US sex specific fetal- and infant mortality. Data quality and relationships

Jonas Schöley

August 18, 2016

This short report looks into possible relationships between fetal- and infant mortality rates, sex differences thereof and the secondary sex ratio.

```
library(dplyr)
library(tidyr)
library(lubridate)
library(ggplot2)

load("../../priv/data/pre-harmonized/us_fdeath_1989-2010.RData")
load("../../priv/data/pre-harmonized/us_ideath_1989-2010.RData")
```

Data preparation

```
# merge data on births, fetal- and infant deaths
fideath <- bind_rows(infant = bind_rows(ideath),
                     fetus = bind_rows(fdeath),
                     .id = "type")
rm(fdeath, ideath)

# prepare
fideath %>%
  mutate(
    date_of_delivery_ym =
      ymd(paste(date_of_delivery_y, date_of_delivery_m, "01", sep = "-")),
    date_of_conception_ym =
      round_date(date_of_delivery_ym - weeks(gestation_at_delivery_w), unit = "month"),
    date_of_conception_y =
      year(date_of_conception_ym),
    death = # death indicator (always TRUE for data from the fetal death file)
      !is.na(age_at_death_d) | type == "fetus"
  ) -> fideath

# summarise
glimpse(fideath)
```

```
## Observations: 78,736,159
## Variables: 10
## $ type                   <chr> "infant", "infant", "infant", "infant"...
## $ date_of_delivery_y     <int> 1989, 1989, 1989, 1989, 1989, 1989, 19...
## $ date_of_delivery_m     <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ gestation_at_delivery_w <int> 38, 39, 37, 41, 38, 39, 39, 41, 41, 35...
## $ sex                    <fctr> Female, Female, Female, Male, Male, M...
```

```
## $ age_at_death_d          <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ date_of_delivery_ym     <date> 1989-01-01, 1989-01-01, 1989-01-01, 1...
## $ date_of_conception_ym   <date> 1988-04-01, 1988-04-01, 1988-05-01, 1...
## $ date_of_conception_y    <dbl> 1988, 1988, 1988, 1988, 1988, 1988, 19...
## $ death                   <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FAL...
```

```
summary(fideath)
```

```
##      type           date_of_delivery_y date_of_delivery_m
##  Length:78736159    Min.   :1989       Min.   : 1.000
##  Class :character   1st Qu.:1996       1st Qu.: 4.000
##  Mode  :character   Median :2001       Median : 7.000
##                     Mean   :2001       Mean   : 6.557
##                     3rd Qu.:2006       3rd Qu.: 9.000
##                     Max.   :2010       Max.   :12.000
##
##  gestation_at_delivery_w     sex           age_at_death_d
##  Min.   : 0              Male  :39995959   Min.   :  0
##  1st Qu.:38              Female:38110722   1st Qu.:  0
##  Median :39              NA's  :  629478   Median :  4
##  Mean   :38                                Mean   : 44
##  3rd Qu.:40                                3rd Qu.: 59
##  Max.   :47                                Max.   :365
##  NA's   :4850227                           NA's   :78185752
##  date_of_delivery_ym  date_of_conception_ym date_of_conception_y
##  Min.   :1989-01-01   Min.   :1988-02-01    Min.   :1988
##  1st Qu.:1996-10-01   1st Qu.:1996-01-01    1st Qu.:1996
##  Median :2001-08-01   Median :2001-05-01    Median :2001
##  Mean   :2001-01-16   Mean   :2000-06-14    Mean   :2000
##  3rd Qu.:2006-05-01   3rd Qu.:2005-11-01    3rd Qu.:2005
##  Max.   :2010-12-01   Max.   :2010-12-01    Max.   :2010
##                       NA's   :4850227       NA's   :4850227
##    death
##  Mode :logical
##  FALSE:76846419
##  TRUE :1889740
##  NA's :0
##
##
##
```

```
# pattern of missing values in sex and gestation at delivery for fetus and infants
fideath %>% group_by(type, date_of_delivery_y, death) %>%
  summarise(gestation_at_delivery_na = sum(is.na(gestation_at_delivery_w)) / n(),
            sex_na = sum(is.na(sex)) / n())
```

```
## Source: local data frame [60 x 5]
## Groups: type, date_of_delivery_y [?]
##
##     type date_of_delivery_y death gestation_at_delivery_na    sex_na
##    <chr>              <int> <lgl>                    <dbl>     <dbl>
## 1  fetus               1989  TRUE               0.06677543 0.4634962
## 2  fetus               1990  TRUE               0.06424342 0.4728900
```

```
## 3  fetus               1991  TRUE              0.06161385 0.4824469
## 4  fetus               1992  TRUE              0.05388487 0.5350280
## 5  fetus               1993  TRUE              0.05047695 0.5562861
## 6  fetus               1994  TRUE              0.04959828 0.5367902
## 7  fetus               1995  TRUE              0.04810828 0.5278297
## 8  fetus               1996  TRUE              0.04975216 0.5434986
## 9  fetus               1997  TRUE              0.04892035 0.5433736
## 10 fetus               1998  TRUE              0.05101044 0.5319525
## # ... with 50 more rows
```

```
# pattern of missing values in sex over gestation for fetal deaths
fideath %>% filter(type == "fetus") %>%
  group_by(gestation_at_delivery_w) %>%
  summarise(sex_na = sum(is.na(sex)) / n())
```

```
## # A tibble: 48 x 2
##    gestation_at_delivery_w    sex_na
##                      <int>     <dbl>
## 1                        0 0.0000000
## 2                        2 0.9186278
## 3                        3 0.9191890
## 4                        4 0.9265504
## 5                        5 0.9227497
## 6                        6 0.9257814
## 7                        7 0.9177231
## 8                        8 0.9156595
## 9                        9 0.9104245
## 10                      10 0.9193855
## # ... with 38 more rows
```

