Take-home Exercise 01 phase 2

Yang Lu

2025-05-09

## Overview Task

In this exercise, we select a peer submission of Take-home Exercise 1.

* Critic good design principles
* Recommendations for further improvement
* Present a makeover version of the data

## Getting started

### Load required packages

pacman::p\_load(tidyverse,dplyr,  
 ggplot2, forcats,   
 scales,patchwork)

### Import data

The data used in this exercise is the same as the data used in the Take-home Exercise 1

population\_age\_sex\_2024 <- read\_csv("respopagesex2024/respopagesex2024.csv")  
population\_age\_sex\_2024

# A tibble: 60,424 × 6  
 PA SZ Age Sex Pop Time  
 <chr> <chr> <chr> <chr> <dbl> <dbl>  
 1 Ang Mo Kio Ang Mo Kio Town Centre 0 Males 10 2024  
 2 Ang Mo Kio Ang Mo Kio Town Centre 0 Females 10 2024  
 3 Ang Mo Kio Ang Mo Kio Town Centre 1 Males 10 2024  
 4 Ang Mo Kio Ang Mo Kio Town Centre 1 Females 10 2024  
 5 Ang Mo Kio Ang Mo Kio Town Centre 2 Males 10 2024  
 6 Ang Mo Kio Ang Mo Kio Town Centre 2 Females 10 2024  
 7 Ang Mo Kio Ang Mo Kio Town Centre 3 Males 10 2024  
 8 Ang Mo Kio Ang Mo Kio Town Centre 3 Females 10 2024  
 9 Ang Mo Kio Ang Mo Kio Town Centre 4 Males 30 2024  
10 Ang Mo Kio Ang Mo Kio Town Centre 4 Females 10 2024  
# ℹ 60,414 more rows

### Data Processing

str(population\_age\_sex\_2024)

spc\_tbl\_ [60,424 × 6] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
 $ PA : chr [1:60424] "Ang Mo Kio" "Ang Mo Kio" "Ang Mo Kio" "Ang Mo Kio" ...  
 $ SZ : chr [1:60424] "Ang Mo Kio Town Centre" "Ang Mo Kio Town Centre" "Ang Mo Kio Town Centre" "Ang Mo Kio Town Centre" ...  
 $ Age : chr [1:60424] "0" "0" "1" "1" ...  
 $ Sex : chr [1:60424] "Males" "Females" "Males" "Females" ...  
 $ Pop : num [1:60424] 10 10 10 10 10 10 10 10 30 10 ...  
 $ Time: num [1:60424] 2024 2024 2024 2024 2024 ...  
 - attr(\*, "spec")=  
 .. cols(  
 .. PA = col\_character(),  
 .. SZ = col\_character(),  
 .. Age = col\_character(),  
 .. Sex = col\_character(),  
 .. Pop = col\_double(),  
 .. Time = col\_double()  
 .. )  
 - attr(\*, "problems")=<externalptr>

sort(unique(population\_age\_sex\_2024$Age))

[1] "0" "1" "10" "11" "12"   
 [6] "13" "14" "15" "16" "17"   
[11] "18" "19" "2" "20" "21"   
[16] "22" "23" "24" "25" "26"   
[21] "27" "28" "29" "3" "30"   
[26] "31" "32" "33" "34" "35"   
[31] "36" "37" "38" "39" "4"   
[36] "40" "41" "42" "43" "44"   
[41] "45" "46" "47" "48" "49"   
[46] "5" "50" "51" "52" "53"   
[51] "54" "55" "56" "57" "58"   
[56] "59" "6" "60" "61" "62"   
[61] "63" "64" "65" "66" "67"   
[66] "68" "69" "7" "70" "71"   
[71] "72" "73" "74" "75" "76"   
[76] "77" "78" "79" "8" "80"   
[81] "81" "82" "83" "84" "85"   
[86] "86" "87" "88" "89" "9"   
[91] "90\_and\_Over"

#### Creating a Numeric Age Column

df <- population\_age\_sex\_2024 %>%  
 mutate(  
 AgeNum = suppressWarnings(  
 ifelse(Age == "90\_and\_Over", 90, as.numeric(Age))  
 )  
 )

