Comparing Period TFR for 2 groups

Carl Schmertmann

10 Nov 2019

TFR for one group

Suppose that there are A possible single-year reproductive ages $x = 1 \ldots A$ in which individuals may have children. At some of these ages fertility rates may be zero for some groups (for example, for women at $x > 50$).

Suppose also that there are a total of B births in a given year to some group of interest (such as "all women", "all men", "Asian-American men", "residents of Montana" ...). Call π_x the fraction of the B births that happened to group members who were age x. That is, births disaggregated by age are

$$B_x = B\,\pi_x \quad x = 1, 2, \ldots, A$$

If there were N_x group members at age x at mid-year, then period TFR for this group is

$$TFR = \sum_x \frac{B_x}{N_x} = B \sum_x \pi_x \left(\frac{1}{N_x}\right) = B\,E_\pi\left(N_x^{-1}\right) \tag{1}$$

where the expectation in (1) is taken over the observed distribution of parents' ages for this year's births.

Comparing TFR for 2 groups

If there are two groups then the ratio of their period TFRs is

$$\frac{TFR_1}{TFR_2} = \frac{B_1 \cdot E_{\pi_1}\left(N_{1x}^{-1}\right)}{B_2 \cdot E_{\pi_2}\left(N_{2x}^{-1}\right)} \tag{2}$$

In some cases (for example when the groups are "women" and "men") the total births are the same for both groups. For those cases:

$$\frac{TFR_1}{TFR_2} = \frac{E_{\pi_1}\left(N_{1x}^{-1}\right)}{E_{\pi_2}\left(N_{2x}^{-1}\right)} \tag{3}$$

Using the results derived in the Appendix below, the ratio in (3) can be approximated fairly well by

$$\frac{TFR_1}{TFR_2} \approx \frac{N_{2,m_2}}{N_{1,m_1}}$$

where m_1 and m_2 are the mean ages of parents in the two groups for this year's observed births.

USA data

USA *male* fertility data from Dudel, C. and S. Klüsener. Male fertility data for high-income countries [unpublished data]. Submitted to the HFC by C. Dudel # on 15.05.2019. downloaded from Human Fertility Collection 2 Nov 2019 https://www.fertilitydata.org/data/RAW_DATA/m_USA_51.zip

