Week 1 Lab

SOC6708 ADA

Alysia De Melo

```r
library(tidyverse)
library(here)
library(readxl)
library(janitor)
```

```r
d_male <-suppressWarnings(read_xlsx(here
    ("data/WPP2024_POP_F01_2_POPULATION_SINGLE_AGE_MALE.xlsx"), skip = 16))
d_male$sex <- "Male"
d_male <- d_male |> drop_na(Year)
d_female <- suppressWarnings(read_xlsx(here
   ("data/WPP2024_POP_F01_3_POPULATION_SINGLE_AGE_FEMALE.xlsx"), skip = 16))

d_female$sex <- "Female"
d_female <- d_female |> drop_na(Year)

d <- rbind(d_male, d_female)
rm(d_male, d_female)

d <- d  |>
  clean_names() |>
  select(region_subregion_country_or_area, iso3_alpha_code, year, x0:sex) |>
  rename(region = region_subregion_country_or_area) |>
  mutate(across(x0:x100, as.numeric))

head(d)
```

```
# A tibble: 6 x 105
  region iso3_alpha_code  year      x0      x1      x2      x3      x4      x5      x6
  <chr>  <lgl>           <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
1 World  NA               1950  41603.  36997.  34245.  32359.  30007.  28065.  27335.
```

```
2 World  NA               1951 42615. 38588. 35577. 33456. 31876. 29664. 27829.
3 World  NA               1952 43984. 39615. 37182. 34798. 32980. 31545. 29427.
4 World  NA               1953 45183. 41001. 38243. 36421. 34326. 32646. 31320.
5 World  NA               1954 45934. 42199. 39641. 37489. 35955. 33999. 32422.
6 World  NA               1955 47059. 42972. 40848. 38889. 37023. 35636. 33784.
# i 95 more variables: x7 <dbl>, x8 <dbl>, x9 <dbl>, x10 <dbl>, x11 <dbl>,
#   x12 <dbl>, x13 <dbl>, x14 <dbl>, x15 <dbl>, x16 <dbl>, x17 <dbl>,
#   x18 <dbl>, x19 <dbl>, x20 <dbl>, x21 <dbl>, x22 <dbl>, x23 <dbl>,
#   x24 <dbl>, x25 <dbl>, x26 <dbl>, x27 <dbl>, x28 <dbl>, x29 <dbl>,
#   x30 <dbl>, x31 <dbl>, x32 <dbl>, x33 <dbl>, x34 <dbl>, x35 <dbl>,
#   x36 <dbl>, x37 <dbl>, x38 <dbl>, x39 <dbl>, x40 <dbl>, x41 <dbl>,
#   x42 <dbl>, x43 <dbl>, x44 <dbl>, x45 <dbl>, x46 <dbl>, x47 <dbl>, ...
```

```r
d_male <- suppressWarnings(read_xlsx(here
 ("data/WPP2024_MORT_F01_2_DEATHS_SINGLE_AGE_MALE.xlsx"), skip = 16))
d_male$sex <- "Male"
d_male <- d_male |> drop_na(Year)
d_female <- suppressWarnings(read_xlsx(here
 ("data/WPP2024_MORT_F01_3_DEATHS_SINGLE_AGE_FEMALE.xlsx"), skip = 16))
d_female$sex <- "Female"
d_female <- d_female |> drop_na(Year)

dm <- rbind(d_male, d_female)
rm(d_male, d_female)

dm <- dm  |>
  clean_names() |>
  select(region_subregion_country_or_area, iso3_alpha_code, year, x0:sex) |>
  rename(region = region_subregion_country_or_area) |>
  mutate(across(x0:x100, as.numeric))
```

