Take Home Exercise - classmate

Tai Qiu Yan

2025-05-09

# 1 Original submission

The original visualization was created by our classmate Calvin Tan Song Hao. Please refer to [this link](https://isss608-ay2024-25t3.netlify.app/take-home_ex/take-home_ex01-1/take-home_ex01-1) for the original submission.

# 2 Libraries and Data

Originally, five R packages will be used for the preparation:

pacman::p\_load(tidyverse, patchwork, ggthemes, knitr, cowplot)

The data file - respopagesex2024 will be used:

sg\_data <- read\_csv("data/respopagesex2024.csv")

Fixing Age Column Inconsistency:

sg\_data$Age <- as.character(sg\_data$Age)  
  
sg\_data$Age[sg\_data$Age == "90\_and\_Over"] <- "90"  
  
sg\_data$Age <- as.numeric(sg\_data$Age)

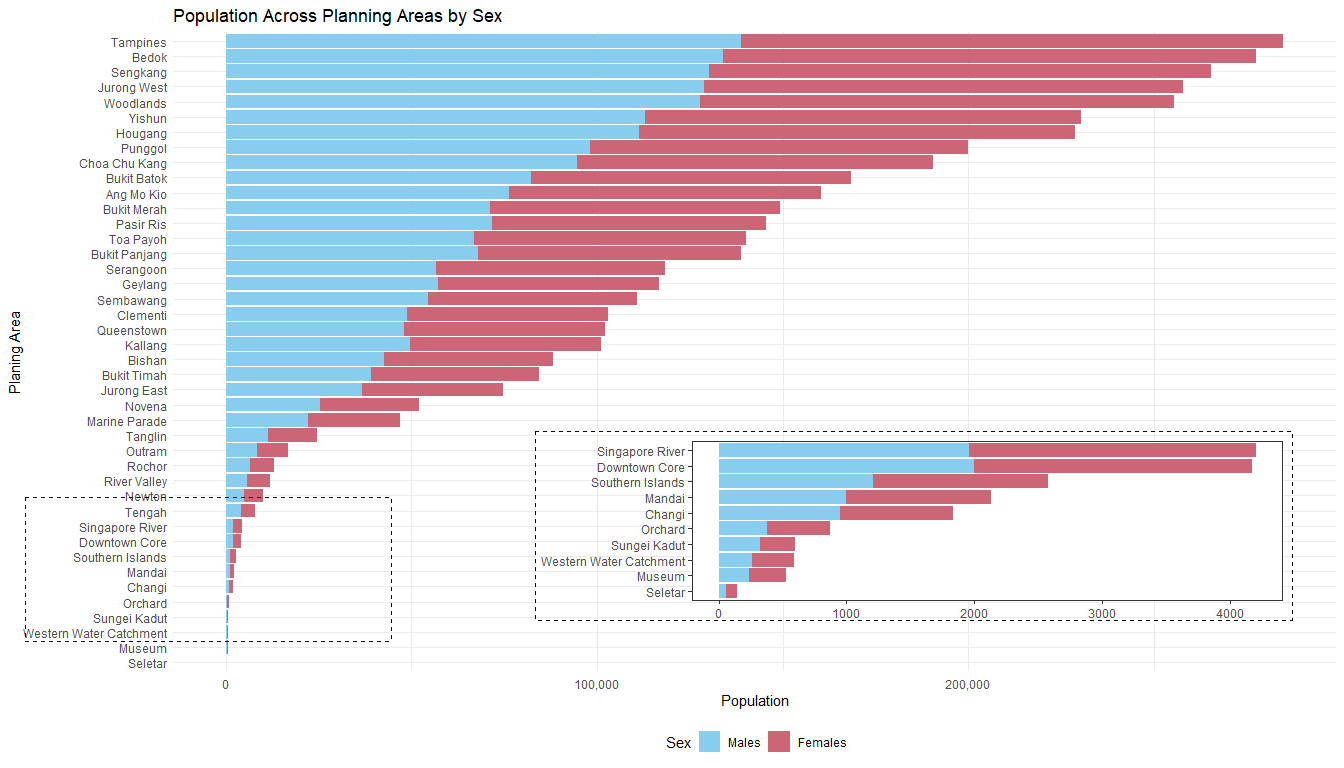
Binning Age Column:

