Making Macroeconomic Indicators Accessible

CSC-475: Seminar in Computer Science
Fall 2023

Joshua Hutson
joshua.hutson@furman.edu
Department of Computer Science
Furman University
joshua.hutson@furman.edu

Abstract

The paper details the data and methods used to create a dashboard for macroeconomic indicators such as inflation, unemployment, and average wage. As the world becomes more data driven, there is a growing demand for clearly communicated economic data. While dashboards exists for individual macroeconomic indicators, few exist that display multiple. This dashboard aims to convey how economic indicators are impacting the average American by presenting county and state level data for each one. The paper also analyzes the preliminary visualizations present in the dashboard and how nuanced information is conveyed to the audience.

1 Introduction

Americans are more avid consumers of news than one might expect. According to the American Press Institute, approximately 59% of Americans self report looking at news several times a day with 90% reporting that the news was at least moderately important to them [1]. Approximately 40% of Americans report closely following one news story and around 75% of Americans watch or listen to opinion pieces infrequently (less than once a day) [1]. These statistics paint a picture of Americans being interested in news stories that report the facts. Additionally, the Pew Research Institute reports that the majority of Americans prefer their news in digital mediums (52%) as opposed to TV (35%), radio (7%) and print (5%) [2]. If Americans are actively seeking out news and prefer it digitally, steps should be taken to make data and trustworthy statistics more accessible to them. One area of news reporting that is very data driven is the economy. Hence, this project seeks to develop a data dashboard to display economic data in a way that is understandable for the average American.

Dashboards are interactive tools that allow users to explore data specifically and broadly, making them perfect for widely disseminating data. To illustrate that point, the opportunity atlas [1] is a well known dashboard that allows users to visualize economic opportunity and potential upward mobility across the United States. The website displays a map of the United States and then shades different areas of the country different colors depending on the potential for upward mobility, measured in household income, incarceration rate, teenage birth rate, and high school graduation rate, among other indicators. The data can further be split on race, gender, and parent's income for an even more in-depth look at economic mobility. Notably, the opportunity atlas communicates that the potential for economic mobility varies depending on where someone lives. In other words, it displays regional differences that tell a richer story than if the data were aggregated at the national level. Using the opportunity atlas as inspiration, this dashboard will explore inflation trends, unemployment rates, employment numbers, and wages, splitting the data by geographic location, product (for inflation), and job sector (for employment and wage) to tell a more complete and relevant story for each indicator.

All of the data for this dashboard comes from the Bureau of Labor Statistics [2]. As a division of the United States government dedicated to data collection, the Bureau of Labor Statistics has wide

[1] `https://www.opportunityatlas.org`
[2] `https://www.bls.gov/`

reaching, robust, and detailed data for the proposed economic indicators. They have the tools to collect large amounts of data everywhere in the United States, updating several statistics on a monthly basis. The data can be split on almost anything, including but not limited to job market, product, region, state, county, specific job, and union or non-unionized work, depending on the statistic. The size and splitting capabilities of the data makes it ideal for a dashboard, which will be discussed further in the Data Section

The methods section discusses the pipeline of data collection, cleaning, and display. It describes the code structure used to download data tables from the Bureau of Labor Statistic's API and create the Bokeh visualizations uploaded to the dashboard. This section also notes changes to the methodology as the project progressed.

Finally, the results section explores how the dashboard presents granular data to tell a richer story about the macroeconomic indicators. It will explore two different visualizations and argue that the granularity presents relevant and rich information to users, which will help them understand the economic factors at play in their lives.

2 Related Work

At its core, data visualization is about making data accessible to interested people, and there is much literature investigating how people perceive and interact with its various elements. For example, Huynh, Ahmad, and Chvalier [3] conducted a study on how non-data analysts perceive color in visualization, focusing mainly on whether participants follow data visualization conventions, such as darker shading indicating more and lighter shading indicating less. They concluded that novices use colors that "encourage readability and interpretability, but [other] competing factors [may] interplay with the selection of perceptually-motivated color palettes, with semantic associations being particularly noteworthy." [3]. Even with some mixed results, the researchers still suggest adhering to common data visualization color practices.

