Male-Female Differences

Henrik Schubert

2023-06-05

Abstract

This file is going to compare male and female fertility rates in Mexico. We will use graphical tools, distributional comparison measures, decomposition methods and statistical tests. The comparisons will be both over time, across regions, and at the national level

Time trend

First, we compare the time-trend of male and female total fertility rates.

```r
# Load the national level data
load("Data/tfr_national_mexico.Rda")

# What is the trend in the national tfr s
ggplot(tfr_nat, aes(year)) +
  geom_line(aes(y = tfr_f, colour = "female"), linewidth = 1.4) +
  geom_line(aes(y = tfr_m, colour = "male"), linewidth = 1.4) +
  annotate(geom = "text",
           x = 2001, y = 3.9,
           label = "Male",
           colour = MPIDRblue, size = 8, family = "serif") +
  annotate(geom = "text",
           x = 1998, y = 3,
           label = "Female",
           colour = MPIDRred, size = 8, family = "serif") +
  scale_colour_manual(values = c(MPIDRred, MPIDRblue)) +
  ylab("Total fertility rate") +
  xlab("Year") +
  guides(colour = "none")
```
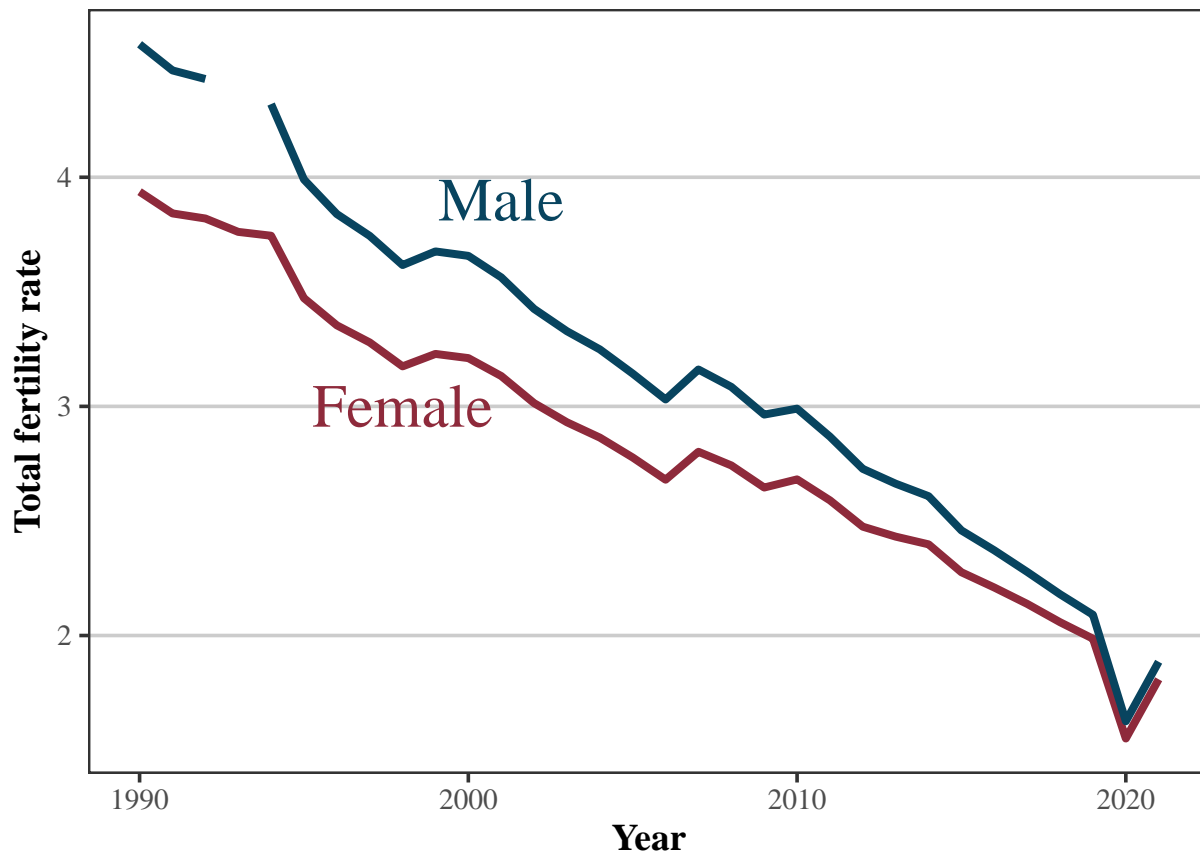
Age-Specific Differences

