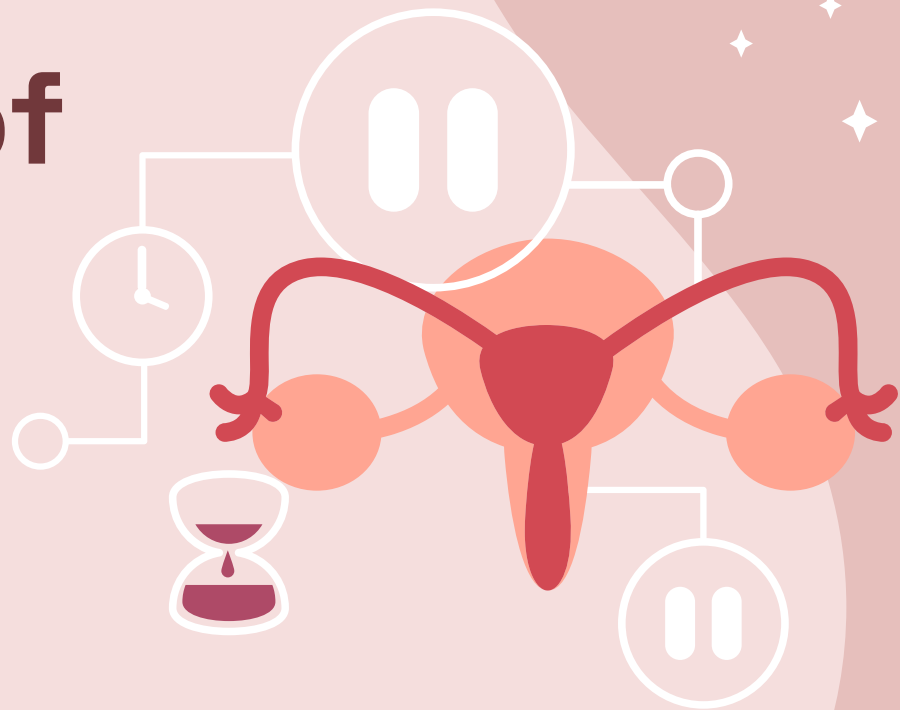




Identifying Predictors of Receiving Insufficient OB-GYN Healthcare



By: Githika Annapureddy and Awa Ndom



Project objective

Not receiving sufficient OB-GYN healthcare delays treatment and causes further complications

To overcome healthcare inequality, we need to be made aware of what factors serve as barriers to treatment.

Effects
292 Participants
or
10% of
Participants

Reported reasons for not receiving needed healthcare

Value	Label	Unweighted Frequency	%
1	Not covered by ins/health plan	27	1.0 %
2	Cannot afford	48	1.7 %
3	Travel distance/lack of transport	2	0.1 %
4	No health care provider	16	0.6 %
5	Too busy/no time	105	3.8 %
6	Other	76	2.8 %

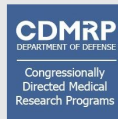




01

Our Data

SWAN is co-sponsored by:



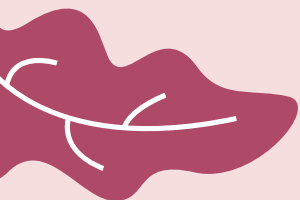
- a.** Study of Women's Health Across the Nation (SWAN), 1998-2001: Family Medical History From Visits 02, 03, and 04 (ICPSR 30181)

data: Family History of general health problems
Extensive family history (mother, sisters, aunts, grandparents) of OB-GYN health problems

- b.** Study of Women's Health Across the Nation (SWAN): Visit 02 Dataset, [United States], 1998-2000 (ICPSR 29401)

data: Individual History of Mental Health, Daily Medicine, Exercise, General Health Problems, OB-GYN Health Problems, Work life, Sex life, Income, Note of major life events and how much it affected participant, General OB-GYN Health Information.

