Barcelona Summer School of Demography

Module 2. Demography with R

1. Basic Demographic Measures

Basic Demographic measures

Tim Riffe
tim.riffe@gmail.com

5 July 2021

1 Exercises

1.1 Load the data `PopulationSpain.txt` and `BirthsSpain.txt`

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.2     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(readr)
#Birth counts in 2014

B  <- read_delim("BirthsSpain.txt",
                 delim = " ",
                 trim_ws = TRUE,
                 skip = 2) %>%
```

```
      mutate(Age = parse_number(Age)) %>%
      select(Year, Age, Births = Total)
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   Year = col_double(),
##   Age = col_character(),
##   Total = col_double()
## )
```

```
P  <- read_delim("PopulationSpain.txt",
                 delim = " ",
                 trim_ws = TRUE,
                 skip = 2,
                 col_types = "ccddd") %>%
      mutate(Age = parse_number(Age)) %>%
      pivot_longer(Female:Total,
                   names_to = "Sex",
                   values_to = "Population") %>%
      arrange(Year, Sex, Age)
```

Note: there was some sort of universe adjustment in the Spanish data in 1975. Possibly a switch from *de facto* to *de jure* population definition. HMD codes such changes to a moment in time, with `-` suffix on the *before* populaton and a `+` suffix on the *after* population. Accounting for this adds some additional logic to the code, but we can pass over that. Actually we could also just remove 1975 altogether if so inclined.

1.2 Calculate age-specific fertility rates from 1950 until 2014 (annually) from age 12 to 55. Skip the year 1975 (if you want to make this easier!).

I'm going to do this the tidy way, first calculating exposure, then merging with births, then calculating rates using `mutate()`. Two helper functions are `nchar()`, which tells you the number of characters in string (`Year` in our case) and `grepl()` which gives logical character pattern matching (i.e. `TRUE` if there's a match, otherwise `FALSE`).

```
# Jan 1 pops
P1 <-
  P %>%
  filter(Sex == "Female",
         nchar(Year) == 4 | Year == "1975+") %>%
  mutate(side = "left",
         Year = parse_number(Year))
# Dec 31 pops
P2 <-
  P %>%
  filter(Sex == "Female",
         nchar(Year) == 4 | Year == "1975-") %>%
  mutate(side = "right",
         Year = parse_number(Year),
```

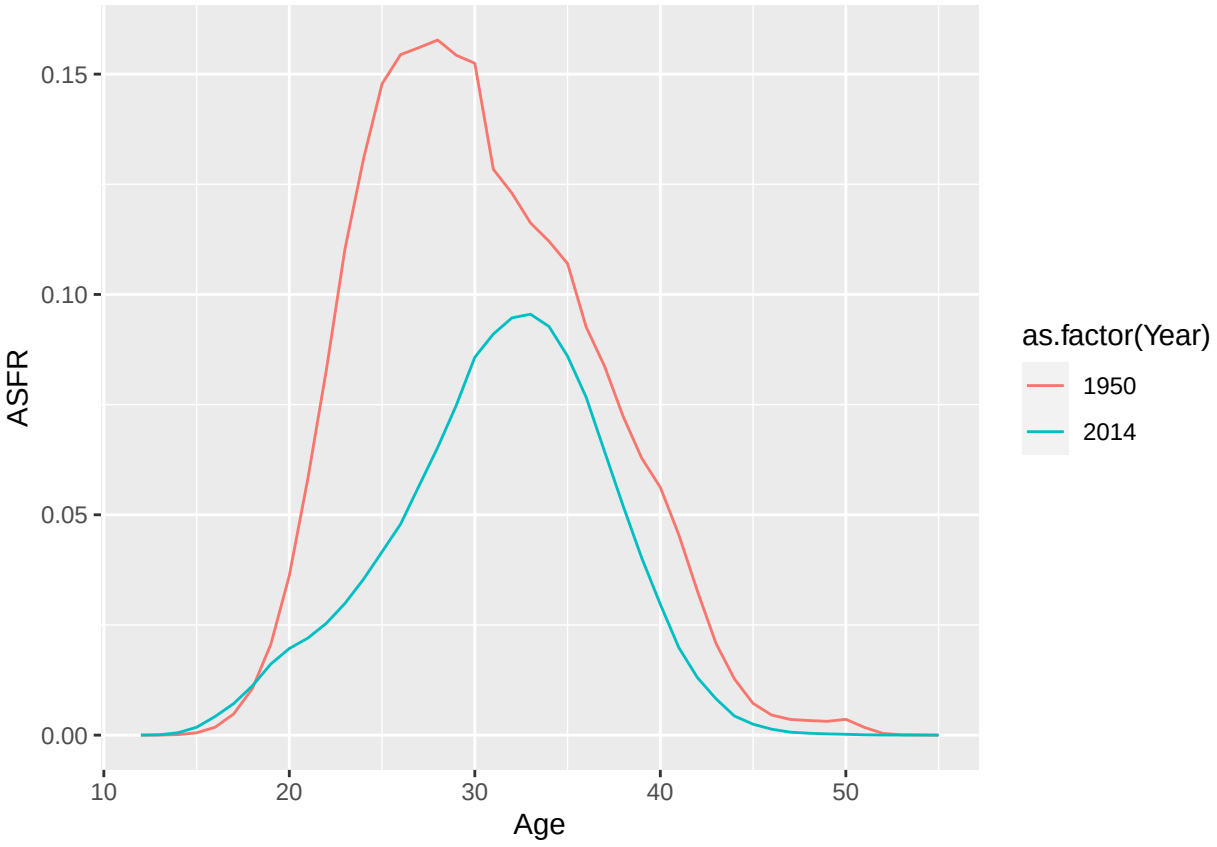
```
         Year = Year - 1)
# stick together, put side by side.
# could have also just joined...)
ES <-
  P1 %>%
  bind_rows(P2) %>%
  pivot_wider(names_from = "side", values_from = "Population", names_prefix = "P_") %>%