* **Young**: Singapore residents aged 0 to 24
* **Economically\_Active**: Singapore residents aged 25 to 64
* **Aged**: Singapore residents aged 65 and above

sg\_data$Age\_Group <- cut(  
 sg\_data$Age,  
 breaks = c(-1, 24, 64, Inf),  
 labels = c("Young", "Economically\_Active", "Aged"),  
 right = TRUE  
)

# 3 Visualisations

## 3.1 Population Across Planning Areas by Sex



## 3.1.1 Good design principles

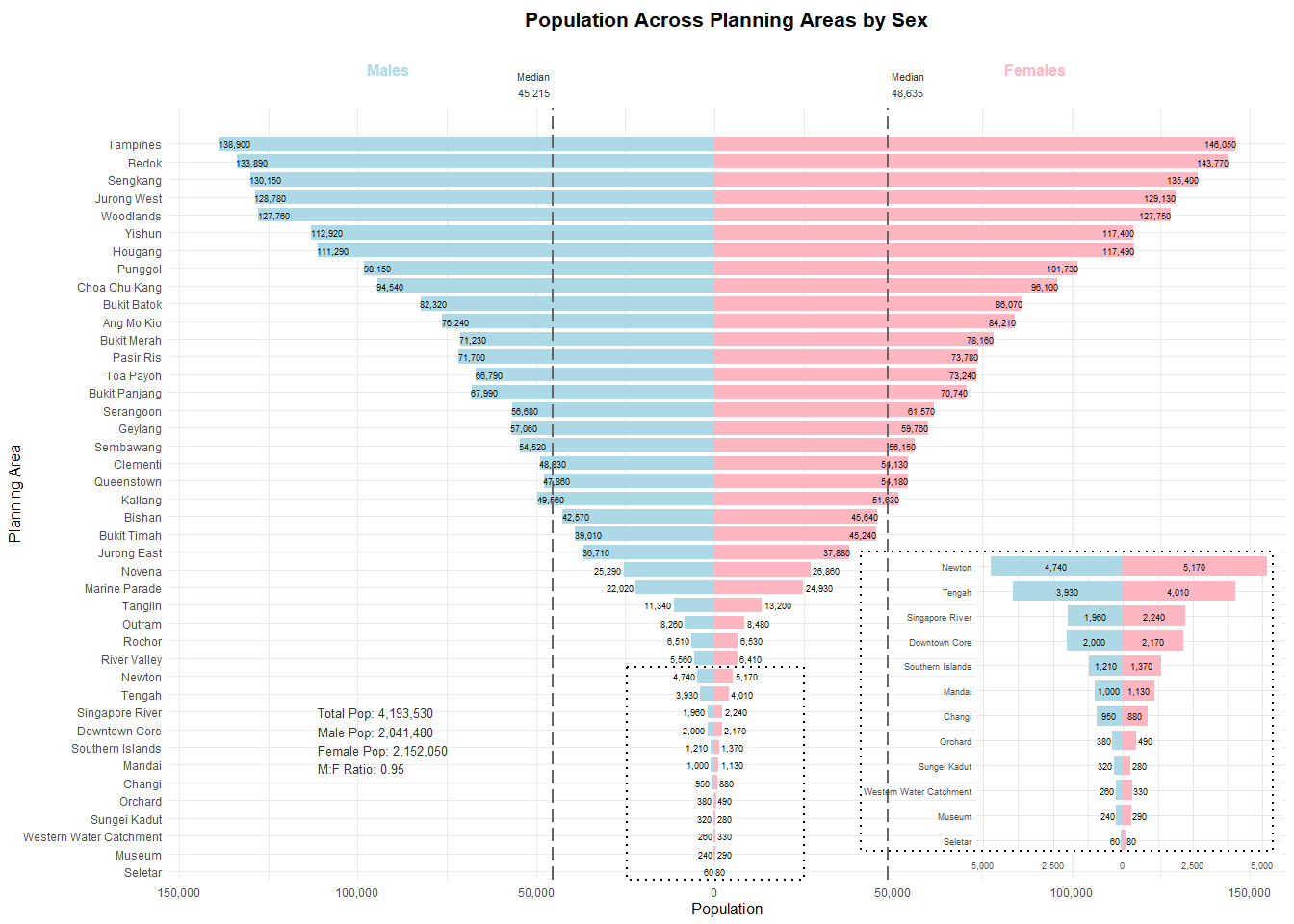
* **Use of inset plot:** The use of an inset plot is particularly effective. It provides a magnified view of planning areas with relatively small populations that would otherwise appear almost invisible in the main chart. This ensures that no data is excluded from analysis, promoting inclusiveness and allowing for scale sensitivity without distorting the overall pattern of the chart.
* **Clear color coding:** Color coding by sex is implemented using familiar and easily distinguishable hues—blue for males and red for females. This choice allows viewers to quickly interpret the distribution of population by gender across different areas without needing to consult the legend repeatedly.
* **Horizontal bar chart with proper sorting:** The horizontal bar chart is well-structured, with planning areas sorted in descending order of total population. This arrangement supports effective comparative analysis and, importantly, accommodates long area names in a way that avoids label clutter, improving overall readability.

