PSY6422 Data Management & Visualisation: Module Project

2023-05-16

Project Focus

I wanted to choose a project which could ultimately present data about an important issue, but that could still have a light hearted and interesting spin put on it. I remembered seeing a tweet from someone who had worked out how many Freddos an hour Doctors could afford now, compared to 13 years ago (https://twitter.com/Huw_Corness/status/1609957206757081090?cxt=HHwWhICwxZbj29csAAAA). This inspired me. After some deliberation I decided to focus on the Freddo Index, looking at this in relation to the national minimum and living wage.

Data Origins

```
# Load libraries
library(plyr) # For working with data
library(tidyverse) # For working with data
library(here) # Set working directory
library(readxl) # Read xlsx files
library(plotly) # To use ggplotly
library(dplyr)
library(ggplot2)
```

I put together two datasets, one for Freddo price information and one for minimum wage information, both for the time period from 2020 to 2023. I put together the Freddo price data by searching for the prices on the internet and finding these through news articles. The minimum wage information was found through a UK government webpage and two UK government datasets.

Freddo prices information sources:

2020 - 2018 prices: https://www.vouchercloud.com/resources/the-freddo-index

2019 price: https://www.mirror.co.uk/money/shopping-deals/freddos-10p-again-tesco-one-13875080

2020 price: https://www.bigissue.com/news/fact-fiction-do-freddos-show-that-the-minimum-wage-should-be-18-per-hour/

2021 price: https://metro.co.uk/2021/08/26/joe-lycett-wades-into-freddo-war-as-investigates-rising-prices-15157463/

2022 price: https://www.walesonline.co.uk/news/cost-of-living/now-how-cost-freddo-changed-23633251

2023 price: https://www.thecourier.co.uk/fp/education/schools/4049089/freddos-teachers-wages-fife-video/

National Minimum and National Living Wage sources:

2000 - 2019 wages: https://www.gov.uk/government/publications/20-years-of-the-national-minimum-wage

2020 wage: https://www.gov.uk/national-minimum-wage-rates

2021 - 2023 wages: https://www.gov.uk/government/publications/the-national-minimum-wage-in-2023

In the Freddo Prices dataset, there is a column for year and a column for the price of a Freddo as a whole number.

```
# Load Freddo prices data
freddo <- read_xlsx(here("data", "Freddo Prices.xlsx"))

# Show first 6 rows of the data
head(freddo)
```

```
## # A tibble: 6 x 2
##    Year FreddoCostPence
##   <dbl>          <dbl>
## 1  2000             10
## 2  2001             10
## 3  2002             10
## 4  2003             10
## 5  2004             10
## 6  2005             10
```

For looking at the national minimum wage, as there is varying degrees of this for different age groups, I decided to approach this looking at the wages targeted at 25 year olds, this being my age. In 2016, the national minimum wage become the national living wage for UK workers aged 23 and over. This information should be considered when observing my visualisation. In the original datasets, some years provide two amounts, one for April and one for October of the same year. In these cases, I have used the April figure, so each minimum wage value is from the April of the year it belongs to.

In the National Minimum Wage dataset, there is a column for year, a column identifying whether the national minimum wage or national living wage was in use and a column for the wage of the particular wage.

```
# Load National Minimum Wage data
N_M_Wage <- read_xlsx(here("data", "National_Minimum_wage.xlsx"))

# Show first 6 rows of the data
head(N_M_Wage)
```

```
## # A tibble: 6 x 3
##    Year NMW_or_NLW WagePounds
##   <dbl> <chr>          <dbl>
## 1  2000 NMW              3.7
## 2  2001 NMW              4.1
## 3  2002 NMW              4.2
## 4  2003 NMW              4.5
## 5  2004 NMW             4.85
## 6  2005 NMW             5.05
```

Research Question

How has the number of Freddos an individual can afford, with an hours worth of the national minimum/living wage, changed over the past 23 years?

