

Lab 4: Fertility

SOC6708 ADA

Alysia De Melo

```
library(tidyverse)
library(here)
library(readxl)
library(janitor)
df <- read_csv("data/WPP2024_Fertility_by_Age5.csv")
dl <- read_csv("data/WPP2024_Life_Table_Abridged_Medium_1950-2023.csv")
dp <- read_csv("data/wpp_pop_clean.csv")
```

Exercise 1

Choose three countries in varying demographic stages and calculate and plot the NRR over time. Also plot population pyramids for these same countries (one recent, one from 1970) and discuss how population growth/decline and NRR are related.

```
fx_mex <- df |>
  filter(Location == "Mexico", Variant == "Medium") |>
  select(Location, Time, AgeGrp, ASFR) |>
  mutate(ASFR = ASFR/1000)

lx_mex <- dl |>
  filter(Location == "Mexico", Sex == "Female") |>
  select(Location, Time, AgeGrp, Lx) |>
  mutate(Lx = Lx / 100000)

ffab <- 0.4886

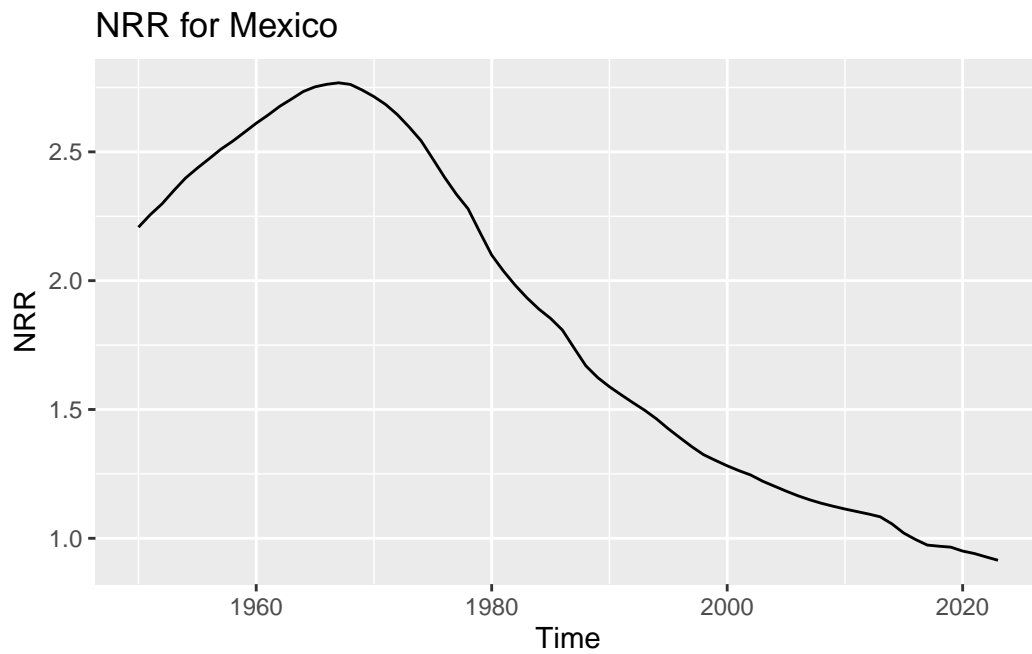
nrrs_mex <- fx_mex |>
  left_join(lx_mex) |>
  mutate(prod = ASFR*Lx) |>
  group_by(Time) |>
```

```

summarize(NRR = sum(prod)*ffab) |>
drop_na()

nrrs_mex |>
  ggplot(aes(Time, NRR)) +
  geom_line()+
  labs(title = "NRR for Mexico")

