Assingment-1 Part 2

Smitesh Patil

2023-02-13

```r
library(ggplot2)
library(dplyr)
library(gghighlight)
library(forcats)
df <- read.csv("Counsumer_Price_Index.csv")

#renaming the column Location
df <- df %>%
  rename(Location = ï..Location)


sav <- df %>%
  #selecting 8 countries and 1 group of nations
  filter(Location %in% c("IRL", "CAN", "JPN",
                         "USA", "FRA", "DEU",
                         "ESP", "ITA", "NLD", "OECD")) %>%
  # filtering data by time
  filter(Time == "2022-12") %>%
  #changing country codes to names for readability
  mutate(Location = recode(Location,
                           "IRL" = "IRELAND",
                           "JPN" = "JAPAN",
                           "USA" = "UNITED STATES",
                           "DEU" = "GERMANY",
                           "CAN" = "CANADA",
                           "ESP" = "SPAIN",
                           "ITA" = "ITALY",
                           "FRA" = "FRANCE",
                           "NLD" = "NETHERLANDS")) %>%
  #removing unwanted data
  filter(Subject != "Energy")

# creatign a order for levels attributes used to other the plot
order <- sav %>%
      filter(Subject == "Food") %>%
      arrange(Percentage) %>%
      select(Location) %>%
      unlist() %>%
      unname()

order <- replace(order, c(1,2), order[c(2,1)])
```
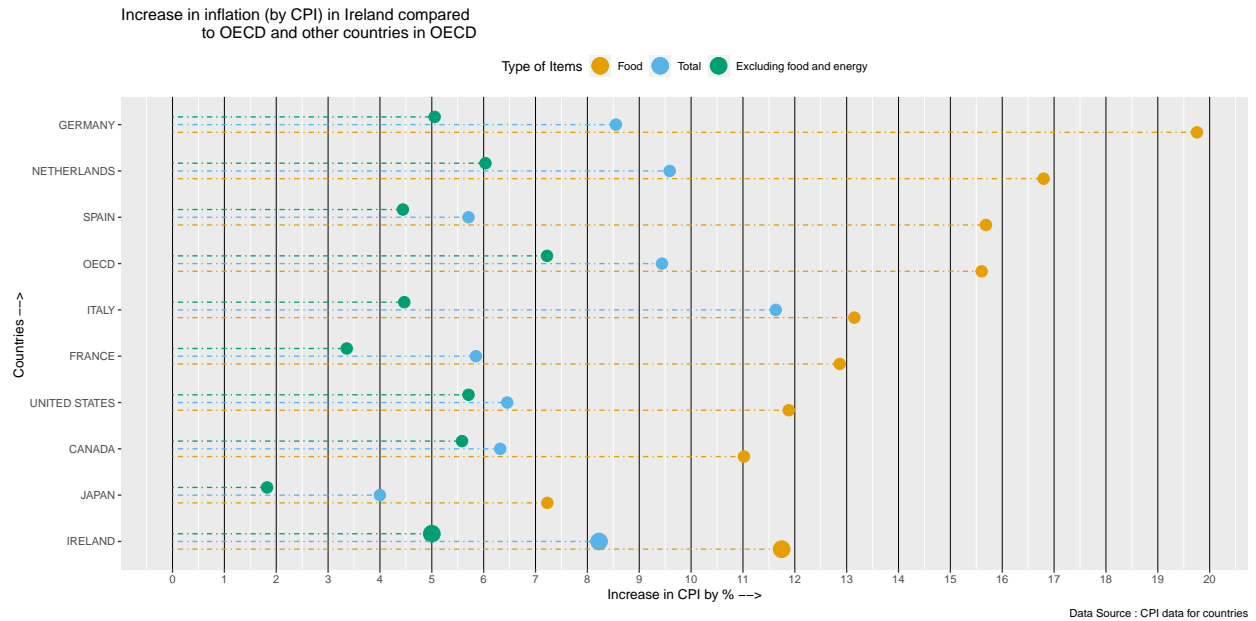
1. Implement the dot plot approach using the ggplot2 library

```r
# codes for colorblind colors
colorblind_colors <- colorBlindness::availableColors()

order <- c( "IRELAND", "JAPAN", "CANADA" ,"UNITED STATES", "FRANCE", "ITALY",
"OECD", "SPAIN", "NETHERLANDS", "GERMANY")

sav %>%
  # adding order to the plot
  mutate(Location = factor(Location, levels = order)) %>%
  # ggplot intialize aesthesis for x,y axis and coloring
  ggplot(aes(x = Location, y = Percentage, color = Subject)) +
  #for countries other than Ireland dot plot with size 4
  geom_point(data = ~ mutate(., Percentage = if_else(Location == "Ireland",
             Percentage[NA], Percentage)),
             size = 4, position = position_dodge(0.5), na.rm = TRUE) +
  #for Ireland dot plot with size 6
  geom_point(data = ~ filter(., Location == "IRELAND"),
             size = 6, position = position_dodge(0.5)) +
  #for the dotted line from y-axis
  geom_linerange(data = ~ filter(., Location == "IRELAND"),
  aes(ymin = 0, ymax = Percentage), position = position_dodge(0.5),linetype = "dotdash") +
  geom_linerange(data = ~ filter(., Location != "IRELAND"),
                 aes(ymin = 0, ymax = Percentage),
                 position = position_dodge(0.5), linetype = "dotdash") +
  #flipping axes
  coord_flip()+
  #labels for data
  labs(caption = "Data Source : CPI data for countries")+
  xlab("Countries --> ") +
  ylab("Increase in CPI by % -->")+
  # y axis scale with 1 unit of gap
  scale_y_continuous(breaks = round(seq(0, 20, by = 1),1))+
  # adding defined colors
  scale_color_manual(name = "Type of Items",
  labels = c("Food", "Total", "Excluding food and energy"),
  values=c(unname(colorblind_colors[2]),
           unname(colorblind_colors[3]),
           unname(colorblind_colors[4])))+
  #title for the graph
  ggtitle(label = "Increase in inflation (by CPI) in Ireland compared
                   to OECD and other countries in OECD")+
  # accentuate the gridline
  theme(panel.grid.major.x = element_line(linewidth =.01, color="black"),
  panel.grid.major.y = element_blank(),
  legend.position = "top"
)
```

