Take-home Exercise 1: Demographic structures and distribution of Singapore in 2024

Vanessa Riadi

2025-05-09

# 1 Overview

A local online media company that publishes daily content on digital platforms is planning to release an article on **demographic structures and distribution of Singapore in 2024**.

# 2 Objective

Assuming the role of the graphical editor of the media company, you are tasked to prepare at most three data visualisations for the article.

# 3 Analytical Toolkit: RStudio

RStudio and Quarto are used as the primary analytical toolkit for this project. The data is processed using appropriate **tidyverse** family of packages and the data visualisation prepared using **ggplot2** and its extensions.

Before we get started, it is important for us to ensure that the required R packages have been installed.

|  |
| --- |
| Install [**pacman**](https://cran.r-project.org/web/packages/pacman/index.html) package |
| If you have yet to install pacman, install itby typing below in the Console:  options(repos = c(CRAN = "https://cloud.r-project.org")) install.packages("pacman") |

We then load the following R packages using the pacman::p\_load() function:

- [**tidyverse**](https://www.tidyverse.org/), a family of modern R packages specially designed to support data science, analysis and communication task including creating static statistical graphs.  
- [**patchwork**](https://patchwork.data-imaginist.com/) for combining multiple ggplot2 graphs into one figure.  
- [**plotly**](https://plotly.com/r/), R library for plotting interactive statistical graphs.  
- **ggrepel:** a R package provides geoms for ggplot2 to repel overlapping text labels.  
- **ggthemes:** a R package provides some extra themes, geoms, and scales for ggplot.  
- **hrbrthemes**: a R package provides typography-centric themes and theme components for ggplot2.  
- [**qreport**](https://cran.r-project.org/web/packages/qreport/index.html): Provides statistical components, tables, and graphs. - [**ggiraph**](https://davidgohel.github.io/ggiraph/): for making ‘ggplot’ graphics interactive.

pacman::p\_load(tidyverse, patchwork,  
 plotly, ggrepel,  
 ggthemes, hrbrthemes, ggiraph, DT, qreport)

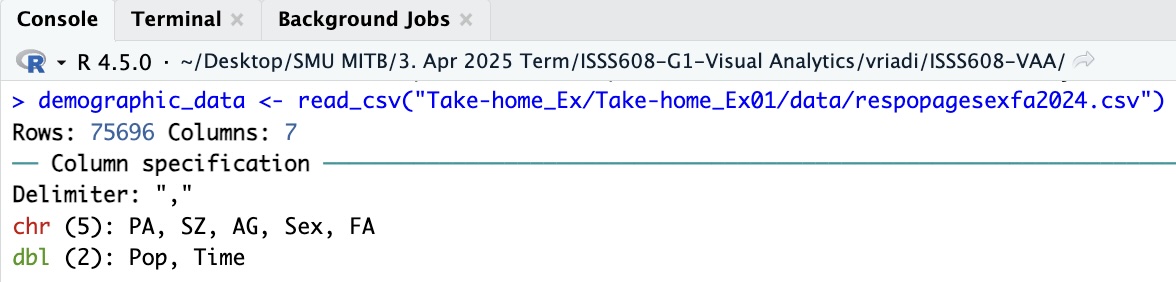
# 4 Data

Singapore Residents by Planning Area / Subzone, Single Year of Age and Sex, June 2024 dataset shared by [Department of Statistics, Singapore(DOS)](https://www.singstat.gov.sg/find-data/search-by-theme/population/geographic-distribution/latest-data)

## 4.1 Load the Data

First we load the data.

demographic\_data <- read\_csv("data/respopagesexfa2024.csv")



## 4.2 Check the Data

jib # THIS IS NOT RUN BECAUSE DT::datatable() produces HTML output #| echo: false #DT::datatable(demographic\_data , options = list( # columnDefs = list(list(className = 'dt-center', targets = 5)), # pageLength = 5, # lengthMenu = c(5, 10, 15, 20) #))

