Demographic structures and distribution of Singapore in 2024

## Getting start

### **Installing and loading the required libraries**

pacman::p\_load(ggrepel, patchwork,   
 ggthemes, hrbrthemes,  
 tidyverse, ggiraph, plotly,   
 DT, readxl, gifski, gapminder,  
 gganimate, webshot2)

### **Importing data**

Popdata <- read.csv("data/respopagesex2024.csv")

## 1. Setting the sense

Singapore’s changing demographics have become a pressing topic, especially in the context of an aging population. Understanding the composition and spatial distribution of residents is increasingly important to both policymakers and the public. This report aims to offer a clear, data-driven overview of Singapore’s demographic landscape in 2024, with a particular focus on differences across planning areas and subzones, as well as variations by age and gender. Drawing on the latest official statistics, this analysis provides insights into the trends shaping Singapore’s communities today.

## 2. Dataset

The dataset used in this report was obtained from the **Singapore Department of Statistics (DOS)** and provides population counts as of June 2024. It contains detailed information on Singapore residents, broken down by Planning Area, Subzone, single year of age, and sex. This high-resolution demographic data enables granular analysis of population structures and spatial distribution across the city. The dataset comprises a total of **60,424 records**. For more details, please refer to the metadata below.

| Variables | Definition |
| --- | --- |
| PA | Planning Area |
| SZ | Subzone |
| Age | Single Year of Age |
| Sex | Sex |
| Pop | Resident Count |
| Time | Time/Period |

## 3. Data processing

To ensure the validity and reliability of this report, it is essential to preprocess the dataset and assess its quality. This includes checking for completeness, consistency, and potential anomalies that may affect the accuracy of our analysis.

### 3.1. Checking for missing values

colSums(is.na(Popdata))

PA SZ Age Sex Pop Time   
 0 0 0 0 0 0

As observed from output of code-chunk above, there’s no missing value in this dataset.

### 3.2. Checking for duplicates

Popdata[duplicated(Popdata)]

data frame with 0 columns and 60424 rows

As observed from output of code-chunk above, there’s no duplicates in this dataset.

### 3.2. Checking for datatype

str(Popdata)

'data.frame': 60424 obs. of 6 variables:  
 $ PA : chr "Ang Mo Kio" "Ang Mo Kio" "Ang Mo Kio" "Ang Mo Kio" ...  
 $ SZ : chr "Ang Mo Kio Town Centre" "Ang Mo Kio Town Centre" "Ang Mo Kio Town Centre" "Ang Mo Kio Town Centre" ...  
 $ Age : chr "0" "0" "1" "1" ...  
 $ Sex : chr "Males" "Females" "Males" "Females" ...  
 $ Pop : int 10 10 10 10 10 10 10 10 30 10 ...  
 $ Time: int 2024 2024 2024 2024 2024 2024 2024 2024 2024 2024 ...

As observed from the output of the code-chunk above, all column data types are appropriate. One point to note is that the datatype of the variable **Age** is character instead of integer. This is because the dataset does not provide age-specific breakdowns beyond 90 years old; instead, individuals aged 90 and above are grouped under the label **“90\_and\_Over”**. This classification is reasonable and consistent with common demographic reporting practices.

### 3.3. Removing unnecessary columns

Since the dataset has already been clearly defined as representing statistics from the year 2024, and upon inspection the variable **Time** contains only a single unique value (2024), this column is no longer necessary and can be safely removed.

unique(Popdata$Time)

[1] 2024

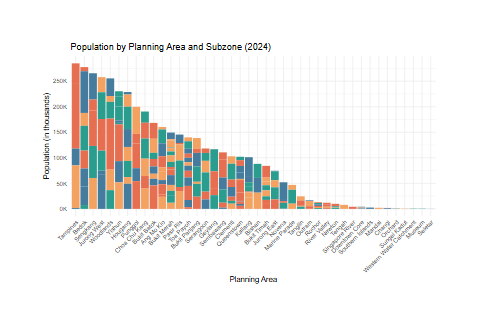
popdata <- subset(Popdata, select = -Time)

## 4. Key Findings

### 4.1. Interactive Population Plot

pa\_sz\_summary <- popdata %>%  
 group\_by(PA, SZ) %>%  
 summarise(Pop = sum(Pop), .groups = "drop") %>%  
 mutate(  
 Pop\_thousand = Pop / 1000,  
 tooltip\_text = str\_replace\_all(  
 paste0("Planning Area: ", PA,  
 "\nSubzone: ", SZ,  
 "\nPopulation: ", Pop),  
 "'", ""))  
  
pa\_order <- pa\_sz\_summary %>%  
 group\_by(PA) %>%  
 summarise(TotalPop = sum(Pop\_thousand)) %>%  
 filter(TotalPop > 0) %>%   
 arrange(desc(TotalPop)) %>%  
 pull(PA)  
  
pa\_sz\_summary <- pa\_sz\_summary %>%  
 filter(PA %in% pa\_order) %>%  
 mutate(PA = factor(PA, levels = pa\_order))  
  
