CT5100 Data Visualisation: Assignment 1

Jiarong Li 20230033, 1MAI1

The School of Computer Science
University of Galway
j.li11@nuigalway.ie

1 Part 1

- 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.
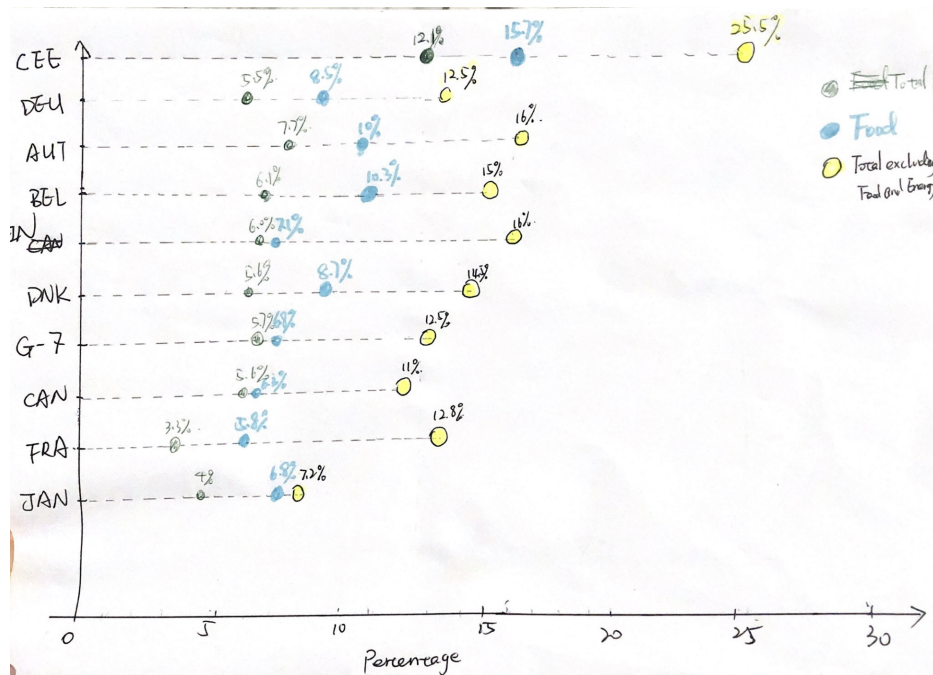

Figure 1: The ordering grouped dot chart of the latest CPI over the following 3 measures: Total, Food and Total excluding Food and Energy.

- Produce a colour drawing* of an equivalent bar chart approach that visualises the same information.
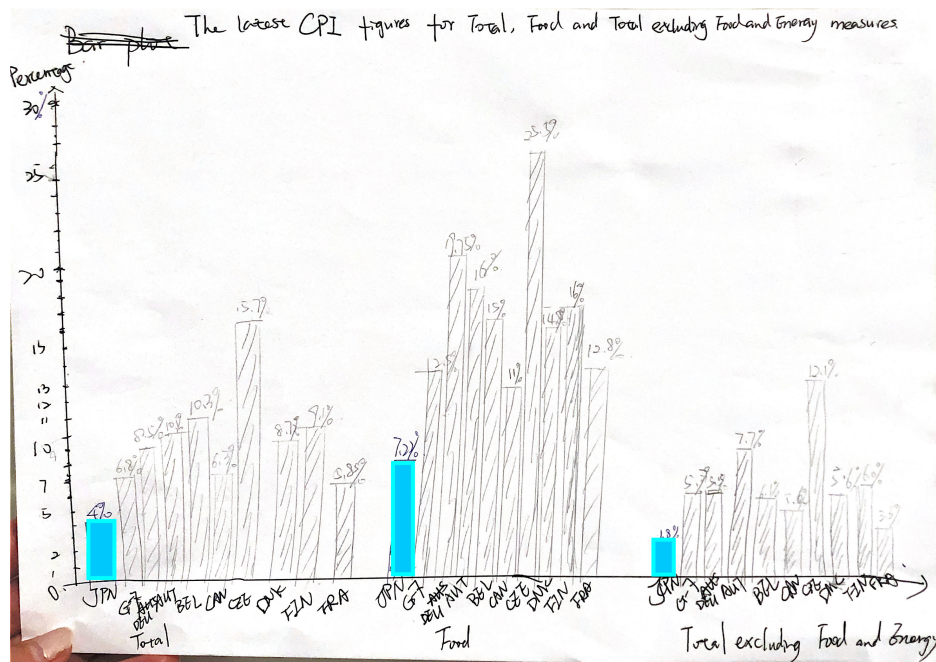
Figure 2: The bar chart of the latest CPI figures for the following 3 measures: Total, Food and Total excluding Food and Energy.