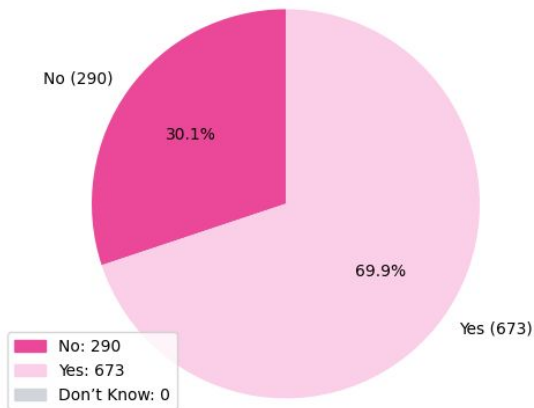
Datasets link across SWANID. 96% of participants in Study A (2729 participants) are also in Study B.



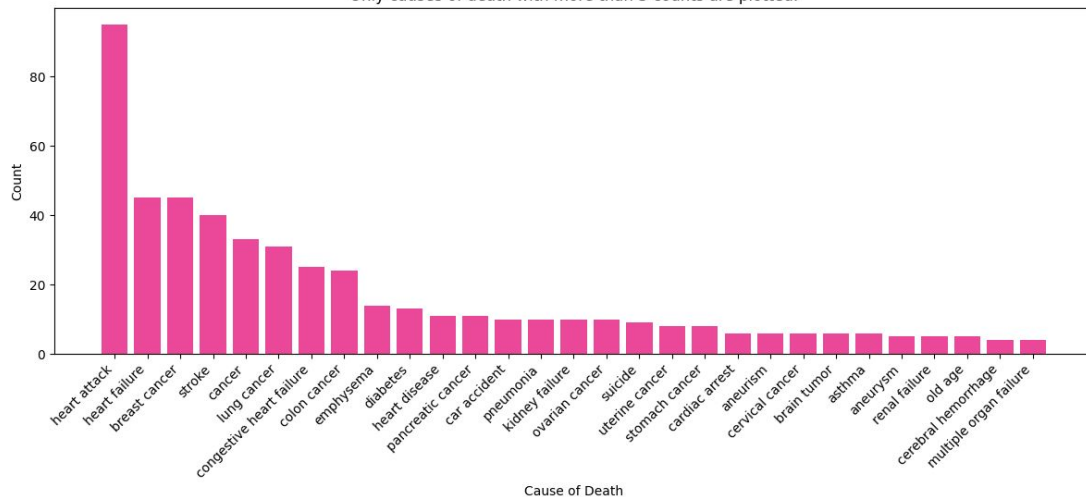
02

Exploratory Data Analysis

HEVYPER2 Chart



MOMCOD2 Distribution
Only causes of death with more than 3 counts are plotted.



Sutton-Tyrell. Study of Women's Health Across the Nation (SWAN), 1998-2001: Family Medical History From Visits 02, 03, and 04. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2014-02-13. <https://doi.org/10.3886/ICPSR30181.v1>

Sutton-Tyrell. Study of Women's Health Across the Nation (SWAN): Visit 02 Dataset, [United States], 1998-2000. Inter-university Consortium for Political and Social Research [distributor], 2019-05-02. <https://doi.org/10.3886/ICPSR29401.v4>



