Homework 2

Brayan Duran Medina

Load the necessary packages

```r
install_and_load <- function(packages) {
  new_packages <- packages[!(packages %in% installed.packages()[, "Package"])]
  if (length(new_packages)) install.packages(new_packages)
  sapply(packages, require, character.only = TRUE)
}

packages_needed <- c("dplyr", "tidyr", "readxl")

install_and_load(packages_needed)

library(dplyr)
library(tidyr)
library(readxl)
```

Load the data

```r
file_path <- "data-denmark.xlsx"
deaths_data <- read_excel(file_path, sheet = "deaths")
pop_data <- read_excel(file_path, sheet = "pop")
```

Ex 1. Using deaths data from Denmark answer the following question:

Q: Which region had the largest number of deaths among men in the year 2003?

```r
deaths_2003_men <- deaths_data %>%
  filter(year == "y2003" & sex == "m")

deaths_2003_men <- deaths_2003_men %>%
  group_by(region) %>%
  summarise(total_deaths = sum(value)) %>%
  arrange(desc(total_deaths))
```

Answer

```r
print(paste("The region with the largest number of deaths among men in 2003 was", deaths_2003_men$region[1], "with", deaths_2003_men$total_deaths[1], "deaths."))
```

```
## [1] "The region with the largest number of deaths among men in 2003 was DK01 with
16610 deaths."
```

Ex 2. Using pop data from Denmark answer the following question:

Q: In which region the sex ration (SR) is highest at ages 15, 45, over75 (coded as "open") in the year 2004?

```
pop_2004 <- pop_data %>%
  filter(year == "y2004")

pop_2004_wide <- pop_2004 %>%
  pivot_wider(names_from = sex, values_from = value)

pop_2004_wide <- pop_2004_wide %>%
  select(-b)

pop_2004_wide <- pop_2004_wide %>%
  mutate(sex_ratio = m / f * 100)

ages <- c("a15", "a45", "open")

highest_sex_ratios <- pop_2004_wide %>%
  filter(age %in% ages) %>%
  group_by(age) %>%
  summarise(region = region[which.max(sex_ratio)],
            sex_ratio = max(sex_ratio, na.rm = TRUE))
```

Answer

```
print(paste("The region with the highest sex ratio at age 15 is", highest_sex_ratios
$region[1], "with a ratio of", highest_sex_ratios$sex_ratio[1]))
```

```
## [1] "The region with the highest sex ratio at age 15 is DK01 with a ratio of 106.9
88436400201"
```

```
print(paste("The region with the highest sex ratio at age 45 is", highest_sex_ratios
$region[2], "with a ratio of", highest_sex_ratios$sex_ratio[2]))
```

```
## [1] "The region with the highest sex ratio at age 45 is DK03 with a ratio of 104.6
30897547616"
```

```
print(paste("The region with the highest sex ratio at age over 75 is", highest_sex_ra
tios$region[3], "with a ratio of", highest_sex_ratios$sex_ratio[3]))
```

```
## [1] "The region with the highest sex ratio at age over 75 is DK05 with a ratio of
65.0347819582616"
```

Ex 3. Joined dataframe. Perform the following tasks:

- Join the two dataframes (left_join OR inner_join) into a new one called df.
- Remove rows where age is equal to total.
- Change the names of columns value.x and value.y to pop and deaths.
- Convert column age to a number (Note: open = 75). Hint: substr and if_else.
- Calculate the Age Specific Death Ratios in a column called ASDR. Hint: mutate.

Now that you have df filter only the ages 15-59 and year 2001 and answer the following question:

Q: Which is the average ratio of male ASDR to female ASDR in each region?

```r
df <- left_join(deaths_data, pop_data, by = c("year", "region", "sex", "age"))

df <- df %>%
  filter(age != "total")

df <- df %>%
  rename(pop = value.y, deaths = value.x)

df <- df %>%
  mutate(age = if_else(age == "open", 75, as.numeric(substr(age, 2, nchar(age)))))

df <- df %>%
  mutate(ASDR = deaths / pop)

df <- df %>%
  select(-deaths, -pop)

df_filtered <- df %>%
  filter(age >= 15 & age <= 59 & year == "y2001")

df_ratio <- df_filtered %>%
  pivot_wider(names_from = sex, values_from = ASDR) %>%
  mutate(ASDR_ratio = m / f) %>%
  mutate(ASDR_ratio = na_if(ASDR_ratio, Inf)) %>%
  group_by(region) %>%
  summarise(average_ratio = mean(ASDR_ratio, na.rm = TRUE))
```

Answer

```r
print("Average ratio of male ASDR to female ASDR in each region (ages 15-59, year 200
1):")
```

```
## [1] "Average ratio of male ASDR to female ASDR in each region (ages 15-59, year 20
01):"
```

```r
print(df_ratio)
```

```
## # A tibble: 5 × 2
##   region average_ratio
##   <chr>          <dbl>
## 1 DK01            1.77
## 2 DK02            2.22
## 3 DK03            2.34
## 4 DK04            2.06
## 5 DK05            1.93
```

Ex 4. Joined dataframe (second part) over your dataframe called df perform the

following tasks:

- Filter only both sex (sex= b).
- Drop columns pop and deaths.
- Transform to wide format using the column "year" (pivot_wider).
- Calculate the growth of the ASDR between 2005 and 2001.

Now answer the following question:

Q: which regions experienced the largest average growth and decrease in the ASDR?

```
df <- left_join(deaths_data, pop_data, by = c("year", "region", "sex", "age"))

df <- df %>%
  filter(age != "total")

df <- df %>%
  rename(pop = value.y, deaths = value.x)

df <- df %>%
  mutate(age = if_else(age == "open", 75, as.numeric(substr(age, 2, nchar(age)))))

df <- df %>%
  filter(sex == "b")

df <- df %>%
  mutate(ASDR = deaths / pop)

df <- df %>%
  select(-deaths, -pop)

df_wide <- df %>%
  pivot_wider(names_from = year, values_from = ASDR)

df_wide <- df_wide %>%
  mutate(ASDR_growth = `y2005` - `y2001`)

average_ASDR_growth <- df_wide %>%
  group_by(region) %>%
  summarise(average_growth = mean(na_if(ASDR_growth, Inf), na.rm = TRUE))

print(average_ASDR_growth)
```

```
## # A tibble: 5 × 2
##   region average_growth
##   <chr>           <dbl>
## 1 DK01        -0.000709
## 2 DK02        -0.000678
## 3 DK03        -0.000858
## 4 DK04        -0.000828
## 5 DK05        -0.000913
```

```
largest_growth_region <- average_ASDR_growth %>%
  filter(average_growth == max(average_growth, na.rm = TRUE))

largest_decrease_region <- average_ASDR_growth %>%
  filter(average_growth == min(average_growth, na.rm = TRUE))
```

Answer

```
print(paste("Region with the largest average ASDR growth:", largest_growth_region$region))
```

```
## [1] "Region with the largest average ASDR growth: DK02"
```

```
print(paste("Region with the largest average ASDR decrease:", largest_decrease_region$region))
```

```
## [1] "Region with the largest average ASDR decrease: DK05"
```