Take-Home Exercise 1 Part 2

Liaw Ying Ting, Celin

2025-05-07

# 1 Overview

For part 2 of the Take-Home Exercise 1, I will be evaluating one of my peer’s [Ng Wee Tinn Shermainn’s](https://isss608-sherm.netlify.app/take-home_ex/take-home_ex01/take-home_ex01) website for her Take-Home Exercise 1. I will be providing 3 good design principles and 3 areas for further improvements for the data visualisation plots done by Shermainn

# 2 Critique of Plots

## 2.1 Top 10 subzones of each age group

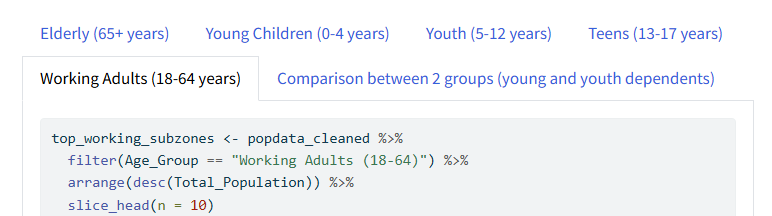
**3 Good Design Principles:**

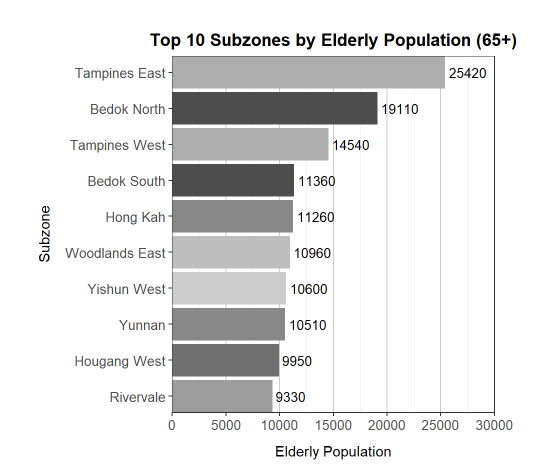
* Analysis is split into the different population age groups which makes it easier to analyse the graphs based on the age groups
* The analysis identifies the top 10 subzones for each age group and makes it easier to identify the facilities that are required for each subzone based on age group
* The graphs are ordered in descending order according to their population (from the largest population to the 10th largest population), making it easier for viewing based on the position of the bar graph, making it easy to identify which subzone has the largest population

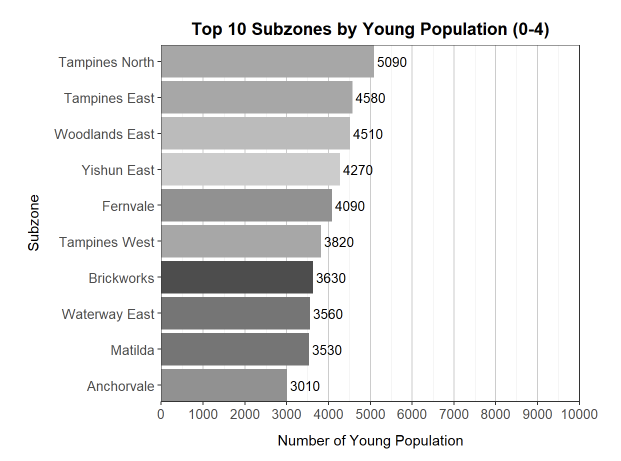
**3 Areas for Further Improvements:**

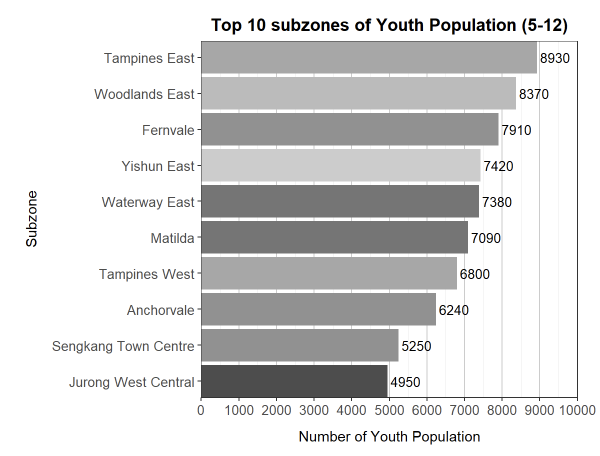
* The colours for the bar charts are in shades of grey which makes it hard to differentiate the different subzones. It would be better if specific colours are used for each area
* There is only 1 tab for comparison between 2 groups (young and youth dependents). It would be better if more comparisons can be made
* The chart also only shows the top 10 subzones but the other subzones were not shown. It would be better if the lowest 10 subzones or all subzones can be shown.