color\_palette <- c("#E76F51", "#457B9D", "#F4A261", "#2A9D8F")  
n\_subzones <- length(unique(pa\_sz\_summary$SZ))  
color\_values <- rep(color\_palette, length.out = n\_subzones)  
sz\_levels <- sort(unique(pa\_sz\_summary$SZ))  
names(color\_values) <- sz\_levels  
  
p <- ggplot(pa\_sz\_summary, aes(x = PA, y = Pop\_thousand, fill = SZ)) +  
 geom\_bar\_interactive(  
 aes(tooltip = tooltip\_text, data\_id = tooltip\_text),  
 stat = "identity", show.legend = FALSE  
 ) +  
 scale\_fill\_manual(values = color\_values) +  
 scale\_y\_continuous(  
 breaks = seq(0, 250, by = 50),  
 labels = function(x) paste0(x, "K")) +  
 labs(title = "Population by Planning Area and Subzone (2024)",  
 x = "Planning Area", y = "Population (in thousands)") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))  
  
girafe(  
 ggobj = p,  
 width\_svg = 8,  
 height\_svg = 8 \* 0.618,  
 options = list(  
 opts\_hover("fill-opacity:1;stroke:black;stroke-width:0.5px;cursor:pointer;"),  
   
 opts\_hover\_inv(css = "opacity:0.2;"),   
 opts\_tooltip()  
 ))

file:///C:/Users/Dell/AppData/Local/Temp/RtmpMbVA3e/file9bc011e065dc/widget9bc036f216fd.html screenshot completed



This chart provides a comprehensive overview of the population distribution across Singapore’s planning areas and subzones in 2024.

From the planning area angle, it shows that population is heavily concentrated in areas such as Tampines, Bedok, and Sengkang. Notably, seven planning areas—Tampines, Bedok, Sengkang, Jurong West, Woodlands, Yishun, and Hougang—have populations exceeding 200,000, indicating high residential density. In contrast, areas such as Seletar and Museum have very low or near-zero populations, suggesting they may be non-residential or under development.

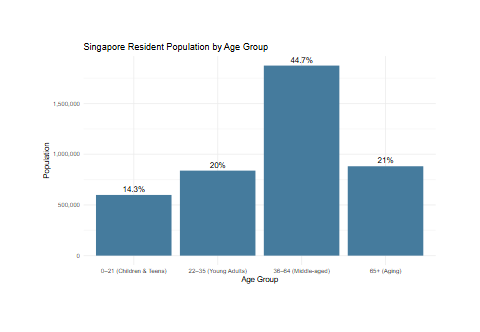
From the subzone angle, Tampines East is the most inhabited subzone, with 128,320 residents, compared to subzones such as Boon lay’s Shipyard which is completely uninhabited.

### 4.2. Distribution of elderly population in Singapore

In Singapore, people aged 65 and above are defined as aging people, directly translated from mandarin, this group of people are expected to have a A happy, joyful and enjoyable retirement life. At this stage of life, they are no longer burdened by raising children or dealing with intense job competition, and instead experience an unprecedented sense of freedom and ease. However, from a societal perspective, an excessive number of seniors can inevitably impact the overall dynamism of society, leading to issues such as labour shortages and increased fiscal pressure.

popdata\_agegroup <- popdata %>%  
 mutate(  
 AgeGroup = case\_when(  
 Age <= 21 ~ "0–21 (Children & Teens)",  
 Age <= 35 ~ "22–35 (Young Adults)",  
 Age <= 64 ~ "36–64 (Middle-aged)",  
 TRUE ~ "65+ (Aging)")) %>%  
   
 group\_by(AgeGroup) %>%  
 summarise(TotalPop = sum(Pop), .groups = "drop") %>%  
 mutate(Percentage = TotalPop / sum(TotalPop),  
 tooltip = paste0(  
 AgeGroup, "\nPopulation: ", formatC(TotalPop, format = "d", big.mark = ",")),  
 AgeGroup = factor(AgeGroup, levels = c(  
 "0–21 (Children & Teens)",   
 "22–35 (Young Adults)",   
 "36–64 (Middle-aged)",   
 "65+ (Aging)")))  
  
p <- ggplot(popdata\_agegroup, aes(x = AgeGroup, y = TotalPop)) +  
 geom\_col\_interactive(  
 aes(tooltip = tooltip, data\_id = AgeGroup),  
 fill = "#457B9D") +  
 geom\_text\_interactive(  
 aes(  
 label = paste0(round(Percentage \* 100, 1), "%"),  
 tooltip = tooltip,  
 data\_id = AgeGroup),  
 vjust = -0.5,  
 size = 4,  
 color = "black") +  
 scale\_y\_continuous(labels = function(x) formatC(x, format = "d", big.mark = ",")) +  
 labs(  
 title = "Singapore Resident Population by Age Group",  
 x = "Age Group",  
 y = "Population") +  
 theme\_minimal()  
  
  
girafe(  
 ggobj = p,  
 width\_svg = 8,  
 height\_svg = 8 \* 0.618,  
 options = list(  
 opts\_hover("fill-opacity:1;stroke:black;stroke-width:0.5px;cursor:pointer;"),  
   
 opts\_hover\_inv(css = "opacity:0.2;"),   
 opts\_tooltip()  
 ))

file:///C:/Users/Dell/AppData/Local/Temp/RtmpMbVA3e/file9bc0371d1b2a/widget9bc069b8415d.html screenshot completed

