Final Project Part 1

Kristie Kooken

2022-08-03

Introduction

The research question that I am investigating is to better understand if raising minimum wage would cause inflation. In recent years, from my observations, there has been a greater gap between between those who have financial means and those who do not. As a person who has lived in an urban environment for many years, the city where I live has changed with housing costs dramatically increasing while at the same time, campers are lining city streets and there are many tents on city sidewalks housing homeless people. The extreme nature of these changes have been surprising to me.

Last semester, I began to research the topic of raising minimum wage and was very discouraged to learn that since 2009, the federal minimum wage is held at $7.25 an hour. This topic is important because without a living wage that is standardized to current valuation of the dollar, the divide will continue to grow between those that have and those who don't and the middle class will continue to shrink. The hallmark of our society has been the ability to have social mobility which is part of our American concept of a middle class (Fivecoat-Campbell (2015)). Without sustainable wages, it will not be possible for folks on the financial lower end to enable change for their circumstances. The federal minimum wage is closely tied to the ability to enact such change - a stagnant wage means family income stagnation (Mishel (2015)). Further, this stagnation points to a greater shift where policy decisions have been made to benefit corporate owners and managers instead of low and middle class workers (Mishel (2015)). This is a data science problem because there are many data sources but there isn't one that answers this question, thus I will use a variety of sources to show the necessity of increasing minimum wage.

Research Questions

In order to investigate if raising wages would increase inflation, I am investigating the following questions:

1. What is the relationship between hourly compensation and productivity and median income
2. Are there differences in median income to housing prices from past years to this year?

    1. What is the relationship between the two variables?

3. What is the relationship between the top, the lower upper and the bottom earning groups and inflation rates? 1. Do inflation rates increase as the gap between the top, the lower upper and the bottom increases over time?
4. Has the ratio between income and housing change over time?

    1. Does that change move in the same direction as changes in wages for top, middle and low earners?

5. Is there a relationship between home value change, income change and inflation rates?

    1. Do home value change and income change predict inflation rates?

Approach

I plan to address this problem statement by building a dataframe from different sources with different economic markers as well as markers of financial stability (i.e., home buying) and using visualizations to explore trends over time as well as correlation to determine the strength of relationship between different pairs of variables.

Approach addresses the problem?

I plan to use R to run exploratory data analysis, looking distributions by plotting the data and running correlations to determine relationships as appropriate on my individual data sources. Likewise, most of data is by year, I plan to combine my data sources by matching years. My goal is to do this merging in R, however I will explore my success with this, I may resort to using Excel. At least three of my data sources are represented as percentages, and according to an on-line source, I can log-ratio transform these data and use methods such as Pearson's correlation to determine if any relationships exists between the variables. I am planning on creating two simple regression models as outlined above.

Data

For my data source, I am using data that has been subset from a larger source and often from a government site. There has been a level of manipulation done to most of the sources in terms of creating a variable definition, or using a method to determine inflation rates in 1939, as an example, which may different than that from 1976 because a change in how was derived or collection method at the government office level. For this effort, I have the below data sources.

1. ASPUS.xls. This is Federal Reserve Economic Data, July 26, 2022. website: https://fred.stlouisfed.org, from FRED Graph Observations, showing Average Sales Price of Houses Sold for the United States, Dollars, Quarterly, Not Seasonally Adjusted by year.

    1. There are two variables in this data, Year/Quarter (1963-2022) (numeric, continuous) and Average Sales Price of Houses Sold (numeric, continuous)

2. economicpolicy_prod_work.xlsx. Economic Policy Institute from February 20, 2020. website: https://www.epi.org/publication/swa-wages-2019/. EPI analysis of unpublished Total Economy Productivity data from the Bureau of Labor Statistics (BLS) Labor Productivity and Costs program.

    1. There are three variables in this data: Year (numeric, continuous), hourly compensation (as a percentage), and net productivity (as a percentage).