Data Preparation

To prepare the data for visualisation, I joined the two data sets together, created two new columns (one to show Freddo prices in decimals and one to show number of Freddos afforded with an hours minimum wage pay) and rounded down the values in one of the columnsa new column with new information and rounded down the values in the new column, from decimals to whole numbers. Some columns were renamed too, to make them easier to work with going forward.

```
# Join Freddo and National minimum wage tables together

Freddo_and_N_M_Wage <- inner_join(freddo, N_M_Wage, by = "Year")

# Making new columns

# Use FreddoCostWhole column to make a FreddoCostDec column, to show Freddo cost in decimals

Freddo_and_N_M_Wage <- mutate(Freddo_and_N_M_Wage, FreddoCostPence_Dec = FreddoCostPence/100)

# Make new column to show Freddos afforded with minimum wage

Freddo_and_N_M_Wage <- mutate(Freddo_and_N_M_Wage, FreddosAfforded = WagePounds/FreddoCostPence_Dec)

# Update FreddosAfforded column with rounded down values of FreddosAfforded column

Freddo_and_N_M_Wage <- Freddo_and_N_M_Wage %>%
  mutate_at(vars(FreddosAfforded), ~ floor(.))
```

Now the data is prepared, we will make a new dataframe, keeping columns needed for the data visualisation only

```
FA_Data <- Freddo_and_N_M_Wage %>%
  select(Year, WagePounds, FreddosAfforded)

FA_Data <- FA_Data %>%
  rename(FreddosAfford = FreddosAfforded)
```

Now we are ready to start on the data visualisations!

Visualisation 1

For the first visualisation, I wanted to plot the individual data points as well as a line to reflect the changes in them.

```
p <- ggplot(data = FA_Data,
            mapping = aes(x = Year,
                          y = FreddosAfford))
```

```r
p + geom_point(color = "springgreen4") +
  geom_smooth(method = "gam", color = "purple4") +
  labs(x = "Year", y = "Number of Freddos afforded with hours worth of Minimum Wage",
       title = "Freddos Afforded With The Minimum Wage:
       From 2000 to 2023") +
  theme(plot.title = element_text(color = "purple4", size = 15, hjust = 0.5, face = "bold")) +
  theme(panel.background = element_blank(), axis.line = element_line(colour = "purple4")) +
  theme(panel.border = element_rect(color = NA, fill = NA)) +
  theme(text = element_text(color = "purple4"),
        axis.text = element_text(color = "purple4"))
```
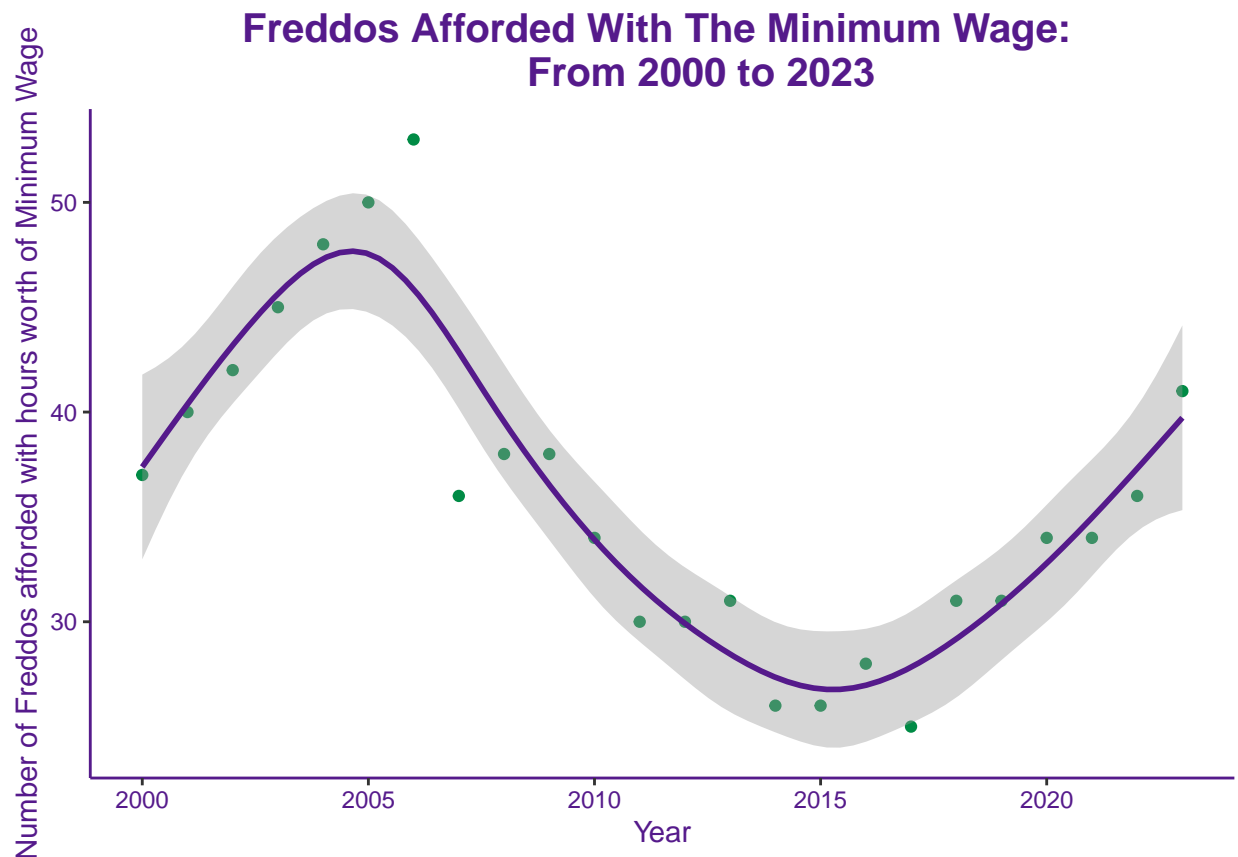

