

295 Data Wrangling Final Project

Alexander Galvan

2025-05-09

Introduction

In the recent months (as of May 2025), the US stock market has seen substantial sell offs in response to Donald Trump's escalating trade tariffs and rising uncertainty in the oval office. In this period we've also seen US Bond Yields spike alongside a watchful Federal Reserve. With that in mind, we might look to previous events and Fed actions to possibly understand what might lie ahead. The goal of this R notebook is to collect and display data that showcase these macro components. My only intention is to convey this data using conventional R tools (tidyverse and ggplot), and not delve too deeply into the topics of inflation, yields, or stocks. I'll briefly discuss the methodology used in the data wrangling and try to relate the data to the larger economic landscape. We'll examine the Federal Reserve's dual mandate: keeping unemployment and inflation low (~2-3%), as well as Covid-19 and the Ukraine Russia conflict. I'll rely on data from the Bureau of Labor Statistics, various government websites, and 3rd party market sites.

The Stock Market - S&P500

We'll start simple by downloading the latest market data and using `read_csv` to load it in. Cleaning its typing, the data in particular is the S&P500, an index that measures the stock performance of the top 500 companies in the US, and accounts for 80% of the US market cap. 10 years of this market data is shown below.

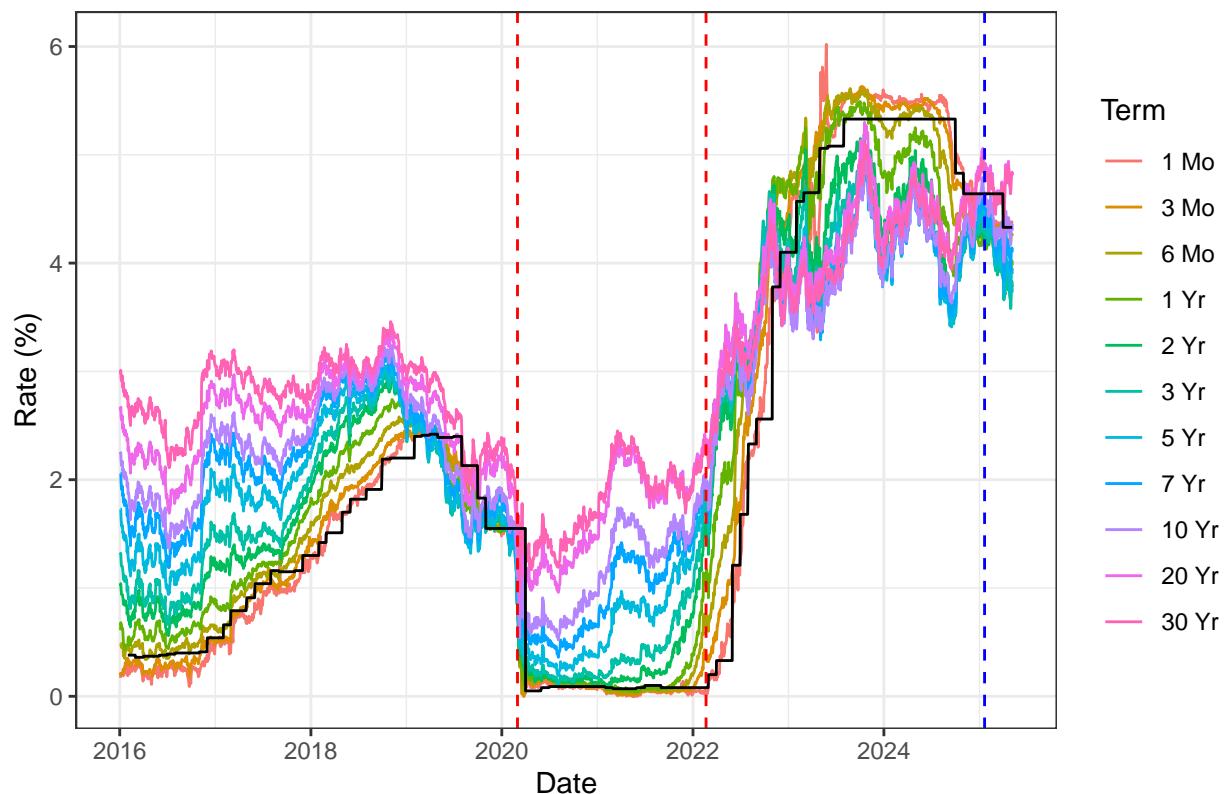


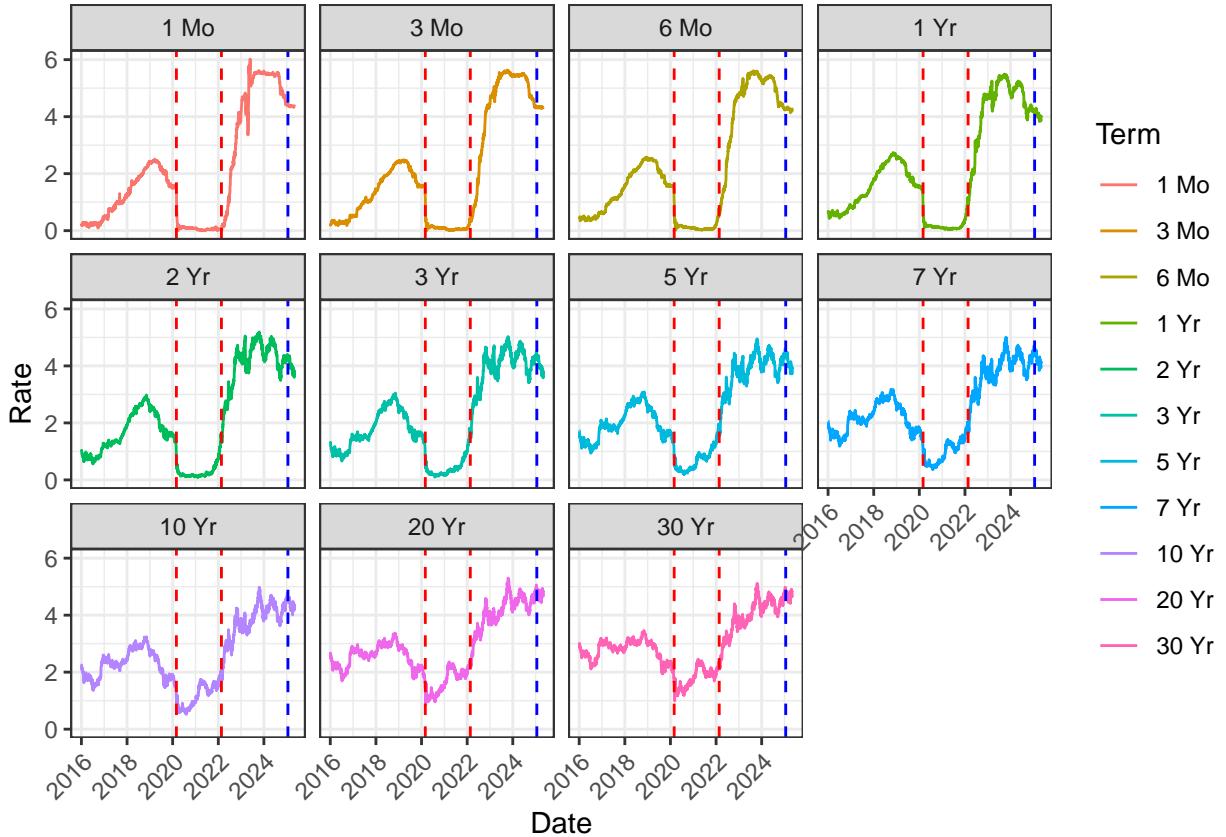
You'll notice three lines marked at February 2020, March 2022, and January 2025, each to mark historic recent events. You might also notice significant variable market pullbacks after each event occurs. Safe to say, we might interpret some connection here between market activity and widespread events. But what I wish to stress the most, is that stocks have always trended upwards, even in the worst of crisis. Whether it be the Federal Reserve's decisions, supply shocks, or trade policy, the stock market, against all odds, has gone up. As we'll see ahead, the Fed took extraordinary measures to combat pandemic slowdowns and red hot inflation using interest rates.

