Assignment E140

Patrik Lenz and Sebastian Hanika

09/10/2021

Question 1

```
### loading data
url_dta <- "https://www.diw.de/documents/dokumentenarchiv/17/
            diw_01.c.412698.de/soep_lebensz_en.zip" # url for download
download(url = url_dta, dest="dataset.zip", mode="wb") # downloads zip folder into current directory
unzip ("dataset.zip", exdir = ".") # unzips file into current working directory

dt <- read.dta(file ="soep_lebensz_en.dta",  # reads dta file from current directory
               convert.factors = F)
```

Question 1b

```
# number of unique individuals in data set according to id
u_ind <- length(unique(x=dt$id))
print(paste0("There are: ", u_ind, " unique individuals in the data set."),
      quote = F)
```

```
## [1] There are: 3550 unique individuals in the data set.
```

Question 1c

```
# number of observations per year
tab_1c <- dplyr::count(dt, year)
```

The number of observations are presented in the table below.

Year	Observations
2000	3198
2001	2690
2002	2485
2003	2299
2004	2250

Question 1d Sex proportions

```
#filtering data
dt_04 <- dt %>%
  filter(year==max(year)) # finding most recent year

dt_prop <- dt_04 %>% # this is a dplyr solution
  group_by(sex) %>%
  summarise(n = n()) %>%
  mutate(prop = round(n/sum(n), 4),
         prop_per = prop*100)

#pull data for female proportions
female_props <- dt_prop %>%
  filter(sex == 1) %>% # filter for females
  pull(prop_per)       #pulls value from prop_per column

#pull data for male proportion
male_props <- dt_prop %>%
  filter(sex == 0) %>% # filter for males
  pull(prop_per)       # pulls value from prop_per column

print(paste0("Proportion of females is: ", female_props, "%"), quote = F)
```

```
## [1] Proportion of females is: 54.22%
```

```
print(paste0("Proportion of males is: ", male_props, "%"), quote = F)
```

```
## [1] Proportion of males is: 45.78%
```

Question 1d Average perceived health

```
# calculating average health
dt_health <- dt_04 %>%
  select(sex, health_org) %>% # selecting relevant columns
  group_by(sex) %>%
  summarise(avg_health = round(mean(health_org), digits = 2)) #calculating

#data for female health
female_health <- dt_health %>%
  filter(sex == 1) %>% # filter for females
  pull(avg_health)     # pulls value from dataframe

#data for male healt
male_health <- dt_health %>%
  filter(sex == 0) %>% # filter for males
  pull(avg_health)     # pulls value from dataframe

print(paste0("The averge subjective health of females is: ", female_health), quote = F)
```

```
## [1] The averge subjective health of females is: 3.39
```

```
print(paste0("The averge subjective health of males is: ", male_health), quote = F)
```

```
## [1] The averge subjective health of males is: 3.46
```