- 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)

Consider the case where we want to show the increase in inflation in a business article. Our audience is the specific businessman. The goal of our visualisation is to discuss the increase in inflation. The information to be communicated is 1) show the range of the increase or decrease in the price for a most recent basket of goods and 2) emphasise the high increase values. The data variables are country names, CPI percentages, and category values which represent this information. There are no statistical summaries required. The aesthetics and scales we use are the position and colour. The graph type we are going to use is a bar chart or a dot chart.

We can visualise the data by using an ordered grouped dot chart. In the dot chart, we set the percentage value on the x-axis and countries name on the y-axis. The colour of the points will be defined by the categorical values (Total, Food and Total excluding Food and Energy). We also use a geom line with the aesthetic group and map it to the name value, which tells ggplot to consider all points with the same name value as a separate group. We could also have made a line with the geom segment geom to clearly show the country associated with the points. We order the country names on the y-axis by calculating the mean of the CPI per country over three measures (Food, Total and Total excluding Food and Energy).

We can also visualise the data by using a grouped bar chart. In a grouped bar plot, we draw a group of bars at each measure category (Food, Total and Total excluding Food and Energy) position on the x-axis. The bars in each group are determined by the countries of the category variable. We can clearly show the inflation changes in each country per measure category by using this grouping way because we care more about the

2

overall pattern per measure over countries. In addition, we care more about the inflation changes in our own country and the comparison of inflation between our own country and other highest or lowest countries. Therefore, we only fill the bar representing our country with blue and fill the other bars with grey to foreground the inflation of our own country.

In this case, a dot chart is more effective in communicating the information. Because 1) the bar chart relies on the length of the bar measured from the zero point and it's not useful when the visualising values have a small but important difference. 2) dot chart can solve this problem as each dot represents a category and associates with the position where the scale does not need to begin at zero. The values of the data we visualise have a small but important difference. Therefore, we choose a dot plot to visualise the data.

2 Part 2

- Implement the dot plot approach using the ggplot2 library

```r
library(dplyr)

countries_dot_plot <- read.csv("Counsumer_Price_Index.csv") %>%
  # select countries and a group for comparison.
  filter(Location %in% c("JPN", "G-7", "DEU", "AUT", "BEL", "CAN",
    "CZE", "DNK", "FIN", "FRA")) %>%
  # select categories for grouping
  filter(Subject %in% c("Total", "Food", "Total_Minus_Food_Energy"))
    %>%
  # select the month
  filter(Time=="2022-12")

# select each country with the mean value of categories requested.
mean_percentage <- countries_dot_plot %>%
  group_by(Location) %>%
  summarise(count=n(),
            avrpercentage = mean(Percentage, na.rm=T)) %>%
  arrange(-avrpercentage)

# locations sorted
locationorder <- mean_percentage$Location

# We turn location into a factor, with levels in the order of
  location order
# This ordering determines the order of the locations presented on
  the y-axis
countries_dot_plot$Location <- factor(countries_dot_plot$Location,
  levels=locationorder)

```

```r
25    # define a highlight_df to store JPN and G-7
26    highlight_df <- countries_dot_plot %>%
27                    filter(Location == 'JPN' | Location == 'G-7')
28
29    # plot
30
31    library(ggplot2)
32
33    theme_set(theme_classic())
34
35    # Now ggplot will display the lacations in the order specified by
      ↪   the factor
36    ggplot(countries_dot_plot,
37           aes(x=Percentage, y=Location)) +
38      geom_segment(aes(yend = Location), xend = 0,
39                   colour="grey", size=0.1, linetype="dotted") +
40      geom_segment(data = highlight_df, aes(yend = Location), xend = 0,
      ↪   colour="grey", size=0.5, linetype="dotted") +
41      geom_point(size = 2.5, aes(colour=Subject)) +
42      # geom_point(data = highlight_df, aes(colour=Subject), size = 3) +
43      scale_colour_manual(values=c("#ce5a6c","#212f85","#5acec0"),
      ↪   name="") +
44      scale_x_continuous(limits = c(0, 26),
45                         expand = c(0, 0),
46                         breaks = seq(2.5, 25.5, by = 2.5),
47                         name = "percentage") +
48      labs(title="The latest CPI of JPN, G-7 and other 8 countries in
      ↪   Total, Food and Total excluding Food and Energy measures",
      ↪   caption="data source: CT5100 Data Visualisation") +
49      theme(axis.title.y = element_blank(),
50            panel.grid.major.x = element_line(size=0.07, colour="grey60",
51                                              linetype="dotted"),
52            panel.grid.minor.x = element_blank(),
53            panel.grid.major.y = element_blank(),
54            axis.line.y = element_blank(),
55            axis.line.x = element_blank(),
56            axis.ticks.y = element_blank(),
57            axis.ticks.x = element_blank(),
58            legend.position = c(0.85, 0.92),
59            legend.text = element_text(size=7),
60            legend.key.size = unit(0.6, "lines"),
61            legend.title = element_blank(),
62            plot.title = element_text(size=10),
63            plot.title.position = "plot",
64            plot.caption = element_text(size = 7, color = "black",
      ↪   face="italic"),
```

```
65        plot.margin = margin(t=3,
66                              r=15,
67                              b=3,
68                              l=3,
69                              unit="pt"))
70
```

