Visualizing Inequalities in Life Expectancy

Elvin

2024-01-16

1. United Nations life expectancy data

```
# This sets plot images to a nice size
options(repr.plot.width = 6, repr.plot.height = 6)

# Loading packages
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
library(ggplot2)
library(readr)

# Loading data
life_expectancy <- read_csv("UNdata.csv", show_col_types = FALSE)

# Taking a look at the first few rows
head(life_expectancy,10)
```

```
## # A tibble: 10 x 7
##    `Country or Area` Subgroup Year      Source     Unit  Value `Value Footnotes`
##    <chr>             <chr>    <chr>     <chr>      <chr> <dbl>             <dbl>
##  1 Afghanistan       Female   2000-2005 UNPD_Worl~ Years    42                NA
##  2 Afghanistan       Female   1995-2000 UNPD_Worl~ Years    42                NA
##  3 Afghanistan       Female   1990-1995 UNPD_Worl~ Years    42                NA
##  4 Afghanistan       Female   1985-1990 UNPD_Worl~ Years    41                NA
##  5 Afghanistan       Male     2000-2005 UNPD_Worl~ Years    42                NA
##  6 Afghanistan       Male     1995-2000 UNPD_Worl~ Years    42                NA
##  7 Afghanistan       Male     1990-1995 UNPD_Worl~ Years    42                NA
##  8 Afghanistan       Male     1985-1990 UNPD_Worl~ Years    41                NA
##  9 Albania           Female   2000-2005 UNPD_Worl~ Years    79                NA
## 10 Albania           Female   1995-2000 UNPD_Worl~ Years    76                NA
```

2. Life expectancy of men vs. women by country

Let's manipulate the data to make our exploration easier. We will build the dataset for our first plot in which we will represent the average life expectancy of men and women across countries for the last period recorded in our data (2000-2005).

