Assignment 2

Deconstruct, Reconstruct Web Report

Phuong Vo Vinh Pham

Assessment declaration checklist

Please carefully read the statements below and check each box if you agree with the declaration. If you do not check all boxes, your assignment will not be marked. If you make a false declaration on any of these points, you may be investigated for academic misconduct. Students found to have breached academic integrity may receive official warnings and/or serious academic penalties. Please read more about academic integrity here (https://www.rmit.edu.au/students/student-essentials/assessment-and-exams/academic-integrity). If you are unsure about any of these points or feel your assessment might breach academic integrity, please contact your course coordinator for support. It is important that you DO NOT submit any assessment until you can complete the declaration truthfully.

By checking the boxes below, I declare the following:

- ☑ I have not impersonated, or allowed myself to be impersonated by, any person for the purposes of this assessment
- ☑ This assessment is my original work and no part of it has been copied from any other source except where due acknowledgement is made. Due acknowledgement means the following:
    - The source is correctly referenced in a reference list
    - The work has been paraphrased or directly quoted
    - A citation to the original work's reference has been included where the copied work appears in the assessment.
- ☑ No part of this assessment has been written for me by any other person except where such collaboration has been authorised by the lecturer/teacher concerned.
- ☑ I have not used generative "AI" tools for the purposes of this assessment.
- ☑ Where this work is being submitted for individual assessment, I declare that it is my original work and that no part has been contributed by, produced by or in conjunction with another student.
- ☑ I give permission for my assessment response to be reproduced, communicated, compared and archived for the purposes of detecting plagiarism.

- ☑ I give permission for a copy of my assessment to be retained by the university for review and comparison, including review by external examiners.

I understand that:

- ☑ Plagiarism is the presentation of the work, idea or creation of another person or machine as though it is your own. It is a form of cheating and is a very serious academic offence that may lead to exclusion from the University. Plagiarised material can be drawn from, and presented in, written, graphic and visual form, including electronic data and oral presentations. Plagiarism occurs when the origin of the material used is not appropriately cited.
- ☑ Plagiarism includes the act of assisting or allowing another person to plagiarise or to copy my work.

I agree and acknowledge that:

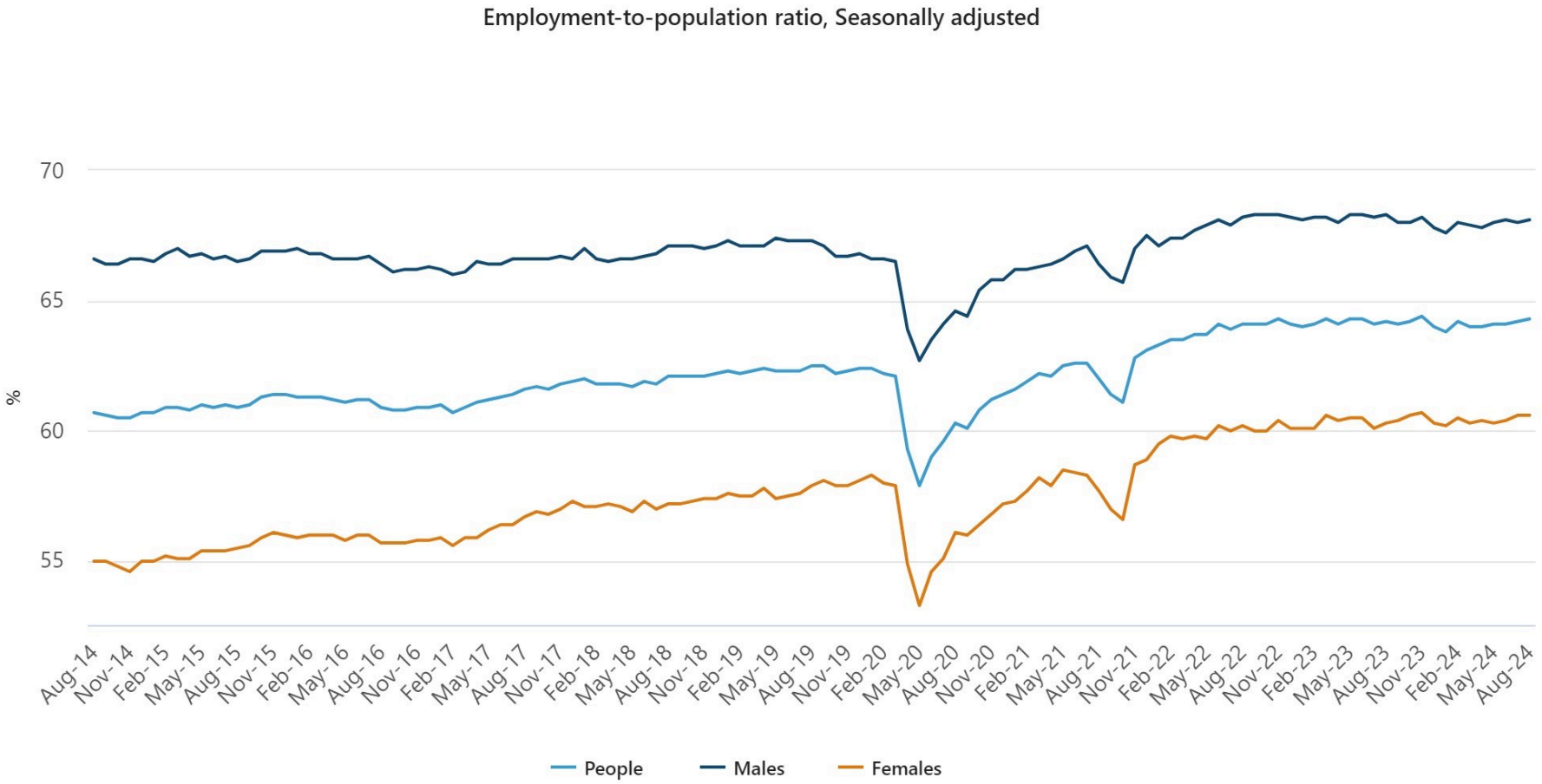
- ☑ I have read and understood the Declaration and Statement of Authorship above.
- ☑ If I do not agree to the Declaration and Statement of Authorship in this context and all boxes are not checked, the assessment outcome is not valid for assessment purposes and will not be included in my final result for this course.

