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

checking if our r is working

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

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

Analysis of Table 15

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)
tab14_eng <- read_excel("~/Documents/assignments/keira/cleaned/tab14-eng.xls", skip = 4, n_max = 19)
```

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

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

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

```
head(tab14_eng)
```

```
## # A tibble: 6 x 10
##   `Region of Work`        `All Employees` Women ...4  `Aboriginal Peoples` ...6
##   <chr>                   <chr>           <chr> <chr> <chr>                <chr>
## 1 <NA>                    <NA>            <NA>  <NA>  <NA>                 <NA>
## 2 <NA>                    #               #     %     #                    %
## 3 Newfoundland and Labra~ 98              40    40.7~ 10                   10.1~
## 4 Prince Edward Island    47              23    48.8~ 0                    0
## 5 Nova Scotia             394             143   36.2~ 12                   3
## 6 New Brunswick           438             282   64.4~ 13                   3
## # i 4 more variables: `Persons  with Disabilities` <chr>, ...8 <chr>,
## #   `Persons in a Visible Minority Group` <chr>, ...10 <chr>
```

```
print(dim(tab14_eng))
```

```
## [1] 19 10
```

```
tab14_eng <- clean_names(tab14_eng)
print(colnames(tab14_eng))
```

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

```
selected_colnames <- c("region_of_work", "all_employees", "women", "persons_in_a_visible_minority_group
head(tab14_eng)
```

```
## # A tibble: 6 x 10
##   region_of_work            all_employees women x4      aboriginal_peoples x6
##   <chr>                     <chr>         <chr> <chr>   <chr>              <chr>
## 1 <NA>                      <NA>          <NA>  <NA>    <NA>               <NA>
## 2 <NA>                      #             #     %       #                  %
## 3 Newfoundland and Labrador 98            40    40.799~ 10                 10.1~
## 4 Prince Edward Island      47            23    48.899~ 0                  0
## 5 Nova Scotia               394           143   36.299~ 12                 3
## 6 New Brunswick             438           282   64.400~ 13                 3
## # i 4 more variables: persons_with_disabilities <chr>, x8 <chr>,
## #   persons_in_a_visible_minority_group <chr>, x10 <chr>
```

```
subset_data <- tab14_eng[, selected_colnames]
subset_data <- subset_data[complete.cases(tab14_eng$region_of_work), ]
head(subset_data)
```

```
## # A tibble: 6 x 6
##   region_of_work            all_employees women persons_in_a_visible_minority_~1
##   <chr>                     <chr>         <chr> <chr>
## 1 Newfoundland and Labrador 98            40    0
## 2 Prince Edward Island      47            23    0
## 3 Nova Scotia               394           143   18
## 4 New Brunswick             438           282   16
## 5 Quebec (without the NCR†) 759           365   75
## 6 NCR† (Quebec)             1236          718   182
## # 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

