Technical Report

Alzheimer's Disease & Healthy Aging

(United States, 2015-2022)

Khadija Abdulla Al-Hay

DAB 16

Table of Content

Introduction

Alzheimer's Disease (AD) is a progressive neurological disorder that is the most common cause of Dementia (Kumar et al., 2025). The cognitive decline in Alzheimer's Dementia results from damage to the nerves, and its progression leads to a gradual decline in the quality of life of its patients; as it affects functions such as Memory, Comprehension, Language, Attention, Reasoning, and Judgment. Although it is incurable, early detection and treatment could manage its symptoms, which delays its progression.

Thus, an analysis is required on Alzheimer's Disease and Healthy Aging of **Older Adults** (50+ years Old) in the **United States**, to identify the **highest risk demographics and region** and to highlight some of the **top Alzheimer's risk indicators** in that specific region.

Problem Statement

Between **2015** and **2022**, the number of Alzheimer's patients in the **United States** have been increasing and reached around **6.07M Patients**. Other than aging, multiple factors could contribute to this increase; such as Mental Health, Cognitive Decline, Obesity, etc. This has resulted in a growing **shortage of care workers**, and incurring **high economic costs**.

Approaches

Across the United States:

1. Over the years, are Older Adults across **all the US** aging healthily?
2. Which **Age Group/Gender/Ethnicity/Race** are more at risk of developing Alzheimer's?
3. Which **region** has a **higher risk** of having more patients in the future?

In the Highest Risk Region:

4. What are the **top risk indicators**?
5. How are older adults doing in terms of **Cognitive Health**?
6. Over the years, how many people have talked to a Health Care professional about **Subjective Cognitive Decline** or Memory Loss?
7. How are older adults doing in terms of **Mental Health**?
8. Over the years, are more people experiencing **Frequent Mental Distress**?

9. Do older adults have **healthy eating** and **physical activity** habits?
10. Are people more **obese** in recent years?

Target Audience

In the US, care facilities are a mix of Public, Private for-profit, and Private non-profit. This analysis focuses on Public Care Facilities in the United States. It is usually operated by the Federal Government for Veterans (VA), the Department of Defense (DA), and state/local government hospitals.

At the **national level:** the **VHA Office of Geriatrics and Extended Care** provides guidance and policy for long-term care programs, which include dementia care.

At the **individual level**: the **Associate Chief of Staff for the Geriatrics Department** at Health Care Facilities provides leadership for long-term dementia patients and works with other departments to achieve facility goals.

Datasets

General Information

Rows: 284K, Columns: 31
Data Source: Kaggle
Original Data Source: BRFSS, CDC

Description

The dataset originates from the U.S. Centers for Disease Control and Prevention's (CDC) BRFSS system, which stands for the **Behavioral Risk Factor Surveillance System.** It is the nation's premier system of health-related telephone surveys that collect state data about US residents regarding their health-related risk behaviors, chronic health conditions, health-care access, and use of preventive services (Centers for Disease Control and Prevention, 2024).

The purpose of the data is to **explore the Alzheimer's Disease and Healthy Aging Data** to gain access to state and national level data on a number of indicators regarding health and well-being for older U.S. adults (Centers for Disease Control and Prevention, 2024). Furthermore, the data covers the United States, specifically the **50 US States**,

Washington D.C. (a.k.a District of Columbia), and **3 US Territories**. Refer to Appendix A for the full list. Also, it includes the aggregated data for the 4 U.S. regions, which are the **West**, **Midwest**, **Northeast**, and **South**. As well as, the data across **all US locations** included in the BRFSS survey for the period **2015-2022** (CDC Division of Population Health, 2016). Refer to Appendix B for the list of locations and their classifications as a US region according to the CDC (Centers for Disease Control and Prevention, 2024).

Data Dictionary

Column Name	Description	Data Type
RowId	Dataset row identifier	Text
YearStart	Year Start	Integer
YearEnd	Year End	Integer
LocationAbbr	Location Abbreviation	Text
LocationDesc	Location Description	Text
Datasource	Data Source	Text
Class	Class description (Also referred to as Indicator Type)	Text
Topic	Topic description (Also referred to as Indicator)	Text
Question	Question	Text
Data_Value_Unit	The unit, such as "%" for percentage	Text
DataValueTypeID	Identifier for the Data Value Type	Text
Data_Value_Type	The data value type, such as age-adjusted prevalence or crude prevalence	Text
Data_Value_Interpolated	Data Value, such as 14.7 after dealing with the nulls using time interpolating per location	Float
Data_Value	Data Value, such as 14.7	Float

Data_Value_Alt	Equal to data value, but format is numeric	Float
Data_Value_Footnote_Symbol	Footnote Symbol	Text
Data_Value_Footnote	Footnote Text	Text
Low_Confidence_Limit	Low Confidence Limit	Float
High_Confidence_Limit	High Confidence Limit	Float
StratificationCategory1	Lookup Identification value, such as Age Group, Sex, or Race/Ethnicity	Text
Stratification1	Data stratified by this value, such as 50-54, 55-59, 60-64, 65 years or older	Text
StratificationCategory2	Lookup Identification value, such as Sex or Race/Ethnicity	Text
Stratification2	Data stratified by this value, such as Male, Female, Asian/Pacific Islander, Black, non-Hispanc, Hispanic, Native Am/Alaskan Native, or White, non-Hispanic	Text
Geolocation	Geolocation data for mapping purposes	Tuple
ClassID	Identifier for Class	Text
TopicID	Identifier for topic	Text
QuestionID	Identifier for question or indicator	Text
LocationID	Location number value corresponding to geographic location like state	Integer
StratificationCategoryID1	Identifier for the first category stratification	Text
StratificationID1	Identifier for the first stratification	Text

StratificationCategoryID2	Identifier for the second category stratification	Text
StratificationID2	Identifier for the second stratification	Text
Population	The location's Population according to the YearStart	Integer
Type	Location Type such as, State, Region, Territory, All (sum of Total Regions)	Text
Data_Value_Interpolated_over100	Percentage of older adults divided by 100	Float(Percentage)
Est.Count_Per_Location	The estimated number of people for each indicator, based on the population of each location in the specific year	Integer
Per Capita Ratio	Calculation of the percentage of older adults based on the location/population	Float
Per 100,000 People	The number of cases per 100,000 people	Integer

Data Handling

Kindly Refer to the following Jupyter Notebook files for the Data Handling steps:

1. **Alzheimers_Data_cleaning.ipynb -** EDA & General Cleaning
2. **Alzheimers_Data_Cleaning_Cont.ipynb -** Calculations on Interpolated Values
3. **Alzheimers_Data_Cleaning_Intr_0.2.ipynb -** Data Interpolation

Also, refer to the **Alzheimers_Healthy_Aging_Dataset_withCalc.xlsx** Excel file for the final clean version*.

*More information on the Data Cleaning will be available below in **Appendix C**.

