

Analyzing the growth in consumer spending on Food

DATASCI 200 - Project 2

Team Members:

- Michael Eisenberg , Edwin Fleurant , Wesley Thomas

GitHub Repository

- https://github.com/UC-Berkeley-I-School/DATASCI200_Eisenberg_Fleurant_Thomas

Primary Datasets

- <https://www.ers.usda.gov/data-products/food-expenditure-series/>
 - State food sales, with taxes and tips, for all purchasers
 - State food sales per capita, with taxes and tips, for all purchasers
- Look for a dataset relating to state salaries over the years
 - <https://www.bea.gov/data/income-saving/personal-income-by-state>
 - This dataset contains data for median per capita income for all states and Washington D.C.
- World bank data for the population of the United States
 - <https://data.worldbank.org/indicator/SP.POP.TOTL?locations=US>

Data Structure

There are 1330 observations and 13 variables. The variables are:

- Year
 - Ex: 2022
 - This helps identify the year the data was collected
- State
 - Ex: Georgia
 - This helps identify where the data represents
- Nominal Sales - Calculation without inflation (Millions of US Dollars)
 - FAH
 - Ex: 31,625.34
 - How much money is spent on food that is taken home, such as Groceries
 - FAFH
 - Ex. 32,754.83
 - How much money is spent on food away from home, such as Restaurants
- Total Nominal Food Sales
 - Ex. 64,380.17
 - This is the total amount for the state over a year

- Constant Dollar Sales(1988 = 100)(Million of US Dollars) - Calculation with inflation
 - FAH
 - Ex: 14,368.32
 - How much money is spent on food that is taken home, such as Groceries
 - FAFH
 - Ex. 13,148.96
 - How much money is spent on food away from home, such as Restaurants
- Total Constant Dollar food sales
 - This is the total amount for the state over a year
- Per Capita Sales of Food
 - Amount spent per person on food by state
 - Nominal
 - FAH
 - FAFH
 - Constant dollar sales per capita
 - FAH
 - FAFH

Introduction

Across the United States, grocery bills are rising steadily, driven by a growing population and a complex web of economic factors. This trend is particularly pronounced in our bustling metropolises, where larger populations translate to a higher demand for food. While this surge in spending might seem positive for the economy, a closer look reveals a troubling consequence: the rising cost of food is increasingly outpacing the ability of low-income earners to afford necessities.

Some Questions we are looking to explore:

1. How does each state's nominal and constant dollar sales compare to that at the national level?
2. What states have the highest and lowest food sales?
 - a. Which states have the largest increase in food sales between 1997 and 2022?
3. How have salaries and food prices increased over time?
 - a. Is there a relationship between the two variables? If so, what is this relationship?
 - b. Do some states more closely correlate food sales and salaries than others?
 - c. Which states spend the most of their income on food?
4. What is the per capita food sales for each state?
 - a. How is it different between states and the national level?
 - b. Do per capita food sales increase or decrease based on major economic or world events, such as the 2008 recession?

Data Cleaning and Sanity Checking

To successfully analyze the data, we must ensure the data contains the expected information and that we verify its validity. We will limit the sanity check just to the variables which will be analyzed. In the primary dataset, the state sales per capita, we will consider all columns in the data set to evaluate both how food spending has increased with and without consideration of inflation. By keeping the entire dataset, as well as year and state, we can easily compare across states and determine whether some have a different rate of spending change over the past 26 years, from 1997 to 2022.

Before diving into the variables, a quick note about the data. The dataset contains measurements in terms of two types of sales:

- Nominal sales: unadjusted value of sales by that year's currency
- Constant sales: value of sales in 1988 US dollars

The purpose of constant sales measurements is to evaluate the data without having to adjust for inflation, which can help determine if consumer spending has truly increased in the past 26 years, just the price of food has increased, or a combination of both. For each of the nominal and constant sales, there are three measurements:

- FAH: spending on food at home
- FAFH: spending on food away from home
- Total: sum of FAH and FAFH sales

In the state sales per capita dataset, we have the following variables. The column names for many of these variables are verbose, so they are shortened to an acronym to decrease the verbosity and reduce the probability of error when attempting to access values in different columns. Variables whose corresponding column names were shortened are identified as such along with the new column name.