USA *female* fertility data from H**F**D

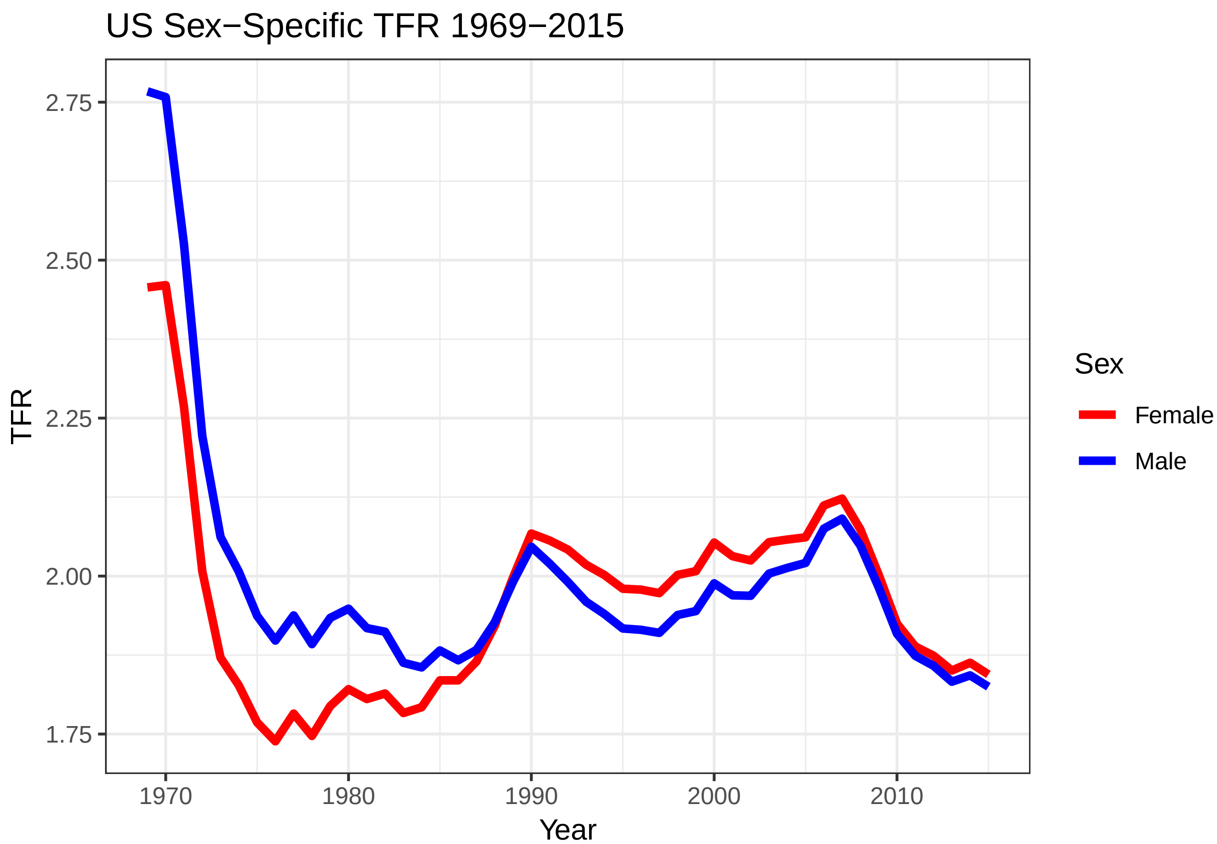
USA *male* and *female* population data from H**M**D

```r
# construct a data frame with annual USA data, from the listed sources

# rates for 1969-2015 x ages 15-59
asfr_male = read.table(file='m_ASFR_USA.txt', header=TRUE,
                       colClasses='numeric') %>%
            add_column(Sex='Male')

# rates for 1969-2015 x ages 12-55
asfr_female = read.table(file='USAasfrTR.txt', skip=2, header=TRUE) %>%
       mutate(Age = 11 + as.numeric(Age)) %>%
       filter(Year %in% unique(asfr_male$Year)) %>%
       group_by(Year,Age) %>%
       summarize(ASFR = mean(ASFR)) %>%
       add_column(Sex='Female')

ASFR = bind_rows(
         asfr_male,
         asfr_female
        )

pop = read.table('Population.txt',skip=2, header=TRUE) %>%
        mutate( Year = as.numeric(as.character(Year)),
                Age  = as.numeric(as.character(Age))) %>%
        filter(Age %in% unique(ASFR$Age),
               Year %in% unique(ASFR$Year)) %>%
        select(Year,Age,Female,Male) %>%
        gather(key=Sex, value=pop, -Year, -Age)
```