Similarly, the problem of missing data plagues data scientists, which prompted Fu, Saket, Song and Stasko to study how the representation of missing data impacts people's decision making processes when conducting exploratory data analysis [4]. They gave two groups of participants the same data set to explore with one key difference; one did not visualize missing data and the other did. They found that visualizing missing data "assists participants to reason about the data and [change] their decision making process." [4]. Participants were also more confident in their conclusions when they knew about the presence of missing data. These findings are convincing evidence to visualize missing data in some way to gain the audience's trust.

Interpretability is also an important part of visualization accessibility. To this end, Tufte recommends a very minimalist approach, wanting to draw a viewer's attention to meaningful insights in the data and not to other distracting elements, which may misrepresent the data [5]. However, the current view of graphical embellishment varies from Tufte's. Parsons and Shukla, using quotes from their survey of data visualizers, describe a movement towards the acceptance of embellishments in visualization [6].They argue that embellishment can help the audience achieve a greater understanding of the data and improve recall. Knaflic presents a middle ground, saying that embellishments can be used if they are purposeful and add to the data's story [7]. Her philosophy of data visualization is similar to the one adhered to in this project, as each visualization is constructed to tell a story. Since the project is comprised of dynamic data visualizations, it will benefit less from embellishments than a static visualization would.

While the aforementioned concerns are valid aspects of data visualization, they do not holistically measure the accessability and impact of visualizations. As the discussion of embellishment [6] highlighted, not everyone agrees on what makes a visual good. However, that did not stop some researchers from trying to develop a heuristic evaluation of visualizations [8]. Although they do not explicitly state what questions they ask to evaluate visualizations, they discuss the main categories that their heuristics fall into. They expand upon Tufte's belief that graphs need to clearly communicate data insight [5] by adding that data visuals should provoke insights, minimize the time needed to answer questions, convey the main point of the data, and generate confidence in the data [8].

Taking into account all the evaluations and studies, a desirable approach for displaying macroeconomic trends is to create a data dashboard using Bureau of Labor Statistics data. In studying dashboards, Sarikaya, Correll, Bartman, Tori, and Fisher distinguish between "the visual genre of dashboards ..., structured as a tiled layout of simple charts and/or large numbers, and the functional genre, an interactive display that enables real time monitoring of dynamically updating data" [9]. The dashboard outlined here falls into the functional genre, even though the visuals can only be updated when the Bureau of Labor Statistics updates their data rather than in real time. These same researchers go on to describe different dashboard structures, and their discussion of each dashboard mimics Knaflic's discussion of basic visualizations. Both sources emphasize the audience of the visualizations and their functional affordances being the driver behind the design [9][7]. In the second half of their paper, Sarikaya, Correll, Bartman, Tori, and Fisher provide dashboard examples for various purposes and audiences, which provide useful templates for this project [9].

Additionally, economic literature indicates a shift towards the integration of big data in economics. In their book, Abraham, Jarmin, Moyer, and Shapiro assert that "users [of economic statistics] are demanding more timely and granular data" leading to "interest in alternative sources... to better address users' demand for information." [10]. With an increases in demand for new data, more demand for new understandable data visualization will follow. In fact, some economists have already begun developing granular data visualizations to tell stories, a major example in recent years being the opportunity atlas [11]. Using just Census data, Chetty, Friedman, Hendren, Jones, and Porter were able create an incredibly granular visualization of economic mobility across the United States [12] to supplement their research, which indicated that "neighborhoods have substantial causal effects on children's long-term outcomes at a granular level." [11]. These results motivate the choice to provide the most granular data possible for a more complete and insightful story. The Bureau of Labor Statistics has also experimented with some dashboards, which is further proof of the demand for insightful visualizations of their data. Their most popular dashboard was the International comparison of annual labor force statistics, which, true to its name, compared unemployment, GDP, productivity, and compensation across various countries [13]. Unfortunately, this dashboard was only available in Excel and was discontinued in 2013. The Bureau has recently added visualizations back into their website, but they are limited to just unemployment. This dashboard will supplement the current work by including inflation, wage, and employment data, where applicable.