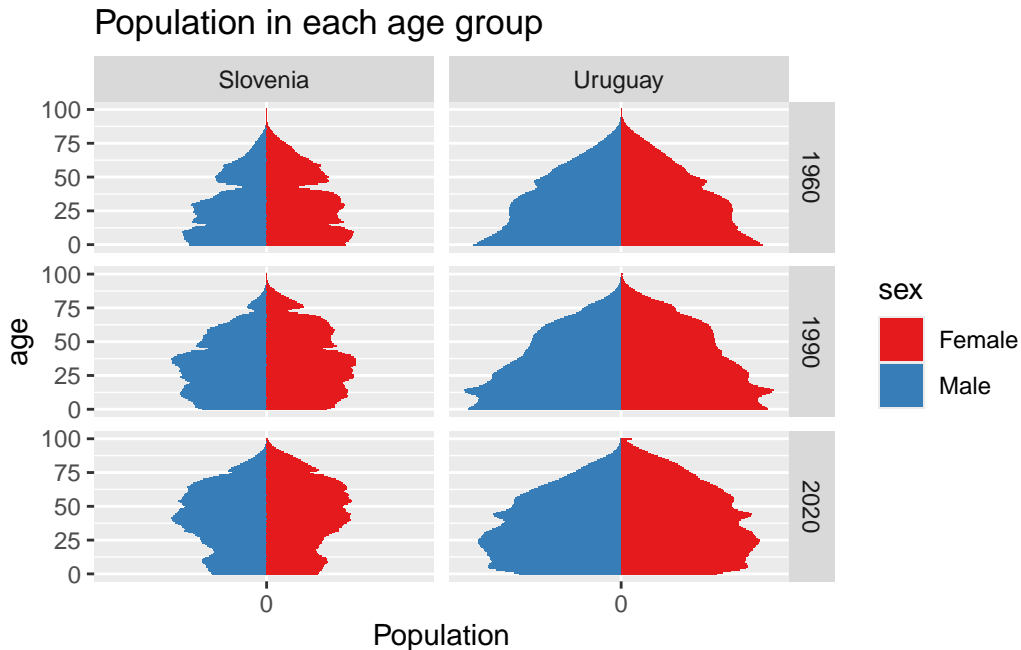1. Pick two countries. Plot their population pyramids in 1960, 1990, and 2020.

```r
d_long <- d  |>
  pivot_longer(x0:x100, names_to = "age", values_to = "pop") |>
  mutate(age = as.numeric(str_remove(age, "x")))

d_long|>
  filter(region=="Uruguay"|region=="Slovenia",
         year%in% c(1960, 1990, 2020))  |>
  mutate(population=ifelse(sex=="Male", -pop, pop)) %>%
ggplot(aes(x = age, y = population, fill = sex)) +
```

```
    facet_grid(year~region)+
    geom_bar(stat="identity")+
    ggtitle("Population in each age group")+
    ylab("Population")+
    coord_flip() +
    scale_y_continuous(breaks = seq(-4000, 4000, 1000),
                       labels = c(seq(4000, 0, -1000), seq(1000, 4000, 1000))) +
    scale_fill_brewer(palette = "Set1")
```


2. Based on the shape and change in population pyramids in 1, do you think the population growth rate in each of your chosen countries is positive or negative in recent years?
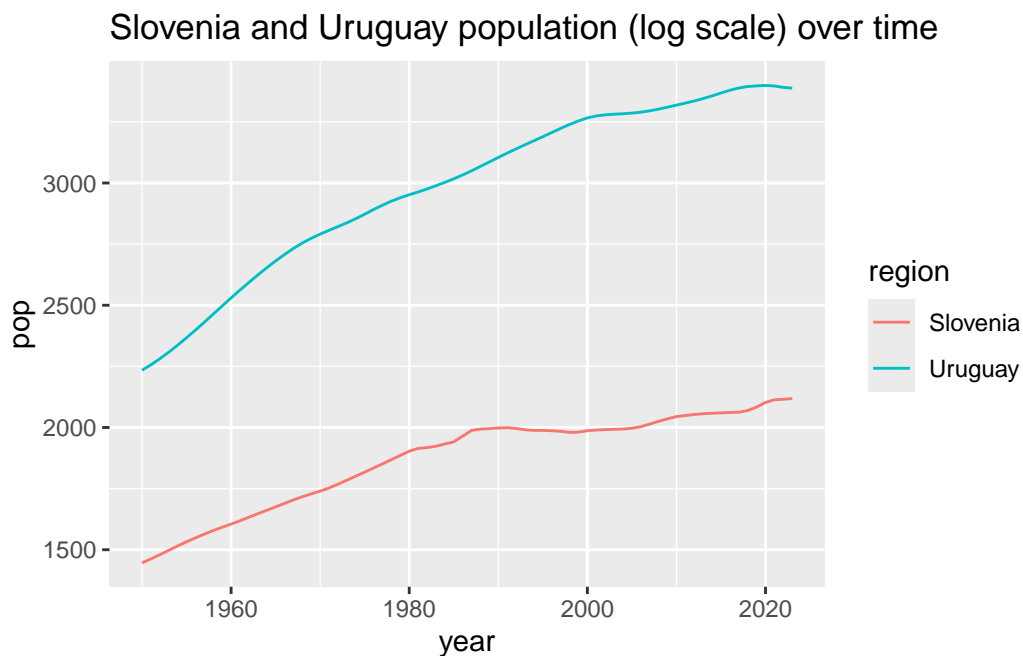
Based on the shape and changes in the population pyramids, population growth in Slovenia appears to be positive but very low in recent years. The 2020 pyramid has a narrow base, a thicker middle, and a narrow top. This suggests that while population growth has slowed, population from the middle cohorts seem to keep the growth slightly positive. In contrast, the 1990 pyramid shows a relatively wider base, and shows that there was stronger population growth in earlier decades. If this earlier trend had continued, population growth would likely be higher than in 2020.