From the first glance, we notice that there are ‘0’ Pop in the dataset. For this exercise, we will be focusing on the top-level visualization per Planning Areas and granular details like Population per Subzone and Floor Area aren’t necessary. We should exclude those rows with zero population values at the Planning Area level. This will help clean up the data and make the visualizations clearer by removing unnecessary zeros. We will do it in [Chapter 4.3](#dataprep)

|  |
| --- |
| Did you know? |
| There are more R packages that can help you to view or describe data. E.g. [Hmisc](https://cran.r-project.org/web/packages/Hmisc/index.html), [psych](https://cran.r-project.org/web/packages/psych/index.html), [**qreport**](https://cran.r-project.org/web/packages/qreport/index.html) package. I will be using qreport here  If you have yet to install qreport, install it by typing below in the Console:  options(repos = c(CRAN = "https://cloud.r-project.org")) install.packages("qreport") |

Here I am using qreportpackage’s [dataOverview](https://search.r-project.org/CRAN/refmans/qreport/html/dataOverview.html) that I already pre-load earlier.

dataOverview(demographic\_data, plot = c("none"),)

demographic\_data has 75696 observations (75696 complete) and 7 variables (7 complete)  
  
  
|Variable |Type | Distinct| Info| Symmetry| NAs|Rarest Value | Frequency of Rarest Value|Mode | Frequency of Mode|  
|:--------|:----------|--------:|-----:|--------:|---:|:-----------------------|-------------------------:|:-----------|-----------------:|  
|PA |Nonnumeric | 55| 0.999| 0.991| 0|Central Water Catchment | 228|Bukit Merah | 3876|  
|SZ |Nonnumeric | 332| 1.000| 1.000| 0|Admiralty | 228|Admiralty | 228|  
|AG |Discrete | 19| 0.997| 1.000| 0|0\_to\_4 | 3984|0\_to\_4 | 3984|  
|Sex |Discrete | 2| 0.750| 1.000| 0|Females | 37848|Females | 37848|  
|FA |Discrete | 6| 0.972| 1.000| 0|<= 60 | 12616|<= 60 | 12616|  
|Pop |Continuous | 183| 0.831| 6.953| 0|1260 | 1|0 | 41742|  
|Time |Discrete | 1| 0.000| 1.000| 0|2024 | 75696|2024 | 75696|

Let’s also count what’s the total Pop

cntpop <- demographic\_data %>%  
 summarise(Pop = sum(Pop, na.rm = TRUE))   
print(cntpop$Pop)

[1] 4187720

#### **Observation**

* The data shows Singapore Residents by Planning Area / Subzone, Single Year of Age and Sex as of June 2024 with total population of 4,187,720.
* We observe that there are 75,696 rows and 7 columns. No missing values are observed. Refer to the column legend in [Appendix A](#appendixa)
* There are a total of seven attributes. 5 of them are categorical data type and the other three are in numerical data type.
  + The categorical attributes are: PA, SZ, AG, Sex, FA.
  + The numerical attributes are: Pop, Time.
* We can also observe how many distinct values for each Variable. This will help us think what to use for our visualization.

## 4.3 Data Preparation

### 4.3.1 Cleaning Data

As mentioned earlier, we will be focusing on the top-level visualization per Planning Areas and granular details like Population per Subzone and Floor Area aren’t necessary. Here we will exclude those rows with zero population values at the Planning Area level by using filter.

## Cleaned Data

## Code

Clean using filter and display as data table

demographic\_data\_clean <- demographic\_data %>%  
 filter(Pop > 0)  
  
#DT::datatable(demographic\_data\_clean , options = list(  
# columnDefs = list(list(className = 'dt-center', targets = 5)),  
# pageLength = 5,  
# lengthMenu = c(5, 10, 15, 20)))

### 4.3.2 Age Group recoding

We will be recoding the Age Group to three groups for clearer visualization: Below 20 years, 20-64 Years, 65 years & Over following [National Statistical Standards Recommendations on Definition and Classification of Age](https://www.singstat.gov.sg/-/media/files/standards_and_classifications/nsa.ashx) (More details in [Appendix B](#appendixb) ).

