Which State Will Become Obese Next?

An Analysis on Patterns of Obesity Spread on States in the United States

for Citadel Asia Datathon 2024

Amy Ling Yuhui, National University of Singapore

Ho Kai Ting, National University of Singapore

Ethan Wei Yuxin, National University of Singapore

24th March 2024

1 Executive Summary

In this report, we aim to identify the states in the US which are most likely to see an adverse increase in the rate of obesity over the next decade. This analysis will better inform state governments as they prepare their healthcare infrastructure budgets and public health intervention plans to mitigate the adverse effects in the years ahead. We utilised Streamlit, a data visualisation tool to create a dashboard for convenient viewing and tinkering with data (refer to Appendix A to try out the dashboard).

1.1 Background

Living in the land of the free means being able to eat whatever you want, whenever you want, wherever you want. But the human's basal inclination for pleasure and dopamine chasing inches the American population towards chaos, in particular, in health.

Socially engineered are the advertisements and icing-laced mouth-watering foods by the big corps, they have been designed to appeal to freedom-to-live and dopamine-chasing Americans.

It would be a blessing if states knew when these undercurrents of cultural habits and increase in interest in unhealthy foods are arriving in their jurisdiction, and be able to proactively protect their people they are sworn to protect and support their people with the healthcare they require.

Fortunately, we have found that the distribution and determinants of obesity across the nation are not uniform, influenced by complex interactions between socioeconomic status, access to healthy foods, physical activity opportunities, and prominently migration patterns.

Since time immemorial, Americans have been undergoing waves of internal migration.

The first wave was westward and rural-to-urban migration beginning with colonisation as places became more industrialised.

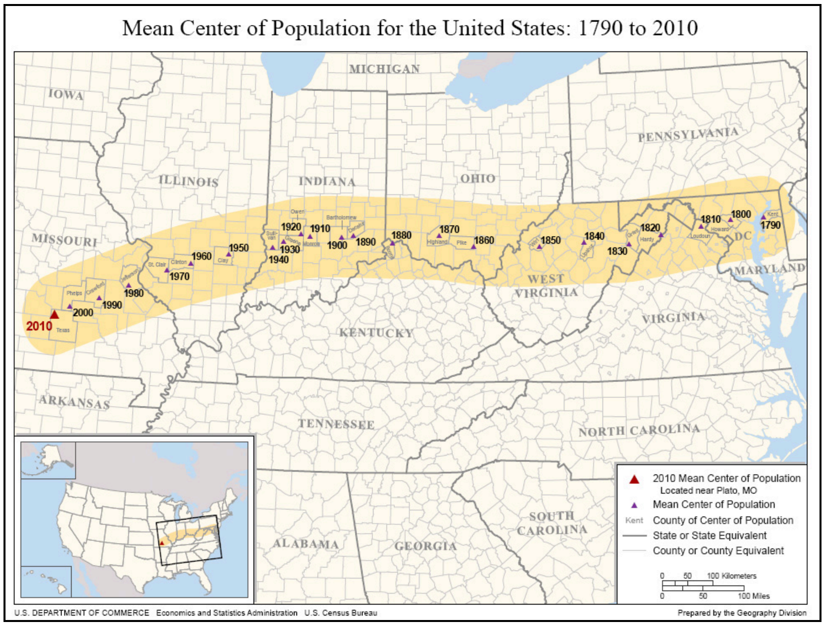


Fig 1. Mean Center of Population for the United States: 1790 to 2010

The second wave was the Great Migration of African Americans from the rural South to cities in other regions from 1916-1970, mostly for economic reasons like jobs in northern factories paying higher wages and escaping racial discrimination.[1]

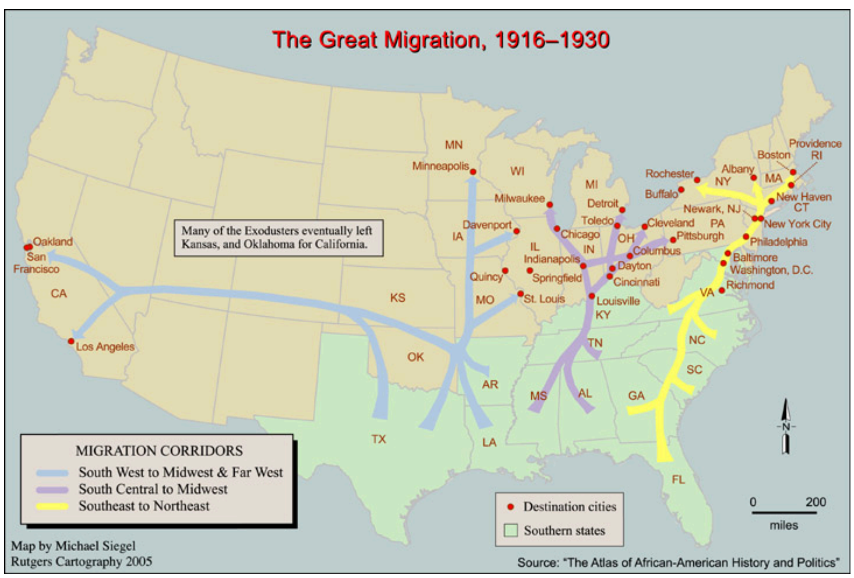


Fig 2. Migration corridors in the United States during The Great Migration

[1] Ms. Silvius's AP Human Geography. (2024). *U.S. Internal Migration*. [online] Available at: https://silviusaphg.weebly.com/us-internal-migration.html [Accessed 23 Mar. 2024].

The third wave was the shift from the Rustbelt to the Sunbelt regions from the post-WWII era to the present. The Rustbelt regions also experienced closing factories and poor economies pushing people to migrate South and West for new opportunities.[2]

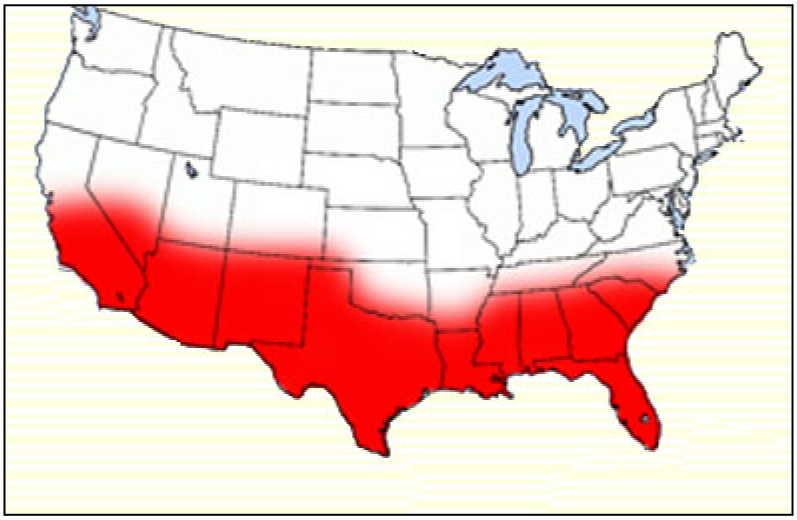


Fig 3. Migration patterns in the United States from the Rustbelt to Sunbelt regions.

Now, we witness a fourth wave of migration from the Central Eastern region toward the West, South and North East, bringing not only wealth but also cultural and dietary habits.

The pattern of obesity's spread is thus both a reflection of existing inequalities and a predictor of future health disparities. By identifying states at the beginning of this trend, targeted interventions can be designed to curb the epidemic's growth and address its root causes. This can come in the form of policies like subsidies and programmes that incentivise a healthier lifestyle for the population.

[2] Ms. Silvius's AP Human Geography. (2024). *U.S. Internal Migration*. [online] Available at: https://silviusaphg.weebly.com/us-internal-migration.html [Accessed 23 Mar. 2024].