This visualisation presents the data reasonably well - but it felt a bit basic to me. I wanted to appropriately reflect how the number of Freddos afforded has changed dramatically over the years. A smooth curve didn't seem appropriate - I decided to instead make a graph with a line that connected up all the data points, to represent some of the steep changes in a better way. I also decided to introduce another element to the graph: formatting the size of the data points to represent the minimum wage amount increase throughout the years.

Visualisation 2

```r
p2 <- ggplot(data = FA_Data,
             mapping = aes(x = Year,
                           y = FreddosAfford))
```

```
p2 + geom_line(size = 1, color = "springgreen4") +
  geom_point(color = "purple4", aes(size = WagePounds)) +
  scale_size_continuous(range = c(1, 5.5)) +
  labs(x = "Year", y = "Number of Freddos afforded with hours worth of Minimum Wage",
       title = "Freddos Afforded With The Minimum Wage:
       From 2000 to 2023", size = "Wage
(Pounds)") +
  theme(plot.title = element_text(color = "purple4", size = 15, hjust = 0.5, face = "bold")) +
  theme(panel.background = element_blank(), axis.line = element_line(colour = "purple4")) +
  theme(panel.border = element_rect(colour = "purple4", fill = NA, size = 3)) +
  theme(text = element_text(color = "purple4"),
        axis.text = element_text(color = "purple4"))
```
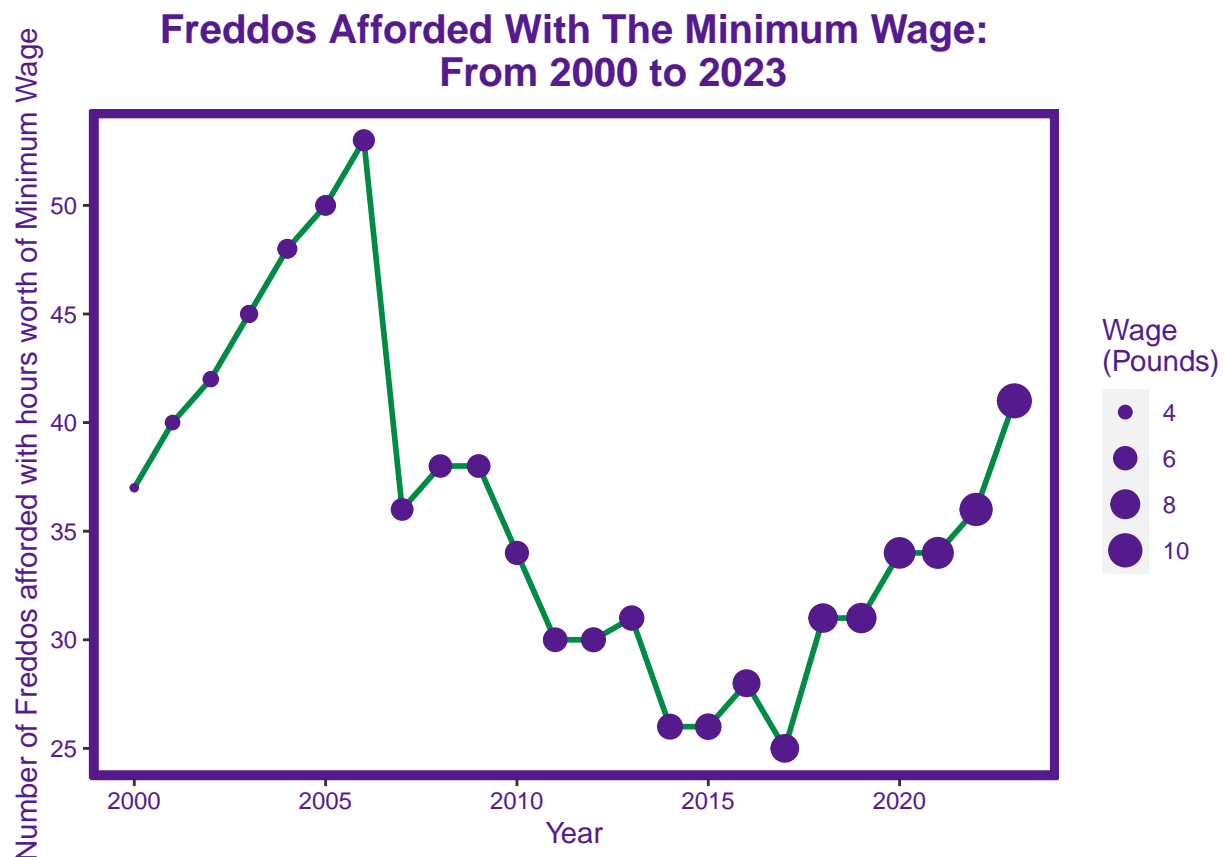