3 Data

As previously mentioned, all the data for this dashboard came from the Bureau of Labor Statistics. Specifically. the employment and wage data came from the Quarterly Census of Employment and Wage (ENU) [3], the inflation data came from the Consumer Price Index for All Urban Consumers (CPI-U) [4], and the unemployment data came from the Local Area Unemployment Statistics Survey (LAUS) [5]. Each of these data sets follows a similar structure. Every instance is identifiable by the series ID, year, and month and reports the requested value, either a CPI value for inflation, a percentage for unemployment, or number of people for employment. After being downloaded, the data was cleaned and arranged.

In each of the data sets, the date, location, and report specific fields were not present as distinct columns, but the information was encoded in the downloaded data. The date field was created using the year and month columns to initialize a DateTime object. Information for the remaining fields was present in every series ID. For example, ENU010011051012 is a series ID for an employment data set containing information about the county, data type, and industry. The three letters at the start of the code (ENU) indicate the survey, in this case the Quarterly Census of Employment and Wage. The next five numbers (01001) encode the county, the following three (105) encode the type of data, which in this case is wage data for the private sector, and the last four (1012) encode the industry, in this case construction. The series IDs for the inflation and unemployment data sets follow a similar structure.

[3] `https://www.bls.gov/cew/data.htm`
[4] `https://www.bls.gov/cpi/data.htm`
[5] `https://www.bls.gov/lau/data.htm`

3.1 CPI data (CPI-U)

Year	Month	Value	Date	Location	Products
2019-2023	Month of recorded observation	CPI value	Dates ranging from January 1st, 2019 to September 1st, 2023	Region	Cereal/Wheat Meat Fruit and Vegetable Non-alcoholic Beverages Dairy Housing Eating Out Household Energy Clothes/Apparel New Cars Used Cars Gasoline Medical Commodities Medical Services Tuition

Table 1: CPI-U Data Descriptions including a list of all product categories represented

As given by Table 1, the inflation data set includes years from 2019 to 2023, including every month in each of those years. Some products have data going further back, but for the purposes of standardization, every product is presented over the same time frame. The macro-economic indicator for the inflation data set is CPI value, the most commonly used statistic to track inflation. The Bureau of Labor Statistics only keeps track of CPI at the regional level, meaning the inflation data set's location field only contains four values: Northeast, South, Midwest, and West. Figure 1 displays the Bureau's definition for these regions. The only other field in the data set is product, whose values are given in Table 1. The visualized data set contains $13,677$ instances.

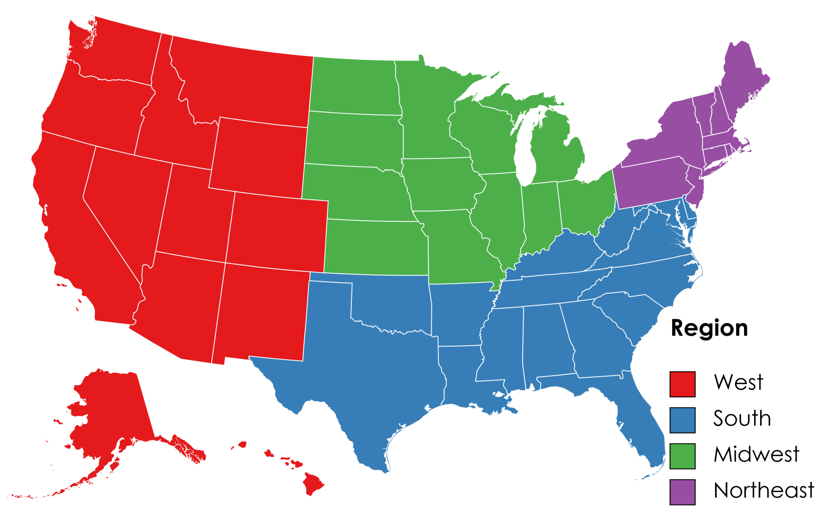


Figure 1: Regional breakdown for the CPI data set as defined by the Bureau of Labor Statistics