03

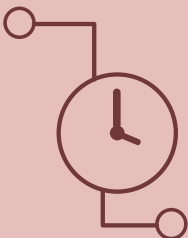
Technical Approach/Methodology

Encoding Features

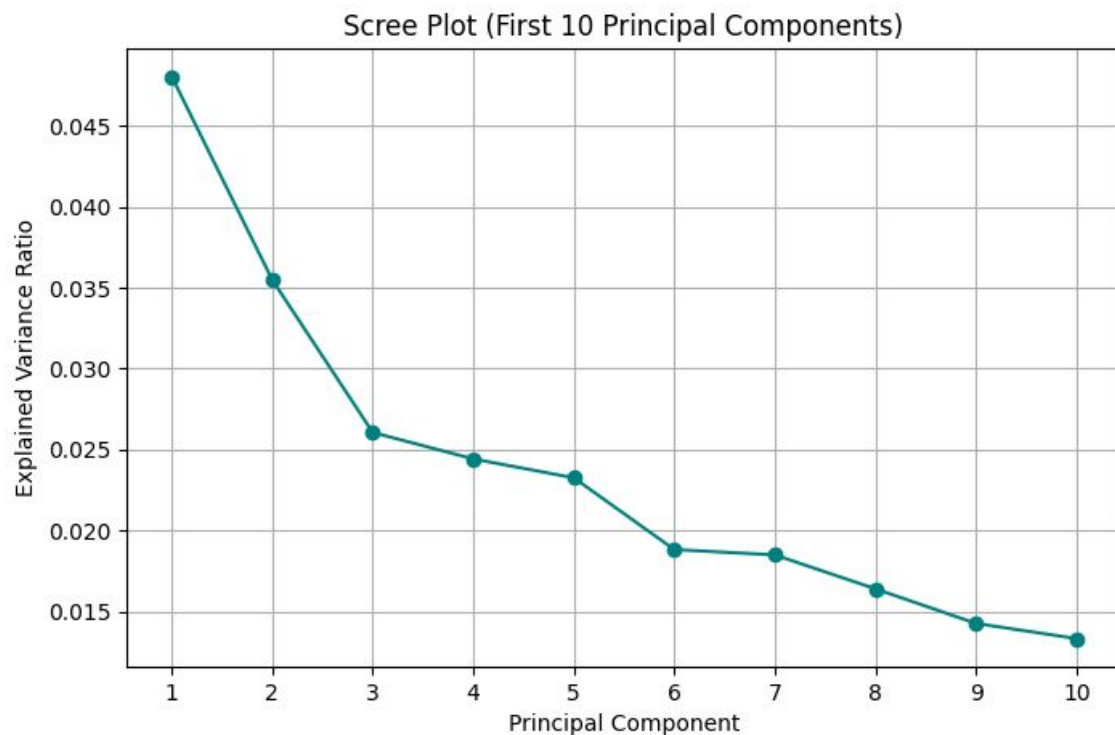
Principal Component Analysis

Decision Tree

Chi-Square Test



PCA Results - SWAN2 Only



PCA 1: Primary Care Provider (visit frequency, OB/GYN visit frequency)

