Promotions Within the Public Service of Canada by Designated Group and Occupational Category

checking if our r is working

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

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

Analysis of Table 15

we are loading required libraries

Loading required libraries

```r
#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
```

```r
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
```

```r
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

```r
library(readxl)
tab15_eng <- read_excel("~/Documents/assignments/keira/cleaned/tab15-eng.xls", skip = 4, n_max = 5)
```

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

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

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

```r
head(tab15_eng)
```

```
## # A tibble: 5 x 10
##   'Occupational Category' 'All Employees' Women ...4  'Aboriginal Peoples' ...6
##   <chr>                   <chr>           <chr> <chr> <chr>                <chr>
## 1 <NA>                    <NA>            <NA>  <NA>  <NA>                 <NA>
## 2 <NA>                    #               #     %     #                    %
## 3 Executive†              441             227   51.5  14                   3.20~
## 4 Scientific and Profess~ 2042            1101  53.8~ 65                   3.20~
## 5 Administrative and For~ 4589            2915  63.5  209                  4.59~
## # i 4 more variables: 'Persons  with Disabilities' <chr>, ...8 <chr>,
## #   'Persons in a  Visible Minority Group' <chr>, ...10 <chr>
```

```r
print(dim(tab15_eng))
```

```
## [1]  5 10
```

```r
tab15_eng <- clean_names(tab15_eng)
print(colnames(tab15_eng))
```

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

```r
selected_colnames <- c("occupational_category", "all_employees", "women", "persons_in_a_visible_minorit
head(tab15_eng)
```

```
## # A tibble: 5 x 10
##   occupational_category       all_employees women x4   aboriginal_peoples x6
##   <chr>                       <chr>         <chr> <chr> <chr>              <chr>
## 1 <NA>                        <NA>          <NA>  <NA>  <NA>               <NA>
## 2 <NA>                        #             #     %     #                  %
## 3 Executive†                  441           227   51.5  14                 3.20~
## 4 Scientific and Professional 2042          1101  53.8~ 65                 3.20~
## 5 Administrative and Foreign~ 4589          2915  63.5  209                4.59~
## # i 4 more variables: persons_with_disabilities <chr>, x8 <chr>,
## #   persons_in_a_visible_minority_group <chr>, x10 <chr>
```

```
subset_data <- tab15_eng[, selected_colnames]
subset_data <- subset_data[complete.cases(tab15_eng$occupational_category), ]
head(subset_data)
```

```
## # A tibble: 3 x 6
##   occupational_category           all_employees women persons_in_a_visible_~1
##   <chr>                           <chr>         <chr> <chr>
## 1 Executive†                      441           227   37
## 2 Scientific and Professional     2042          1101  336
## 3 Administrative and Foreign Service 4589       2915  726
## # i abbreviated name: 1: persons_in_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 distributions of employees categories across ocupational categories

