

**Assignment 1****Part 1****Aim:** To understand and implement:

- Produce a colour drawing\* of a dot plot approach that shows the latest CPI figures for the following 3 measures: Total, Food and Total excluding Food and Energy.
- Produce a colour drawing\* of an equivalent bar chart approach that visualises the same information.
- Discuss your approach and whether the dot or bar approach is more effective in communicating the information required to support your article. (Half page required)

**Theory/Working:**

The CPI for countries *Great Britain (GBR)*, *United States of America (USA)*, *Austria (AUT)*, *Mexico (MEX)*, *France (FRA)*, *Spain (ESP)*, *Canada (CAN)* and group of nations *G-7* was taken to be compared against our country, ***Ireland (IRL)***, with respect to the 3 measures: **Food, Total & Total Excluding Food and Energy**. The data related to December 2022 was taken using the filter method on the attribute "Time" to select the data of that month/year. The result is shown below in figures A and B when drawn in an approach of dot plot and bar plot respectively, and their visualization is expected to be rendered by the computer as seen:

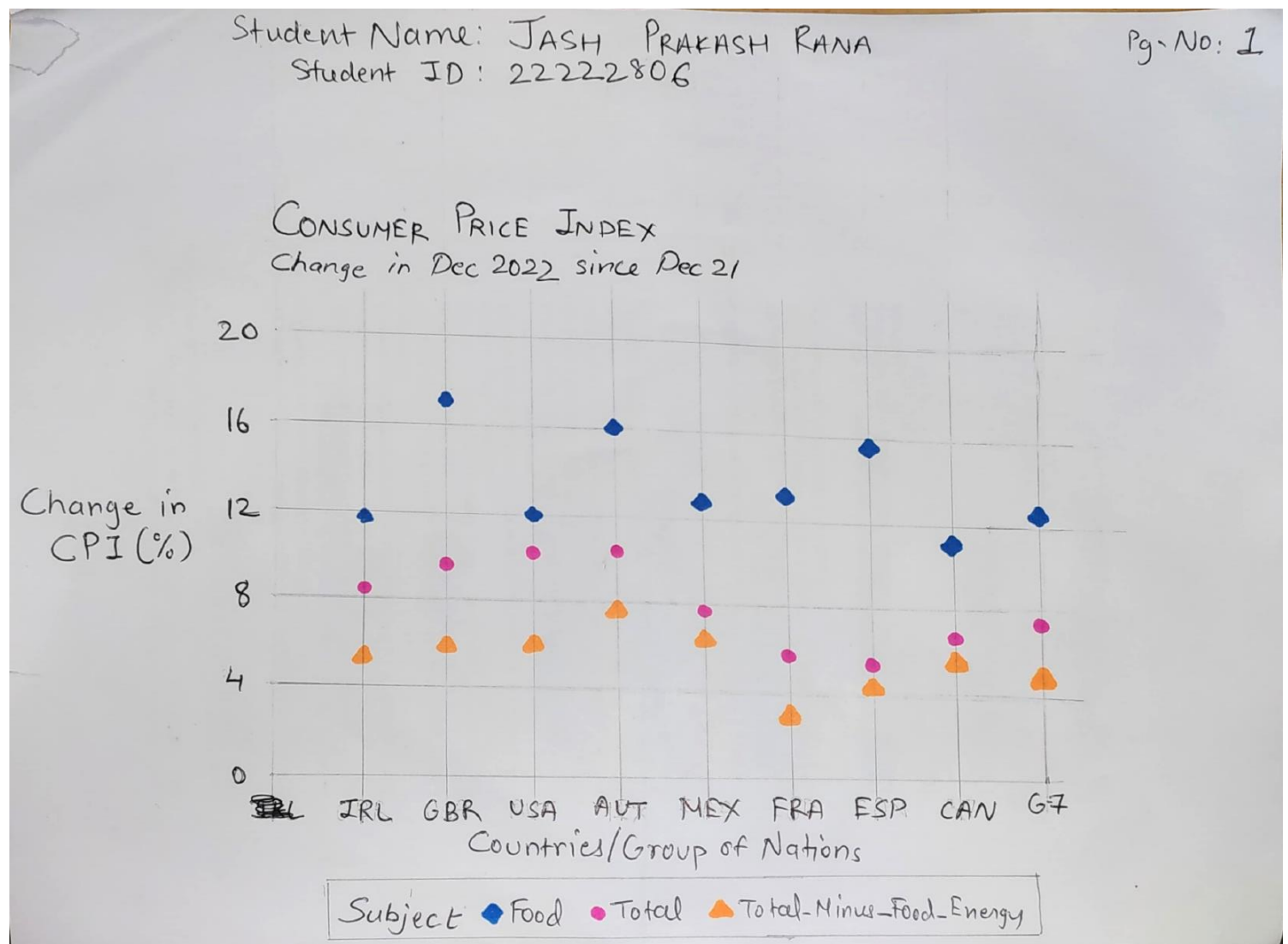


Figure A: Dot Plot Drawing of CPI showing change in Dec 2022 since 2021.