Interest Rates - Fed Funds and Yields

Pulling from the US treasury website, I downloaded yield data and used `read_csv` to load it in. I stitched the files together using `bind_rows`, and pivoted its columns into a longer format so to get individual rows for each term and date. Next, I joined the Fed Funds rate to the treasury rates by Date, and displayed the fed funds rate as a black line over colored yield curves.

Treasury Rates by Term



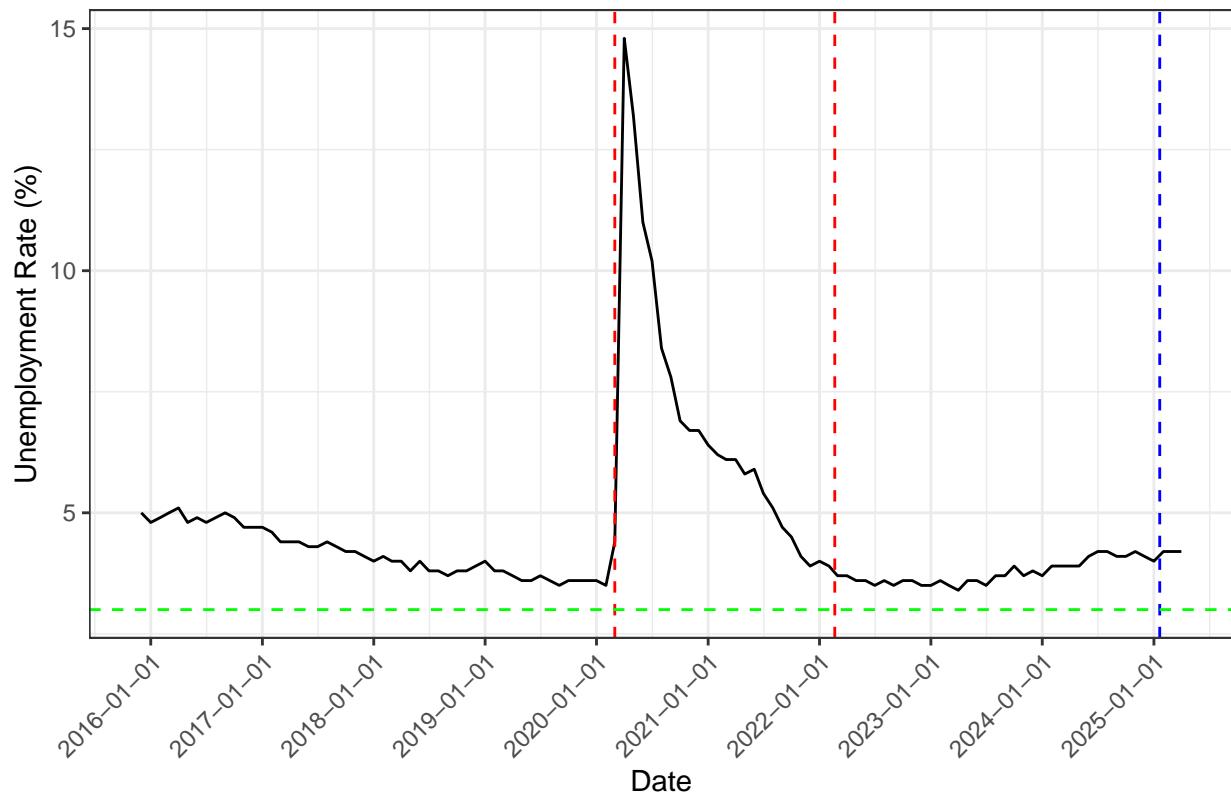


When the pandemic happened, the Federal Reserve knew the implications to the economy when unemployment hit 15%. Immediately, rates were lowered, and measures were taken to keep the economy alive. While the market crashed March 2020, stocks were quick to rebound with renewed investor sentiment in the lower interest environment. Thus, the market soared until February 2022. And just to note the graph, the Fed Funds rate usually defines the baseline yield for the shortest term bonds, while greater terms seem to lag or lead depending on the economic environment.

Unemployment

To showcase unemployment, I requested data from the Bureau of Labor Statistics using their API calls and JSON file unpacking techniques. To do this, I had to install the government's API package to use its call. And to unpack the JSON file, I used the `do.call` and `map` functions to traverse and bind the data. `do.call` was especially useful in this situation as I needed a single dataframe at the end, and not a list of dataframes like `map` returns. Standard cleaning and tidying applied, I added a green line to signify 3% unemployment.