3.2 Unemployment Data (LAUS)

Year	Month	Value	Date	Location	Other Fields
2005-2023	Month of recorded observation	Unemployment rate and Labor force participation	Dates ranging from January 1^{st}, 2005 to September 1^{st}, 2023	County and State	None

Table 2: LAUS unemployment data descriptions

The unemployment data set includes years from 2005 to 2023 with every month in each of those years. Unemployment rate and labor force participation rate are the two macroeconomic indicators, as displayed in Table 2. The Bureau of Labor Statistics gives data at the county and state level, so both the county and state rates were downloaded. No other fields apply to the unemployment data set. If someone is unemployed, they are not unemployed in a particular industry, like construction or mining. Since someone who is unemployed can move between industries, it makes no sense to categorize unemployed people with any other fields. The visualized data set contains $23,226$ instances.

3.3 Employment Data (ENU)

Year	Month	Value	Date	Location	Industries
2005-2023	Month of recorded observation	Employment numbers and, Weekly Wage	Dates ranging from January 1^{st}, 2005 to September 1^{st}, 2023	County, State	Natural Resources and Mining Manufacturing Professional and Business Services Information Financial Activities Education and Health Services Trade Transportation and Utilities Leisure and Hospitality Construction Other Services Unclassified

Table 3: ENU employment data descriptions including a list of all industries represented

Like the unemployment data set, the employment data spans from 2005 to 2023 and includes two indicators, employment, measured in number of people employed, and average weekly wage, shown in Table 3. Once again, data was obtained at the county and state level. Unlike unemployment, this data set includes an industry field, whose values are given in Table 3. The visualized data set contains $164,924$ instances.

The unemployment and employment data sets were granular to the county level, but visualization complications, which will be discussed in the methods section, limited the spatial resolution of the dashboard. Instead of being granular to the county level, the unemployment and employment pages of the dashboard could only be granular to the state level.

4 Methods

The Bureau of Labor Statistics has a public API for downloading data. Without registering, the Bureau will let any computer make 25 queries of 25 tables per day. A registered computer can make 500 queries with 50 tables per query. To speed up the process, the data was downloaded on a registered computer with an API key. The calls to the API produced JSON tables, which were then converted into pandas [6] data frames with python code. Because downloading data at the county level

[6] `https://pandas.pydata.org/docs/`

produced millions of instances, several small tables were downloaded, one for each county, and later combined. The inflation table was small enough to download as one table, but the unemployment and employment tables had to be retrieved in small chunks. Once downloaded, each survey was cleaned for easier data storage. Months were converted from strings to integers, the date field was created from the year and month fields, and information from the Series ID field was extracted to make location and report specific fields. The clean data was saved as three tables, one for inflation, one for employment, and one for unemployment.

Each of the combined tables were uploaded to Google Colab as a pandas data frame for visualization using the Bokeh [7]package in a python environment. The integer values associated with data type, product, and industry were replaced with informative strings describing each category. County and state codes were also replaced with their associated county and state names. The pandas data frames were converted to Bokeh column data source objects, which were used to create two visualizations, a line graph and a heat map, for each of the indicators. Each visualization was designed to update when a user gave one of the selectors a new input. For example, if a user changed the state in a drop down menu, the dashboard would automatically update to display data for that state. JavaScript code updates the visualizations in a web environment. Each visualization was downloaded as an HTML file, modified for visual appeal, and uploaded to a GitHub repository. The website with the dashboard is `https://penguin-del.github.io/CSC-475_Project/` and the GitHub repository with all of the method code and HTML files is `https://github.com/penguin-del/CSC-475_Project`. Figure 2 visually conveys the structure of the pipeline.

