Take-home\_Ex01-phaseII

Lucas Luo

2025-05-09

## ✅ Review of Peer Submission: David Chiam

refer:https://isss608-davidchiam.netlify.app/take-home\_ex/take-home\_ex01/take-home\_ex01

### 🌟 Three Good Design Principles

**1. Effective Use of Visual Encodings**  
David effectively applies core visualisation techniques by using mirrored bar charts for population pyramids and boxplots for median age distribution. In particular:

* The mirrored bars in the pyramid chart enable immediate comparison between male and female cohorts.
* The use of boxplots is apt for showing variability and central tendency (e.g., median age) across planning areas.
* In the stacked bar chart, gender and age group differences are clearly segmented, leveraging hue and spatial separation.

**2. Logical Flow and Visual Hierarchy**  
The visualisation follows a logical and structured flow:

* It begins with a high-level overview (e.g., gender share), then progressively narrows to specific planning area comparisons.
* This top-down design facilitates cognitive flow and narrative pacing.
* Clear section titles and cohesive subplot sequencing enhance usability.

**3. Consistent and Minimalist Visual Theme**  
David’s work maintains high aesthetic discipline:

* Muted pastel tones reduce cognitive overload.
* Consistent typography and plot margin spacing lend a polished appearance.
* Axes, titles, and legends are harmonised across charts, which ensures clarity without redundancy.

### ⚠️ Three Areas for Improvement (and How We Address Them)

**1. Lack of Annotations and Contextual Labels**  
While David’s charts are clean, they often lack direct data labels or annotations. This limits interpretability:

* ✅ *Fix:* The makeover scatterplot adds dynamic labeling for outliers (e.g., Outram with high elderly share), improving insight at-a-glance.

**2. Axis and Formatting Inconsistencies**  
There are inconsistencies in axis label units (e.g., some show absolute numbers, others show percentages):

* ✅ *Fix:* In our makeover, all values are formatted uniformly using percent\_format() and /1000 suffixes for population.

**3. Redundant Charts and Missed Insight Opportunities**  
Some charts display similar gender breakdowns without revealing new dimensions. At the same time, insightful metrics like the elderly/youth ratio are missing:

* ✅ *Fix:* We introduced a derived metric (Elderly\_to\_Youth) and plotted it against population size and aging ratio, unveiling patterns not covered in the original submission.

### 🛠️ Makeover: Intergenerational Analysis (Static Version)

Load required packages 📦

pacman::p\_load(  
 tidyverse, # Core data manipulation and ggplot2 plotting  
 knitr, # Dynamic report generation and rendering  
 patchwork, # Combines multiple ggplot2 plots  
 ggthemes, # Additional themes and visual styles  
 scales, # Formatting axes, scales, and labels  
 ggtext, # Rich text formatting in ggplot titles/labels  
 ggalt, # Dumbbell, lollipop, and other advanced geoms  
 ggridges, # Ridgeline plots to visualize distributions  
 ggpubr, # Publication-ready ggplot2 enhancements  
 ggExtra, # Marginal plots for added insight  
 cowplot, # Publication-quality figure alignment  
 ggnewscale, # Allows multiple color/fill scales in one ggplot  
 ggrepel # Allows  
)

Loading the Data and Taking a Glimpse of the Dataset 📂

# Load the dataset  
df <- read\_csv("data/respopagesex2024.csv")

The dataset contains 60,424 rows and 6 columns. Each row represents the count of residents for a specific combination of planning area, subzone, age, sex, and year.

glimpse(df)

Rows: 60,424  
Columns: 6  
$ PA <chr> "Ang Mo Kio", "Ang Mo Kio", "Ang Mo Kio", "Ang Mo Kio", "Ang Mo K…  
$ SZ <chr> "Ang Mo Kio Town Centre", "Ang Mo Kio Town Centre", "Ang Mo Kio T…  
$ Age <chr> "0", "0", "1", "1", "2", "2", "3", "3", "4", "4", "5", "5", "6", …  
$ Sex <chr> "Males", "Females", "Males", "Females", "Males", "Females", "Male…  
$ Pop <dbl> 10, 10, 10, 10, 10, 10, 10, 10, 30, 10, 20, 10, 20, 30, 30, 10, 3…  
$ Time <dbl> 2024, 2024, 2024, 2024, 2024, 2024, 2024, 2024, 2024, 2024, 2024,…