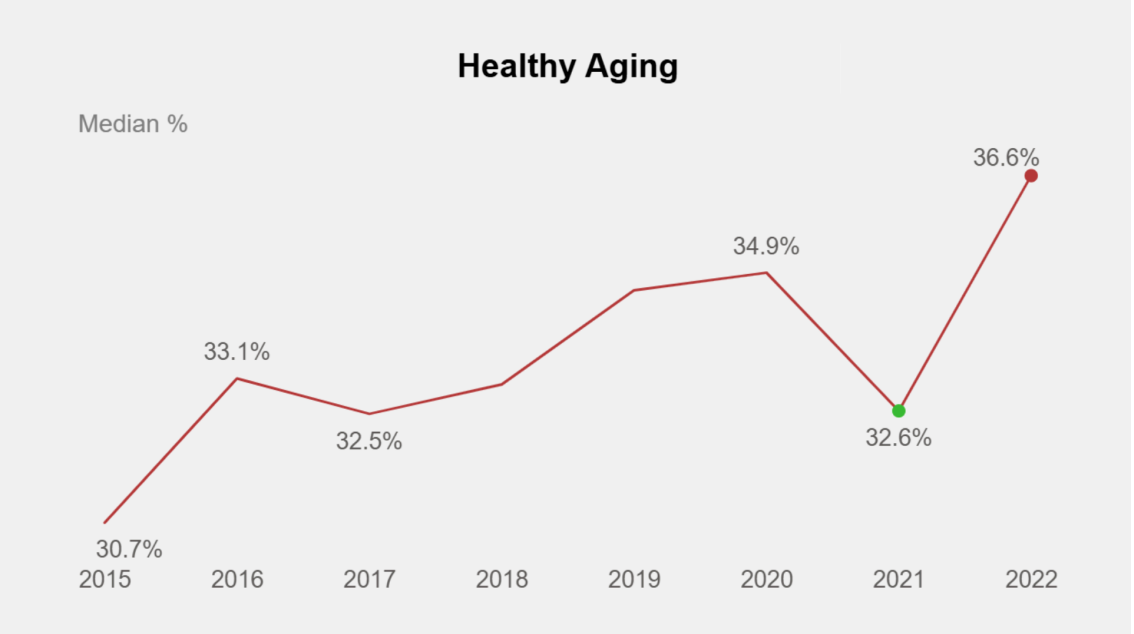
Analysis & Key Findings

Key Findings

1. An **increased risk** of having more Alzheimer's Patients due to a steady **decrease** in the overall Health of older adults in the US.
2. Across the United States and the South region, **65+ year olds**, **Females**, and **Native Americans/Alaskans, Black(Non-Hispanic) & Hispanic** people are more likely to develop Alzheimer's.
3. The **South** Region has the **highest** Alzheimer's risk, compared to the other regions.
4. Between 2015 and 2022, more older adults in the **South** Region are reporting and experiencing **Subjective Cognitive Decline, increasing** their likelihood of developing the disease.
5. A general **decline** in the **Mental Wellbeing** of Citizens of the **South** Region as people are experiencing more **frequent stress**, with some even having a **Lifetime Diagnosis of Depression**.
6. **Obesity** poses a **high risk** in terms of Nutrition in the **South** region as **36.9%** of older adults are considered obese by 2022.

Detailed Analysis

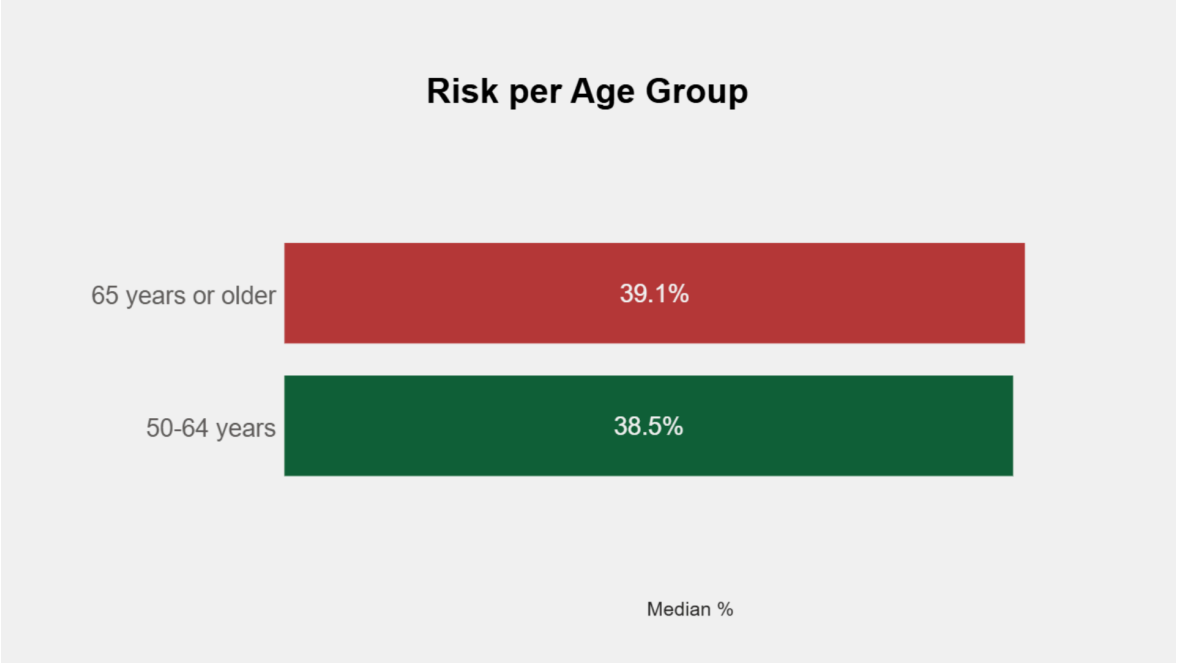
Q1. Over the years, are Older Adults across **all the United States** aging healthily?



The increase in the percentage over the years signals a **decline in the overall health of Older Adults** in the United States as it reached **%36.6** Older Adults by 2022. This indicates an **increased risk** of having more Alzheimer's Patients in the future.

On the other hand, there was a slight drop in the percentage around 2021, which **could** be due to the **COVID-19 pandemic**; as the sudden occurrence of the pandemic had exposed flaws in the United States Public Health data collection infrastructure. Thus, it forced the United States to switch from the Traditional Healthcare data collection ways into adopting digital health technologies such as **TeleHealth**.

