

# Homework 8

GONG Kuiyuan

Student ID: 39-246182

Student name: GONG Kuiyuan

Preferred name: Eddie

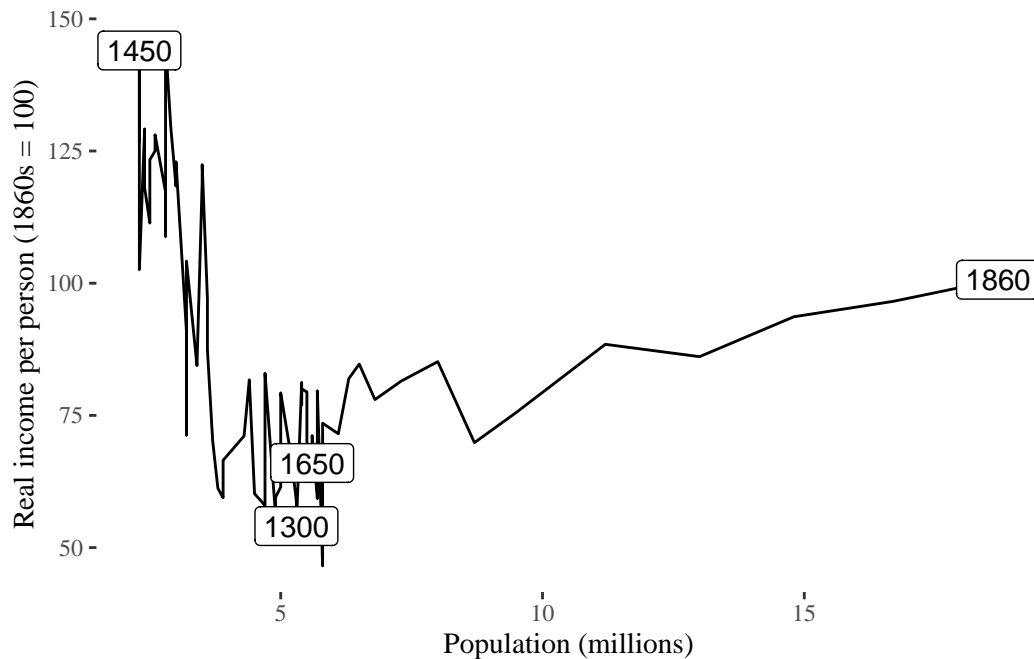
## Answers:

### Task 1:

```
library(pacman)
p_load(griffendata, ggplot2, ggthemes, hrbrthemes, png, magick, grid, readxl, dplyr, stringr)

p <- clark |>
  ggplot(
    aes(x = population, y = realwage)
  ) +
  geom_line() +
  labs(
    x = "Population (millions)",
    y = "Real income per person (1860s = 100)"
  ) +
  theme_tufte() +
  geom_label(
    data = clark[clark$year == 1300 | clark$year == 1450 | clark$year == 1650 | clark$year == 1800],
    aes(label = year))

print(p)
```



## Task 2:

```
setwd("/Users/edisonkung/desktop/R for empirical research/HW")
ggsave("clark.png", p, width = 6, height = 4)
```

**Task 3:** I use two argument `path =` and `pattern =` to show that the file `clark.png` is indeed saved on my system.

```
list.files(path = "/Users/edisonkung/desktop/R for empirical research/HW")
```

```
[1] "39246182_GONG_Kuiyuan_hw1.pdf"
[2] "39246182_GONG_Kuiyuan_hw1.qmd"
[3] "39246182_GONG_Kuiyuan_hw2.pdf"
[4] "39246182_GONG_Kuiyuan_hw2.qmd"
[5] "39246182_GONG_Kuiyuan_hw3.pdf"
[6] "39246182_GONG_Kuiyuan_hw3.qmd"
[7] "39246182_GONG_Kuiyuan_hw4.pdf"
[8] "39246182_GONG_Kuiyuan_hw4.qmd"
[9] "39246182_GONG_Kuiyuan_hw5.pdf"
[10] "39246182_GONG_Kuiyuan_hw5.qmd"
[11] "39246182_GONG_Kuiyuan_hw6.pdf"
[12] "39246182_GONG_Kuiyuan_hw6.qmd"
```