Now, we plot the age-specific differences in fertility between men and women.

```
# Load the national level data
load("Data/asfr_national_mexico.Rda")

# What is the trend in the national TFR for women
a <- ggplot(asfr_nat, aes(x = age, alpha = year)) +
  geom_line(aes(y = asfr_f, colour = "female", group = year),
            linewidth = 1.4, colour = MPIDRred) +
  annotate(geom = "text",
           x = 38, y = 0.17,
           label = "Female",
           colour = MPIDRred, size = 8, family = "serif") +
  scale_y_continuous(expand = c(0, 0)) +
  ylab("Age-specific fertilitly rate") +
  xlab("Age") +
  guides(colour = "none", alpha = "none")

# Plot the trend in the national TFR for men
b <- ggplot(asfr_nat, aes(x = age, alpha = year)) +
  geom_line(aes(y = asfr_m, colour = "male", group = year),
            linewidth = 1.4, colour = MPIDRblue) +
  annotate(geom = "text",
           x = 40, y = 0.17,
```
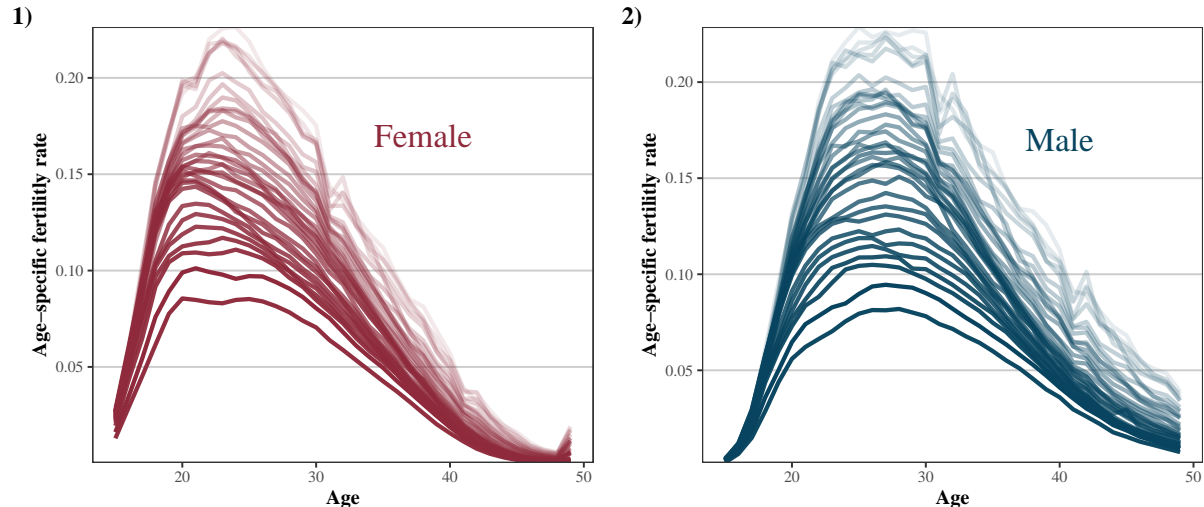
```
            label = "Male",
            colour = MPIDRblue, size = 8, family = "serif") +
  scale_y_continuous(expand = c(0, 0)) +
  ylab("Age-specific fertilitly rate") +
  xlab("Age") +
  guides(colour = "none", alpha = "none")


a + b + plot_annotation(tag_levels = c("1"), tag_suffix = ")")
```


Cumulative distribution

Moreover, to understand better the cumulative difference, we are going to estimate the cumulative age-specific fertility rate.