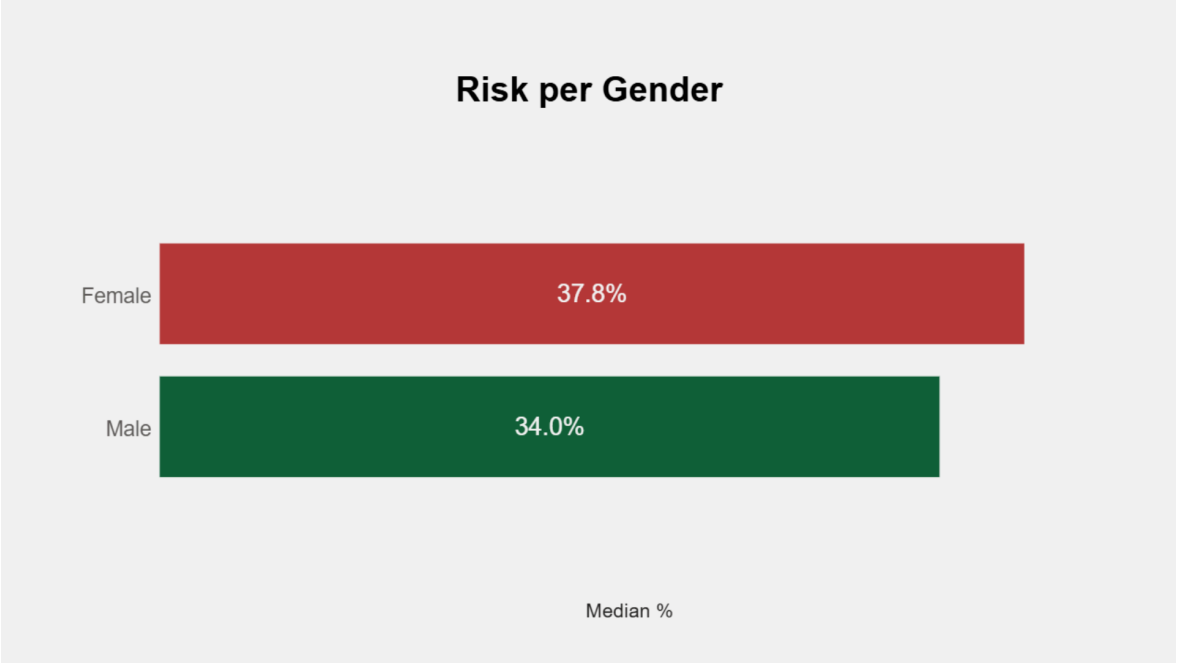
Q2.1. Which **age group** are more at risk of developing Alzheimer's?



Across the United States, **65+ year olds** are **more prone to having Alzheimer's** as they age, compared to other age groups, as they have the highest average percentage of Older Adults across all risk indicators.

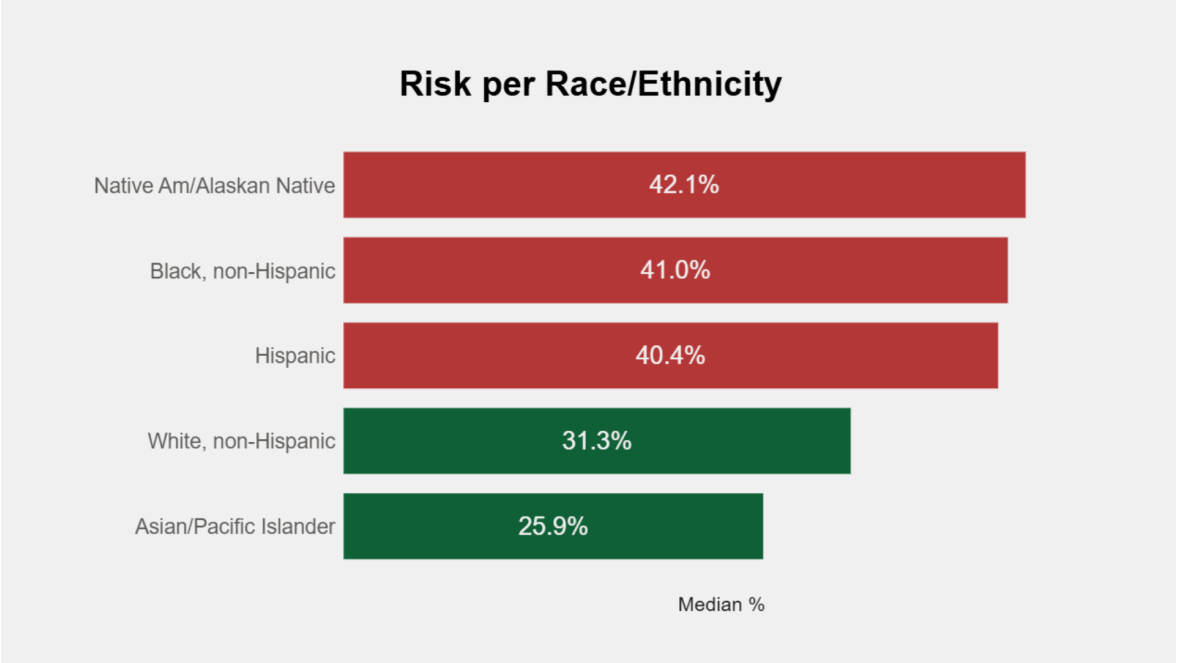
However, it is important to note that although age is a way to indicate a higher risk of the disease, this does not immediately mean that every older adult who reaches this age is going to have Alzheimer's. **Other factors must be present** as well to increase the likelihood of developing it.

Q2.2. Which **gender** is more at risk of developing Alzheimer's?



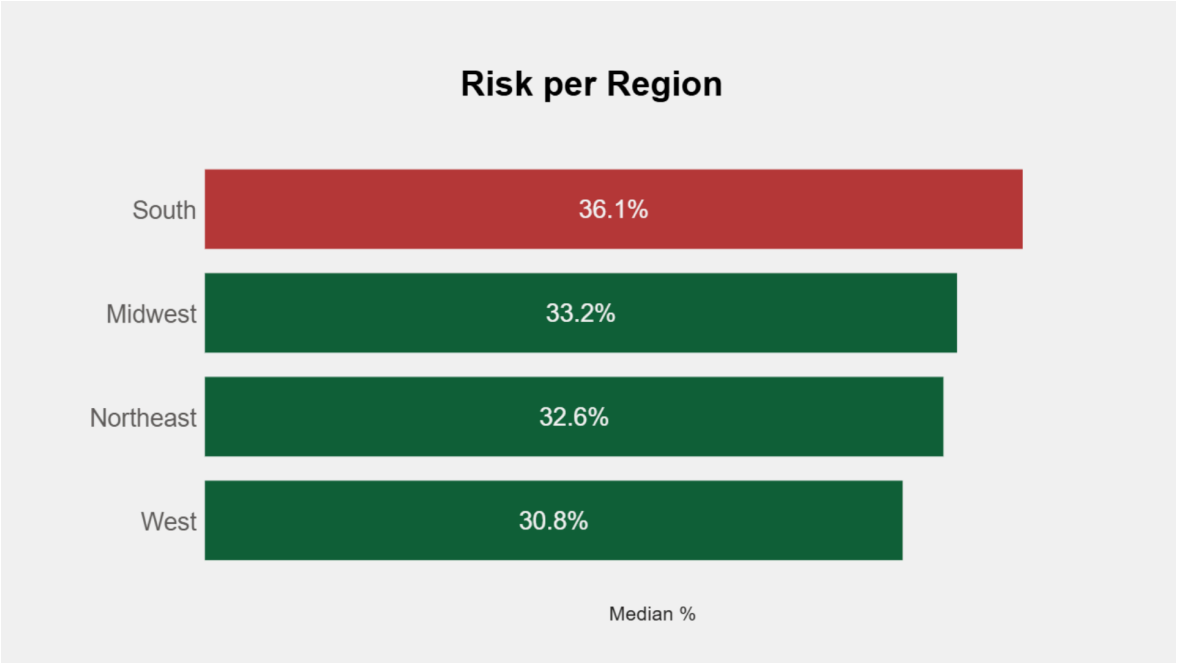
Between the two genders in the United States, **females** are **most at risk** of developing Alzheimer's disease. This could be due to many **biological** and **sociocultural** factors. For example, women tend to have a **longer lifespan** than men, and **hormonal changes** during menopause lead to a sharp decrease in estrogen levels that supports healthy brain functions.