## Data

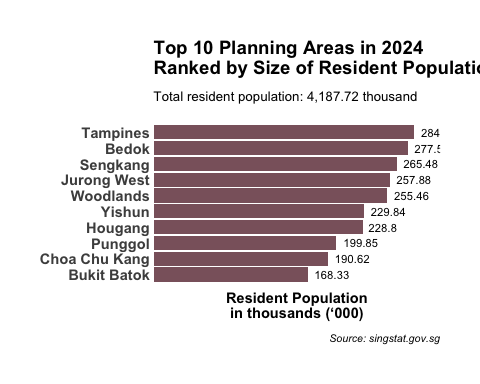
## Code

demographic\_data\_recode <- demographic\_data\_clean %>%  
 mutate(  
 AG\_recode = recode(  
 AG,  
 "0\_to\_4" = "Below 20 Years",  
 "5\_to\_9" = "Below 20 Years",  
 "10\_to\_14" = "Below 20 Years",  
 "15\_to\_19" = "Below 20 Years",  
 "20\_to\_24" = "20-64 Years",  
 "25\_to\_29" = "20-64 Years",  
 "30\_to\_34" = "20-64 Years",  
 "35\_to\_39" = "20-64 Years",  
 "40\_to\_44" = "20-64 Years",  
 "45\_to\_49" = "20-64 Years",  
 "50\_to\_54" = "20-64 Years",  
 "55\_to\_59" = "20-64 Years",  
 "60\_to\_64" = "20-64 Years",  
 "65\_to\_69" = "65 Years and Above",  
 "70\_to\_74" = "65 Years and Above",  
 "75\_to\_79" = "65 Years and Above",  
 "80\_to\_84" = "65 Years and Above",  
 "85\_to\_89" = "65 Years and Above",  
 "90\_and\_over" = "65 Years and Above"  
 )  
 )  
  
#DT::datatable(demographic\_data\_recode , options = list(  
# columnDefs = list(list(className = 'dt-center', targets = 5)),  
# pageLength = 5,  
# lengthMenu = c(5, 10, 15, 20)))

# 5 Data Visualisation, Observation, and Insights

## 5.1 Top 10 Planning Areas (PA) Ranked by Size of Resident Population (Pop)

# 🎀 Plot 1



# Code

top10PA <- demographic\_data\_clean %>%  
 group\_by(PA) %>%  
 summarise(Pop = sum(Pop, na.rm = TRUE)) %>%  
 slice\_max(order\_by = Pop, n = 10)  
  
top10plot <- ggplot(data = top10PA,   
 aes(y = reorder(PA, Pop/1000), x = Pop/1000)) + # reorder PA by Pop  
 geom\_col(show.legend = FALSE, fill = "pink4") +  
 geom\_text(aes(label = (Pop/1000)),  
 hjust = -0.2, color = "black", size = 3) +  
 ggtitle("Top 10 Planning Areas in 2024\nRanked by Size of Singapore Resident Population",  
 subtitle = paste("Total resident population:",   
 format(round(cntpop$Pop / 1000, 2), big.mark = ","),   
 "thousand")) +  
 labs(  
 y = NULL,  
 x = "Resident Population\nin thousands (‘000)",  
 caption = "Source: singstat.gov.sg"  
 ) +  
 theme\_ipsum(base\_family = "Arial",  
 plot\_title\_size = 14,  
 subtitle\_size = 10,  
 caption\_size = 8,  
 plot\_title\_face = "bold",  
 caption\_face = "italic",  
 grid = "",  
 axis\_title\_face = "bold",  
 axis\_title\_size = 11) +  
 theme(axis.text.x = element\_blank(),  
 axis.text.y = element\_text(size=11, face="bold"),  
 axis.title.x = element\_text(hjust = 0.5)  
 )+  
 scale\_x\_continuous(expand = expansion(mult = c(0, 0.1)))  
  
top10plot

# Deepdive for more insights

# Total population of the top 10 Planning Area  
cntpoptop10 <- top10PA %>%  
 summarise(Pop = sum(Pop, na.rm = TRUE))   
cat(cntpoptop10$Pop)

2358550

# The percentage of the population of the top 10 most populous Planning Area relative to the total population.  
perc\_pop\_top10 <- (cntpoptop10 / cntpop) \* 100  
cat(perc\_pop\_top10$Pop)