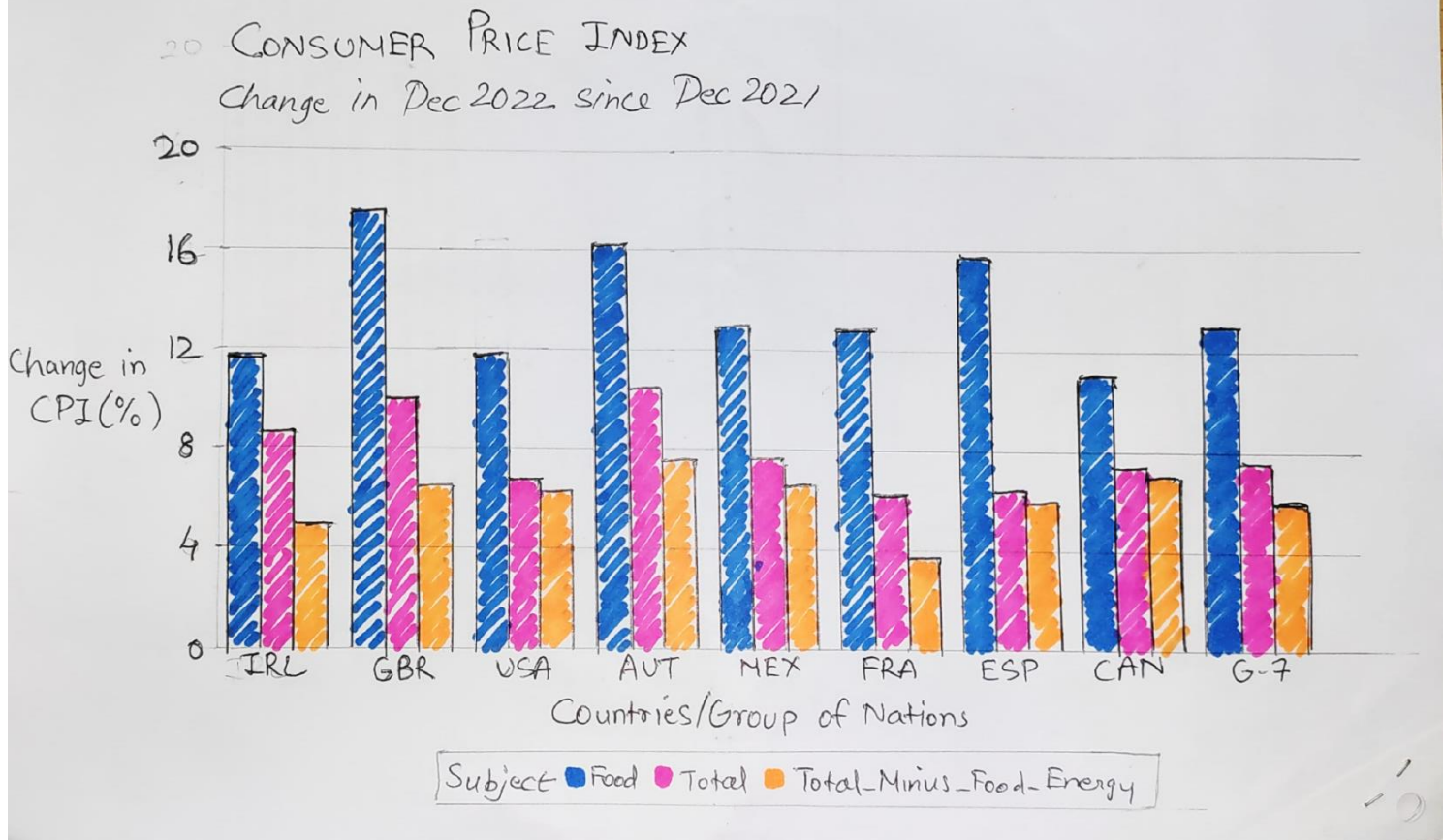


Figure B: (Grouped) Bar Chart Drawing of CPI showing change in Dec 2022 since 2021.