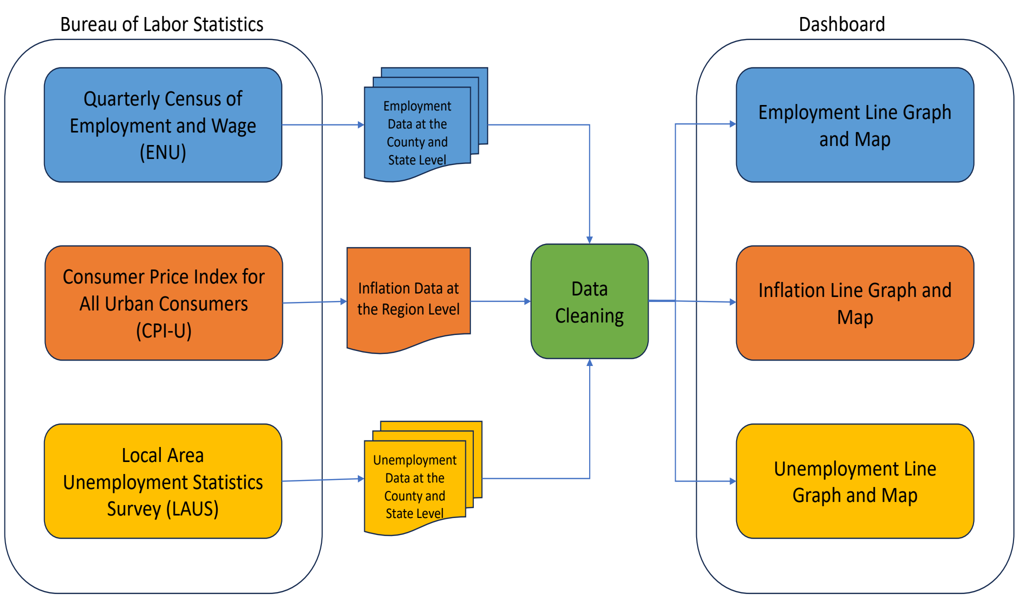


Figure 2: The Data Pipeline for the dashboard. The reports from the Bureau of Labor Statistics were downloaded into smaller tables which were cleaned. Then using Bokeh, the data was visualized and uploaded to the Dashboard website.

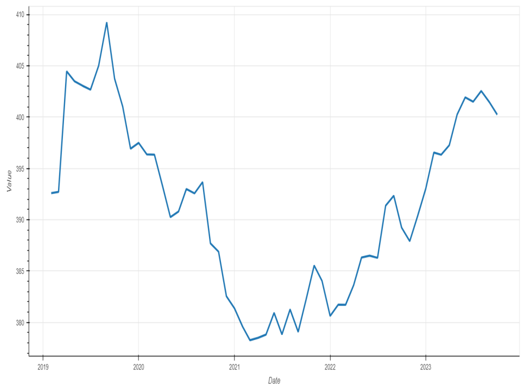
Initially, data was collected at the county level in accordance with the goal of the project: to present the most granular data possible. However, since Google Colab gives limited resources to each of its users and JavaScript code is not designed to handle large amounts of data, the spatial resolution of the dashboard had to be reduced. When given data at the state level, Google Colab was able to handle creating the visualizations.

[7]`https://docs.bokeh.org/en/latest/index.html`

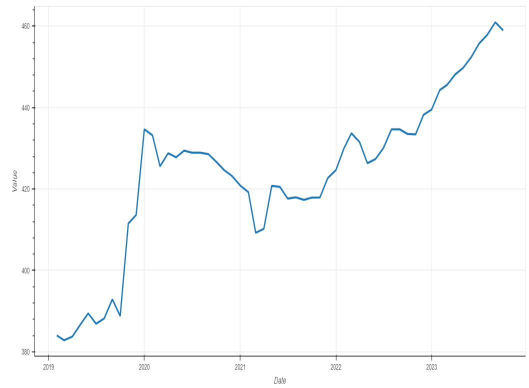
5 Results

Despite having to reduce spatial resolution, the dashboard still communicates nuanced information at regional and state levels. As examples, these section highlight trends from the dashboard that could spark further research.