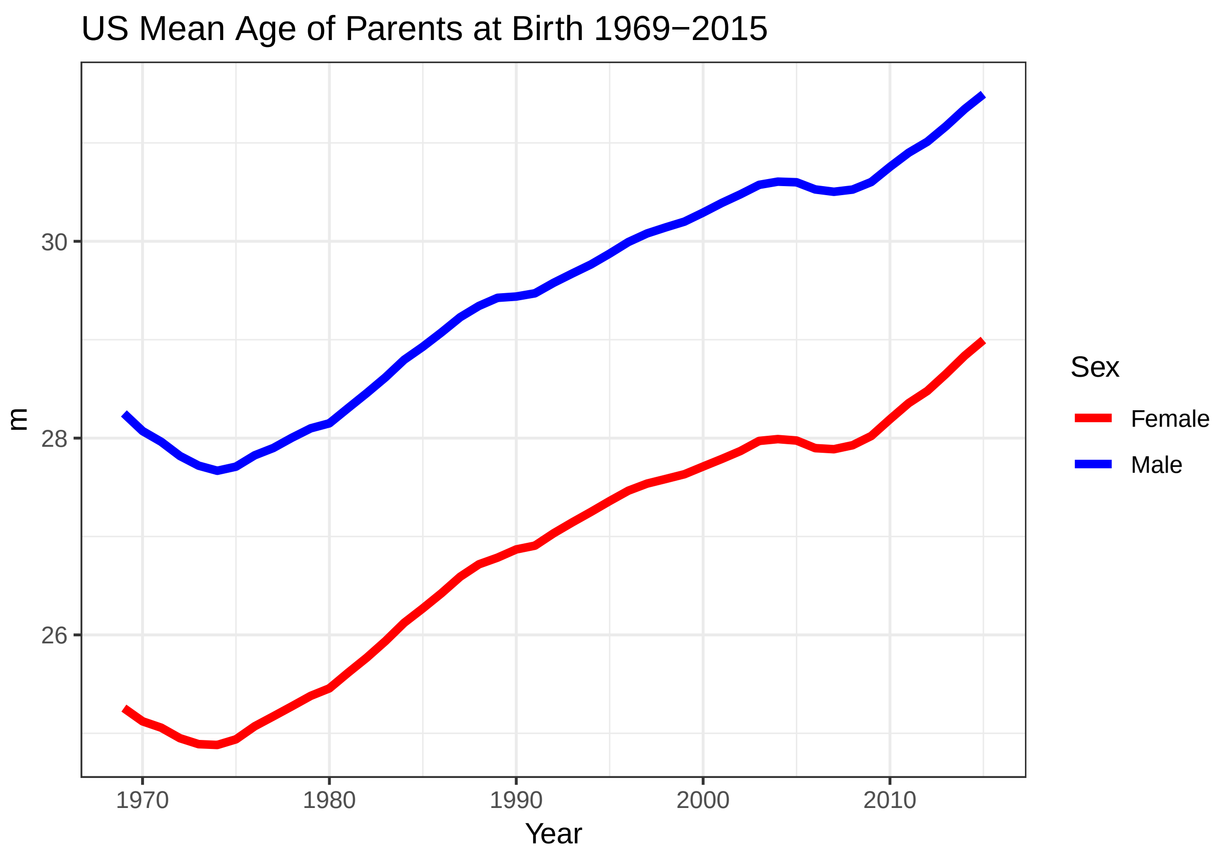
```
## Warning: NAs introduced by coercion

## Warning: NAs introduced by coercion
```

```r
## add estimated births, fraction of births by age, etc.

big = full_join(pop, ASFR, by=c('Year','Age','Sex')) %>%
         mutate(births = pop * ASFR)

big$births[is.na(big$births)] = 0

# summarize each year x sex

mean_expos = function(A,P,M) {
  approx(x=A, y=P, xout=M )$y
}

df = big %>%
       group_by(Year,Sex) %>%
       summarize( TFR = sum(ASFR, na.rm=TRUE),
                  B   = sum(births),
                  m   = weighted.mean( Age+.50   , births),
                  Nm  = mean_expos(Age+.50,pop,m),
                  approx_TFR = B/Nm)
```

```r
ggplot(data=df) +
   aes(x=Year, y=TFR, color=Sex) +
   geom_line(lwd=1.5) +
   scale_color_manual(values=c('red','blue')) +
   labs(title='US Sex-Specific TFR 1969-2015') +
  theme_bw()
```

