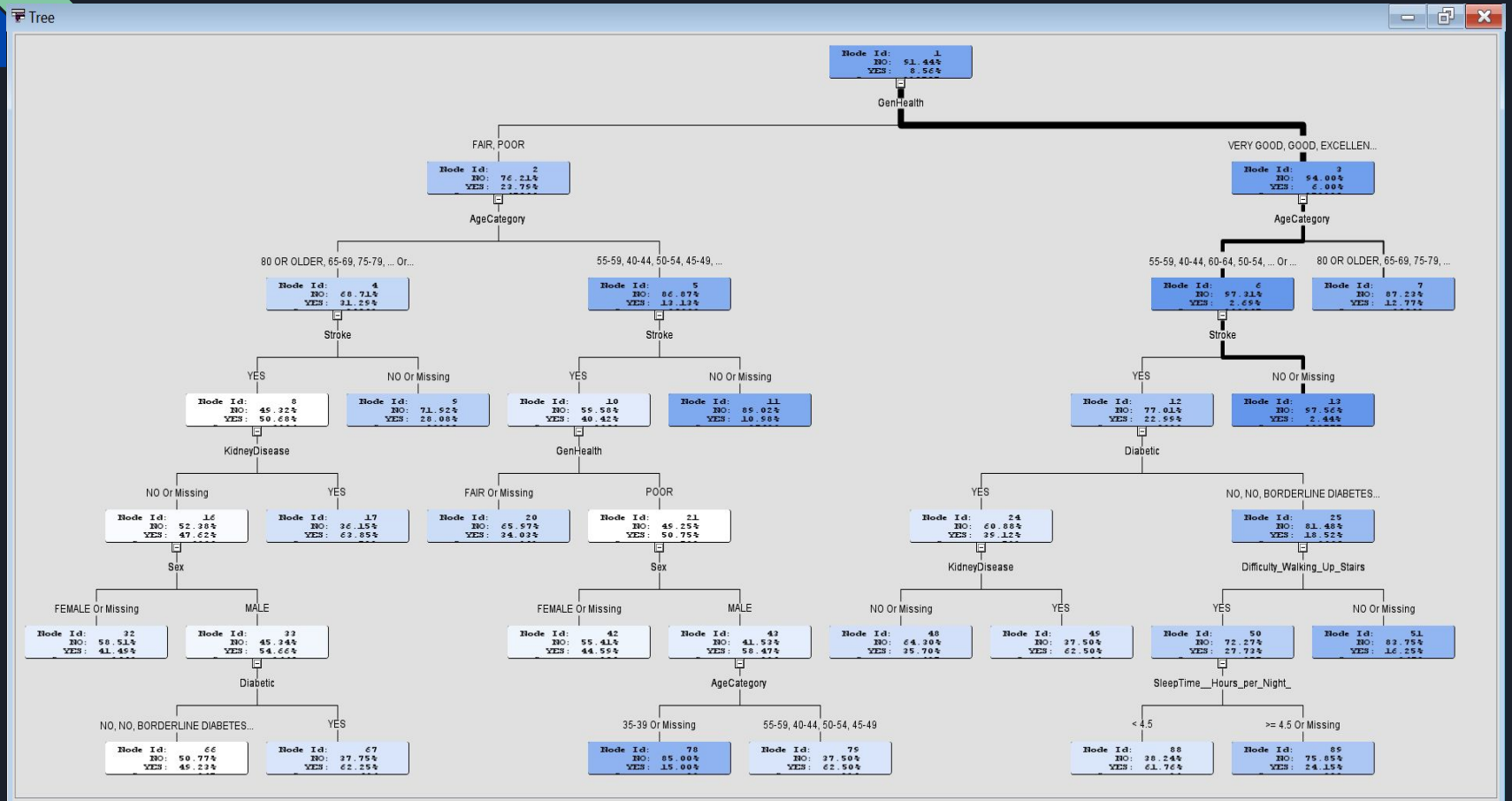
Cluster Analysis (Hierarchical Centroid)

- Group 1:
  - Participants: 265,038
  - BMI: 27.70 (Overweight)
  - Days of poor mental health: 2.95
  - Days of poor physical health: 1.09
- Group 2:
  - Participants: 54,757
  - BMI: 31.33 (Obese)
  - Days of poor mental health: 8.47
  - Days of poor physical health: 14.42