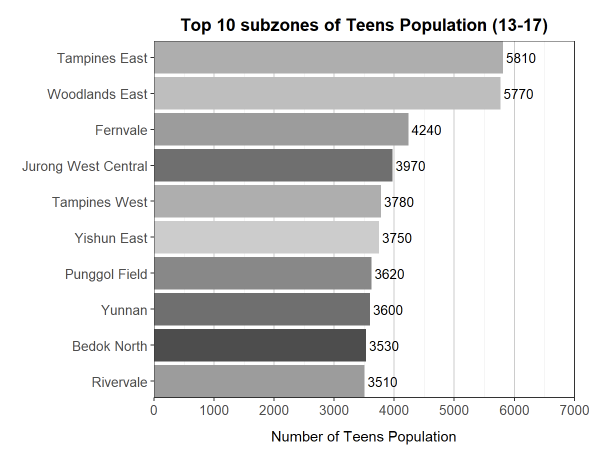
**I have added screenshots of her graphs below:**

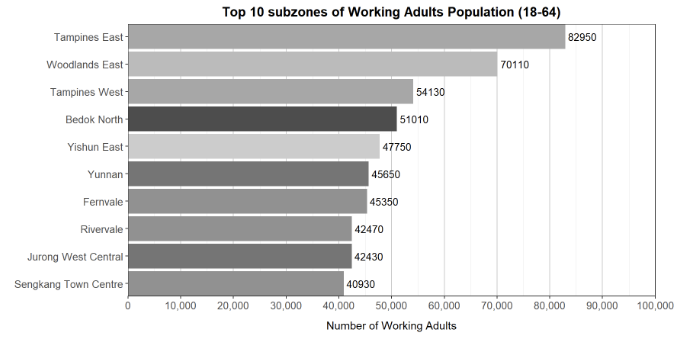


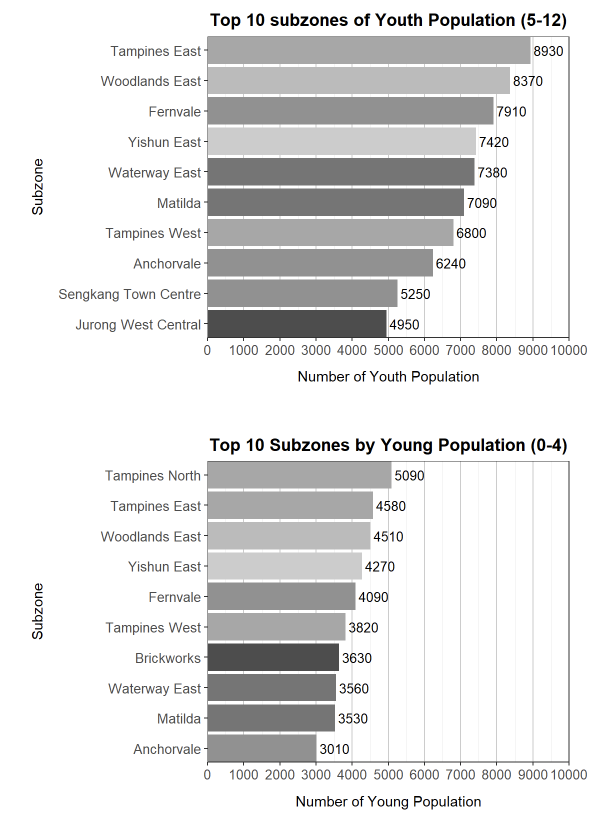












## 2.2 Data Wrangling

### 2.2.1 Load Packages

pacman::p\_load(tidyverse,patchwork, ggthemes, dplyr)

### 2.2.2 Import Data

popdata <- read\_csv("respopagesex2024.csv")