- Year: Year for which the food spending is recorded
- State: State for which the food spending is recorded
- FAH sales per capita nominal U.S. dollars with taxes and tips
 - shortened to FAHSPCN (FAH Sales Per Capita Nominal)
- FAFH sales per capita nominal U.S. dollars with taxes and tips
 - shortened to FAFHSPCN (FAFH Sales Per Capita Nominal)
- Total sales per capita nominal U.S. dollars with taxes and tips
 - shortened to TSPCN (Total Sales Per Capita Nominal)
- FAH sales per capita constant 1988 U.S. dollars with taxes and tips
 - shortened to FAHSPCC (FAH Sales Per Capita Constant)
- FAFH sales per capita constant 1988 U.S. dollars with taxes and tips
 - shortened to FAFHSPCC (FAFH Sales Per Capita Constant)
- Total sales per capita constant 1988 U.S. dollars with taxes and tips
 - shortened to TSPCC (Total Sales Per Capita Constant)

We now must evaluate the dataset to determine if there are any issues with the data and if any cleaning is required. Looking at the dataset, we must reduce the verbosity of the column names. We can also see that the columns representing sales data have an object datatype.

```

RangeIndex: 1326 entries, 0 to 1325
Data columns (total 8 columns):
#   Column                                                                 Non-Null Count  Dtype
---  -
0   Year                                                                    1326 non-null   int64
1   State                                                                    1326 non-null   object
2   FAFH sales per capita nominal U.S. dollars with taxes and tips          1326 non-null   object
3   FAFH sales per capita nominal U.S. dollars with taxes and tips          1326 non-null   object
4   Total sales per capita nominal U.S. dollars with taxes and tips         1326 non-null   object
5   FAFH sales per capita constant 1988 U.S. dollars with taxes and tips     1326 non-null   object
6   FAFH sales per capita constant 1988 U.S. dollars with taxes and tips     1326 non-null   object
7   Total sales per capita constant 1988 U.S. dollars with taxes and tips    1326 non-null   object
dtypes: int64(1), object(7)

```

However, given expectations of the datatype of the data and as we can see below, the sales columns should be represented as floats to accurately represent the data types and make operations and chart building easier.

	Year	State	FAFH sales per capita nominal U.S. dollars with taxes and tips	FAFH sales per capita nominal U.S. dollars with taxes and tips	Total sales per capita nominal U.S. dollars with taxes and tips	FAFH sales per capita constant 1988 U.S. dollars with taxes and tips	FAFH sales per capita constant 1988 U.S. dollars with taxes and tips	Total sales per capita constant 1988 U.S. dollars with taxes and tips
0	1997	Alabama	1,340.03	802.19	2,142.22	1,005.24	612.08	1,617.32
1	1998	Alabama	1,393.75	882.8	2,276.55	1,030.28	656.3	1,686.58
2	1999	Alabama	1,466.51	938.48	2,404.99	1,067.14	681.04	1,748.18
3	2000	Alabama	1,516.64	977.59	2,494.23	1,074.75	693.7	1,768.44
4	2001	Alabama	1,543.35	1030.54	2,573.89	1,063.28	712.16	1,775.44

While the state sales per capita data set contains useful information regarding state sales, it is difficult to use the data for consideration of national sales, as we would need the population for each state in each year from 1997 to 2022. Instead, two more data sets are required. The first is a state sales dataset representing the same data as the state sales per capita but is the aggregate sales, not per capita. The second is a data set representing the US population from each year 1997 to 2022.