The latest CPI of JPN, G–7 and other 8 countries in Total, Food and Total excluding Food and Energy measures

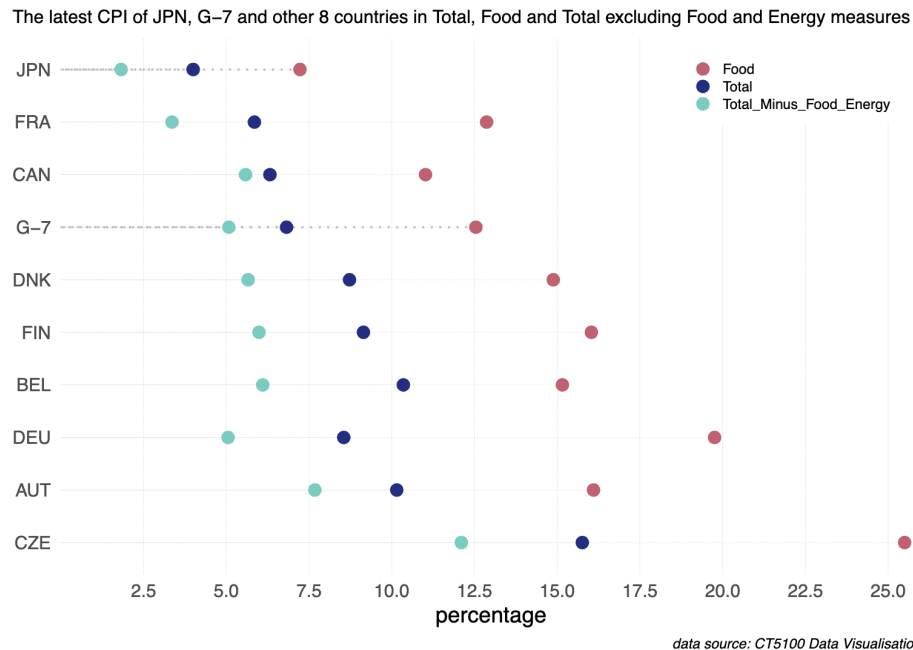data source: CT5100 Data Visualisation

Figure 3: The dot plot of the latest CPI of JPN, G-7 and other 8 countries in Total, Food and Total excluding Food and Energy measures.