1.2 Summary of Findings:

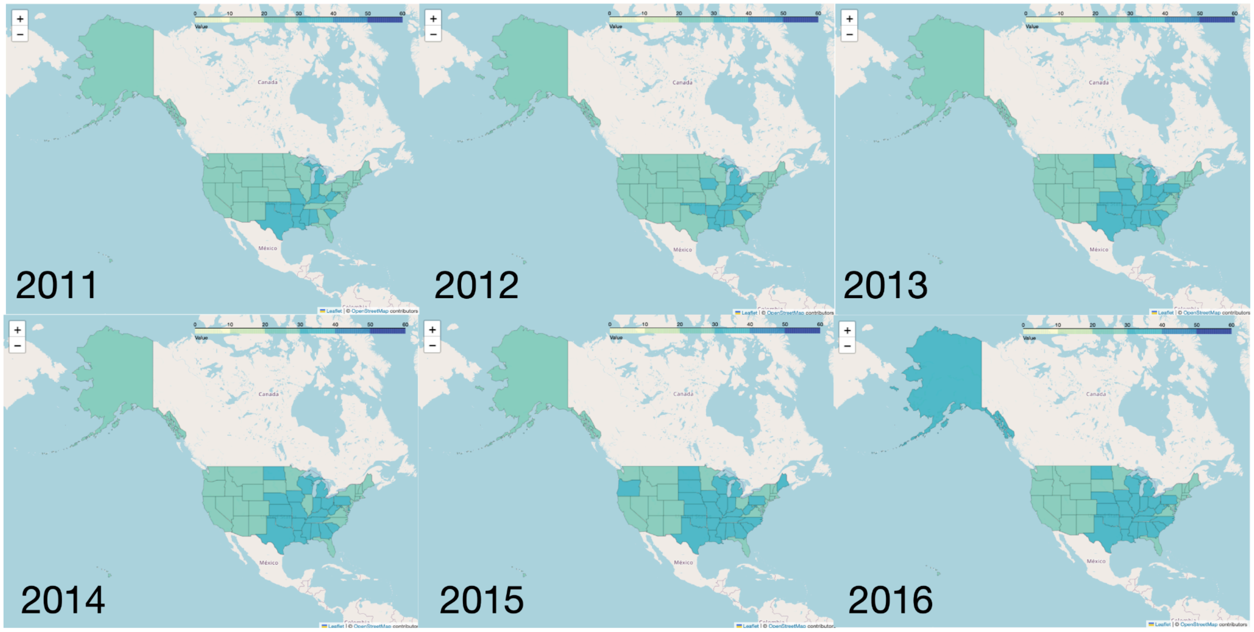


Fig. 4: Map of the United States showing the percentage of obesity by state from 2011 to 2016.

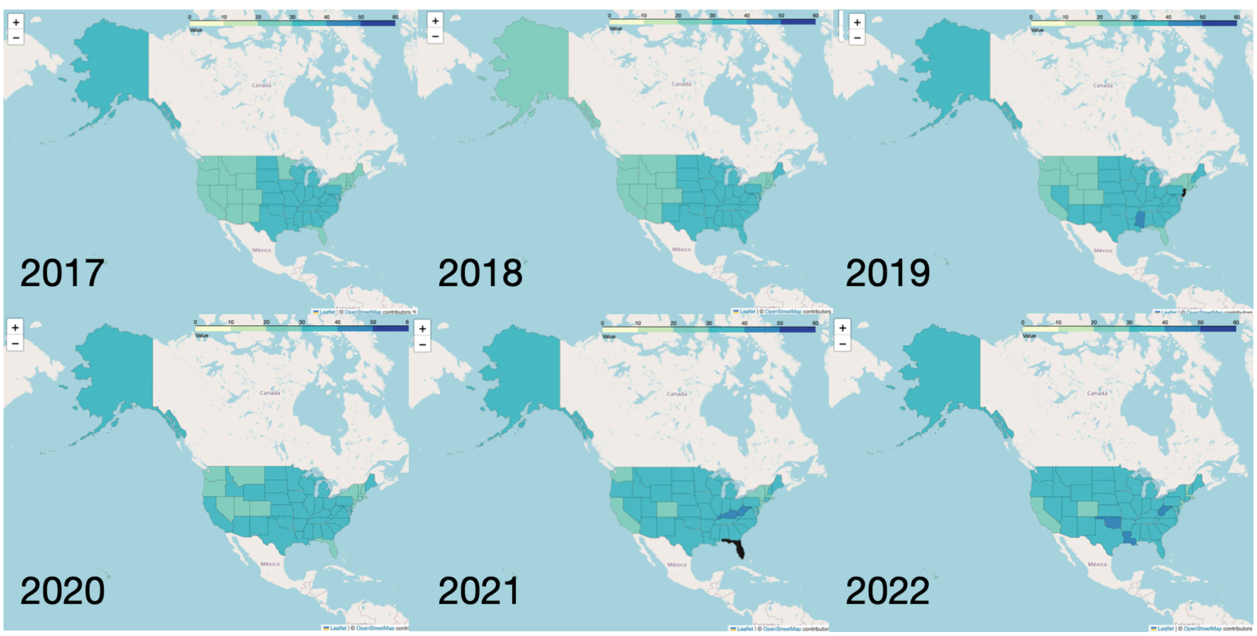


Fig. 5: Map of the United States showing the percentage of obesity by state from 2017 to 2022.

The two figures above show state-level heatmaps with **dark blue** indicating a **higher** percentage of **obesity** and **light green** indicating a **lower** percentage of **obesity**. We identify that over 6 years, states with more than 50% of obese people spread causes the obesity percentages to increase in neighbouring states that share a common border while non-bordered states show limited change in obesity percentages. This piqued our interest in investigating if the migration trends from 2011 to 2022 can explain this pattern of obesity.

2 Technical Exposition

2.1 Data Collection & Cleaning

2.1.1 Description of Dataset

Nutrition, Physical activity, and Obesity Dataset

The first dataset we used was given as part of the datathon datasets. It details nutrition, physical activity, and obesity data, and was derived from the CDC Behavioral Risk Factor Surveillance System. This comprehensive dataset spans from 2001 to 2022 and encompasses a wide array of variables related to obesity metrics, demographic information, and behaviours affecting obesity, such as hours of TV watched per day.

Encompassing approximately 133,000 entries and spanning 34 columns, this dataset provides an in-depth look at obesity and reported physical activity across U.S. states. Key variables include the start and end years of reporting, state abbreviations and descriptions, detailed questions related to obesity and physical activity, data values presented as percentages, confidence limits, and stratifications based on age, education, gender, income, and race/ethnicity, among others. It offers a granular view into the prevalence and distribution of obesity within the United States, making it a valuable resource for public health analysis and policy-making.

State-to-State Migration Flows Dataset

As we were interested in investigating migration flows in relation to obesity rates, we self-sourced the state-to-state migration flow data from the United States Census Bureau (https://www.census.gov/data/tables/time-series/demo/geographic-mobility/state-to-state-migration.html). The data bank consists of the estimates of people migrating from an origin to a destination state, and has data from 2004 to 2022. While it is fairly complete across time, there were several inconsistencies in some years' records. For example, the 2020 migration data is unavailable (likely due to COVID-19), and the records of 2004 and 2005 were combined.

2.1.2 Preprocessing Steps

Nutrition, Physical activity, and Obesity Dataset

The steps we took to clean the data are as follows:

1. Filtering down to data from 2011 to 2022
    a. The data before 2011 had substantial missing data so we could not perform meaningful analysis on them
2. Filtering down to Obesity data only
    a. Since obesity rates are our dependent variable
3. Filtering down to Stratification Category "Total"
    a. Since we are interested in the overall state obesity rates, as opposed to smaller stratification categories.
4. Pruning away 3 of 4 questions:
    a. Pruned questions: The data from these questions were used purely for visualisation purposes. This is because they had a lot of missing data and there was no significant trend observed over time.
        i. Percent of students in grades 9-12 who have obesity.
        ii. Percent of students in grades 9-12 who have an overweight classification.
        iii. Percent of adults aged 18 years and older who have an overweight classification.
    b. Kept question: We focused our analysis on the data for this question. The data for this question was almost complete. And there was an observable trend of obesity rates spreading by the proximity of regions to "more obese" states.
        i. Percent of adults aged 18 years and older who have obesity.
5. Handling missing data for the kept question
    a. For the data we will be focusing our analysis on, there were 2 main missing data points: 2019 New Jersey, and 2021 Florida. We chose to fill these missing data points with the average of the previous and subsequent years' data points. For example, 2021 Florida's data value was filled with the average of 2020 and 2022 Florida data values.