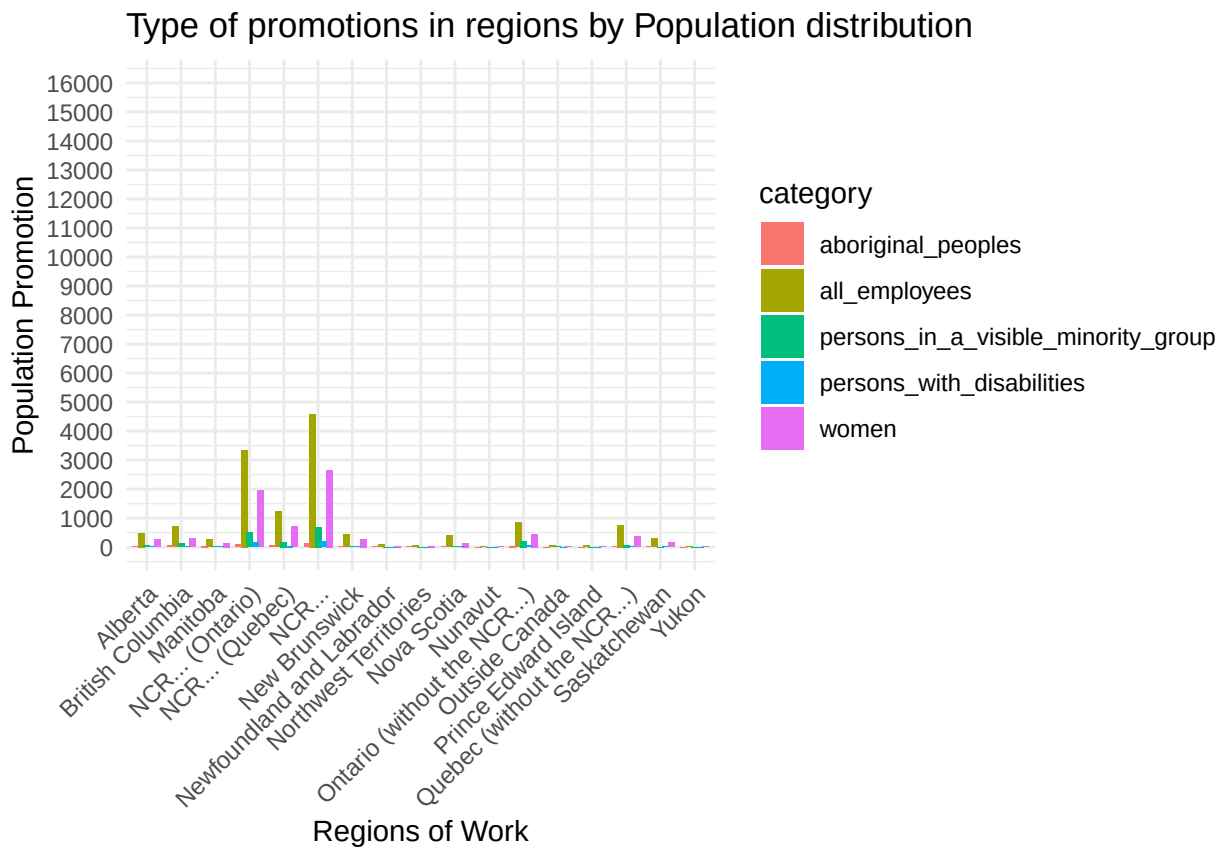
1. Drawing a bar graph showing the different distributions of employees categories across regions of work

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", -region_of_work)

ggplot(subset_data_long, aes(x = region_of_work, y = value, fill = category)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8)) +
  labs(title = "Type of promotions in regions by Population distribution",
       y = "Population Promotion", x = "Regions of Work") +
  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))
```

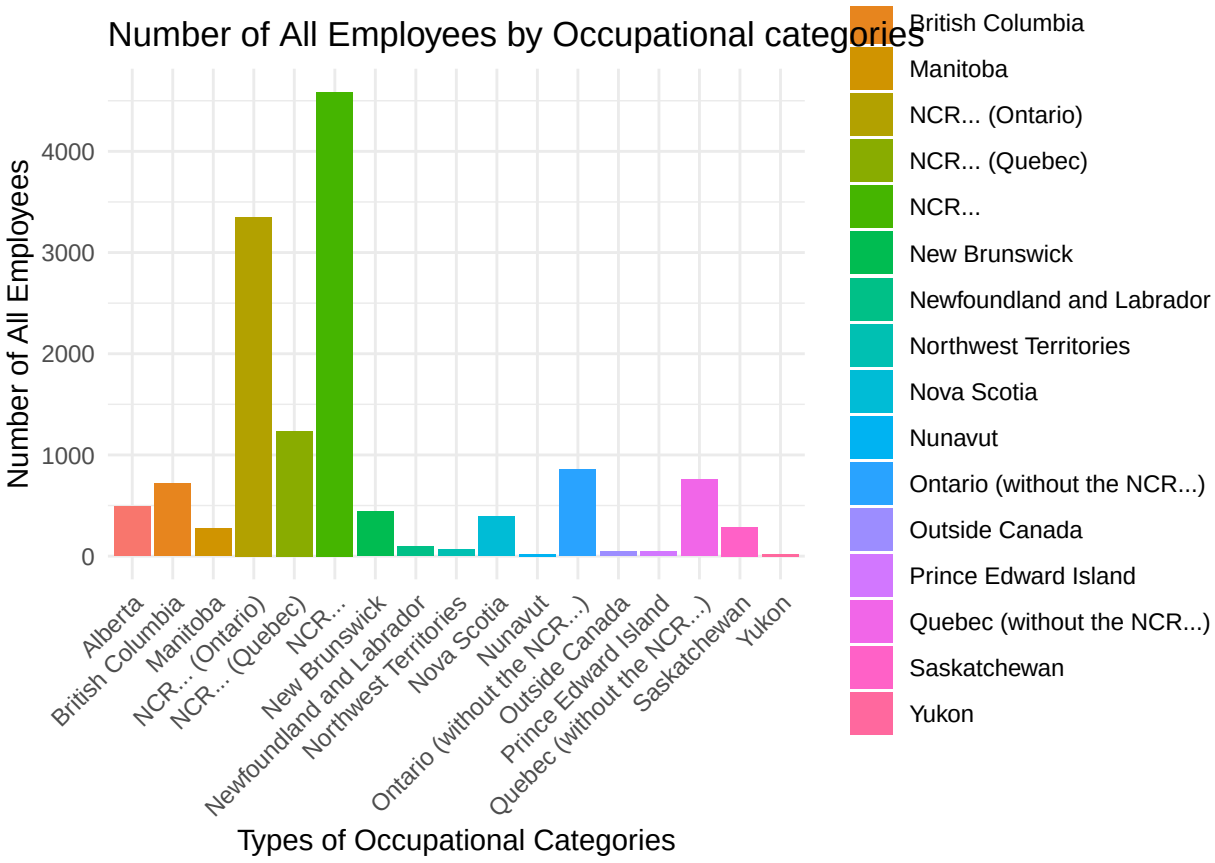


from the bar graph above you can see:

- women have an averagely distributed rate of promotion

1. distribution of employment across regions