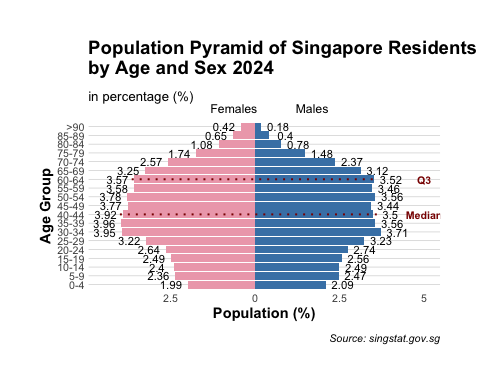
56.32062

#### 🎀 **Insights Plot 1**

* Slightly over half (56.3%) of the 4,187.72 thousand (~4.19 million) residents in Singapore stayed in the top 10 planning areas of residence.
* There were five planning areas with more than 250,000 residents each, namely Tampines, Bedok, Sengkang, Jurong West, and Woodlands.
* Tampines was the most populated with 284,720 residents.

## 5.3 Age Group and Sex distribution

# 🎀 Plot 2

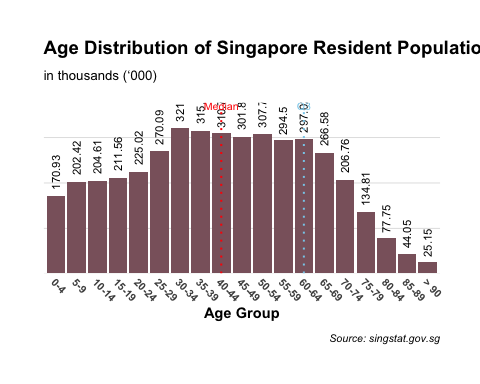


# Code

# Define age group levels and labels  
AG\_levels <- c(  
 "0\_to\_4", "5\_to\_9", "10\_to\_14", "15\_to\_19", "20\_to\_24",  
 "25\_to\_29", "30\_to\_34", "35\_to\_39", "40\_to\_44", "45\_to\_49",  
 "50\_to\_54", "55\_to\_59", "60\_to\_64", "65\_to\_69", "70\_to\_74",  
 "75\_to\_79", "80\_to\_84", "85\_to\_89", "90\_and\_over"  
)  
  
AG\_labels <- c(  
 "0-4", "5-9", "10-14", "15-19", "20-24", "25-29", "30-34", "35-39",  
 "40-44", "45-49", "50-54", "55-59", "60-64", "65-69", "70-74", "75-79",  
 "80-84", "85-89", ">90"  
)  
  
AG\_lookup <- setNames(AG\_labels, AG\_levels)  
  
# Compute population shares by age group  
AGsum2 <- demographic\_data\_clean %>%  
 group\_by(AG) %>%  
 summarise(Pop = sum(Pop, na.rm = TRUE), .groups = "drop") %>%  
 mutate(  
 AG = factor(AG, levels = AG\_levels),  
 Pop\_share = Pop / sum(Pop),  
 cum\_share = cumsum(Pop\_share),  
 percentile = round(cum\_share \* 100, 1)  
 ) %>%  
 arrange(AG)  
  
# Median and Q3 age groups  
median\_AG <- AGsum2 %>% filter(cumsum(Pop) >= sum(Pop) / 2) %>% slice(1) %>% pull(AG)  
q3\_AG <- AGsum2 %>% filter(cumsum(Pop) >= sum(Pop) \* 0.75) %>% slice(1) %>% pull(AG)  
  
# Convert median and Q3 to labels  
median\_label <- AG\_lookup[as.character(median\_AG)]  
q3\_label <- AG\_lookup[as.character(q3\_AG)]  
  
# Prepare population pyramid data  
pyramid\_data <- demographic\_data\_clean %>%  
 group\_by(AG, Sex) %>%  
 summarise(Population = sum(Pop, na.rm = TRUE) / 1000, .groups = "drop") %>%  
 mutate(  
 AG = factor(AG, levels = AG\_levels, labels = AG\_labels, ordered = TRUE),  
 PopPercentage = ifelse(Sex == "Females", -Population, Population) / sum(Population) \* 100,  
 PopPercentage = round(PopPercentage, 2),  
 Signal = ifelse(Sex == "Females", -1, 1)  
 )  
  