Wrangling, Cleaning & Transformation 🧹

df <- df %>%  
 mutate(Age = as.integer(Age)) %>% # Convert to numeric (100+ becomes NA)  
 drop\_na(Age) # Remove missing values caused by conversion  
  
df <- df %>%  
 mutate(AgeBand = cut(Age,  
 breaks = c(0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55,  
 60, 65, 70, 75, 80, 85, 90, 95, 100, Inf),  
 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", "90-94", "95-99", "100+"),  
 right = FALSE))  
  
# For population pyramid (Age-Sex)  
pyramid\_data <- df %>%  
 group\_by(Age, Sex) %>%  
 summarise(Pop = sum(Pop), .groups = "drop")  
  
# For top 5 planning areas  
top5\_pa <- df %>%  
 group\_by(PA) %>%  
 summarise(Pop = sum(Pop), .groups = "drop") %>%  
 slice\_max(order\_by = Pop, n = 5)  
  
df <- df %>%  
 mutate(  
 PA = str\_trim(str\_to\_upper(PA)),  
 Region = case\_when(  
 PA %in% c("BISHAN", "BUKIT MERAH", "BUKIT TIMAH", "CENTRAL AREA", "DOWNTOWN CORE",  
 "GEYLANG", "KALLANG", "MARINA EAST", "MARINA SOUTH", "MUSEUM", "NEWTON",  
 "NOVENA", "ORCHARD", "OUTRAM", "PAYA LEBAR", "QUEENSTOWN", "RIVER VALLEY",  
 "ROCHOR", "SINGAPORE RIVER", "SOUTHERN ISLANDS", "STRAITS VIEW", "TANGLIN",  
 "TOA PAYOH", "MARINE PARADE") ~ "Central",  
 PA %in% c("BEDOK", "CHANGI", "CHANGI BAY", "PASIR RIS", "TAMPINES") ~ "East",  
 PA %in% c("CENTRAL WATER CATCHMENT", "MANDAI", "SEMBABWANG", "SIMPANG",  
 "SUNGEI KADUT", "WOODLANDS", "YISHUN", "SEMBAWANG") ~ "North",  
 PA %in% c("ANG MO KIO", "HOUGANG", "NORTH-EASTERN ISLANDS", "PUNGGOL",  
 "SELETAR", "SENGKANG", "SERANGOON") ~ "North-East",  
 PA %in% c("BOON LAY", "BUKIT BATOK", "BUKIT PANJANG", "CHOA CHU KANG",  
 "CLEMENTI", "JURONG EAST", "JURONG WEST", "LIM CHU KANG",  
 "PIONEER", "TENGAH", "TUAS", "WESTERN ISLANDS",  
 "WESTERN WATER CATCHMENT") ~ "West",  
 TRUE ~ NA\_character\_  
 )  
 )  
  
df\_median <- df %>%  
 arrange(Age) %>%  
 mutate(  
 cum\_pop = cumsum(Pop),  
 total\_pop = sum(Pop)  
 ) %>%  
 filter(cum\_pop >= total\_pop / 2) %>%  
 slice(1)  
  
median\_age <- df\_median$Age  
  
df <- df %>%  
 mutate(  
 GenerationBand = case\_when(  
 Age <= 24 ~ "Alpha & Gen Z (0–24)",  
 Age <= 34 ~ "Young Adults (25–34)",  
 Age <= 54 ~ "Working Adults (35–54)",  
 Age <= 64 ~ "Silver Generation (55–64)",  
 Age <= 74 ~ "Merdeka Generation (65–74)",  
 TRUE ~ "Pioneer Generation (75+)"  
 ),  
 GenerationBand = factor(GenerationBand, levels = c(  
 "Alpha & Gen Z (0–24)", "Young Adults (25–34)", "Working Adults (35–54)",  
 "Silver Generation (55–64)", "Merdeka Generation (65–74)", "Pioneer Generation (75+)"  
 ))  
 )

🔍 Interpretation

This interactive scatterplot allows users to:

Identify areas with high elderly population shares despite small total populations (e.g., Outram).

Compare elderly shares across all areas by hovering over each dot.

Understand population pressure points for aging infrastructure and services.

📌 Conclusion

David’s work is well-designed and informative. With just a few enhancements such as interactivity, annotation, and layout consistency, the overall impact of the visuals can be significantly increased. The proposed makeover demonstrates how a single interactive visual can provide richer engagement while maintaining visual clarity.