State-to-State Migration Flows Dataset

As this dataset was taken raw from the US Census Bureau, it required a lot of pre-processing in order to make it compatible with our analysis.

Firstly, the databank consists of one datasheet for each year. Moreover, each datasheet was given non-standard format, with extra columns and rows throughout the spreadsheet (as seen in Fig. 6)

Table 1. State-to-State Migration Flows[1]: 2022
Dataset: 2022 American Community Survey 1-Year Estimates
Universe: Population 1 year and over

	Population 1 year and over		Same house 1 year ago		Same state of residence 1 year ago		Different state of residence 1 year ago			
							Total		Alabama	
Current residence	Estimate	MOE	Estimate	MOE	Estimate	MOE	Estimate	MOE	Estimate	MOE
United States[2]	329,820,603	+/- 31,045	288,415,913	+/- 187,620	31,031,895	+/- 173,674	8,230,953	+/- 67,325	102,894	+/- 10,640
Alabama	5,022,366	+/- 4,086	4,397,470	+/- 21,660	471,530	+/- 19,721	139,263	+/- 10,142	N/A	N/A
Alaska	724,196	+/- 1,151	614,091	+/- 8,896	68,351	+/- 5,851	36,563	+/- 5,991	151	+/- 125
Arizona	7,285,247	+/- 5,106	6,264,863	+/- 32,211	690,856	+/- 26,086	282,729	+/- 15,227	727	+/- 448
Arkansas	3,007,872	+/- 3,599	2,618,120	+/- 16,164	294,640	+/- 13,727	86,375	+/- 8,179	1,911	+/- 1,779
California	38,629,179	+/- 10,318	34,335,186	+/- 57,597	3,515,077	+/- 48,926	475,803	+/- 21,187	5,204	+/- 2,639
Colorado	5,781,381	+/- 3,527	4,834,425	+/- 28,072	682,450	+/- 23,806	229,876	+/- 11,539	700	+/- 532
Connecticut	3,588,965	+/- 3,572	3,172,833	+/- 17,815	245,100	+/- 14,054	145,315	+/- 10,526	1,005	+/- 1,141
Delaware	1,008,173	+/- 1,919	899,454	+/- 8,649	57,963	+/- 6,154	46,162	+/- 5,074	203	+/- 253
District of Columbia	660,942	+/- 2,008	515,345	+/- 9,328	68,363	+/- 8,248	64,506	+/- 5,229	419	+/- 363
Florida	22,043,900	+/- 8,811	19,029,143	+/- 52,322	2,008,758	+/- 47,284	738,969	+/- 27,539	14,734	+/- 3,244
Georgia	10,791,161	+/- 7,077	9,348,359	+/- 33,252	1,054,380	+/- 29,926	327,795	+/- 16,598	21,031	+/- 5,430
Hawaii	1,425,611	+/- 1,842	1,251,057	+/- 10,923	106,642	+/- 8,443	56,209	+/- 6,444	1,611	+/- 1,645
Idaho	1,919,357	+/- 1,777	1,633,120	+/- 11,362	189,164	+/- 11,119	87,949	+/- 7,501	0	+/- 195
Illinois	12,455,441	+/- 5,747	11,002,931	+/- 32,911	1,150,055	+/- 30,073	228,308	+/- 13,363	2,389	+/- 1,105
Indiana	6,757,160	+/- 4,612	5,886,762	+/- 28,260	691,986	+/- 25,311	149,331	+/- 11,366	792	+/- 538
Iowa	3,166,734	+/- 2,669	2,760,076	+/- 12,059	323,444	+/- 11,589	72,231	+/- 5,843	180	+/- 190
Kansas	2,906,619	+/- 2,657	2,480,630	+/- 13,894	320,201	+/- 13,110	94,208	+/- 6,737	1,371	+/- 1,165
Kentucky	4,462,146	+/- 3,954	3,892,713	+/- 19,644	439,944	+/- 17,580	113,197	+/- 8,594	1,218	+/- 880
Louisiana	4,537,185	+/- 4,410	4,006,126	+/- 19,111	435,538	+/- 20,212	75,330	+/- 7,049	1,332	+/- 691
Maine	1,372,172	+/- 1,657	1,230,229	+/- 8,250	95,823	+/- 6,223	41,618	+/- 4,329	274	+/- 390
Maryland	6,100,234	+/- 5,254	5,395,850	+/- 19,513	511,449	+/- 19,836	139,784	+/- 8,687	1,256	+/- 827
Massachusetts	6,918,482	+/- 4,087	6,065,294	+/- 22,939	606,626	+/- 19,958	171,077	+/- 9,818	1,082	+/- 533
Michigan	9,936,710	+/- 5,396	8,786,959	+/- 25,314	949,469	+/- 20,510	157,955	+/- 10,391	1,846	+/- 1,422
Minnesota	5,654,602	+/- 3,147	4,940,488	+/- 19,839	571,566	+/- 16,933	117,016	+/- 6,849	732	+/- 581
Mississippi	2,907,327	+/- 3,124	2,584,891	+/- 13,612	238,542	+/- 14,259	69,948	+/- 6,563	5,444	+/- 2,118
Missouri	6,111,432	+/- 4,353	5,320,580	+/- 22,369	609,551	+/- 18,967	163,254	+/- 9,218	570	+/- 703
Montana	1,111,641	+/- 1,372	964,155	+/- 8,376	96,436	+/- 7,364	48,165	+/- 4,543	168	+/- 273
	Population 1 year and over		Same house 1 year ago		Same state of residence 1 year ago		Different state of residence 1 year ago			
							Total		Alabama	
Current residence	Estimate	MOE	Estimate	MOE	Estimate	MOE	Estimate	MOE	Estimate	MOE

Table 1. State-to-State Migration Flows[1]: 2022
Dataset: 2022 American Community Survey 1-Year Estimates
Universe: Population 1 year and over