  # calculate exposure
  mutate(Exposure = (P_left + P_right) / 2) %>%

  # now RIGHT joint to births
  # (cuts down exposure to just ages where there is fertility)
  right_join(B, by = c("Year","Age")) %>%
  # Calculate ASFR
  mutate(ASFR = Births / Exposure)
```

1.3 Plot the age-specific rates in 1950 and 2014.

```
ES %>%
  filter(Year %in% c(1950, 2014)) %>%
  ggplot(aes(x = Age, y = ASFR, color = as.factor(Year), group = Year)) +
  geom_line() +
  scale_color_discrete()
```

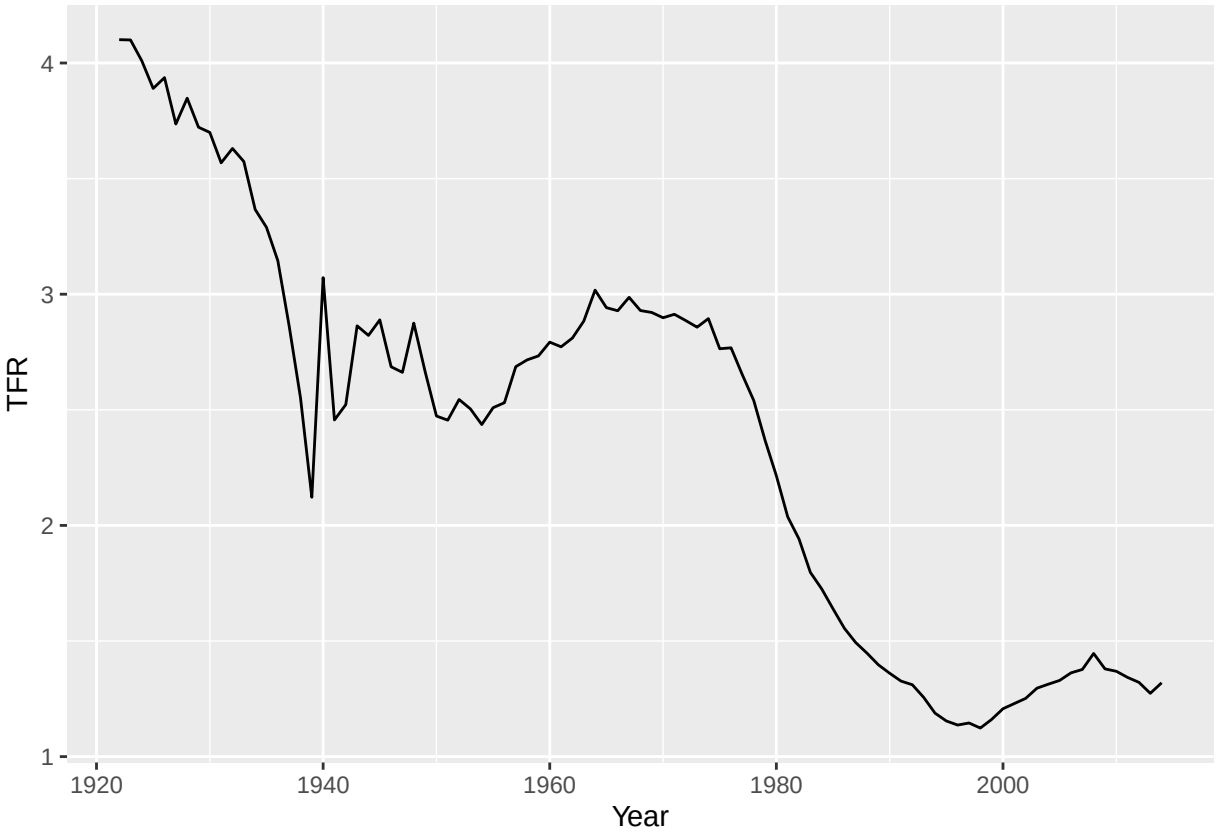


1.4 Calculate the TFR at each year.