This is a template file. The following example included is not considered a good example to follow for Assignment 2. Remove this warning prior to submitting.

Deconstruct

Original

The data visualization below is collected from Australian Bureau of Statistics website for Labour Force:



Source: Australian Bureau of Statistics, Labour Force, Australia August 2014.

Objective and Audience

Objective

The primary objecttive corresponding to the chosen visualization is to display the employment ratio of Australian population starting from 2014 to 2024. There are three variables including the percentage for male, female and the employment rate of the whole population.

Using line chart as the means of visualization, it emphasizes on how the employment ratio change within the specified ten-year period to record the significant trend or pattern. Furthermore, the graph also compare the employment percentage between two genders and the general population of Australia.

It can be seen that during the year 2020, there was a significant drop in the statistics for all the associated groups. The data lines then slowly climb up the ladder to reach the initial peaks after the start of the year 2022 and is maintaining until present. notably, the visualization does not provide any forecast but focuses on examining the historical trend. and patterns.

As a result, the data visualization is looking forward to assisting in contributing to the policy decision-making procedures, especially in explain the potential gender disparity in labor market or enacting specific employment programs.

Audience

The target audience of this analysis involves a wide range of stakeholders from different fields and sectors. For instance, policymakers or governors can utilise this data to generate insights supporting them improve the overall ratios through certain programs or encourage the employability of female groups. Business' Executive Board as well as Human Resource department may find it is useful in enhancing their recruitment phase while introducing strategies to adapt to the employment trend. Moreover, since employment can affect the economic performance of a city, state, or even a whole country, economists are responsible to study the long-term prediction based on the data visualization to avoid the potential event impacting the dynamics of economic situation.

Critique

Unfortunately, the visualization chosen encounters three issues that are required further substantial improvement to achieve a more precise message:

- *Deceptive Method:*

Even though the graph has shown the trend of employment ratios in Australia after 10 years for each months, the visualization only shows the statistic for only some of the months within each year. The missing data may cause misleading impression to audience since there might be some possibility that those missing month may contains meaningful information. Also, the visualization can be considered lack informative as the actual number of people employed is not mentioned.

In this situation, the data will be summarize to get the average annual employment ratios. It will represent the overall performance of that accounting year. Secondly, an additional dataset containing the total population with the status "Employed" is joined with the original one to give further insights.

- *Color issue:*

Although each groups has been assigned with different color to better distinction, the lines' colors for male group and population are both in blue palette, which may lead to difficulty in differentiating groups. As a result, the color for the lines are modified with numerous options. Saturated ones are specifically avoided.

