Distribution of Public Service of Canada Employees by Designated Group and Age Range

checking if our r is working

```
print("Hello R world")
```

```
## [1] "Hello R world"
```

Analysis of Table 6

we are loading required libraries

Loading required libraries

```
#install.packages(c("readxl", "dplyr", "ggplot2", "tidyr"))
library(readxl)
library(janitor)
```

```
##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(tidyr)
```

loading the data and cleaning the names

we are going to load the data for table 1 and display the first few rows, just to ensure that our data is loaded successfully

we also cleaned the data to use numbers only, excluding the percentages

```
library(readxl)
tab06_eng <- read_excel("~/Documents/assignments/keira/cleaned/tab06-eng.xls", skip = 4, n_max = 7)
```

```
## New names:
## * `` -> `...3`
## * `` -> `...5`
## * `` -> `...6`
## * `` -> `...8`
## * `` -> `...9`
## * `` -> `...11`
## * `` -> `...12`
## * `` -> `...14`
## * `` -> `...15`
```

```
Sys.setlocale(category = "LC_CTYPE", locale = "en_US.UTF-8")
```

```
## [1] "en_US.UTF-8"
```

```
tab06_eng <- clean_names(tab06_eng)
print(colnames(tab06_eng))
```

```
##  [1] "age_range"                           "all_employees"
##  [3] "x3"                                  "women"
##  [5] "x5"                                  "x6"
##  [7] "aboriginal_peoples"                  "x8"
##  [9] "x9"                                  "persons_with_disabilities"
## [11] "x11"                                 "x12"
## [13] "members_of_a_visible_minority_group" "x14"
## [15] "x15"
```

```
selected_colnames <- c("age_range", "all_employees", "women", "members_of_a_visible_minority_group","pe
subset_data <- tab06_eng[, selected_colnames]
subset_data <- subset_data[complete.cases(tab06_eng$age_range), ]
head(subset_data)
```

```
## # A tibble: 6 x 6
##   age_range   all_employees women members_of_a_visible_~1 persons_with_disabil~2
##   <chr>       <chr>         <chr> <chr>                  <chr>
## 1 Under 35    33830         18755 5599                   871
## 2 35 to 39    26338         14740 4221                   1009
## 3 40 to 44    27032         15030 3957                   1317
## 4 45 to 49    28931         15971 3695                   1758
## 5 50 to 54    32710         17723 3065                   2476
## 6 55 and over 32515         15859 3382                   2959
## # i abbreviated names: 1: members_of_a_visible_minority_group,
## #   2: persons_with_disabilities
## # i 1 more variable: aboriginal_peoples <chr>
```

visualization of the data

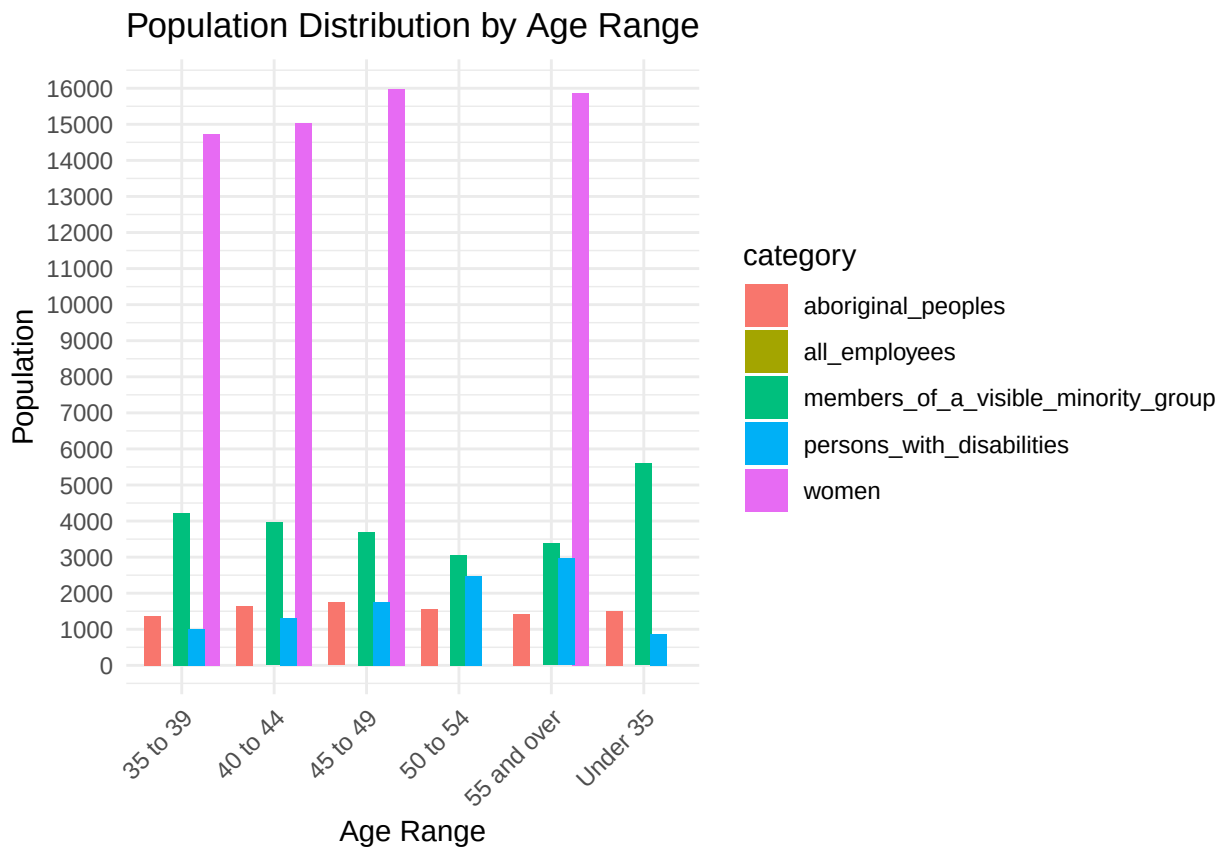
1. Drawing a bar graph showing the different distributions of employees against their age ranges

