Take-home\_Ex01b

Shermainn

2025-05-05

This is on the Phase ll of Take-home exercise 1 where I will be evaluating and critiquing on one of my peer’s, Celin, Take-home exercise 1 from her Netlify shown [here](https://isss608celin.netlify.app/take-home_ex/take-home_ex01/take-home_ex01).

# 1. Overview

In this phase 2, I will be providing 3 good design principles and 3 areas for further improvements for data visualization of plots done by Celin.

# 2. Critique of Plots

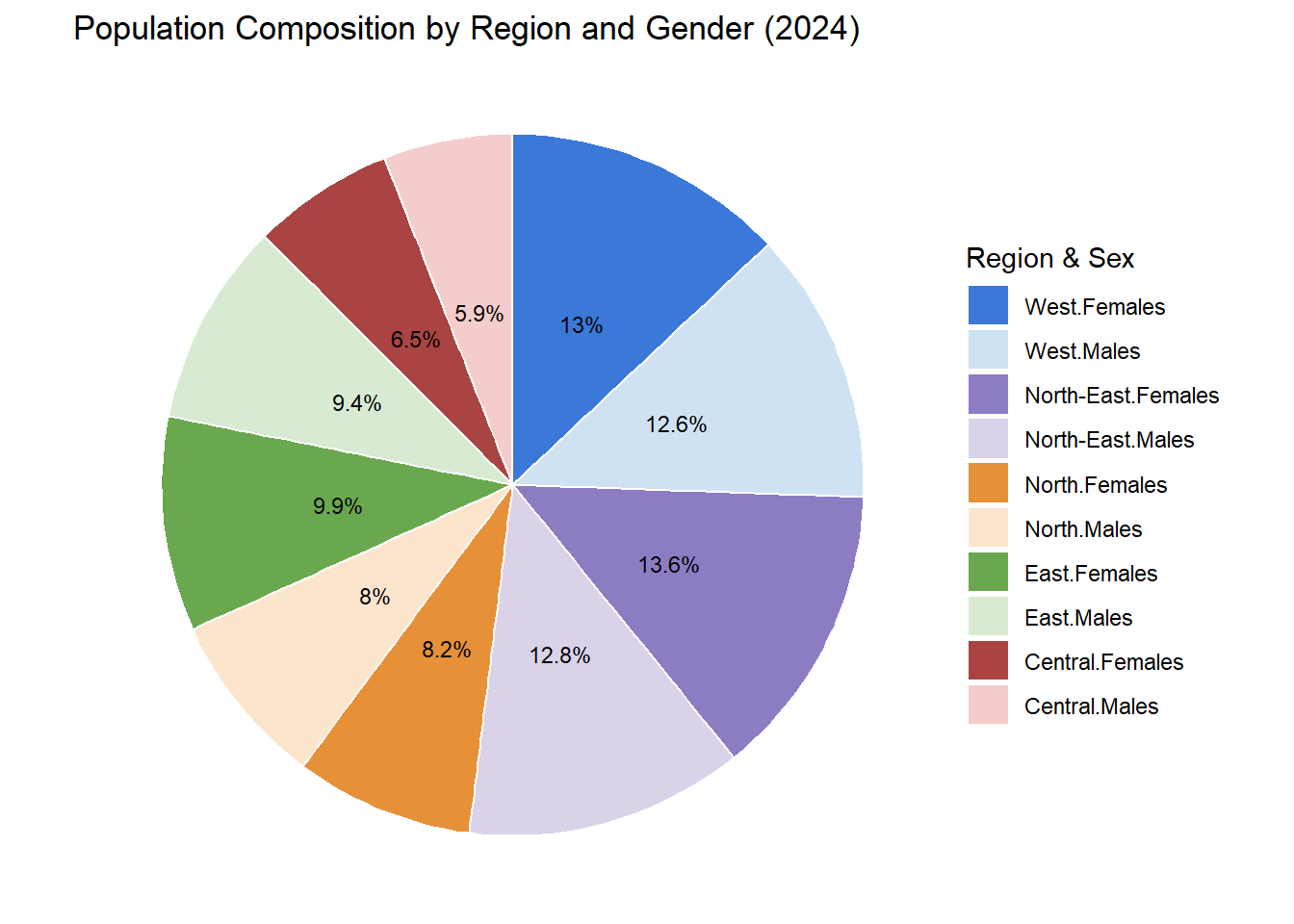
## Pie Chart by Region

For the pie chart shown below, there are some good design principles such as:

* Labeling of slices: The chart labels shown on the legend directly corresponds with the category and percentage shown.
* Flat 2D design: This avoids the distortion of the pie chart that could happen when using a 3D plot.
* Sorted Slices in descending order: As the slices are arranged sequentially from biggest to smallest in a clockwise direction, this helps me to quickly grasp the distribution of values.

