Take-home Exercise 1 Part 2

Vanessa Riadi

2025-05-09

# 1 Introduction, Package and Data Load, Data Preparation

I will be reviewing this pie chart from [LIAW YING TING CELIN](https://isss608celin.netlify.app/take-home_ex/take-home_ex01/take-home_ex01#pie-chart-by-regions)

First we load all the necessary packages and data preparation that she has done.

# Data

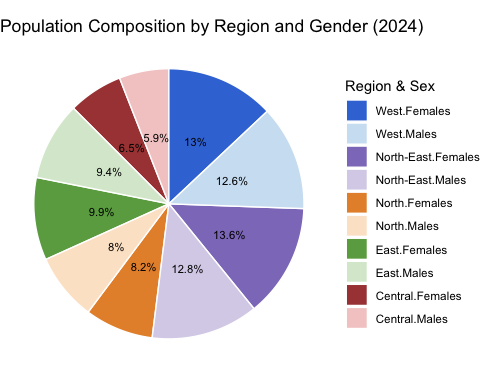
# A tibble: 6 × 8  
 Planning\_Area Subzone Age Sex Population Time Age\_numeric Region  
 <chr> <chr> <fct> <chr> <dbl> <dbl> <dbl> <chr>   
1 Ang Mo Kio Ang Mo Kio Town… 0 Males 10 2024 0 North…  
2 Ang Mo Kio Ang Mo Kio Town… 0 Fema… 10 2024 0 North…  
3 Ang Mo Kio Ang Mo Kio Town… 1 Males 10 2024 1 North…  
4 Ang Mo Kio Ang Mo Kio Town… 1 Fema… 10 2024 1 North…  
5 Ang Mo Kio Ang Mo Kio Town… 2 Males 10 2024 2 North…  
6 Ang Mo Kio Ang Mo Kio Town… 2 Fema… 10 2024 2 North…

# Code

pacman::p\_load(tidyverse, ggrepel,   
 ggthemes, hrbrthemes,  
 patchwork, dplyr)   
  
resident\_data <- read\_csv("data/respopagesex2024.csv")  
  
resident\_data\_clean <- resident\_data %>%  
 rename(  
 Planning\_Area = PA,  
 Subzone = SZ,  
 Population = Pop  
 ) %>%  
 # 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)