$$cumulative f(x) = \sum_{12}^{x} f(x)$$

```
# Estimate the cumulative distribution
asfr_nat <- asfr_nat %>%
  group_by(year) %>%
  arrange(age) %>%
  mutate(cum_f = cumsum(asfr_f),
         cum_m = cumsum(asfr_m))

# What is the trend in the national tfr s
c <- ggplot(asfr_nat, aes(x = age, alpha = year)) +
  geom_line(aes(y = cum_f, colour = "female", group = year),
            linewidth = 1.4, colour = MPIDRred) +
  annotate(geom = "text",
           x = 22, y = 3.2,
           label = "Female",
           colour = MPIDRred, size = 8, family = "serif") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 4.5)) +
  scale_x_continuous(expand = c(0, 0)) +
```

```
  ylab("Cumulative age-specific fertilitly rate") +
  xlab("Age") +
  guides(colour = "none", alpha = "none")


d <- ggplot(asfr_nat, aes(x = age, alpha = year)) +
  geom_line(aes(y = cum_m, colour = "male", group = year),
            linewidth = 1.4, colour = MPIDRblue) +
  annotate(geom = "text",
           x = 24, y = 3.2,
           label = "Male",
           colour = MPIDRblue, size = 8, family = "serif") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 4.5)) +
  scale_x_continuous(expand = c(0, 0)) +
  ylab("Cumulative age-specific fertilitly rate") +
  xlab("Age") +
  guides(colour = "none", alpha = "none")


c + d + plot_annotation(tag_levels = c("1"), tag_suffix = ")")
```
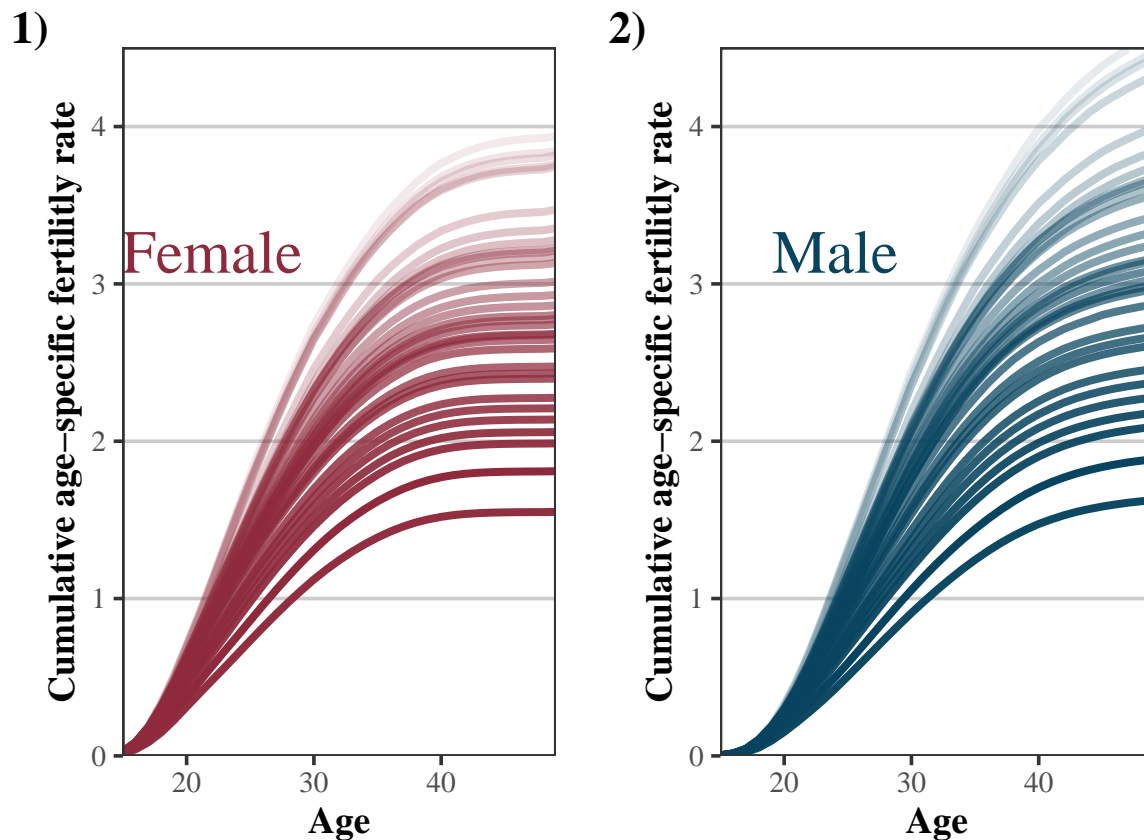

Decomposing the difference in the total fertility rate

Another tool to study the difference between the male and female total fertility rate is decomposition methods. The total fertility rate is the sum of the age-specific fertility rates. Because of the additive nature, it is feasible to use standard demographic decomposition methods (Preston, Heuveline, and Guillot (2008), p. 28)

$$\Delta TFR_{f,m} = \sum_i \left(\frac{1}{C_i^f} - \frac{1}{C_i^m} \right) \left[\frac{b_i^f + b_i^m}{2} \right] + \sum_i (b_I^f - b_i^m) \cdot \frac{1}{\left[\frac{C_i^f + C_i^m}{2} \right]}$$

The total fertility rate can be decomposed using the following equation:

$$\Delta TFR_{f,m} = \sum_i s_i^f \Delta_i^{f,m} \left(\frac{B}{P} \right)_{i,f} + \sum_i \left(\frac{B}{P} \right)_{i,f} \Delta s_i^{f,m} + \sum_i \Delta c_i^{f,m} \Delta \left(\frac{B}{P} \right)$$

where B is the number of births, P is the population size, s is the share of overall population, i indexes the different age-groups, and f and m are the beginning and ending years.