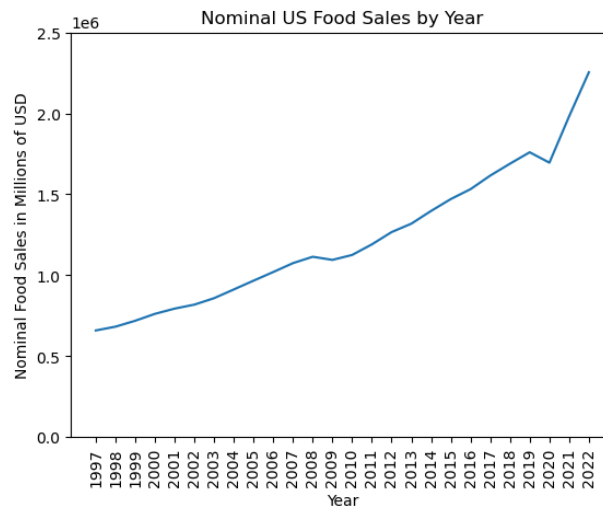
The aggregate state sales data contains the same type of data, so it can be cleaned in the same manner as the per capita state sales data.

The population data, from the World Bank, contains population data for every country, so all countries except the United States must be removed. We are also removing population data for every year before 1997 since the earliest year of sales data is 1997 as well as the column representing 2023 population since sales data ends at year 2022 as well as other columns which do not contain any useful information.

Lastly, we must gather data regarding per capita median income by state and at the national level. This data, from the Bureau of Economic Analysis, which is part of the US Chamber of Commerce, contains per capita nominal median income for every year from 1929 until 2022.

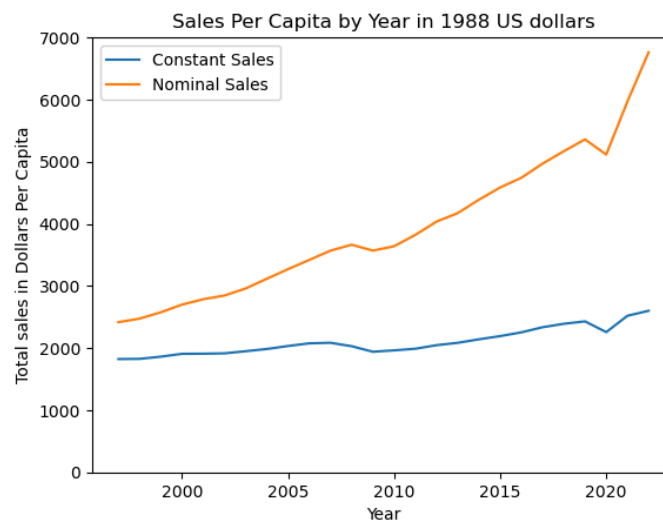
Initial Exploration

After cleaning and validating the data, we begin the initial exploration. First, we consider how nominal sales have increased over the years to gain a surface-level understanding of how food spending has changed.

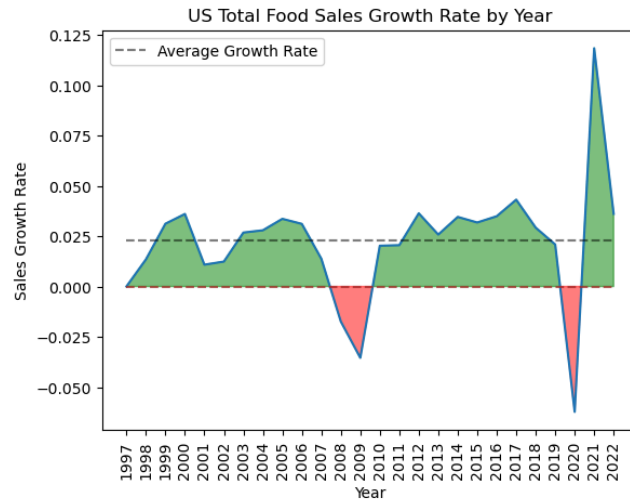


With this, you can see the growth over the years and how it has more than tripled in the past 25 years.

We then take the aggregate and per capita nominal sales in each year. Using the constant and nominal aggregate sales and the World Bank population count for the United States, we can calculate the national per capita sales by year in 1988 US dollars.



