

# are grocery prices attainable\*

## Analysing grocery prices with inflation and wages

Cristina Burca

December 1, 2024

### Abstract

We analyse the probability of Kamala Harris winning in the 7 swing states of USA using bayesian modeling; the states we look at are Arizona, Michigan, Pennsylvania, Nevada, Georgia, Wisconsin and North Carolina. USA is a large country with its national economy affecting the global economic conditions, which is why predicting the future President of USA will help understand the future economic condition of the world. Through bayesian modelling we found that North Carolina (47.26% support for Harris), Nevada (46.43% support for Harris), Wisconsin (48.38% support for Harris), Michigan (47.30% support for Harris), Pennsylvania (48.02% support for Harris) will vote Kamala Harris, while the rest of the swing states might vote Donald Trump, meaning that majority of the swing states will vote Kamala Harris. Provided this we predict that Kamala Harris will win the election. By creating a bayesian model based on ‘polls of polls’ where we compare the results from different polls, we are able to make this prediction.

## Table of contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Data Overview</b>	<b>2</b>
2.1	Citations . . . . .	2
2.2	Variables . . . . .	3
<b>3</b>	<b>Data Analysis</b>	<b>3</b>
<b>4</b>	<b>Model</b>	<b>5</b>
4.1	Model Setup . . . . .	6
4.2	Model Justification . . . . .	6
4.3	Model Results . . . . .	6
4.4	very vague Predicted Model . . . . .	7
4.5	Discussion . . . . .	7
4.6	Projected Affordability Index . . . . .	8

---

\*Code and data are available at: [Are Groceries Attainable](#).

<b>5</b>	<b>Discussion</b>	<b>8</b>
5.1	Limitations . . . . .	8
5.2	Next Steps . . . . .	8
<b>6</b>	<b>Appendix</b>	<b>9</b>
6.1	Data Cleaning . . . . .	9
6.2	Methodological Exploration of CPI Data Collection . . . . .	9
6.2.1	Grocery Data Collection . . . . .	9
6.3	Wage Data Collection . . . . .	10
6.4	CPI Data . . . . .	10
6.5	Linear Model Summary Statistics . . . . .	10
6.6	Posterior Predictive Check . . . . .	11

# 1 Introduction

Inflation in Cananda has increased rampantly in all sectors, growing concerns about affordability. Within this, the roles of inflation and stagnant wages have made groceries less attainable for Canadians. Grocery prices have risen 22.5% since May of 2020, and continue to rise each month, creating panic for households (News (2024)). In this paper, we explore the inflation rates of groceries in Canada from 2017 to 2023, relating it to wages. We then aim to predict the cost of groceries over the coming years, assessing if the increasing rate of wages is enough for groceries as an average Canadian.

“How have grocery prices in Canada changed over time, and how do these changes, relative to inflation and wages, affect the accessibility of groceries across socioeconomic classes?”

Focus on a subset of comparable products (e.g., fruits vs. snacks, vegetables vs. processed foods). Analyze how prices for healthy and unhealthy groceries have changed over time.

Overview paragraph

Estimand paragraph We want to predict grocery prices based on past grocery prices, inflation and wages? to determine if Canadians will be able to afford groceries in the coming years, assuming inflation and wages are increasing at a steady rate.

Results paragraph

Why it matters paragraph

Telegraphing paragraph: The remainder of this paper is structured as follows. ?@sec-data....

# 2 Data Overview

## 2.1 Citations

The datasets used in this analysis are retrieved from Statistics Canada ((cite?)). The main dataset is acquired from (Statistics Canada 2024), containing grocery prices of set units for each month of the year. This dataset is recorded beginning January 2017, and thus analysis with this data set will only be from 2017. The second dataset ((cite?)) contains monthly records of Consumer Price Index (CPI), all-items excluding the effect of indirect taxes. CPI is used as an indicator of change for consumer prices, or in other words, rate of inflation. It is measuring the prices, and thus price difference, of a set number of common groceries and

household goods at a set quantity(unit) representing the average Canadian household. The third dataset contains records of yearly wages of Canada. The three datasets were merged together for analysis.