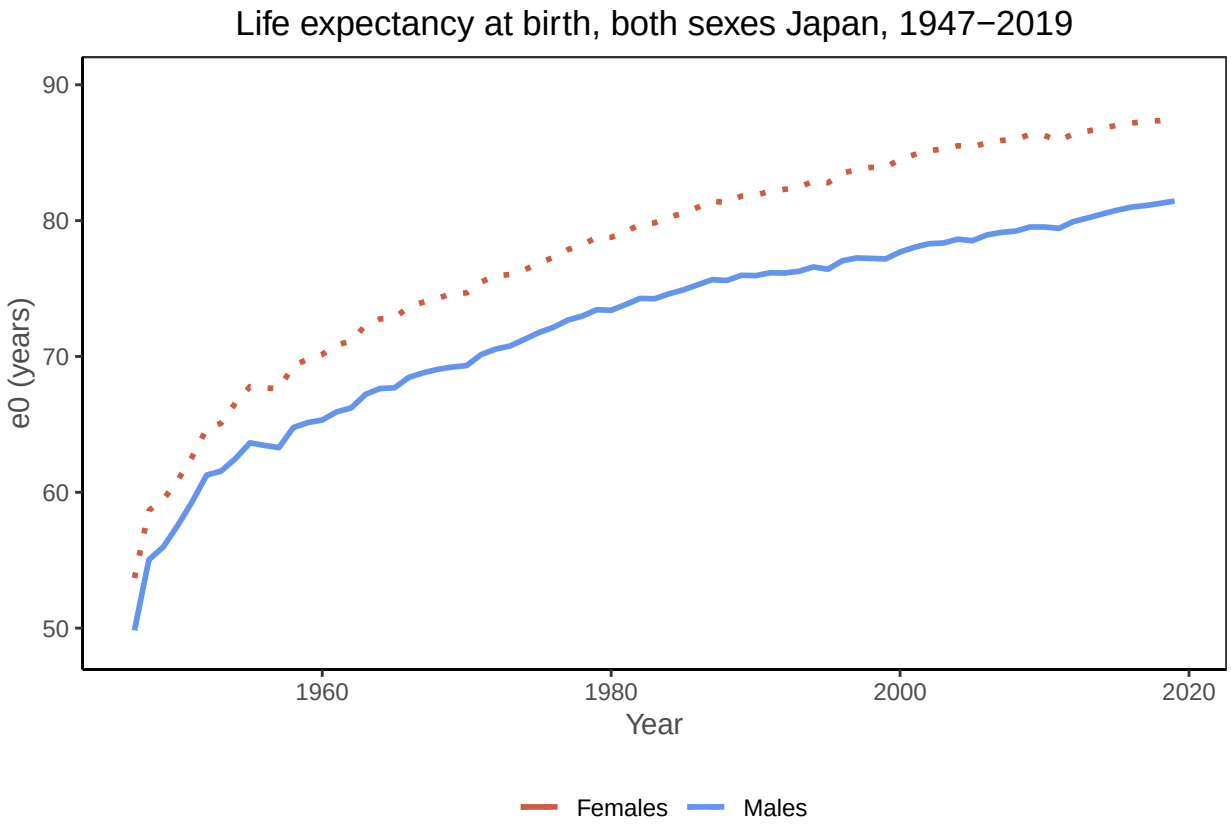
Therefore subjective health is higher for males than females.