## 3.1.2 Areas for further improvement

* **Stacked bars obscure comparison across genders:** The female bars are stacked on top of the male bars, making it difficult to compare female populations across planning areas since they do not share a common baseline. This limits interpretability and distorts direct comparisons. A better approach would be to use a population pyramid format, with males and females on opposite sides of a central axis. Additionally, labeling each bar with the exact population figure would provide immediate numerical clarity and make the chart more informative.
* **Redundant legend placement and layout inefficiency:** Although the legend is useful, it is somewhat redundant given the clear color-coding. It could be relocated near the inset chart to save space and reduce clutter. In the original layout, overlapping labels made it hard to read planning area names. Removing the legend from its current position would allow the chart to be taller and cleaner, improving overall readability.
* **Lack of summary indicators:** The chart would benefit from the addition of vertical median lines for both male and female populations. These lines would serve as reference benchmarks, allowing users to quickly identify which planning areas have populations above or below the median. This simple addition would enhance interpretability and provide a clearer statistical context to the viewer.

## 3.1.3 Makeover

## The plot



## The code

library(ggplot2)  
library(dplyr)  
library(scales)  
  
# 1 Data prep  
demography\_sex <- sg\_data %>%  
 group\_by(PA, Sex) %>%  
 summarise(Pop = sum(Pop), .groups = "drop") %>%  
 group\_by(PA) %>%  
 mutate(Total = sum(Pop)) %>%  
 filter(Total > 0) %>%  
 mutate(Pop = ifelse(Sex == "Males", -Pop, Pop)) %>%  
 ungroup()  
  
stats <- demography\_sex %>%  
 summarise(  
 total = sum(abs(Pop)),  
 male = sum(abs(Pop[Sex=="Males"])),  
 female = sum(Pop[Sex=="Females"]),  
 ratio = sum(abs(Pop[Sex=="Males"])) / sum(Pop[Sex=="Females"])  
 )  
stats\_text <- paste0(  
 "Total Pop: ", comma(stats$total), "\n",  
 "Male Pop: ", comma(stats$male), "\n",  
 "Female Pop: ", comma(stats$female), "\n",  
 "M:F Ratio: ", round(stats$ratio,2)  
)  
  