In the figure below, we see that

```r
# Estimate the decomposition
comp_asfr <- asfr_nat %>%
                group_by(year) %>%
                mutate(pop_share_f = pop_share(exposure_f),
                       pop_share_m = pop_share(exposure_m),
                       delta_pop = difference(pop_share_f, pop_share_m),
                       delta_rate = difference(asfr_f, asfr_m),
                       change_rate = pop_share_f * delta_rate,
                       change_pop  = asfr_f * delta_pop,
                       change_inter = delta_pop * delta_rate)


# Make the decomposition
decomposition <- comp_asfr %>%
  select(year, age, change_rate, change_pop, change_inter) %>%
  pivot_longer(cols = starts_with("change"),
               names_to = "component",
               values_to = "contribution")


# Plot the decomposition
decomposition %>%
  filter(year %in% c(1990, 2000, 2010, 2020)) %>%
  ggplot(aes(age, contribution, fill = component)) +
  geom_col(colour = "white", linewidth = 0.01) +
  geom_hline(yintercept = 0, colour = "black") +
  facet_wrap(~ year) +
  scale_fill_manual(values = c(MPIDRred, MPIDRorange, MPIDRblue),
                    labels = c(expression(paste(Delta, "Population and rate")),
                               expression(paste(Delta, "Population")),
                               expression(paste(Delta, "Rate"))),
                    name = "Component: ") +
