December 16, 2022 CST 4704

BREAST CANCER DATA ANALYSIS

Business Intelligence + Data Warehousing Project

Presented By

Natacha Chetty

CONTENT OF THE DATASET

The dataset includes information from **6,788,436 mammograms** in the BCSC between **January 2005** and **December 2017**.

The dataset includes participant characteristics previously shown to be associated with breast cancer risk.

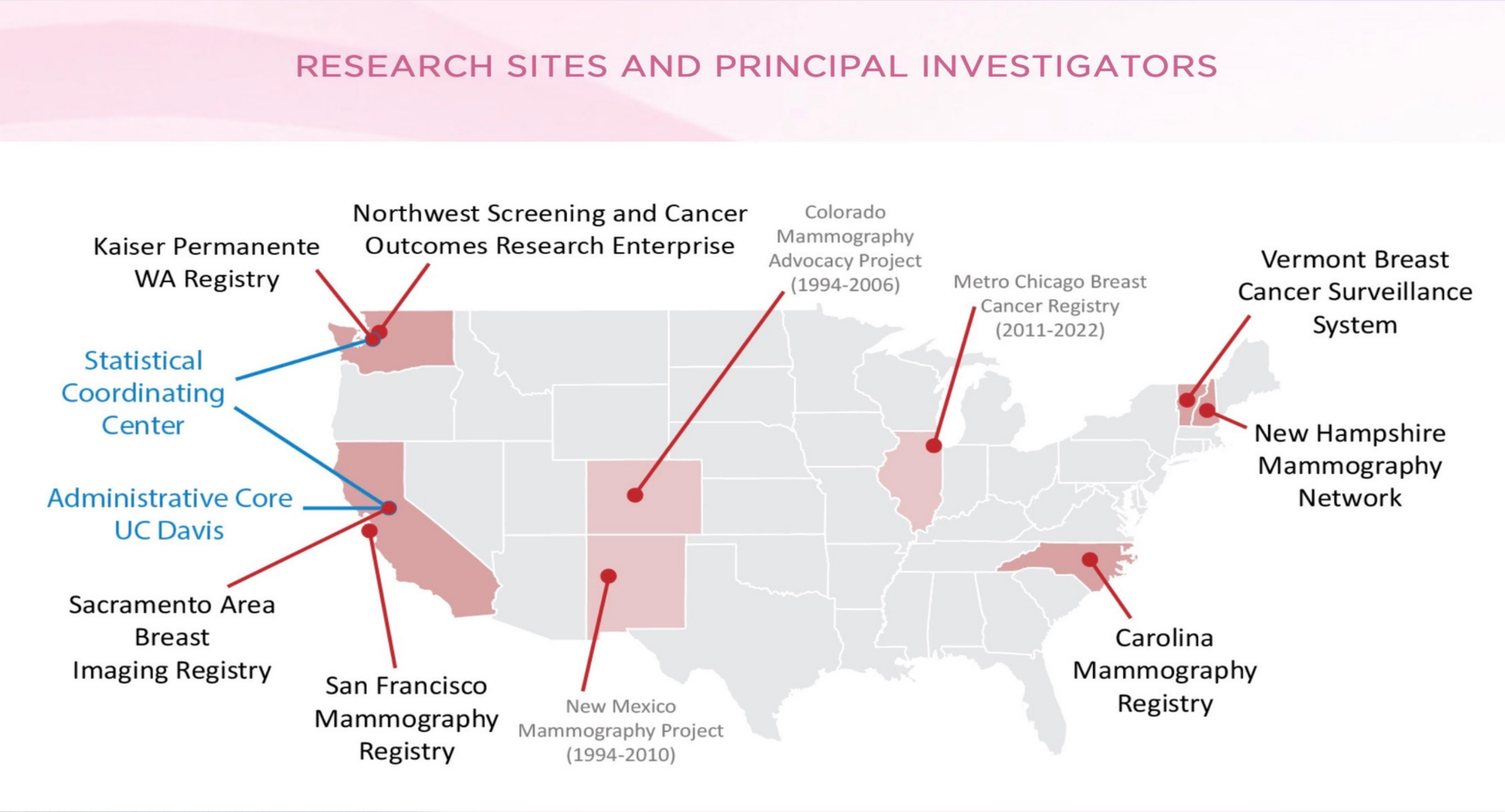
CONTENT OF THE DATASET

Preview of Raw Dataset

A	B	C	D	E	F	G	H	I	J	K	L	M
year	age_group_5_years	race_eth	first_degree_hx	age_menarche	age_first_birth	BIRADS_breast_density	current_hrt	menopaus	bmi_group	biophx	breast_cance	count
2013	7	1	0	9	3	1	1	2	3	0	0	7
2013	7	1	0	9	3	1	1	2	3	1	0	3
2013	7	1	0	9	3	1	1	2	4	0	0	6
2013	7	1	0	9	3	1	1	2	4	1	0	1
2013	7	1	0	9	3	1	1	2	4	1	1	1
2013	7	1	0	9	3	1	1	2	9	0	0	1
2013	7	1	0	9	3	1	9	1	9	0	9	1
2013	7	1	0	9	3	1	9	2	1	0	0	1
2013	7	1	0	9	3	1	9	2	2	0	0	2
2013	7	1	0	9	3	1	9	2	2	9	0	1
2013	7	1	0	9	3	1	9	2	3	0	0	5
2013	7	1	0	9	3	1	9	2	4	0	0	1
2013	7	1	0	9	3	1	9	2	4	1	1	1
2013	7	1	0	9	3	1	9	2	9	0	0	1

SOURCE

Data are collected from these different BCSC Registries



THE RISK FACTORS

age_group Age within an age group, defined in a 5 year range (e.g. 30-34, 50-54, 75-79...)

first_degree_hx History of breast cancer in a first-degree relative. First-degree relatives include parent, sibling, or child