## 2.2 Variables

The important variables that are focused on throughout this paper:

- **Date:** Contains the date of the recording of the observations, formatted as YYYY-MM-DD. Days are set to 01, since daily data is not recorded.
- **CPI:** Contains monthly values of recorded CPI. CPI is a unitless measure, where the average price level in the base year is set to 100.
- **Wage:** Yearly average wage of all employees from all industries.
- **CPI\_Normalized and Wage\_Normalized:** calculated columns to measure.
- **Average\_price:** calculates the average price based off selected items that make up a Canadians' "essential grocery basket".
- **Affordability :** A variable created as an Affordability Index, calculated by Average price divided by Wage.

The Affordability Index is a measure that reflects the ease in which households or individuals can afford a list of essential goods, relative to income– in this context we consider average wage. Higher values of the index indicate better affordability, while lower values suggest affordability struggles.

Items included in Average\_Price include... The total average is a very rough estimate of popular items purchased by Canadians. The average is not particularly important, but the impact and correlation between the two variables is.

More about data cleaning in the appendix section appendix (data?) cleaning.

## 3 Data Analysis

The first variables examined are CPI and Wage. In Figure 1, we analyse normalized values of yearly average wages and monthly CPI from 2000 to 2023, as they are measured on different scales. This graph illustrates the trend between CPI and wages. The period of 2000 to 2020 shows a predominantly economically steady correlation between the two variables, with predictable inflation rates. However, the Covid-19 pandemic began in 2020, causing restricted movement globally, subsequently disruption global economies and changing consumer spending behaviour, shown by the sharp growth after 2020. This resulted in a decline in demand of goods and services, which is reflected in the CPI. By 2022, CPI rapidly inflates due to increase consumer demand as economies reopened, but wages remain on a somewhat steady increase.

in Figure 2 between 2020-2022, covid. CPI went down because people were buying things, and wages went up surprisingly, since many people lost their jobs during covid, suggesting that companies were saving more money, people were getting paid more perhaps through bonuses and government covid funding. It is clear that the two variables are correlated, but may not necessarily depend on each other.

Introducing Average\_Price, Figure 3 depicts the trend of average grocery prices from 2017 onwards. It is obvious that grocery prices increase as inflation increases, but it is confounding how much inflation has risen in the past 7 years. A grocery cart that costed \$120 in 2017 can cost well above \$160 in 2023 – that is a 22% inflation. This number brings concern about the affordability of groceries in Canada, in addition to other inflating expenses, such as rent and utilities.

To investigate this concern, we analyze the Affordability Index created in ?@fig-affordability. Graphed over the years beginning in 2017, we notice that the general trend of affordability is decreasing, implying affordability challenges.