```



```

fx_chi <- df |>
  filter(Location == "China", Variant == "Medium") |>
  select(Location, Time, AgeGrp, ASFR) |>
  mutate(ASFR = ASFR/1000)

lx_chi <- dl |>
  filter(Location == "China", Sex == "Female") |>
  select(Location, Time, AgeGrp, Lx) |>
  mutate(Lx = Lx / 100000)

ffab <- 0.4886

nrrs_chi <- fx_chi |>
  left_join(lx_chi) |>

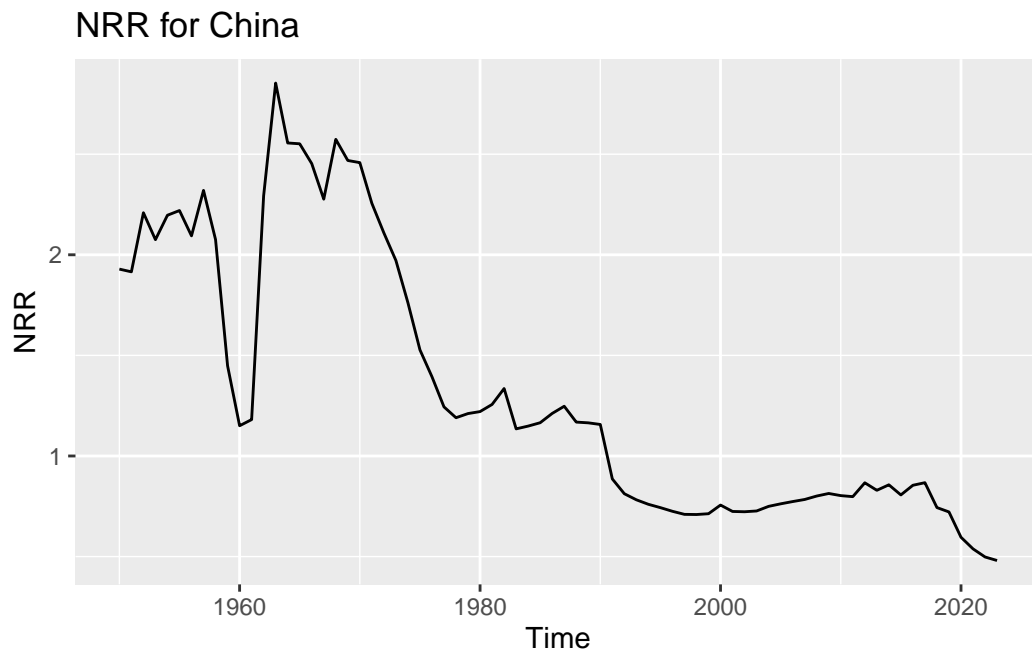
```

```

mutate(prod = ASFR*Lx) |>
group_by(Time) |>
summarize(NRR = sum(prod)*ffab) |>
drop_na()

nrrs_chi |>
  ggplot(aes(Time, NRR)) +
  geom_line()+
  labs(title = "NRR for China")

```



```

fx_sud <- df |>
  filter(Location == "Sudan", Variant == "Medium") |>
  select(Location, Time, AgeGrp, ASFR) |>
  mutate(ASFR = ASFR/1000)

lx_sud <- dl |>
  filter(Location == "Sudan", Sex == "Female") |>
  select(Location, Time, AgeGrp, Lx) |>
  mutate(Lx = Lx / 100000)

ffab <- 0.4886

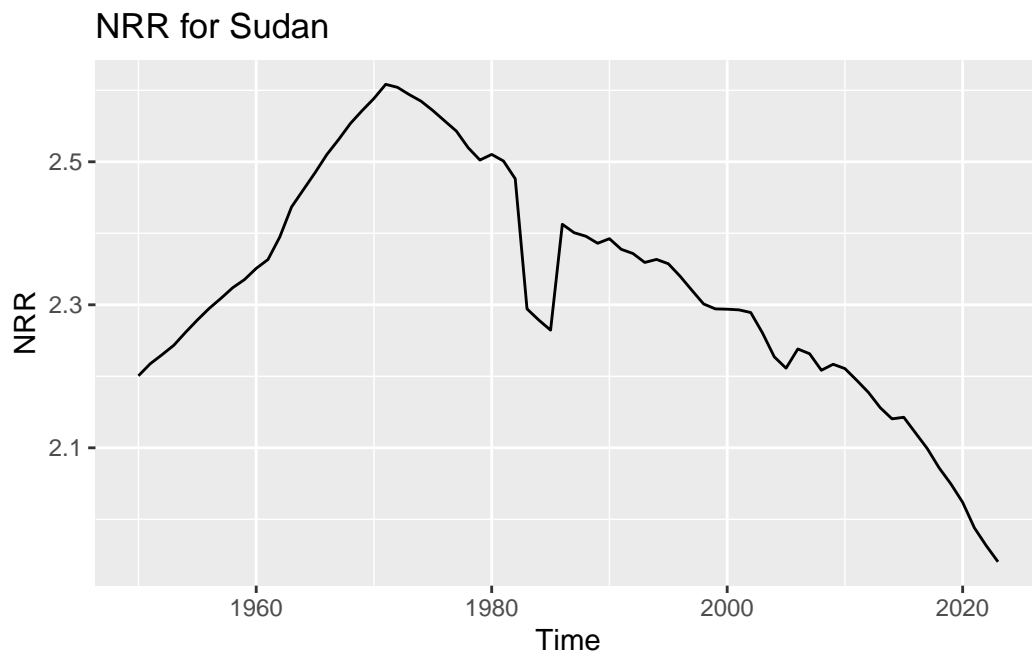
```

```

nrrs_sud <- fx_sud |>
  left_join(lx_sud) |>
  mutate(prod = ASFR*Lx) |>
  group_by(Time) |>
  summarize(NRR = sum(prod)*ffab) |>
  drop_na()

nrrs_sud |>
  ggplot(aes(Time, NRR)) +
  geom_line()+
  labs(title = "NRR for Sudan")