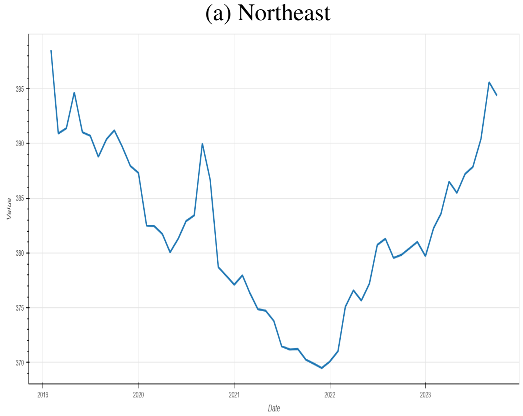
5.1 Temporal Trends



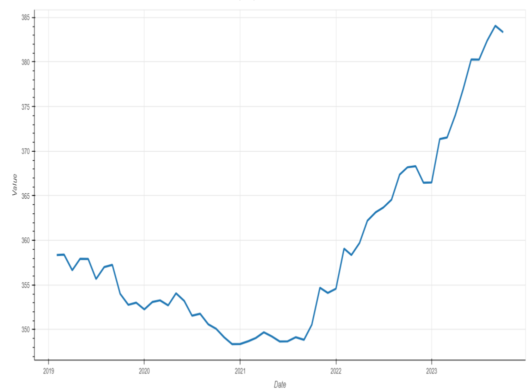
(a) Northeast



(b) West



(c) Midwest



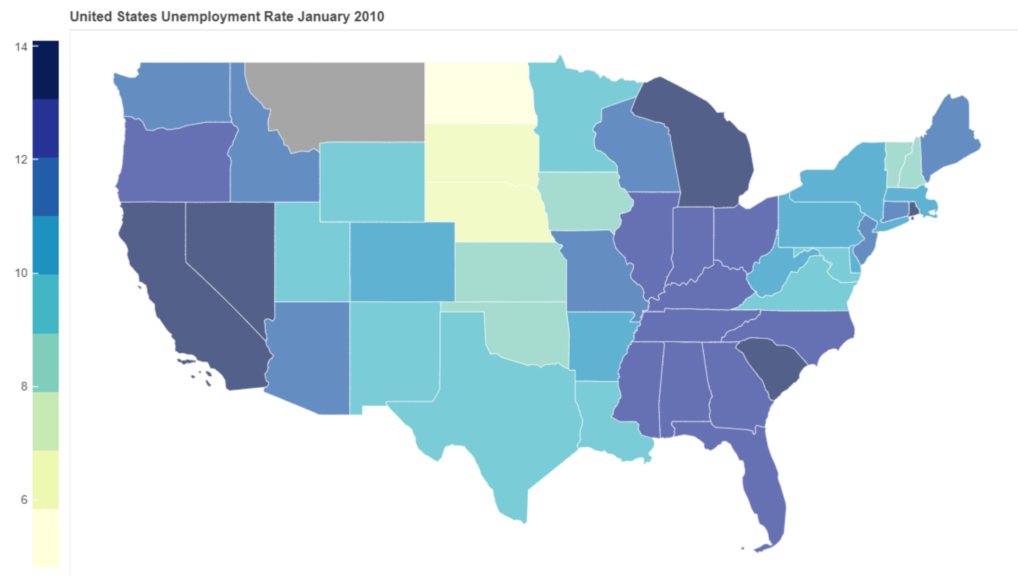
(d) South

Figure 3: CPI values for medical commodities for the four US regions defined by the Bureau of Labor Statistics