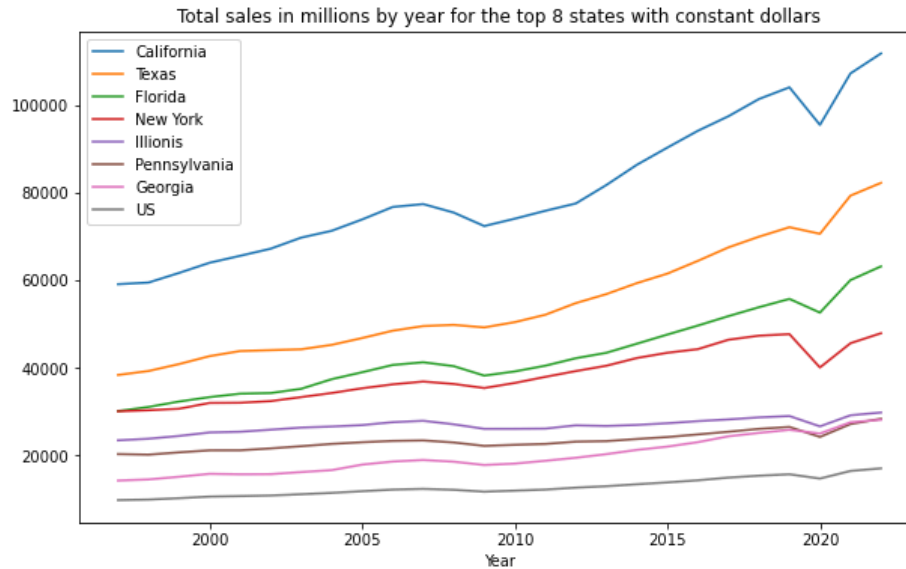
From the above graph, we can see that constant sales have increased in the past 26 years. However, while this indicates people are buying more food and would be spending more money on food regardless of inflation, the rate of increase in nominal food sales per capita is much quicker and appears to be increasing at superlinear rates. The nominal food sales increase is not surprising given inflation typically increases over time and the only drops in food spending coincide with the 2008 recession and start of the COVID-19 pandemic in 2020.



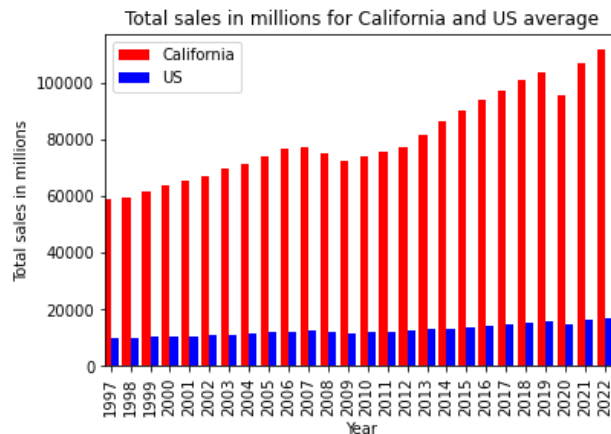
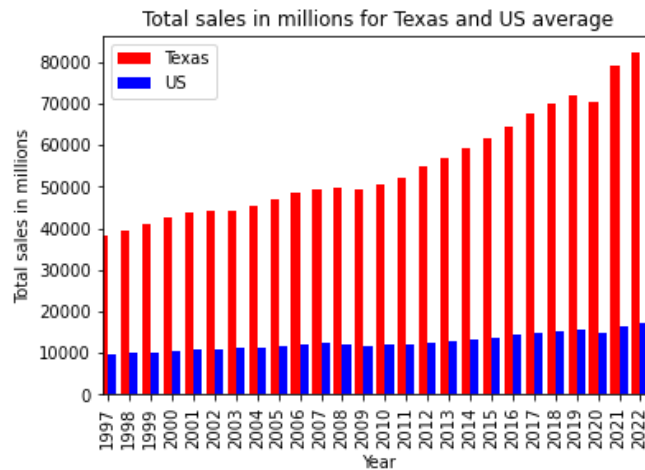
The above graph further highlights the consistent growth rate of US food sales, with only three of the 26 years ending with a smaller amount of food sales than the previous year, following the trend of food sales increasing year over year with the slight dips resulting from negative economic events. Most years of sales growth hovered between one and four percent, with one outlier year of sales growth percentage of 11.8 percent. That year can be attributed to a post-pandemic return of sales to pre-pandemic expectations.