Unemployment Rate over Time

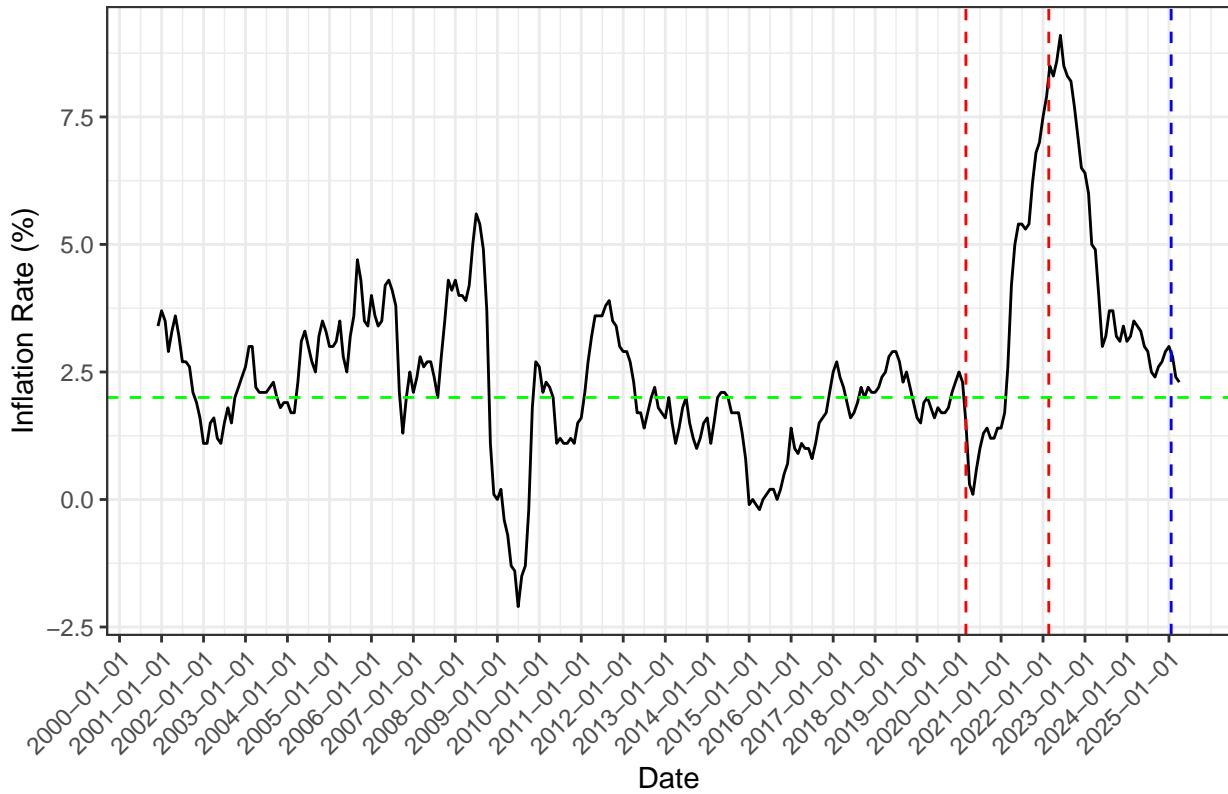


Seeing clearly here, unemployment has been low since the start of the Ukraine war.

Inflation

Scraping the data off a URL website, I cleaned the inflation data by renaming columns, preserving/re-adding incomplete rows, and pivoting it into a longer format to proper display. Cleaning for data types, I added a green horizontal line to show the Federal Reserve's 2% target.