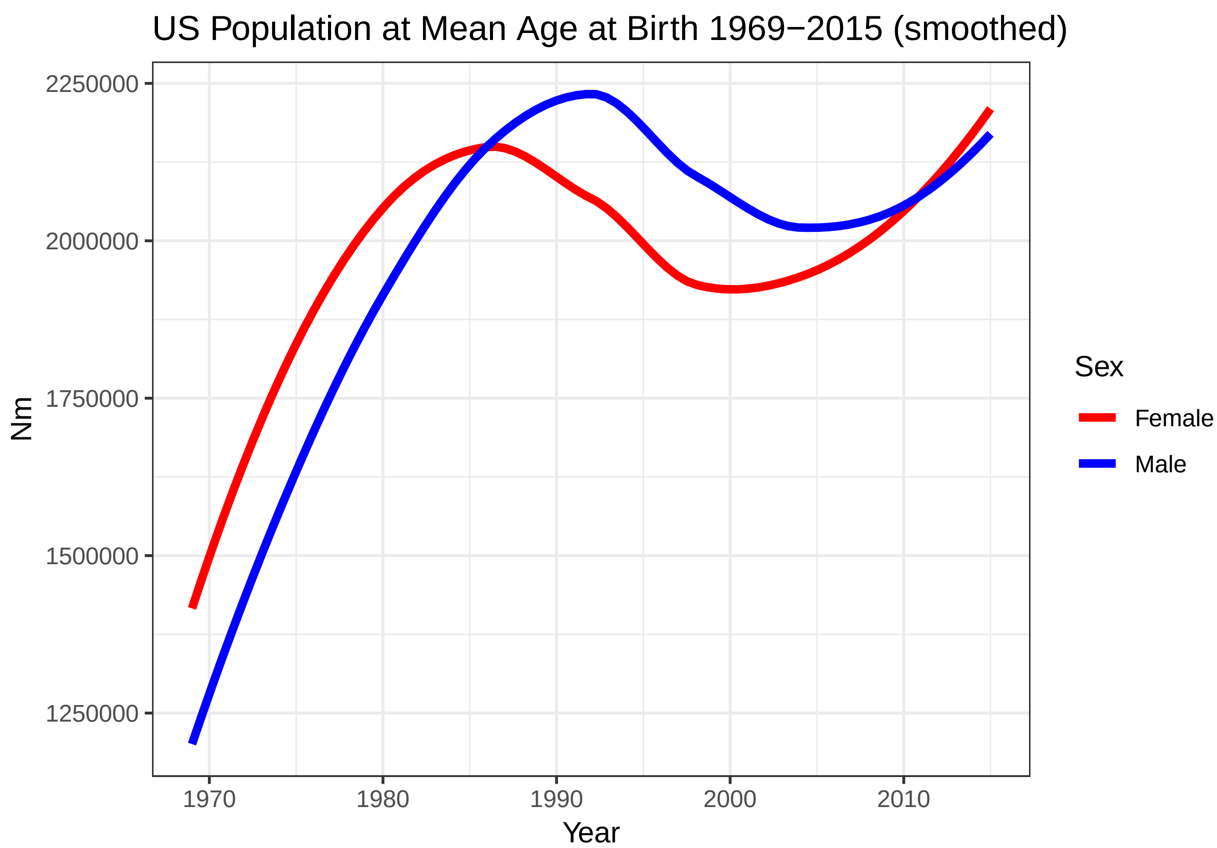



```r
ggplot(data=df) +
   aes(x=Year, y=m, color=Sex) +
   geom_line(lwd=1.5) +
   scale_color_manual(values=c('red','blue')) +
   labs(title='US Mean Age of Parents at Birth 1969-2015') +
  theme_bw()
```




```r
ggplot(data=df) +
   aes(x=Year, y=Nm, color=Sex) +
   geom_smooth(lwd=1.5,se=FALSE) +
   scale_color_manual(values=c('red','blue')) +
   labs(title='US Population at Mean Age at Birth 1969-2015 (smoothed)') +
  theme_bw()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Example: US TFRs in 1995

Select 1995 data and analyze in more detail.

```
tmp.big = filter(big, Year == detail_year)
tmp.df  = filter(df,  Year == detail_year)

TFR          = tmp.df$TFR
approx_TFR = tmp.df$approx_TFR
m            = tmp.df$m
Nm           = tmp.df$Nm

B = mean(tmp.df$B) #avg of (slightly difft) male and female total births

names(TFR) = names(approx_TFR) = names(m) = names(Nm) = c('Female','Male')