The sex of the dead fetus is not known in about 50% of the cases throughout the years. However, the NAs are solely concentrated prior to a gestation of 20 weeks and sex is always known thereafter, except for cases in which the gestational age is undetermined. For infants the sex is always known. From 1995 until 2002 the gestation at delivery is unknown for about 15% of the infants (no significant difference between those who died and those who survived). In the other periods it is known in more than 99% of the cases. I am not sure about the reason for this pattern.

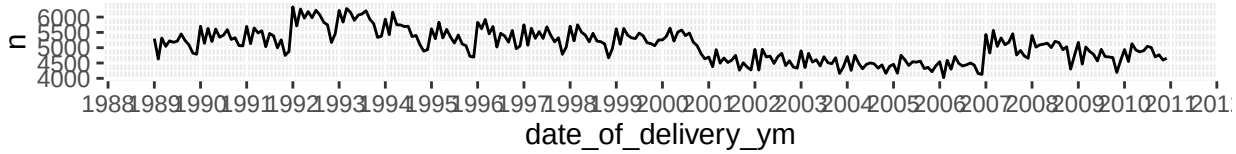
We exclude fetal deaths prior to week 20 from the analysis. We exclude cases where gestation at delivery or sex is unknown.

```
# subset to gestation > 20 weeks,
# remove cases where sex or gestation at delivery is unknown
fideath %>%
  filter(gestation_at_delivery_w >= 20,
         !is.na(sex), !is.na(gestation_at_delivery_w)) -> fideath_sub
```

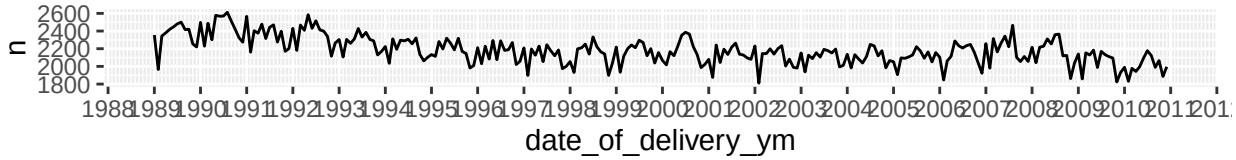
The number of registered fetal deaths varies throughout time with 5 periods clearly segmented by different levels: 1989-1992, 1992-1995, 1995-2001, 2001-2007, 2007-2011. Excluding cases with unknown sex, gestation at delivery and gestation at delivery less than 20 weeks corrects for the segmentation. Time series of births and infant deaths show no segmentation (analysis not shown).

```
# fetal deaths over time complete data
fideath %>%
```

```
  filter(type == "fetus") %>%
  count(date_of_delivery_ym) %>%
  ggplot(aes(x = date_of_delivery_ym, y = n)) +
  geom_line() +
  scale_x_date(date_breaks = "year", date_labels = "%Y", date_minor_breaks = "month") +
  theme(aspect.ratio = 0.07)
```




```
# fetal deaths over time for cases with known sex and gestation
fideath_sub %>%
  filter(type == "fetus") %>%
  count(date_of_delivery_ym) %>%
  ggplot(aes(x = date_of_delivery_ym, y = n)) +
  geom_line() +
  scale_x_date(date_breaks = "year", date_labels = "%Y", date_minor_breaks = "month") +
  theme(aspect.ratio = 0.07)
```