  
# Create the population pyramid plot  
pyramid\_plot <- ggplot(pyramid\_data, aes(x = AG, y = PopPercentage, fill = Sex)) +  
 geom\_bar(stat = "identity") +  
 geom\_text(aes(y = PopPercentage + Signal \* 0.5, label = abs(PopPercentage)),  
 size = 3, color = "black") +  
 coord\_flip() +  
 scale\_fill\_manual(values = c("Females" = "pink2", "Males" = "steelblue"),  
 guide = guide\_legend(override.aes = list(fill = NA))) +  
 scale\_y\_continuous(labels = abs) +  
 annotate("segment", x = median\_label, xend = median\_label, y = -4, yend = 3.6,  
 color = "red4", linewidth = 0.7, linetype = "dotted") +  
 annotate("text", x = median\_label, y = 5, label = "Median",  
 color = "red4", size = 2.8, fontface = "bold") +  
 annotate("segment", x = q3\_label, xend = q3\_label, y = -3.65, yend = 3.65,  
 color = "red4", linewidth = 0.7, linetype = "dotted") +  
 annotate("text", x = q3\_label, y = 5, label = "Q3",  
 color = "red4", size = 2.8, fontface = "bold") +  
 ggtitle("Population Pyramid of Singapore Residents\nby Age and Sex 2024",  
 subtitle = "in percentage (%)") +  
 labs(  
 y = "Population (%)",  
 x = "Age Group",  
 fill = "Sex",  
 caption = "Source: singstat.gov.sg"  
 ) +  
 theme\_ipsum(base\_family = "Arial",  
 plot\_title\_size = 14,  
 subtitle\_size = 10,  
 caption\_size = 8,  
 plot\_title\_face = "bold",  
 caption\_face = "italic",  
 grid = "Y",  
 axis\_title\_face = "bold",  
 axis\_title\_size = 11,  
 axis\_text\_size = 8) +  
 theme(  
 strip.text = element\_text(face = "bold"),  
 axis.title.x = element\_text(hjust = 0.5),  
 axis.title.y = element\_text(hjust = 0.5),  
 legend.position = "top",  
 legend.title = element\_blank(),  
 legend.justification = c(0.45, 0),  
 legend.margin = margin(t = -20, r = 0, b = -10, l = 0, unit = "pt")  
 )  
  
pyramid\_plot

# Alternative Plot

Alternatively for more simpler visualisation we can also use this plot, though we can only infer Age Distribution Insights from this plot.



# Deepdive for more insights

# Details of the population pyramid distribution  
print(pyramid\_data, n = 38)

# A tibble: 38 × 5  
 AG Sex Population PopPercentage Signal  
 <ord> <chr> <dbl> <dbl> <dbl>  
 1 0-4 Females 83.4 -1.99 -1  
 2 0-4 Males 87.5 2.09 1  
 3 10-14 Females 100. -2.4 -1  
 4 10-14 Males 104. 2.49 1  
 5 15-19 Females 104. -2.49 -1  
 6 15-19 Males 107. 2.56 1  
 7 20-24 Females 110. -2.64 -1  
 8 20-24 Males 115. 2.74 1  
 9 25-29 Females 135. -3.22 -1  
10 25-29 Males 135. 3.23 1  
11 30-34 Females 166. -3.95 -1  
12 30-34 Males 156. 3.71 1  
13 35-39 Females 166 -3.96 -1  
14 35-39 Males 149. 3.56 1  
15 40-44 Females 164. -3.92 -1  
16 40-44 Males 147. 3.5 1  
17 45-49 Females 158. -3.77 -1  
18 45-49 Males 144. 3.44 1  
19 50-54 Females 158. -3.78 -1  
20 50-54 Males 149. 3.56 1  
21 55-59 Females 150. -3.58 -1  
22 55-59 Males 145. 3.46 1  
23 5-9 Females 98.9 -2.36 -1  
24 5-9 Males 104. 2.47 1  
25 60-64 Females 150. -3.57 -1  
26 60-64 Males 147. 3.52 1  
27 65-69 Females 136. -3.25 -1  
28 65-69 Males 130. 3.12 1  
29 70-74 Females 107. -2.57 -1  
30 70-74 Males 99.3 2.37 1  
31 75-79 Females 72.9 -1.74 -1  
32 75-79 Males 61.9 1.48 1  
33 80-84 Females 45.1 -1.08 -1  
34 80-84 Males 32.7 0.78 1  
35 85-89 Females 27.2 -0.65 -1  
36 85-89 Males 16.9 0.4 1  
37 >90 Females 17.4 -0.42 -1  
38 >90 Males 7.73 0.18 1