ratio        = TFR['Male']/TFR['Female']
approx_ratio = approx_TFR['Male']/approx_TFR['Female']
```

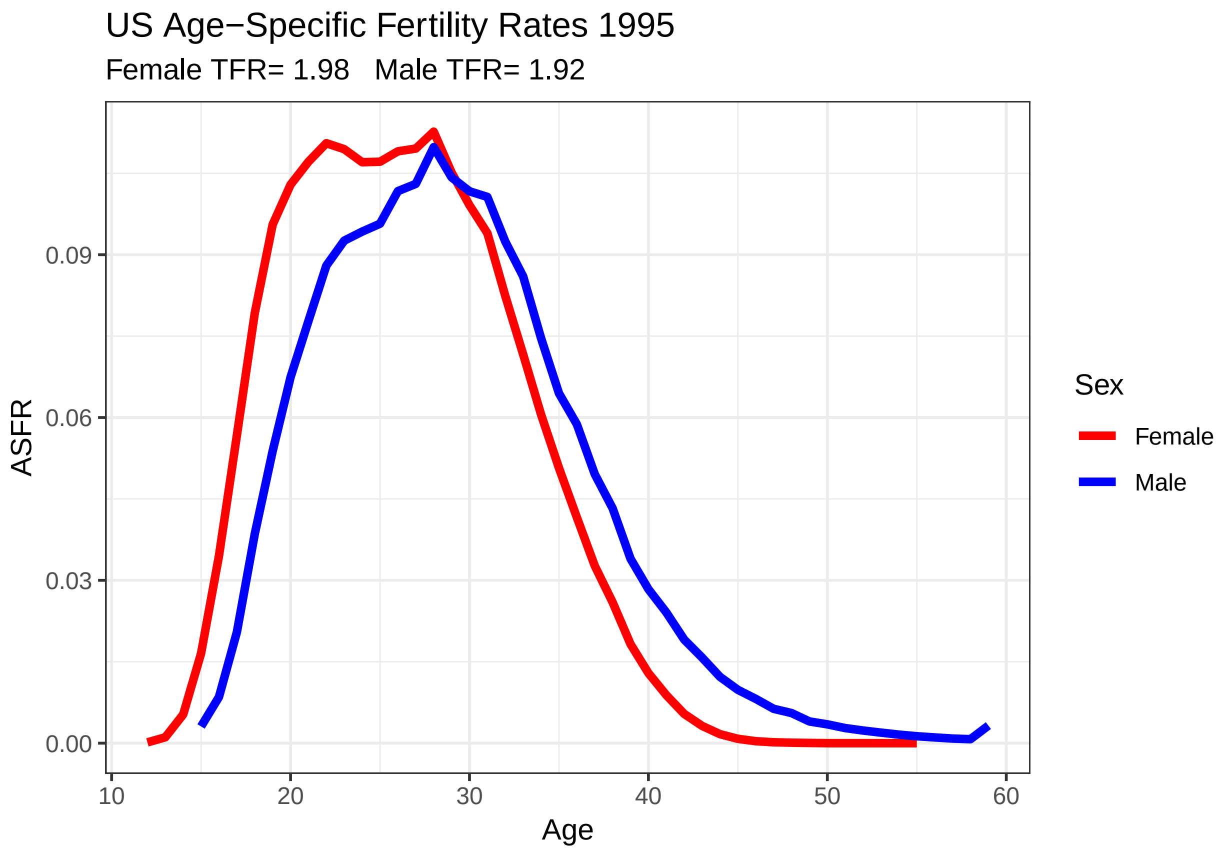
For the US in 1995, female TFR was **1.98** and male TFR was **1.92**.

Now illustrate 1995 fertility rates, births, and populations, by age and sex.

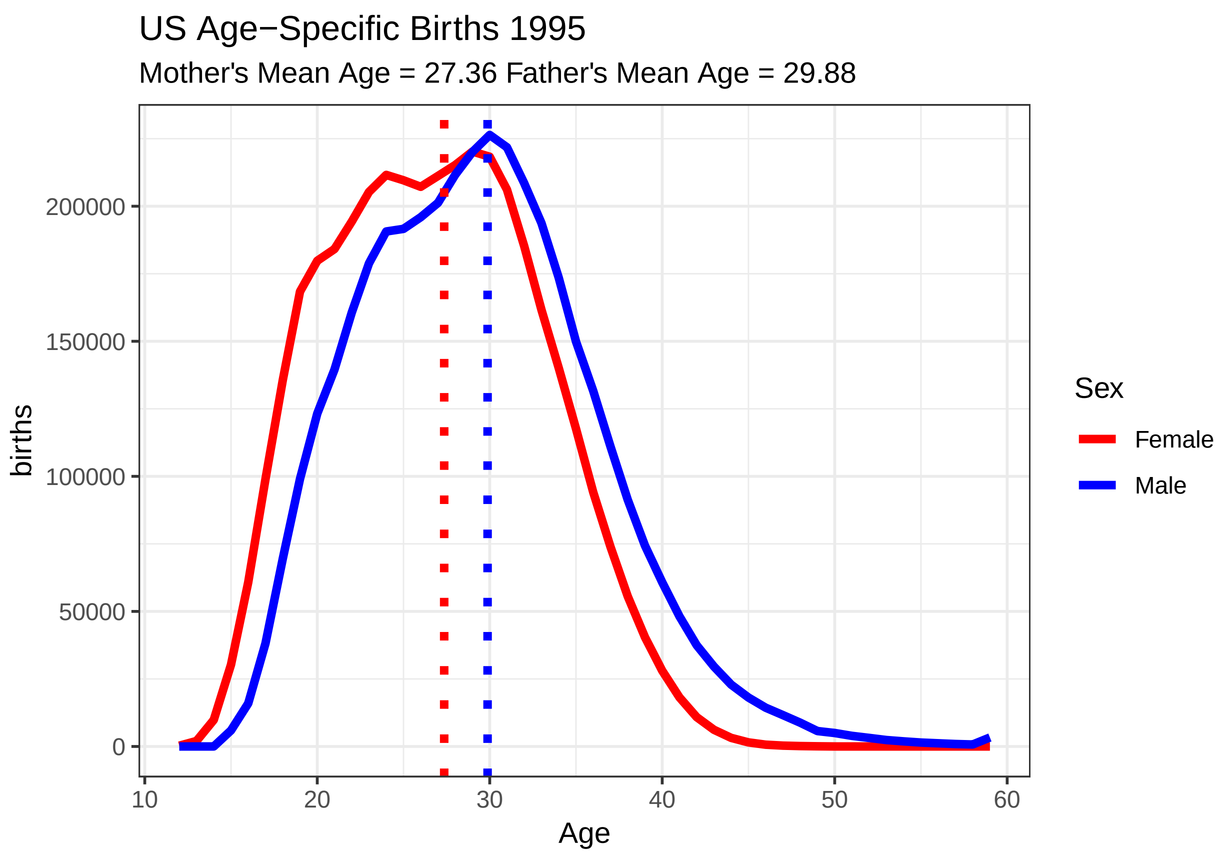
```
ggplot(data=tmp.big) +
  aes(x=Age, y=ASFR, color=Sex) +
  geom_line(lwd=1.5) +
  scale_color_manual(values=c('red','blue')) +
```