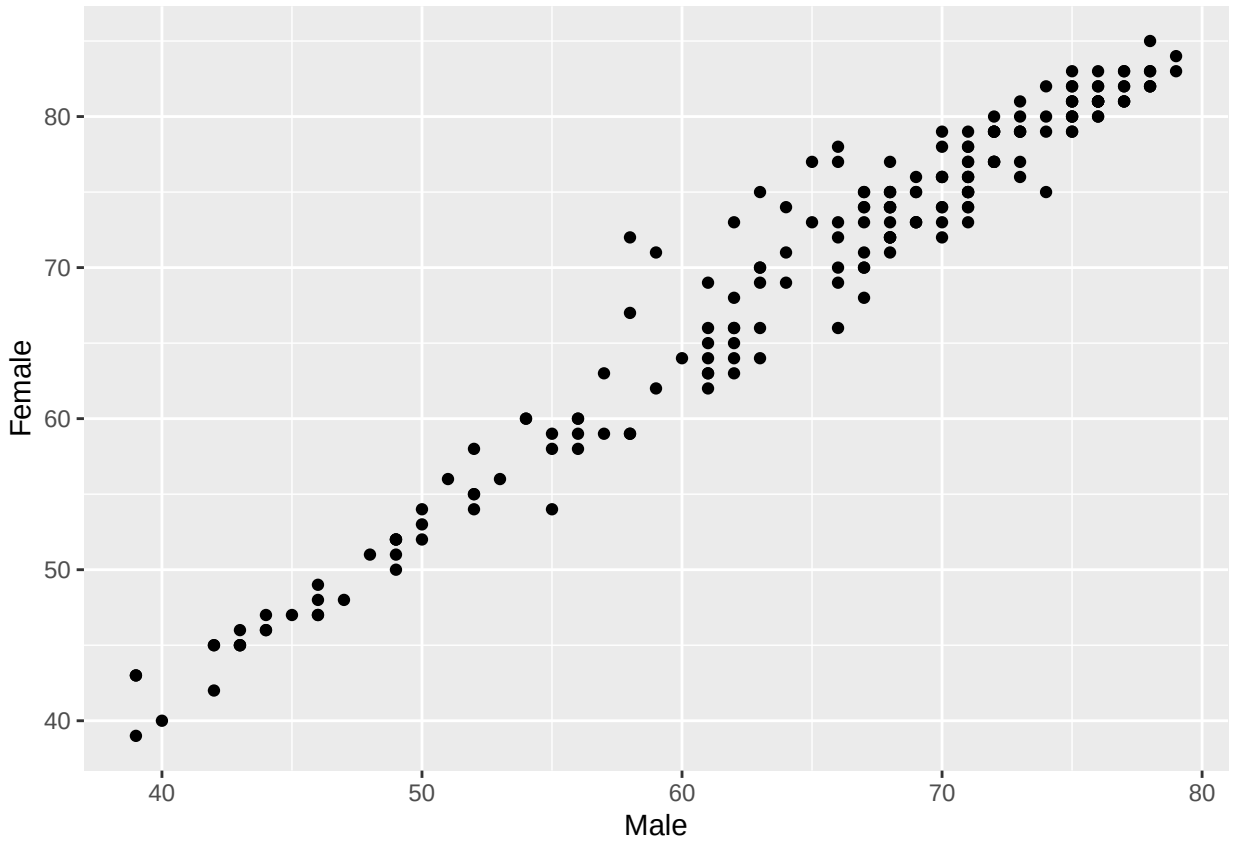
```
# Subsetting and reshaping the life expectancy data
subdata <- life_expectancy  %>%
filter(Year=="2000-2005") %>%
select(`Country or Area`, Subgroup, Value) %>%
spread(Subgroup, Value)

# Taking a look at the first few rows
head(subdata,10)
```

```
## # A tibble: 10 x 3
##    `Country or Area` Female  Male
##    <chr>              <dbl> <dbl>
##  1 Afghanistan           42    42
##  2 Albania               79    73
##  3 Algeria               72    70
##  4 Angola                43    39
##  5 Argentina             78    71
##  6 Armenia               75    68
##  7 Aruba                 76    70
##  8 Australia             83    78
##  9 Austria               82    76
## 10 Azerbaijan            70    63
```

3. Visualize I

A scatter plot is a useful way to visualize the relationship between two variables. It is a simple plot in which points are arranged on two axes, each of which represents one of those variables. Let's create a scatter plot using ggplot2 to represent life expectancy of males (on the x-axis) against females (on the y-axis). We will create a straightforward plot in this task, without many details. We will take care of these kinds of things shortly.