### 2.2.3 Categorizing into Main zones and Subzones

# Clean and process the data  
popdata\_cleaned <- popdata %>%  
 rename(  
 Planning\_Area = PA,  
 Subzone = SZ,  
 Population = Pop  
 ) %>%  
 filter(Sex != "Total", Age != "Total", Subzone != "Total") %>%  
 mutate(  
 Population = as.numeric(Population),  
 Age = as.numeric(Age),  
 Age\_Group = case\_when(  
 Age >= 0 & Age <= 4 ~ "Young (0-4)",  
 Age >= 5 & Age <= 12 ~ "Youth (5-12)",  
 Age >= 13 & Age <= 17 ~ "Teens (13-17)",  
 Age >= 18 & Age <= 64 ~ "Working Adults (18-64)",  
 Age >= 65 ~ "Elderly (65+)",  
 TRUE ~ NA\_character\_  
 )  
 ) %>%  
 filter(!is.na(Age\_Group)) %>%  
 group\_by(Planning\_Area, Subzone, Age\_Group) %>%  
 summarise(Total\_Population = sum(Population, na.rm = TRUE), .groups = "drop")  
  
# View results  
head(popdata\_cleaned)

# A tibble: 6 × 4  
 Planning\_Area Subzone Age\_Group Total\_Population  
 <chr> <chr> <chr> <dbl>  
1 Ang Mo Kio Ang Mo Kio Town Centre Elderly (65+) 880  
2 Ang Mo Kio Ang Mo Kio Town Centre Teens (13-17) 290  
3 Ang Mo Kio Ang Mo Kio Town Centre Working Adults (18-64) 3170  
4 Ang Mo Kio Ang Mo Kio Town Centre Young (0-4) 120  
5 Ang Mo Kio Ang Mo Kio Town Centre Youth (5-12) 420  
6 Ang Mo Kio Cheng San Elderly (65+) 6580

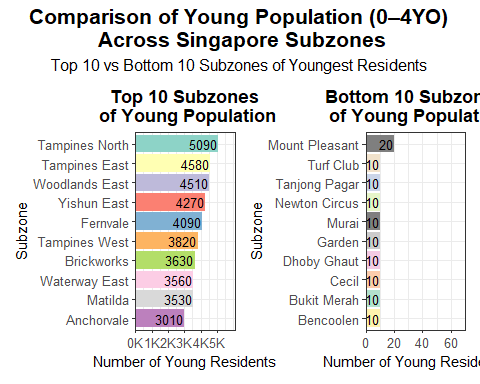
### 2.2.4 Top 10 and Bottom 10 Subzones of each age group

In this code, i have added colours to the different Subzones and also put the top 10 and bottom 10 subzones side-by-side to make it more easier to compare.

The code #| code-fold: true is also added at the top of the codes to make sure the codes are collapsible and the other graphs can be easily comparable after collapsing the codes. They have also been displayed from the youngest to the oldest.

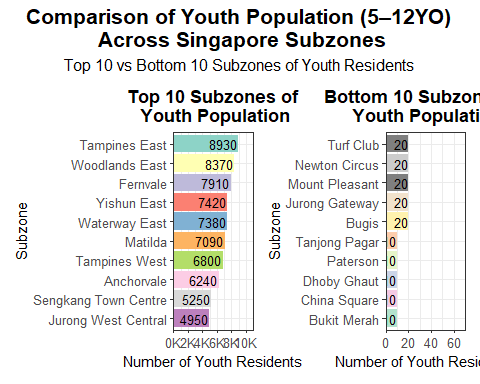
**Top 10 vs Bottom 10 Subzones of Young Population (0-4 Years Old)**

# Load necessary libraries  
pacman::p\_load(tidyverse, patchwork, RColorBrewer)  
  
# Get top 10 and bottom 10 subzones of young population  
top\_young\_subzones <- popdata\_cleaned %>%  
 filter(Age\_Group == "Young (0-4)") %>%  
 arrange(desc(Total\_Population)) %>%  
 slice\_head(n = 10)  
  
bottom\_young\_subzones <- popdata\_cleaned %>%  
 filter(Age\_Group == "Young (0-4)", Total\_Population > 0) %>%  
 arrange(Total\_Population) %>%  
 slice\_head(n = 10)  
  