**Inference:** A dot plot compares various measures over a single line in the X-axis, while a bar plot provides different bars over X-axis for each measure. Technically, in this case, dot plot looks less cluttered as the data points are not closer to each other, but if a cluttered data comes along the way, it would be better to approach them with the bar plot. The current charts only show the countries if factored in the way shown in the X-axis, but an even better approach would be to take Subject in the X-axis and then reorder them in the descending order of percentages for each country. In both cases, my bias goes towards the bar chart out of both, as it is easy to read a visual when more colour is used to show something; and more the colour on the page, the more amount of CPI is interpreted by the brain. For e.g., GBR is highest in 'Food' measure and is easily perceived in Bar chart, but we need to find the dot in the Dot Plot to interpret the same information; same goes for AUT being the highest when it comes to 'Total' measure. Bar plot also tends to require less information and theme designs required to gather information, as we see that it was essential to plot the major x axis line in the dot plot to track the dots for each country, while the same isn't required in the bar chart. It is possible to capture small details and changes if we truncate a dot plot to a certain limit on Y-axis, but the same is not possible in bar charts as it gives a wrong message, so it becomes a little nuisance to understand the small differences between multiple bars, but as we need to compare our country with others, it gives us an edge to normally understand and differentiate between countries to find out where our chosen country lies between them. From what I see, it becomes clear that there is a moderate increase in the CPI for IRL in December 2022 since the last year, with Total coming near to 8.5% while the Food almost reaching 12% and Total without Food and Energy around 4.8 to 5%. This information was easy to understand and interpret in both the plots, but coming down to the crux of the matter, it becomes easy when comparing it to either GBR or USA or the whole X-axis.

**Conclusion:** Thus, a bar chart is the preferred method of visualization in this case over the dot plot due to its efficient comparability and easy data readability even for a non-technical human being.

# Assignment 1

Name: Jash Prakash Rana, ID: 22222806

16th Feb, 2022

## Part 2

Implementation of Part 1 using ggplot2 in RStudio, along with some modifications explained below.

### 1. Importing the data

```
cpi_data <- read.csv("Counsumer_Price_Index.csv")  
  
#Ref: https://www.geeksforgeeks.org/delete-rows-with-empty-cells-from-excel-using-r/  
cpi_data <- cpi_data[!apply(is.na(cpi_data) | cpi_data == "", 1, all),]
```

### 2. Creating a plotting data using dplyr::filter package.

- Filtering the data according to Dec 2022
- Changing the column name of Percentage to Per\_2022\_12
- Removing the Time and Frequency columns
- Changing 'Total\_Minus\_Food\_Energy' to TMFE
- Factoring Subject column to show Total, Food and TMFE in that level.

```
library(dplyr)  
  
sub_req = c("Total", "Food", "Total_Minus_Food_Energy")  
loc_req = c("IRL", "GBR", "USA", "AUT", "MEX", "FRA", "ESP", "CAN", "G-7")  
  
plot_data <- cpi_data%>%  
  dplyr::filter(Time == "2022-12",  
                Subject %in% sub_req,  
                Location %in% loc_req)  
colnames(plot_data)[5] <- "Per_2022_12"  
plot_data <- subset(plot_data, select = -c(Time, Frequency))  
plot_data$Subject <- gsub("Total_Minus_Food_Energy", "TMFE", plot_data$Subject)  
plot_data$Subject <- factor(plot_data$Subject, levels = c("Total", "Food", "TMFE"))  
head(plot_data)
```

```
##   Location Subject Per_2022_12  
## 1      AUT      Food  16.097560  
## 2      AUT     Total  10.151800
```

## 3	CAN	Food	11.020150
## 4	CAN	Total	6.319445
## 5	FRA	Food	12.866780
## 6	FRA	Total	5.850718

### 3. Creating a Dot Plot

The dot plot was created in focusing the Location IRL into the foreground, while the other countries and G-7 into the background. The major change that we see from part 1 is the focus on our country IRL, and instead of the Location in the X-axis, I decided to use Subject in the X-axis as that is something which can be grouped/reordered by the Percentage (Per\_2022\_12) attribute. While I also have coloured the Group of Nations' shape to a blue shade and our country by the pink shade, I decided to keep all the other countries by the grey colour as shown in the Legend, and put a caption which is displayed directly below the Legend which talks about the countries included in the 'Other' section. I have given shape 21 (circle with colour fill) to IRL while we have given shape 4 (X or cross) to all the other countries. I have disabled only the minor gridline at X-axis as our data is categorical and not a continuous value. I have also changed the Y-axis title to rotate it, so that we can see it without any neck tilt. And finally, the legend was taken below the X-axis, in the center, to save space and also to make the graph bigger in size by the width.