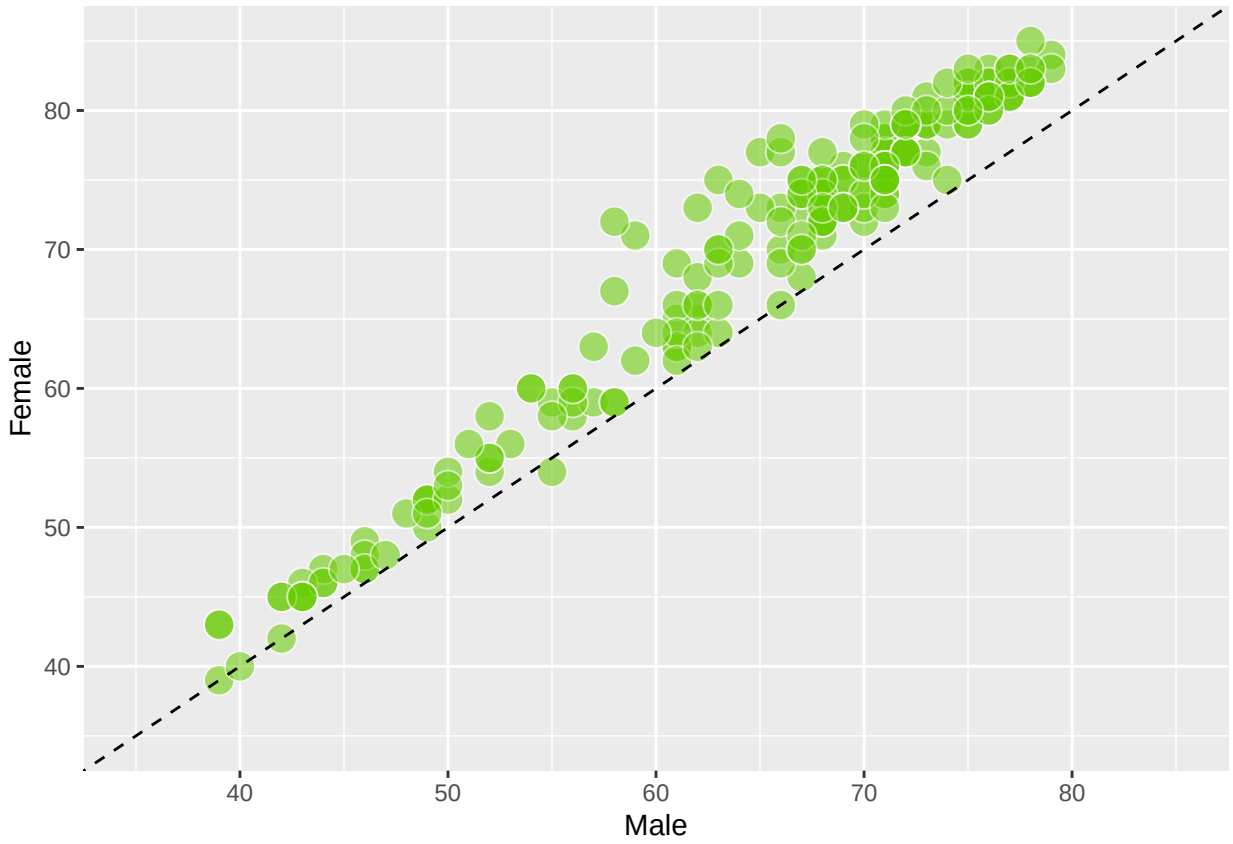
```
# Plotting male and female life expectancy
ggplot(subdata, aes(x = Male, y = Female)) + geom_point()
```



4. Reference lines I

A good plot must be easy to understand. There are many tools in ggplot2 to achieve this goal and we will explore some of them now. Starting from the previous plot, let's set the same limits for both axes as well as place a diagonal line for reference. After doing this, the difference between men and women across countries will be easier to interpret. After completing this task, we will see how most of the points are arranged above the diagonal and how there is a significant dispersion among them. What does this all mean?