Q2.3. Which **ethnicity/race** are more at risk of developing Alzheimer's?



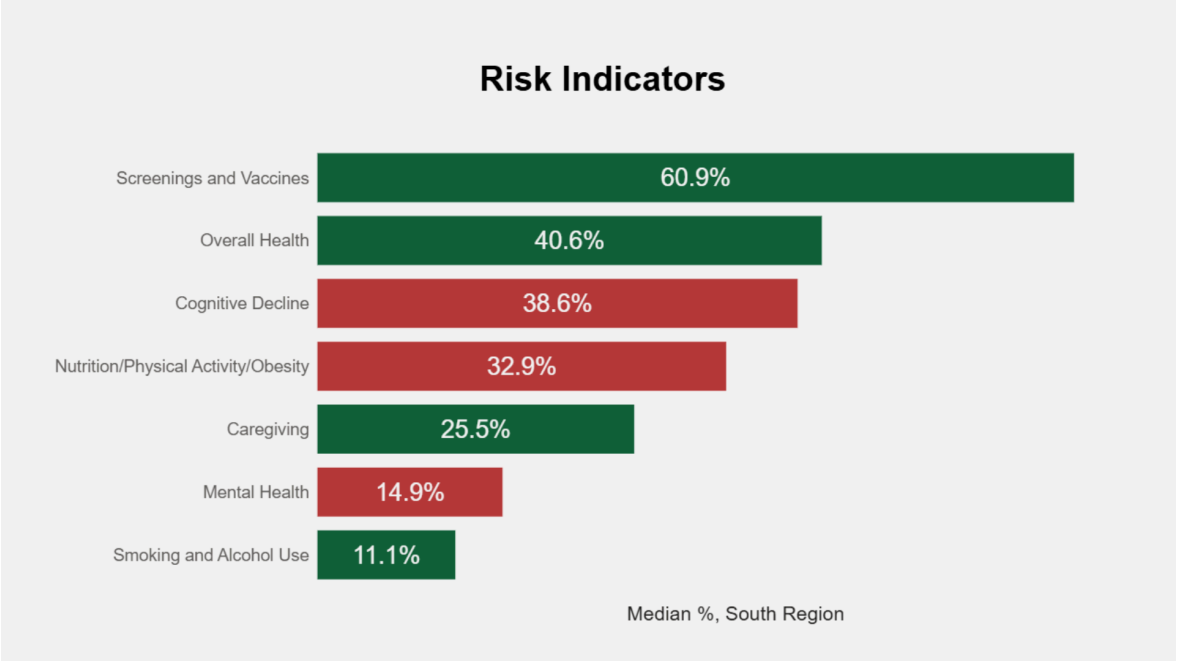
Native Americans/Alaskans, Black(Non Hispanic), and **Hispanic** people are the top three demographics that are more prone to having Alzheimer's compared to the others. This could be the result of these groups being more **genetically predisposed to health problems** like high blood pressure, obesity, and other cognitive related issues. Also, other social factors could contribute such as having **lower access to education and healthcare**.

Q3. Which **region** has a **higher risk** of having more patients in the future?



The **South** region has the **highest risk** of Alzheimer's disease, compared to the other regions across all risk indicators. A deeper view is provided below to look into the top factors in the South.

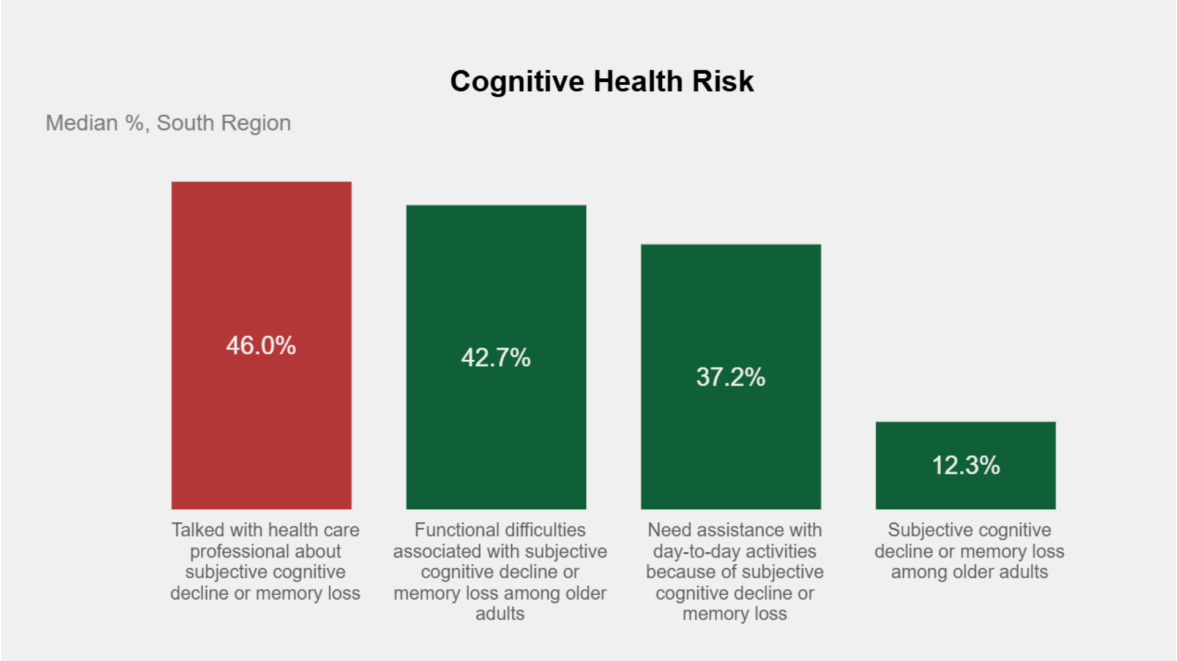
Q4. What are the **top risk indicators** in the South Region?



Amongst all risk indicators, in the South region, Screenings and Vaccines, and Overall health are the top. However, other factors affect the likelihood of having Alzheimer's more, which will be focused on in upcoming graphs such as **Cognitive Decline**, **Nutrition & obesity**, and **Mental Health**.

