STATS/CSE 780
Homework Assignment 1

Konrad Swierczek

January 31, 2023

Introduction

Throughout 2022, Canadians experienced higher than usual increases in prices of consumer goods. However, the reality of these prices increases varies depending on province as well as the type of products. This report visualizes trends in the price of consumer products across Canada during 2022 and show that while prices did increase, both province and prouduct type are mediating factors in how much prices may have increased and when.

Methods

The dataset "Mothly average retail prices for selected products" (**???**) was retrieved from Statstics Canada. This data is available from 2017 and is collected monthly. For the purposes of this report, only data from 2022 was used. Data for all provinces and all of Canada was used. 110 products ranging from produce to personal hygiene products are displayed. All products were used, as well as the mean of all 110 products combined. Data is at present unavailble for December of 2022. Some data are not available in certain provinces due to data or product availablity (for instance, four litre milk bags are not availble in Newfoundland & Labrador at scale).

Results

Consumer products did increase in price across 2022, with an average price of $10 in January and $10 in November (see (**???**)). However, the data indicates that a significant decrease in prices occured in October.

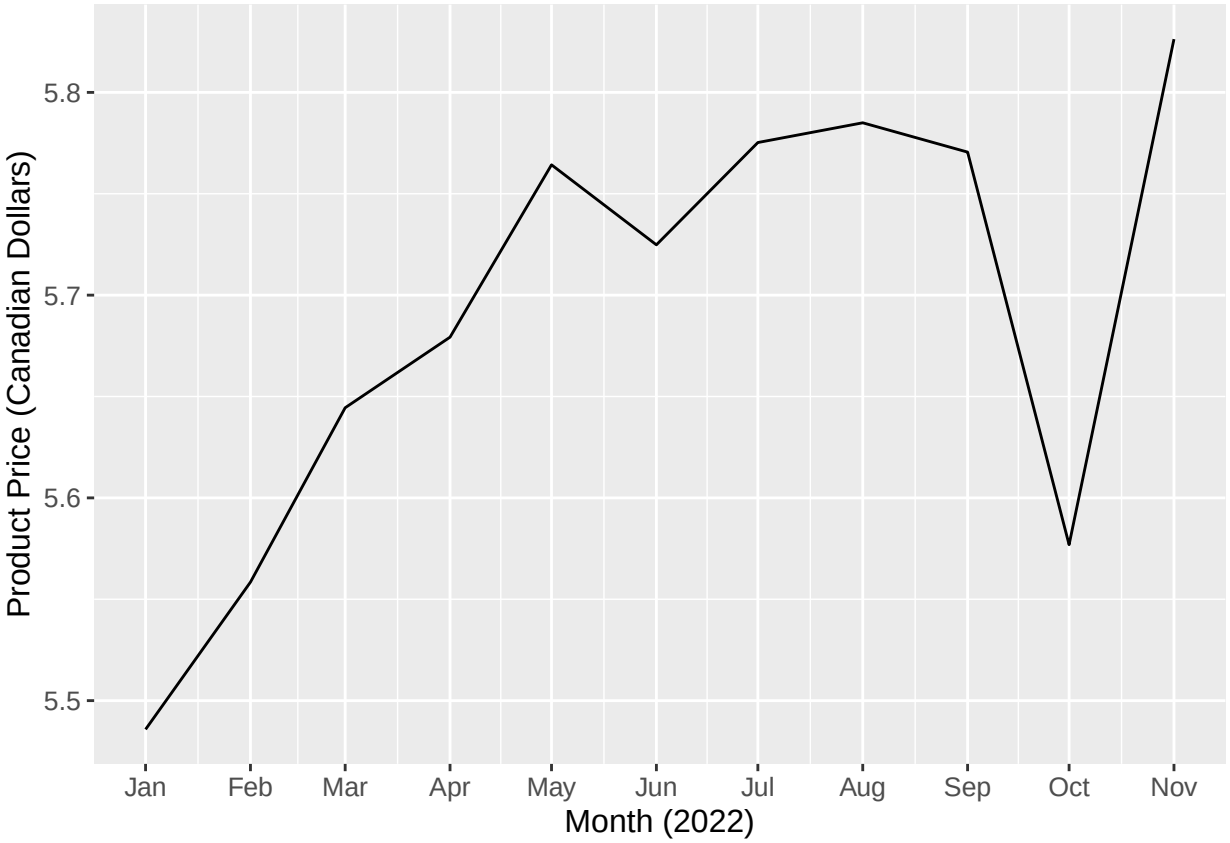

Figure 1: Average price of 110 consumer products in Canada. Prices are in Canadian dollars. Data is collected monthly.

Provincial differences @ref(fig:fig1).