```
ggplot(subset_data, aes(x = region_of_work, y = all_employees, fill = region_of_work)) +
  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))
```

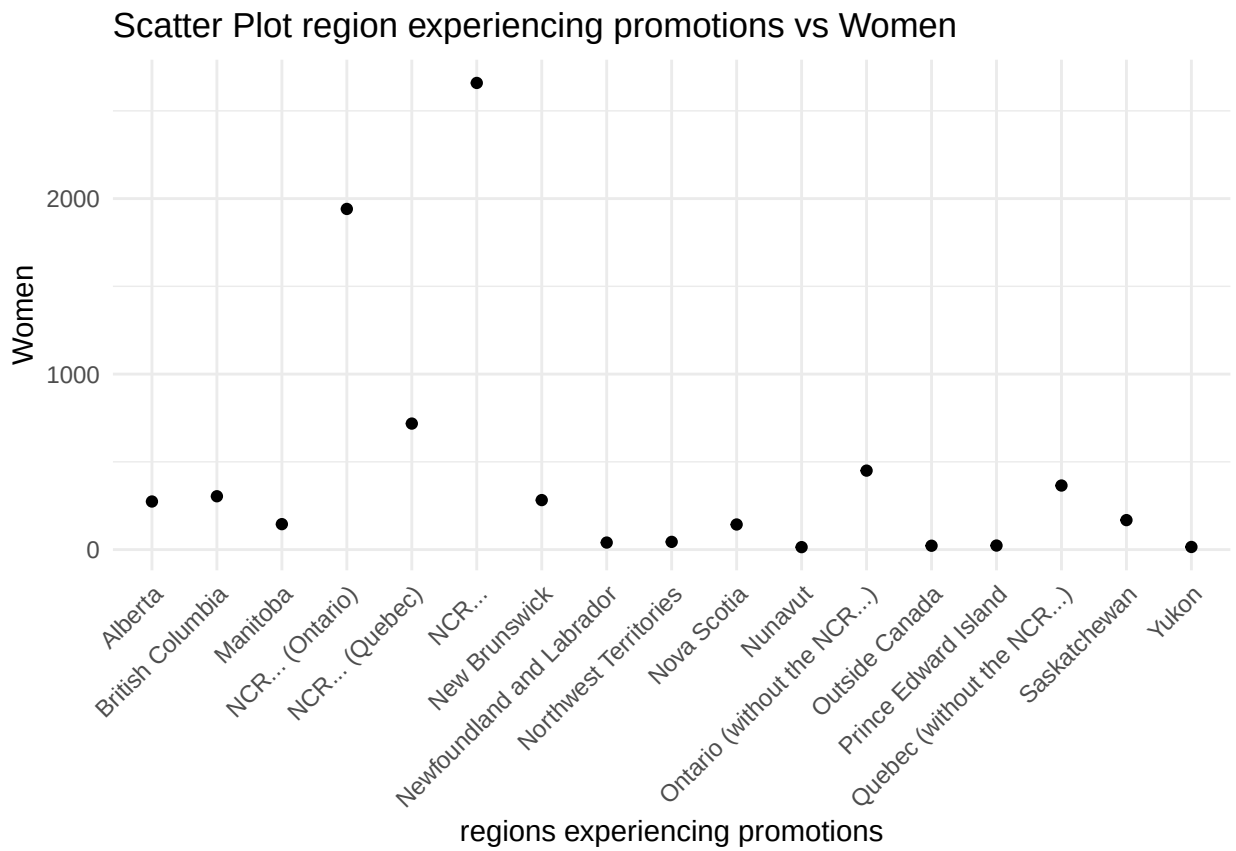


from the bar graph above you can deduce:

- NCR posts a high record of promotions to its employees
- low rate of promotions in Yukon

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