I agree that the dot plot is cluttered from Part 1 but I found this approach to be a better way to show as we can finally see the pattern showing where our country lies in the Location data, while ranking it in each different measure of Subject column. We can make an easy comparison this way between our country and the others this way according to the measures, and it becomes clear on how much change did our country have compared to other. As said in Part 1, the better approach in this scenario was to use bar charts and as we progress, we will see why that approach clears many things visually and how it is easy to understand our data using that approach.

```
library(ggplot2)

shapes <- c(4, 21)
Others <- c("GBR", "USA", "AUT", "MEX", "FRA", "ESP", "CAN")

dotplot <- ggplot(plot_data, aes(x = Subject, y = Per_2022_12,
                                group = -Per_2022_12,
                                colour = Location,
                                fill = Location,
                                alpha = Location == "IRL",
                                shape = Location == "IRL")) +

  geom_point(size = 3.2) +
  scale_shape_manual(values = c(4, 21)) +
  scale_fill_manual(values = c("IRL" = "#D81B60", "G-7" = "#1E88E5", Others = "lightgrey")) +
  scale_colour_manual("Countries/Group of Nation",
                      values = c("IRL" = "#D81B60", "G-7" = "#1E88E5", Others = "lightgrey")) +
  scale_alpha_manual(values = c(0.7, 1), guide = "none") +
  labs(title = "Consumer Price Index (CPI) Comparison",
       subtitle = "Change in December 2022 since December 2021",
       caption = "Other Countries = USA, AUT, MEX, FRA, ESP & CAN") +
  xlab("Countries/Group of Nations") +
  ylab("Change in \nCPI (%)") +
  guides(
    colour = guide_legend("Countries/Group of Nation"),
    fill = guide_legend("Countries/Group of Nation"),
    shape = "none") +
  scale_y_continuous(breaks = seq(0, 20, by = 4), limits = c(2, 18)) +
```

```

theme(panel.background = element_rect(fill = "white"),
      panel.grid.major.x = element_line(size = 0.1, linetype = "dashed",
                                         colour = 'lightgrey'),
      panel.grid.major.y = element_line(size = 0.12, linetype = "solid",
                                         colour = 'lightgrey'),
      panel.grid.minor.y = element_line(size = 0.068, linetype = "dotted",
                                         colour = 'lightgrey'),

      axis.ticks = element_blank(),
      axis.text = element_text(size = 9),
      axis.title = element_text(size = 11),
      axis.title.y = element_text(size = 11,
                                   angle = 0,
                                   vjust=0.5,
                                   hjust=0,
                                   margin = margin(r=4)),

      legend.title = element_text(face="italic", size=11),
      legend.text = element_text(face = "italic", size = 9),
      legend.box.background = element_rect(fill = 'lightgrey', colour = 'lightgrey'),
      legend.key = element_rect(fill = "white"),
      legend.position = "bottom",
      legend.direction = "horizontal",
      legend.justification = "center",
      legend.box.margin = margin(0.5, 0.5, 0.5, 0.5),

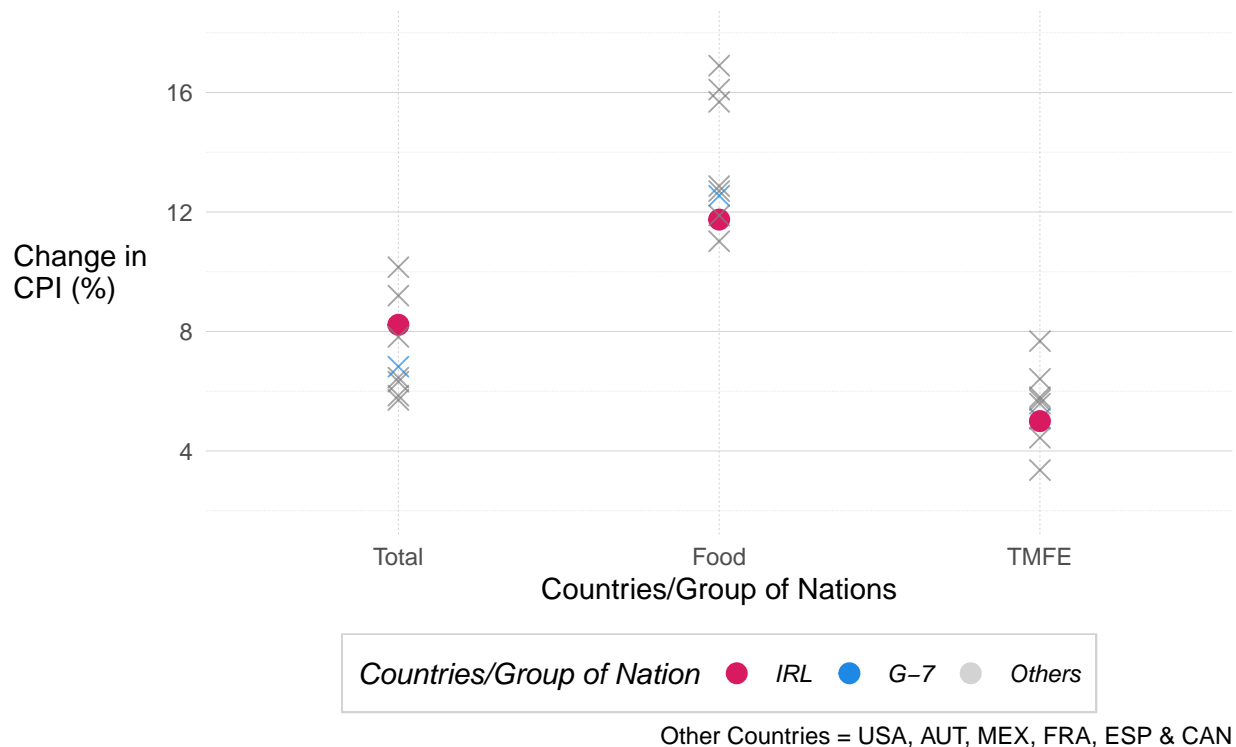
      plot.title = element_text(face = "bold", size = 14),
      plot.subtitle = element_text(face = 'italic', size = 12))