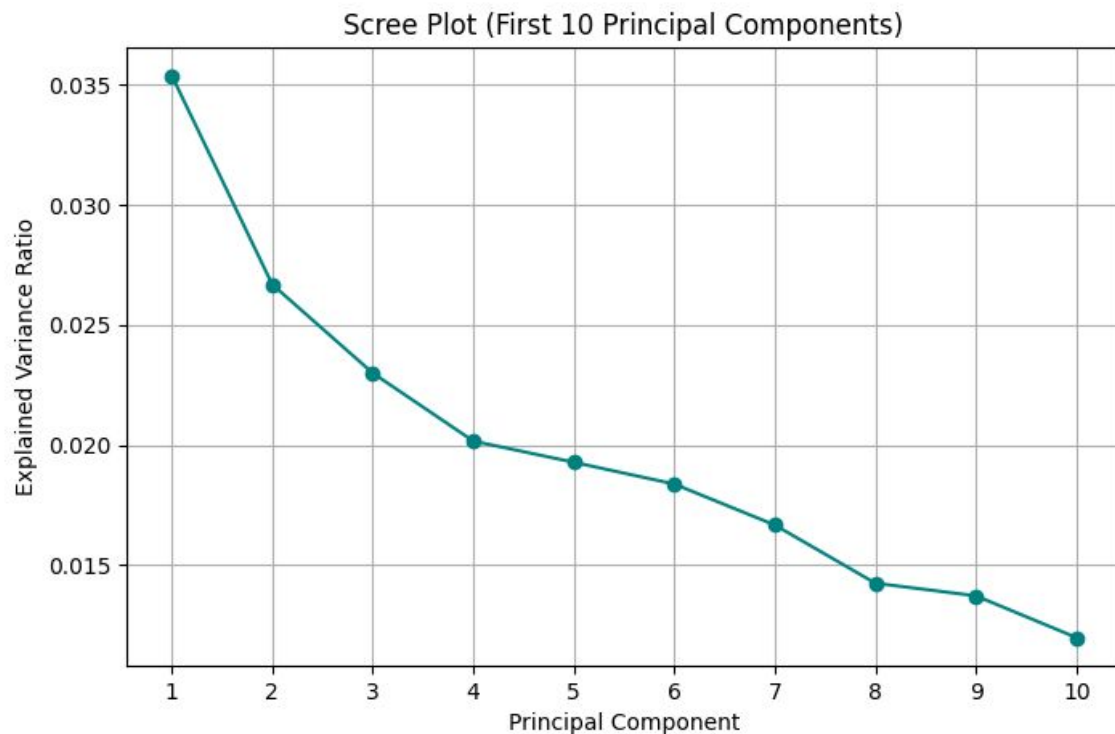
PCA 2: Why participants take hormones

PCA 3: General Health (thyroid, osteoporosis, fertility medication, heart attack, pregnant)

PCA 4: Over the Counter Medication

PCA 5: Over the Counter Medication

PCA Results - SWAN1 and SWAN2



PCA 1: Primary Care Provider (visit frequency, OB/GYN visit frequency)