#### Creating Age Grouping

df <- df %>%  
 mutate(  
 AgeGroup = case\_when(  
 AgeNum <= 12 ~ "Child",  
 AgeNum <= 24 ~ "Youth",  
 AgeNum <= 64 ~ "Adult",  
 TRUE ~ "Senior"  
 )  
 )

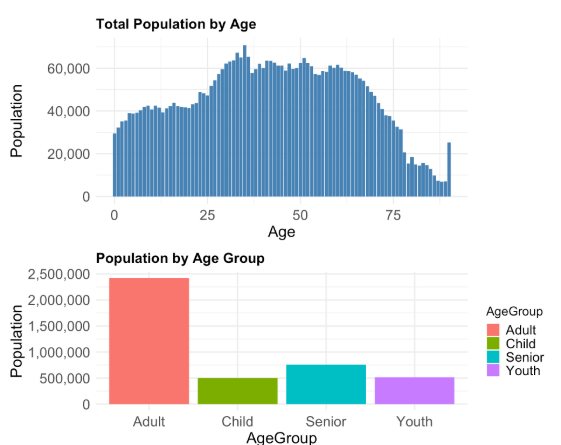
## Visualization: **Population by Age / Age Group**

### Original Design

This visualization shows two charts for Singapore’s resident population (2024):

1. **Total Population by Age** – population by single-age bar chart
2. **Population by Age Group** – a 4-category bar chart (Child, Youth, Adult, Senior)

The original design is shown below.



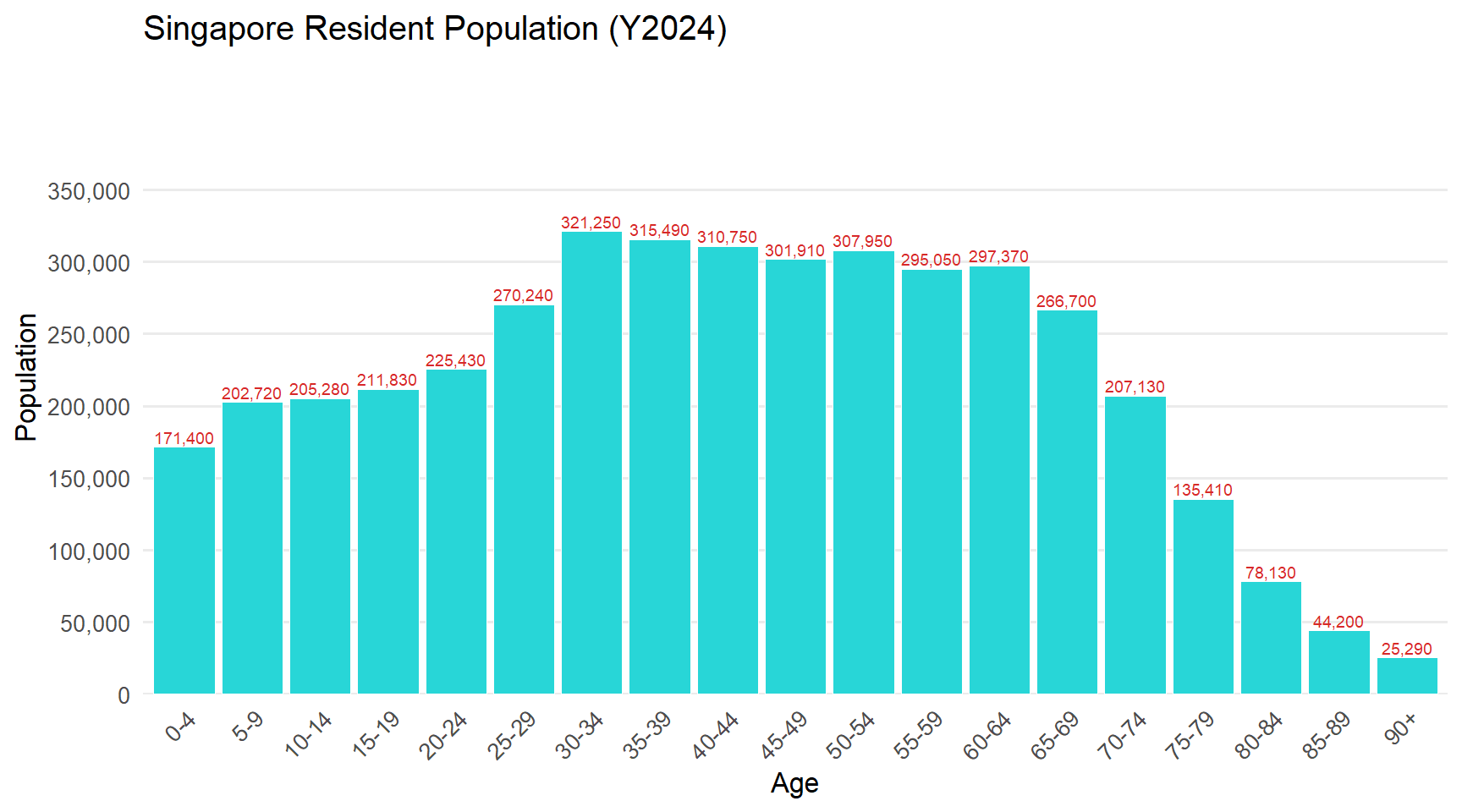
### **Critique for Population by Age / Age Group**