age_menarche Age of the first occurrence of menstruation (before age 12, between age 12-13, or after age 14)

current_hrt Currently using Hormone Replacement Therapy; a form of hormone therapy used to treat common menopausal symptoms

age_first_birth Age of first pregnancy within an age range (e.g. before age 20, between age 25-29, Nulliparous/never given birth)

race/ethnicity Self-identified racial category (e.g. Hispanic, Asian/Pacific Islander, Native American)

THE RISK FACTORS

CONT..

BIRAIDS_breast_density BI-RADS breast density category, which is a visually estimated description of the volume of dense breast tissue on the mammogram.
(e.g. Almost entirely fat, Heterogeneously dense, Extremely dense..)

menopause Menopause is defined as the permanent cessation of ovulation, marked by the end of menstruationThere are three stages of menopause: perimenopause, menopause and postmenopause.

bmi_group Body mass index, a measurement of body size. Uses standard weight status categories, a person's weight and height, that can help doctors to track weight status and identify potential issues in individuals.

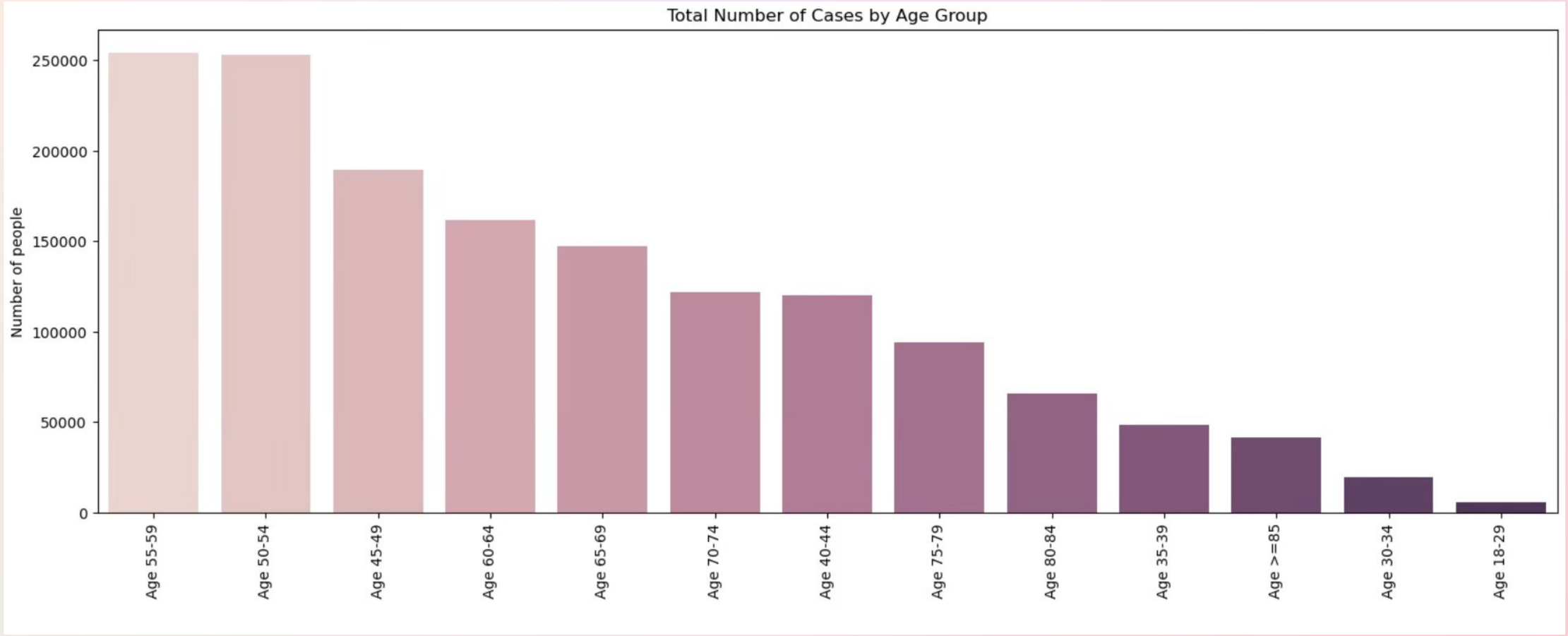
biophx Previous breast biopsy or aspiration. A breast biopsy is a procedure to remove a sample of breast tissue for testing to see if they contain cancer cells.

