Distribution of Public Service of Canada Employees by Designated Group and Occupational Category

checking if our r is working

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

Analysis of Table 2

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)
tab03_eng <- read_excel("~/Documents/programming/R/velma/keira/tab03-eng.xls", skip = 4, n_max = 10)
```

```
## New names:
## * `` -> `...4`
## * `` -> `...6`
## * `` -> `...8`
## * `` -> `...10`
```

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

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

```
tab03_eng <- clean_names(tab03_eng)
print(colnames(tab03_eng))
```

```
##  [1] "occupational_group"                 "all_employees"
##  [3] "women"                              "x4"
##  [5] "aboriginal_peoples"                 "x6"
##  [7] "persons_with_disabilities"          "x8"
##  [9] "members_of_a_visible_minority_group" "x10"
```

```
selected_colnames <- c("occupational_group", "all_employees", "women", "members_of_a_visible_minority_g
subset_data <- tab03_eng[, selected_colnames]
subset_data <- subset_data[complete.cases(tab03_eng$occupational_group), ]
head(subset_data)
```

```
## # A tibble: 6 x 6
##   occupational_group                all_employees women members_of_a_visible_~1
##   <chr>                             <chr>         <chr> <chr>
## 1 Executives                        5252          2423  447
## 2 EX (Executive)                    4955          2256  429
## 3 LC (Law Management) †             297           167   18
## 4 Scientific and Professional       31854         16357 5491
## 5 Administrative and Foreign Service 82710        52009 11638
## 6 Technical                         12593         3140  974
## # i abbreviated name: 1: members_of_a_visible_minority_group
## # i 2 more variables: persons_with_disabilities <chr>, aboriginal_peoples <chr>
```

visualization of the data

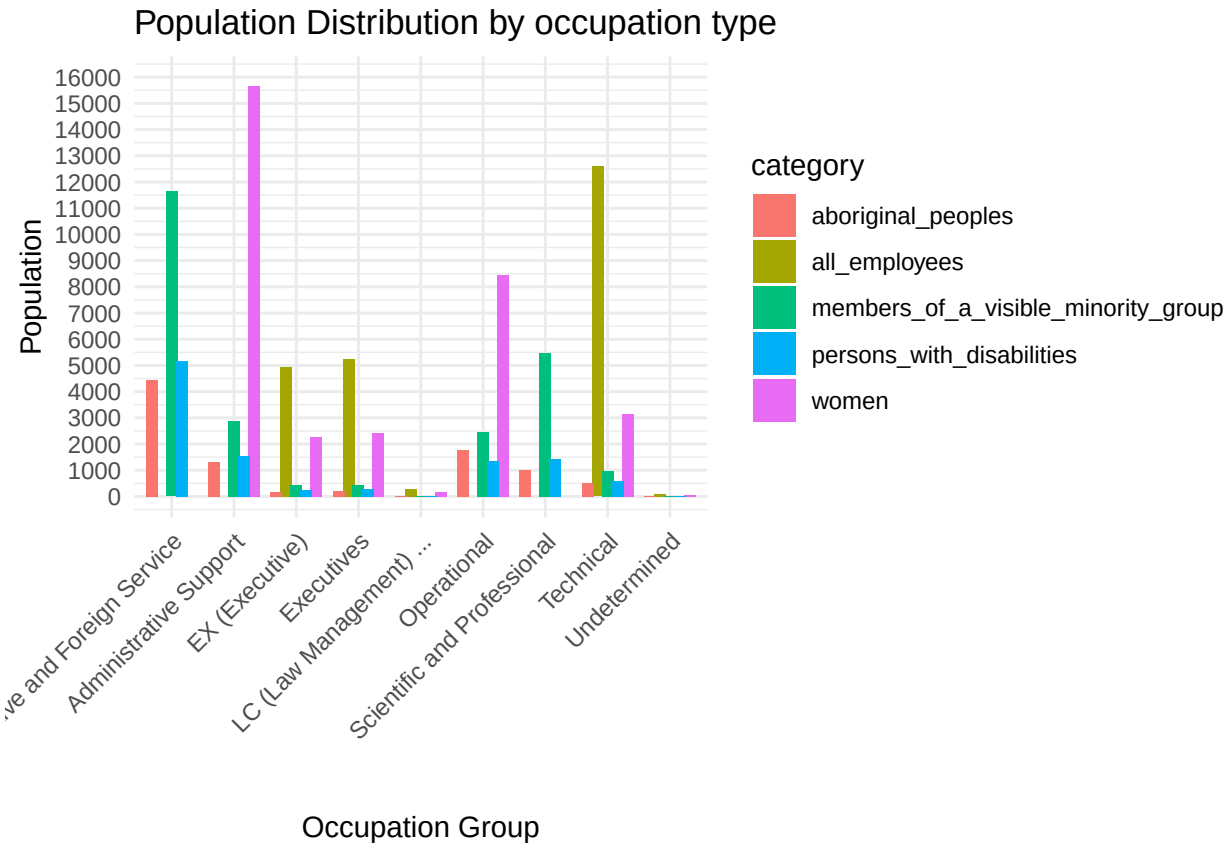
1. Drawing a bar graph showing the different distrbutions of employees categories across different job types

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", -occupational_group)

ggplot(subset_data_long, aes(x = occupational_group, y = value, fill = category)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8)) +
  labs(title = "Population Distribution by occupation type",
       y = "Population", x = "Occupation Group") +
  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))
```