|  |
| --- |
| Clarity |
| **Clear Aspects**   * Bars are labeled and vertically aligned, allowing easy mapping to axis values. * Axis titles (“Age” / “Age-Group” and “Population”) are clearly stated. * Insights below the charts succinctly summarize the key takeaways.   **Confusing Aspects**   * For Total Population by Age chart   + The x-axis includes every single age, making it difficult to identify exact values for each one.   + The age on the x-axis cuts off after age 75, making the data appear incomplete.   + The y-axis has a maximum of 60,000, but some age bars exceeded above the limit and appear visually clipped. * Population by Age Group chart   + Age groups are labeled as “Child,” “Youth,” etc., without their numeric ranges (e.g. 0–14, 15–24), making the groupings harder to interpret.   + The order of age groups “Adult, Child, Senior, Youth,” does not follow a logical age progression, making it harder to grasp at a glance which can confuse viewers on first impression. |

|  |
| --- |
| Aesthetics |
| **Visually Appealing**   * The minimalist design keeps the focus on the bars by reducing distractions. * Bold chart titles and axis labels enhance’s readability.   **Visually Distracting**   * For Total Population by Age chart   + The wide age range makes it hard to distinguish where one age group ends and another begings.   + Bars do not display actual population values for easier understanding at first glance. * For Population by Age Group chart   + The absence of data labels on bars forces viewers to rely on the y-axis to estimate values. |

### Remake

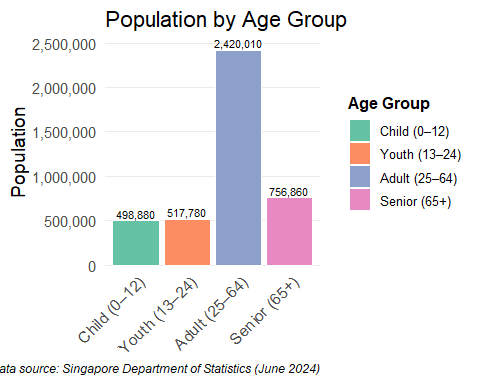
### Total Population by Age (The Plot)



### Total Population by Age (The code chunk)

library(tidyverse)  
library(scales)  
  
bin\_labels <- c(paste0(seq(0, 85, by = 5), "-", seq(4, 89, by = 5)), "90+")  
  
pop\_5yr <- df %>%  
 mutate(  
 AgeNum = suppressWarnings(  
 ifelse(Age %in% c("90\_and\_over","90\_and\_above"), 90, as.numeric(Age))  
 ),  
 AgeBin = cut(  
 AgeNum,  
 breaks = c(seq(0, 90, by = 5), Inf),  
 labels = bin\_labels,  
 right = FALSE  
 )  
 ) %>%  
 replace\_na(list(AgeBin = "90+")) %>%  
 group\_by(AgeBin) %>%  
 summarise(Pop = sum(Pop, na.rm = TRUE), .groups = "drop") %>%  
 mutate(AgeBin = factor(AgeBin, levels = bin\_labels))  
  