# 2 Ordering  
pa\_order <- demography\_sex %>%  
 distinct(PA, Total) %>%  
 arrange(desc(Total)) %>%  
 pull(PA)  
demography\_sex$PA <- factor(demography\_sex$PA, levels = rev(pa\_order))  
  
# 3 Identify inset PAs (Newton → Seletar)  
pa\_seq <- pa\_order  
start\_idx <- which(pa\_seq == "Newton")  
end\_idx <- which(pa\_seq == "Seletar")  
inset\_order <- pa\_seq[start\_idx:end\_idx]  
inset\_data <- demography\_sex %>%  
 filter(PA %in% inset\_order) %>%  
 mutate(PA = factor(PA, levels = rev(inset\_order)))  
  
# 4 Medians  
male\_med <- demography\_sex %>% filter(Sex=="Males") %>% pull(Pop) %>% abs() %>% median()  
female\_med <- demography\_sex %>% filter(Sex=="Females") %>% pull(Pop) %>% median()  
  
# 5 Highlight region  
highlight\_pos <- match(inset\_order, levels(demography\_sex$PA))  
ymin\_hl <- min(highlight\_pos) - 0.5  
ymax\_hl <- max(highlight\_pos) + 0.5  
  
box\_pops <- demography\_sex %>%  
 filter(PA %in% c("Newton","Tengah")) %>%  
 pull(Pop)  
  
xmin\_box <- min(box\_pops)   
xmax\_box <- max(box\_pops)   
  
pad <- (xmax\_box - xmin\_box) \* 2  
xmin\_box <- xmin\_box - pad  
xmax\_box <- xmax\_box + pad  
  
# DEFINE INSIDE vs OUTSIDE for main plot  
inside\_main <- pa\_order[1:which(pa\_order=="Jurong East")]  
outside\_main <- setdiff(pa\_order, inside\_main)  
  
# DEFINE INSIDE vs OUTSIDE for inset plot  
inside\_inset <- inset\_order[1:which(inset\_order=="Changi")]  
outside\_inset <- setdiff(inset\_order, inside\_inset)  
  
# MAIN PYRAMID  
main\_plot <- ggplot(demography\_sex, aes(x = Pop, y = PA, fill = Sex)) +  
   
 annotate("rect",  
 xmin = xmin\_box, xmax = xmax\_box,  
 ymin = ymin\_hl, ymax = ymax\_hl,  
 colour = "black", linetype = "dotted",  
 size = 1, fill = NA) +  
   
 geom\_col(width = 0.8) +  
   
 # summary text in left blank area  
 annotate("text",  
 x = min(demography\_sex$Pop) \* 0.8,  
 y = nlevels(demography\_sex$PA) \* 0.2,  
 label= stats\_text,  
 hjust= 0, size = 3.5, colour = "grey20") +  
   
 # inside labels (Tampines → Jurong East), shifted to bar ends but still inside  
 geom\_text(  
 data = filter(demography\_sex, PA %in% inside\_main),  
 aes(label = comma(abs(Pop))),  
 position = position\_identity(),  
 hjust = ifelse(  
 filter(demography\_sex, PA %in% inside\_main)$Sex == "Males",  
 0.02, 0.98  
 ),  
 vjust = 0.5,  
 size = 2.5, colour = "black"  
 ) +  
   