Inflation Rates over Time



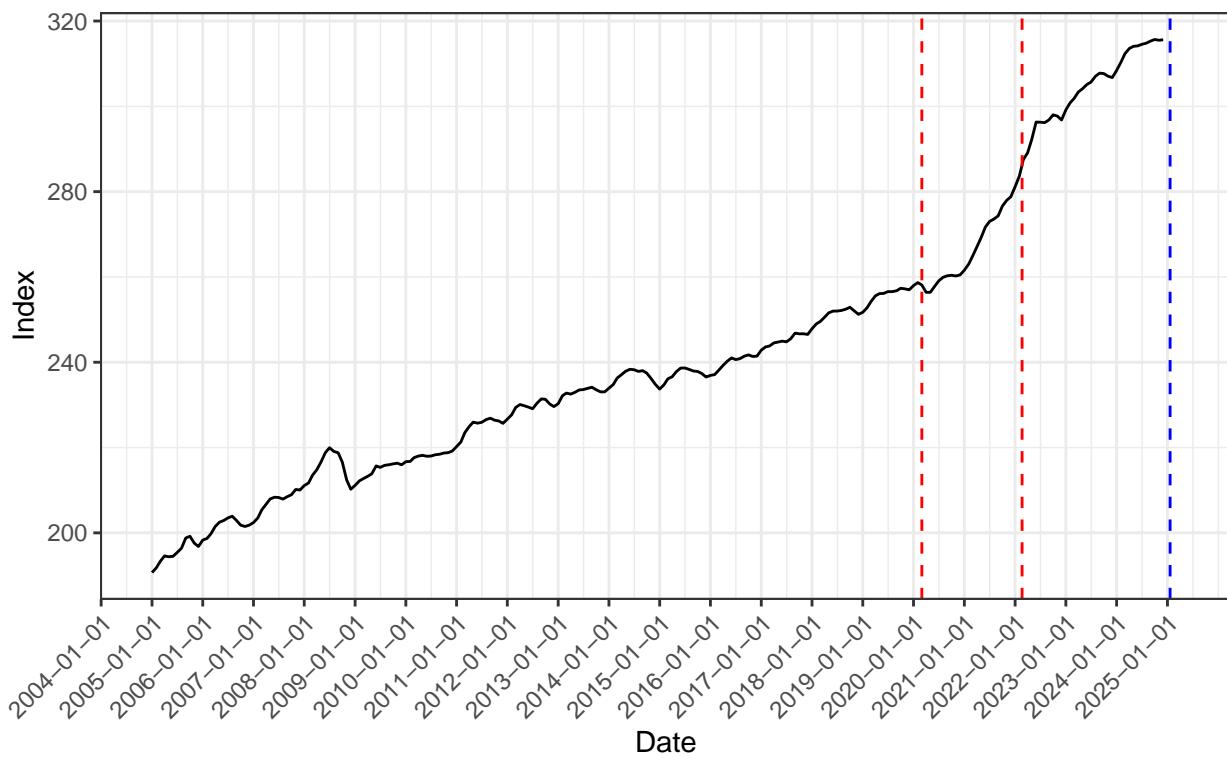
Immediately you can notice that first, inflation is typically kept to its 2% target, and that second, significant events indeed have measurable impact on inflation, most notably the 2008 Financial crisis and Covid-19 pandemic. Once Covid shutdowns began, we saw an immediate deflationary effect in response to the supply shocks, only for it to rise up to 7.5% inflation by June 2022. This, and other macro indicators, ultimately led to the Fed's decision to begin raising interest rates.

CPI Urban and Eggs

When talking about inflation, we have to remember that it represents a change in prices, and has many different ways to be measured. The CPI Urban measures the average prices of all goods (including Food and Energy) from every major US City, without seasonal adjustment. It's generally used to decide economic policy and contract price adjustments. Calling and unpacking it from the BLS' API, along with some basic cleaning and standardizing, I plotted it below.