```
TFR <-
  ES %>%
  group_by(Year) %>%
  summarize(TFR = sum(ASFR))
```

1.5 Plot TFR over time.

```
TFR %>%
  ggplot(aes(x = Year, y = TFR)) +
  geom_line()
```

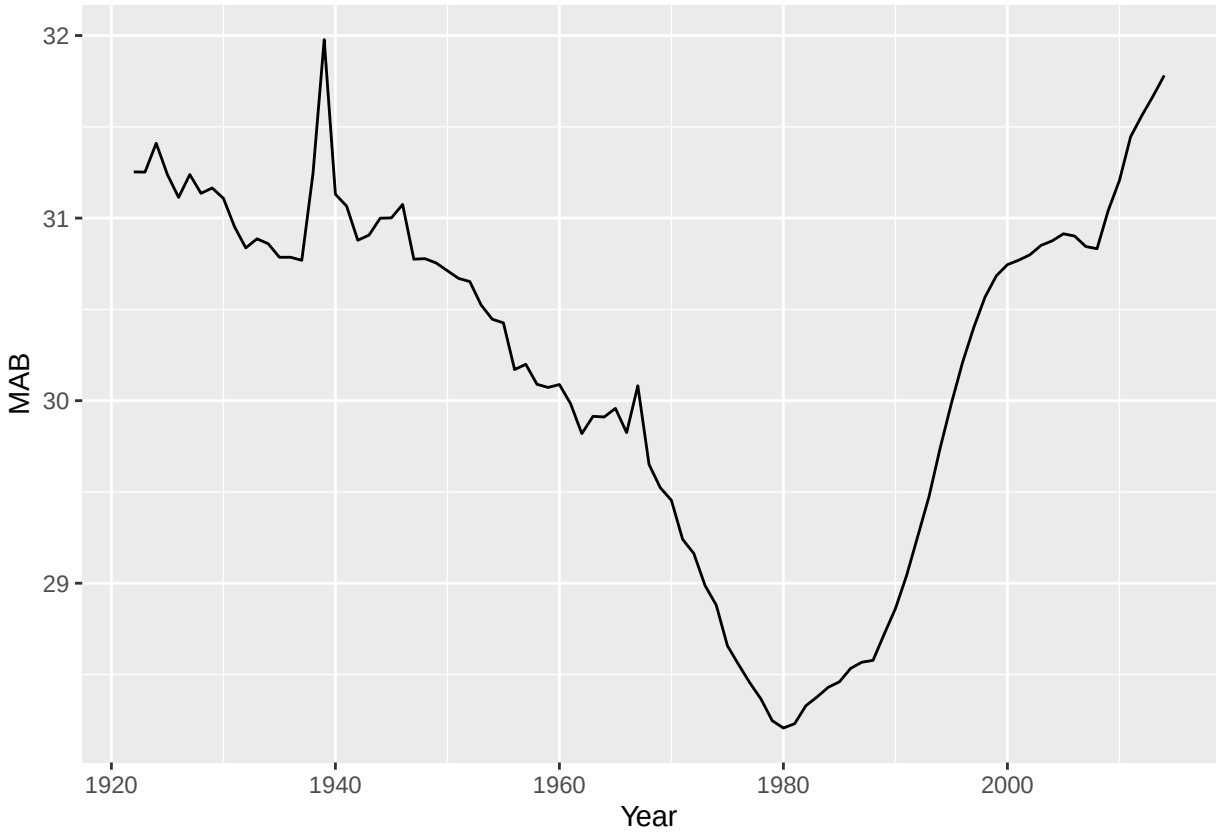


1.6 Calculate the mean age at birth over time.

```
MAB <-
  ES %>%
  mutate(Agemid = Age + .5) %>%
  group_by(Year) %>%
  summarize(MAB = sum(Agemid * ASFR) / sum(ASFR))
```

1.7 Plot the mean age over time.

```
MAB %>%
  ggplot(aes(x = Year, y = MAB)) +
  geom_line()
```


32
31
30
29
MAB
1920
1940
1960
1980
2000
Year