Areas to improve on:

* Limit the chart to one variable: The pie chart should only show one factor instead of comparing two factors as it is quite difficult to interpret visually. It would be clearer if it is focused on one factor - such as gender and use colour variations to represent the regions within that category.
* Unclear percentage meaning: I am unable to distinguish what the total percentage actually represent - whether they are referirng to the distribution by region within the total population or by gender when I look at the chart visually. This ambiguity makes the chart pretty confusing and would need further clarification
* Improve colour transparency: This makes the chart harder to distinguish between sections. Hence, reduce the opacity or use more contrasting colours to enhance the readability



# 3. Makeover design of the Pie Chart

## Cleaned Data

pacman::p\_load(tidyverse, ggrepel,   
 ggthemes, hrbrthemes,  
 patchwork, dplyr)   
  
library(readxl)  
  
resident\_data <- read\_excel("C:\\shermainn\\ISSS608new\\Take-home\_Ex\\Take-home\_Ex01\\data\\respopagesex2024e.xlsx")  
  
any(duplicated(resident\_data))

[1] FALSE

colSums(is.na(resident\_data))

Planning Area Subzone Age Sex 2024   
 1 6 6 6 6

resident\_data\_clean <- resident\_data %>%  
 rename(  
 Planning\_Area = `Planning Area`,  
 Subzone = Subzone,  
 Population = `2024`  
 ) %>%  
 # Convert Age to "90+" if it's "90\_and\_Over"  
 mutate(  
 Age = ifelse(Age == "90\_and\_Over", "90", Age),  
 Population = as.numeric(Population),  
 Age = factor(Age, levels = as.character(0:90)),  
 Age\_numeric = as.numeric(as.character(Age))   
 )  
  
# 1. Step 1  
region\_map <- list(  
 North = c(  
 "Central Water Catchment", "Lim Chu Kang", "Mandai",  
 "Sembawang", "Simpang", "Sungei Kadut", "Woodlands", "Yishun"  
 ),  
 `North-East` = c(  
 "Ang Mo Kio", "Hougang", "Punggol", "Sengkang",   
 "Serangoon", "Seletar"  
 ),  
 East = c(  
 "Bedok", "Changi", "Changi Bay",   
 "Pasir Ris", "Paya Lebar", "Tampines"  
 ),  
 West = c(  
 "Boon Lay", "Bukit Batok", "Bukit Panjang", "Choa Chu Kang",  
 "Clementi", "Jurong East", "Jurong West", "Pioneer",   
 "Tengah", "Tuas", "Western Water Catchment"  
 ),  
 Central = c(  
 "Bishan", "Bukit Merah", "Bukit Timah", "Central Area",  
 "Downtown Core", "Marina East", "Marina South", "Museum",  
 "Newton", "Novena", "Orchard", "Outram", "River Valley",  
 "Rochor", "Singapore River", "Straits View", "Tanglin"  
 )  
)  
  
# Step 2  
resident\_regioned <- resident\_data\_clean %>%  
 mutate(  
 Region = case\_when(  
 Planning\_Area %in% region\_map$North ~ "North",  
 Planning\_Area %in% region\_map$`North-East` ~ "North-East",  
 Planning\_Area %in% region\_map$East ~ "East",  
 Planning\_Area %in% region\_map$West ~ "West",  
 Planning\_Area %in% region\_map$Central ~ "Central",  
 TRUE ~ NA\_character\_  
 )  
 ) %>%  
 filter(!is.na(Region))  
  
head(resident\_regioned)

# A tibble: 6 × 7  
 Planning\_Area Subzone Age Sex Population Age\_numeric Region   
 <chr> <chr> <fct> <chr> <dbl> <dbl> <chr>   
1 Ang Mo Kio Total <NA> Total 159340 NA North-East  
2 Ang Mo Kio Total <NA> Males 75600 NA North-East  
3 Ang Mo Kio Total <NA> Females 83740 NA North-East  
4 Ang Mo Kio Total 0 Total 810 0 North-East  
5 Ang Mo Kio Total 0 Males 410 0 North-East  
6 Ang Mo Kio Total 0 Females 410 0 North-East

## Modified Pie Chart

pie\_all\_regions <- resident\_regioned %>%  
 group\_by(Region, Sex) %>%  
 summarise(Population = sum(Population, na.rm = TRUE), .groups = "drop") %>%  
 filter(!is.na(Population) & Population > 0)  
  
male\_data <- pie\_all\_regions %>%  
 filter(Sex == "Males") %>%  
 mutate(  
 label = paste0(Region, "\n", round(Population / sum(Population) \* 100, 1), "%")  
 )  
  
female\_data <- pie\_all\_regions %>%  
 filter(Sex == "Females") %>%  
 mutate(  
 label = paste0(Region, "\n", round(Population / sum(Population) \* 100, 1), "%")  
 )  
  
region\_colors <- c(  
 "Central" = "#f4cccc",  
 "East" = "#d9ead3",  
 "North" = "#fce5cd",  
 "North-East" = "#d9d2e9",  
 "West" = "#cfe2f3"  
)  
  
# Plot: Two pie charts with direct labels  
male\_pie <- ggplot(male\_data, aes(x = "", y = Population, fill = Region)) +  
 geom\_col(width = 1, color = "white") +  
 geom\_text(aes(label = label), position = position\_stack(vjust = 0.5), size = 3) +  
 coord\_polar(theta = "y") +  
 theme\_void(base\_size = 11) +  
 labs(title = "Population by Region (Males)") +  
 scale\_fill\_manual(values = region\_colors) +  
 theme(legend.position = "none")  
  
female\_pie <- ggplot(female\_data, aes(x = "", y = Population, fill = Region)) +  
 geom\_col(width = 1, color = "white") +  
 geom\_text(aes(label = label), position = position\_stack(vjust = 0.5), size = 3) +  
 coord\_polar(theta = "y") +  
 theme\_void(base\_size = 11) +  
 labs(title = "Population by Region (Females)") +  
 scale\_fill\_manual(values = region\_colors) +  
 theme(legend.position = "none")  
  
male\_pie + female\_pie