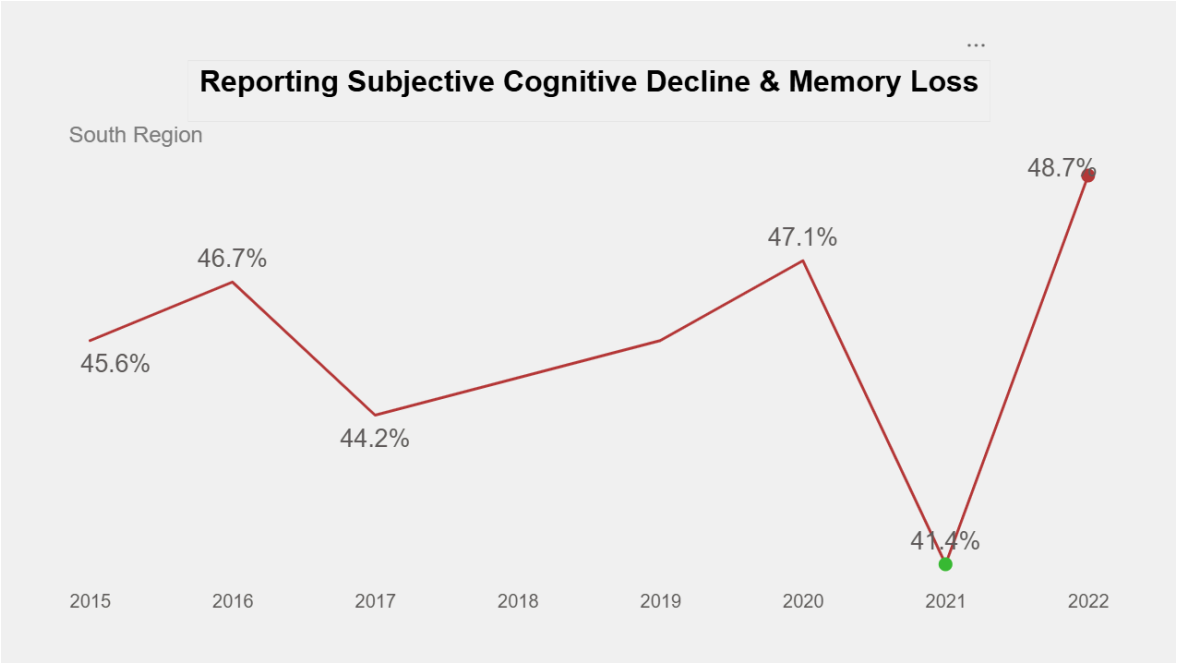
People in the South have **Low education attainment** which has an impact later in life on their cognitive health which is the primary symptom and defining characteristic of the disease itself. Also, high rates of **physical inactivity** and **unhealthy food** lead to obesity which in turn increase the chances of developing the disease.

Q5. How are older adults doing in terms of **Cognitive Health**?



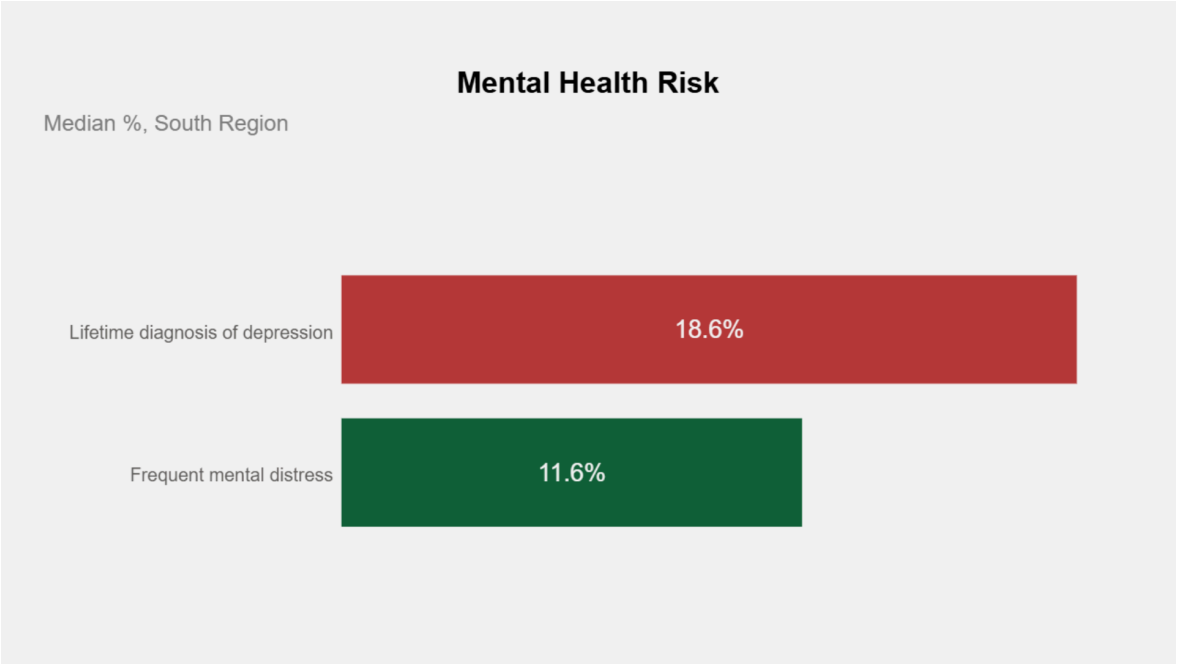
An alarming **46%** of older adults in the **South** have reported having Subjective Cognitive Decline(SCD) to a Health care professional. While **42.7%** are already having functional difficulties with SCD. Both are considered potential clinical symptoms to objective Cognitive impairment and eventual Alzheimer's dementia.

Q6. Over the years, how many people have talked to a Health Care professional about **Subjective Cognitive Decline** or Memory Loss?



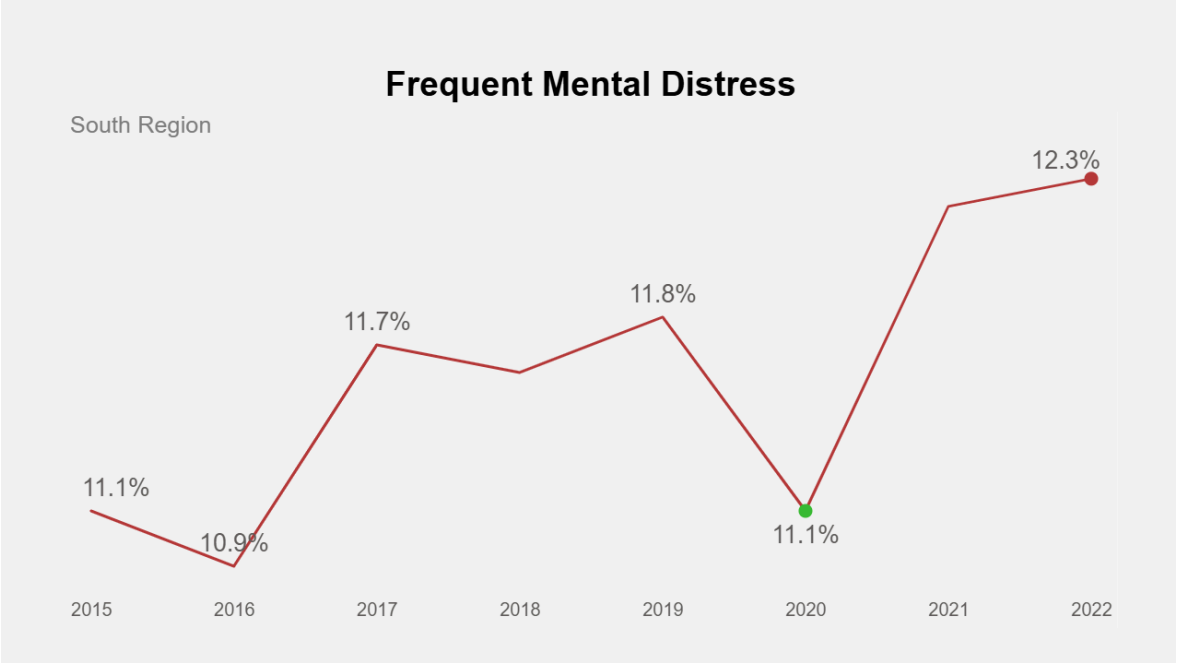
The number of people reporting Subjective cognitive decline has **increased** gradually over the years in the South region. Which indicates that the number of Alzheimer's patients in the region might increase as SCD is considered a risk factor that **predicts the onset or progression of the disease**. And similar to a previous graph, the drop in 2021 could be affected by the COVID-19 Pandemic as mentioned previously.