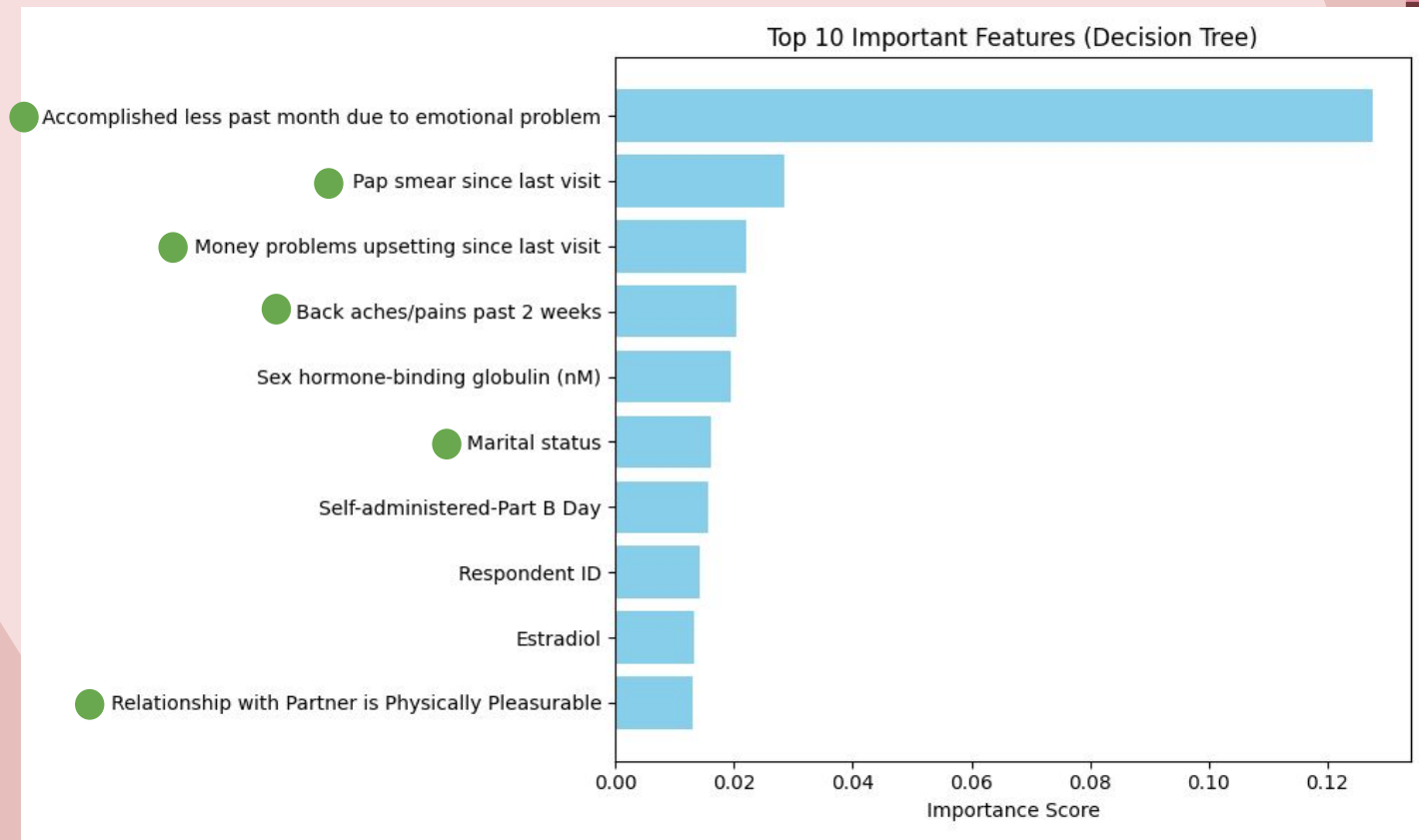
PCA 2: Why participants take hormones

PCA 3: Sisters' Health

PCA 4: Sisters' Health

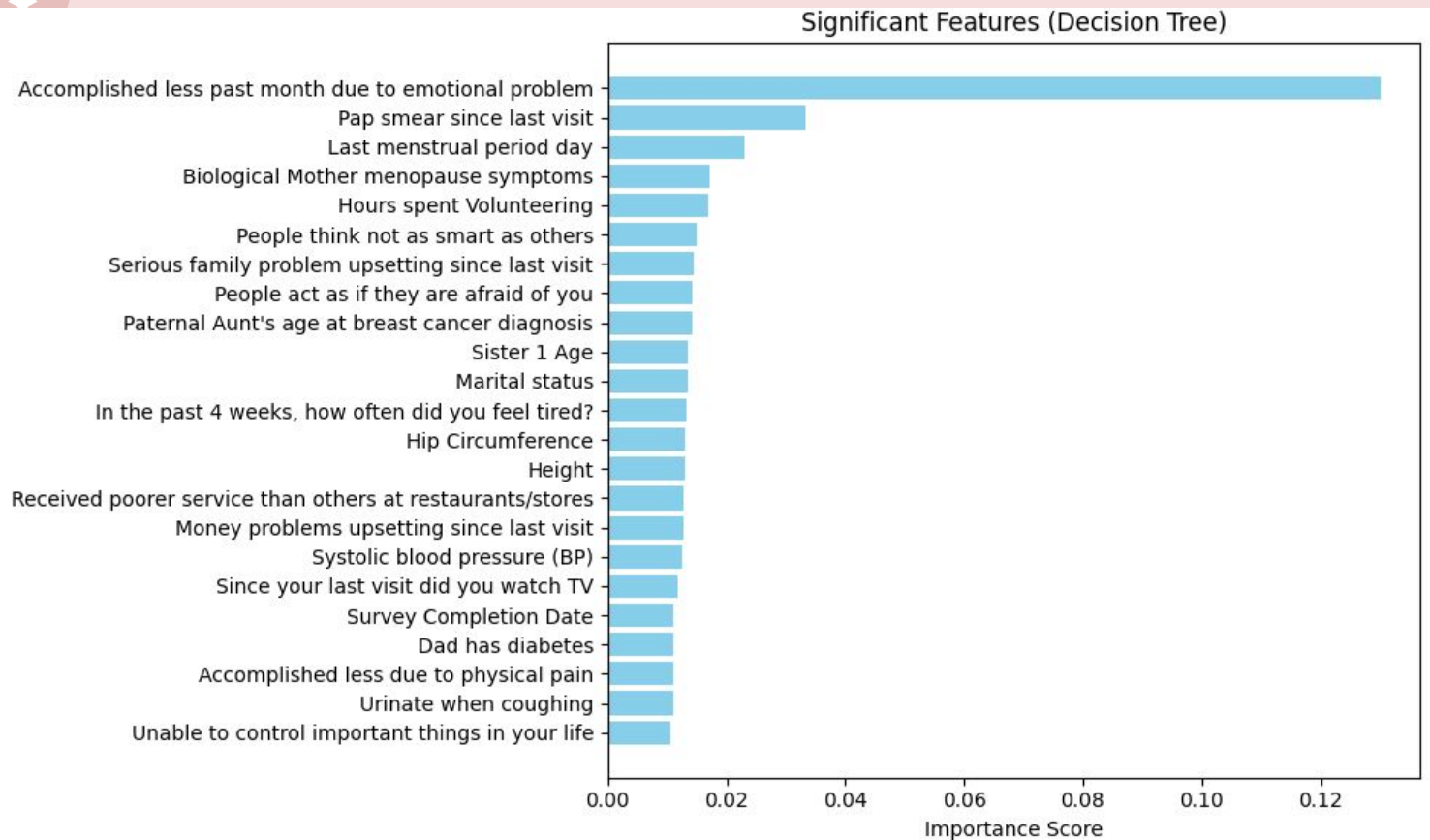
PCA 5: Sisters' Health

Decision Tree Results - SWAN2 Only



Some non-relevant features were correlated to the response variable by the Decision Tree because they vary across features (unique outcome for each participant). This could also be due to overfitting. Green dots are features that intuitively could correspond to response variable.

Decision Tree Results - SWAN1 and SWAN2



Features removed:

- SWANID -
- Participant ID
- BMI2 - BMI
- DHAS2 -
- Dehydroepiandrosterone sulfate
- E2AVE2 - Estradiol
- SPBMDT2 - Total Spine BMD w/cross-calibration applied
- HPBMDT2 - Total Hip BMD w/cross-calibration applied

Problem: Features with high cardinality were over emphasized in the tree.

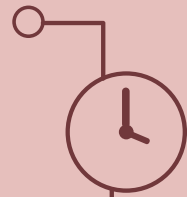
Solution: We removed features where the proportion of unique values was $\geq 50\%$ of the number of entries.

Chi-Square Test - SWAN1 and SWAN2

405 Features were significant ($p < 0.01$)

Decision Tree Top 10 Features

- Survey Completion Date
 - Smoke cigarettes more than once a day.
 - Number of times you talked to a doctor to since last visit.
 - Breast exam since last visit
- Pap smear since last visit
- Physical health/emotional problem interfered w/normal social activity.
- Currently married/committed relationship
- Cut down on activity/work past month due to emotional problem
- Cut down activity/work due to physical health
- Total family income
- Mammogram since last visit



Green Dots - features that were also identified by Decision Tree. Blue Dots - features that are very similar to features identified by Decision Tree.