This graph gives the kind of line I had in mind, showing the sharp drops between years. I added in formatting to make the data point circles vary in size according to the price of Freddos in each year. However, this idea did not translate as well as I thought it would onto the plot itself - it looked a bit clunky. Therefore, I decided to keep the data point formatting to represent the minimum wage, but include a line of best fit instead of lines connecting the data points. I also wanted to bring an element of interactivity to the visualisation, so included the use of ggplotly.
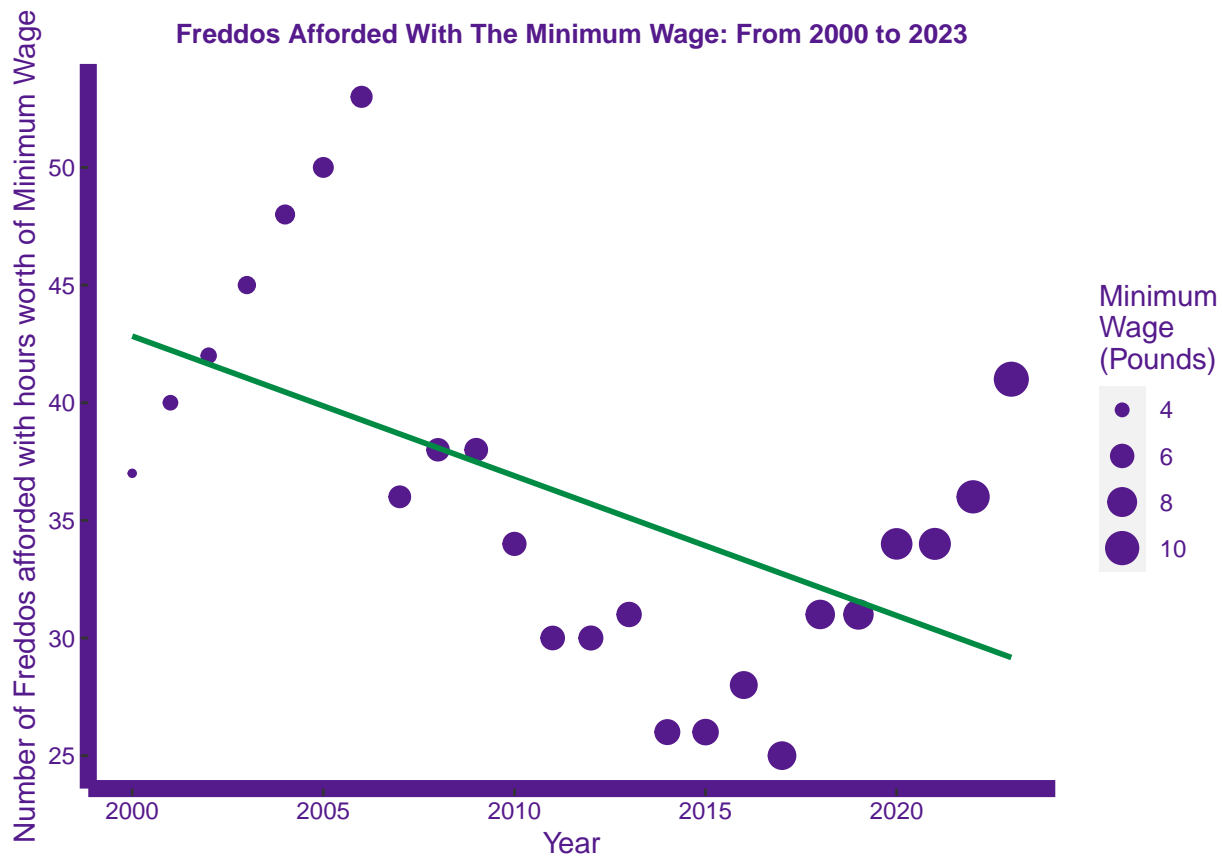
Visualisation 3

```r
p3 <- ggplot(data = FA_Data,
             mapping = aes(x = Year,
                           y = FreddosAfford)) +
  geom_point(aes(size = WagePounds), color = "purple4") +
  scale_size_continuous(range = c(1, 5.5)) +
  geom_smooth(method = "lm", color = "springgreen4", se = F) +
  labs(x = "Year", y = "Number of Freddos afforded with hours worth of Minimum Wage",
       title = "Freddos Afforded With The Minimum Wage: From 2000 to 2023",
       size = "Minimum
Wage
(Pounds)") +
  theme(plot.title = element_text(color = "purple4", size = 10, hjust = 0.5, face = "bold")) +
  theme(panel.background = element_blank(), axis.line = element_line(colour = "purple4", size = 3)) +
  theme(panel.border = element_rect(colour = NA, fill = NA, size = 3)) +
  theme(text = element_text(color = "purple4"),
        axis.text = element_text(color = "purple4"))

print(p3)
```



```r
ggplotly(p3)
```

```r
ggsave(here("plots", "freddograph.png"))
```

Summary

The visualisations present an interesting history of the 'Freddo index' and the minimum wage since 2000. The minimum wage has risen steadily throughout, but the number of Freddos an individual can afford with the minimum wage has not. From 2000 to 2005 there was a continuous increase, before this gradually declined (albeit with some slight increases) until 2017, when the increases started again, but not the same level as there was in the first several years since 2000.

I was pleased with final visualisation I made. In particular, I am glad I made a interactive graph, with hovering over the data points providing the specific year, minimum wage and Freddos afforded amount. Including a linear model line felt appropriate, to show an overall trend line, which is interesting when this is judged against the data points.

If I were to run this project again, I would wish to explore more options with formatting, using animation and finding the best way to include the cost of the Freddo itself in the graph. I decided not to include this information in the graph so as to avoid information overload, though addition of this could help to improve context.