	Different state of residence 1 year ago (continued)									
	Alaska		Arizona		Arkansas		California		Colorado	
Current residence	Estimate	MOE	Estimate	MOE	Estimate	MOE	Estimate	MOE	Estimate	MOE
United States[2]	32,755	+/- 5,463	204,734	+/- 13,929	74,408	+/- 8,399	817,669	+/- 27,227	239,200	+/- 12,017
Alabama	714	+/- 660	1,202	+/- 684	1,536	+/- 1,130	8,012	+/- 2,839	3,329	+/- 1,679
Alaska	N/A	N/A	1,645	+/- 1,041	10	+/- 32	3,855	+/- 1,312	665	+/- 693
Arizona	2,091	+/- 1,283	N/A	N/A	1,147	+/- 676	74,157	+/- 8,989	12,232	+/- 2,678
Arkansas	64	+/- 4	1,776	+/- 1,012	N/A	N/A	7,783	+/- 2,545	2,562	+/- 1,218
California	2,040	+/- 1,066	27,412	+/- 4,964	963	+/- 534	N/A	N/A	19,970	+/- 4,135
Colorado	1,961	+/- 1,107	7,386	+/- 2,214	2,557	+/- 1,244	33,213	+/- 5,144	N/A	N/A
Connecticut	97	+/- 159	1,267	+/- 909	625	+/- 703	7,097	+/- 2,014	1,992	+/- 568
Delaware	0	+/- 203	949	+/- 742	0	+/- 203	660	+/- 417	123	+/- 162
District of Columbia	0	+/- 238	712	+/- 485	21	+/- 2	4,539	+/- 1,469	1,187	+/- 699
Florida	1,418	+/- 955	11,901	+/- 3,380	3,204	+/- 1,334	50,701	+/- 6,891	20,980	+/- 4,316
Georgia	210	+/- 215	1,798	+/- 931	531	+/- 373	25,960	+/- 4,784	5,419	+/- 1,728
Hawaii	773	+/- 676	2,750	+/- 1,606	75	+/- 148	10,562	+/- 2,423	2,750	+/- 1,056
Idaho	411	+/- 376	3,633	+/- 1,532	299	+/- 370	26,887	+/- 4,189	1,655	+/- 834
Illinois	434	+/- 442	5,377	+/- 1,502	1,468	+/- 892	20,573	+/- 4,774	6,720	+/- 1,907
Indiana	1,118	+/- 1,093	4,953	+/- 2,181	2,888	+/- 2,856	7,769	+/- 2,309	3,586	+/- 1,827
Iowa	118	+/- 144	2,264	+/- 992	580	+/- 395	2,349	+/- 645	1,932	+/- 764
Kansas	218	+/- 199	1,519	+/- 743	1,016	+/- 839	5,520	+/- 2,495	7,359	+/- 1,943
Kentucky	182	+/- 214	2,618	+/- 1,066	343	+/- 357	5,985	+/- 1,690	2,182	+/- 1,251
Louisiana	432	+/- 376	315	+/- 235	2,429	+/- 1,022	4,847	+/- 2,190	3,329	+/- 2,077
Maine	11	+/- 20	138	+/- 123	27	+/- 46	3,552	+/- 1,512	1,332	+/- 1,000
Maryland	364	+/- 414	1,439	+/- 988	67	+/- 87	7,198	+/- 2,226	1,148	+/- 590
Massachusetts	154	+/- 156	2,251	+/- 1,261	904	+/- 1,106	18,543	+/- 3,424	2,186	+/- 1,531
Michigan	487	+/- 356	3,772	+/- 1,276	640	+/- 464	13,939	+/- 3,247	4,589	+/- 1,959
Minnesota	269	+/- 274	4,234	+/- 1,412	713	+/- 325	6,908	+/- 2,253	3,620	+/- 1,285
Mississippi	90	+/- 161	1,848	+/- 807	1,290	+/- 638	1,638	+/- 797	2,701	+/- 1,232
Missouri	414	+/- 544	3,576	+/- 1,527	9,590	+/- 2,328	12,107	+/- 2,963	8,042	+/- 2,213
Montana	498	+/- 390	2,415	+/- 1,119	32	+/- 43	4,660	+/- 1,623	3,440	+/- 1,121
	Different state of residence 1 year ago (continued)									
	Alaska		Arizona		Arkansas		California		Colorado	
Current residence	Estimate	MOE	Estimate	MOE	Estimate	MOE	Estimate	MOE	Estimate	MOE

Fig. 6: Original Datasheet Format

Therefore we had to remove all the inconsistencies and combine all separate data into a single standard dataframe, with columns [To, From, Year, Type, Value].

Column Name	Data Type	Explanation
To	str	The destination state of the migration
From	str	The origin state of the migration
Year	int	The year of the migration flow
Type	str	Data type of the migration flow measure, could be an estimate or a margin of error
Value	str	The numerical value of the migration flow measure

Table 1: Explanation of preprocessed migration dataframe column names

Dealing with missing data:

Secondly, the time frame we are interested in i.e. 2011-2022, lacked 2020's state-to-state migration flows data. We decided to fill this data point instead of leaving it null as it will give us a better idea of the trends over time.

We chose to handle this missing data point by duplicating 2021's data for 2020. This is because 2020 and 2021 were both years where migration was similarly impacted by COVID-19. Hence, we made the reasonable assumption that 2020's migration rates would be the most similar to 2021's migration rates.

Lastly, for neighbouring migration, there are certain states in the US, such as Alaska, that have no direct neighbours. Since no meaningful analysis can be performed for such states, we chose to exclude them from our study entirely (although the data is still present in our dashboard for visualisation purposes only), which brings the total number of states in our study to 47.

2.2 Methodology

Our analysis is broken down into three main stages.

2.2.1 Trend observed in obesity rates by location over time

First, we mapped out the obesity rates by location over time to visualise any trends. Using the LocationDesc parameter in our dataset, we mapped the state name over to the location coordinates of the state using a GeoJson dataset. We then plotted the obesity rate to that region on the map. The same legend was used across all maps, in order to visually observe trends across different maps.

We had a different map for each year and each of the following questions.

1) Percent of students in grades 9-12 who have obesity.
2) Percent of students in grades 9-12 who have an overweight classification.
3) Percent of adults aged 18 years and older who have an overweight classification.
4) Percent of adults aged 18 years and older who have obesity.

2.2.2 Correlation Study between Neighbouring Migration and Obesity Rates

Our findings from the above data visualisation process led us to do a correlation study between neighbouring migration rates and obesity rates. **Neighbouring migration** is defined in this report as migration from a neighbouring state. Neighbouring states are defined as sharing a common land border as the destination state. The state-to-state migration flow data had estimates of the number of people migrating from an origin state to a destination state. We filtered the data to only consider people migrating from neighbouring states into a destination state.

- Dependent variable: Obesity rate
- Independent variables: Neighbouring migration rate (which is directly tied to location)

Since neighbouring migration might not have an immediate effect on obesity rates, defined as the **lag effect**, we investigated two possible interpretations of this hypothesised lag effect. Firstly, we conducted a cross-correlation study between neighbouring migration and obesity rates, taking into account different time lags. This could explain if there is a direct time lag of obesity rates with changes in migration rates. Secondly, we did a cross-correlation study with the simple moving average of migration rates. The rationale behind this is that obesity rates could be more affected by the aggregation of migration rates over a period of time, rather than the migration rate of only one year. We varied the window size of the moving average to observe changes in the correlation coefficients between neighbouring migration and obesity rates, as well as the number of states that meet the bar of a strong positive correlation coefficient of over 0.5.

2.2.3 Correlation Study between Foreign Migration and Obesity Rates

To prove that the correlation found between neighbouring migration and obesity rates is statistically significant, we conducted a similar correlation study between foreign migration and obesity rates as a control. **Foreign migration** is defined as the influx of immigrants from a country outside of the United States. This is to ensure that the correlation found between neighbouring migration and obesity rates was simply not a baseline trend, where both increase as a result of natural circumstances.

- Dependent variable: Obesity rate
- Independent variables: Foreign migration rate (which is directly tied to location)

Comparisons were made between the correlation coefficient between neighbouring migration and obesity rates, as well as the correlation coefficient between foreign migration and obesity rates for each specified state.

2.3 Main Findings:

2.3.1 Trend observed in obesity rates by location over time

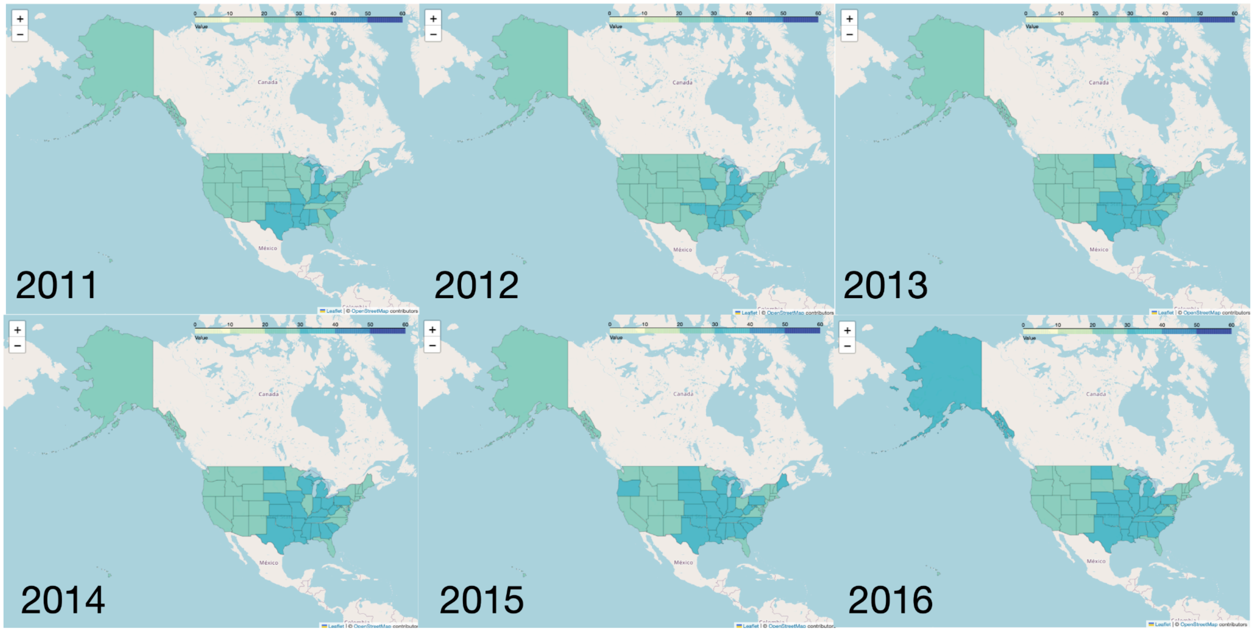


Fig. 7: Map of the United States showing the percentage of obesity by state from 2011 to 2016.