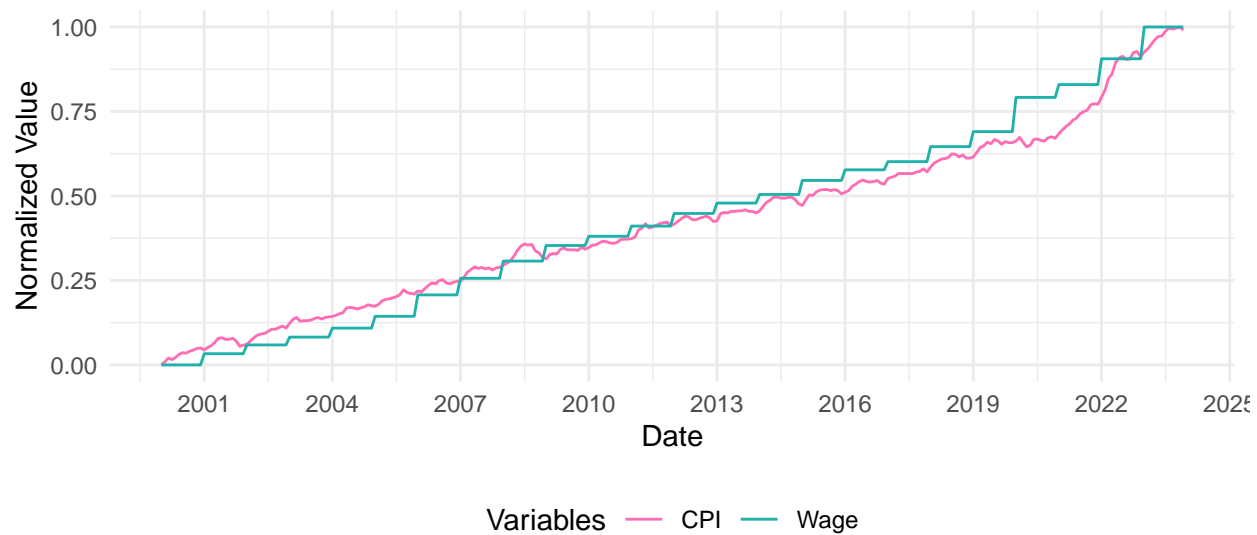


Figure 1: Normalized CPI and Wages Over Time from 2000 to 2023

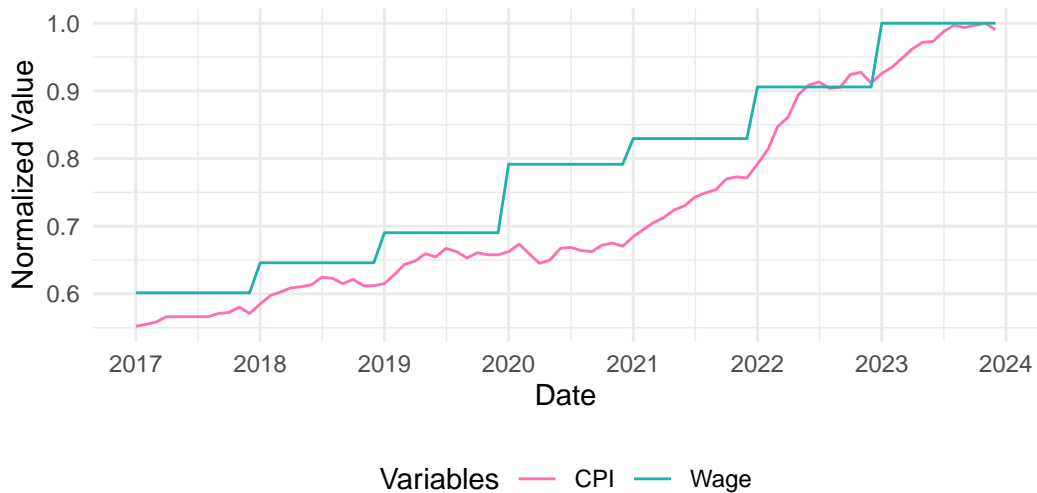


Figure 2: Normalized CPI and Wages Over Time from 2017 to 2023

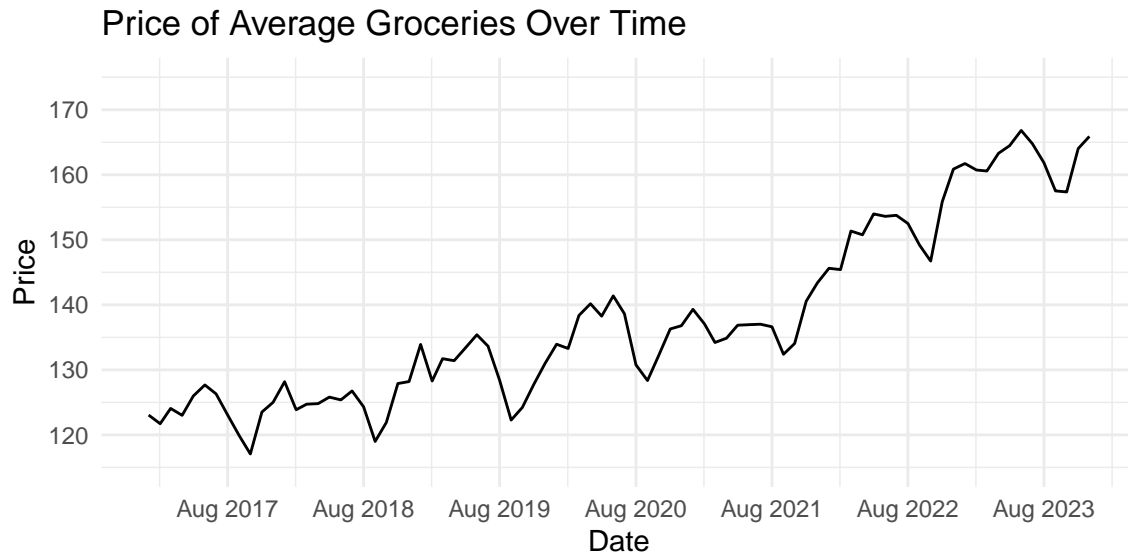
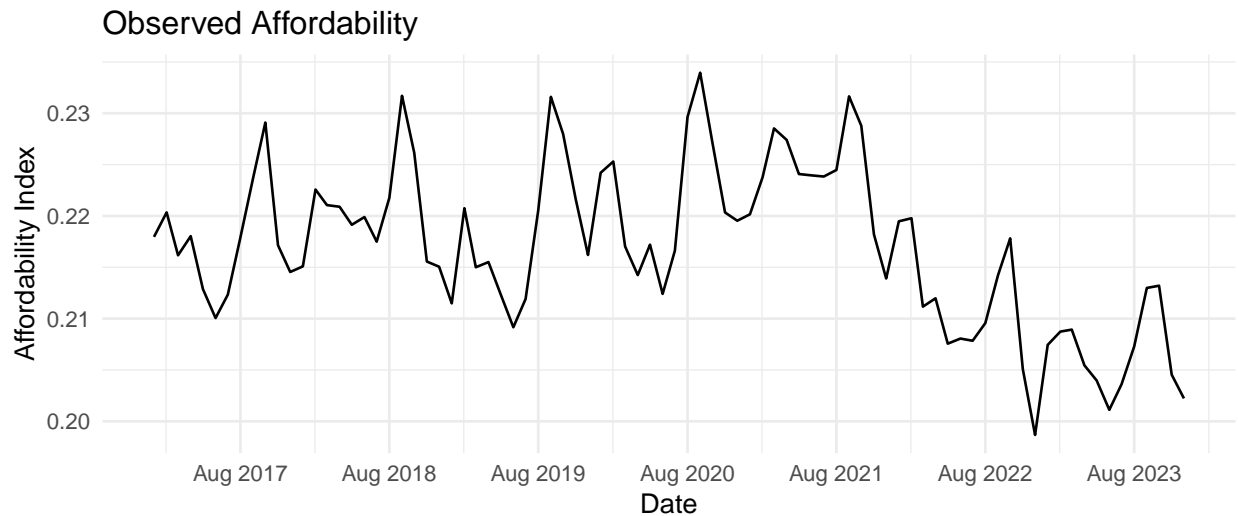


Figure 3: Normalized CPI and Wages Over Time from 2017 to 2023

even with increasing wages as CPI increases, there suggests that wage increasing is not enough to sustain the CPI increase.



We further explore this correlation through a model....

## 4 Model

Introduce model. We run it using (**rstanarm?**)...

Table 1: Summary statistics of model

Elasticity Model	
(Intercept)	1.202** (0.373)
log(CPI)	-0.885*** (0.153)
log(Wage)	0.484*** (0.127)
Num.Obs.	84
R2	0.431
RMSE	0.03
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001	

#### 4.1 Model Setup

$$\log(y_i) | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \quad (1)$$

$$\mu_i = \beta_0 + \beta_1 \times \log(\text{CPI})_i + \beta_2 \times \log(\text{Time})_i \quad (2)$$

$$\beta_0 \sim \text{Normal}(0, 2.5) \quad (3)$$

$$\beta_1 \sim \text{Normal}(0, 2.5) \quad (4)$$

$$\beta_2 \sim \text{Normal}(0, 2.5) \quad (5)$$

$$\sigma \sim \text{Exponential}(1) \quad (6)$$

The outcome variable of this model is a prediction of the affordability index, represented as  $y_i$ . The outcome is modeled as a function of CPI and Time. We choose not to use **Average\_Price** as a predictor because it would cause redundancy, as affordability is defined in terms of wage and average price. This would cause multicollinearity issues, and produce almost exact results. Instead, we focus on the elastic

- $y_i$  is the dependent variable, representing the Affordability Index,
- $\beta_0$  is the intercept value, when all predictors are set to zero,
- $\beta_1$  is a numerical variable, represent the elasticity
- $\beta_2$  is a numerical variable of Date turned into a cumulative column of days for each month of the year. Captures the change in the mean response  $\mu_i$  over time.