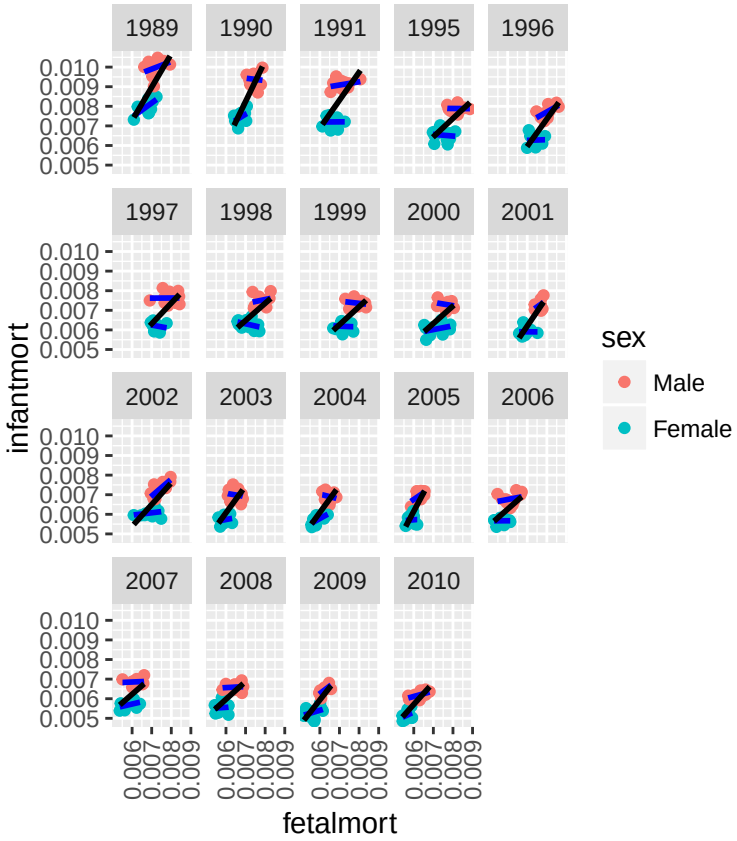
Infant- and fetal period mortality

What are the infant- and fetal mortality rates by sex and period? Following the "National Center of Health Statistics" in the US, we calculate fetal mortality as the number of fetal deaths per fetal deaths and births in period p; and infant mortality as the number of infant deaths per numer of births in period p. Having no information on the number of births or infant deaths in years 1992, 1993, and 1994 we exclude these periods from the calculation.

We can't find any obvious relationship between the levels of fetal- and infant mortality in a given period.