# SAFE color assignment for top and bottom (brewer.pal only works up to 12 for Set3 and 8 for Pastel2)  
top\_n <- nrow(top\_young\_subzones)  
bottom\_n <- nrow(bottom\_young\_subzones)  
  
top\_palette <- brewer.pal(max(3, min(top\_n, 12)), "Set3")  
bottom\_palette <- brewer.pal(max(3, min(bottom\_n, 8)), "Pastel2")  
  
top\_colors <- setNames(top\_palette[1:top\_n], top\_young\_subzones$Subzone)  
bottom\_colors <- setNames(bottom\_palette[1:bottom\_n], bottom\_young\_subzones$Subzone)  
  
# Plot: Top 10  
y1 <- ggplot(top\_young\_subzones,   
 aes(x = reorder(Subzone, Total\_Population),   
 y = Total\_Population,   
 fill = Subzone)) +  
 geom\_col() +  
 geom\_text(aes(label = Total\_Population),   
 hjust = 1.05, size = 3.5, color = "black") +  
 coord\_flip() +  
 labs(  
 title = "Top 10 Subzones\n of Young Population",  
 x = "Subzone",  
 y = "Number of Young Residents"  
 ) +  
 scale\_fill\_manual(values = top\_colors) +  
 scale\_y\_continuous(  
 breaks = seq(0, max(top\_young\_subzones$Total\_Population), by = 1000),  
 labels = function(x) paste0(x / 1000, "K"),  
 limits = c(0, max(top\_young\_subzones$Total\_Population) + 1000),  
 expand = c(0, 0)  
 ) +  
 theme\_bw() +  
 theme(  
 legend.position = "none",  
 plot.title = element\_text(hjust = 0.5, face = "bold"),  
 axis.text = element\_text(size = 10)  
 )  
  
# Plot: Bottom 10  
y2 <- ggplot(bottom\_young\_subzones,   
 aes(x = reorder(Subzone, Total\_Population),   
 y = Total\_Population,   
 fill = Subzone)) +  
 geom\_col() +  
 geom\_text(aes(label = Total\_Population),   
 hjust = 1.05, size = 3.5, color = "black") +  
 coord\_flip() +  
 labs(  
 title = "Bottom 10 Subzones\n of Young Population",  
 x = "Subzone",  
 y = "Number of Young Residents"  
 ) +  
 scale\_fill\_manual(values = bottom\_colors) +  
 scale\_y\_continuous(  
 limits = c(0, max(bottom\_young\_subzones$Total\_Population) + 50),  
 labels = function(x) ifelse(x >= 1000, paste0(x / 1000, "K"), x),  
 expand = c(0, 0)  
 ) +  
 theme\_bw() +  
 theme(  
 legend.position = "none",  
 plot.title = element\_text(hjust = 0.5, face = "bold"),  
 axis.text = element\_text(size = 10)  
 )  
  
# Combine the plots side by side  
y1 + y2 +  
 plot\_layout(ncol = 2) +  
 plot\_annotation(  
 title = "Comparison of Young Population (0–4YO)\n Across Singapore Subzones",  
 subtitle = "Top 10 vs Bottom 10 Subzones of Youngest Residents",  
 theme = theme(  
 plot.title = element\_text(hjust = 0.5, face = "bold", size = 16),  
 plot.subtitle = element\_text(hjust = 0.5, size = 12)  
 )  
 )



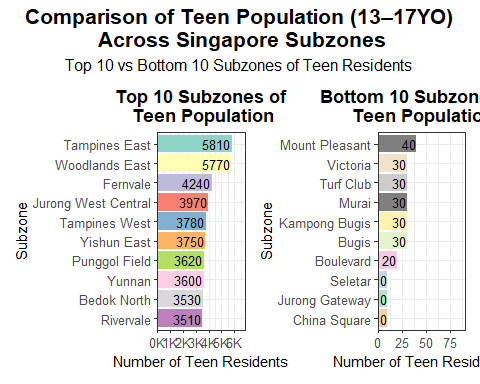
**Top 10 vs Bottom 10 Subzones of Youths (5-12 Years Old)**

# Load required packages  
pacman::p\_load(tidyverse, patchwork, RColorBrewer)  
  
# Get top 10 and bottom 10 subzones of youth population  
top\_youth\_subzones <- popdata\_cleaned %>%  
 filter(Age\_Group == "Youth (5-12)") %>%  
 arrange(desc(Total\_Population)) %>%  
 slice\_head(n = 10)  
  
bottom\_youth\_subzones <- popdata\_cleaned %>%  
 filter(Age\_Group == "Youth (5-12)", Total\_Population > 0) %>%  
 arrange(Total\_Population) %>%  
 slice\_head(n = 10)  
  