# 2 Original Visualisation: Pie Chart By Regions

## The Plot

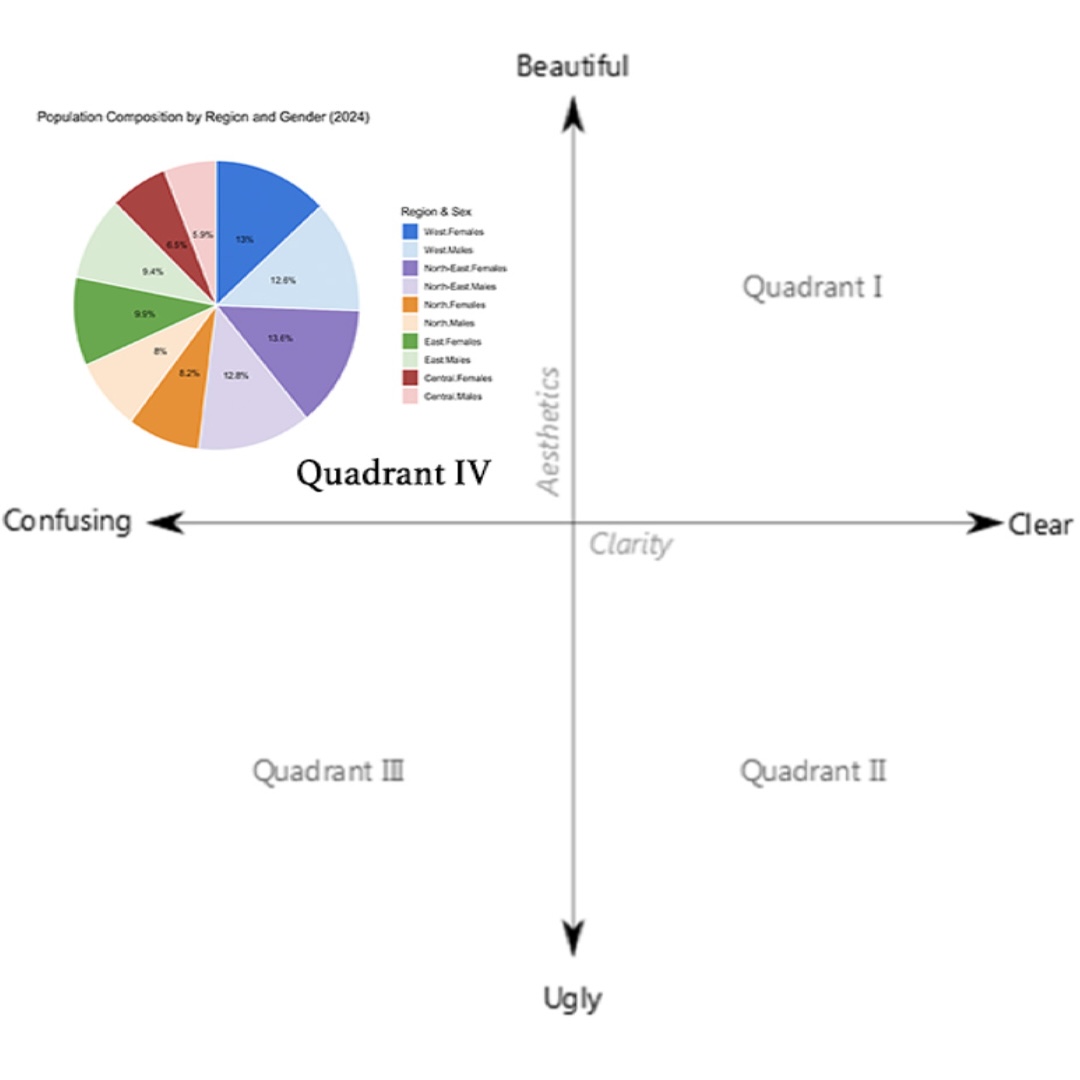


## The Code

# Summarise population by Region and Sex  
pie\_all\_regions <- resident\_regioned %>%  
 group\_by(Region, Sex) %>%  
 summarise(Population = sum(Population), .groups = "drop")  
  
  
# Define the correct stacking order for pie slices  
ordered\_levels <- c(  
 "Central.Males", "Central.Females",  
 "East.Males", "East.Females",  
 "North.Males", "North.Females",  
 "North-East.Males", "North-East.Females",  
 "West.Males", "West.Females"  
)  
  
# Prepare the data  
pie\_all\_regions <- pie\_all\_regions %>%  
 mutate(  
 RegionSex = factor(paste(Region, Sex, sep = "."), levels = ordered\_levels),  
 label = paste0(round(Population / sum(Population) \* 100, 1), "%")  
 )  
  
# Plot: simple, clean pie with percentages inside  
ggplot(pie\_all\_regions, aes(x = "", y = Population, fill = RegionSex)) +  
 geom\_col(width = 1, color = "white") +  
 geom\_text(  
 aes(label = label),  
 position = position\_stack(vjust = 0.5),  
 size = 3,  
 color = "black"  
 ) +  
 coord\_polar(theta = "y") +  
 theme\_void(base\_size = 11) +  
 labs(  
 title = "Population Composition by Region and Gender (2024)",  
 fill = "Region & Sex"  
 ) +  
 scale\_fill\_manual(  
 values = c(  
 "Central.Males" = "#f4cccc",  
 "Central.Females" = "#a94442",  
 "East.Males" = "#d9ead3",  
 "East.Females" = "#6aa84f",  
 "North.Males" = "#fce5cd",  
 "North.Females" = "#e69138",  
 "North-East.Males" = "#d9d2e9",  
 "North-East.Females"= "#8e7cc3",  
 "West.Males" = "#cfe2f3",  
 "West.Females" = "#3c78d8"  
 ),  
 breaks = rev(ordered\_levels)  
 )

# 3 Learning from the Original Visualisation

We learned from [Lesson 2](https://isss608-ay2024-25apr.netlify.app/lesson/lesson02/lesson02-designing%20graphs%20to%20enlighten#a-tale-of-four-quadrants) about [**The Four Quadrant**](https://dataremixed.com/2012/05/data-visualization-clarity-or-aesthetics/) of Data Visualization and the consideration between **Clarity** or **Aesthetics**.