  
ggplot(pop\_5yr, aes(x = AgeBin, y = Pop)) +  
 geom\_col(fill = "#28D6D7", color = "white") +  
 geom\_text(  
 aes(label = comma(Pop)),  
 vjust = -0.3,   
 size = 5,   
 color = "#D72928"  
 ) +  
 scale\_y\_continuous(  
 breaks = seq(0, 350000, 50000),  
 limits = c(0, 440000),   
 labels = comma,  
 expand = expansion(add = c(0, 0))  
 ) +  
 labs(  
 title = "Singapore Resident Population (Y2024)",  
 x = "Age",  
 y = "Population"  
 ) +  
 theme\_minimal(base\_size = 25) +  
 theme(  
 axis.text.x = element\_text(angle = 45, hjust = 1),  
 panel.grid.major.x = element\_blank(),  
 panel.grid.minor = element\_blank()  
 )

### population by age group (The Plot)



### population by age group (The code chunk )

library(tidyverse)  
library(RColorBrewer)  
library(scales)  
  
df2 <- df %>%  
 mutate(  
 AgeNum = suppressWarnings(  
 ifelse(Age %in% c("90\_and\_over","90\_and\_above"),  
 90,  
 as.numeric(Age))  
 )  
 ) %>%  
  
 mutate(  
 AgeGroup = case\_when(  
 AgeNum <= 12 ~ "Child (0–12)",  
 AgeNum <= 24 ~ "Youth (13–24)",  
 AgeNum <= 64 ~ "Adult (25–64)",  
 TRUE ~ "Senior (65+)"  
 ),  
 AgeGroup = factor(  
 AgeGroup,  
 levels = c("Child (0–12)",  
 "Youth (13–24)",  
 "Adult (25–64)",  
 "Senior (65+)")  
 )  
 )  
  
p2\_new <- df2 %>%  
 group\_by(AgeGroup) %>%  
 summarise(Pop = sum(Pop, na.rm = TRUE), .groups = "drop") %>%  
 ggplot(aes(x = AgeGroup, y = Pop, fill = AgeGroup)) +  
 geom\_col(color = "white") +  
 geom\_text(aes(label = comma(Pop)), vjust = -0.3, size = 3) +  
 scale\_y\_continuous(labels = comma, expand = expansion(mult = c(0, .05))) +  
 scale\_fill\_brewer(palette = "Set2", name = "Age Group") +  
 labs(  
 title = "Population by Age Group",  
 x = NULL,  
 y = "Population",  
 caption = "Data source: Singapore Department of Statistics (June 2024)"  
 ) +  
 theme\_minimal(base\_size = 14) +  
 theme(  
 legend.position = "right",  
 legend.title = element\_text(size = 12, face = "bold"),  
 legend.text = element\_text(size = 10),  
 axis.text.x = element\_text(angle = 45, hjust = 1, size = 12),  
 panel.grid.major.x = element\_blank(),  
 panel.grid.minor = element\_blank(),  
 plot.caption = element\_text(size = 9, hjust = 1, face = "italic")  
 )  
  
print(p2\_new)

|  |
| --- |
| Refinements |
| * **Clearer Age Grouping**: Replaced 91 single-age bins with broader 5-year intervals (e.g., 0–4, 5–9, …, 85–89, 90+), making cohort sizes easier to interpret at a glance. * **Labeled Group Names**: Added numeric ranges to group labels — e.g., “Child (0–12)”, “Youth (13–24)” — so each label conveys full meaning without requiring external reference. * **Value Labels on Bars**: Population counts are now displayed directly above each bar in a contrasting color, reducing reliance on the y-axis for exact values. * **Improved Text Legibility**: Increased base font sizes (including for titles, axis labels, and bar labels), ensuring readability even in high-resolution displays or print. |

## Summary and conclusion

By grouping ages into 5-year bins and enriching the labels, we transformed a busy histogram and a loosely ordered bar chart into concise, self-contained visuals. These redesigned plots not only improve readability and accessibility but also make the demographic story of Singapore—its strong working-age base, shrinking youth cohorts, and emerging aging population—more transparent for policymakers and stakeholders.

## **Reference**

<https://cabbage-wonderland.netlify.app/takehome_ex/takehome_ex01/takehome_ex01#overview>