```



```

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))

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)

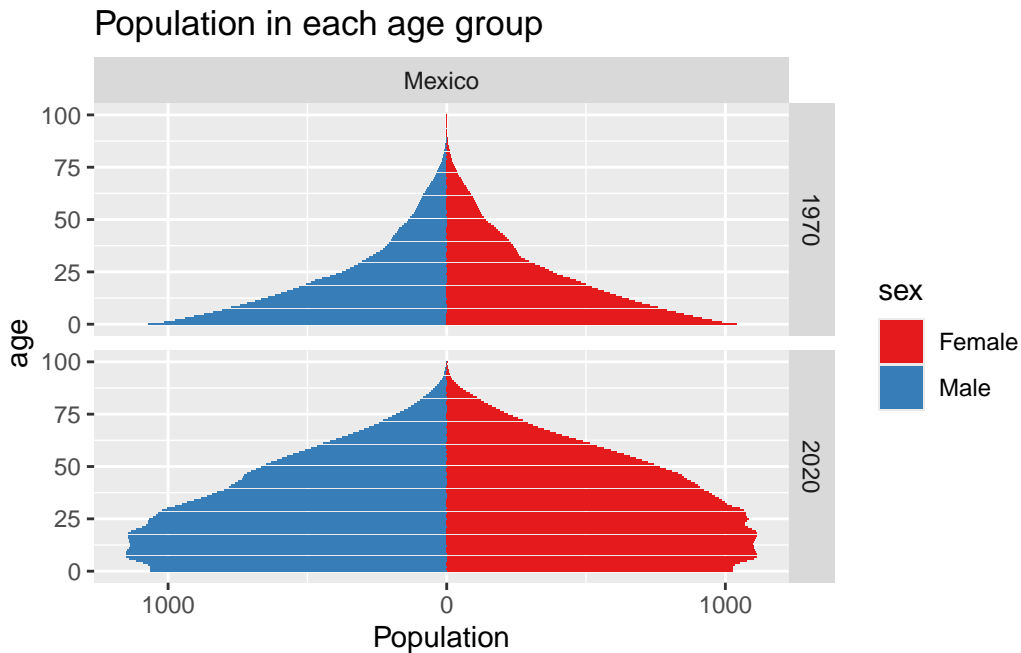
dm <- dm |>
  clean_names() |>
  select(region_subregion_country_or_area, iso3_alpha_code, year, x0:sex) |>
  rename(region = region_subregion_country_or_area) |>
  mutate(across(x0:x100, as.numeric))

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

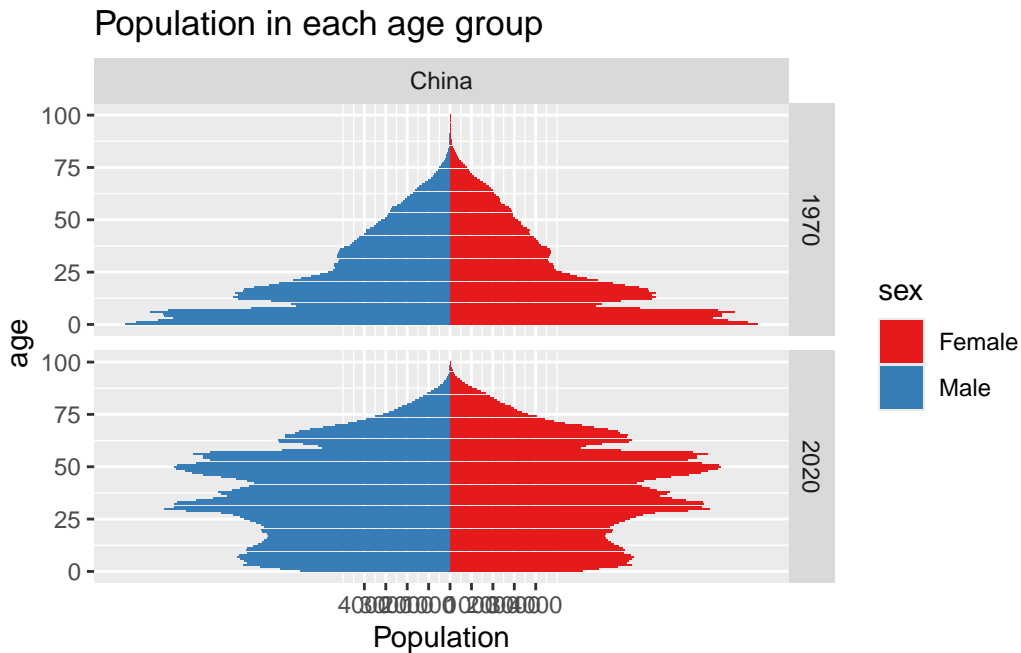
d_long|>
  filter(region=="Mexico",
    year%in% c(1970, 2020)) |>
  mutate(population=ifelse(sex=="Male", -pop, pop)) %>%
ggplot(aes(x = age, y = population, fill = sex)) +
  facet_grid(year~region)+
  geom_bar(stat="identity")+
  ggtitle("Population in each age group")+
  ylab("Population")+
  coord_flip() +