# Details of the AG distribution  
AGsum2

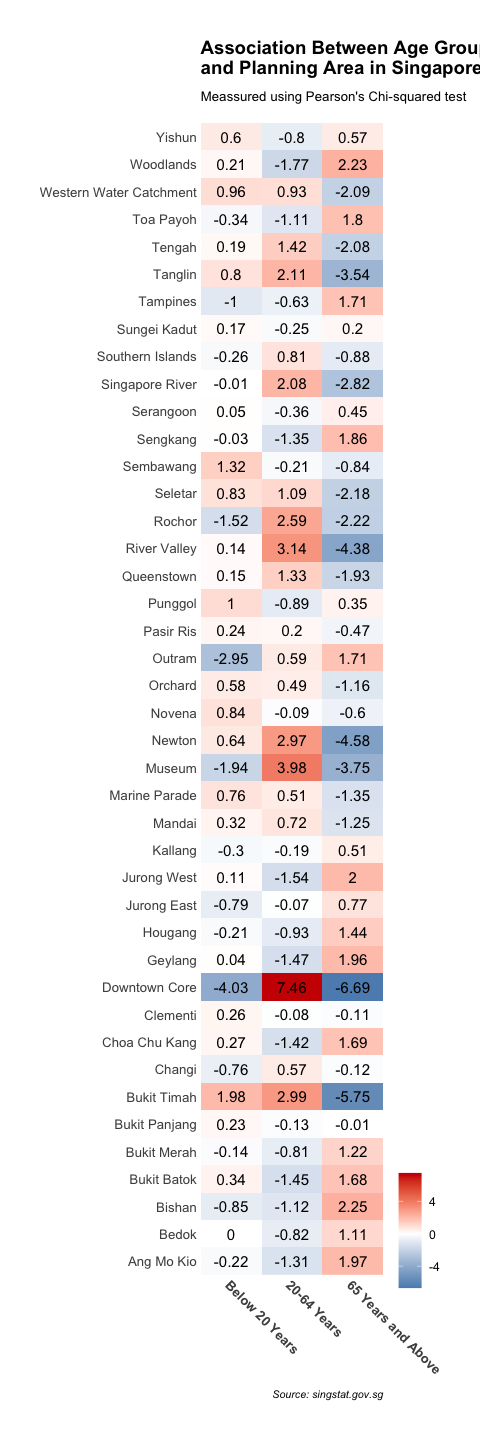
# A tibble: 19 × 5  
 AG Pop Pop\_share cum\_share percentile  
 <fct> <dbl> <dbl> <dbl> <dbl>  
 1 0\_to\_4 170930 0.0408 0.0408 4.1  
 2 5\_to\_9 202420 0.0483 0.0892 8.9  
 3 10\_to\_14 204610 0.0489 0.138 13.8  
 4 15\_to\_19 211560 0.0505 0.189 18.9  
 5 20\_to\_24 225020 0.0537 0.242 24.2  
 6 25\_to\_29 270090 0.0645 0.307 30.7  
 7 30\_to\_34 321010 0.0767 0.383 38.3  
 8 35\_to\_39 315180 0.0753 0.459 45.9  
 9 40\_to\_44 310700 0.0742 0.533 53.3  
10 45\_to\_49 301820 0.0721 0.605 60.5  
11 50\_to\_54 307760 0.0735 0.678 67.8  
12 55\_to\_59 294500 0.0703 0.749 74.9  
13 60\_to\_64 297020 0.0709 0.820 82   
14 65\_to\_69 266580 0.0637 0.883 88.3  
15 70\_to\_74 206760 0.0494 0.933 93.3  
16 75\_to\_79 134810 0.0322 0.965 96.5  
17 80\_to\_84 77750 0.0186 0.983 98.3  
18 85\_to\_89 44050 0.0105 0.994 99.4  
19 90\_and\_over 25150 0.00601 1 100