```r
  labs(title=paste('US Age-Specific Fertility Rates',detail_year),
       subtitle=paste('Female TFR=',round(TFR['Female'],2),
                      '   Male TFR=',round(TFR['Male']   ,2))) +
  theme_bw()
```

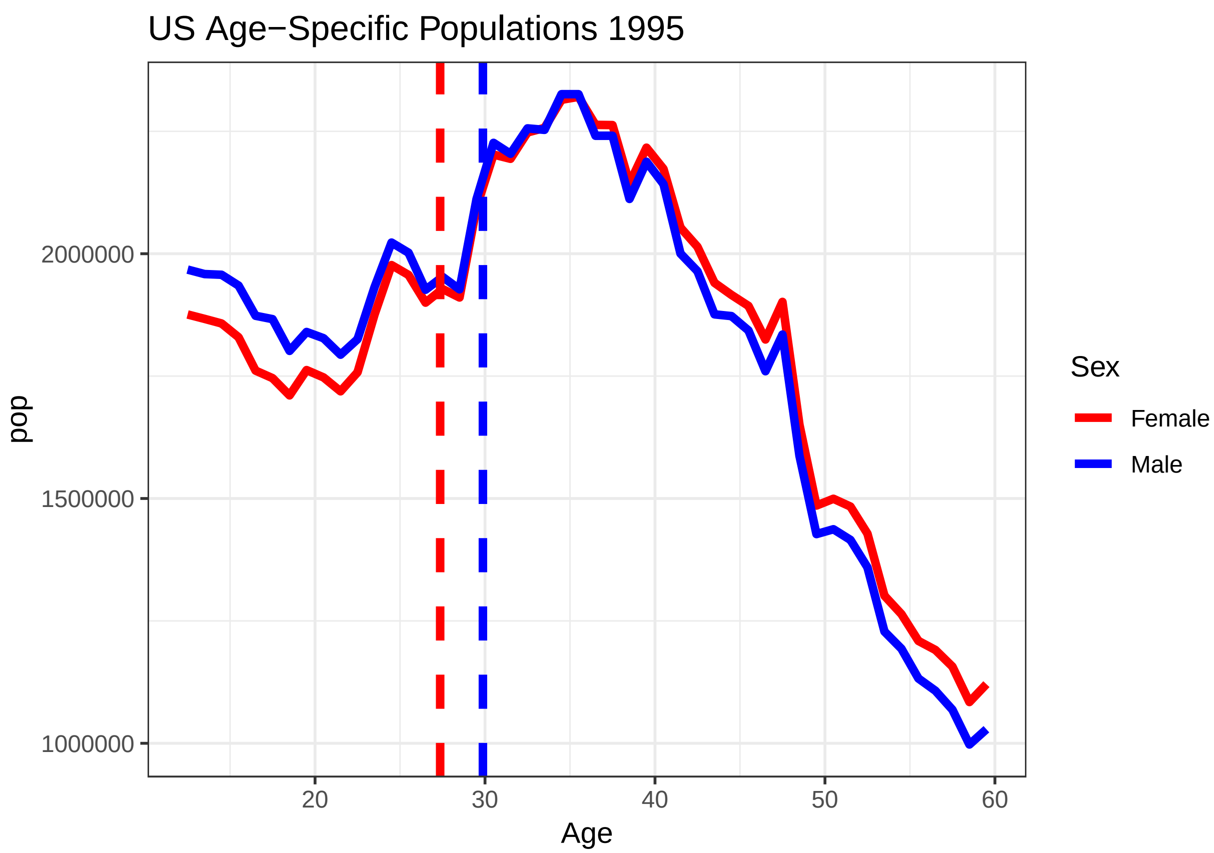
```
## Warning: Removed 7 rows containing missing values (geom_path).
```




