Distribution of Public Service of Canada Employees by Designated Group and Region of Work

## 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)  
tab02\_eng <- read\_excel("~/Documents/programming/R/velma/keira/tab02-eng.xls", skip = 4, n\_max = 18)

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

head(tab02\_eng)

## # A tibble: 6 × 10  
## `Region of Work` `All Employees` Women ...4 `Aboriginal Peoples` ...6   
## <chr> <chr> <chr> <chr> <chr> <chr>  
## 1 <NA> Number Numb… % Number %   
## 2 Newfoundland and Labra… 2784 1177 42.2… 173 6.21…  
## 3 Prince Edward Island 1554 980 63.0… 44 2.83…  
## 4 Nova Scotia 8349 3567 42.7… 412 4.93…  
## 5 New Brunswick 6387 3572 55.9… 233 3.64…  
## 6 Quebec (without the NC… 20427 10669 52.2… 420 2.05…  
## # ℹ 4 more variables: `Persons with Disabilities` <chr>, ...8 <chr>,  
## # `Members of a Visible Minority Group` <chr>, ...10 <chr>

print(dim(tab02\_eng))

## [1] 18 10

tab02\_eng <- clean\_names(tab02\_eng)  
selected\_colnames <- c("region\_of\_work", "all\_employees", "women", "members\_of\_a\_visible\_minority\_group","persons\_with\_disabilities" , "aboriginal\_peoples")  
head(tab02\_eng)

## # A tibble: 6 × 10  
## region\_of\_work all\_employees women x4 aboriginal\_peoples x6   
## <chr> <chr> <chr> <chr> <chr> <chr>  
## 1 <NA> Number Number % Number %   
## 2 Newfoundland and Labrador 2784 1177 42.2… 173 6.21…  
## 3 Prince Edward Island 1554 980 63.0… 44 2.83…  
## 4 Nova Scotia 8349 3567 42.7… 412 4.93…  
## 5 New Brunswick 6387 3572 55.9… 233 3.64…  
## 6 Quebec (without the NCR) † 20427 10669 52.2… 420 2.05…  
## # ℹ 4 more variables: persons\_with\_disabilities <chr>, x8 <chr>,  
## # members\_of\_a\_visible\_minority\_group <chr>, x10 <chr>

subset\_data <- tab02\_eng[, selected\_colnames]  
#subset\_data <- filter(subset\_data, !is.na(region\_of\_work))  
subset\_data <- subset\_data[complete.cases(tab02\_eng$region\_of\_work), ]  
head(subset\_data)

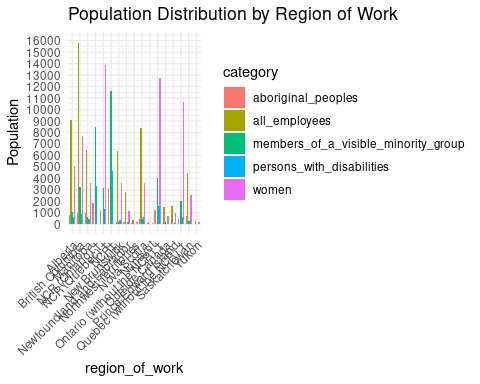
## # A tibble: 6 × 6  
## region\_of\_work all\_employees women members\_of\_a\_visible\_minority…¹  
## <chr> <chr> <chr> <chr>   
## 1 Newfoundland and Labrador 2784 1177 52   
## 2 Prince Edward Island 1554 980 39   
## 3 Nova Scotia 8349 3567 490   
## 4 New Brunswick 6387 3572 154   
## 5 Quebec (without the NCR) † 20427 10669 1999   
## 6 NCR (Quebec) † 24333 13927 3162   
## # 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 regions
2. Converted the data to numerical data
3. Created a bar graph

subset\_data <- subset\_data %>%  
 mutate\_at(vars(all\_employees, women, members\_of\_a\_visible\_minority\_group, persons\_with\_disabilities, aboriginal\_peoples),  
 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 = "Population Distribution by Region of Work",  
 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))

## Warning: Removed 7 rows containing missing values (`geom\_bar()`).

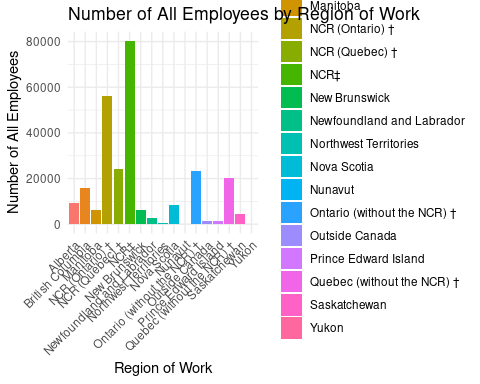


from the bar graph above you can see:

* that women are the second most employed across the different regions

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 Region of Work",  
 x = "Region of Work",  
 y = "Number of All Employees") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))

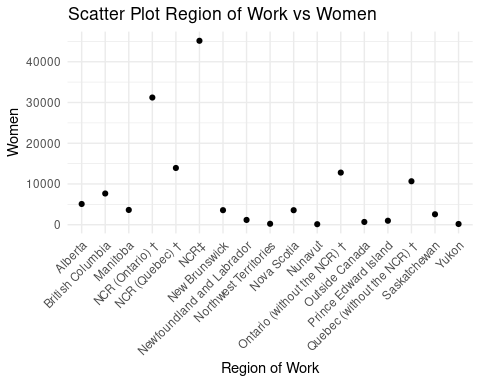


from the bar graph above you can deduce:

* there is a high rate of employment in NCR+ compared to other regions of work

1. 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 of Work vs Women",  
 x = "Region of Work",  
 y = "Women") +  
 theme\_minimal() + theme(axis.text.x = element\_text(angle = 45, hjust = 1))

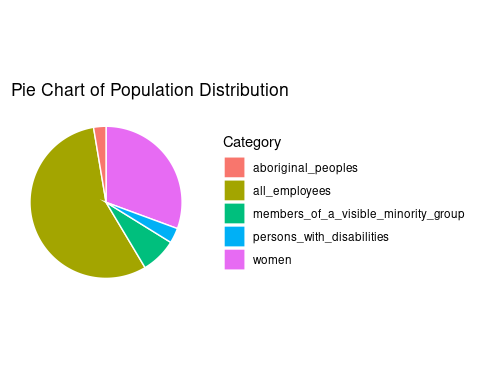


from the scatter plot above we can deduce:

* women employment is under 20,000 in various regions except for NCR‡

1. 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