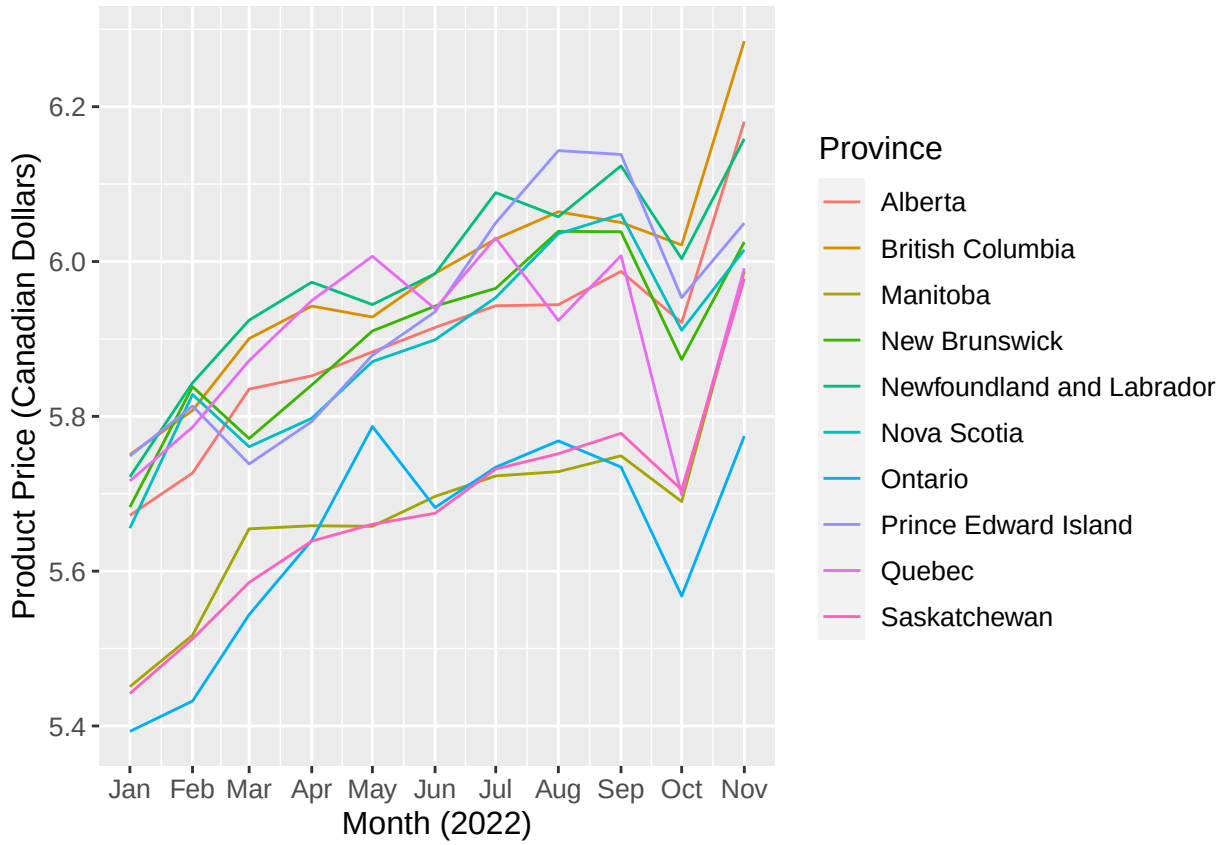

Figure 2: Average price of 110 consumer products across all Canadian provinces. Prices are in Canadian dollars. Data is collected monthly.

Finally in this *Shiny app*, each individual product is displayed.

Conclusions

What do we conclude?

ALl materials are available at.

References

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Statistics Canada. (2020). *Monthly average retail prices for selected products, by province* [Data set]. Government of Canada. https://doi.org/10.25318/1810024501-ENG

Wickham, H., & Grolemund, G. (2016). *R for data science: Import, tidy, transform, visualize, and model data*. " O'Reilly Media, Inc.".

Xie, Y., Dervieux, C., & Riederer, E. (2020). *R markdown cookbook*. CRC Press.

Supplementary Materials

```r
# knitr setup
knitr::opts_chunk$set(echo = TRUE)
knitr::opts_chunk$set(fig.pos = "H", out.extra = "")

# imports
library(ggplot2)
library(readr)
library(shiny)
library(dplyr)
library(knitr)

# "Monthly average retail prices for selected products"
# https://doi.datacite.org/dois/10.25318%2F1810024501-eng

# Pull Data
temp <- paste(tempfile(), ".zip", sep = "")
download.file("https://www150.statcan.gc.ca/n1/tbl/csv/18100245-eng.zip", temp)
data <- read_csv(unz(temp, "18100245.csv"))
metadata <- read_csv(unz(temp, "18100245_MetaData.csv"))
unlink(temp)

# Tidy Data
tidyData <- data
# Subset data for only 2022
tidyData$date <- lubridate::ymd(paste(data$REF_DATE, "-01", sep = ""))
tidyData <- subset(tidyData, date > "2021-12-31" & date < "2023-01-01")
# Add entries for means of all products
means <- tidyData %>%
  group_by(GEO, date) %>%
  summarise_at(vars(VALUE), list(VALUE = mean))
```

```r
means$Products <- "All Products"
tidyData <- bind_rows(tidyData, means)

# Write .RData for shinyapp
save(tidyData, file = "shiny/tidyData.RData")

# Figure 1: Price in  Canada have increased
ggplot(data = subset(subset(tidyData, Products == "All Products"), GEO == "Canada")) +
        geom_line(aes(x = date, y = VALUE)) +
        xlab("Month (2022)") +
        ylab("Product Price (Canadian Dollars)") +
        scale_x_date(date_breaks = "months", date_labels = "%b") +
        theme(text = element_text(size = 12))

# Figure 2: These trends vary by province
ggplot(data = subset(subset(tidyData, Products == "All Products"), GEO != "Canada")) +
        geom_line(aes(x = date, y = VALUE, colour = GEO)) +
        labs(colour = "Province") +
        xlab("Month (2022)") +
        ylab("Product Price (Canadian Dollars)") +
        scale_x_date(date_breaks = "months", date_labels = "%b") +
        theme(text = element_text(size = 12))
```