i) Converted the data to numerical data
ii) Created a bar graph

```
subset_data <- subset_data %>%
  mutate_at(vars(all_employees, women, members_of_a_visible_minority_group, persons_with_disabilities, a
            as.numeric)

subset_data_long <- subset_data %>%
  gather(key = "category", value = "value", -age_range)

ggplot(subset_data_long, aes(x = age_range, y = value, fill = category)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8)) +
  labs(title = "Population Distribution by Age Range",
       y = "Population", x = "Age Range") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_y_continuous(limits = c(0, 16000), breaks = seq(0, 16000, by = 1000))
```



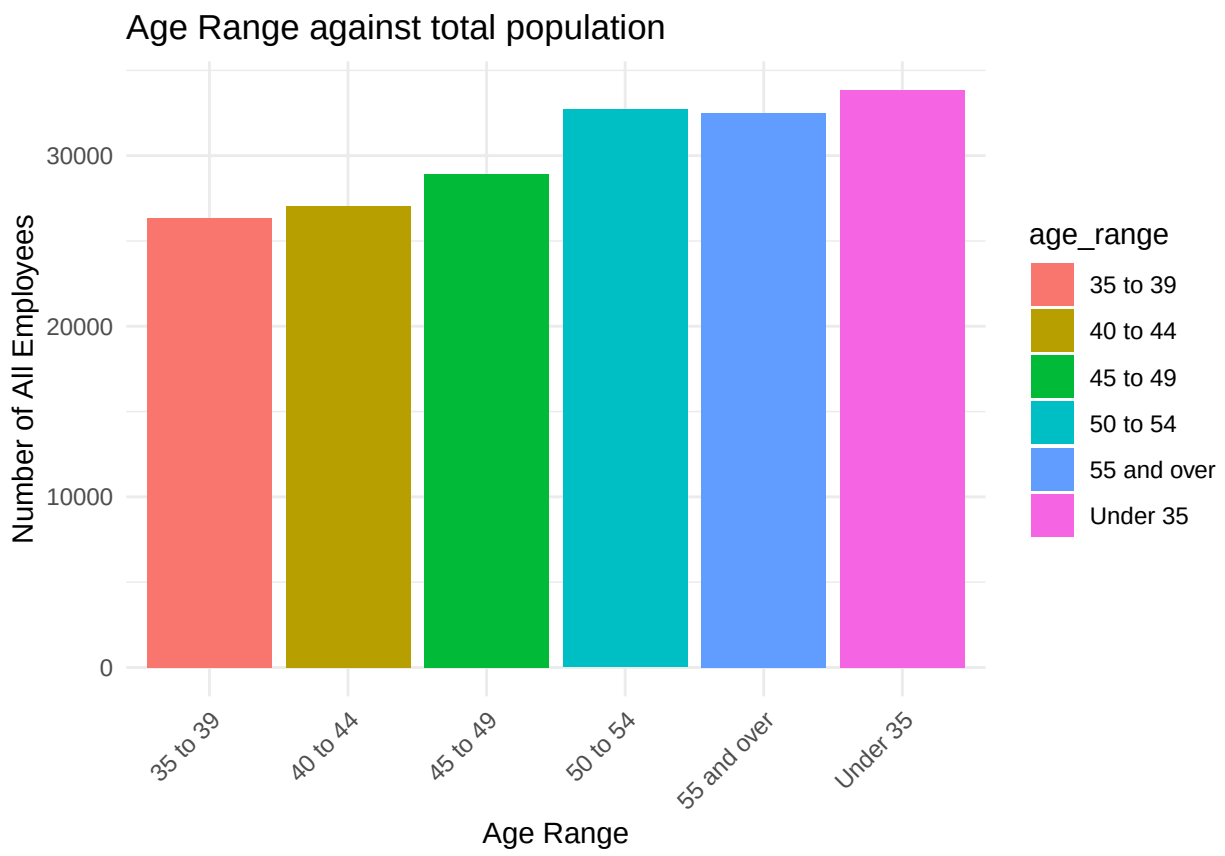
from the bar graph above you can see:

- Aboriginal groups are average across the age ranges

- women display a high number across the different ranges

1. distribution of age across populations

```
ggplot(subset_data, aes(x = age_range, y = all_employees, fill = age_range)) +
  geom_bar(stat = "identity") +
  labs(title = "Age Range against total population",
       x = "Age Range",
       y = "Number of All Employees") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

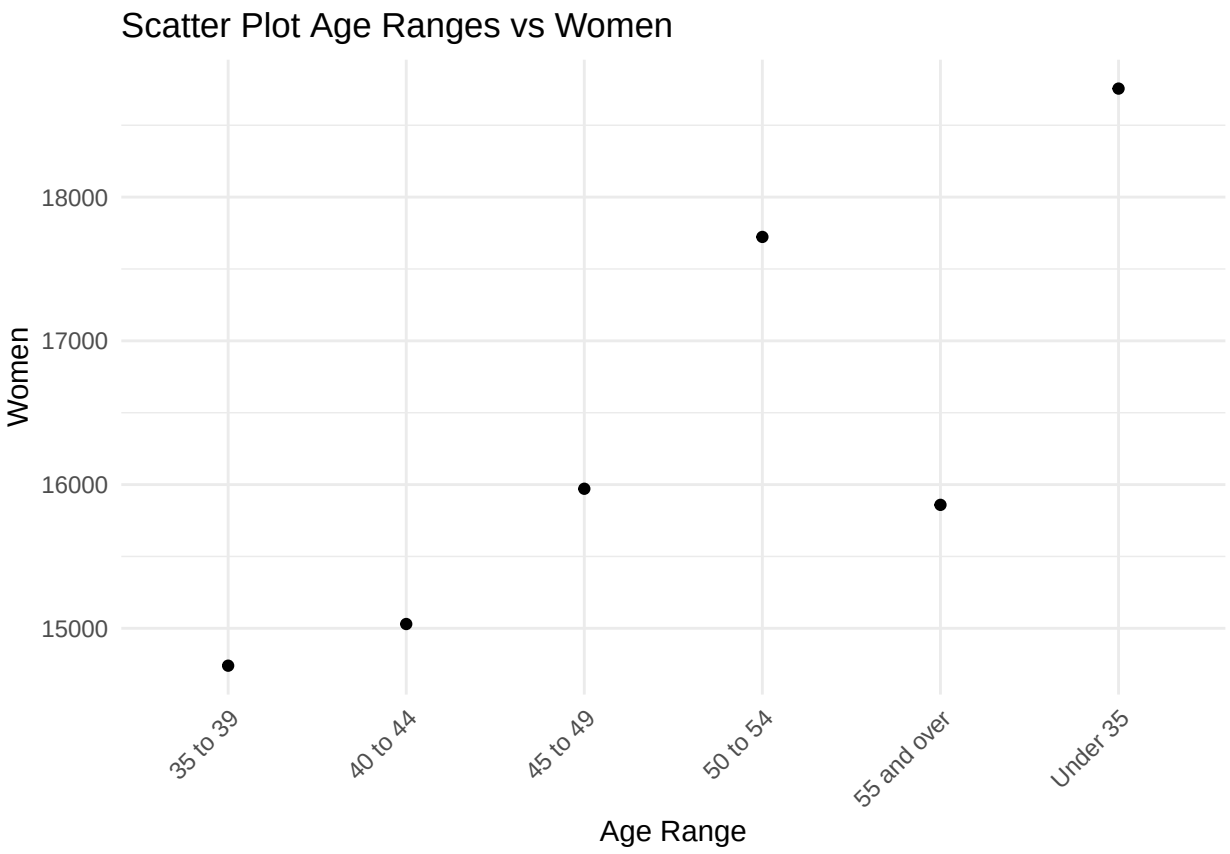


from the bar graph above you can deduce:

- A slighter working class is under the age of 35
- A set of high number of employees of the age of 50 is represented

2. Scatter plot for distribution of women across different age ranges