  scale_x_continuous(expand = c(0, 0)) +
  ylab("Contribution") + xlab("Age") +
  theme(legend.key.width = unit(0.3, "cm"),
        legend.key.height = unit(0.2, "cm"))
```
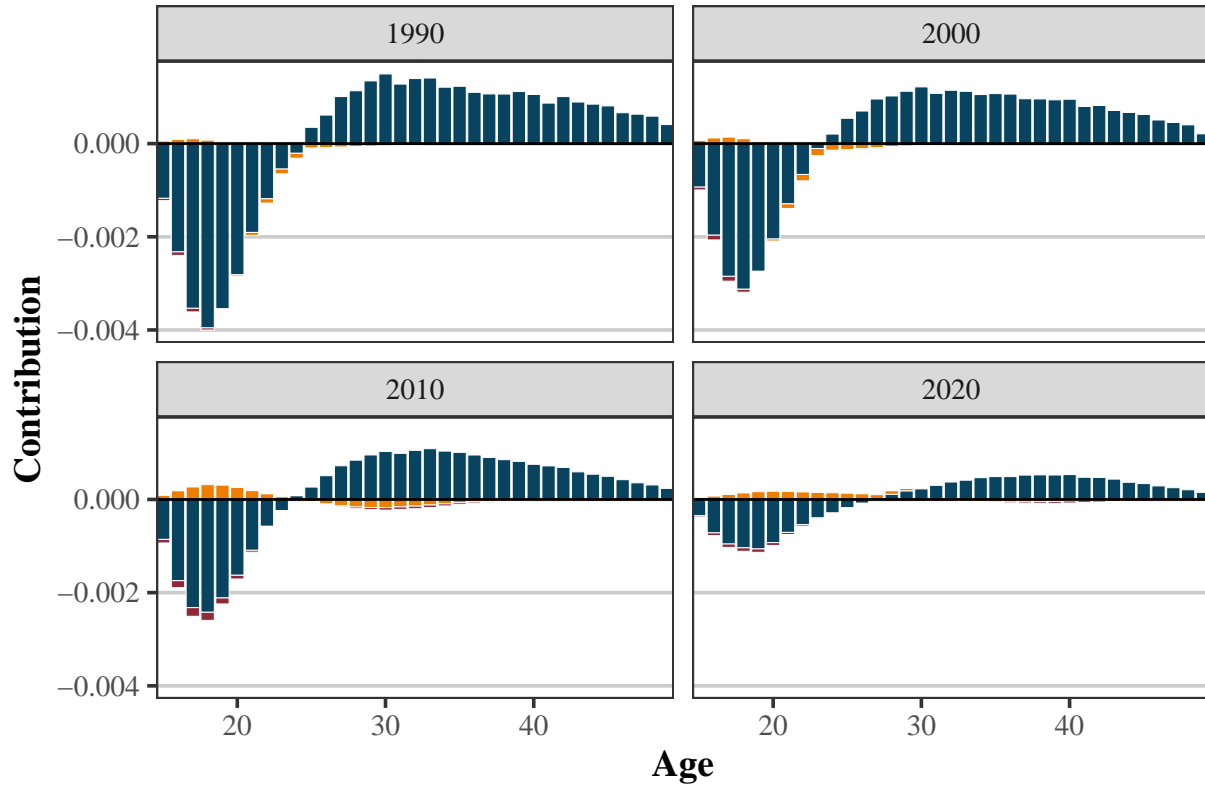
Moreover, since we have the additive components for single years, we can also investigate how the different components changed over time in order to better explain the changes in the gap between male and female fertility rate.

```
# Plot the time trend for populatin estimates
decomposition %>%
  filter(age %in% c(15, 25, 35, 45)) %>%
  ggplot(aes(x = year, contribution, colour = component, group = component)) +
  geom_line(linewidth = 1.2) +
  geom_hline(yintercept = 0, colour = "black") +
  scale_colour_manual(values = c(MPIDRred, MPIDRorange, MPIDRblue),
                      labels = c(expression(paste(Delta, "Population and rate")),
                                 expression(paste(Delta, "Population")),
                                 expression(paste(Delta, "Rate"))),
                      name = "Component: ") +
  scale_y_continuous(labels = scales::label_number_si()) +
  scale_x_continuous(expand = c(0, 0)) +
  facet_wrap( ~ age) +
  theme(legend.key.width = unit(0.3, "cm"),
        legend.key.height = unit(0.2, "cm"),
        axis.text.x = element_text(angle = 45, hjust = 1) )
```
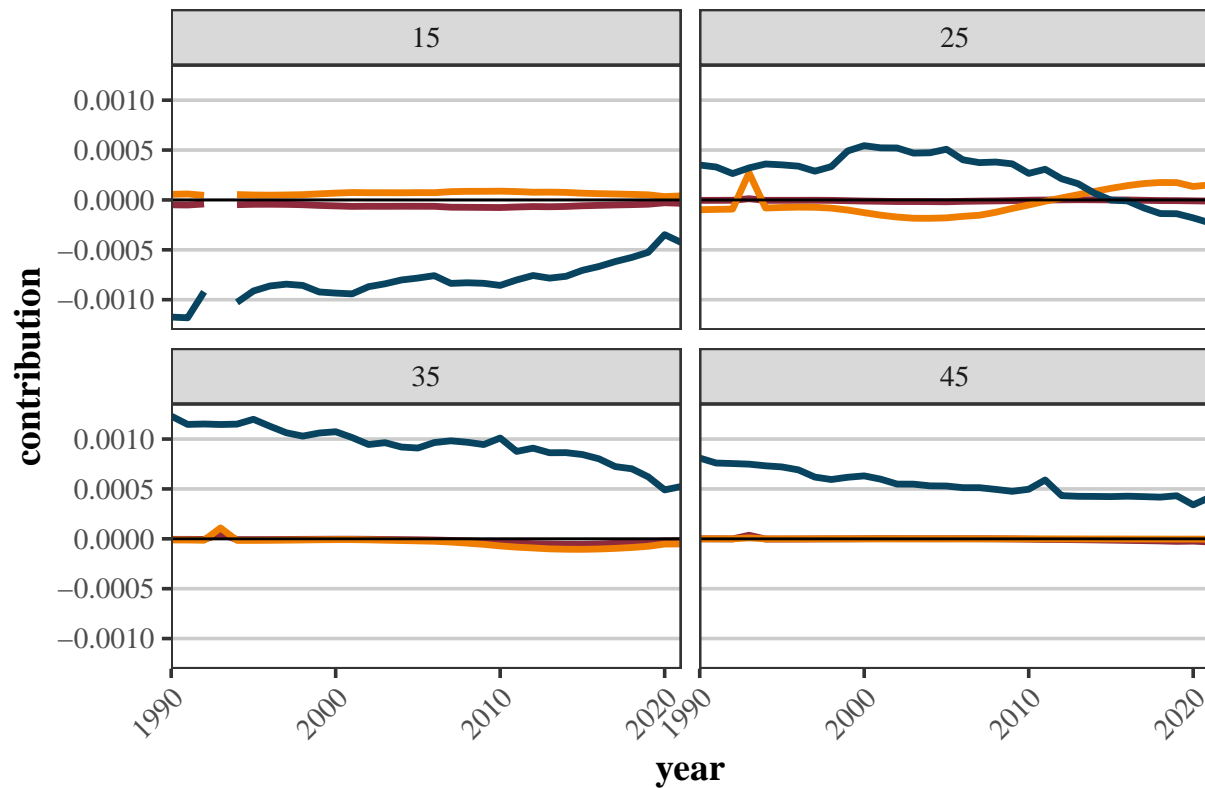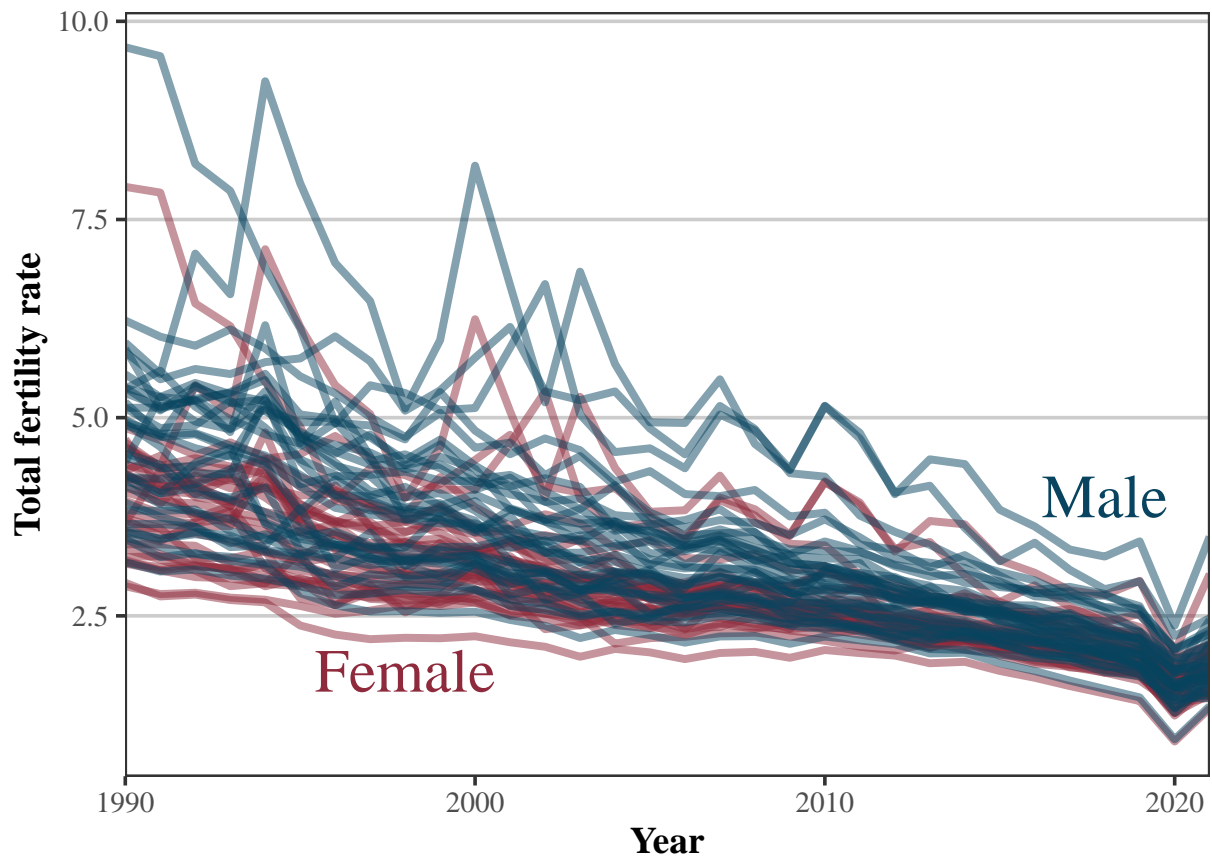
Sub-national level