#### 🎀 **Insights Plot 2**

* The median age falls within 40 to 44 age group. About 25% are aged 60 and above, with 18% aged 65 and older, highlighting [aging population trend](https://www.weforum.org/stories/2023/09/life-expectancy-countries-ageing-populations/).
* Children (0–14) account for only about 13.8% of the population, indicating low birth rates. Slightly more males than females are observed in this group.
* The working-age population (15–64 years), as defined by the [Ministry of Manpower](https://stats.mom.gov.sg/Statistics/Pages/labourforce.aspx), comprises around 68.2% of the population. This reflects a strong labour force, though future demographic challenges may arise as this group continues to age.
* The 25-64 age group shows a near-equal gender balance.
* The gender gap widens in the older age cohorts, with majority female in the 80+ age groups. This show that females live longer than males on average, consistent with the life expectancy at birth between the different gendersfrom [2023 report by the Singapore Department of Statistics](https://www.singstat.gov.sg/-/media/files/publications/population/lifetable22-23.ashx).

## 5.3 Association Between Age Group Distribution and Planning Area in Singapore

# 🎀 Plot 3



# Code

table\_data <- table(demographic\_data\_recode$AG\_recode, demographic\_data\_recode$PA)  
  
chi\_test <- chisq.test(table\_data)  
  
residuals\_df <- as.data.frame(as.table(chi\_test$residuals))  
  
AG\_levels <- c("Below 20 Years", "20-64 Years", "65 Years and Above")  
  
AG\_labels <- c("Below 20 Years", "20-64 Years", "65 Years and Above")  
  
residuals\_df <- residuals\_df %>%  
 mutate(Var1 = factor(Var1, levels = AG\_levels, labels = AG\_labels, ordered = TRUE))  
  
ggplot(residuals\_df, aes(Var1, Var2, fill = Freq)) +  
 geom\_tile() +  
 geom\_text(aes(label = round(Freq, 2)), size = 4) +  
 scale\_fill\_gradient2(low = "steelblue", high = "red3", mid = "white", midpoint = 0) +  
 labs(x = "AG", y = "PA", fill = "Residuals") +  
 ggtitle("Association Between Age Group\nand Planning Area in Singapore 2024",  
 subtitle = "Meassured using Pearson's Chi-squared test") +  
 labs(caption = "Source: singstat.gov.sg")+  
 theme\_ipsum(base\_family = "Arial",  
 plot\_title\_size = 14,  
 subtitle\_size = 10,  
 caption\_size = 8,  
 plot\_title\_face = "bold",  
 caption\_face = "italic",  
 grid = "Y",  
 axis\_title\_face = "bold",  
 axis\_title\_size = 11,  
 axis\_text\_size = 8) +  
 theme(  
 strip.text = element\_text(face = "bold"),  
 axis.text.x = element\_text(size=10, face="bold", angle = -45, hjust = 0),  
 axis.text.y = element\_text(size=10),  
 axis.title.x = element\_blank(),  
 axis.title.y = element\_blank(),  
 legend.title = element\_blank(),  
 legend.justification = c(0.45, 0),  
 legend.margin = margin(t = -20, r = 0, b = -10, l = 0, unit = "pt"))+  
 scale\_x\_discrete(expand = expansion(add = 0.5)) +  
 scale\_y\_discrete(expand = expansion(add = 0.5))