Q7. How are older adults doing in terms of **Mental Health**?



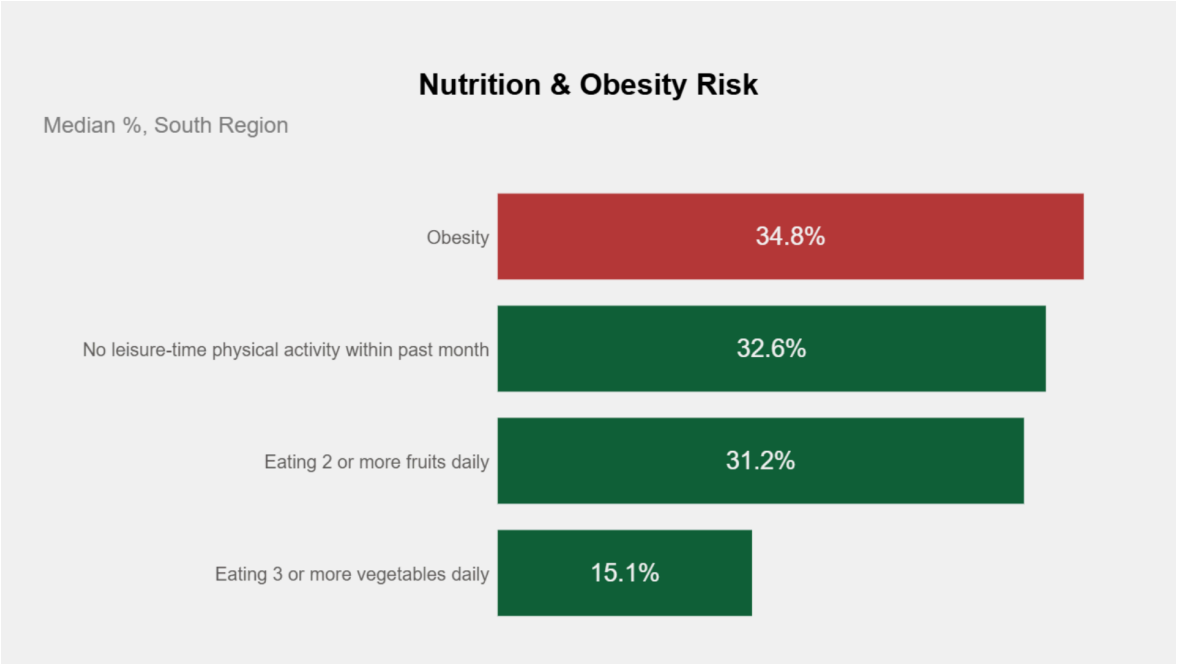
18.6% of older adults in the south region have been diagnosed with Lifetime Depression. Also, **11.6%** experience frequent mental distress which refers to having poor mental health for 14 or more days in the past 30 days. Both Depression and social isolation are identified as **top contributing risk factors for SCD,** which eventually leads to Alzheimer's Dementia.

Q8. Over the years, are more people experiencing **Frequent Mental Distress**?



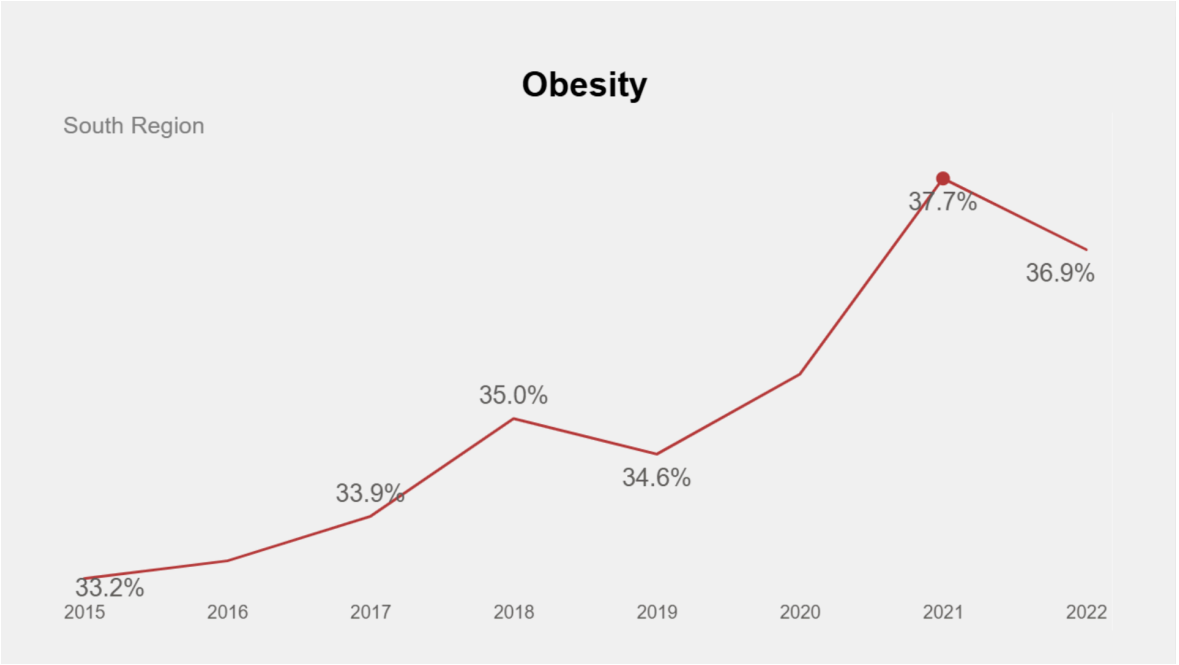
More people are experiencing poor mental health over the years in the **South** region. A **sharp increase** occurred from 2020 until 2022, which could be due to the COVID-19 pandemic. People experienced unfamiliar circumstances and were isolated which impacted their mental health severely.

Q9. Do older adults have **healthy eating** and **physical activity** habits?



In terms of Nutritions, obesity and low physical activity levels are the **top two risk factors** in the South. **32.6%** of older adults do not do **regular physical activity** which is a must to reduce the risk of having Alzheimer's disease. As well as, **obesity and bad eating habits** could lead to further physical health issues as well contributing greatly to the likelihood of having the disease as one ages.

Q10. Are people more **obese** in recent years?



The number of obese people in the **South** have increased over the years. With the peak being in 2021, which was around **37.7%** of older adults. Obesity is considered a high risk indicator of Alzheimer's as it is particularly linked to damage in areas of the brain associated with Alzheimer's and can lead to greater disease progression.

