

# Hirings, Promotions Into the Public Service of Canada by Designated Group and Separations From the Public Service of Canada by Designated Group

checking if our r is working

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

## Analysis of Table 5

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)
tab05_eng <- read_excel("~/Documents/assignments/keira/cleaned/tab05-eng.xls", skip = 4, n_max = 4)
```

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

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

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

```
tab05_eng <- clean_names(tab05_eng)
print(colnames(tab05_eng))
```

```
## [1] "action_type"          "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("action_type", "all_employees", "women", "members_of_a_visible_minority_group", "persons_with_disabilities")
subset_data <- tab05_eng[, selected_colnames]
subset_data <- subset_data[complete.cases(tab05_eng$action_type), ]
head(subset_data)
```

```
## # A tibble: 3 x 6
##   action_type all_employees women members_of_a_visible~1 persons_with_disabil~2
##   <chr>      <chr>      <chr> <chr>      <chr>
## 1 Hirings    7146          3948 1140          238
## 2 Promotions 9106          4948 1261          389
## 3 Separations 15184         8502 1501          1348
## # 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

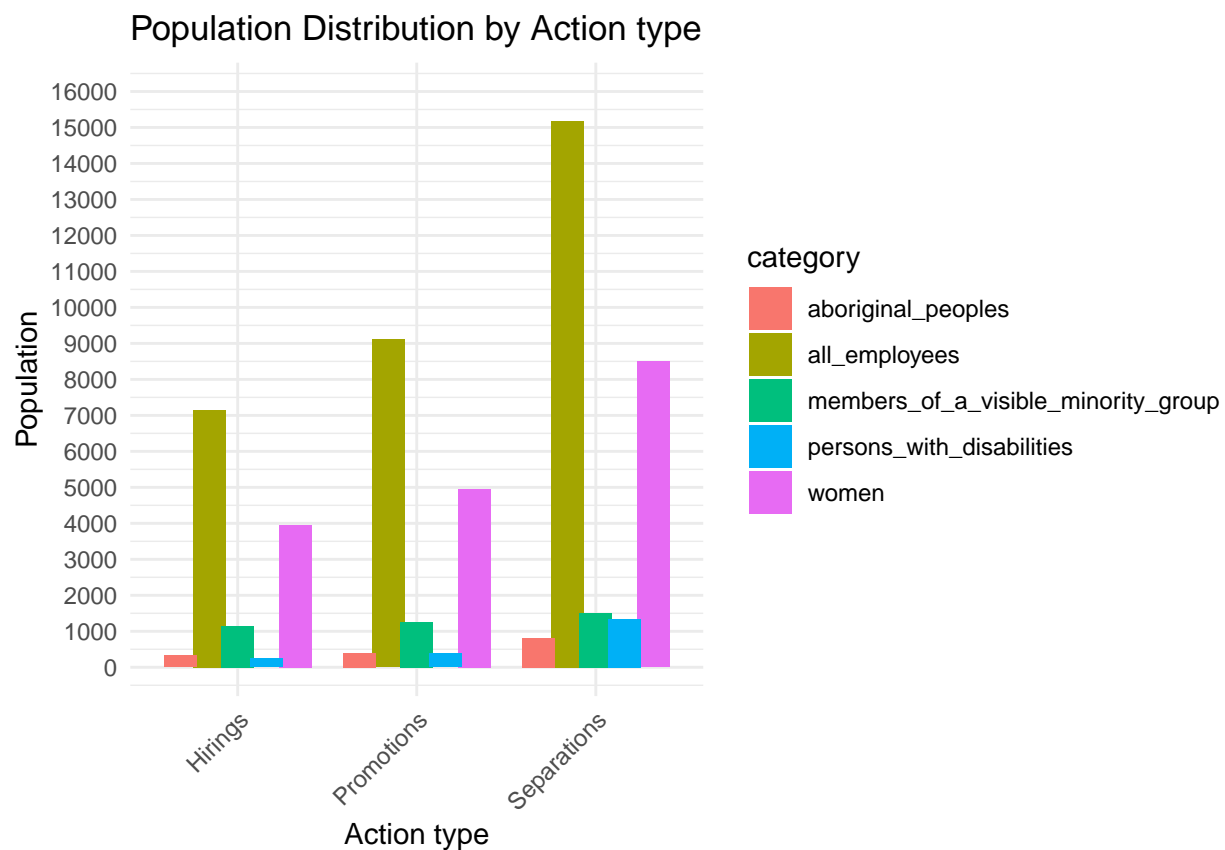
1. Drawing a bar graph showing the different distributions of employees depending on action taken

- 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,
                 as.numeric))