- *Perceptual problem:*

The main title of the graph mentions the “seasonally adjusted”, yet there is no variable to indicate if there is any modification in the employment ratios between groups while it does not refer to the data per genders. Besides, since there are too many observation on the x-axis, it is struggle to keep track with the data without the vertical projections, and the y-axis is not provided appropriately. The grid lines, axes as well as the main label of the graphs will be included in order to get rid of the perceptual issue.

Reconstruct

Code

```
#Packages required
library(magrittr)
library(knitr)
library(dplyr)
library(readr)
library(tidyr)
library(ggplot2)
```

Here below is the original dataset for the employment ratios. Because the data was not formatted properly with inappropriate first and last rows, it must be cleaned and tidied before applying to the visualization.

```
#Dataset1
emp_df <- read_csv("C:/Users/ASUS/Desktop/Employment-to-population ratio, Seasonally adjusted.csv")
head(emp_df)
```

```
## # A tibble: 6 × 1
##   `Employment-to-population ratio, Seasonally adjusted`
##   <chr>
## 1 ",\"People (%)\",\"Males (%)\",\"Females (%)\""
## 2 "Aug-14,60.7,66.6,55.0"
## 3 "Sep-14,60.6,66.4,55.0"
## 4 "Oct-14,60.5,66.4,54.8"
## 5 "Nov-14,60.5,66.6,54.6"
## 6 "Dec-14,60.7,66.6,55.0"
```

```
#Data Cleaning
names(emp_df)[1] <- "merged"
emp_df <- emp_df %>% separate(col = merged,
                              into = c("Period", "People (%)", "Males (%)", "Females (%)"),
                              sep = ",")

emp_df <- emp_df[-c(1), ] #Remove 1st row
emp_df <- emp_df[-c(122), ] #Remove last row
head(emp_df)
```

```
## # A tibble: 6 × 4
##   Period `People (%)` `Males (%)` `Females (%)`
##   <chr>  <chr>        <chr>       <chr>
## 1 Aug-14 60.7         66.6        55.0
## 2 Sep-14 60.6         66.4        55.0
## 3 Oct-14 60.5         66.4        54.8
## 4 Nov-14 60.5         66.6        54.6
## 5 Dec-14 60.7         66.6        55.0
## 6 Jan-15 60.7         66.5        55.0
```

Then, the additional dataset is imported to solve the potential deceptive method specified in the previous subsection. Encountering the similar issue happened to the first source, this dataset is required to go through cleaning and tidying procedures.

```
#Additional Dataset2
emp_ppl <- read.csv("C:/Users/ASUS/Desktop/Employed people.csv", sep = ";")
head(emp_ppl)
```

```
##                          Employed.people..
## 1 ,Trend ('000),Seasonally adjusted ('000)
## 2                 Aug-14,11546.00,11549.20
## 3                 Sep-14,11552.90,11546.80
## 4                 Oct-14,11563.80,11542.20
## 5                 Nov-14,11579.00,11542.20
## 6                 Dec-14,11598.60,11597.30
```

```
#Data Cleaning
names(emp_ppl)[1] <- "merged"
emp_ppl <- emp_ppl %>% separate(col = merged,
                                into = c("Period", "Trend ('000)", "Seasonally adjusted ('000)"),
                                sep = ",")

emp_ppl <- emp_ppl[-c(1), ] #Remove 1st row
emp_ppl <- emp_ppl[-c(122:126), ] #Remove last rows
head(emp_ppl)
```

```
##   Period Trend ('000) Seasonally adjusted ('000)
## 2 Aug-14     11546.00                   11549.20
## 3 Sep-14     11552.90                   11546.80
## 4 Oct-14     11563.80                   11542.20
## 5 Nov-14     11579.00                   11542.20
## 6 Dec-14     11598.60                   11597.30
## 7 Jan-15     11622.00                   11609.90
```

After cleaning to receive proper datasets, those are joined together to form a new set that contains the preferred information of number of employed people.

```
#Joining Dataset 1 & 2
emp_ppl <- emp_ppl %>% select("Period", "Seasonally adjusted ('000)")
emp_ppl <- emp_ppl %>% rename("Population" = "Seasonally adjusted ('000)")
emp_rate <- left_join(emp_df, emp_ppl, by = "Period")
head(emp_rate)
```

```
## # A tibble: 6 × 5
##   Period `People (%)` `Males (%)` `Females (%)` Population
##   <chr>  <chr>        <chr>       <chr>         <chr>
## 1 Aug-14 60.7         66.6        55.0          11549.20
## 2 Sep-14 60.6         66.4        55.0          11546.80
## 3 Oct-14 60.5         66.4        54.8          11542.20
## 4 Nov-14 60.5         66.6        54.6          11542.20
## 5 Dec-14 60.7         66.6        55.0          11597.30
## 6 Jan-15 60.7         66.5        55.0          11609.90
```