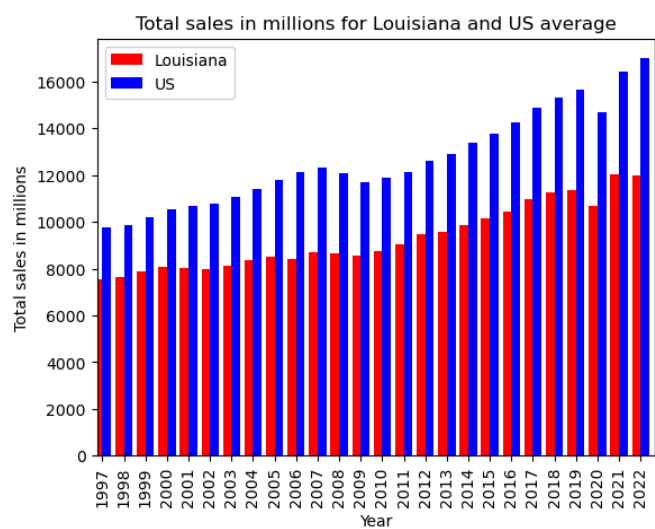
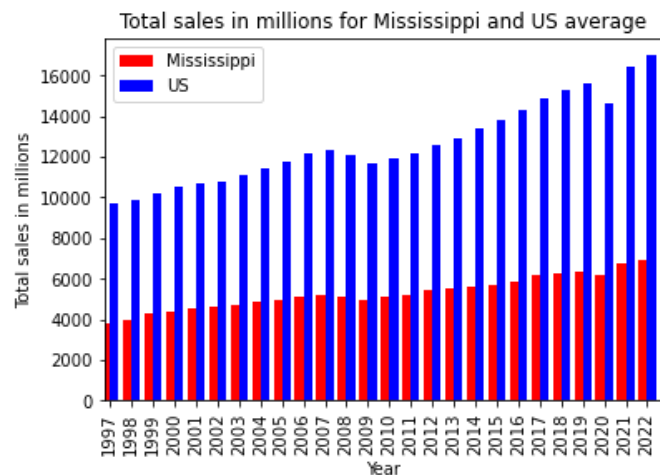
Food Sales at the State and National Level

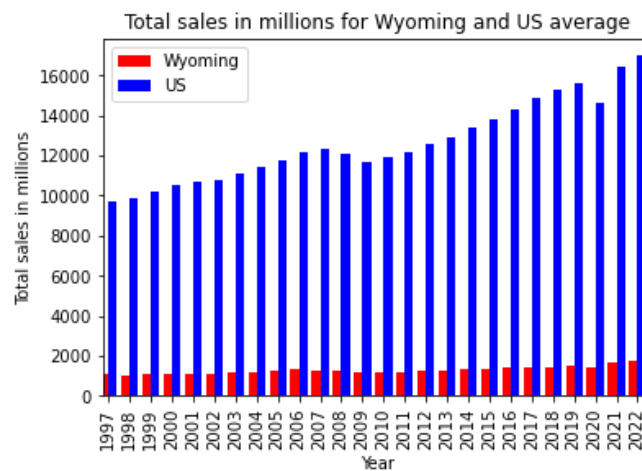
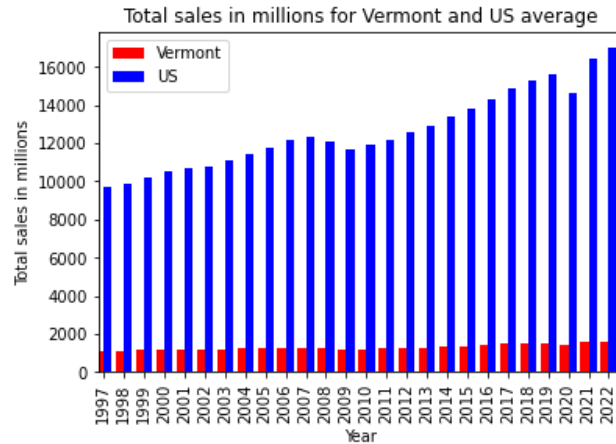
From the graph above detailing the nominal sales increase between 1997 and 2022, we can tell that the nominal food spend has tripled in the past 26 years. Since the top states contribute much more to the total food sales than the rest, the graph below shows the top 7 states in constant dollar food sales with one line representing the average constant dollar food sales by year for the entire country.