- Implement the bar plot using the ggplot2 library

```
1   library(tidyr)
2
3   # convert data from long form to wide form
4   wide_countries_dot_plot <- countries_dot_plot %>%
5     pivot_wider(names_from=Subject, values_from=Percentage)
6
7   # order the location categories so that they display in descending
    ↪  order based on the value of percentage in the category.
8   location_sum <- rowSums(wide_countries_dot_plot[,c(-1, -2, -3)])
9   location_order_bar <- order(location_sum, decreasing = T)
10
11  # this is the order the categories will be displayed in
12  location_order_name <-
    ↪  levels(c(wide_countries_dot_plot[,1])[location_order_bar] $
    ↪  Location)
```

```
13
14   # the location field is made into a factor with the ordered levels
     ↪   calculated in the previous lines
15   countries_dot_plot$Location <- factor(countries_dot_plot$Location,
     ↪   levels = location_order_name)
16
17   # plot
18   library(ggplot2)
19   library(colorspace)
20
21   ggplot(countries_dot_plot, (aes(x=Subject, y=Percentage,
     ↪   fill=Location))) +
22     geom_col(colour="grey85", size=0.2, position="dodge", alpha=0.85) +
23     scale_y_continuous(
24       breaks = seq(0, 27, by=3),
25       limits = c(0, 27),
26       expand = c(0, 0),
27       name = "Percentage") +
28     xlab(label = "2022-12") +
29     scale_fill_manual(values=c("lightgrey","lightgrey","lightgrey",⌋
     ↪   "lightgrey","lightgrey","lightgrey","#ce5a6c","lightgrey",⌋
     ↪   "lightgrey","#212f85"), name="Location")
     ↪   +
30     scale_x_discrete(labels=c("Food", "Total",
     ↪   "Total_Minus_Food_Energy")) +
31     labs(title="The latest CPI of JPN, G-7 and other 8 countries in
     ↪   Total, Food and Total excluding Food and Energy measures",
     ↪   captions = "Data source: CT5100 Data Visualisation") +
32
33     theme_bw() +
34
35     theme(
36       axis.line.y = element_blank(),
37       axis.ticks.y = element_blank(),
38       axis.line.x = element_blank(),
39       axis.ticks.x = element_blank(),
40       axis.title.x = element_text(size=12),
41       axis.title.y = element_text(size = 12,
42                                   angle = 0,
43                                   vjust = 0.5,
44                                   hjust = 0,
45                                   margin = margin(r=4)),
46       axis.text = element_text(size=10),
47       panel.grid.major.y = element_line(size=0.1,
48                                   linetype='solid',
49                                   colour='lightgrey'),
```

```
50    panel.grid.minor.y = element_blank(),
51    panel.grid.major.x = element_blank(),
52    panel.grid.minor.x = element_blank(),
53
54    legend.position = "bottom",
55    legend.direction="horizontal",
56    legend.text = element_text(size=6),
57    legend.key.size = unit(0.9, "lines"),
58    legend.title = element_text(size = 10),
59    plot.caption = element_text(size = 7, color = "black",
      ↪ face="italic"),
60    plot.title = element_text(size=9, vjust = 6),
61    plot.title.position = "plot",
62    plot.margin = margin(t=15,
63                         r=15,
64                         b=3,
65                         l=3,
66                         unit="pt"),
67    axis.line = element_blank(),
68    axis.ticks = element_blank(),
69    panel.border = element_blank()
70  ) +
71  # display legend in one row
72  guides(fill = guide_legend(nrow = 1))
```
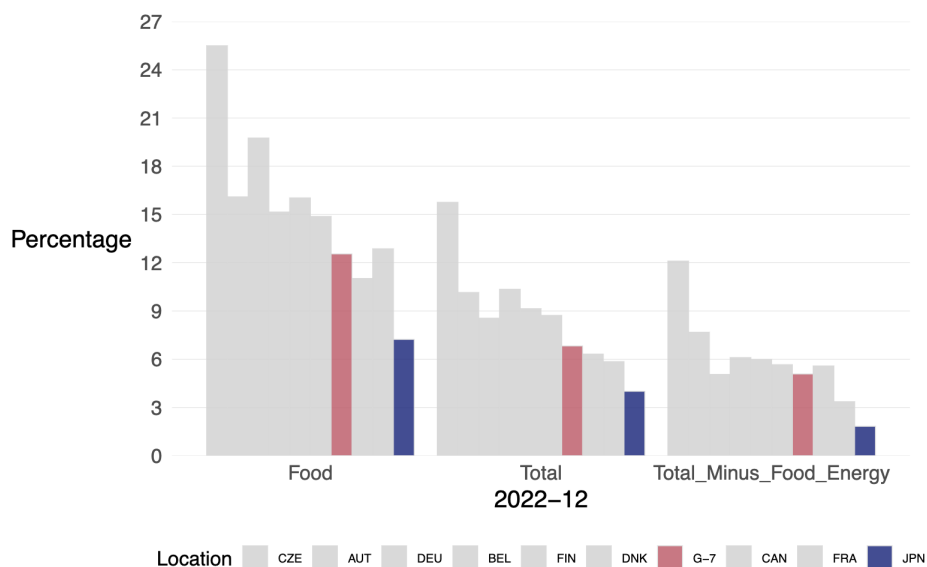

Figure 4: The bar plot of the latest CPI of JPN, G-7 and other 8 countries in Total, Food and Total excluding Food and Energy measures.

7

- Comment upon changes you have made since part 1 The communication features I would like to deliver is the comparison of JPN, G-7 and other 8 countries.

  The changes I made on the dot plot above are: 1) ascending the order of locations on the y-axis because I want to emphasise the inflation changes of JPN and ascending order can help me to place the JNP at the first place on the y-axis, 2) I highlight the geom_segment of JPN and G-7 as JNP is the target country and G-7 is the only one of the group of nations, therefore I make the stronger dot lines connected to these two locations, 3) I add the title of the dot plot.

  The changes I made on the bar plot above are: 1) I order the locations by calculating the sum values of each location, and group the locations with the order by three categories, 2) I remove the number which is on the top of each bar in the previous version of the bar plot as I want to display the inflation changes instead of the exact number of inflation, 3) I highlight both JPN and G-7 in this case and make the colours of the other 8 locations light grey to foreground the inflation of JPN and G-7, 4) I put the legend in the bottom within one row horizontally corresponding to the order of locations per category to ensure deliver the information clearly with least ink.