```

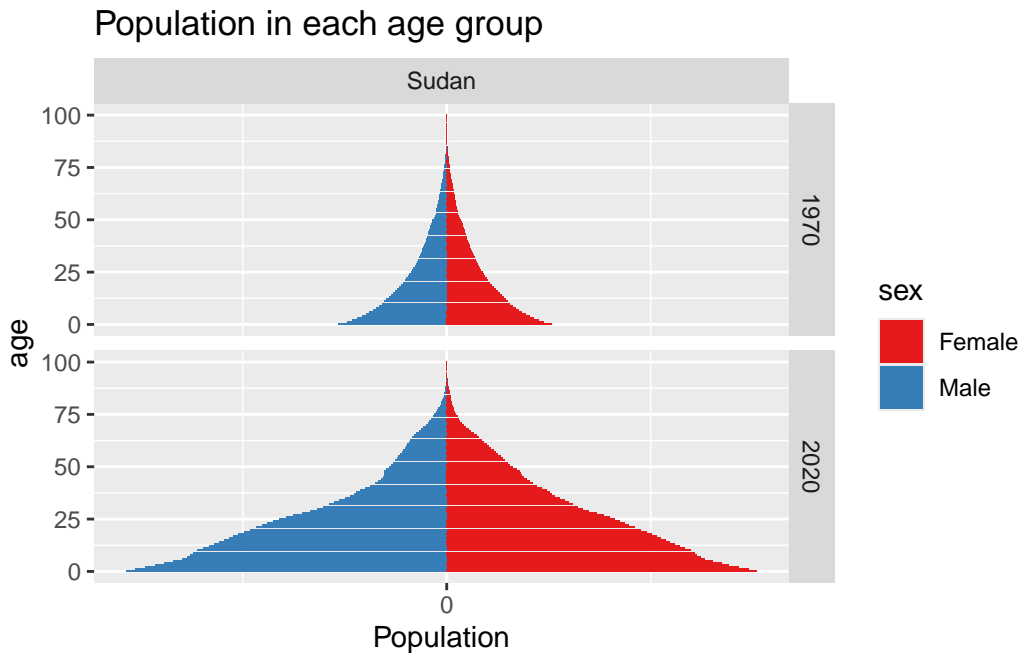
```
scale_y_continuous(breaks = seq(-4000, 4000, 1000),
                   labels = c(seq(4000, 0, -1000), seq(1000, 4000, 1000))) +
scale_fill_brewer(palette = "Set1")
```



```
d_long|>
  filter(region=="China",
         year%in% c(1970, 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")
```



```
d_long|>
  filter(region=="Sudan",
         year%in% c(1970, 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")
```