Source: [Data Visualisation: Clarity or Aesthetics](http://dataremixed.com/2012/05/data-visualization-clarity-or-aesthetics/)

Using The Four Quadrants, we evaluate three good design principles and three areas for further improvements.

Using the principals, we can classify this in the: **Quadrant IV – Confusing yet Beautiful**

## 3.1 Three good design principles Why is it “beautiful”?

* Color coding is visually appealing.
* Legend is complete and match the segment. It’s also located where it’s easy to find.
* Clean and clear font choice with percentage labels are neatly placed inside the slices.

## 3.2 Three areas for further improvement Why is it “confusing”?

* There’s no clear visual grouping of the same region. For example, “Central.Males” and “Central.Females” are separated visually, making it hard to compare within regions.
  + It also blends region and gender into a single variable without grouping, which prevents high-level insights (e.g., total by gender or region).
* Pie Chart might not be ideal for this kind of “demographic” data with so many regions.
  + If you still want to stick with a pie chart, you can create a [**sunburst plot**](https://www.simplesheets.co/blog/sunburst-chart) to show the grouping. In this case, we can show the splits between Males and Females for each region. Which is what I will implement in the makeover version.
* Though the legend is complete, it’s not very clear and aesthethic with the . in the naming.

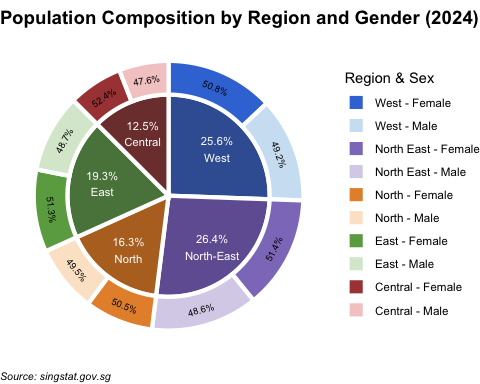
# 4 Makeover version

Next we prepare the makeover version of the data visualisation with the learning above.

|  |
| --- |
| Note |
| Although from the learning above we note that Pie Chart might not be ideal for this kind of “demographic” data with so many regions, the task is to improve and **not** change the entire idea completely but to **improve** on it as mentioned in class in session 3. |

I will be implementing changes directly using the original code and mark the lines changes with # VR CHANGE:

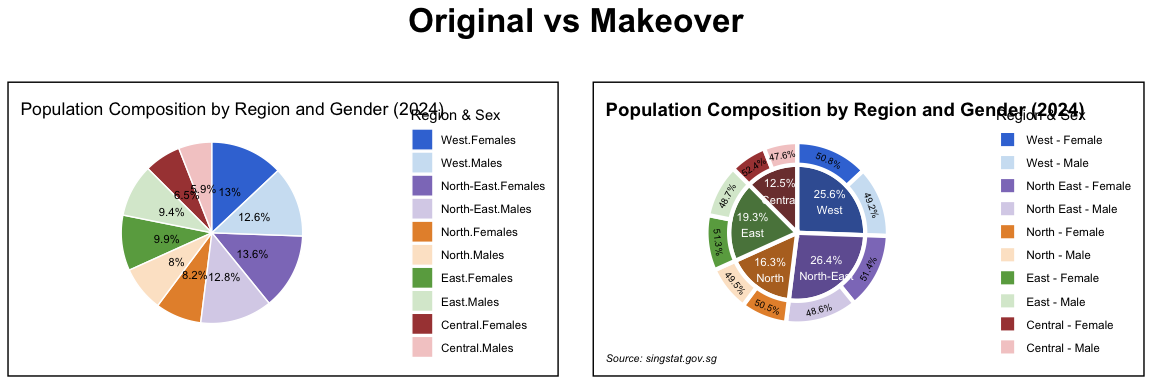
# Makeover Plot



# Code

# Summarise population by Region and Sex  
pie\_all\_regions <- resident\_regioned %>%  
 group\_by(Region, Sex) %>%  
 summarise(Population = sum(Population), .groups = "drop") %>%  
 mutate( # VR CHANGE: combine together from the second part  
 RegionSex = factor(paste(Region, Sex, sep = ".")), #, levels = ordered\_levels), VR CHANGE: not needed as we will level later after combined  
 ring = "outer", # VR CHANGE: this "ring" column is to mark which layer  
 x = 2 # VR CHANGE: this x is to mark which layer and for position later  
 )  
  