```

dotplot

## Consumer Price Index (CPI) Comparison

*Change in December 2022 since December 2021*



### 4. Creating a Bar Chart

Using the same colour scale we used in the dot plot above, our country IRL was given a pink shade, G-7 the blue shade, and the rest with grey shade. The alpha values were modified to keep our country at the forefront and solid in the chart, while keep everything else in the background as light as possible. The bar chart is grouping descending by the Percentage (Per\_2022\_12) values. Major change from Part 1 is that I am using Subject column in the X-axis as opposed to Location column. I have kept only the Major and Minor Y-axis gridlines to keep this chart minimal. To keep our country in foreground, the alpha value was set lower for other countries. I decided to use Subject in the X-axis as that is something which can be grouped/reordered by the Percentage (Per\_2022\_12) attribute. Also the captions mentions the other countries marked by the grey colour in the chart. I have also changed the Y-axis title to rotate it, so that we can see it without any neck tilt. And finally, the legend was taken below the X-axis, in the center, to save space and also to make the graph bigger in size by the width.

This is a much different graph from Part 1 as it focuses on our country when all the measures were separated on X-axis to show descending values of all the countries when compared to our country. This is definitely a better chart in terms of pattern-finding and appeal, while this becomes very easy to read chart even for a non-related person of the field who can infer basic things from this graph, and that was the goal to plotting a graph in Part 2. This approach is way beneficial if compared even to the dot plot as this shows better readability and less clutter.

```
library(RColorBrewer)
library(ggplot2)
```

```

library(dplyr)

Others <- c("GBR","USA","AUT","MEX","FRA","ESP","CAN")