After comparing state sales with the national average we see that only a few states have a higher amount of sales than the national average. This can be due to many factors such as the growth of the population, tourism, or even the fluctuation of salary amongst the population. We took a deeper look into the higher, middle, and lower-end states.



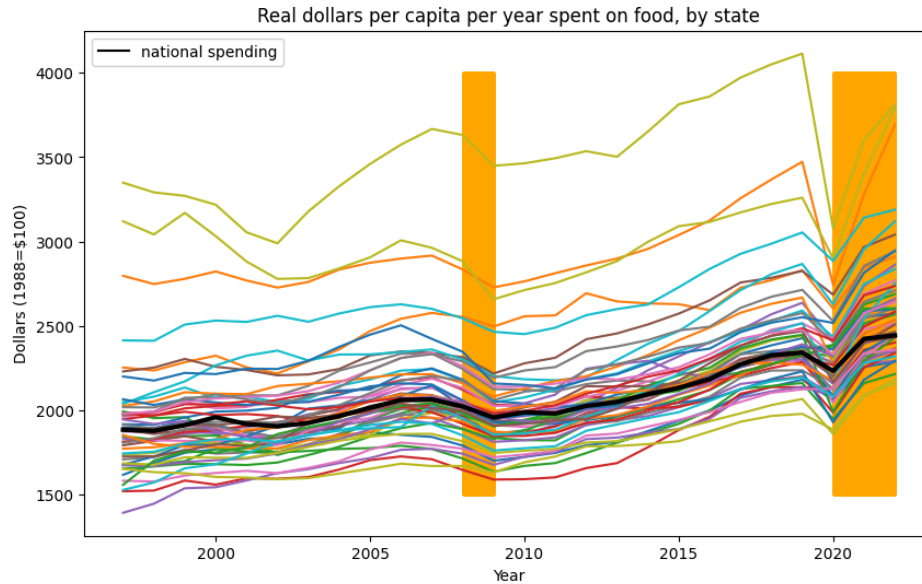




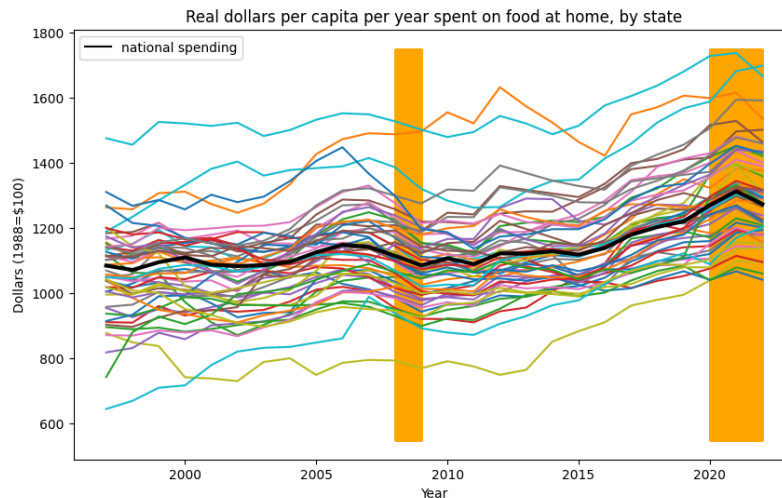
From these charts, when there was an economic event or disaster we saw that states with higher than average food costs reacted similarly or more aggressively while states that have way lower than average spending reacted minimally compared to the national average.