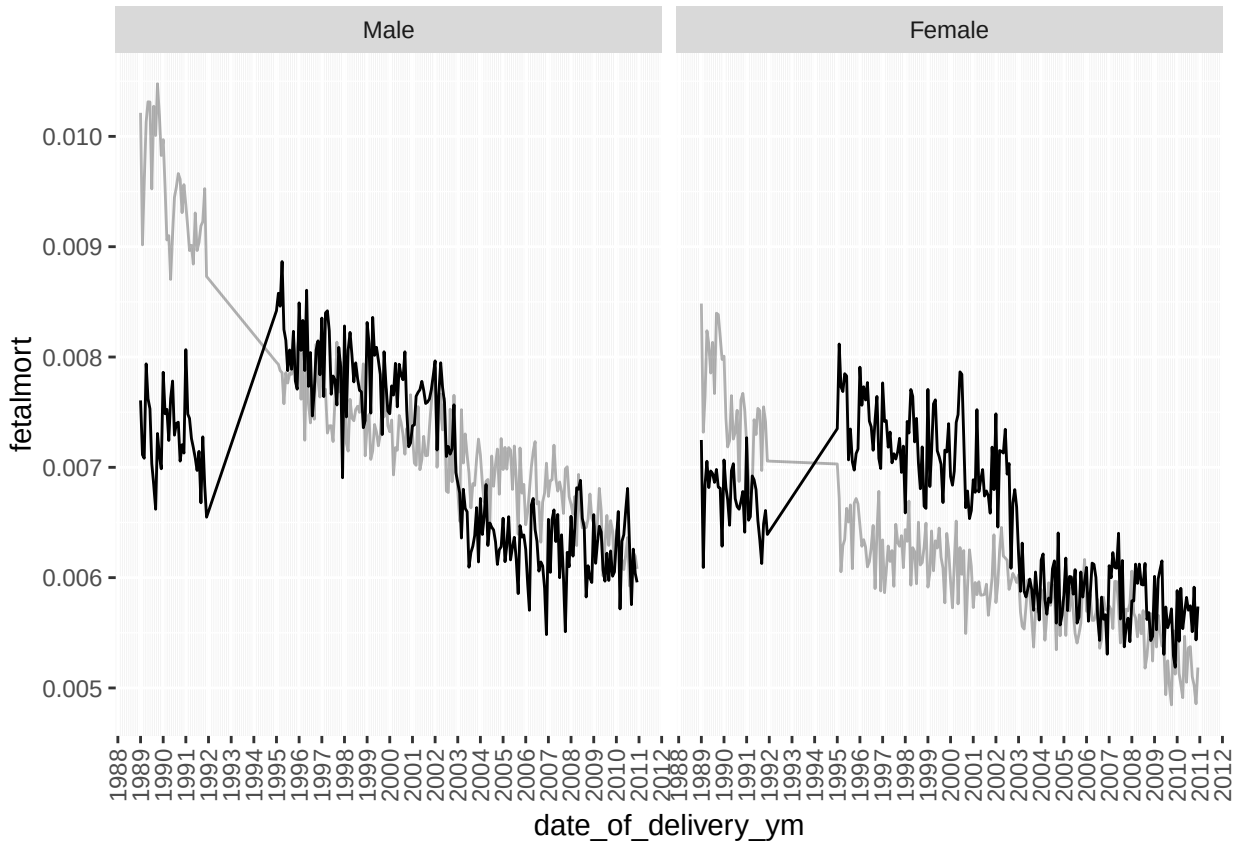
```
# infant and fetal period mortality rates
fideath_sub %>%
  filter(!date_of_delivery_y %in% 1992:1994) %>%
  group_by(date_of_delivery_y, date_of_delivery_ym, sex) %>%
  summarise(
    fetalmort  = sum(type == "fetus") / n(),
    infantmort = sum(type == "infant" & death == TRUE) / sum(type == "infant")
  ) -> fimort_period

# infant versus fetal period mortality rates by sex and period
ggplot(fimort_period, aes(x = fetalmort, y = infantmort, colour = sex)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, colour = "blue", aes(group = sex)) +
  geom_smooth(method = "lm", group = 1, se = FALSE, colour = "black") +
  facet_wrap(~date_of_delivery_y, scales = "fixed") +
```

```
  coord_fixed() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5))
```