Eventually, the newly-created data frame will be subseted for further visualizing process.

```
#Data Transformation
emp_sub <- emp_rate %>%
  pivot_longer(cols = c("People (%)", "Males (%)", "Females (%)", "Population"),
               names_to = "Groups",
               values_to = "Ratios") %>%
  separate(Period, into = c("Month", "Year"), sep = "-")

emp_sub$Ratios <- as.numeric(emp_sub$Ratios) #Convert to numeric
head(emp_sub, 8)
```

```
## # A tibble: 8 × 4
##   Month Year  Groups       Ratios
##   <chr> <chr> <chr>         <dbl>
## 1 Aug   14    People (%)     60.7
## 2 Aug   14    Males (%)      66.6
## 3 Aug   14    Females (%)    55
## 4 Aug   14    Population  11549.
## 5 Sep   14    People (%)     60.6
## 6 Sep   14    Males (%)      66.4
## 7 Sep   14    Females (%)    55
## 8 Sep   14    Population  11547.
```

```
#Summarize to annually data
emp_sub2 <- emp_sub %>%
  arrange(Year, Groups) %>%
  group_by(Year, Groups) %>%
  summarize(Ratios = round(mean(Ratios, na.rm = TRUE),2))

#Create Facet Groups
emp_sub3 <- emp_sub2 %>%
  mutate(Facet_Group = ifelse(Groups == "People (%)" | Groups == "Population",
                              "Overall Population",
                              "Genders")) %>%
  mutate(Facet_Group = factor(Facet_Group, levels = c("Overall Population", "Genders")))
head(emp_sub3)
```