Each coefficient  $\beta$  is assigned a Normal prior distribution, centered at zero, with a standard deviation of 2.5 to prevent overfitting. Exponential prior ensures that variance  $\sigma > 0$ .

#### 4.2 Model Justification

what do we expect the results to be...

#### 4.3 Model Results

Interpreting Table 1, this table summarized the result of the elasticity model, which quantifies how changes in one variables affect changes in another through logarithmic transformations. Log transformations linearize

relationships between variables, and we use a log-log model to directly represent elasticities— defined as a 1% percent change in the dependent for a 1% change in the independent variable. In this equation,  $\beta_1$  represents elasticity of Affordability with respect to CPI, and  $\beta_2$  represents the elasticity of Affordability with respect to Wage. The summary statistics are understood to be as followed:

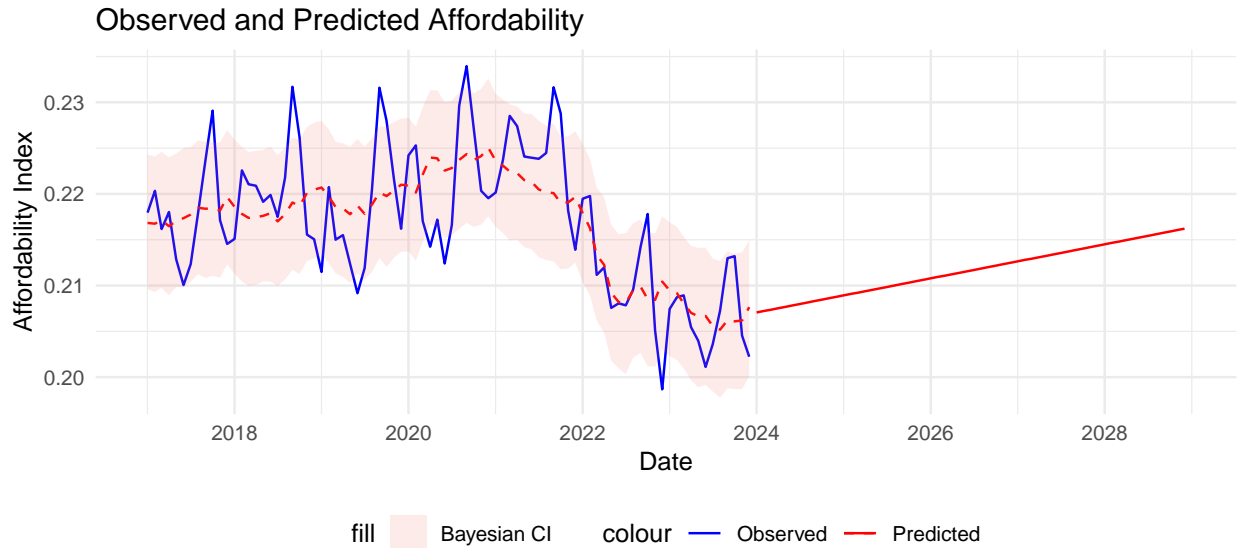
- $\beta_0$  is the intercept value of 1.202, which is when both  $\log(\text{CPI})$  and  $\log(\text{Wage})$  are set to 0,
- $\beta_1$  is  $\log(\text{CPI})$  with a coefficient of -0.885, meaning a 1% increase in CPI leads to a 0.885% decrease in affordability, when Wages are constant,
- $\beta_2$  is  $\log(\text{Wage})$  with coefficient 0.484, meaning a 1% increase in wages leads to a 0.484% increase in affordability, when CPI is constant.

Thus, wage growth positively affects affordability but does not fully offset inflation’s negative impact.

Model fit metrics will be discussed in (appendix?).

#### 4.4 very vague Predicted Model

- hard to forecast short or long term events that affect CPI, such a strikes, government bills, diseases/sickness, etc, thus accurate model is very ahrd to predict. However, it is informative to have a general sense of where affordability will be in coming years.



#### 4.5 Discussion

The interpretation of the model is shown in Figure 4, where the linear relationships is plotted between log of average price and log of CPI. The red regression line and lack of extreme outliers suggests a positive relationship between CPI and average grocery prices. The gradient of points indicates higher CPI values correspond to more recent years.