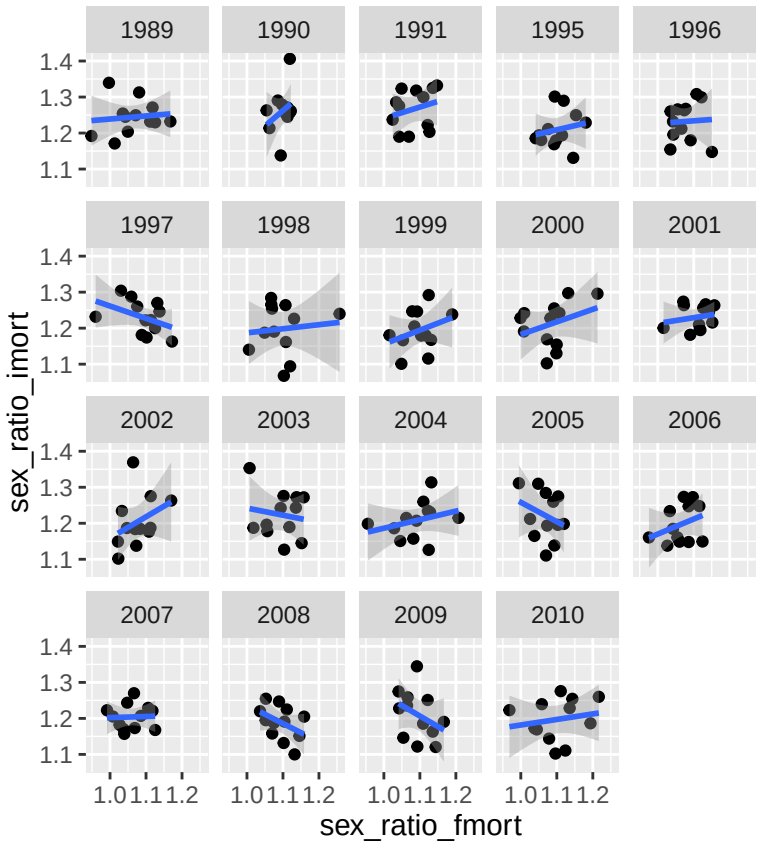
Fetal mortality shows some unusual characteristics. It increased between 1991 and 1995, declining thereafter with a steep decline between 2003 and 2004. The reason for the steep decline is unclear. Infant mortality decreases in a regular fashion.

```
# fetal and infant mortality over time by sex
ggplot(fimort_period,
       aes(x = date_of_delivery_ym)) +
  geom_line(aes(y = fetalmort)) +
  geom_line(aes(y = infantmort), alpha = 0.3) +
  scale_x_date(date_breaks = "year", date_labels = "%Y", date_minor_breaks = "month") +
  facet_wrap(~sex) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5))
```



No obvious relationship between fetal- and infant mortality sex-ratio in a given period can be found.

```
# fetal and infant mortality sex ratio by period
fimort_period %>%
  gather(... = fetalmort:infantmort) %>%
  unite(var, sex, key) %>%
  spread(var, value) %>%
  mutate(sex_ratio_fmort = Male_fetalmort / Female_fetalmort,
         sex_ratio_imort = Male_infantmort / Female_infantmort) -> fimort_period_sex_ratio

# plot fetal and infant mortality sex ratio by period
ggplot(fimort_period_sex_ratio, aes(x = sex_ratio_fmort, y = sex_ratio_imort)) +
  geom_point() +
  geom_smooth(method = "lm") +
  facet_wrap(~date_of_delivery_y) +
  coord_equal()
```



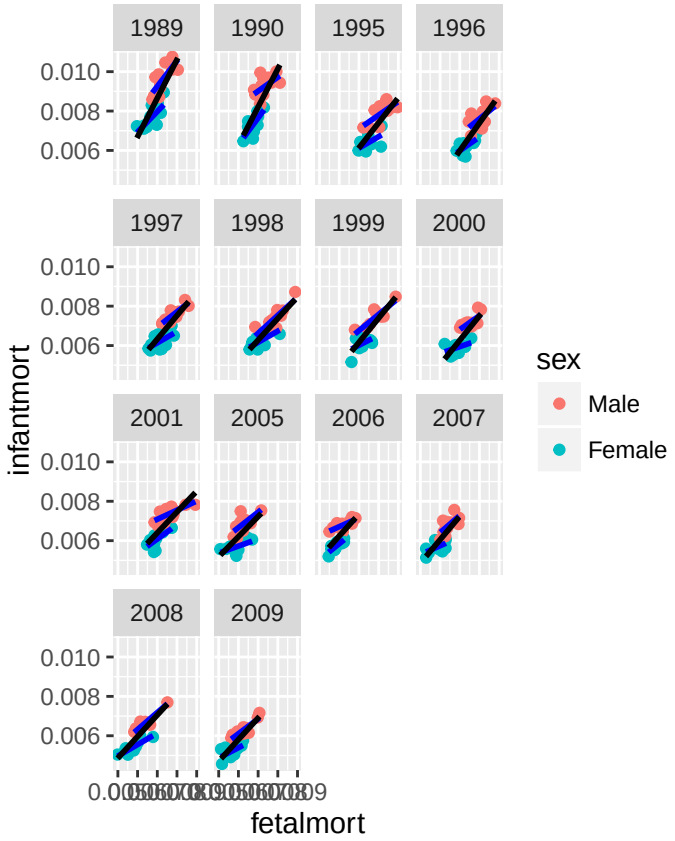
Infant- and fetal conception cohort mortality

We repeat the same analysis, but this time with mortality rates calculated for monthly conception cohorts. The cohort of conception is defined as the month of delivery minus the months of gestation at delivery. Each conception cohort c includes individuals from delivery cohort p and $p+1$. Therefore we leave out conception cohorts which are incomplete due to missing information on births/deaths in the following years.