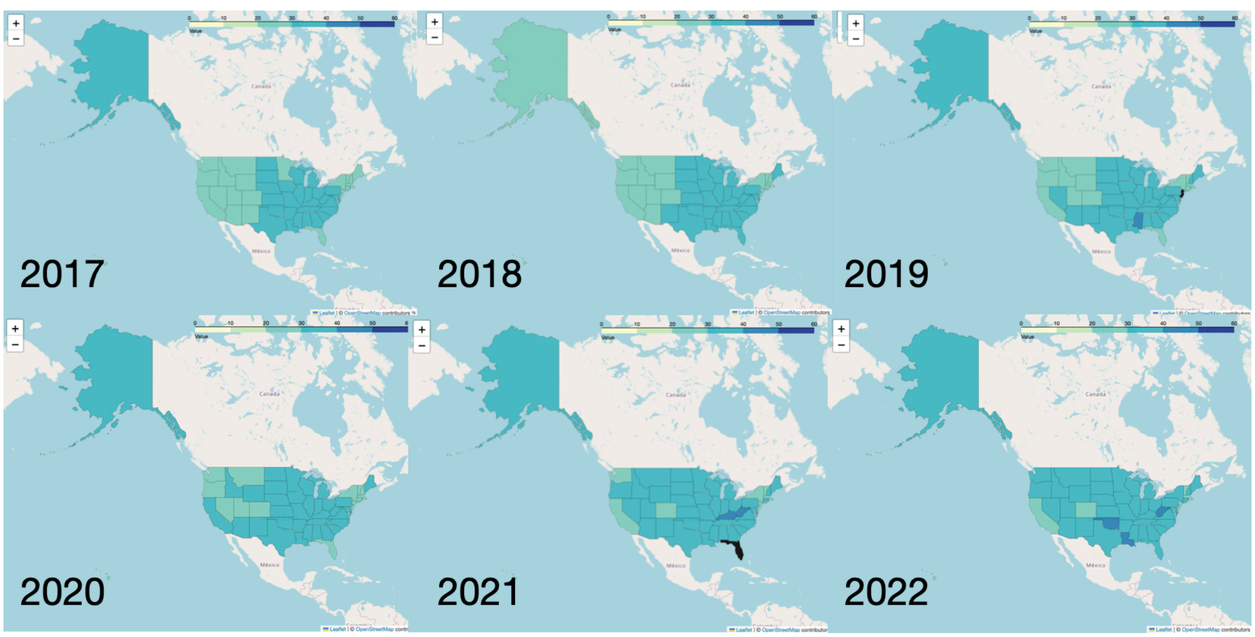
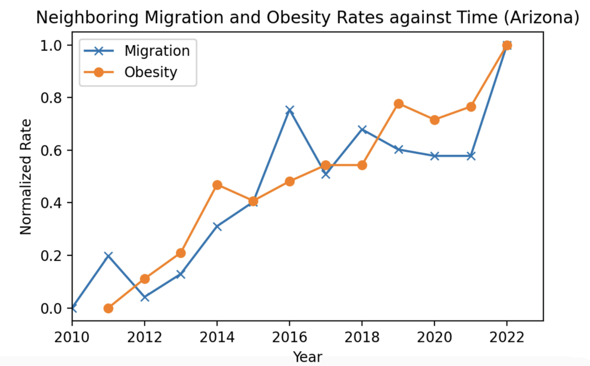


Fig. 8: Map of the United States showing the percentage of obesity by state from 2017 to 2022.

From the heatmaps above, we observe that states with closest proximity to the states with higher obesity rates see an increase in their own obesity rates over time. Moreover, states that had the highest obesity levels in 2011, such as West Virginia and Oklahoma, continued to have the highest obesity levels in 2022.

From this observation, we hypothesise that obesity rates were strongly correlated with migration from neighbouring states. We wanted to investigate the validity and extent of this correlation, as well as if all states experienced the same extent of this correlation.

2.3.2 Correlation Study between Neighboring Migration and Obesity Rates



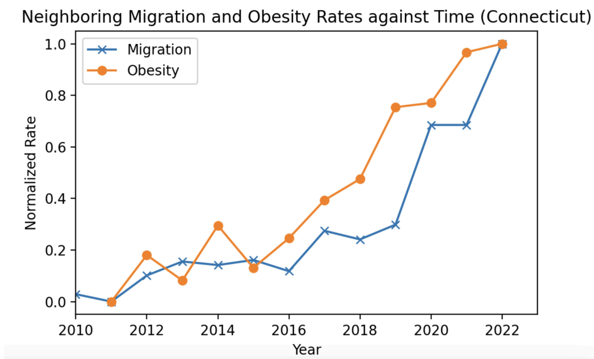
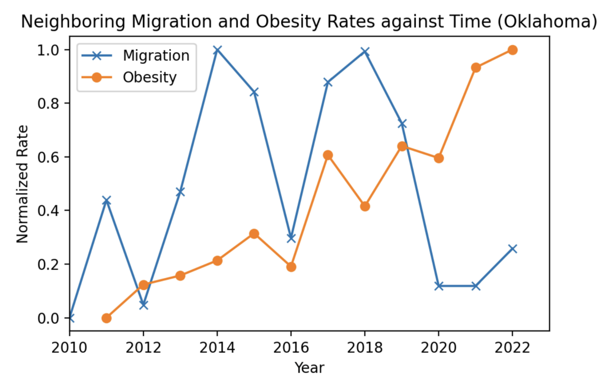


Fig. (9 and 10) Graphs of Neighbouring Migration and Obesity Rates (Normalised) against Time (Strong positive correlation)



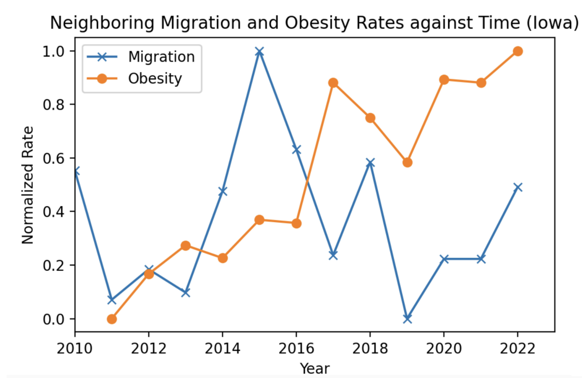
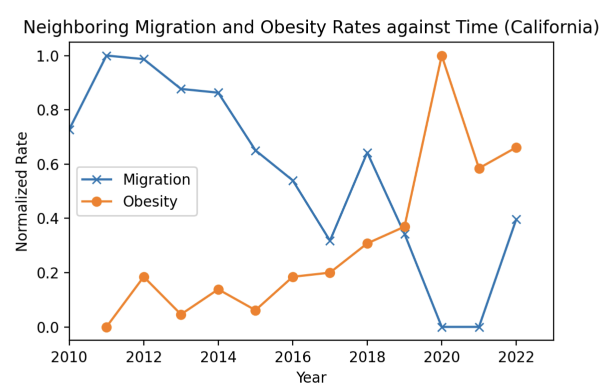