3. inflation_overtime.xlsx. The balance website, from July 27, 2022. website: https://www.thebalance.com/u-s-inflation-rate-history-by-year-and-forecast-3306093. U.S. Inflation Rate History and Forecast.

    1. There are four variables in this data: Year (numeric, continuous), Inflation Rate YOY (percentage), Fed Funds Rate (percentage), Event Affecting Inflation (character)

4. medianincome_tableD1.xlsx. U.S. Census Bureau, Current Population Survey, 1968 through 2021 Annual Social and Economic Supplements (CPS ASEC), SEPTEMBER 2021. website: https://www.census.gov/data/tables/2021/demo/income-poverty/p60-273.html. (Table 1-D from site). Historical Median Income Using Alternative Price Indices: 1967 to 2020

    1. Variables in this source are: Year (numeric, continuous), Current Dollars for variables Estimate, Margin of error and CPI-U-RS/current method for variables Estimate, Margin of error. I am not planning on using the similar estimate variables.

5. homechg_incomechg.xlsx. Real Estate Witch House Price to Income Ratio Study, 2021. website: *https://www.realestatewitch.com/house-price-to-income-ratio-2021/*. Home Value Change vs. Income Change (1965-2021)

    1. There are three variables in this data source: Year (numeric, continuous), Home Value Change (percentage), Income Change (percentage).

6. topmidbot.xlsx. Economic Policy Institute, February 20, 2020. website: https://www.epi.org/publication/swa-wages-2019/. Top 0.1% earnings grew 15 times as fast as 90% earnings: Cumulative percent change in real annual earnings, by earnings group, 1979–2018.

    1. There are four variables in the data source, Year (numeric, cumulative), Bottom 90% (percentage), Top 1% (percentage), Top 0.1% (percentage).

R Packages

The R packages that I am familiar with are from our weekly readings (Field (2012)). This list may be updated after analyses are conducted.

ggplot2

knitr (using kable for any tables)

readxl

QuantPsyc

car

plyr

psych

clr: Centered log ratio transform

GGally

purr

dplyr

Plots and Table Needs

Because my data is time series, I will create many line plots of a variable over time. Likewise, I will explore the data, creating histograms and other plots in order to examine the distribution of each variable. Also, I will plot the different non-year variables against each other as a scatter plot in order to examine the spread and shape of the data points.

I will create tables of correlations after I transform the percentage values to see if inflation correlation with change of income (as an example). Also, I am investigating what testing I could do to determine significantly different time periods from each other. For example, to determine if one year's income change is significant over another year's change.

If I run my models, I will include a variety of tabled results from the model as well as exploratory analysis used to assist which variables should be used in a statistical model (i.e., examining correlations to ensure a relationship exists).

Questions for future steps

Questions I have for future steps are around my usage of R and my ability to combine these different data sources successfully (we did not cover merging in a way that I typically do merging so I may need to learn this). Also, though I have used data transformations before, I am not sure the one I am suggesting is the best to use. I worry that my data sources are percentages and thus will make it hard to do analyses in a meaningful way and that my analysis plan would work better on the source data of my sources. I have a concern that my analysis plan is not thought out enough and I will need to change my approach midway from something that I learn once I am working with the data. I may need to adjust my approach as this topic is very big and maybe my sources or my analysis is not refined enough. However, I am very interested to find out and then make improvements to this analysis.

References

Field, Miles, A. 2012. *Discovering Statistics Using r*. 1st ed. Thousand Oaks, CA: Sage Publications. https://studysites.sagepub.com/dsur/main.htm.

Fivecoat-Campbell, K. 2015. "Who Is Middle Class and Why It Matters." https://www.forbes.com/sites/nextavenue/2015/08/04/who-is-middle-class-and-why-it-matters/?sh=303876c53054.

Mishel, L. 2015. "Causes of Wage Stagnation." https://www.epi.org/publication/causes-of-wage-stagnation/.