Recommendations

For the United States to:

1. Encourage **routine screening** for cognitive and physical health across the United States.
2. Create **training** opportunities to increase the number of qualified Alzheimer's Dementia specialists
3. Provide **incentives** to encourage qualified specialists to work in high-need communities.

For the South Region to:

1. Utilize **telehealth** to better reach **rural** and **underserved communities** for Physical, Mental and Cognitive Health.
4. Advocate for creating **Alzheimer's-friendly communities** with accessible transportation and structured living arrangements.
5. Improve Access to **Healthy Foods** to manage Obesity, and facilitate **walking groups**, **running clubs**, and other physical activity clubs for all ages to improve Physical and Mental Health.

Limitations

- **Missing Data**: 30% missing data in the Alzheimer's original dataset on the State Level, have prevented deeper analysis.
- **Time Limitations**: with more time, a different approach to interpolating the data would have been done attempting to consider the change in the demographic groups as well as the time.

Future Work

1. To use Machine Learning to interpolate missing data values aiming for a more accurate interpolation.
2. To do a State Level Analysis instead of a Regional Level Analysis to get deeper insights into the risk of Alzheimer's Disease in different US states.
3. To look into the risk of indicators on different Age Groups, Genders, and Races/Ethnicities.

References

CDC Division of Population Health. (2016, May 16). *Alzheimer's Disease and Healthy Aging Dataset*. Alzheimer's Disease and Healthy Aging Data. Retrieved Nov 23, 2025, from https://data.cdc.gov/Healthy-Aging/Alzheimer-s-Disease-and-Healthy-Aging-Data/hfr9-rurv/about_data

Centers for Disease Control and Prevention. (2024). *Behavioral Risk Factor Surveillance System*. About BRFSS. Retrieved Nov 23, 2025, from https://www.cdc.gov/brfss/about/index.htm

Centers for Disease Control and Prevention. (2024, Feb 8). *Alzheimer's Disease and Healthy Aging Data*. Alzheimer's Disease and Healthy Aging Data Portal. Retrieved Nov 23, 2025, from https://www.cdc.gov/healthy-aging-data/data-portal/index.html

Centers for Disease Control and Prevention. (2024, Jul 24). *National Center for Health Statistics*. Geographic division or region. Retrieved Nov 23, 2025, from https://www.cdc.gov/nchs/hus/sources-definitions/geographic-region.htm#:~:text=Northeast:%20Includes%20Connecticut%2C%20Maine%2C,Utah%2C%20Washington%2C%20and%20Wyoming.

Kumar, A., Jaskirat, S., Lui, F., & Tsao, J. W. (2025, Jan). *Alzheimer Disease* [Latest updates and evidence-based practices for diagnosing and managing Alzheimer's disease]. Alzheimer Disease - StatPearls - NCBI Bookshelf. Retrieved Nov 23, 2025, from https://nursing.jhu.edu/wp-content/uploads/2025/07/Alzheimer-Disease.pdf

Appendices

Appendix A

Population per location for the period of (2015-2022). Please note that the *United States, DC & Territories* in the below table, **includes all United States, DC and only 3 of its territories** that are included in the dataset and covered by the CDC's survey, as mentioned in the Dataset Description section above.

Geographical Area	2015	2016	2017	2018	2019	2020	2021	2022	Type
United States, DC & Territories	324,373,548	326,610,003	328,569,433	330,135,836	331,684,398	335,109,142	335,613,406	337,489,121	All
United States	**320,635,163**	**322,941,311**	**324,985,539**	**326,687,501**	**328,239,523**	**331,577,720**	**332,099,760**	**334,017,321**	All
All regions	**320,635,163**	**322,941,311**	**324,985,539**	**326,687,501**	**328,239,523**	**331,577,720**	**332,099,760**	**334,017,321**	Total Regions
Northeast	56,034,684	56,042,330	56,059,240	56,046,620	55,982,803	57,431,458	57,252,533	57,159,597	Region
Midwest	67,860,583	67,987,540	68,126,781	68,236,628	68,329,004	68,984,258	68,872,831	68,903,297	Region
South	120,997,341	122,351,760	123,542,189	124,569,433	125,580,448	126,476,549	127,368,010	129,037,849	Region
West	75,742,555	76,559,681	77,257,329	77,834,820	78,347,268	78,685,455	78,606,386	78,916,578	Region
Alabama	4,852,347	4,863,525	4,874,486	4,887,681	4,903,185	5,033,094	5,049,196	5,076,181	State
Alaska	737,498	741,456	739,700	735,139	731,545	733,017	734,420	734,442	State

Arizona	6,829,676	6,941,072	7,044,008	7,158,024	7,278,717	7,187,135	7,274,078	7,377,566	State
Arkansas	2,978,048	2,989,918	3,001,345	3,009,733	3,017,804	3,014,546	3,026,870	3,047,704	State
California	38,918,045	39,167,117	39,358,497	39,461,588	39,512,223	39,521,958	39,142,565	39,142,414	State
Colorado	5,450,623	5,539,215	5,611,885	5,691,287	5,758,736	5,787,129	5,814,036	5,850,935	State
Connecticut	3,587,122	3,578,141	3,573,297	3,571,520	3,565,287	3,579,918	3,606,607	3,617,925	State
Delaware	941,252	948,921	956,823	965,479	973,764	991,928	1,005,062	1,020,625	State
District of Columbia	675,400	685,815	694,906	701,547	705,749	670,917	669,256	676,725	DC
Florida	20,209,042	20,613,477	20,963,613	21,244,317	21,477,737	21,592,035	21,831,949	22,379,312	State
Georgia	10,178,447	10,301,890	10,410,330	10,511,131	10,617,423	10,732,888	10,792,060	10,931,805	State
Hawaii	1,422,052	1,427,559	1,424,393	1,420,593	1,415,872	1,451,252	1,447,029	1,440,359	State
Idaho	1,651,059	1,682,380	1,717,715	1,750,536	1,787,065	1,849,415	1,904,848	1,944,299	State
Illinois	12,858,913	12,820,527	12,778,828	12,723,071	12,671,821	12,799,088	12,700,641	12,621,821	State
Indiana	6,608,422	6,634,304	6,658,078	6,695,497	6,732,219	6,790,497	6,815,907	6,844,545	State
Iowa	3,120,960	3,131,371	3,141,550	3,148,618	3,155,070	3,191,141	3,198,613	3,202,820	State
Kansas	2,909,011	2,910,844	2,908,718	2,911,359	2,913,314	2,938,172	2,938,338	2,937,324	State