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

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

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

ggplot(subset_data_long, aes(x = occupational_category, y = value, fill = category)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8)) +
  labs(title = "Type of Employment Category by Population distribution",
       y = "Population") +
  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))
```

Type of Employment Category by Population distribution
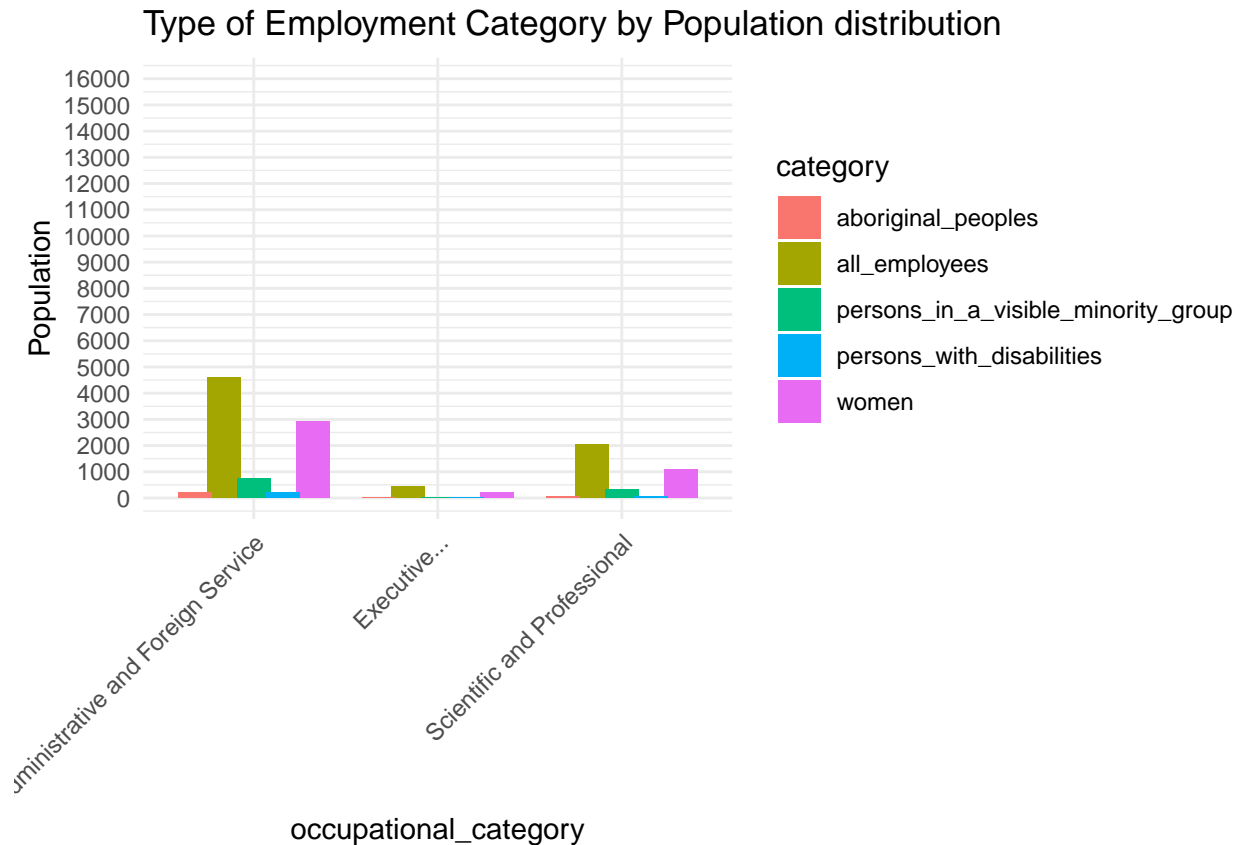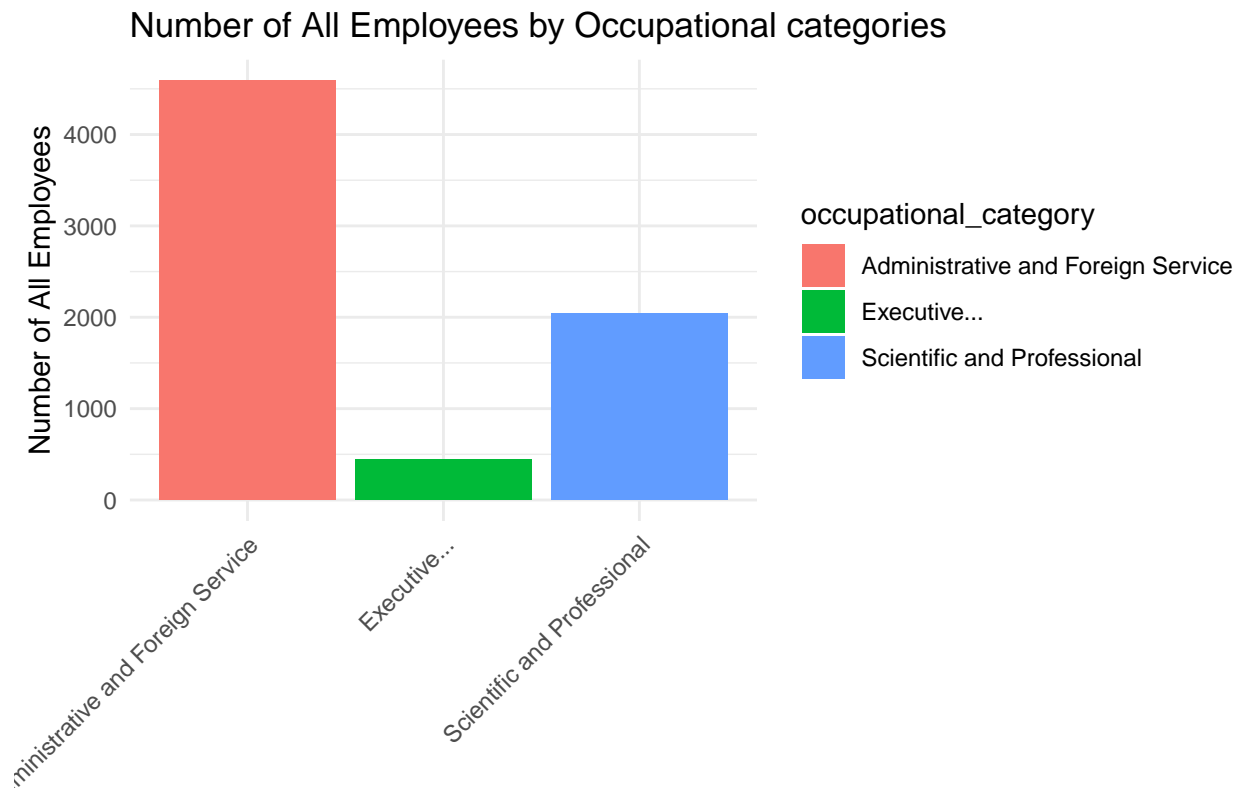

from the bar graph above you can see:

- persons with disabilities are poorly distributed across the categories

1. distribution of employment across regions