The relationship between population growth or decline and the net reproduction rate (NRR) is clearly reflected in these graphs. When NRR is high, as in Sudan, large birth cohorts create a wide base in the population pyramid and sustained population growth. This is evident in 2020, where the large cohorts born around 1970 have aged into the 15–50 year range and, because NRR remains high, have continued to produce even larger birth cohorts. When NRR remains below replacement for an extended period, as in China, younger cohorts become smaller than older ones, leading to population aging and eventual population decline. In contrast, when NRR declines toward replacement, as in Mexico, the base of the pyramid narrows and population growth slows, but the population can continue to grow for some time due to momentum from earlier high fertility. Thus, although Mexico's NRR has recently fallen below 1, this decline is recent, and the population structure has not yet shifted toward the pronounced aging seen in China.

Exercise 2

Repeat the graph above (PPR from 2 to 3, based on sex of first two children), but plot by age group of mother. Comment on observations in one sentence.

Graph is done and just need to inretrept

```
d <- read_csv("data/canada_parity.csv")
```



```
dy <- d |>
  filter(child_age_1<21)
```

```
ppr2_m <- dy |>
  filter(child_age_2<16) |>
  filter(child_sex_1==1) |>
  group_by(age_group, child_sex_2) |>
  summarise(ppr2 = sum(nchild>2)/n()) |>
  rename(sex = child_sex_2, age = age_group) |>
  mutate(sex1 = "first born is male")
```

```
ppr2_f <- dy |>
  filter(child_age_2<16) |>
  filter(child_sex_1==2) |>
  group_by(age_group, child_sex_2) |>
  summarise(ppr2 = sum(nchild>2)/n()) |>
  rename(sex = child_sex_2, age = age_group) |>
  mutate(sex1 = "first born is female")
```

```
ppr2_m |>
  bind_rows(ppr2_f) |>
  mutate(sex = ifelse(sex == 1, "Male", "Female")) |>
  ggplot(aes(age, ppr2, color = factor(sex))) + facet_grid(~sex1) +
  geom_point() +
  ggtitle("Parity progression ratios to third child by sex of first two children \nCanada 20
  scale_color_discrete(name = "sex of second born") +
  ylab("PPR")+xlab("age of mother")
```

Parity progression ratios to third child by sex of first two children Canada 2011



There appears to be a relationship between the sex composition of the first two children and women's decision to have a third child. Among women aged 35–39 whose first child is female, those with two female children are about 33% likely to have a third child, compared to about 28% among those whose second child is male. Similar patterns are observed across other age groups, with roughly a 5 percentage point lower likelihood of having a third child when the first two children are female and then male compared to when they are both female. A comparable pattern emerges when the first child is male: women with two male children are generally more likely to progress to a third birth than those whose second child is female. This suggests that parents may be more likely to have a third child when the first two children are the same sex, potentially reflecting a preference for having at least one child of each sex. An exception appears among women aged 50 and older. When the first-born child is male, there is approximately a 10 percentage point difference in parity progression depending on whether the second child is male or female. In contrast, when the first-born child is female, the pattern reverses. The reason for this divergence in the 50+ age group is unclear.