State spending on food per capita

We next wanted to see how people in different states behave differently in their food spending habits. To do this we took data on per capita spending on food at home and plotted each state to get an idea of states behavior as a whole and to be able to see individual differences.

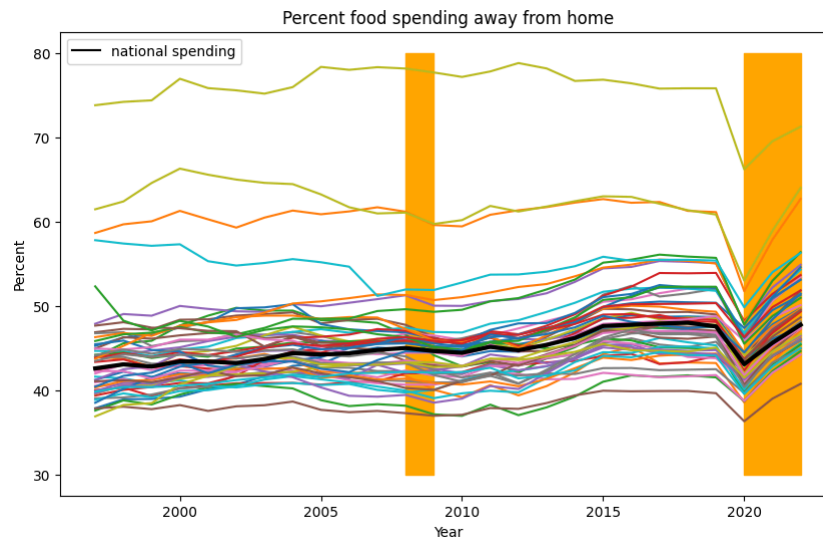


To do this we looked at total spending per person separated by state. We plotted the real per capita expenditure on food at home for each state to get a sense of how food spending has changed throughout the years. We also plotted the national spending to see how each state compares to the national average. One hiccup is we had data on state spending per capita, but only national spending per household. We used a somewhat crude method of dividing household spending by 2.53, the 2022 census estimate of household size. We see that most spend around \$1500-2200 per person in a year, however one group that stands out is the District of Columbia as the highest spender on food, who will show up again later.



To see how food spending changes, see that between 1997 and 2015 spending remained fairly constant and there has been a steady increase between 2015 and 2022 on food relative to other members of the CPI. Some interesting periods of time are 2007-2009 and 2020-2022. The first orange bar highlights the 2008-2009 recession which we believe explains the trend of decreased spending on food. As people had less money available to spend due to

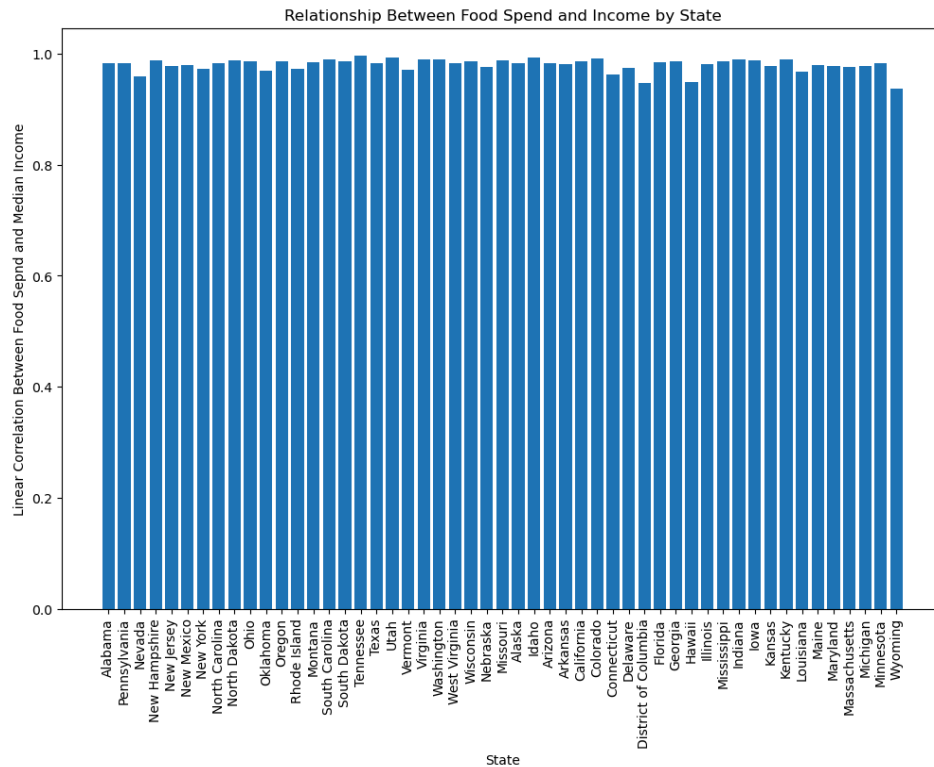
unemployment and foreclosure they spent less on food. An especially interesting trend is the sharp increase and decrease between 2020 and 2022. There was well recorded sharp inflation of food costs during the 2020 pandemic, but why the decrease between 2021 and 2022? We believe this may be explained by higher inflation of other goods relative to food causing the price of food to appear to decrease.