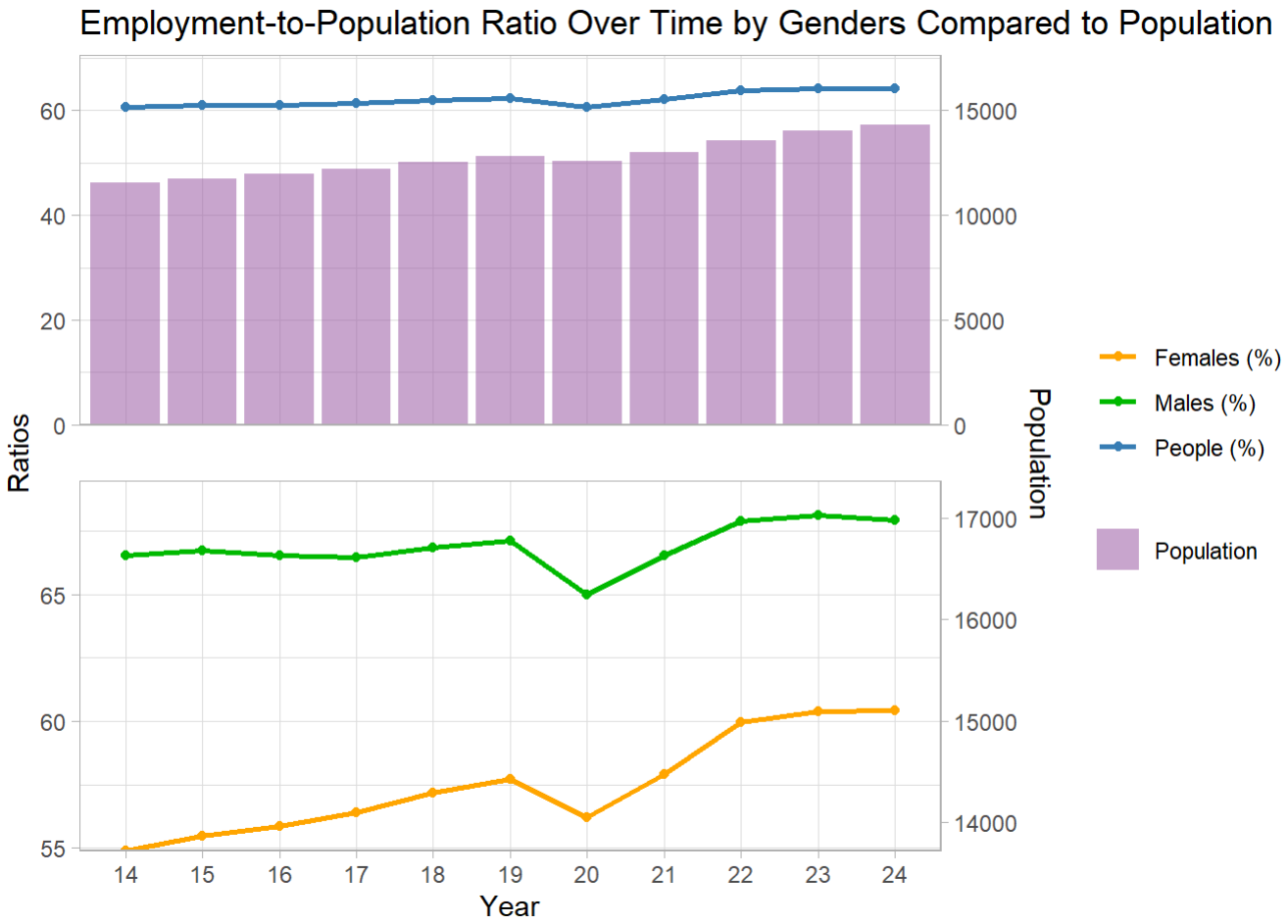
```
## # A tibble: 6 × 4
## # Groups:   Year [2]
##   Year  Groups       Ratios Facet_Group
##   <chr> <chr>         <dbl> <fct>
## 1 14    Females (%)    54.9 Genders
## 2 14    Males (%)      66.5 Genders
## 3 14    People (%)     60.6 Overall Population
## 4 14    Population  11556.  Overall Population
## 5 15    Females (%)    55.5 Genders
## 6 15    Males (%)      66.7 Genders
```

```
viz <- ggplot(emp_sub3, aes(x = Year)) +
  # Line and Bar Charts
  geom_line(data = filter(emp_sub3, Groups != "Population"),
            aes(y = Ratios, color = Groups, group = Groups), linewidth = 1) +
  geom_point(data = filter(emp_sub3, Groups != "Population"),
             aes(y = Ratios, color = Groups)) +
  geom_col(data = filter(emp_sub3, Groups == "Population" & Facet_Group == "Overall Population"),
           aes(y = Ratios / 250, fill = Groups), alpha = 0.5) +

  # Primary and secondary y-axis
  scale_y_continuous(
    name = "Ratios",
    sec.axis = sec_axis(~ . * 250, name = "Population"), #scale the y-axis
    expand = expansion(mult = c(0, 0.1))  # Adjust for spacing
  ) +

  # Labels
  labs(title = "Employment-to-Population Ratio Over Time by Genders Compared to Population",
       x = "Year") +

  # Lines Colors
  scale_color_manual(name = "", values = c(
    "People (%)" = "#377eb8",
    "Males (%)" = "#00bd00",
    "Females (%)" = "#FFA500"
  )) +

  # Bar Colors
  scale_fill_manual(name = "", values = c(
    "Population" = "#984ea3"  # Purple for the Population bars
  )) +

  # Facet
  facet_grid(Facet_Group ~ ., scales = "free_y",
             labeller = as_labeller(c("Overall Population" = "Overall Population", "Genders" = "Genders"))) +

  #Theme
  theme_light() +
```

```
theme(
  strip.text.y = element_blank(),  # Remove facet text labels
  strip.background = element_blank(),  # Remove the background box of facet labels
  panel.spacing = unit(1.5, "lines"),
  legend.position = "right",
  legend.title = element_blank()  # Remove legend title
)
```

Reconstruction

The following plot is the final version that are able to fix the three main issues in the critiques. It can be seen that even though there are some slight fluctuation before the significant drop in the year 2020, the number of people whose status “employed” kept increasing annually except the one during Covid-19 pandemic.



References

- Australian Bureau of Statistics. (2024). *Labour Force, Australia*. Retrieved September 19, 2024, from: https://www.abs.gov.au/statistics/labour/employment-and-unemployment/labour-force-australia/latest-release (https://www.abs.gov.au/statistics/labour/employment-and-unemployment/labour-force-australia/latest-release)
- Australian Bureau of Statistics. (2024). *Labour Force, Australia*. Retrieved September 19, 2024, from: https://www.abs.gov.au/statistics/labour/employment-and-unemployment/labour-force-australia/latest-release (https://www.abs.gov.au/statistics/labour/employment-and-unemployment/labour-force-australia/latest-release)