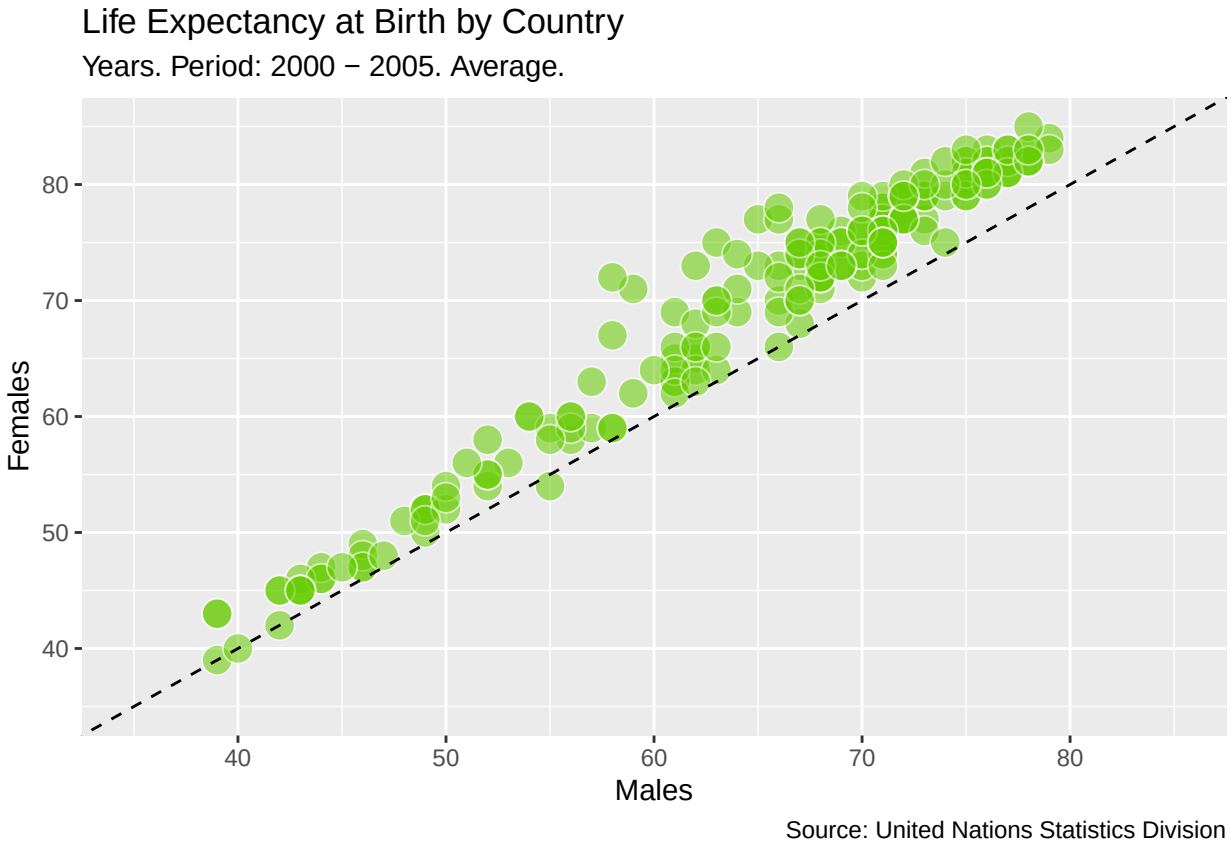
```
ggplot(subdata, aes(x = Male, y = Female)) +
  geom_point(color ="white", fill = "chartreuse3", shape = 21, alpha = 0.55, size =5) +
  geom_abline(intercept = 0, slope = 1, linetype = 2)+
  scale_x_continuous(limits = c(35,85)) +
  scale_y_continuous(limits = c(35,85))
```



5. Plot titles and axis labels

A key point to make a plot understandable is placing clear labels on it. Let's add titles, axis labels, and a caption to refer to the source of data. Let's also change the appearance to make it clearer.

```
ggplot(subdata, aes(x = Male, y = Female)) +
  geom_point(color ="white", fill = "chartreuse3", shape = 21, alpha = 0.55, size =5) +
  geom_abline(intercept = 0, slope = 1, linetype = 2)+
  scale_x_continuous(limits = c(35,85)) +
  scale_y_continuous(limits = c(35,85))+
  labs(title = "Life Expectancy at Birth by Country",
       subtitle = "Years. Period: 2000 - 2005. Average.",
       caption="Source: United Nations Statistics Division",
       x="Males",
       y="Females")
```