Now, we are looking at the variation at the sub-national level.

```r
# Load the national level data
load("Data/tfr_regional_mexico.Rda")

# What is the trend in the national tfr s
ggplot(tfr_reg, aes(year)) +
  geom_line(aes(y = tfr_f, colour = "female", group = entity), linewidth = 1.4, alpha = 0.5) +
  geom_line(aes(y = tfr_m, colour = "male", group = entity), linewidth = 1.4, alpha = 0.5) +
  annotate(geom = "text",
           x = 2018, y = 4,
           label = "Male",
           colour = MPIDRblue, size = 8, family = "serif") +
  annotate(geom = "text",
           x = 1998, y = 1.8,
           label = "Female",
           colour = MPIDRred, size = 8, family = "serif") +
  scale_colour_manual(values = c(MPIDRred, MPIDRblue)) +
  scale_x_continuous(expand = c(0, 0)) +
  ylab("Total fertility rate") +
  xlab("Year") +
  guides(colour = "none")
```

```r
# Load the national level data
load("Data/asfr_regional_mexico.Rda")

# What is the trend in the national TFR for women
a <- ggplot(asfr_reg, aes(x = age, alpha = year)) +
  geom_line(aes(y = asfr_f, colour = entity_name, group = interaction(year, entity)),
            linewidth = 1.4) +
  ylab("Age-specific fertilitly rate") +
  xlab("Age") +
  facet_wrap(~ entity_name) +
  guides(alpha = "none", colour = "none")

# What is the trend in the national TFR for women
b <- ggplot(asfr_reg, aes(x = age, alpha = year)) +
  geom_line(aes(y = asfr_m, colour = entity_name, group = interaction(year, entity)),
            linewidth = 1.4) +
  ylab("Age-specific fertilitly rate") +
  xlab("Age") +
  facet_wrap(~ entity_name) +
  guides(alpha = "none", colour = "none")


a / b + plot_annotation(tag_levels = c("1"), tag_suffix = ")")
```