Fig. (11 and 12) Graphs of Neighbouring Migration and Obesity Rates (Normalised) against Time (Weak correlation)



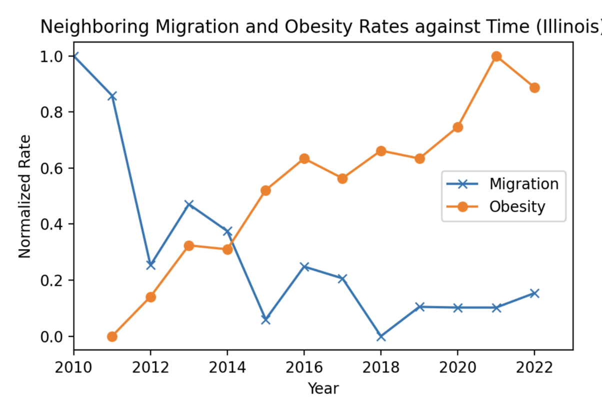


Fig. (13 and 14) Graphs of Neighbouring Migration and Obesity Rates (Normalised) against Time (Strong negative correlation)

First, we plotted the graphs of neighbouring migration and obesity rates against time. The migration and obesity rates were **normalised** to obtain a better visual representation of the relative changes of the two rates. We observe three categories of correlation among all the states considered in our study - those with strong positive correlation (Fig. 9-10) , those with weak correlation (Fig. 11-12) and those with strong negative correlation (Fig 13-14).

We then proceeded to perform a cross-correlation study between the two rates, and compute the correlation coefficients for each state.

	State	Correlation Coefficient
6	Connecticut	0.9053
2	Arizona	0.8621
37	Pennsylvania	0.8618
18	Maine	0.8564
29	New Jersey	0.6991
9	Georgia	0.6847
7	Delaware	0.6573
39	South Carolina	0.6499
11	Idaho	0.6343
15	Kansas	0.5965
13	Indiana	0.5466
46	Washington	0.5367
36	Oregon	0.4866
27	Nevada	0.407
20	Massachusetts	0.3657
0	Alabama	0.3085
32	North Carolina	0.2754
38	Rhode Island	0.1259
48	Wisconsin	0.0934
44	Vermont	0.0747

Fig 15. Correlation coefficients of the top 20 states

We observe that 12 out of 47 states meet the bar of strong positive correlation coefficients of over 0.5.

Furthermore, we plot the obesity rates against migration rates for a state in each category to ascertain our findings.

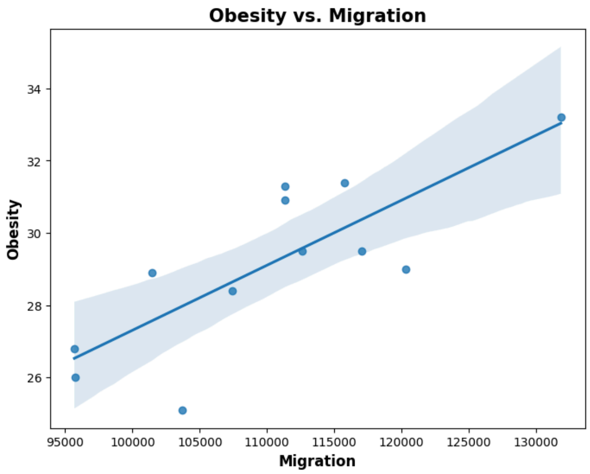


Fig. 16 Obesity vs Migration Rates for Arizona

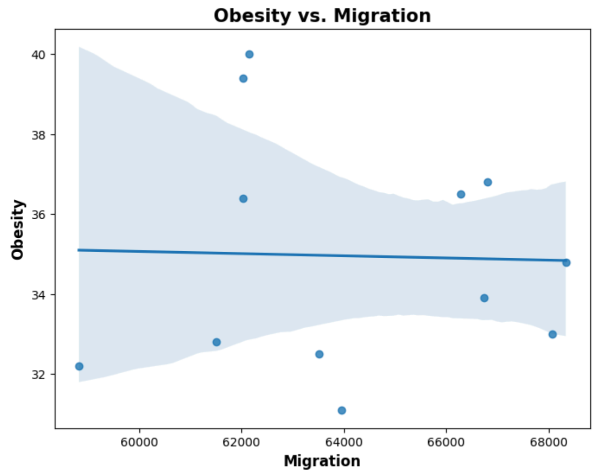


Fig. 17 Obesity vs Migration Rates for Oklahoma

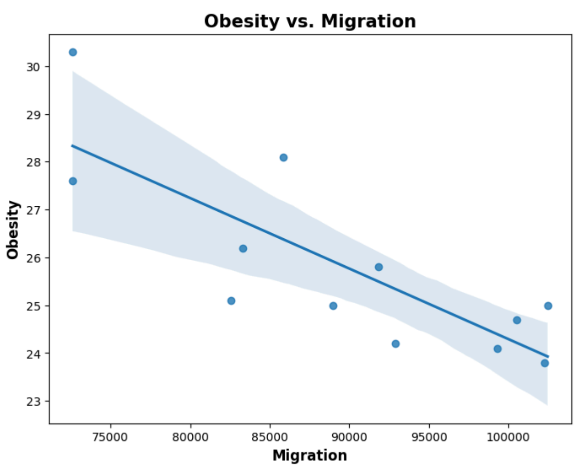


Fig. 18 Obesity vs Migration Rates for California

For states in the first category such as Arizona, there is an evident strong positive correlation between obesity and migration rates. For states in the second category such as Oklahoma, there is no clear correlation between the two rates. And for states in the third category such as California, there is a strong negative correlation between the two rates.

2.3.2.1 Lag Effect

Exploring the direct lag effect

Due to the possibility of a lead-lag relationship between obesity and rate of migration, in which a change in rate of obesity may have been correlated to state migrations that had occurred before or after, we plotted a time-lagged cross correlation graph to further explore their relationship.

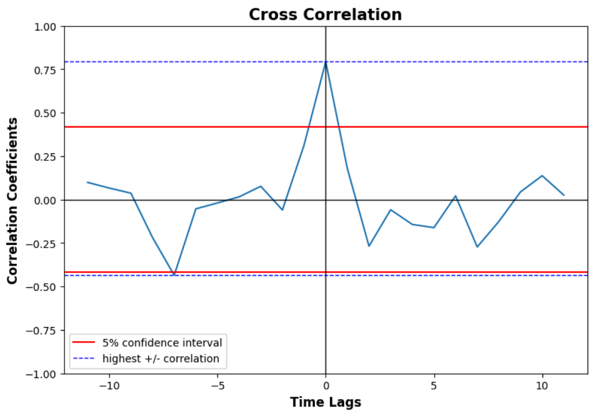


Fig. 19: Time-lagged Cross Correlation Graph for Arizona

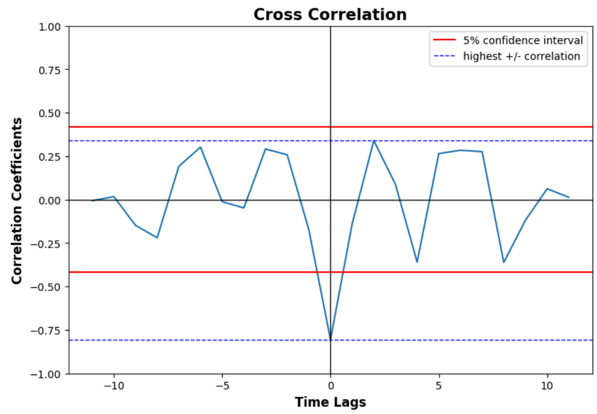


Fig. 20: Time-lagged Cross Correlation Graph for Oklahoma

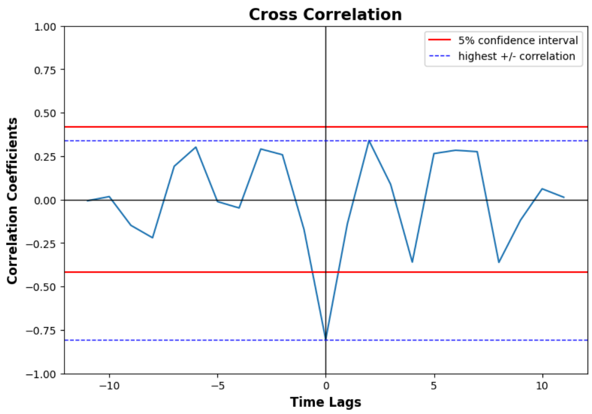


Fig. 21: Time-lagged Cross Correlation Graph for California

For Arizona, the time-lagged cross-correlation observed the largest rise in correlation at 0 years, which suggests that influx of immigrants from neighbouring states led to an increase in the obesity rate within the same year.