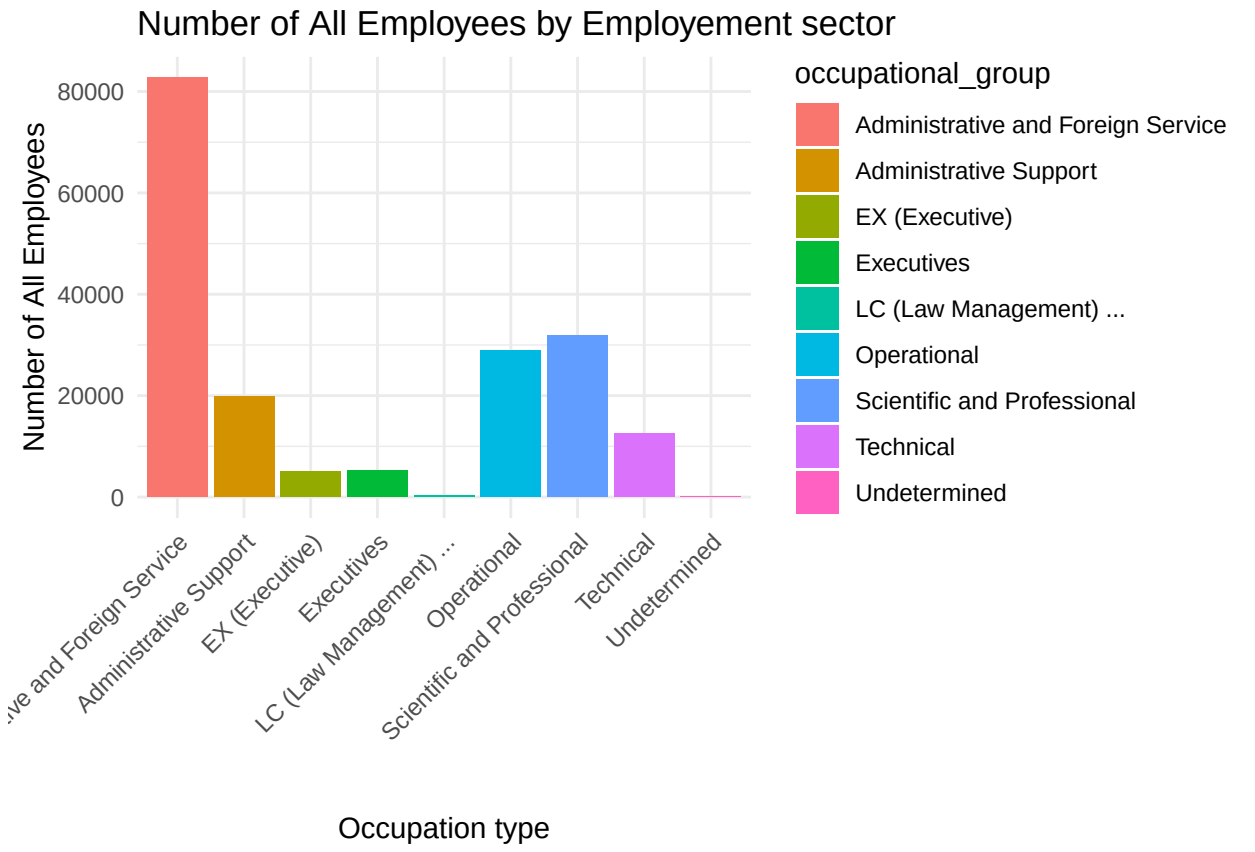


from the bar graph above you can see:

- that women are the most employed in admnistrative Support
- members of the visible minority group are represented a lot in the Administrative and Foreign Service

1. distribution of employment across job types

```r
ggplot(subset_data, aes(x = occupational_group, y = all_employees, fill = occupational_group)) +
  geom_bar(stat = "identity") +
  labs(title = "Number of All Employees by Employement sector",
       x = "Occupation type",
       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:

- alot of employees are employed in the Administrative and Foreign Service compared to other job sectors
- there are a few employees in the Law Management

2. Scatter plot for distribution of women across different job sectors

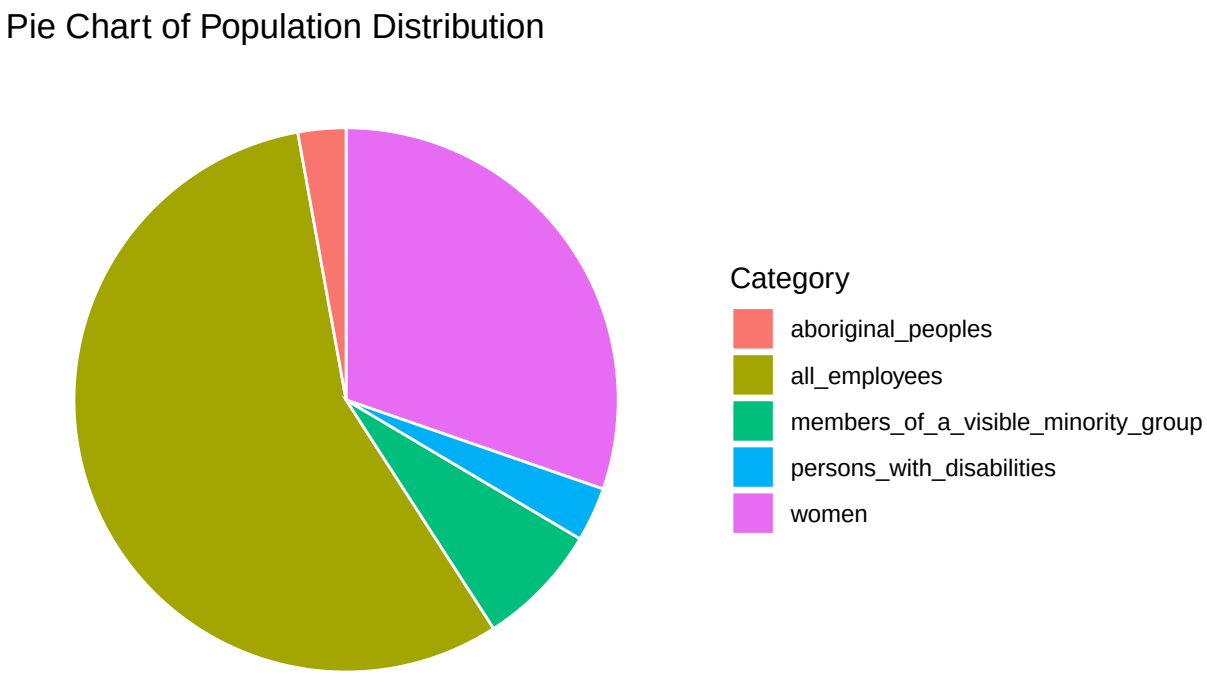
```r
ggplot(subset_data, aes(x = occupational_group, y = women)) +
  geom_point() +
  labs(title = "Scatter Plot occupational group vs Women",
       x = "Occupational Group",
       y = "Women") +
  theme_minimal() + theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

from the scatter plot above we can deduce:

- women employment is high in Adminstrative and foreign service
- there is low or zero women employed in the law management space

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