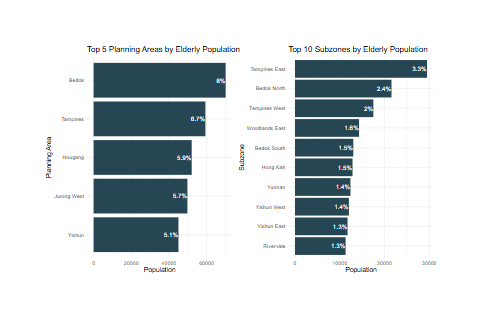


As observed from the chart above, 44.7% of Singapore’s resident population will be in the middle-aged group (36-64 years old), indicating a mature or even aging demographic structure. The children and adolescents (0-21 years old) and the youth (22-35 years old) together account for only 34.3% of the population, resulting in a relatively limited labor supply in the future. The elderly population (65 years old and above) has already reached 21%, a proportion that should not be ignored.

According to international standards, when the proportion of people aged 65 and above in a country or region exceeds 7%, it is considered an aging society; and when it exceeds 14%, it is an aged society. Singapore’s current 21% has clearly entered the “aging society” stage. This means that there are multiple challenges ahead in terms of labor force constraints, rising healthcare needs, and increasing pressure on public finances. The structural trends shown in the charts highlight the need for the government to take effective measures to encourage childbearing, extend the working life and improve the pension system.

elderly\_data <- popdata %>%  
 filter(Age >= 65) %>%  
 group\_by(PA, SZ) %>%  
 summarise(ElderlyPop = sum(Pop), .groups = "drop")  
  
total\_elderly <- sum(elderly\_data$ElderlyPop)  
  
elderly\_by\_PA <- elderly\_data %>%  
 group\_by(PA) %>%  
 summarise(TotalElderly = sum(ElderlyPop)) %>%  
 arrange(desc(TotalElderly)) %>%  
 slice\_max(order\_by = TotalElderly, n = 5) %>%  
 mutate(  
 PA = factor(PA, levels = rev(PA)),  
 Percent = TotalElderly / total\_elderly,  
 tooltip = paste0("Planning Area: ", PA,  
 "\nPopulation: ", TotalElderly))  
  
P1 <- ggplot(elderly\_by\_PA, aes(x = PA, y = TotalElderly)) +  
 geom\_col\_interactive(aes(tooltip = tooltip, data\_id = PA), fill = "#264653") +  
 geom\_text\_interactive(  
 aes(label = paste0(round(Percent \* 100, 1), "%"),  
 tooltip = tooltip, data\_id = PA),  
 hjust = 1.02, color = "white", size = 4) +  
 coord\_flip() +  
 labs(  
 title = "Top 5 Planning Areas by Elderly Population",  
 x = "Planning Area", y = "Population") +  
 theme\_minimal(base\_size = 12)  
  
top\_sz <- elderly\_data %>%  
 arrange(desc(ElderlyPop)) %>%  
 slice\_max(order\_by = ElderlyPop, n = 10) %>%  
 mutate(  
 SZ = reorder(SZ, ElderlyPop),  
 Percent = ElderlyPop / total\_elderly,  
 tooltip = paste0("Subzone: ", SZ,  
 "\nPopulation: ", ElderlyPop))  
  
P2 <- ggplot(top\_sz, aes(x = SZ, y = ElderlyPop)) +  
 geom\_col\_interactive(aes(tooltip = tooltip, data\_id = SZ), fill = "#264653") +  
 geom\_text\_interactive(  
 aes(label = paste0(round(Percent \* 100, 1), "%"),  
 tooltip = tooltip, data\_id = SZ),  
 hjust = 1.02, color = "white", size = 4) +  
 coord\_flip() +  
 labs(title = "Top 10 Subzones by Elderly Population",  
 x = "Subzone", y = "Population") +  
 theme\_minimal(base\_size = 12)  
  
girafe(  
 ggobj = P1 + P2,  
 width\_svg = 10,  
 height\_svg = 6,  
 options = list(  
 opts\_hover("fill-opacity:1;stroke:black;stroke-width:0.5px;cursor:pointer;"),  
 opts\_hover\_inv(css = "opacity:0.2;"),  
 opts\_tooltip()))

file:///C:/Users/Dell/AppData/Local/Temp/RtmpMbVA3e/file9bc0680c879/widget9bc029071711.html screenshot completed



As observed from charts above, planning areas such as Bedok, Tampines, and Hougang have some of the highest elderly populations in Singapore, with Bedok accounting for 8% of the national elderly population, Tampines 6.7%, and Tampines East alone reaching 3.3%. These figures indicate a high concentration of elderly residents in specific mature residential areas. In response to this trend, these regions need to accelerate the development of age-friendly communities, optimize the allocation of care resources, and strengthen neighborhood support systems to address the rapidly growing challenges of an ageing population.