Chi-Square Test - SWAN1 and SWAN2

405 Features were significant ($p < 0.01$). PCA on these 405 Features.

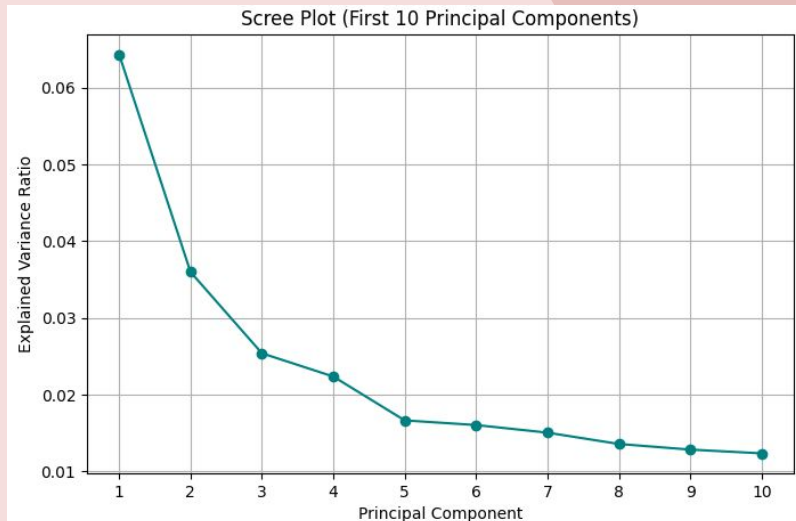
PCA 1: Health Problems (Heart attack, Angina, osteoporosis, thyroid gland removal, fertility medicine)

PCA 2: Talk to doctor (psychologist, nurse practitioner, physician assistant, other) about emotional problems

PCA 3: Feeling sad/depressed/irritable/mood swings

PCA 4: Felt you were treated with less respect than others due to sexual orientation, income level, physical appearance, age, ethnicity.

PCA 5: Interview Language





04

Conclusion

Model Results: Features detected by Decision Tree and Chi-square Test

Survey Completion Date

Pap smear since last visit

Currently married/committed relationship

Physical health/emotional problem interfered w/normal social activity


Cut down on activity/work past month due to emotional problem

Cut down on activity/work past month due to physical health

Analysis

Healthcare providers can screen for these factors regularly and reach out to patients who show more than one of these features (ie. not in a committed relationship, no pap smear since last visit, and physical/emotional health interfering with normal activity) and schedule check ins with them to address health needs.

What We Learned

- How to structure a repo to make our project reproducible
 - Techniques of data cleaning/preparation
 - Picking which model to use and appropriately analyzing the results
- 



05

Limitations/Future Work



Future Work: Use other datasets that focus on other aspects of health to make a more comprehensive analysis. Use newer datasets.

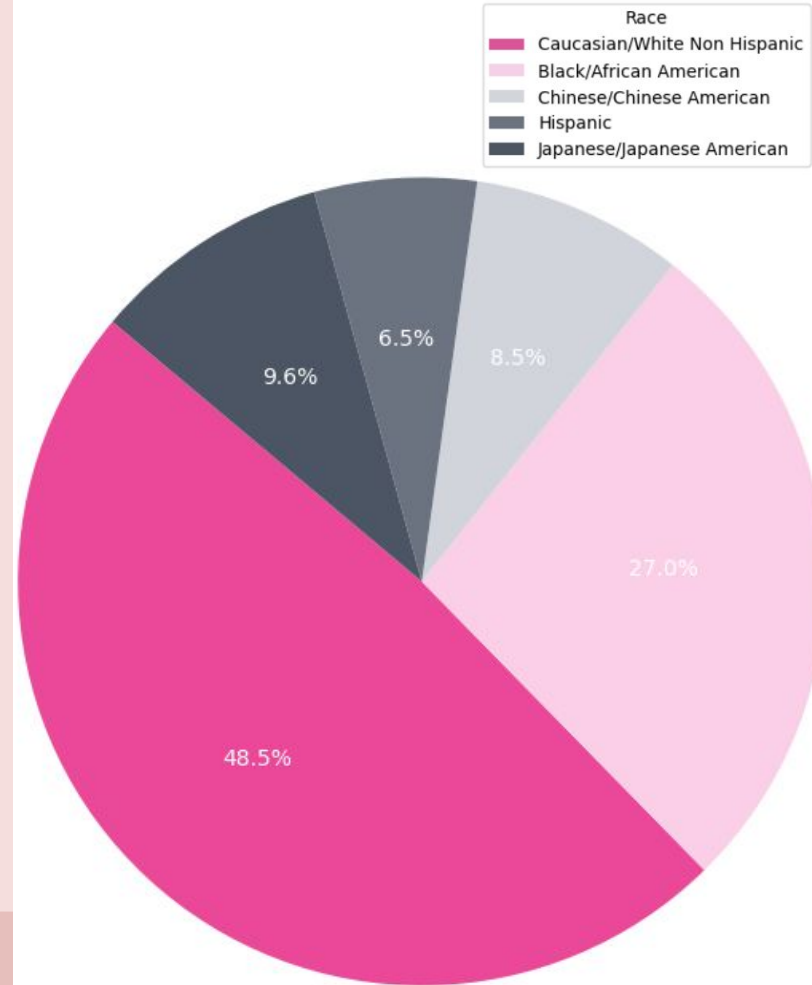
Future Work: Study why the correlated features lead to women not receiving the health care they need.