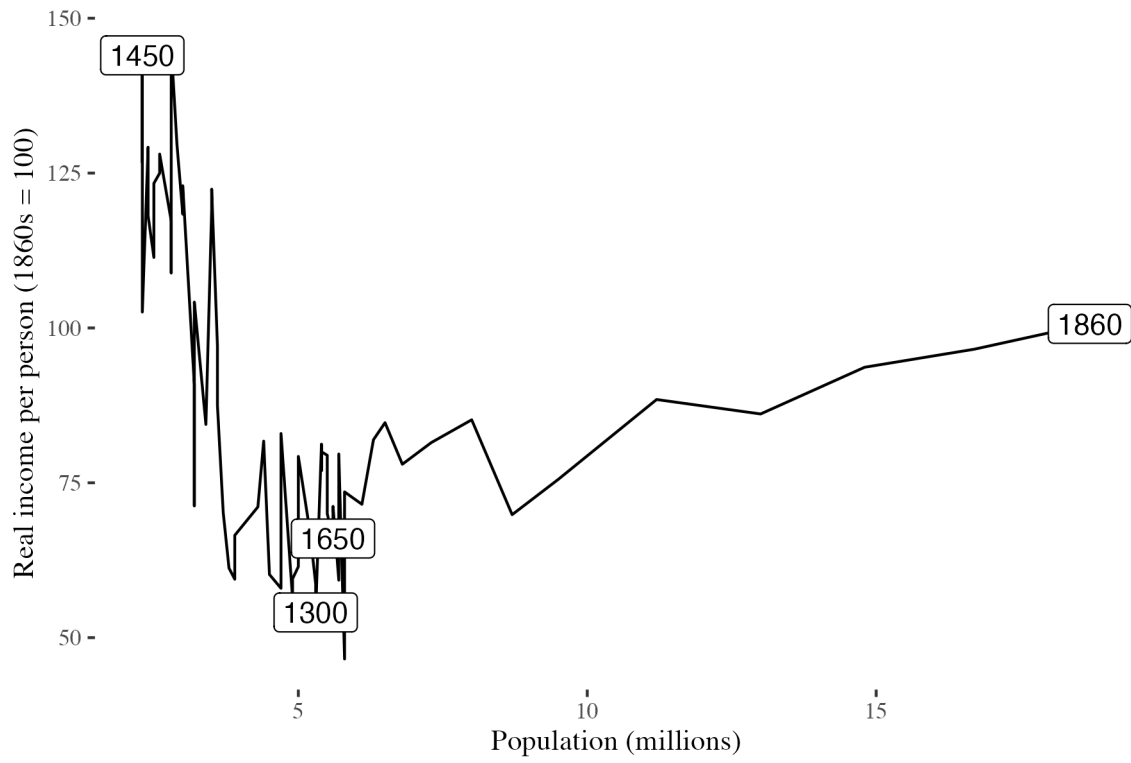
```
[13] "39246182_GONG_Kuiyuan_hw7.pdf"
[14] "39246182_GONG_Kuiyuan_hw7.qmd"
[15] "39246182_GONG_Kuiyuan_hw8_files"
[16] "39246182_GONG_Kuiyuan_hw8_htmlslides_files"
[17] "39246182_GONG_Kuiyuan_hw8_htmlslides.html"
[18] "39246182_GONG_Kuiyuan_hw8_htmlslides.qmd"
[19] "39246182_GONG_Kuiyuan_hw8.pdf"
[20] "39246182_GONG_Kuiyuan_hw8.qmd"
[21] "39246182_GONG_Kuiyuan_hw8.rmarkdown"
[22] "clark.png"
[23] "classroom1.xlsx"
[24] "classroom2.xlsx"
[25] "netflix_titles.csv"
```

```
list.files(pattern = "\\\\.png$")
```