1)

2)
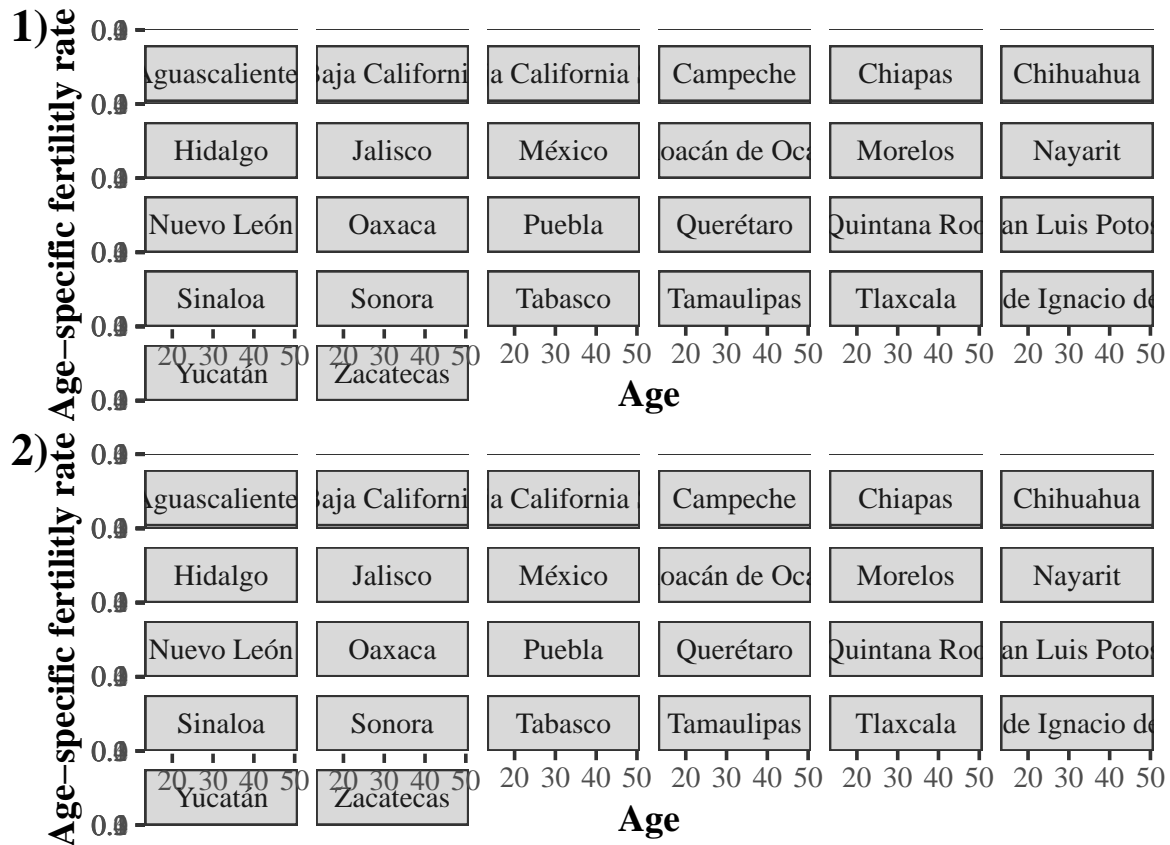
The graphs are overfull. In order to make the graphics comparable, we estimate the male-to-female asfr ratio

```r
# Plot the age-specific fertility rates
asfr_reg %>%
  mutate(asfr_ratio = asfr_m / asfr_f) %>%
  ggplot(aes(x = age, alpha = year)) +
  geom_line(aes(y = asfr_ratio, colour = entity_name, group = interaction(year, entity)),
            linewidth = 1.4) +
  geom_hline(yintercept = 1) +
  ylab("Age-specific fertilitly rate") +
  scale_y_continuous(trans = "log10", labels = scales::label_number_si()) +
  xlab("Age") +
  facet_wrap(~ entity_name) +
  guides(alpha = "none", colour = "none")
```