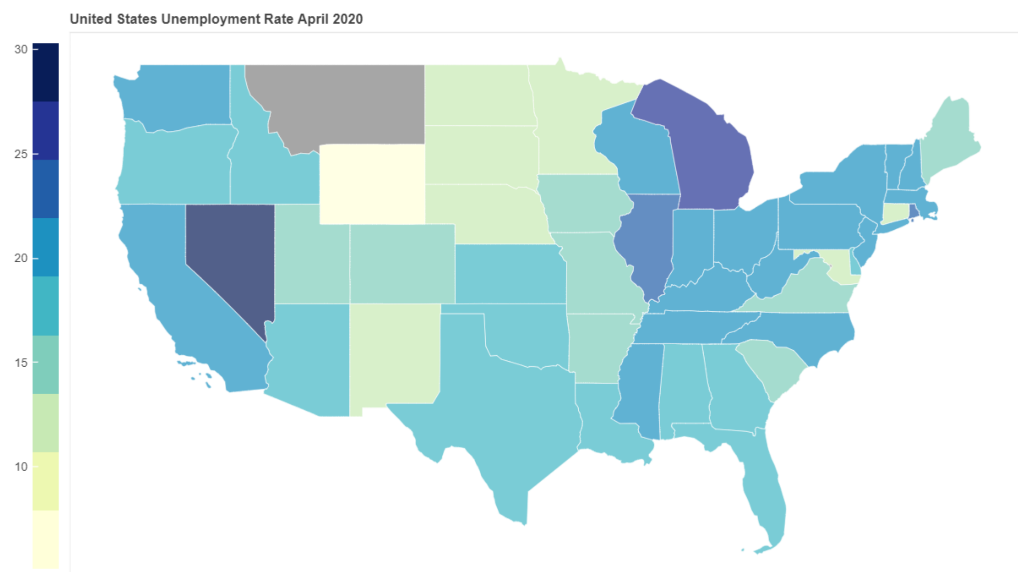
Figure 3 displays four line graphs from the inflation page of the dashboard. Each graph displays CPI values for medical commodities in the four US regions defined in Figure 1. In contrast to all other products in the dashboard, CPI values decreased from 2019 to 2021 in all regions except the West. That means prices for medical commodities deflated during the Covid-19 pandemic. This finding is non-intuitive because one might suggest that a disease would create demand for medicine causing prices to increase. However, that's not the trend. The Northeast and Midwest have CPI spikes in the latter half of 2020, but those spikes are short lived and the trend of deflation continues. Each region exhibits a different trend as well. The Northeast experiences steep inflation for medical commodities in 2019 before prices began deflating. They reach their lowest point in 2021 before inflating again. Even in 2023, the prices did not reach the same level of inflation as fall of 2019. Prices for medical commodities in the Midwest started deflating in 2019 and continued until 2022 with one spike in the fall of 2020. At the beginning of 2022, prices began inflating again, almost to the 2019 level. In stark contrast, the West experienced steep inflation at the end of 2019 and did not see the same deflation trend as the Northeast and Midwest. They experienced a small dip in prices at the beginning of 2021, but otherwise their prices have continued to inflate and their 2023 CPI is much higher than their 2019 CPI value. The South saw a similar trend but did experience slight deflation from 2019 to late 2021. Their 2023 CPI value still exceeded their 2019 CPI value considerably. None of these trends are exactly the same, meaning the dashboard communicates nuanced insight in temporal trends, highlighting regional differences.

5.2 Spatial Trends

The other key component of the dashboard is maps, which display the same data spatially that the line graphs do temporally. The maps can capture relative differences between states in a way that the line graphs cannot. Figure 4 illustrates relative unemployment rates in January 2010, after the 2008 recession hit, and April 2020 following the Covid-19 lock downs.



(a) Unemployment rates for January 2010, the month that employment rates peaked after the 2008 housing crisis



(b) Unemployment rates for April 2020 following the Covid-19 lockdowns

Figure 4: A comparison of relative unemployment rates for the 2008 recession and the Covid-19 lockdown.