For Oklahoma and California, the time-lagged cross-correlation observed the largest dip in correlation at 0 years, which suggests that influx of immigrants from neighbouring states led to a decrease in the obesity rate within the same year.

For all three categories of states, the most extreme time-lagged cross correlation is observed immediately. This suggests that there is no evidence for a direct time lag effect between neighbouring migration rates and obesity rates of the state.

Exploring the aggregated lag effect

Instead of considering a direct time lag of our migration rates, we explore the aggregation of migration rates across a small number of years as a better justification to the hypothesised lag effect. This was performed by calculating the simple moving average of the raw neighbouring migration rates, and using this to calculate the cross correlation with obesity rates.

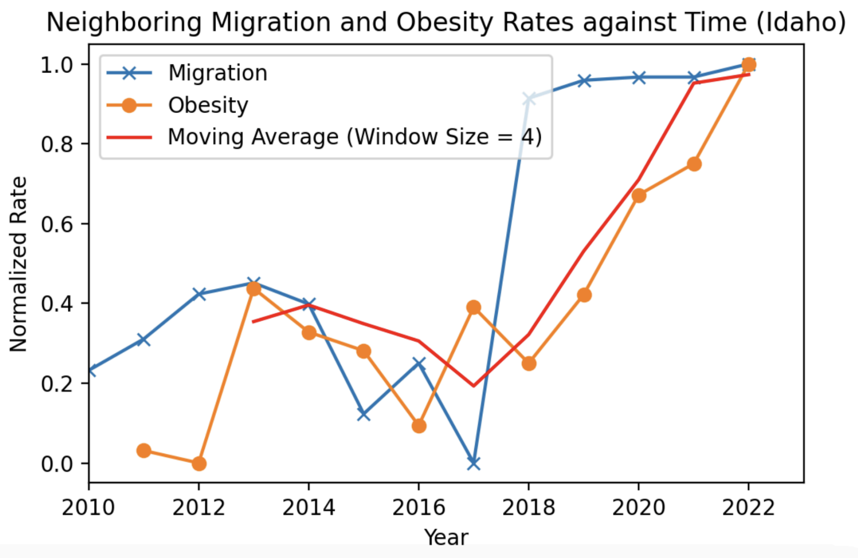


Fig. 22: Visualisation of the moving average line (window size = 4)

We observed that this feature engineering turned out to be a better indicator of migration rates than the raw or time lagged migration rates. As seen in Fig. 23, the number of states that satisfy strong positive correlation increased from 12 with window size = 1 to 17 with window size = 4.

Firstly, this could be because the effect of migration rates on obesity rates in the state spans across approximately four years, instead of being a one-to-one mapping by year. Secondly, the simple moving average smooths out small fluctuations between subsequent years and dampens the effect of the fluctuations.

	State	Correlation Coefficient
6	Connecticut	0.9053
2	Arizona	0.8621
37	Pennsylvania	0.8618
18	Maine	0.8564
29	New Jersey	0.6991
9	Georgia	0.6847
7	Delaware	0.6573
39	South Carolina	0.6499
11	Idaho	0.6343
15	Kansas	0.5965
13	Indiana	0.5466
46	Washington	0.5367
36	Oregon	0.4866
27	Nevada	0.407
20	Massachusetts	0.3657
0	Alabama	0.3085
32	North Carolina	0.2754
38	Rhode Island	0.1259
48	Wisconsin	0.0934
44	Vermont	0.0747

	State	Correlation Coefficient
18	Maine	0.9275
6	Connecticut	0.9037
11	Idaho	0.8972
9	Georgia	0.8911
37	Pennsylvania	0.8861
2	Arizona	0.8842
7	Delaware	0.8272
20	Massachusetts	0.8173
15	Kansas	0.7764
29	New Jersey	0.7655
13	Indiana	0.7211
39	South Carolina	0.7188
27	Nevada	0.6057
46	Washington	0.6027
19	Maryland	0.5767
32	North Carolina	0.5114
36	Oregon	0.5107
8	Florida	0.4571
45	Virginia	0.4498
38	Rhode Island	0.3875

Fig. 23: Correlation coefficients for top 20 states with window size = 1 (left) and window size = 4 (right)

Limitations

Since we used the Pearson correlation coefficient, which is a linear regression correlation metric, the underlying assumption of the model is that we are assuming a roughly linear trend in our data. However, this assumption is violated for a couple of states, such as Arkansas, where we see high fluctuations in migration rates over time. We have mitigated this by using the moving average line to smooth out small fluctuations in migration rates. However, this mitigation is insufficient for states with severe fluctuations in migration rates. Therefore, the correlation we found is more relevant for states whose neighbouring migration patterns follow this assumption, rather than those that do not.

2.3.3 Correlation Study between Foreign Migration and Obesity Rates

We conducted a similar study using foreign migration rates instead. We observe no strong (both positive and negative) correlation between foreign migration rates and obesity rates (Fig. 24). The highest observed correlation coefficient is in Nebraska, at 0.6387, which hardly compares to the very strong correlation observed with neighbouring migration rates at over 0.9. Moreover, only 2 states meet the strong positive correlation criteria of over 0.5, and even when we increase the window size of the simple moving average, the highest attainable number was 6.

	State	Correlation Coefficient Neighbor	Correlation Coefficient Foreign
29	Nebraska	-0.1189	0.6387
45	Louisiana	-0.6179	0.6188
4	New Jersey	0.6991	0.455
34	Mississippi	-0.2951	0.4279
30	Florida	-0.1492	0.3998
18	Wisconsin	0.0934	0.3933
37	Montana	-0.3842	0.3431
10	Indiana	0.5466	0.2928
6	Delaware	0.6573	0.2115
16	North Carolina	0.2754	0.2048
14	Massachusetts	0.3657	0.1948
42	Tennessee	-0.5045	0.1845
39	Texas	-0.4296	0.138
36	Ohio	-0.3768	0.1174
23	North Dakota	-0.0348	0.1076
11	Washington	0.5367	0.0725
8	Idaho	0.6343	0.0712
26	Wyoming	-0.1033	0.0689
7	South Carolina	0.6499	0.0576
13	Nevada	0.407	0.0408

Fig. 24: Comparison between correlation coefficients for neighbouring migration and foreign migration (sorted by foreign migration)

	State	Correlation Coefficient Neighbor	Correlation Coefficient Foreign
0	Connecticut	0.9053	-0.0915
1	Arizona	0.8621	-0.0406
2	Pennsylvania	0.8618	-0.199
3	Maine	0.8564	-0.4148
4	New Jersey	0.6991	0.455
5	Georgia	0.6847	-0.0056
6	Delaware	0.6573	0.2115
7	South Carolina	0.6499	0.0576
8	Idaho	0.6343	0.0712
9	Kansas	0.5965	-0.4054
10	Indiana	0.5466	0.2928
11	Washington	0.5367	0.0725
12	Oregon	0.4866	-0.146
13	Nevada	0.407	0.0408
14	Massachusetts	0.3657	0.1948
15	Alabama	0.3085	-0.5203
16	North Carolina	0.2754	0.2048
17	Rhode Island	0.1259	-0.3306
18	Wisconsin	0.0934	0.3933
19	Vermont	0.0747	-0.0552
20	Maryland	0.0572	-0.1876

Fig. 25: Comparison between correlation coefficients for neighbouring migration and foreign migration (sorted by neighbouring migration)

Additionally, we can also observe in Fig. 25 that states with strong positive correlation between neighbouring migration rates and obesity rates (e.g. Arizona) mostly experience weak positive correlation or negative correlation between foreign migration rates and obesity rates. This change in results observed could be due to the dilution of the number of obese US incumbents by the incoming immigrants from foreign countries. This observation strengthens the statistical evidence that obesity rates for these states are primarily fueled by neighbouring migration in particular.