breast_cancer_history Prior breast cancer diagnosis

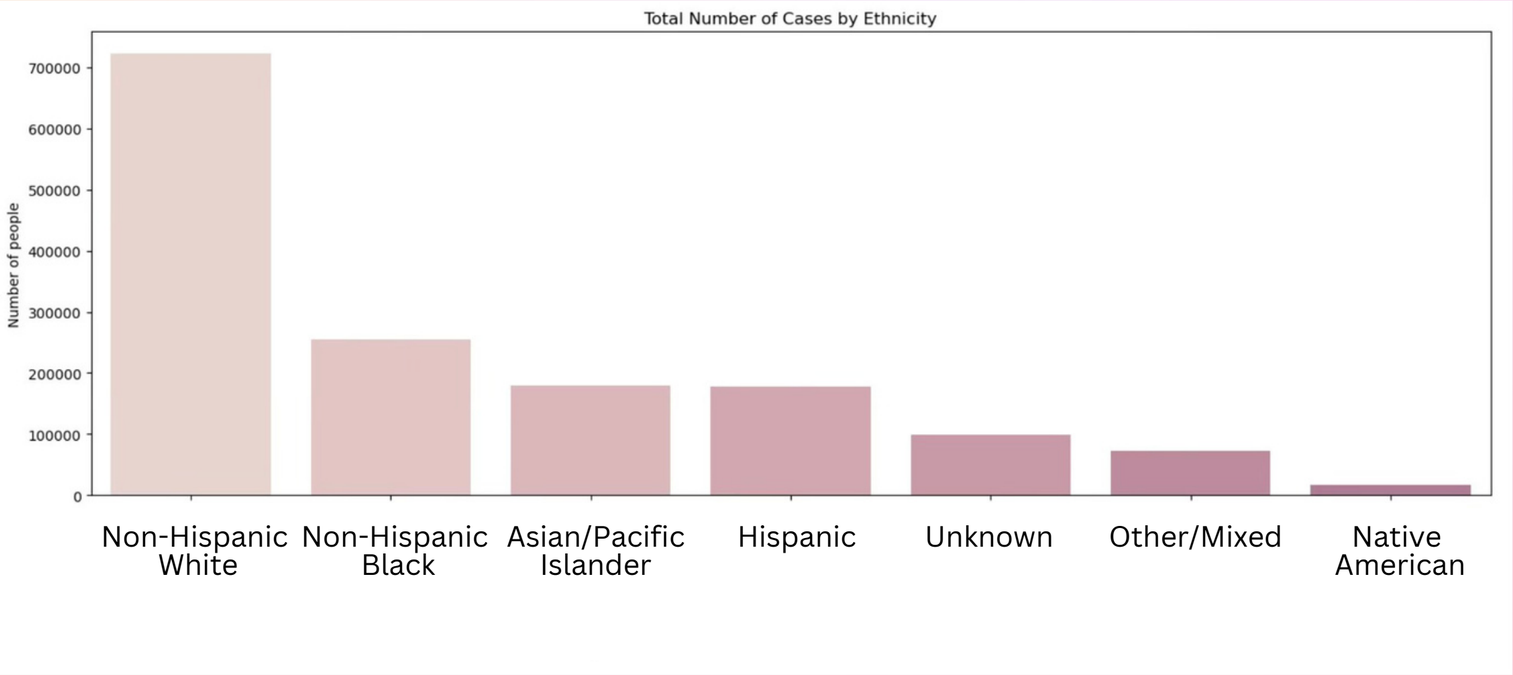
QUESTIONS

- What is the Total Number of Breast cases by **Ethnicity**?
- Which Ethnicity has the most breast **cancer history**?