# Define the correct stacking order for pie slices # VR CHANGE: NOT NEEDED, use new level later after we combine  
#ordered\_levels <- c(  
# "Central.Males", "Central.Females",  
# "East.Males", "East.Females",  
# "North.Males", "North.Females",  
# "North-East.Males", "North-East.Females",  
# "West.Males", "West.Females"  
#)  
  
# VR CHANGE: have moved up and combine together  
# Prepare the data   
#pie\_all\_regions <- pie\_all\_regions %>%   
# mutate(  
# RegionSex = factor(paste(Region, Sex, sep = "."), levels = ordered\_levels),  
# label = paste0(round(Population / sum(Population) \* 100, 1), "%")  
# )  
  
# VR CHANGE: here we will create the "outer layer" which is the overall region  
pie\_inner <- resident\_regioned %>%  
 group\_by(Region) %>%  
 summarise(Population = sum(Population), .groups = "drop") %>%  
 mutate(  
 RegionSex = Region,  
 ring = "inner", # VR CHANGE: this "ring" column is to mark which layer  
 x = 1 # VR CHANGE: this x is to mark which layer and for position later  
 )  
  
# VR CHANGE: Combine both for plotting  
pie\_combined <- bind\_rows(pie\_all\_regions, pie\_inner)  
  
# VR CHANGE: after we combine, we need to define order of slices again  
ordered\_levels\_combined <- c(  
 "Central", "Central.Males", "Central.Females",  
 "East", "East.Males", "East.Females",  
 "North", "North.Males", "North.Females",  
 "North-East", "North-East.Males", "North-East.Females",  
 "West", "West.Males", "West.Females"  
)  
  
# VR CHANGE: Here we will label  
pie\_combined <- pie\_combined %>%  
 group\_by(ring) %>%  
 mutate(  
 percent = Population / sum(Population) \* 100, # this is the overall % for each region  
 label = paste0(round(percent, 1), "%")  
 ) %>%  
 ungroup() %>%  
 group\_by(ring, Region) %>%  
 mutate(  
 percent = Population / sum(Population) \* 100,  
 label = case\_when(  
 is.na(Sex) ~ paste0(label, "\n", Region), # keep original label if sex is NA  
 !is.na(Sex) ~ paste0(round(percent, 1), "%") # this is sex % per region otherwise show the new percentage,  
 ),  
 RegionSex = case\_when(  
 is.na(RegionSex) ~ Region, # keep original label if sex is NA  
 !is.na(RegionSex) ~ RegionSex  
 )) %>%  
 ungroup() %>%  
 mutate(RegionSex = factor(RegionSex, levels = ordered\_levels\_combined)) %>%  
 arrange(RegionSex) # make surethe order of slices are correct  
  
# VR CHANGE: after running the first time, decided to add this so we can adjust the angle of the inner layer label automatically  
pie\_combined <- pie\_combined %>%  
 group\_by(ring) %>%  
 mutate(  
 midpoint = cumsum(Population) - 0.5 \* Population,  
 total = sum(Population),  
 angle = 360 \* midpoint / total,  
 adjusted\_angle = case\_when(  
 angle <= 90 ~ angle,  
 angle > 90 & angle < 290 ~ angle + 180,  
 angle >= 290 ~ angle) # flip upside-down labels  
 ) %>%  
 ungroup()  
  
# VR CHANGE: I want to make the sunburst plot look better and easier to see. Hence maybe a poker chip (thinner outer layer) would differentiate them better  
pie\_combined <- pie\_combined %>%  
 mutate(  
 xmin = case\_when(  
 ring == "inner" ~ 0.5, # Inner ring closer to center  
 ring == "outer" ~ 1.1 # Outer ring further out  
 ),  
 xmax = case\_when(  
 ring == "inner" ~ 1.1, # Inner ring thickness = 0.6  
 ring == "outer" ~ 1.3 # Outer ring thickness = 0.2  
 )  
 ) %>%  
 group\_by(ring) %>%  
 arrange(ring, desc(RegionSex)) %>%  
 mutate(  
 ymax = cumsum(Population),  
 ymin = ymax - Population,  
 x\_text = ( xmax + xmin )/2  
 ) %>%  
 ungroup()  
# The sizing chunk end here  
  