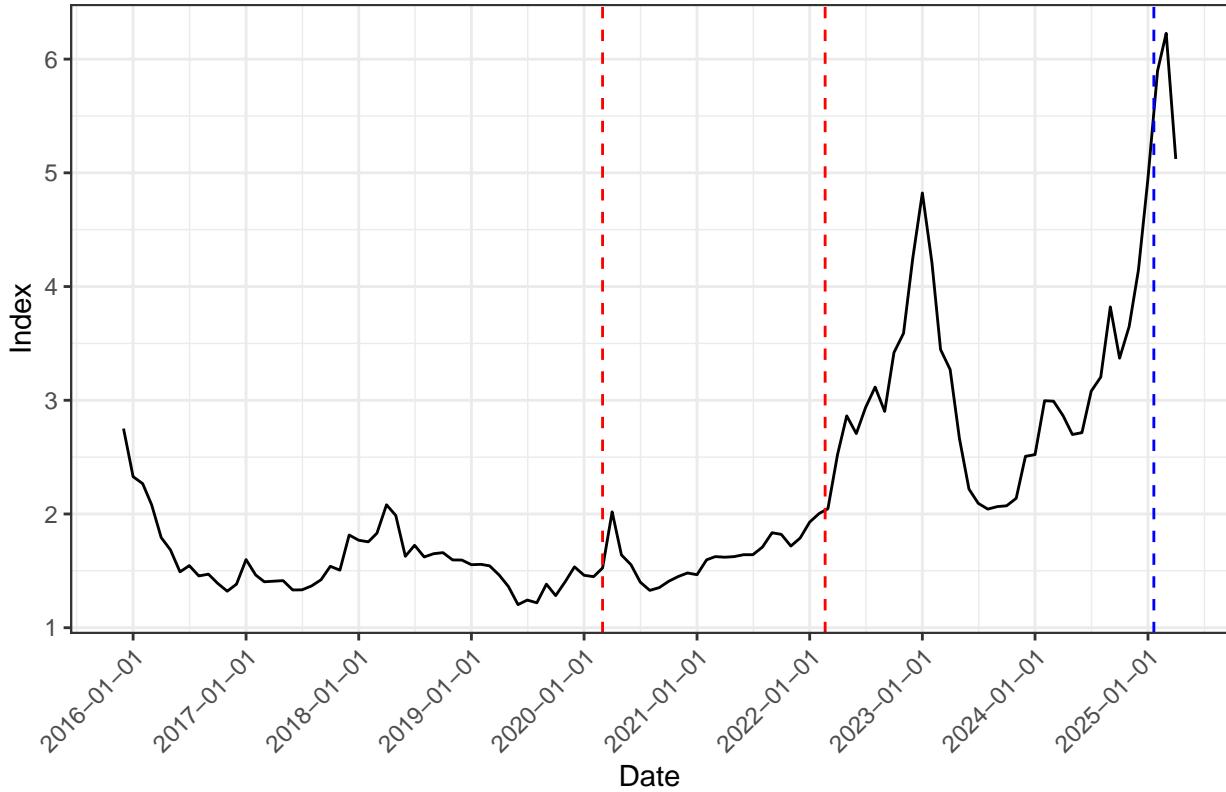
The Consumer Price Index for All Urban Consumers (CPI–U)

Base Period: 1982–84=100



As you can see, everyday prices have steadily risen over the past 20 years, before rapidly accelerating under COVID and the Fed's interest hikes. Also note that even though inflation has cooled since the onset the Ukraine war, prices on average are still at all-time highs.