# Safely assign palette colors  
top\_n <- nrow(top\_youth\_subzones)  
bottom\_n <- nrow(bottom\_youth\_subzones)  
  
top\_palette <- brewer.pal(max(3, min(top\_n, 12)), "Set3")  
bottom\_palette <- brewer.pal(max(3, min(bottom\_n, 8)), "Pastel2")  
  
top\_colors <- setNames(top\_palette[1:top\_n], top\_youth\_subzones$Subzone)  
bottom\_colors <- setNames(bottom\_palette[1:bottom\_n], bottom\_youth\_subzones$Subzone)  
  
# Plot: Top 10  
y\_top <- ggplot(top\_youth\_subzones,   
 aes(x = reorder(Subzone, Total\_Population),   
 y = Total\_Population,   
 fill = Subzone)) +  
 geom\_col() +  
 geom\_text(aes(label = Total\_Population),   
 hjust = 1.05, size = 3.5, color = "black") +  
 coord\_flip() +  
 labs(  
 title = "Top 10 Subzones of\n Youth Population",  
 x = "Subzone",  
 y = "Number of Youth Residents"  
 ) +  
 scale\_fill\_manual(values = top\_colors) +  
 scale\_y\_continuous(  
 breaks = seq(0, 10000, by = 2000),  
 labels = function(x) paste0(x / 1000, "K"),  
 limits = c(0, max(top\_youth\_subzones$Total\_Population) + 2000),  
 expand = c(0, 0)  
 ) +  
 theme\_bw() +  
 theme(  
 legend.position = "none",  
 plot.title = element\_text(hjust = 0.5, face = "bold"),  
 axis.text = element\_text(size = 10)  
 )  
  
# Plot: Bottom 10  
y\_bottom <- ggplot(bottom\_youth\_subzones,   
 aes(x = reorder(Subzone, Total\_Population),   
 y = Total\_Population,   
 fill = Subzone)) +  
 geom\_col() +  
 geom\_text(aes(label = Total\_Population),   
 hjust = 1.05, size = 3.5, color = "black") +  
 coord\_flip() +  
 labs(  
 title = "Bottom 10 Subzones of\n Youth Population",  
 x = "Subzone",  
 y = "Number of Youth Residents"  
 ) +  
 scale\_fill\_manual(values = bottom\_colors) +  
 scale\_y\_continuous(  
 limits = c(0, max(bottom\_youth\_subzones$Total\_Population) + 50),  
 labels = function(x) ifelse(x >= 1000, paste0(x / 1000, "K"), x),  
 expand = c(0, 0)  
 ) +  
 theme\_bw() +  
 theme(  
 legend.position = "none",  
 plot.title = element\_text(hjust = 0.5, face = "bold"),  
 axis.text = element\_text(size = 10)  
 )  
  
# Combine both plots side by side  
y\_top + y\_bottom +  
 plot\_layout(ncol = 2) +  
 plot\_annotation(  
 title = "Comparison of Youth Population (5–12YO)\n Across Singapore Subzones",  
 subtitle = "Top 10 vs Bottom 10 Subzones of Youth Residents",  
 theme = theme(  
 plot.title = element\_text(hjust = 0.5, face = "bold", size = 16),  
 plot.subtitle = element\_text(hjust = 0.5, size = 12)  
 )  
 )



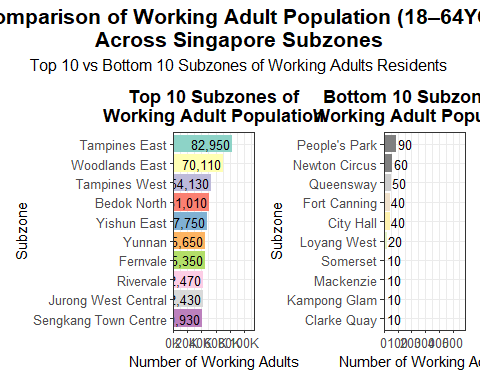
**Top 10 vs Bottom 10 Subzones of Teens (13-17 Years Old)**