In opposition to the period perspective, we see a clear correlation between the levels of fetal and infant mortality when comparing the rates on a conception cohort basis. This makes sense as the conception cohorts concept captures similar historical conditions during development while the period concept only captures similar historical conditionsl during death.

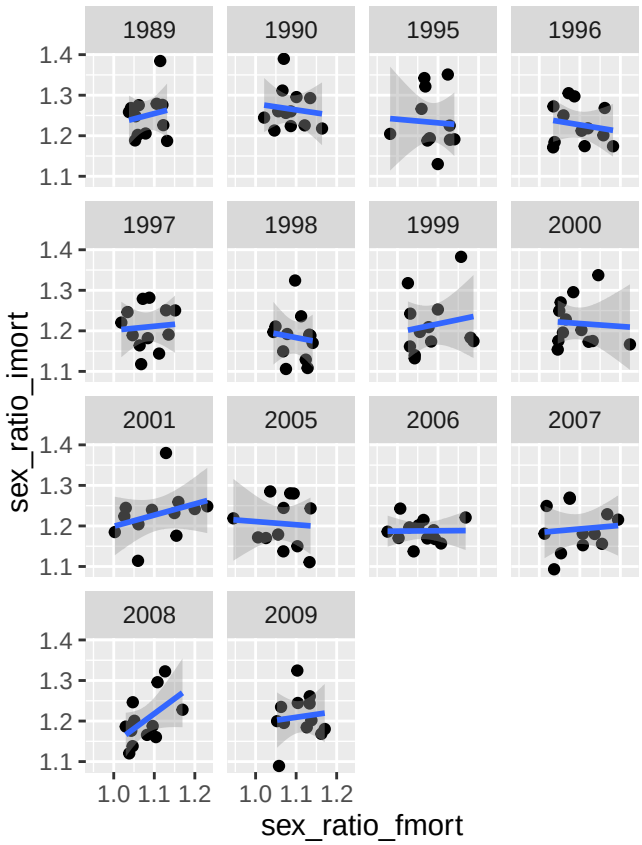
```
# infant and fetal conception cohort mortality rates
fideath_sub %>%
  filter(!date_of_conception_y %in% c(1988, 1991:1994, 2002:2004, 2010)) %>%
  group_by(date_of_conception_y, date_of_conception_ym, sex) %>%
  summarise(
    fetalmort  = sum(type == "fetus") / n(),
    infantmort = sum(type == "infant" & death == TRUE) / sum(type == "infant")
  ) -> fimort_cohort

# infant versus fetal period mortality rates by sex and conception cohort
ggplot(fimort_cohort, aes(x = fetalmort, y = infantmort, colour = sex)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, colour = "blue", aes(group = sex)) +
  geom_smooth(method = "lm", group = 1, se = FALSE, colour = "black") +
```

```
  facet_wrap(~date_of_conception_y, scales = "fixed") +
  coord_fixed()
```



No obvious relationship between fetal- and infant mortality sex-ratio in a given conception cohort can be found.

```
# fetal and infant mortality sex ratio by period
fimort_cohort %>%
  gather(... = fetalmort:infantmort) %>%
  unite(var, sex, key) %>%
  spread(var, value) %>%
  mutate(sex_ratio_fmort = Male_fetalmort / Female_fetalmort,
         sex_ratio_imort = Male_infantmort / Female_infantmort) -> fimort_cohort_sex_ratio

# plot fetal and infant mortality sex ratio by period
ggplot(fimort_cohort_sex_ratio,
       aes(x = sex_ratio_fmort, y = sex_ratio_imort)) +
  geom_point() +
  geom_smooth(method = "lm") +
  coord_equal() +
  facet_wrap(~date_of_conception_y)
```


1989
1990
1995
1996
1997
1998
1999
2000
2001
2005
2006
2007
2008
2009
1.4
1.3
1.2
1.1
1.0 1.1 1.2
sex_ratio_imort
sex_ratio_fmort