- What is the total Number of cases by **Age Menarche**?
- Which **Age Group** has the most breast cancer cases?

WHICH AGE GROUP HAS THE MOST BREAST CANCER CASES?

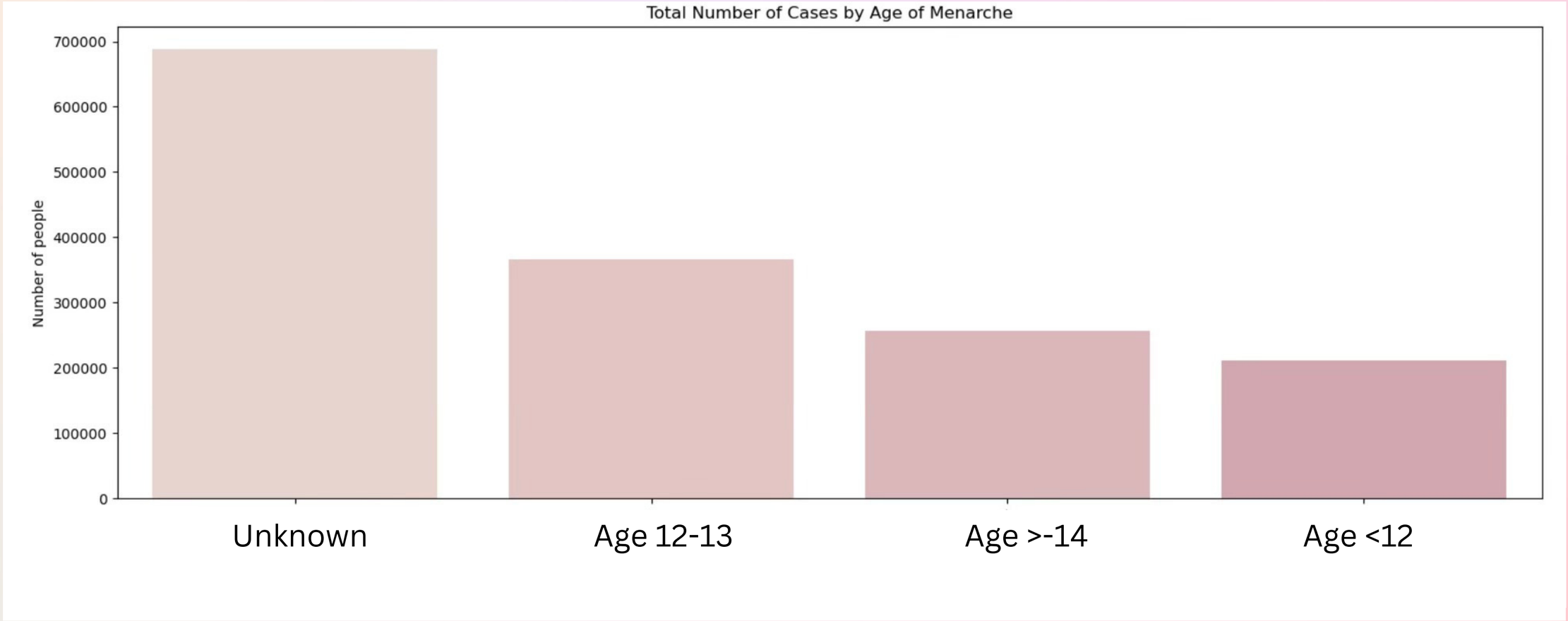


WHICH ETHNICITY HAS THE MOST BREAST CANCER CASES?