Eggs CPI over Time

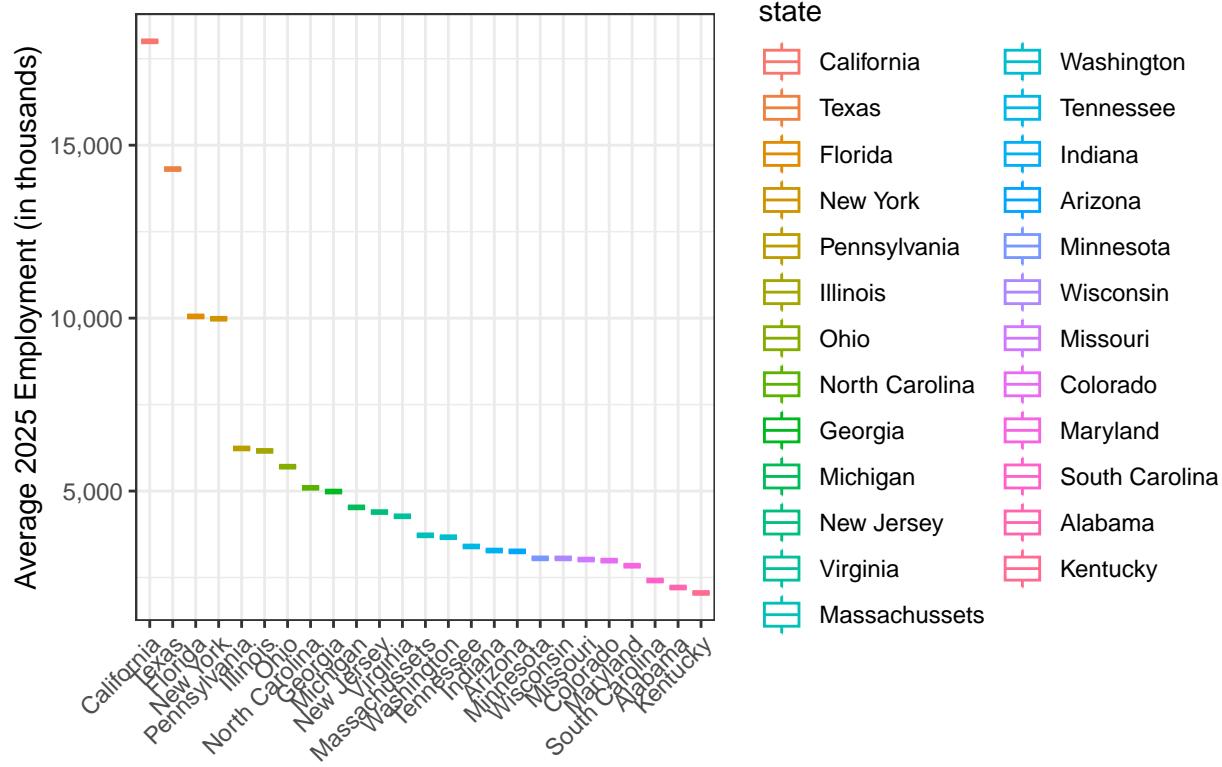


And of course, no talk about prices is complete without looking at eggs. In the same way I wrangled CPIU data, I also called and unpacked the Eggs CPI. Cleaning and tidying it, I show the price index above. As I've outlined in other charts, egg prices reached all-time highs during the start of the Ukraine war and started coming down once the Fed pivoted. Afterwards, prices only briefly cooled momentarily before reaching new all-time highs at the beginning of 2025.

State Employments

And finally, to round out this project, I wanted to show how employment varies from state to state and how they reacted during these past 10 years. We've already seen this data, the added bonus here however is we get to see these update values in real time. As far as the data handling goes, this was the most challenging data set to work with as I needed to unpack multiple series of state ID's from the API call. This made it so one more level of unpacking was required as I needed to traverse through lists of lists. So to do this, I simply added another layer of `do.call map` functions to traverse these inner lists, and labeled each row with a state.

2025 Top 25 States: Average Employment



From the BLS data, we can see that some states are under reported, while others hold the lion's share of US Employment. Unsurprisingly, in 2025, California, Texas, Florida, New York, and Pennsylvania all, on average, hold the most employment out of all US states. If I had to point out one reason why, I would have to say it's because of their sheer population sizes.

Conclusion

Throughout this project, I've touched on some of the leading economic indicators the Federal Reserve uses to guide its decision-making. Examining inflation, unemployment, and interest rates, we saw how they might interact and precipitate the stock market. So, in conclusion, there are many critical indicators we can look at to assess our current economic standing, regardless of the current administration's actions. In any case, through various R techniques, I hope I've done an adequate job of data wrangling and visualization for your ease of understanding!

Appendix:

To make the most out of this project, I had ChatGPT assist me with list traversals and JSON unpacking. Using AI, I was able to ask specific questions about lists, lapply, and do.call, to quickly learning what it was I needed to do. And while useful in providing fast information, I still worked to better my understanding and create solutions to my issues. Aside from that, AI assistance was kept to troubleshooting.

Data Sources:

- https://www.nasdaq.com/market-activity/index/spx/historical?page=1&rows_per_page=10&timeline=y10
- <https://www.usinflationcalculator.com/inflation/current-inflation-rates/>

-https://home.treasury.gov/resource-center/data-chart-center/interest-rates/TextView?type=daily_treasury_yield_curve&field_tdr_date_value=2025 -<https://www.bls.gov/cps/>

-<https://fred.stlouisfed.org/series/FEDFUNDS>