```
ggplot(subset_data, aes(x = age_range, y = women)) +
  geom_point() +
  labs(title = "Scatter Plot Age Ranges vs Women",
       x = "Age Range",
       y = "Women") +
  theme_minimal() + theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

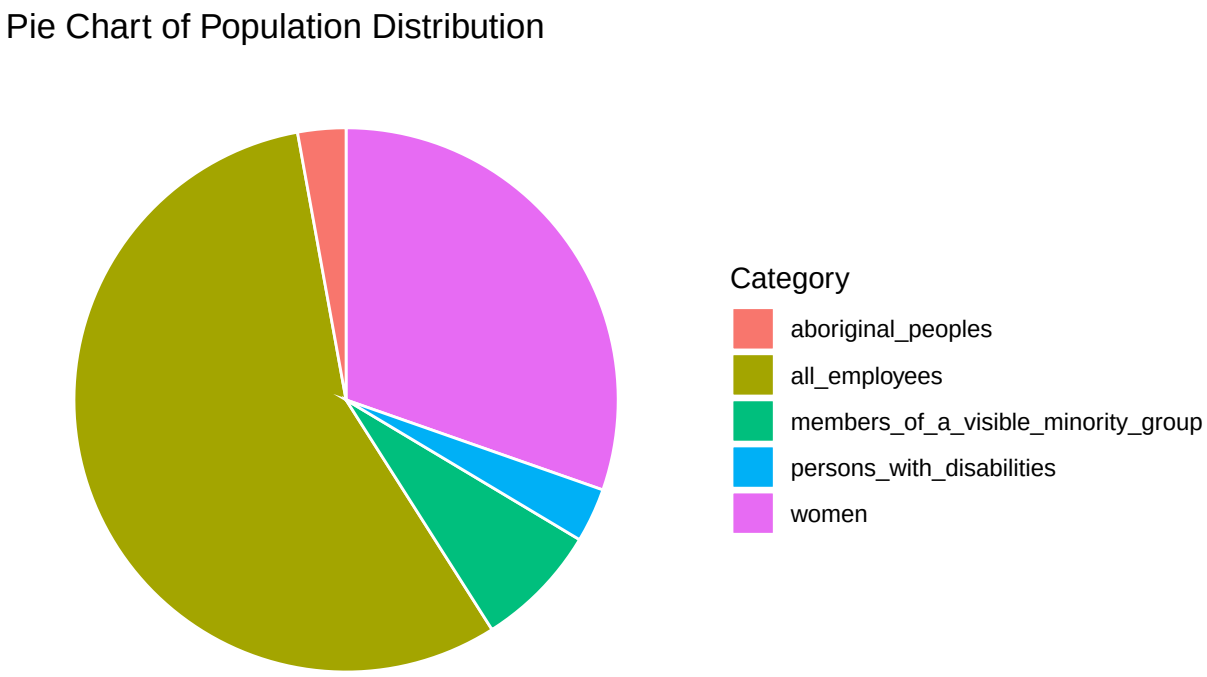


from the scatter plot above we can deduce:

- women under the age of 35 is highly represented
- there is a few representation of women between the age of 35-39

3. summary of how the employees are spread out

```
summary_data <- subset_data %>%
  summarise(
    all_employees = sum(all_employees),
    women = sum(women),
    members_of_a_visible_minority_group = sum(members_of_a_visible_minority_group),
    persons_with_disabilities = sum(persons_with_disabilities),
    aboriginal_peoples = sum(aboriginal_peoples)
  )
summary_data_long <- gather(summary_data, key = "category", value = "value")
ggplot(summary_data_long, aes(x = "", y = value, fill = category)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar("y") +
  labs(title = "Pie Chart of Population Distribution",
       fill = "Category") +
  theme_minimal() +
  theme(axis.text = element_blank(),
        axis.title = element_blank(),
        panel.grid = element_blank())
```



from the pie chart above we can deduce:

- women are the second most employed category
- Aborginal people and person with disabilities have a few representation in the job industry