# Load required packages  
pacman::p\_load(tidyverse, patchwork, RColorBrewer)  
  
# Get top 10 and bottom 10 subzones by teen population  
top\_teen\_subzones <- popdata\_cleaned %>%  
 filter(Age\_Group == "Teens (13-17)") %>%  
 arrange(desc(Total\_Population)) %>%  
 slice\_head(n = 10)  
  
bottom\_teen\_subzones <- popdata\_cleaned %>%  
 filter(Age\_Group == "Teens (13-17)", Total\_Population > 0) %>%  
 arrange(Total\_Population) %>%  
 slice\_head(n = 10)  
  
# Color-safe assignment  
top\_n <- nrow(top\_teen\_subzones)  
bottom\_n <- nrow(bottom\_teen\_subzones)  
  
top\_palette <- brewer.pal(max(3, min(top\_n, 12)), "Set3")  
bottom\_palette <- brewer.pal(max(3, min(bottom\_n, 8)), "Pastel2")  
  
top\_colors <- setNames(top\_palette[1:top\_n], top\_teen\_subzones$Subzone)  
bottom\_colors <- setNames(bottom\_palette[1:bottom\_n], bottom\_teen\_subzones$Subzone)  
  
# Plot: Top 10  
t\_top <- ggplot(top\_teen\_subzones,   
 aes(x = reorder(Subzone, Total\_Population),   
 y = Total\_Population,   
 fill = Subzone)) +  
 geom\_col() +  
 geom\_text(aes(label = Total\_Population),   
 hjust = 1.05, size = 3.5, color = "black") +  
 coord\_flip() +  
 labs(  
 title = "Top 10 Subzones of\n Teen Population",  
 x = "Subzone",  
 y = "Number of Teen Residents"  
 ) +  
 scale\_fill\_manual(values = top\_colors) +  
 scale\_y\_continuous(  
 breaks = seq(0, 6000, by = 1000),  
 labels = function(x) paste0(x / 1000, "K"),  
 limits = c(0, max(top\_teen\_subzones$Total\_Population) + 1000),  
 expand = c(0, 0)  
 ) +  
 theme\_bw() +  
 theme(  
 legend.position = "none",  
 plot.title = element\_text(hjust = 0.5, face = "bold"),  
 axis.text = element\_text(size = 10)  
 )  
  
# Plot: Bottom 10  
t\_bottom <- ggplot(bottom\_teen\_subzones,   
 aes(x = reorder(Subzone, Total\_Population),   
 y = Total\_Population,   
 fill = Subzone)) +  
 geom\_col() +  
 geom\_text(aes(label = Total\_Population),   
 hjust = 1.05, size = 3.5, color = "black") +  
 coord\_flip() +  
 labs(  
 title = "Bottom 10 Subzones of\n Teen Population",  
 x = "Subzone",  
 y = "Number of Teen Residents"  
 ) +  
 scale\_fill\_manual(values = bottom\_colors) +  
 scale\_y\_continuous(  
 limits = c(0, max(bottom\_teen\_subzones$Total\_Population) + 50),  
 labels = function(x) ifelse(x >= 1000, paste0(x / 1000, "K"), x),  
 expand = c(0, 0)  
 ) +  
 theme\_bw() +  
 theme(  
 legend.position = "none",  
 plot.title = element\_text(hjust = 0.5, face = "bold"),  
 axis.text = element\_text(size = 10)  
 )  
  
# Combine both plots side by side  
t\_top + t\_bottom +  
 plot\_layout(ncol = 2) +  
 plot\_annotation(  
 title = "Comparison of Teen Population (13–17YO)\n Across Singapore Subzones",  
 subtitle = "Top 10 vs Bottom 10 Subzones of Teen Residents",  
 theme = theme(  
 plot.title = element\_text(hjust = 0.5, face = "bold", size = 16),  
 plot.subtitle = element\_text(hjust = 0.5, size = 12)  
 )  
 )



**Top 10 vs Bottom 10 Subzones of Working Adults (18-64 Years Old)**

# Load required packages  
pacman::p\_load(tidyverse, patchwork, RColorBrewer, scales)  
  
# Get top 10 and bottom 10 subzones by working adults population  
top\_working\_subzones <- popdata\_cleaned %>%  
 filter(Age\_Group == "Working Adults (18-64)") %>%  
 arrange(desc(Total\_Population)) %>%  
 slice\_head(n = 10)  
  
bottom\_working\_subzones <- popdata\_cleaned %>%  
 filter(Age\_Group == "Working Adults (18-64)", Total\_Population > 0) %>%  
 arrange(Total\_Population) %>%  
 slice\_head(n = 10)  
  