2.4 Conjectures on why this is so

Migration Emotions and Fast Food Nostalgia

Internal migration is associated with anxiety and it also “increase(s) the risk of loneliness”.[3] Fast-food chains seem to be the quickest resolve.

When one migrates to a new neighbouring state for any reason be it economic, housing or job reasons, one would likely prefer a subculture not too different from their own established beliefs and practices. However, as with accents, attitudes of people and familiarity with places, one might feel foreign and on a psychological level, feel homesick.

As a result, one may seek familiarity – mega food franchises, which are dominated by fast food chains brands like McDonalds, Chick-fil-A, Dunkin' Donuts and Taco Bell, just to name a few.[4]

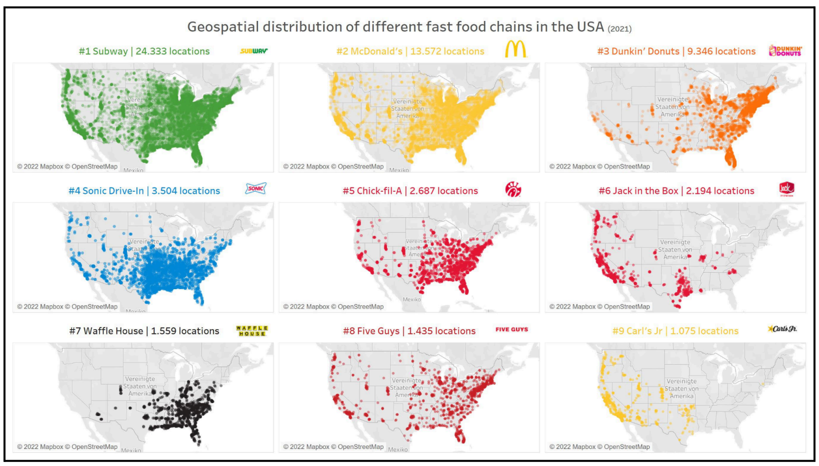


Fig. 27: Obesity rates vs. Migration for Arizona

---

[3] Liao, Z.-X., Tan, X.-M., Zhao, Y.-Y., Sun, X.-C. and Yi, F.-L. (2023). Association between internal migration experience and depressive symptoms: analysis of PSID data. *BMC Public Health*, [online] 23(1). doi:https://doi.org/10.1186/s12889-023-16073-0.

[4] Lubin, G. and Nudelman, M. (2016). *Biggest food chains in America: maps*. [online] Business Insider. Available at: https://www.businessinsider.com/biggest-food-chains-in-america-maps-2016-11#8-dominos-has-5480-stores-in-the-us-8 [Accessed 24 Mar. 2024].

2.5 Recommendations

Governments can prepare policies that can protect their people from this inclination and hence the outcome of obesity by using our research to understand how likely it is for an increase of obesity to occur in their state in the next decade.

We have a few suggestions on what governments can do:

2.5.1. Create Healthy Environment Incentives

Incentivize Urban Planning for Health: Encourage cities to design urban spaces that promote physical activity, such as bike lanes, walking paths, and green spaces. Offer grants or tax incentives to cities that integrate health-promoting features into their infrastructure.

2.5.2. Integrate Fitness Programmes and Rewards

Develop a State-initiated Health App: Create a comprehensive app that provides users with information on local healthy eating options, exercise facilities, and social groups focused on physical activity. Incorporate gamification to encourage participation, such as challenges and rewards for meeting personal health goals. Singapore has an application called Healthy 365 for residents to earn supermarket vouchers and shopping deals[5] when they attend fitness programmes or participate in events like the National Steps Challenge.[6]

2.5.3. Foster Community-based Initiatives

Support Local Food Systems: Fund community gardens and farmers' markets to increase access to fresh produce. Partner with local businesses to offer discounts for healthy food options to residents who participate in community fitness events.

Initiate 'Active Communities' Program: Create a program that supports community-led initiatives to increase physical activity, such as walking clubs, dance classes, or sports leagues. Provide resources and funding to help these initiatives flourish.

---

[5] CNA. (2022). *Singapore residents who enrol in Healthier SG to get S$20 worth of points on Healthy 365 app*. [online] Available at: https://www.channelnewsasia.com/singapore/moh-healthier-sg-healthy-365-app-points-3321451 [Accessed 24 Mar. 2024].

[6] Healthhub.sg. (2023). *nsc*. [online] Available at: https://www.healthhub.sg/programmes/nsc [Accessed 24 Mar. 2024].

2.6 Conclusion

Our key findings:

- Obesity tends to spread from states with higher obesity rates to their neighbouring states.
- For 36% of states in the US, there was a strong and positive correlation between neighbouring migration rates and obesity rates. The extent of this correlation observed varies from state to state.
- A similar correlation between foreign migration rates and obesity rates cannot be established.
- The lag effect is most apparent as an aggregation of the last four years of neighbouring migration data.

Overall, we found that neighbouring migration patterns could be one of the reasons explaining the spread of obesity from “more obese” states to neighbouring states. Moreover, neighbouring migration rates appear to be a strong indicator of obesity rates for a significant number of states, such as Arizona and Connecticut.

3. Appendices

Appendix A: Streamlit dashboard

If you would like to try playing with the data for yourself:

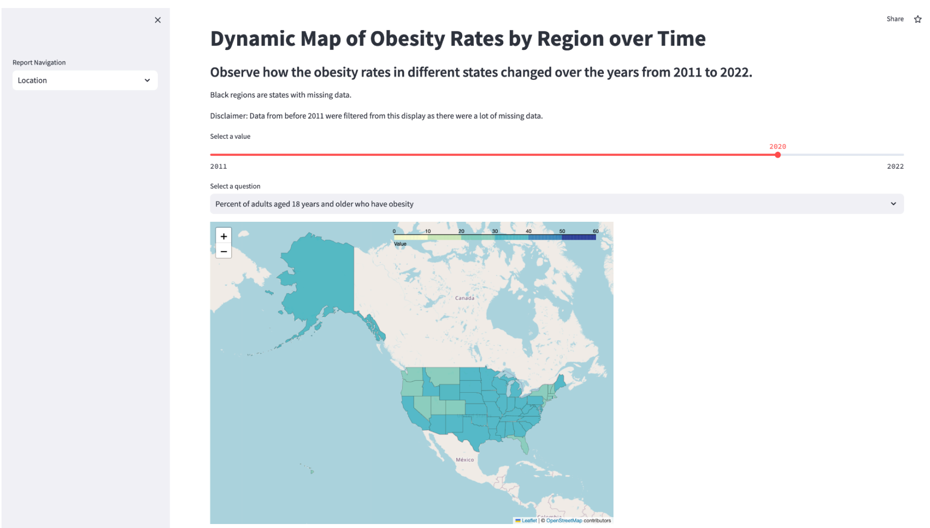
https://2024asiacitadeldatathon-teamkae.streamlit.app/

To see a pattern of obesity "spreading outward", you can perform the following steps.

Step 1: Set the question to "Percent of adults aged 18 years and older who have obesity"

Step 2: Drag timeline incrementally from 2011 to 2022

Step 3: Look as the heat map radiates to neighbouring states!



Appendix B: Links to External Datasets

- GeoJson File: https://raw.githubusercontent.com/python-visualization/folium/master/examples/data/us-states.json
- State-to-state Migration Flows from the US Census Bureau: https://www.census.gov/data/tables/time-series/demo/geographic-mobility/state-to-state-migration.html

Appendix C: Demographic with Substantial Missing Data

The population group for students in grades 9-12 had missing data on their migration patterns. The dataset on migration patterns did not differentiate between adults and school-going children. Hence, they were not analysed.