From Table 1,

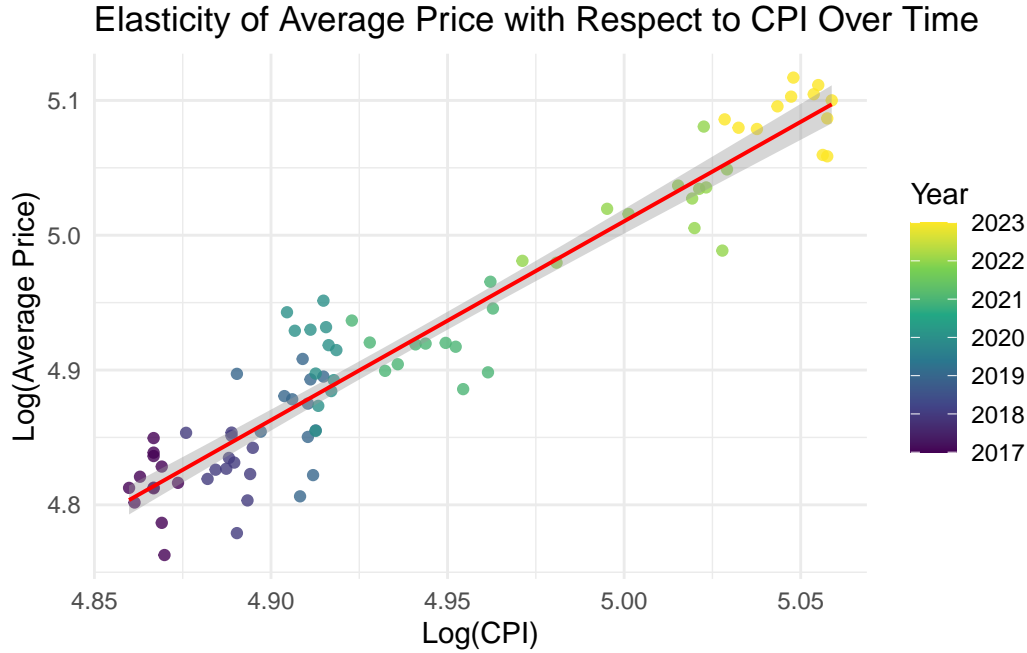


Figure 4: Predicted linear model

#### 4.6 Projected Affordability Index

- Model follows observed Affordability Index, capturing long-term trend
- The Bayesian prediction smooths out short-term fluctuations (e.g., the volatility in observed values during 2018 and 2019). This indicates that while real-world affordability may be influenced by sudden events (e.g., market shocks), the underlying trend is governed by systematic factors such as inflation and wage growth, which the model captures.
  - The smoothing effect helps reveal the underlying, systematic decline in affordability that may not be immediately visible in the raw data, which is influenced by noise or temporary fluctuations.
- The smoothing effect helps reveal the underlying, systematic decline in affordability that may not be immediately visible in the raw data, which is influenced by noise or temporary fluctuations.

### 5 Discussion

#### 5.1 Limitations

- limited produce
- limited data available (only from 2017)
  - more recent data, 2024 has high spikes in groceries

#### 5.2 Next Steps

- Perhaps a better question would be how does affordability compare to all sectors, not just grocery prices ?
- not focus on wages as a whole, but idfnet classes and wage ranges



## 6 Appendix

### 6.1 Data Cleaning

This section will be explaining the data cleaning process documented in `/scripts/03-clean_data`.

I used a 3 different datasets, split between 5 files. `old_inflation_data` and `inflation_data` are from the same source, but `old_inflation_data` contains years 2000 to 2023. The same was done for `old_wage_data`. This decision was done to make a broader analysis over many years easier, since data for `grocery_data` is only available from 2017 and on- variables `avg_wage_data` and `inflation_data` only contain data from 2017 and on as well. Similar cleaning processes were used for each variable. The data is loaded in and the first couple of rows are skipped since they contains subheadings?. One of the rows is kept for the column names using `colnames`, which then are renamed. `gsub` is used to rename the column names to remove spaces and characters that would conflict with the code. `mutate` was used to extract year and month columns for `grocery_data` and `inflation_data`, and `old_inflation_data`, since these variables are recorded monthly. `left_join` is used to merge `inflation_data` to `grocery_data` by variables `year` and `month`.