# VR CHANGE: use the new pie combined and add more attributes   
# Plot: simple, clean pie with percentages inside  
makeover\_plot <- ggplot(pie\_combined, aes(y = Population, xmin = xmin, xmax = xmax, ymin = ymin, ymax = ymax, fill = RegionSex)) + #VR CHANGE: use pie\_combined dataset and change x = "" to x = x. Also assign to a variable name 'makeover\_plot' so we can call it again later for comparison  
 geom\_rect(color = "white", linewidth = 1.5, show.legend = TRUE) + #VR CHANGE: add linewidth so the plots are more separated and use geom\_rect so we can have a nice poker chip looking  
#VR CHANGE: we will have 2 geom\_text, the first one is for the inner ring here  
 geom\_text(  
 data = subset(pie\_combined, ring == "inner"), #VR CHANGE: filter the part of the dataset to use  
 aes(label = label,  
 x = x\_text + 0.1),   
 position = position\_stack(vjust = 0.5),  
 size = 3,  
 color = "white", #change from black to white  
 show.legend = FALSE  
 ) +  
#VR CHANGE: inner ring label end  
#VR CHANGE: this part is for outer ring  
 geom\_text(  
 data = subset(pie\_combined, ring == "outer"),  
 aes(  
 x = x\_text,  
 label = label,  
 angle = adjusted\_angle,  
 hjust = 0.5  
 ),  
 position = position\_stack(vjust = 0.5),  
 size = 2.5,  
 color = "black",  
 show.legend = FALSE  
) +  
#VR CHANGE: outer ring label end  
 coord\_polar(theta = "y") +  
 theme\_void(base\_size = 11) +  
 labs(  
 title = "Population Composition by Region and Gender (2024)",   
 fill = "Region & Sex",  
 caption = "Source: singstat.gov.sg"  
 ) +  
 scale\_fill\_manual(  
 values = c(  
 "Central.Males" = "#f4cccc",  
 "Central.Females" = "#a94442",  
 "East.Males" = "#d9ead3",  
 "East.Females" = "#6aa84f",  
 "North.Males" = "#fce5cd",  
 "North.Females" = "#e69138",  
 "North-East.Males" = "#d9d2e9",  
 "North-East.Females"= "#8e7cc3",  
 "West.Males" = "#cfe2f3",  
 "West.Females" = "#3c78d8",  
 "Central" = "#7D3D3C", # VR CHANGE: ADD COLOR FOR THE REGION  
 "East" = "#5A834A", # VR CHANGE: ADD COLOR FOR THE REGION  
 "North" = "#B67127", # VR CHANGE: ADD COLOR FOR THE REGION  
 "North-East" = "#7160A1", # VR CHANGE: ADD COLOR FOR THE REGION  
 "West" = "#3C5FA2" # VR CHANGE: ADD COLOR FOR THE REGION  
 ),  
 breaks = unique(pie\_combined$RegionSex[!is.na(pie\_combined$Sex)]), #VR CHANGE: to hide the unecessary legend like the region since we label it directly in the plot  
 labels = c( #VR CHANGE: add this to re-label the legend  
 "Central.Females" = "Central - Female",  
 "Central.Males" = "Central - Male",  
 "East.Females" = "East - Female",  
 "East.Males" = "East - Male",  
 "North.Females" = "North - Female",  
 "North.Males" = "North - Male",  
 "North-East.Females" = "North East - Female",  
 "North-East.Males" = "North East - Male",  
 "West.Females" = "West - Female",  
 "West.Males" = "West - Male"  
 )  
 ) +  
 theme( # VR CHANGE: add this to make the title bo,d  
 plot.title = element\_text(face = "bold", size = 14),  
 plot.caption = element\_text(face = "italic", size = 8, hjust = 0),  
 )   
  
makeover\_plot

# 5 Comparison



# 6 Reference

[The Four Quadrant](https://dataremixed.com/2012/05/data-visualization-clarity-or-aesthetics/)

[LIAW YING TING CELIN Take-Home Exercise 1](https://isss608celin.netlify.app/take-home_ex/take-home_ex01/take-home_ex01#pie-chart-by-regions)

[Data Visualisation: Clarity or Aesthetics](http://dataremixed.com/2012/05/data-visualization-clarity-or-aesthetics/)

[Pie-donut-chart-in-r](https://statdoe.com/pie-donut-chart-in-r/)