For Uruguay, population growth also appears to be positive in recent years. The 2020 population pyramid shows a wider base that gradually narrows with age. Although the 1990 pyramid

has a more triangular shape that is typical of faster population growth, the overall structure in 2020 still suggests positive growth.

3. Confirm your guesses in 2 by plotting population over time in both countries.

```
d_long |>
  filter(region=="Slovenia"|region=="Uruguay") |>
  group_by(region, year) |>
  summarize(pop = sum(pop)) |>
  ggplot(aes(year, pop, color = region))+
  geom_line() +
   ggtitle("Slovenia and Uruguay population (log scale) over time")
```

Slovenia and Uruguay population (log scale) over time


```
d_long |>
  filter(region=="Slovenia") |>
  group_by(year) |>
  summarize(pop = sum(pop)) |>
  mutate(log_pop = log(pop))|>
  summarise(
  growth_rate = (log_pop[year==2020] - log_pop[year==1960])/(2020-1960))
```

```
# A tibble: 1 x 1
```

```
     growth_rate
            <dbl>
1      0.00450
```

```
d_long |>
  filter(region=="Uruguay") |>
  group_by(year) |>
  summarize(pop = sum(pop)) |>
  mutate(log_pop = log(pop))|>
  summarise(
  growth_rate = (log_pop[year==2020] - log_pop[year==1960])/(2020-1960))
```

```
# A tibble: 1 x 1
  growth_rate
        <dbl>
1      0.00492
```

My initial guess was correct, as Slovenia's population shows an overall positive trend over time. Population growth is positive and relatively steady until approximately 1980, after which growth slows and includes decline period, though the total population continues to increase through 2020. This pattern is consistent with the population pyramid. When calculated, the population growth rate is positive but small at 0.45% per year

My initial guess for Uruguay was also correct, as the population shows an overall positive trend over time. Population growth appears fairly steady, with no substantial periods of decline, aside from a slight dip in the early 2000s but it is much less pronounced than the declines observed in Slovenia after 1980. This pattern is consistent with the population pyramid, which suggests continued but slowing population growth. The population growth rate is 0.049%. Based on the population pyramids, I initially expected Uruguay to have substantially higher population growth than Slovenia, however, when calculating the growth, Uruguay has only a modestly larger growth (difference of 0.004)

4. Taking the US population in 2000 as the standard population, calculate the standardized mortality rates for US in 2023 and Australia in 2023. How much higher was the death rate in US compared to Australia?

```
pops <- d_long

dm_long <- dm  |>
  pivot_longer(x0:x100, names_to = "age", values_to = "deaths") |>
  mutate(age = as.numeric(str_remove(age, "x")))
```

```r
asmr <- d_long |>
  left_join(dm_long) |>
  mutate(mx = deaths/pop)

US_2000 <- asmr |>
  filter(region == "United States of America", year == 2000) |>
  rename(std_pop = pop) |>
  select(std_pop, sex, age)


AUS_2023 <- asmr |>
  filter(year==2023, region=="Australia")  |>
  rename(AUSmx = mx)   |>
  select(AUSmx, sex, age)



US_2023 <- asmr |>
  filter(year==2023, region=="United States of America")   |>
  rename(USmx = mx) |>
  select(USmx, sex, age)


#used left_join to ensure age and sex match
USAUS_2023 <- left_join(AUS_2023, US_2023,  by = c("age", "sex"))

USAUS_2000_2023 <- left_join(USAUS_2023, US_2000,  by = c("age", "sex"))


std_numbers <- USAUS_2000_2023  |>
  mutate(std_deaths_Australia_2023 = std_pop*AUSmx,
         std_deaths_US_2023 = std_pop*USmx) |>
  summarise(std_rate_AUS23 = sum(std_deaths_Australia_2023)/sum(std_pop),
std_rate_US23 = sum(std_deaths_US_2023)/sum(std_pop))

std_numbers
```

```
# A tibble: 1 x 2
  std_rate_AUS23 std_rate_US23
          <dbl>         <dbl>
1       0.00487       0.00673
```

```
ratio <- std_numbers  |>
  summarize(ratio = std_rate_US23/std_rate_AUS23)
ratio
```

```
# A tibble: 1 x 1
  ratio
  <dbl>
1  1.38
```

Using the US population in 2000 as the standard, the standardized mortality rate in 2023 was approximately 0.67% in the United States and 0.49% in Australia. The US death rate was therefore about 1.38 greater than the death rate in Australia.