Kentucky	4,425,976	4,438,182	4,452,268	4,461,153	4,467,673	4,508,318	4,507,583	4,519,233	State
Louisiana	4,664,628	4,678,135	4,670,560	4,659,690	4,648,794	4,652,301	4,627,971	4,593,687	State
Maine	1,328,262	1,331,317	1,334,612	1,339,057	1,344,212	1,364,571	1,378,931	1,390,922	State
Maryland	5,985,562	6,003,323	6,023,868	6,035,802	6,045,680	6,177,935	6,179,403	6,192,440	State
Massachusetts	6,794,228	6,823,608	6,859,789	6,882,635	6,892,503	6,994,598	7,000,474	7,022,468	State
Michigan	9,931,715	9,950,571	9,973,114	9,984,072	9,986,857	10,072,703	10,041,351	10,050,877	State
Minnesota	5,482,032	5,522,744	5,566,230	5,606,249	5,639,632	5,710,735	5,718,660	5,721,621	State
Mississippi	2,988,471	2,987,938	2,988,510	2,981,020	2,976,149	2,958,536	2,947,209	2,941,939	State
Missouri	6,071,732	6,087,135	6,106,670	6,121,623	6,137,428	6,154,744	6,171,374	6,179,414	State
Montana	1,030,475	1,040,859	1,052,482	1,060,665	1,068,778	1,087,230	1,106,522	1,122,095	State
Nebraska	1,891,277	1,905,616	1,915,947	1,925,614	1,934,408	1,963,387	1,964,537	1,972,246	State
Nevada	2,866,939	2,917,563	2,969,905	3,027,341	3,080,156	3,116,967	3,148,141	3,176,116	State
New Hampshire	1,336,350	1,342,307	1,348,787	1,353,465	1,359,711	1,378,756	1,387,677	1,396,678	State
New Jersey	8,867,949	8,870,827	8,885,525	8,886,025	8,882,190	9,272,794	9,270,541	9,295,227	State
New Mexico	2,089,291	2,091,630	2,091,784	2,092,741	2,096,829	2,118,606	2,117,333	2,113,868	State

New York	19,654,666	19,633,428	19,589,572	19,530,351	19,453,561	20,105,171	19,848,276	19,703,747	State
North Carolina	10,031,646	10,154,788	10,268,233	10,381,615	10,488,084	10,449,652	10,564,320	10,710,793	State
North Dakota	754,066	754,434	754,942	758,080	762,062	779,563	777,966	781,057	State
Ohio	11,617,527	11,634,370	11,659,650	11,676,341	11,689,100	11,798,905	11,767,344	11,777,874	State
Oklahoma	3,909,500	3,926,331	3,931,316	3,940,235	3,956,971	3,965,415	3,992,238	4,026,229	State
Oregon	4,015,792	4,089,976	4,143,625	4,181,886	4,217,737	4,243,779	4,254,280	4,247,372	State
Pennsylvania	12,784,826	12,782,275	12,787,641	12,800,922	12,801,989	12,996,143	13,015,571	12,984,990	State
Rhode Island	1,056,065	1,056,770	1,055,673	1,058,287	1,059,361	1,096,530	1,097,246	1,099,498	State
South Carolina	4,891,938	4,957,968	5,021,268	5,084,156	5,148,714	5,132,249	5,194,274	5,287,935	State
South Dakota	853,988	862,996	872,868	878,698	884,659	887,948	896,492	909,723	State
Tennessee	6,591,170	6,646,010	6,708,799	6,771,631	6,829,174	6,927,904	6,965,740	7,062,217	State
Texas	27,470,056	27,914,410	28,295,273	28,628,666	28,995,881	29,239,570	29,570,351	30,113,488	State
Utah	2,981,835	3,041,868	3,101,042	3,153,550	3,205,958	3,284,077	3,339,738	3,391,011	State
Vermont	625,216	623,657	624,344	624,358	623,989	642,977	647,210	648,142	State
Virginia	8,361,808	8,410,106	8,463,587	8,501,286	8,535,519	8,637,615	8,658,910	8,683,414	State

Washington	7,163,657	7,294,771	7,423,362	7,523,869	7,614,893	7,727,209	7,743,760	7,794,123	State
West Virginia	1,842,050	1,831,023	1,817,004	1,804,291	1,792,147	1,791,646	1,785,618	1,774,122	State
Wisconsin	5,760,940	5,772,628	5,790,186	5,807,406	5,822,434	5,897,375	5,881,608	5,903,975	State
Wyoming	585,613	584,215	578,931	577,601	578,759	577,681	579,636	581,978	State
Guam	164,905	164,233	163,490	162,679	161,808	162,158	163,730	165,180	Territory
Virgin Islands	100,248	97,787	95,118	92,302	89,373	87,707	87,223	86,507	Territory
Puerto Rico	3,473,232	3,406,672	3,325,286	3,193,354	3,193,694	3,281,557	3,262,693	3,220,113	Territory

Appendix B

States/Territories in each Region (Northeast, Midwest, West, South) according to the CDC.

No	Location	Region
1	Connecticut	Northeast
2	Maine	Northeast
3	Massachusetts	Northeast
4	New Hampshire	Northeast
5	New Jersey	Northeast
6	New York	Northeast
7	Pennsylvania	Northeast
8	Rhode Island	Northeast
9	Vermont	Northeast
10	Illinois	Midwest
11	Indiana	Midwest
12	Iowa	Midwest
13	Kansas	Midwest
14	Michigan	Midwest
15	Minnesota	Midwest
16	Missouri	Midwest
17	Nebraska	Midwest
18	North Dakota	Midwest
19	Ohio	Midwest
20	South Dakota	Midwest

21	Wisconsin	Midwest
22	Alabama	South
23	Arkansas	South
24	Delaware	South
25	District of Columbia	South
26	Florida	South
27	Georgia	South
28	Kentucky	South
29	Louisiana	South
30	Maryland	South
31	Mississippi	South
32	North Carolina	South
33	Oklahoma	South
34	South Carolina	South
35	Tennessee	South
36	Texas	South
37	Virginia	South
38	West Virginia	South
39	Alaska	West
40	Arizona	West
41	California	West
42	Colorado	West
43	Hawaii	West
44	Idaho	West

45	Montana	West
46	Nevada	West
47	New Mexico	West
48	Oregon	West
49	Utah	West
50	Washington	West
51	Wyoming	West

Appendix C

Additional Data Cleaning Steps that were done before the Visualization using **Power BI**:

No.	Column Name	Action	Notes
1	Column1	**Removed Column**	Column dropped as it was an **index column** created from data handling in Jupyter Notebook
2	YearStart	**Extracted Year**	Extracted the Year in the **format YYYY** after having to convert into its type into datetime for Time Interpolation using Jupyter Notebook
3	YearStart	**Reordered Column**	Moved YearStart to be the **second column** as RowId is the unique identified for each row
4	Per 100,000 People	**Changed Type**	Changed the type from Float to an **Integer** as it represents a number of people
5	Per Capita Ratio_Per	**Changed Type**	Changed the type to a **Float**
6	StratificationID2	**Renamed Column**	Renamed into **Age_Race_Sub_Category_ID2**
7	StratificationCategoryID2	**Renamed Column**	Renamed into **Age_Race_Category_ID2**
8	Stratification2	**Renamed Column**	Renamed into **Age_Race_Sub_Category_Name**
9	StratificationCategory2	**Renamed Column**	Renamed into **Age_Race_Category_Name**
10	StratificationCategory1	**Renamed Column**	Renamed into **AgeCategoryFilter**
11	Stratification1	**Renamed Column**	Renamed into **Age_Category_Name**
12	Data_Value_Interpolated_over100	**Changed Data Type**	Changed from **Decimal** to **Percentage** with two decimal points
13	Per Capita Ratio_Per	**Dropped Column**	Dropped **unused column**

Kindly refer to the **Data Handling** section for the rest of the steps.