The heat map for the Covid unemployment rates looks much more monochromatic than the 2008 recession map. Unemployment rates were more uniform across the United States during Covid, even though in the majority of states, the unemployment rate in April 2020 was higher than the unemployment rate in January 2010. The 2008 recession map displays regions of the US that were hit harder than others. The Southeast, parts of the Midwest, California and Nevada experienced the highest unemployment rates in the country. In 2020, Nevada was still the hardest hit, with a 30.3% unemployment rate, but the other regions did not experience above average increases in unemployment.

6 Future Work

The main limitation of the dashboard is the spatial resolution. Despite aiming to provide visualizations that were granular to the county level, it was difficult to find the computing resources necessary to produce such granular visualizations. The first step in any future work would be finding ways to produce county level visualizations.

The next limitation was time resolution. As previously mentioned, the inflation dashboard was limited to the years in the range 2019 to 2023 because data existed for every product and every month in these years, but some products have data from earlier than 2019. Bokeh had trouble initializing data where several products were not represented in all years. Another future step would be to increase the time resolution to view trends over a longer time interval.

Finally, the dashboard is not automatically updating. Given the sheer number of queries needed to create the data set, it is almost certain that the Bureau of Labor Statistic's API would experience an internal error that would terminate the download. Additionally, the initial data collection process requires an API key which is only good for a year, meaning the API key would need to be renewed and updated every year. These complications hinder automatic updates in the code's current form. One potential route of exploration is restructuring the data download method. Despite these limitations, the dashboard still accomplishes its purpose of communicating nuanced information about macroeconomic indicators to its audience.

References

[1] Media Insight Project. How Americans describe their news consumption behaviors. `https://americanpressinstitute.org/publications/reports/survey-research/americans-news-consumption/`, 2018. Online; accessed 13-Sep-2023.

[2] Elisa Shearer. More than eight-in-ten Americans get news from digital devices. `https://www.pewresearch.org/short-reads/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices/`, 2021. Online; accessed 13-Sep-2023.

[3] E. Huynh J. Ahmad and F. Chevalier. When red means good, bad, or canada: Exploring people's reasoning for choosing color palettes. *2021 IEEE Visualization Conference (VIS)*, pages 56–60, 2021.

[4] B. Saket H. Song, Y. Fu and J. Stasko. Understanding the effects of visualizing missing values on visual data exploration. *2021 IEEE Visualization Conference (VIS)*, pages 161–165, 2021.

[5] Edward R. Tufte. *The Visual Display of Quantitative Information*. Graphics press, 2001.

[6] P Parsons and P. Shukla. Data visualization practitioners' perspectives on chartjunk. *2020 IEEE Visualization Conference (VIS)*, pages 211–215, 2020.

[7] Cole Nussbaumer Knaflic. *Storytelling with Data: a Data Visualization Guide for Business Professionals*. Wiley, 2015.

[8] Emily Wall, Meeshu Agnihotri, Laura Matzen, Kristin Divis, Michael Haass, Alex Endert, and John Stasko. A heuristic approach to value-driven evaluation of visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):491–500, 2019.

[9] A. Sarikaya, M. Correll, L. Bartram, M. Tory, and D. Fisher. What do we talk about when we talk about dashboards? *IEEE Transactions on Visualization and Computer Graphics*, pages 682–692, 2019.

[10] Kathrine G. Abraham, Ronald S. Jarmin, Brian Moyer, and Matthew D. Shapiro, editors. *Big Data for Twenty-First Century Economic Statistics*. University of Chicago Press, 2022.

[11] Raj Chetty, John N. Friedman, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter. The opportunity atlas: Mapping the childhood roots of social mobility. *National Bureau of Economic Research*, 2018.

[12] `https://www.opportunityatlas.org/`. [Online; accessed 13-Sep-2023].

[13] Bureau of Labor Statistics. International comparisons of annual labor force statistics. https://www.bls.gov/ilc/dashboards.htm, June 2013.