`ref_cpi` is created to calculate the percentage change in CPI, `CPI_Percentage`. Then, I took the values of wages for each year and manually added the value for each year in new variables `inflation_wage_data` and `'old_inflation_wage_data'`, both variables having columns of `cpi`, `wage` and `date`, where `wage` has the same value throughout the year since the record is yearly. Lastly, csv files are written in the folder `/02-Analysis_data` for variables `grocery_data`, `inflation_data`, `avg_wage_data`, `inflation_wage_data` and `'old_inflation_wage_data'`.

### 6.2 Methodological Exploration of CPI Data Collection

#### 6.2.1 Grocery Data Collection

The reliability of affordability analysis is determined by the accuracy of the considered variables. This appendix will highlight the importance of how CPI data, grocery data, and wage data was collected and proceed, to ensure the credibility of the data.

The CPI is an indicator used to measure the rate of inflation, with implementations in: - contracted payments, used to preserve purchasing power of a value, - obtaining constant dollar estimates of variables such as income of expenditure values, - monitoring implementation of economic policy, - research, such as cause and effect of inflation.

The target population includes households in urban and rural households in Canada. Citizens with distinctive circumstances, such as collective households, those living on Indian reserves, official representatives, chronic care patients, prison inmates, etc. are excluded from the target population.

The CPI price sample is derived from a basket of representative foods and goods across all geographical areas and stores. Only stores with high sales revenues are included. The timing of price collection is set for the first two weeks of every month to ensure consistency, but may vary depending on the nature of the good. Season items are only observed in their respective months. To ensure the compliance of international guidelines regarding price index methodology, manual and automated processes are implemented to ensure product characteristics over time.

The general sampling approach for the grocery data involves three main stages to ensure representative data. - stage 1: Areas are divided into subdivisions of municipalities and neighborhoods. Population count is used in the selection process of these subdivisions to prioritize areas with larger populations. - stage 2: Representative outlets (retail stores, small stores, or online retailers) are selected based on largest consumer purchases. - stage 3: Small samples of products are chosen from each product class to represent the broader class, based on popularity and research. Products with volatile prices, significant basket weights, and lower cost are prioritized. Each product class is weighed based on its share of total consumer expenditure.

Price collection is conducted by many methods through retail stores, flyers or online. Units of products are primary as possible, with goods usually measured per unit, per kilogram, or per liter.

### 6.2.1.1 Limitations of Methodolgy and Survey Samples

- focusing on private households may affect analysis, since different households have different consumption patterns, may not represent broader Canadian population.
- There is regional variation in price sampling, limited coverage for specific products in some areas
- a “judgmental” selection of the products
- sales, promotions, out- of-stock items, locally popular products, judgmental selection of products, all affect CPI accuracy. how?