```r
ggplot(data=tmp.big) +
  aes(x=Age, y=births, color=Sex) +
  geom_line(lwd=1.5) +
  geom_vline(xintercept = m[c('Female','Male')], color=c('red','blue'),
             lwd=1.5, lty='dotted') +
  scale_color_manual(values=c('red','blue')) +
  labs(title=paste('US Age-Specific Births',detail_year),
       subtitle=paste('Mother\'s Mean Age =',round(m['Female'],2),
                      'Father\'s Mean Age =',round(m['Male'],2))) +
  theme_bw()
```




```
ggplot(data=tmp.big) +
  aes(x=Age+0.5, y=pop, color=Sex) +
  geom_line(lwd=1.5) +
  geom_vline(xintercept = m[c('Female','Male')], color=c('red','blue'),
             lwd=1.5,lty='dashed') +
  scale_color_manual(values=c('red','blue')) +
  labs(title=paste('US Age-Specific Populations',detail_year),
       x='Age') +
  theme_bw()
```



Now we will check the approximations based on

- mean parental ages for 1995 births: 27.36 for females and 29.88 for males
- 1995 populations at mean ages: 1924013 for females and 2154632 for males

Thus

Sex	True 1995 TFR	m	N_m (000s)	Approx TFR
Male	1.92	29.88	2155	1.81
Female	1.98	27.36	1924	2.03

and

True Male/Female TFR Ratio	Approximate Ratio
0.97	0.89

Appendix: Approximation

We can approximate the TFRs in (1) by considering a Taylor series approximation for N_x^{-1}. For the purposes of this derivation, temporarily ignore the discrete nature of ages in observed data, and treat N_x as a continuous function of exact age x.

Define the (period) mean age of parents in our group as m. The second-order Taylor approximation to N_x^{-1} is

$$N_x^{-1} \approx N_m^{-1} + (x-m)\left[-N_m^{-2} N_m'\right] + \tfrac{1}{2}(x-m)^2\left[2N_m^{-3} N_m' - N_m^{-2} N_m''\right]$$

Thus the expectation in (1) is

$$E(N_x^{-1}) \approx N_m^{-1} + \tfrac{\sigma_x^2}{2}\left[2N_m^{-3} N_m' - N_m^{-2} N_m''\right]$$

where σ_x^2 is the observed variance of parental ages. Typical values for the variance term are in $[25, 50]$, so when exposure N is measured in thousands or millions, the second-order term is many orders of magnitude smaller than the first-order term.

Using only the first term leads to a very simple approximation for TFR, based on (1) that is fairly robust:

$$TFR \;\approx\; \frac{B}{N_m} = \frac{\text{total annual births}}{\text{mid-year population at mean parental age}}$$