Type of Items ● Food ● Total ● Excluding food and energy

GERMANY
NETHERLANDS
SPAIN
OECD
ITALY
FRANCE
UNITED STATES
CANADA
JAPAN
IRELAND

Countries --->

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Increase in CPI by % -->

Data Source : CPI data for countries

2. Implement the bar plot using the ggplot2 library

```
sav %>%
  # ggplot initialize aesthetics for x,y axis
  ggplot(aes(x = factor(Location, levels = order), y = Percentage)) +
  # columns/bars for Ireland
  geom_col(data = filter(sav, Location == "IRELAND"), fill = colorblind_colors[2],
           linewidth=0.2, position="dodge")+
  # columns/bars for OECD
  geom_col(data = filter(sav, Location == "OECD"), fill = colorblind_colors[3],
           linewidth=0.2, position="dodge")+
  # columns/bars for other countries
  geom_col(data = filter(sav, !Location %in% c("OECD", "IRELAND")),
           fill = colorblind_colors[4] ,linewidth=0.2, position="dodge")+
  # labels for axis
  xlab("Countries --> ") +
  ylab("Increase in CPI by % -->")+
  #faceting the plot
  facet_grid(~Subject,
             labeller = as_labeller(c("Food" = "Food",
                                      "Total" = "Total",
                                      "Total_Minus_Food_Energy" = "Excluding Food & Expenses")))+
  #flipping labels
  coord_flip()+
  labs(fill = "Country / Organization", caption = "Data Source : CPI data for countries")+
  # y axis scale with 1 unit of gap
  scale_y_continuous(breaks = round(seq(0, 20, by = 1),1))+
  scale_fill_discrete(name = "Country / Organization",
                      labels = c("OECD", "Other countries in OECD", "IRELAND")) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1),
        legend.position = "top",)+
  # title for the graph
  ggtitle(label = "Increase in inflation (by CPI) in Ireland compared
                   to OECD and other countries in OECD")+
```

```
theme_light()
```