 # outside labels (below Rochor → Seletar)  
 geom\_text(  
 data = filter(demography\_sex, PA %in% outside\_main),  
 aes(label = comma(abs(Pop))),  
 position = position\_identity(),  
 hjust = ifelse(  
 filter(demography\_sex, PA %in% outside\_main)$Sex == "Males",  
 1.1, -0.1  
 ),  
 size = 2.5, colour = "black"  
 ) +  
   
 geom\_vline(xintercept = -male\_med, linetype="longdash", size=1, colour="grey40") +  
 geom\_vline(xintercept = female\_med, linetype="longdash", size=1, colour="grey40") +  
   
 annotate("text",  
 x = -male\_med, y = Inf,  
 label= paste0("Median\n", comma(male\_med)),  
 hjust= 1.1, vjust= -0.5,  
 size = 3, colour = "grey20") +  
 annotate("text",  
 x = female\_med, y = Inf,  
 label= paste0("Median\n", comma(female\_med)),  
 hjust= -0.1, vjust= -0.5,  
 size = 3, colour = "grey20") +  
   
 scale\_fill\_manual(values = c("Males"="#ADD8E6","Females"="#FFB6C1")) +  
 scale\_x\_continuous(  
 labels = function(x) comma(abs(x))  
 ) +  
 scale\_y\_discrete(expand = expansion(add = c(0,2))) +  
   
 labs(x="Population", y="Planning Area") +  
 theme\_minimal() +  
 theme(  
 plot.title = element\_blank(),  
 axis.title = element\_text(size=12),  
 axis.text = element\_text(size=9),  
 legend.position = "none"  
 ) +  
 coord\_cartesian(clip="off")  
  
# INSET PLOT  
inset\_plot <- ggplot(inset\_data, aes(x = Pop, y = PA, fill = Sex)) +  
 geom\_col(width = 0.8) +  
   
 # inside inset labels (Newton → Changi)  
 geom\_text(  
 data = filter(inset\_data, PA %in% inside\_inset),  
 aes(label = comma(abs(Pop))),  
 position = position\_stack(vjust = 0.5),  
 size = 2.5, colour = "black"  
 ) +  
   
 # outside inset labels (Mandai → Seletar)  
 geom\_text(  
 data = filter(inset\_data, PA %in% outside\_inset),  
 aes(label = comma(abs(Pop))),  
 position = position\_identity(),  
 hjust = ifelse(  
 filter(inset\_data, PA %in% outside\_inset)$Sex == "Males",  
 1.1, -0.1  
 ),  
 size = 2.5, colour = "black"  
 ) +  
   
 scale\_fill\_manual(values = c("Males"="#ADD8E6","Females"="#FFB6C1")) +  
 scale\_x\_continuous(  
 labels = function(x) comma(abs(x))  
 ) +  
 theme\_minimal(base\_size=9) +  
 theme(  
 axis.title = element\_blank(),  
 axis.text.y = element\_text(size=7),  
 axis.text.x = element\_text(size=7),  
 panel.background = element\_rect(fill="white", colour=NA),  
 panel.grid.major = element\_line(colour="grey90"),  
 legend.position = "none"  
 )  
  
# DOTTED BOX FOR THE INSET  
zoom\_box <- ggplot() +  
 theme\_void() +  
 annotate("rect",  
 xmin = 0, xmax = 1,  
 ymin = 0, ymax = 1,  
 colour = "black", linetype = "dotted", size = 1, fill = NA)  
  
# ASSEMBLE EVERYTHING  
title\_plot <- ggdraw() +  
 draw\_label("Population Across Planning Areas by Sex",  
 fontface="bold", size=16, hjust=0.3)  
  
label\_plot <- ggdraw() +  
 draw\_label("Males", x=0.3, y=0.5, fontface="bold", size=13, colour="#ADD8E6") +  
 draw\_label("Females", x=0.8, y=0.5, fontface="bold", size=13, colour="#FFB6C1")  
  
final\_plot <- plot\_grid(  
 title\_plot,  
 label\_plot,  
 ggdraw(main\_plot) +  
 draw\_plot(inset\_plot, x=0.665, y=0.06, width=0.33, height=0.40) +  
 draw\_plot(zoom\_box, x=0.65, y=0.07, width=0.35, height=0.40),  
 ncol=1,  
 rel\_heights=c(0.04,0.07,0.89)  
)  
  
print(final\_plot)