barplot_new <- ggplot(plot_data, aes(x = Subject, y = Per_2022_12)) +
  geom_col(position = "dodge2", width = 0.9,
    aes(group = -Per_2022_12, fill = Location, alpha = Location == "IRL"))+
  scale_fill_manual("Countries/Group of Nation",
    values = c("IRL" = "#D81B60", "G-7" = "#1E88E5", Others = "grey"), aes(group = Per_2022_12))
  scale_alpha_manual(values = c(0.38, 1), guide = "none")+
  labs(title = "Consumer Price Index (CPI) Comparison",
    subtitle = "Change in December 2022 since December 2021",
    caption = "Other Countries = USA, AUT, MEX, FRA, ESP & CAN") +
  xlab("Subject") +
  ylab("Change in \nCPI (%)") +
  scale_y_continuous(breaks = seq(0,20, by = 4), limits = c(0,18), expand = c(0,0)) +
  theme(panel.background = element_rect(fill = "white"),
    panel.grid.major.y = element_line(size = 0.12, linetype = "solid",
      colour = 'lightgrey'),
    panel.grid.minor.y = element_line(size = 0.068, linetype = "dashed",
      colour = 'lightgrey'),
    axis.ticks = element_blank(),
    axis.text = element_text(size = 9),
    axis.title = element_text(size = 11),
    axis.title.y = element_text(size = 11,
      angle = 0,
      vjust=0.5,
      hjust=0,
      margin = margin(r=4)),
    legend.title = element_text(face="italic", size=11),
    legend.text = element_text(face="italic", size = 9),
    legend.box.background = element_rect(fill = 'lightgrey', colour = 'lightgrey'),
    legend.key = element_rect(fill = "white"),
    legend.position = "bottom",
    legend.direction = "horizontal",
    legend.justification = "center",
    legend.box.margin = margin(0.5, 0.5, 0.5, 0.5),
    plot.title = element_text(face = "bold", size = 14),
    plot.subtitle = element_text(size = 12),
    panel.border = element_blank(),
    panel.spacing = unit(1, "lines"))

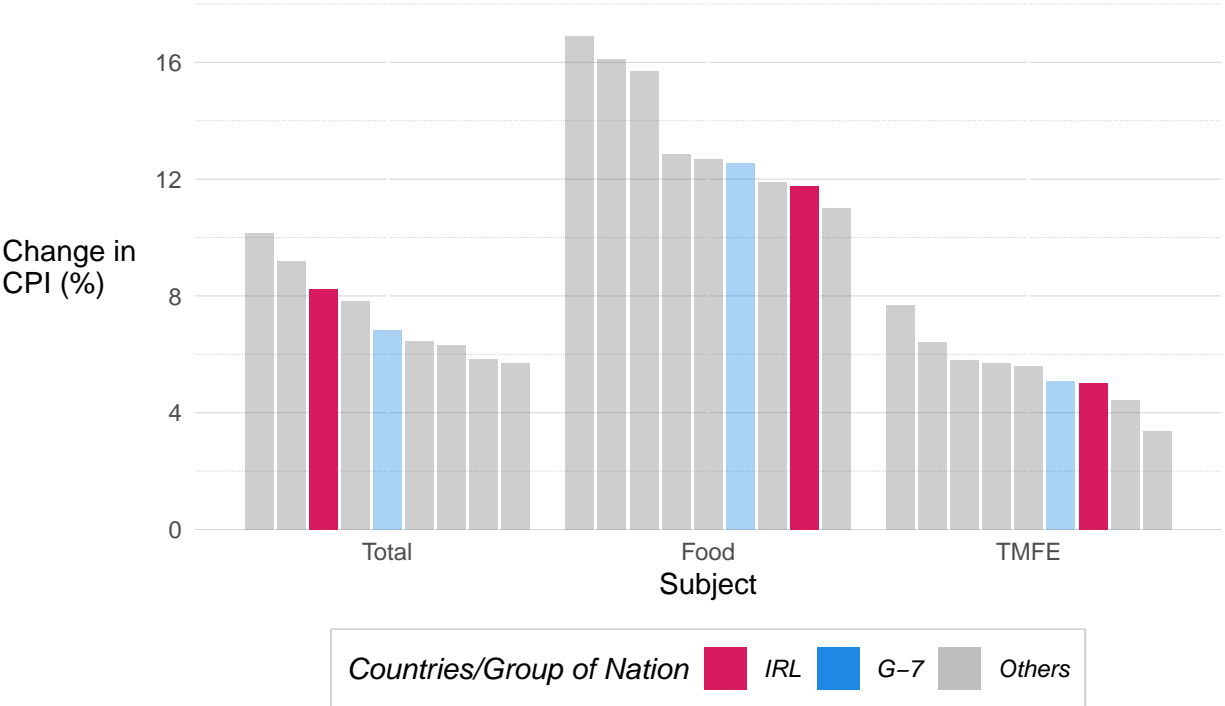
barplot_new

```



# Consumer Price Index (CPI) Comparison

Change in December 2022 since December 2021



Other Countries = USA, AUT, MEX, FRA, ESP & CAN