# Safely assign colors based on count  
top\_n <- nrow(top\_working\_subzones)  
bottom\_n <- nrow(bottom\_working\_subzones)  
  
top\_palette <- brewer.pal(max(3, min(top\_n, 12)), "Set3")  
bottom\_palette <- brewer.pal(max(3, min(bottom\_n, 8)), "Pastel2")  
  
top\_colors <- setNames(top\_palette[1:top\_n], top\_working\_subzones$Subzone)  
bottom\_colors <- setNames(bottom\_palette[1:bottom\_n], bottom\_working\_subzones$Subzone)  
  
# Plot: Top 10  
a\_top <- ggplot(top\_working\_subzones,   
 aes(x = reorder(Subzone, Total\_Population),   
 y = Total\_Population,   
 fill = Subzone)) +  
 geom\_col() +  
 geom\_text(aes(label = comma(Total\_Population)),   
 hjust = 1.05, size = 3.5, color = "black") +  
 coord\_flip() +  
 labs(  
 title = "Top 10 Subzones of\nWorking Adult Population",  
 x = "Subzone",  
 y = "Number of Working Adults"  
 ) +  
 scale\_fill\_manual(values = top\_colors) +  
 scale\_y\_continuous(  
 breaks = seq(0, 100000, by = 20000),  
 limits = c(0, 110000 + 5000),  
 expand = c(0, 0),  
 labels = function(x) paste0(x / 1000, "K")  
 ) +  
 theme\_bw() +  
 theme(  
 legend.position = "none",  
 plot.title = element\_text(hjust = 0.5, face = "bold"),  
 axis.text = element\_text(size = 10)  
 )  
  
# Plot: Bottom 10  
a\_bottom <- ggplot(bottom\_working\_subzones,   
 aes(x = reorder(Subzone, Total\_Population),   
 y = Total\_Population,   
 fill = Subzone)) +  
 geom\_col() +  
 geom\_text(aes(label = comma(Total\_Population)),   
 hjust = -0.1, size = 3.5, color = "black") + # adjusted label position  
 coord\_flip() +  
 labs(  
 title = "Bottom 10 Subzones of\nWorking Adult Population",  
 x = "Subzone",  
 y = "Number of Working Adults"  
 ) +  
 scale\_fill\_manual(values = bottom\_colors) +  
 scale\_y\_continuous(  
 limits = c(0, max(bottom\_working\_subzones$Total\_Population) + 500),  
 expand = c(0, 0),  
 labels = comma  
 ) +  
 theme\_bw() +  
 theme(  
 legend.position = "none",  
 plot.title = element\_text(hjust = 0.5, face = "bold"),  
 axis.text = element\_text(size = 10)  
 )  
  
# Combine both plots side by side  
a\_top + a\_bottom +  
 plot\_layout(ncol = 2) +  
 plot\_annotation(  
 title = "Comparison of Working Adult Population (18–64YO)\nAcross Singapore Subzones",  
 subtitle = "Top 10 vs Bottom 10 Subzones of Working Adults Residents",  
 theme = theme(  
 plot.title = element\_text(hjust = 0.5, face = "bold", size = 16),  
 plot.subtitle = element\_text(hjust = 0.5, size = 12)  
 )  
 )



**Top 10 vs Bottom 10 Subzones of Elderly Population (65+ Years Old)**

top\_elderly\_subzones <- popdata\_cleaned %>%  
 filter(Age\_Group == "Elderly (65+)") %>%  
 arrange(desc(Total\_Population)) %>%  
 slice\_head(n = 10)  
  