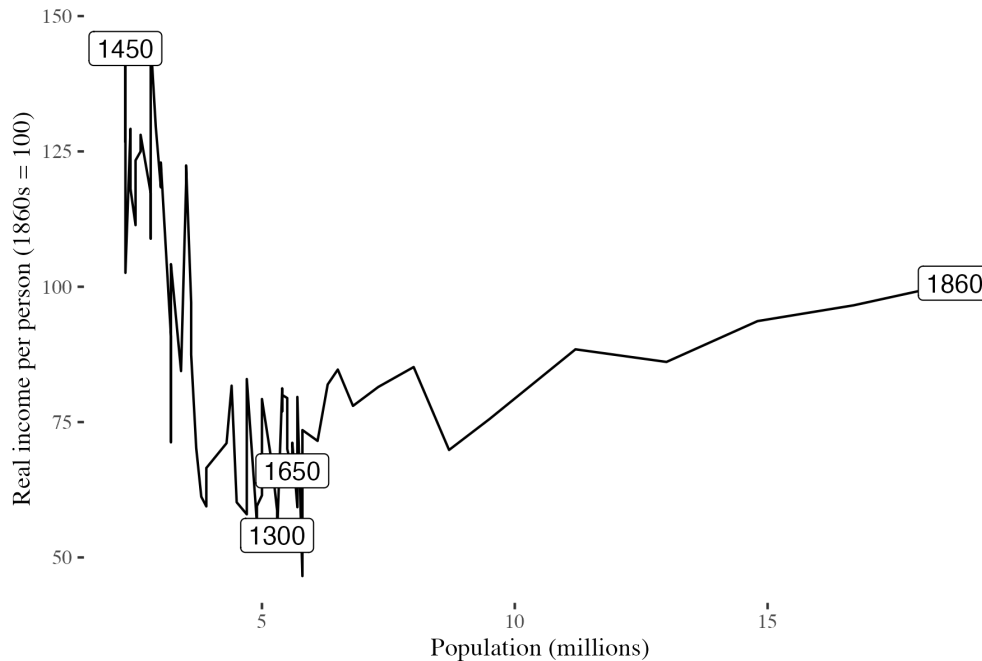
```
[1] "clark.png"
```

**Task 4:** I used two ways to read the png file. The first one is the `readPNG()` function from the package `png`, and the other one is the `image_read()` function from the package `magick`. In order to show the image in this pdf file, `readPNG` alone is enough but include `grid.raster()` from the package `grid`.

```
img1 <- image_read("clark.png")
img1
```



```
img2 <- readPNG("clark.png")  
grid.raster(img2)
```



**Task 5:** The slides I made is submitted along with other files.

**Task 6:** First, I think it ensures the consistency, when we save it as a file, the content of the graph will not change until we overwrite it. In this case, we allow to rewrite the codes and make some other adjustments for other purposes but it doesn't change the graph. Second, we don't want to create the graph in the quarto document that creates the slides is because it makes us difficult to make adjustments and debugging. Making an independent file for generating and saving the graph is a better approach. Third, doing this is more efficient since we don't have to run all the codes for plots over and over again. We will end up having a clear workflow to follow in the first place.

**Task 7:**

```
classroom1 <- read_xlsx("classroom1.xlsx")
classroom2 <- read_xlsx("classroom2.xlsx")
```

**Task 8:** I couldn't bind these two files since `id` in `classroom1` is not entirely numeric, which also happens to `test` in `classroom2`. On the other hand, the names of variables are also different for `test score` and `test`. I also found the variable name of `test score` is not convenient for future operations.

```
typeof(classroom1$id)
```

```
[1] "character"
```

```
typeof(classroom2$id)
```

```
[1] "double"
```

```
typeof(classroom1$`test score`)
```

```
[1] "double"
```

```
typeof(classroom2$test)
```

```
[1] "character"
```

**Task 9:** Here, I changed the name of value column into `test_score` to prevent generating missing data afterwards. There will be a lot of missing data otherwise because the values can't be matched together if the names are different but generates two separate columns.

```
classroom1 <- classroom1 |>
  rename(test_score = `test score`) |>
  mutate(
    id = str_replace_all(id, c("six" = "6", "10" = "10")),
    id = as.numeric(id)
  )

classroom2 <- classroom2 |>
  rename(test_score = test) |>
  mutate(
    test_score = str_remove(test_score, "\\\\"),
    test_score = as.numeric(test_score)
  )

classroom <- bind_rows(classroom1, classroom2)
sum(is.na(classroom))
```

```
[1] 0
```

**Task 10:**

```
print(classroom)
```

```
# A tibble: 20 x 2
      id test_score
  <dbl>     <dbl>
1     1         78
2     2         91
3     3         81
4     4         45
5     5         77
6     6         30
7     7         90
8     8         61
9     9         59
10    10         67
11    11         99
12    12         76
13    13         55
14    14         23
15    15         12
16    16         40
17    17         50
18    18         66
19    19         87
20    20         93
```