Increase in inflation (by CPI) in Ireland compared to OECD and other countries in OECD
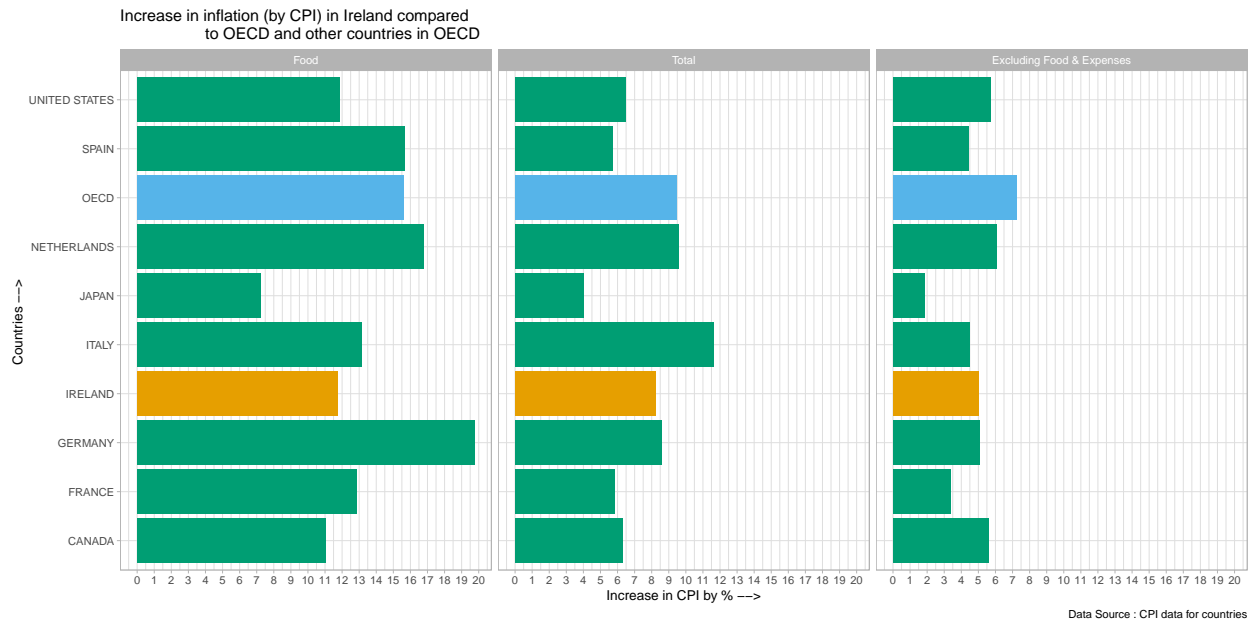
3. Comment upon changes you have made since part 1

Answer:-

Since the review lecture on assigment 1, I concluded that my approach with dot-plot was not correct for two reasons.

First, the dots in the dot plot were not able to differentiate between the countries other than Ireland and the group of nations (OECD).

Second, information was had to interpret as each individual subject category(Food, Total, Excluding food and energy) looked like cluster.

In conclusion, the dot-plot was hard to interpret and it was difficult for the eyes to find patterns.Hence, I have went with a lollipop plot for the final draft. Here, I have sorted the data based on the food cpi value. Then I have kept Ireland separately from others at the bottom. I have also added emphasis on Ireland by making the dot bigger in size for Ireland.Futhermore, I have added a dotted line that would aid the viewer in measuring the value of percentages.

I have sticked with the format of bar-plot that I decided on Part 1. I think coloring Ireland and the group of nations (as group of nations is a separate entity compared to other nations in the plot) different make them visually stand out from other nations. Faceting the category of commodity for CPI values allows for interpretation of inflation for diffrent categories at a time. Keeping the countries/ group of nations on y-axis allows for better comparison amongst them.