Next we wanted to examine how much of food spending is spent on food away from home. We created a new variable of percent spending by dividing food spend away from home by the total amount spent on food in each state and for the nation as a whole. We can see most states spend about 35-50% food budget eating out, with a steady increase between 1997 and 2019. Interestingly the 2008 recession had no noticeable effect. However during the pandemic we see a sharp decrease in spending on food away from home in 2020 with almost complete recovery by 2020.

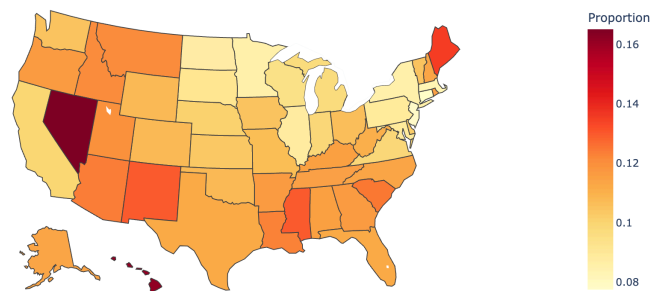
The Relationship Between Food and Salary

To fully understand the impact of rising food prices on the lives of the standard American consumer, we must evaluate how salaries have changed in the last 25 years and if Americans are spending too much of their income on food. People need money to spend on necessities beyond food, such as clothing and housing, and rising food costs can make it challenging to properly allocate payments while leaving money aside for emergencies and retirement.



At the state level, there is a clear positive correlation between median income and food spend. While income and food spend can vary greatly by each person, this indicates that knowing one's income makes it easier to predict their nominal food expenditure.

Proportion of Food Spend to Median Income in 2022



However, residents of some states do spend a much higher proportion of their income on food than in others. In the above graph relating median household income and food expenditure, some states such as New Jersey and Connecticut spent little of their income on food compared to others. On the other hand, food is relatively more expensive in Nevada, the state where residents spent the highest proportion of their income on food than in other states. So, even though food spend and income have risen at similar rates, some areas of the United States have a much more difficult time spending on food and having money left over for other purposes. Regardless, all states are spending at least 8% of their income on food, with many of

them spending over 12 percent of their income on food, an amount that leaves little room for emergencies and leisure after considering other necessities such as rent and clothing.

Conclusion

From the analysis done above, we concluded that consumers have been spending more money on food, and not just due to inflation. Some states, such as California, have historically spent more on food due to their much higher population. However, even with the dips in 2008 and 2020 due to economic disaster events, all states have faced higher spending on food adjusting for inflation. The higher food spend has been somewhat alleviated by the increase in median per capita income at the national and state levels. Despite that however, some states such as Nevada and Hawaii spend a significant portion of their paychecks on food. Food costs are clearly on the rise, even adjusting for inflation, and wages must continue to keep up in order to prevent people from spending what is already a significant portion of their paychecks on food.