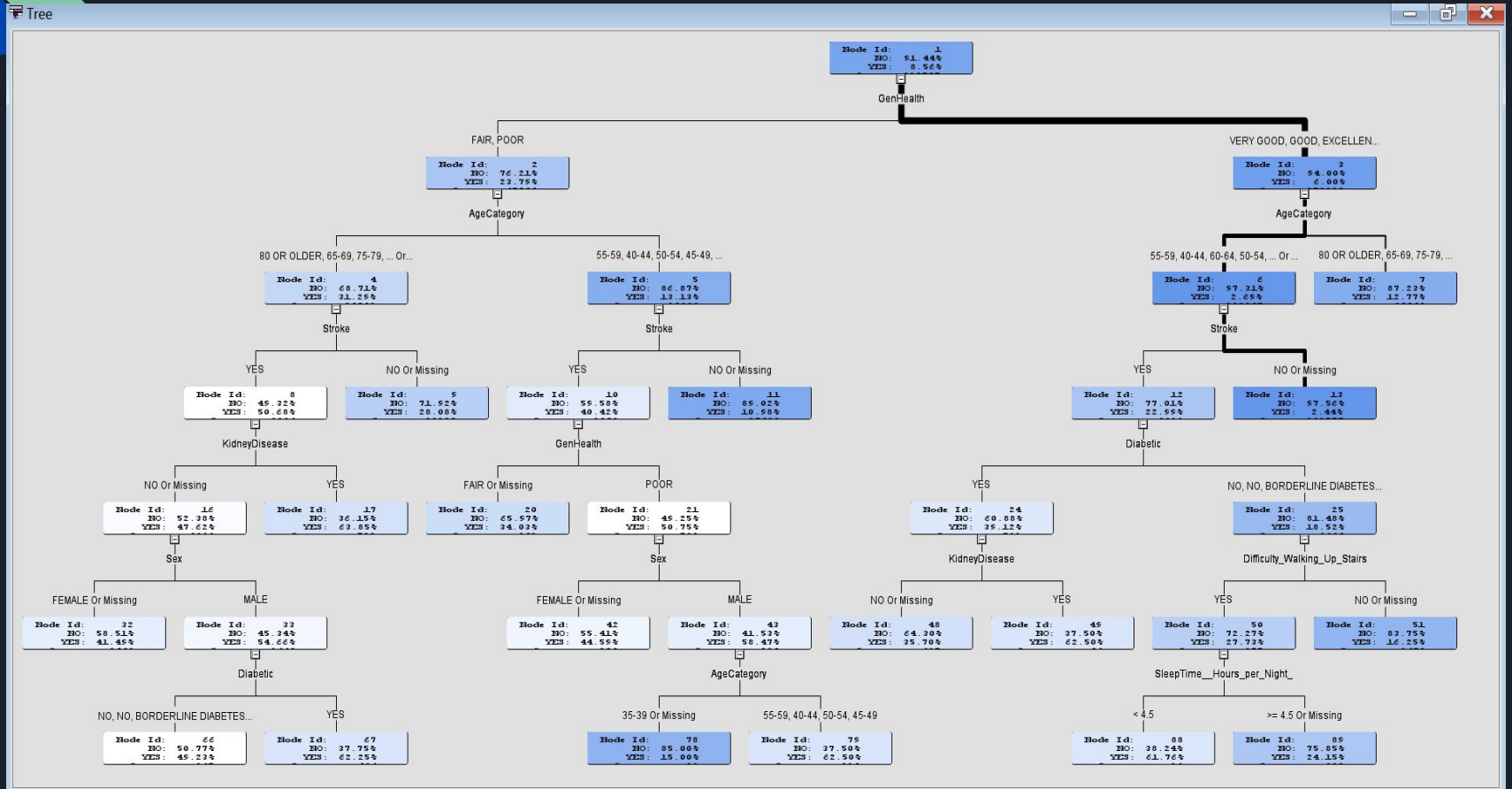
Cluster Analysis: Segment Plot



Decision Tree Analysis (With Physical Activity)



Decision Tree Analysis (Without Physical Activity)



Decision Tree Analysis

Decision Tree Analysis:

- According to the decision tree, those who have good health, are younger, and who have not had a stroke before are the least likely to have heart disease
- Those who had poor health, are older, and have had a stroke before are most likely to have heart disease.
- The “physical activity” variable was removed because it was determined that it was not significant by the regression analysis

Conclusion

- Age is a key predictor of heart disease according to our analysis
    - This is because with age comes a higher likelihood of developing other illnesses like diabetes that contribute to heart disease
- Although physical activity was not very significant in the logistical regression, it does not mean physical activity is not important to preventing heart disease

Citations

- Kaggle: https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease
- CDC: https://www.cdc.gov/brfss/annual_data/annual_2020.html