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

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

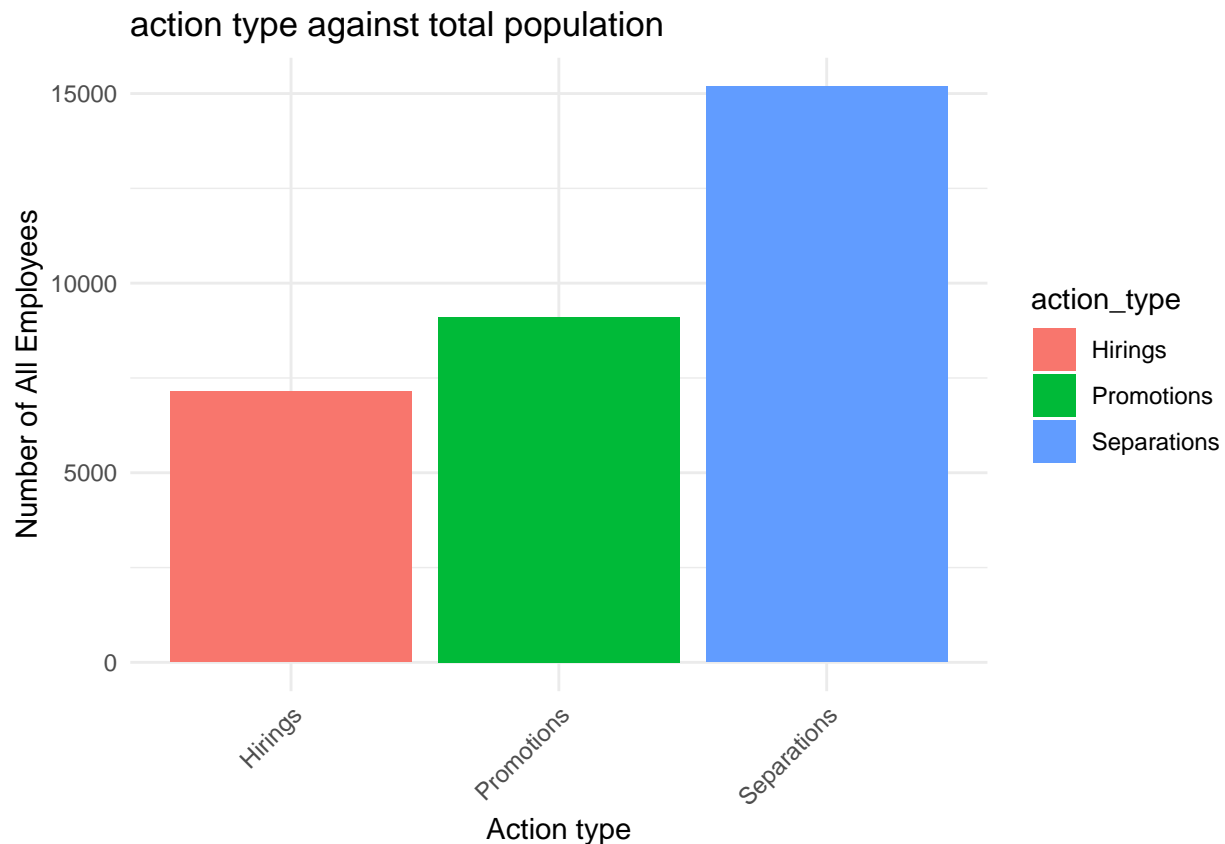


from the bar graph above you can see:

- Aboriginal groups are the least hired, least promoted and least separated
- members of the visible minority group are more likely to get promoted compared persons with disabilities