# Load required libraries  
pacman::p\_load(tidyverse, patchwork, ggthemes, RColorBrewer)  
  
# Load the data  
popdata <- read\_csv("respopagesex2024.csv")  
  
# Clean and process the data  
popdata\_cleaned <- popdata %>%  
 rename(  
 Planning\_Area = PA,  
 Subzone = SZ,  
 Population = Pop  
 ) %>%  
 filter(Sex != "Total", Age != "Total", Subzone != "Total") %>%  
 mutate(  
 Population = as.numeric(Population),  
 Age = as.numeric(Age),  
 Age\_Group = case\_when(  
 Age >= 0 & Age <= 4 ~ "Young (0-4)",  
 Age >= 5 & Age <= 12 ~ "Youth (5-12)",  
 Age >= 13 & Age <= 17 ~ "Teens (13-17)",  
 Age >= 18 & Age <= 64 ~ "Working Adults (18-64)",  
 Age >= 65 ~ "Elderly (65+)",  
 TRUE ~ NA\_character\_  
 )  
 ) %>%  
 filter(!is.na(Age\_Group)) %>%  
 group\_by(Planning\_Area, Subzone, Age\_Group) %>%  
 summarise(Total\_Population = sum(Population, na.rm = TRUE), .groups = "drop")  
  
# Top 10 subzones by elderly population  
top\_elderly\_subzones <- popdata\_cleaned %>%  
 filter(Age\_Group == "Elderly (65+)") %>%  
 arrange(desc(Total\_Population)) %>%  
 slice\_head(n = 10)  
  
# Bottom 10 subzones (filter out 0s)  
bottom\_elderly\_subzones <- popdata\_cleaned %>%  
 filter(Age\_Group == "Elderly (65+)", Total\_Population > 0) %>%  
 arrange(Total\_Population) %>%  
 slice\_head(n = 10)  
  
# Plot: Top 10  
e1 <- ggplot(top\_elderly\_subzones,   
 aes(x = reorder(Subzone, Total\_Population),   
 y = Total\_Population,   
 fill = Subzone)) +  
 geom\_col() +  
 coord\_flip() +  
 geom\_text(aes(label = Total\_Population),   
 hjust = 1.05, size = 3.5, color = "black") + # Move label inside  
 labs(  
 title = "Top 10 Subzones\n by Elderly Population",  
 x = "Subzone",  
 y = "Elderly Population"  
 ) +  
 scale\_fill\_brewer(palette = "Set3") +  
 scale\_x\_discrete(expand = expansion(mult = c(0.05, 0.05))) +  
 # scale\_y\_continuous(  
 # breaks = seq(0, 25000, by = 5000), # Cleaner axis ticks  
 # limits = c(0, max(top\_elderly\_subzones$Total\_Population) + 2000),   
 # expand = c(0, 0)   
 # ) +  
 scale\_y\_continuous(  
 breaks = seq(0, 25000, by = 5000), # Tick marks at 5K intervals  
 limits = c(0, max(top\_elderly\_subzones$Total\_Population) + 2000),  
 labels = function(x) paste0(x / 1000, "K"),  
 expand = c(0, 0)  
 ) +  
  
 theme\_bw() +  
 theme(  
 legend.position = "none",  
 plot.title = element\_text(hjust = 0.5, face = "bold"),  
 axis.text = element\_text(size = 10)  
 )  
  
# Plot: Bottom 10  
e2 <- ggplot(bottom\_elderly\_subzones,   
 aes(x = reorder(Subzone, Total\_Population),   
 y = Total\_Population,   
 fill = Subzone)) +  
 geom\_col() +  
 coord\_flip() +  
 geom\_text(aes(label = Total\_Population),   
 hjust = 1.05, size = 3.5, color = "black") +  
 labs(  
 title = "Bottom 10 Subzones\n by Elderly Population",  
 x = "Subzone",  
 y = "Elderly Population"  
 ) +  
 scale\_fill\_brewer(palette = "Set3") +  
 scale\_x\_discrete(expand = expansion(mult = c(0.05, 0.05))) +  
 scale\_y\_continuous(  
 limits = c(0, max(bottom\_elderly\_subzones$Total\_Population) + 20),   
 expand = c(0, 0)   
 ) +  
 theme\_bw() +  
 theme(  
 legend.position = "none",  
 plot.title = element\_text(hjust = 0.5, face = "bold"),  
 axis.text = element\_text(size = 10)  
 )  
  
# Combine plots side by side  
e1 + e2 +   
 plot\_layout(ncol = 2) +   
 plot\_annotation(  
 title = "Comparison of Elderly Population(65+ YO)\n Across Singapore Subzones",  
 subtitle = "Top 10 vs Bottom 10 Subzones of Elderly Residents (Age 65+)",  
 theme = theme(  
 plot.title = element\_text(hjust = 0.5, face = "bold", size = 16),  
 plot.subtitle = element\_text(hjust = 0.5, size = 12)  
 )  
 )