Age
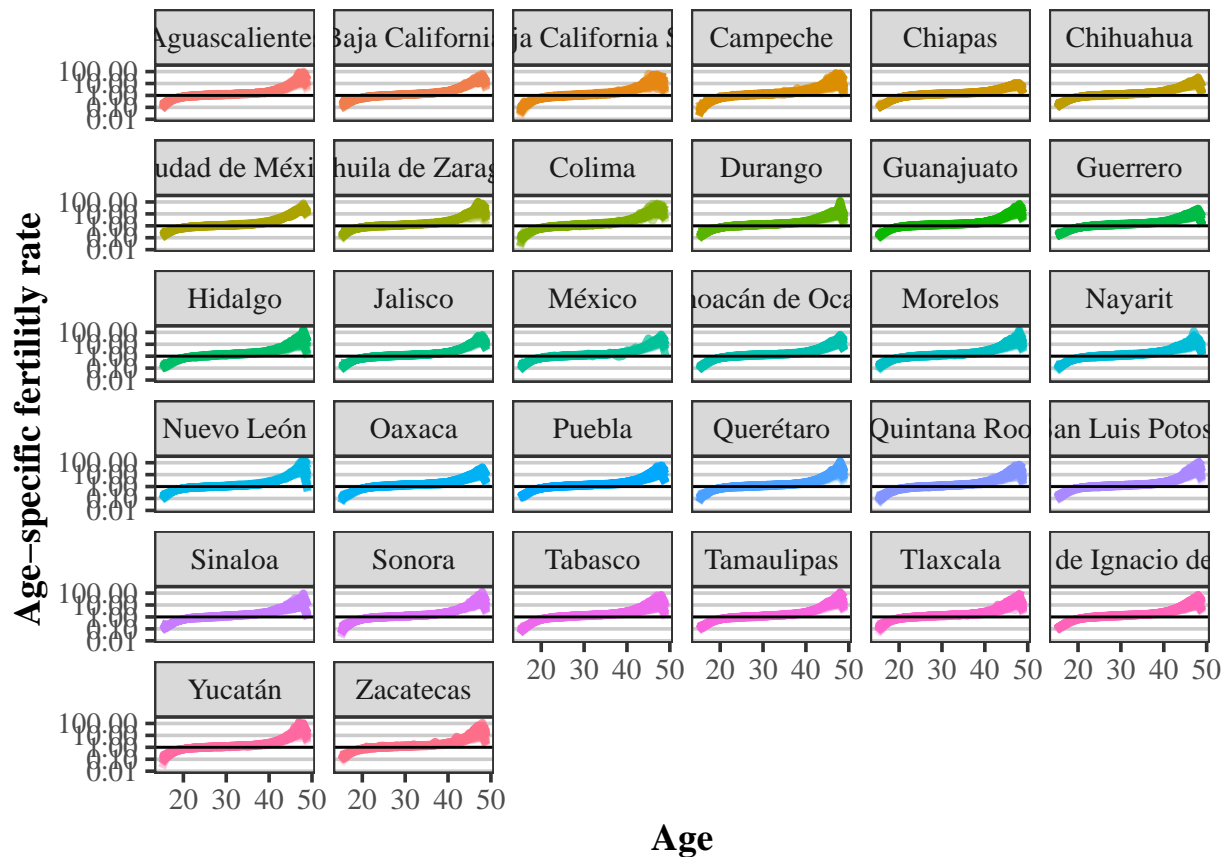
Decomposition of regional differences

```r
# Estimate the decomposition
comp_asfr_reg <- asfr_reg %>%
                group_by(year, entity_name) %>%
                mutate(pop_share_f = pop_share(mid_year_pop_f),
                       pop_share_m = pop_share(mid_year_pop_m),
                       delta_pop = difference(pop_share_f, pop_share_m),
                       delta_rate = difference(asfr_f, asfr_m),
                       change_rate = pop_share_f * delta_rate,
                       change_pop  = asfr_f * delta_pop,
                       change_inter = delta_pop * delta_rate)

# Make the decomposition
decomposition <- comp_asfr_reg %>%
  select(year, age, change_rate,
         change_pop, change_inter, entity_name) %>%
  pivot_longer(cols = starts_with("change"),
               names_to = "component",
               values_to = "contribution")

# Plot the decomposition
decomposition %>%
  filter(year %in% c(1990, 2020) &
         entity_name %in% c("Baja California Sur",
```

```
                        "Guerrero")) %>%
ggplot(aes(age, contribution, fill = component)) +
geom_col(colour = "white", linewidth = 0.01) +
geom_hline(yintercept = 0, colour = "black") +
facet_grid(entity_name ~ year) +
scale_fill_manual(values = c(MPIDRred, MPIDRorange, MPIDRblue),
                  labels = c(expression(paste(Delta, "Population and rate")),
                             expression(paste(Delta, "Population")),
                             expression(paste(Delta, "Rate"))),
                  name = "Component: ") +
scale_x_continuous(expand = c(0, 0)) +
ylab("Contribution") + xlab("Age") +
theme(legend.key.width = unit(0.3, "cm"),
      legend.key.height = unit(0.2, "cm"))
```
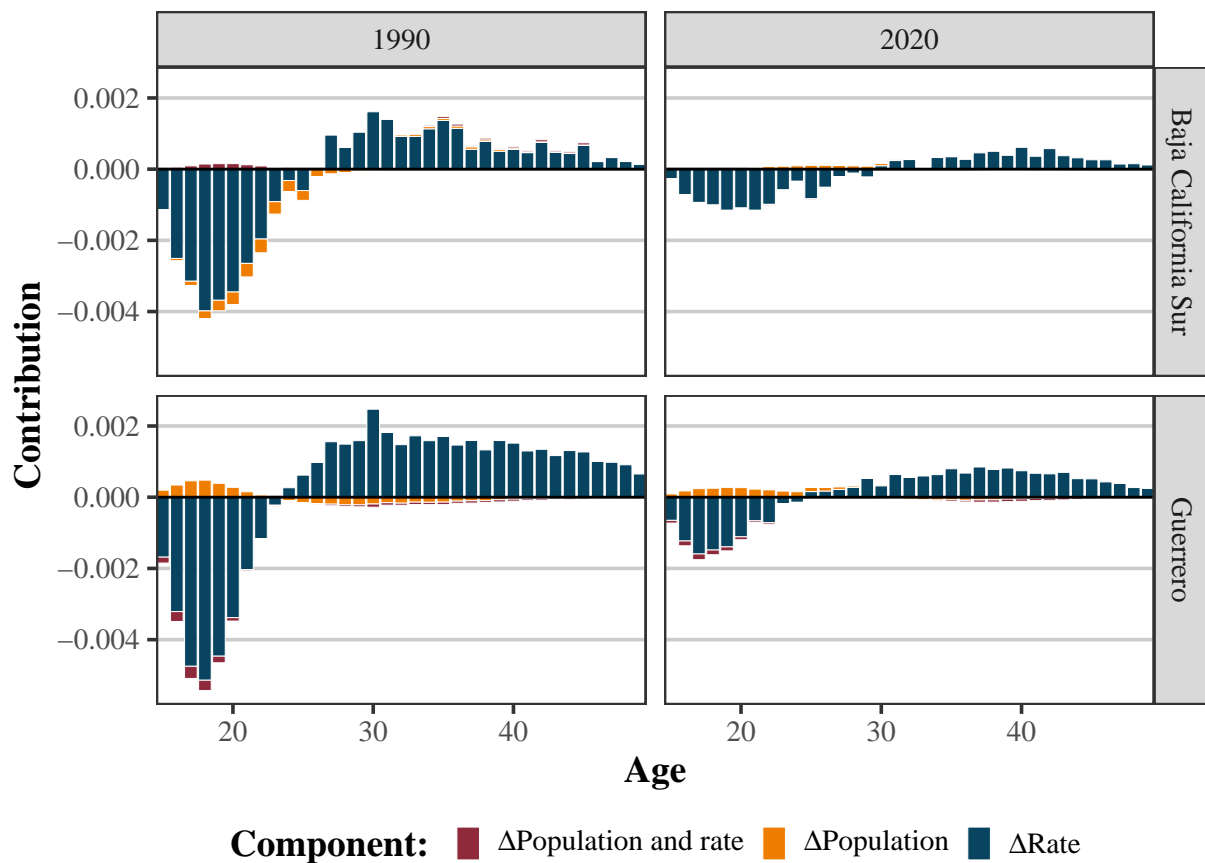

References

Preston, Samuel H., Patrick Heuveline, and Michel Guillot. 2008. *Demography: Measuring and Modeling Population Processes.* 9. [pr.]. Oxford: Blackwell.