#### 🎀 **Insights Plot 3**

* [Central Areas](https://www.hdb.gov.sg/about-us/history/hdb-towns-your-home/central) like Downtown Core are skewed toward working-age group with strong over-representation of the 20–64 group (+7.46) and lesser children (–4.03) and seniors (–6.69), reflecting its role as a Central Business District. Similar patterns appear in Newton, Museum (Bras Basah, Dhoby Ghaut, Fort Canning), and Singapore River (Boat Quay, Clarke Quay, Robertson Quay).
* New BTO areas like [Tengah](https://www.hdb.gov.sg/about-us/history/hdb-towns-your-home/tengah) and [Punggol](https://www.hdb.gov.sg/about-us/history/hdb-towns-your-home/punggol) show more children and lesser seniors, indicating young families moving into new Towns / Planning Areas.
* [Mature Towns](https://www.channelnewsasia.com/singapore/mature-estates-hdb-flats-classified-cna-explains-3581846#:~:text=The%20mature%20towns%20and%20estates,Serangoon%2C%20Tampines%20and%20Toa%20Payoh.) such as Bedok, Ang Mo Kio, and Hougang exhibit neutral residuals, suggesting stable populations with a mix of age groups and long-term residents.
* Affluent Districts such as [Bukit Timah](https://www.propertyasiadirect.com/sg/places/bukit-timah/#google_vignette), we observe more children (+1.98) and 20-65 group (+2.99) with lesser seniors (-5.75). Proximity to “good schools” for *kiasu* parents likely attracts young wealthy families with its abundance of private housing. Similar trends observed in Tanglin.

# 6 Reference

- [ggplot for categorical-data](https://mq-software-carpentry.github.io/r-ggplot-extension/02-categorical-data/index.html)  
- [Describe function](https://www.geeksforgeeks.org/describe-function-in-r/)  
- [gt package](https://gt.rstudio.com/)  
- [theme for ggplot2](https://ggplot2.tidyverse.org/reference/theme.html)  
- [Recode Values with dplyr](https://dplyr.tidyverse.org/reference/recode.html)  
- [Customize tick marks and labels](https://www.sthda.com/english/wiki/ggplot2-axis-ticks-a-guide-to-customize-tick-marks-and-labels)  
- [National Statistical Standards Recommendations on Definition and Classification of Age](https://www.singstat.gov.sg/-/media/files/standards_and_classifications/nsa.ashx)  
- [Cencus of Population 2020](https://www.parliament.gov.sg/docs/default-source/default-document-library/cop2020sr2.pdf)  
- [Population Pyramid Plot](https://www.njtierney.com/post/2022/08/09/ggplot-pyramid/)  
- [Ageing Population](https://www.weforum.org/stories/2023/09/life-expectancy-countries-ageing-populations/)  
- [Heatmap ggplot2](https://r-charts.com/correlation/heat-map-ggplot2/)  
- [Chi Square in r](https://www.datacamp.com/tutorial/chi-square-test-r)

# 7 Appendix

## 7.1 Appendix A

| **Column Name** | **Description** |
| --- | --- |
| PA | Planning Area |
| SZ | Subzone |
| AG | Age Group |
| Sex | Sex |
| FA | Floor Area of Residence |
| Pop | Resident Count (Population) |
| Time | Time / Period |

## 7.2 Appendix B

[National Statistical Standards Recommendations on Definition and Classification of Age](https://www.singstat.gov.sg/-/media/files/standards_and_classifications/nsa.ashx)

| **AG** | **AG\_recode** |
| --- | --- |
| 0\_to\_4 | Below 20 years |
| 5\_to\_9 | Below 20 years |
| 10\_to\_14 | Below 20 years |
| 15\_to\_19 | Below 20 years |
| 20\_to\_24 | 20-64 Years |
| 25\_to\_29 | 20-64 Years |
| 30\_to\_34 | 20-64 Years |
| 35\_to\_39 | 20-64 Years |
| 40\_to\_44 | 20-64 Years |
| 45\_to\_49 | 20-64 Years |
| 50\_to\_54 | 20-64 Years |
| 55\_to\_59 | 20-64 Years |
| 60\_to\_64 | 20-64 Years |
| 65\_to\_69 | 65 years & Over |
| 70\_to\_74 | 65 years & Over |
| 75\_to\_79 | 65 years & Over |
| 80\_to\_84 | 65 years & Over |
| 85\_to\_89 | 65 years & Over |
| 90\_and\_over | 65 years & Over |