Question 2

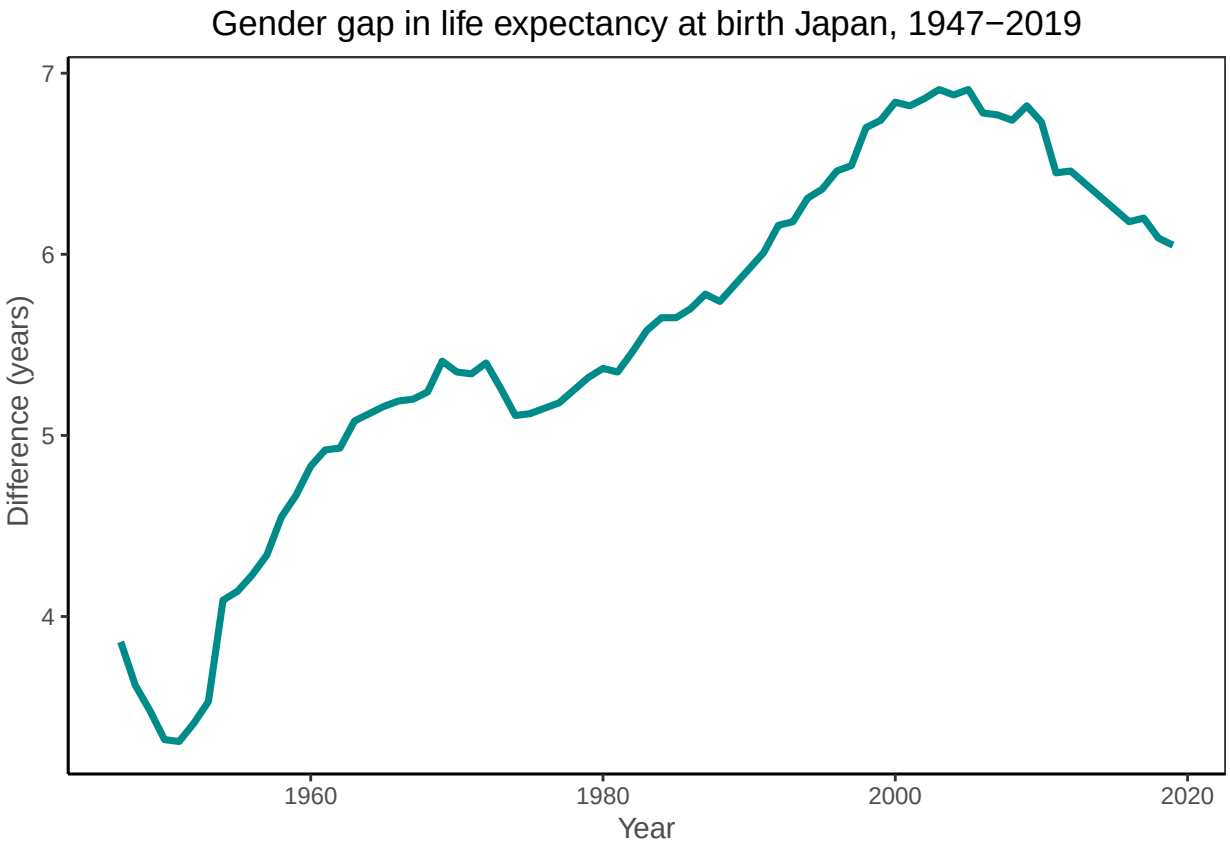
```
# loading data from Human Mortality Database
# Japan,life expectancy at birth by sex

myHMDusername <- "Enter username"
myHMDpassword <- "Enter password"
```

```
jpn <-readHMDweb("JPN","E0per",
                     username = myHMDusername,
                     password = myHMDpassword,
                     fixup = T)
```

```
# visualization of the trend in life expectancy at birth

#specifying colors for graphs
col <- c("Females" = "coral3", "Males" = "cornflowerblue")

fig1 <- jpn %>%
  ggplot(aes(x=Year)) +
  geom_line(aes(y=Female, color = "Females"), size = 1, linetype = 3) +
  geom_line(aes(y=Male, color = "Males"), size = 1) +
  scale_color_manual(values = col) +
  labs(title = "Life expectancy at birth, both sexes Japan, 1947-2019",
       x = "Year",
       y = "e0 (years)",
       color = "") +
  scale_y_continuous(limits = c(49, 90)) +
  theme_bw() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.line = element_line(colour = "black"),
        axis.title = element_text(size=11, color = "grey30"),
        legend.position = "bottom",
        plot.title = element_text(size = 13, hjust = 0.5))

fig1
```




```
# calculating the gender gap
jpn <- jpn %>%
  mutate(diff_e0 = Female - Male)

# visualization of the gender gap in life expectancy at birth
fig2 <- jpn %>%
  ggplot(aes(x = Year, y = diff_e0)) +
  geom_line(color = "darkcyan", size = 1.25) +
  labs(title = "Gender gap in life expectancy at birth Japan, 1947-2019",
       x = "Year",
       y = " Difference (years)") +
  theme_bw() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.line = element_line(colour = "black"),
        axis.title = element_text(size=11, color = "grey30"),
        legend.position = "bottom",
        plot.title = element_text(size = 13, hjust = 0.5))


fig2
```


Gender gap in life expectancy at birth Japan, 1947−2019
Difference (years)
7
6
5
4
1960
1980
2000
2020
Year