## 6.3 Wage Data Collection

## 6.4 CPI Data

## 6.5 Linear Model Summary Statistics

Call:

```
lm(formula = log(Affordability) ~ log(CPI) + log(Wage), data = model_data)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.048924	-0.020340	-0.003576	0.018908	0.064259

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	1.2023	0.3727	3.226	0.001812	**
log(CPI)	-0.8852	0.1527	-5.795	1.26e-07	***
log(Wage)	0.4839	0.1267	3.818	0.000262	***

---

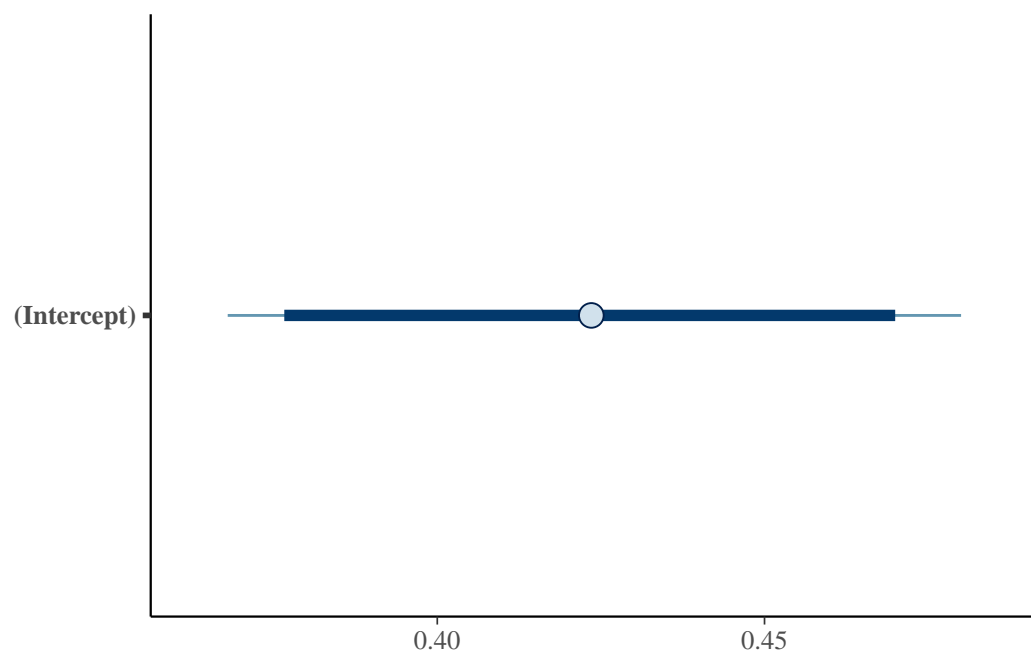
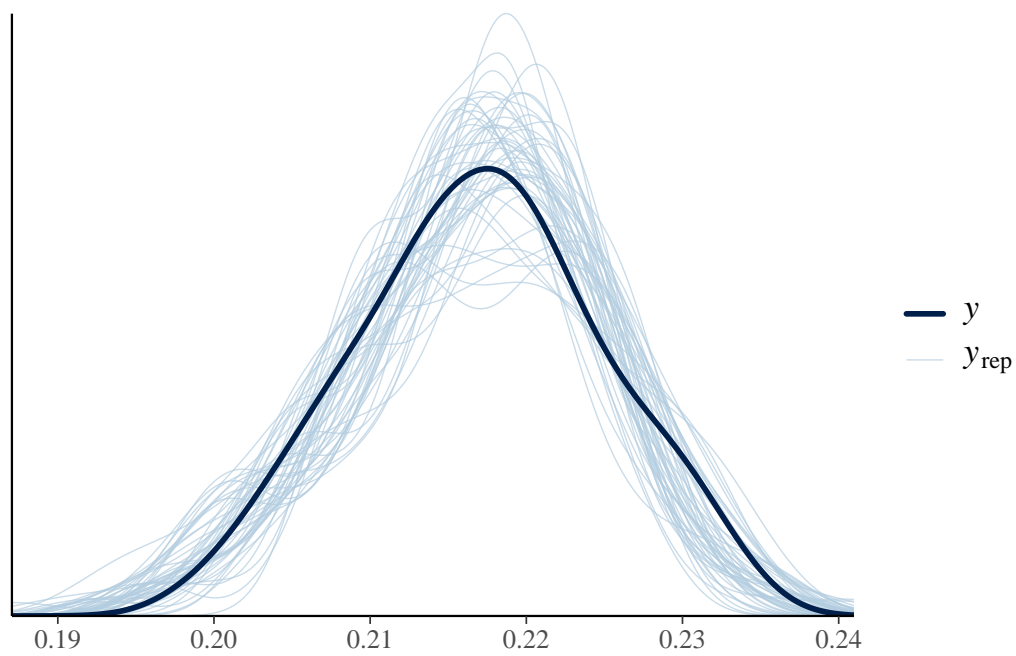
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02773 on 81 degrees of freedom

Multiple R-squared: 0.4315, Adjusted R-squared: 0.4175

F-statistic: 30.74 on 2 and 81 DF, p-value: 1.166e-10

## 6.6 Postrior Predictive Check



News, CTV. 2024. “Canadians Feel the Pinch at Grocery Store as Food Inflation Ticks Higher for First Time in Nearly a Year.” 2024. <https://www.ctvnews.ca/business/canadians-feel-the-pinch-at-grocery-store-as-food-inflation-ticks-higher-for-first-time-in-nearly-a-year-1.6940172>.

Statistics Canada. 2024. “Monthly average retail prices for selected products.” <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810024501>.