```
ggplot(subset_data, aes(x = occupational_category, y = all_employees, fill = occupational_category)) +
  geom_bar(stat = "identity") +
  labs(title = "Number of All Employees by Occupational categories",
       x = "Types of Occupational Categories",
       y = "Number of All Employees") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Number of All Employees by Occupational categories

Types of Occupational Categories

from the bar graph above you can deduce:

- there is a high rate of employment in Adminstrative and foreign services

2. Scatter plot for distribution of women across regions of work

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

Scatter Plot Occupational Category vs Women

from the scatter plot above we can deduce:

- women employment in executive position is low

3. summary of how the employees are spread out

```
summary_data <- subset_data %>%
  summarise(
    all_employees = sum(all_employees),
    women = sum(women),
    persons_in_a_visible_minority_group = sum(persons_in_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
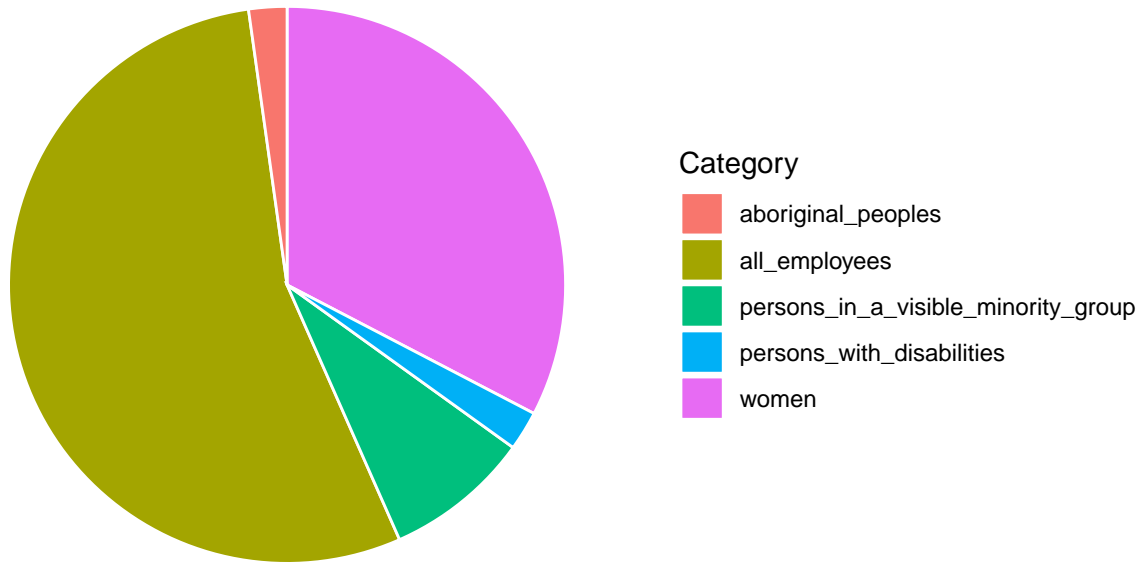
from the pie chart above we can deduce:

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