Value: This is a social science analysis. It doesn't attract investors. It is useful to health institutions to make their care more equitable.

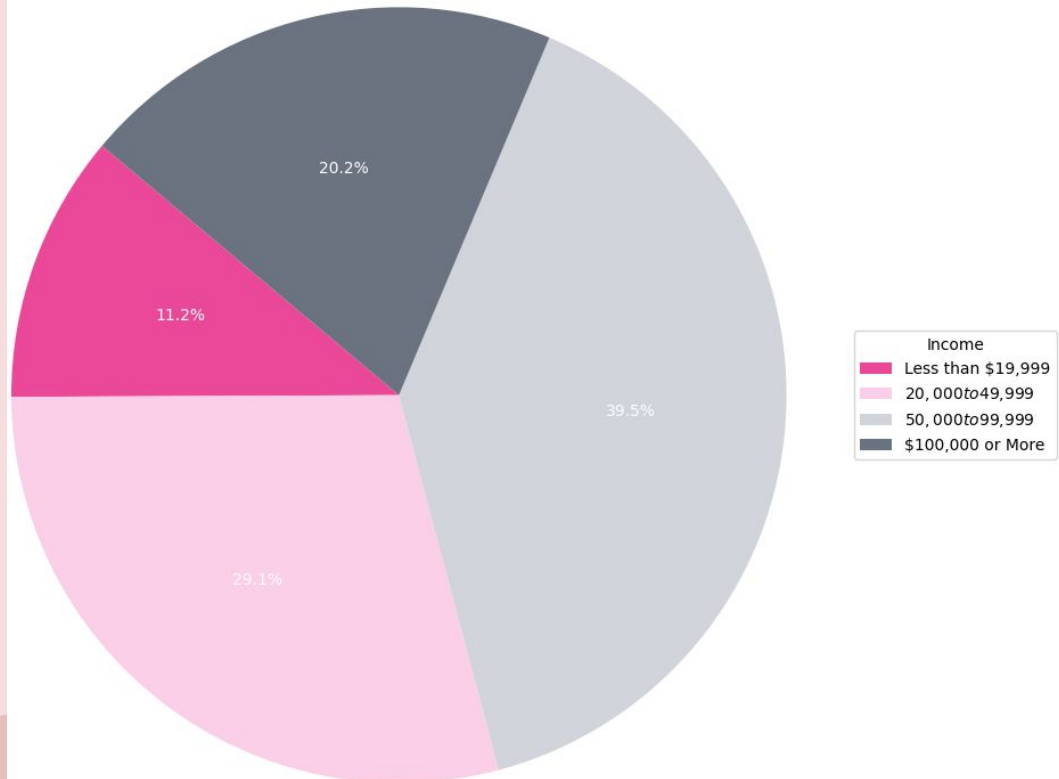
Limitations: Study sample may differ from real population in significant ways. Race and Income are plotted but other differences may exist



Race Distribution



Income Level Distribution





Thanks!

Questions?

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**