```
ggplot(subset_data, aes(x = region_of_work, y = women)) +
  geom_point() +
  labs(title = "Scatter Plot region experiencing promotions vs Women",
       x = "regions experiencing promotions",
       y = "Women") +
  theme_minimal() + theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

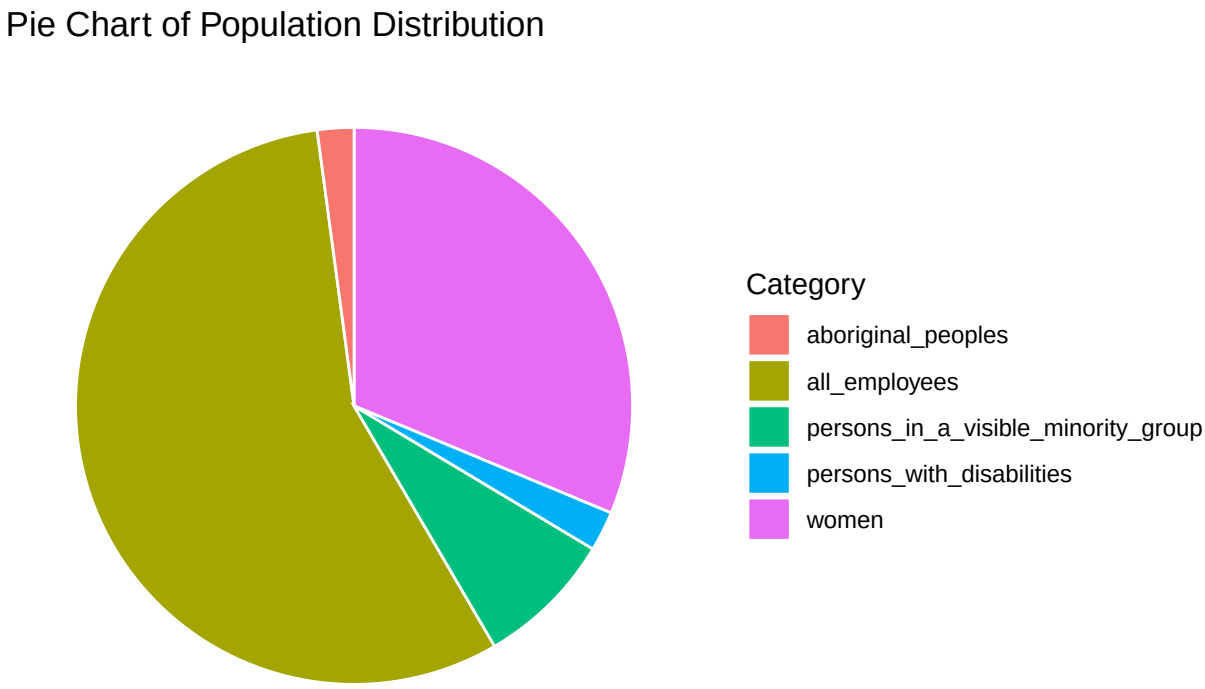


from the scatter plot above we can deduce:

- NCR still poses as a high employer for women and yukon poses the opposite - lowest employement

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())
```



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