1. distribution of actions across populations

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

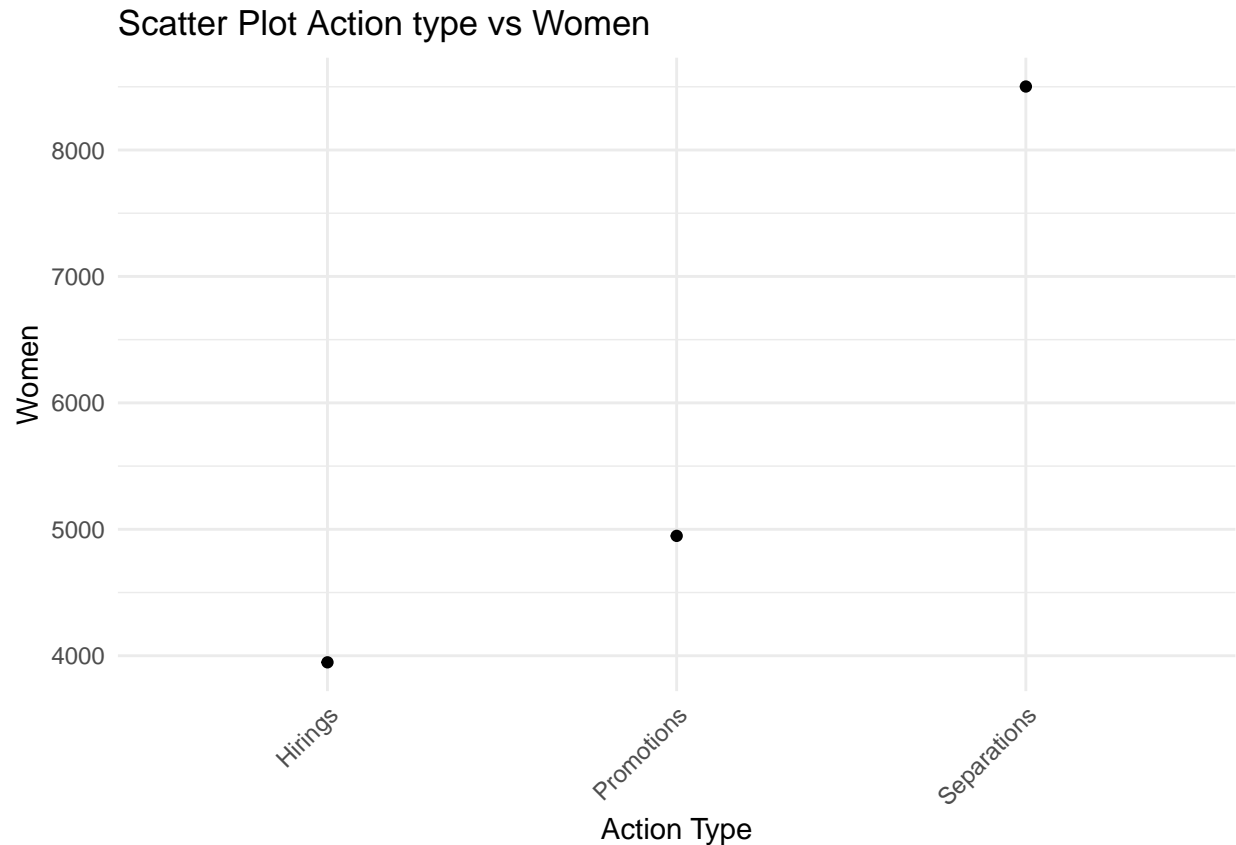


from the bar graph above you can deduce:

- A lot of employees are separated compared to Hirings and promotions
- there is a low rate of hiring

2. Scatter plot for distribution of women across different action types

```
ggplot(subset_data, aes(x = action_type, y = women)) +
  geom_point() +
  labs(title = "Scatter Plot Action type vs Women",
        x = "Action Type",
        y = "Women") +
  theme_minimal() + theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



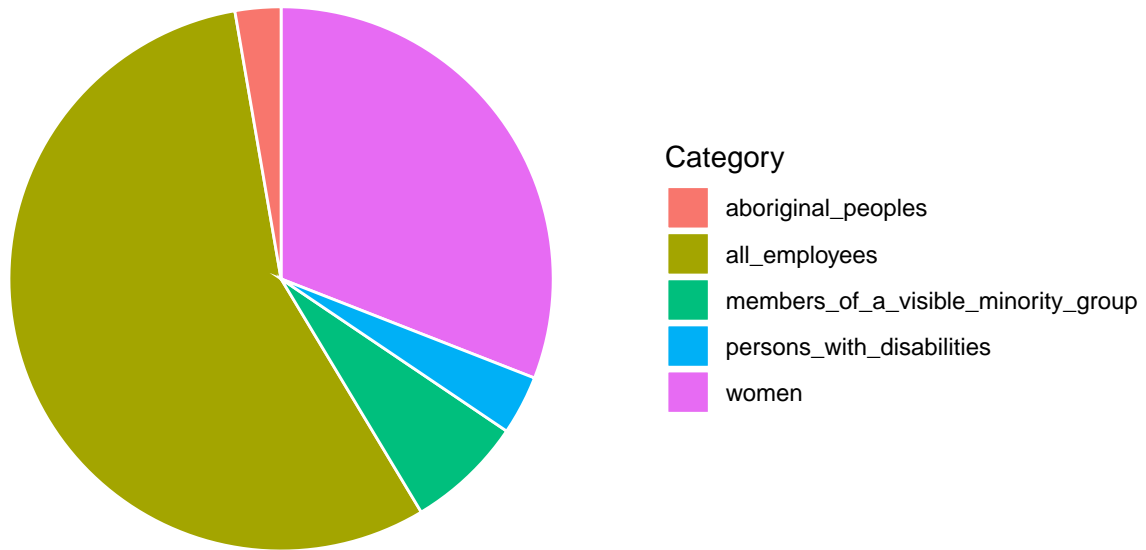
from the scatter plot above we can deduce:

- women separations is high
- there is low or zero women hired

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())
```

Pie Chart of Population Distribution



from the pie chart above we can deduce:

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