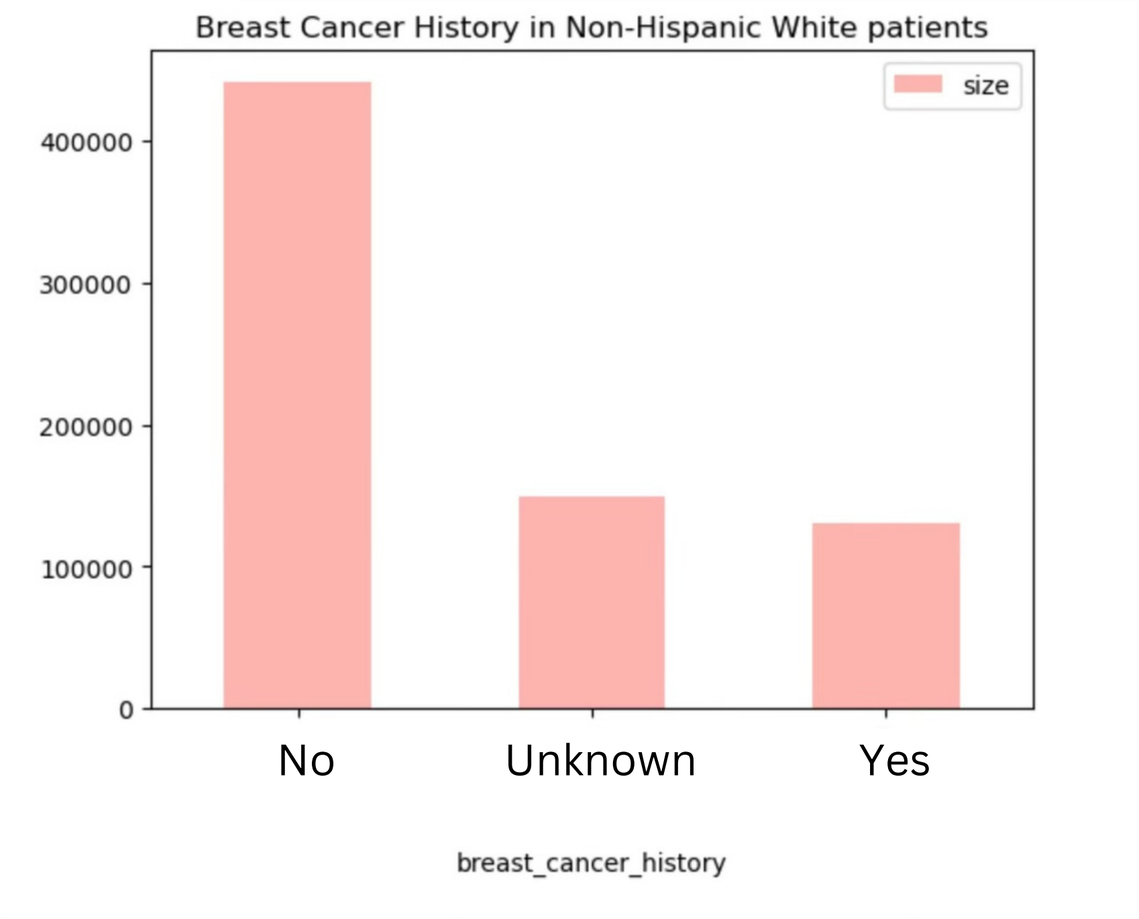


Non-Hispanic White	721859
Non-Hispanic Black	255476
Asian/Pacific Islander	179926
Hispanic	177069
Unknown	98329

AGE OF MENARCHE AND NUMBER OF CASES



ETHNICITY WITH THE HIGHEST RATE OF PREVIOUS BREAST CANCER DIAGNOSIS



All Ethnicity, History of cancer Diagnosis

Non-Hispanic White	721859
Non-Hispanic Black	255476
Asian/Pacific Islander	179926
Hispanic	177069
Unknown	98329

Non-Hispanic White	**No**	441436
	Unknown	149357
	Yes	131066

FUTURE ANALYSIS

- What are the risk factors that were the most common?
- What are the risk factors that are most common by ethnicity?
- Which state has the population with the higher breast cancer rates?