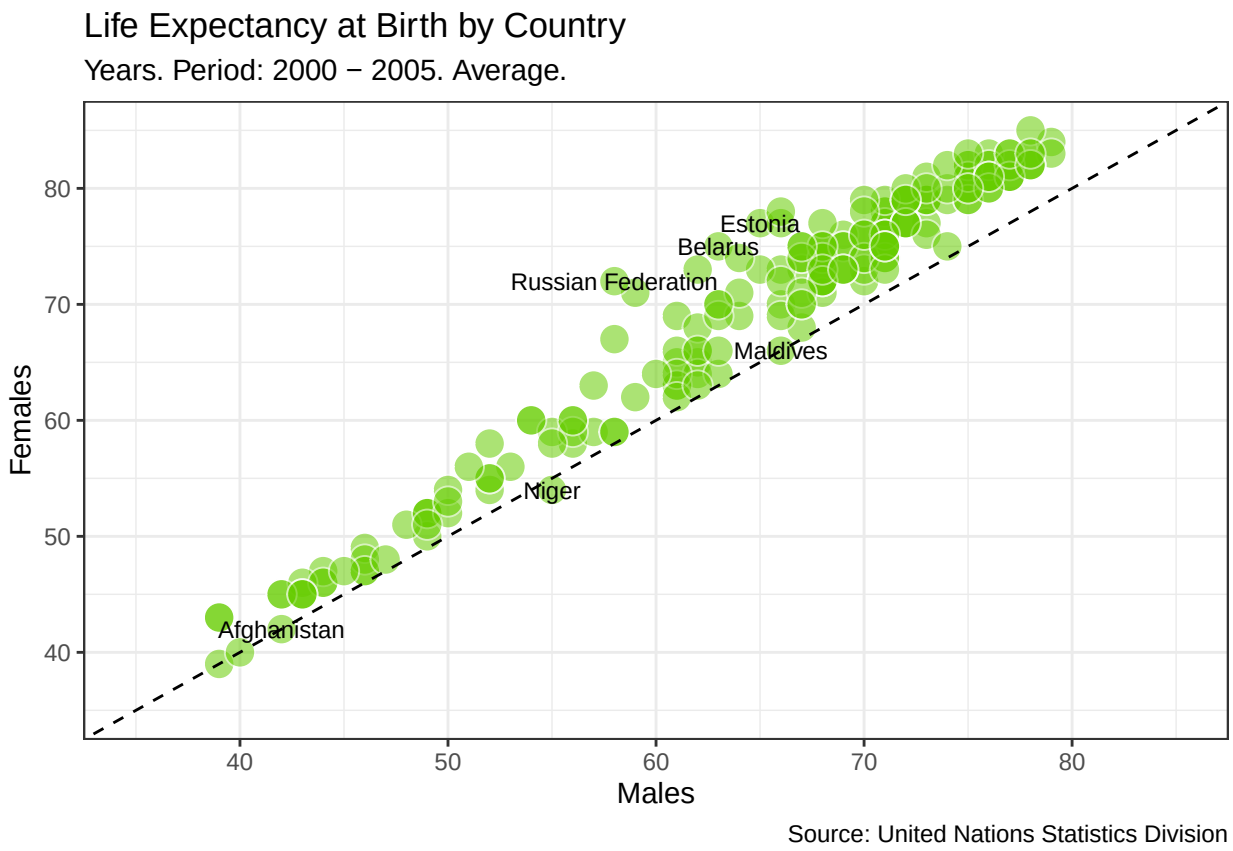


6. Highlighting remarkable countries I

Now, we will label some points of our plot with the name of its corresponding country. We want to draw attention to some special countries where the gap in life expectancy between men and women is significantly high. These will be the final touches on this first plot.

```
top_male <- subdata %>%
  arrange(Male - Female) %>%
  head(3)
top_female <- subdata %>%
  arrange(Female - Male)%>%
  head(3)

ggplot(subdata, aes(x = Male, y = Female)) +
  geom_point(color ="white", fill = "chartreuse3", shape = 21, alpha = 0.55, size =5) +
  geom_abline(intercept = 0, slope = 1, linetype = 2)+
  scale_x_continuous(limits = c(35,85)) +
  scale_y_continuous(limits = c(35,85))+
  labs(title = "Life Expectancy at Birth by Country",
       subtitle = "Years. Period: 2000 - 2005. Average.",
       caption="Source: United Nations Statistics Division",
       x="Males",
       y="Females")+
  geom_text(data = top_male, aes(label= `Country or Area`), size = 3) +
  geom_text(data = top_female, aes(label= `Country or Area`), size = 3) +
```

```
  theme_bw()
```



7. How has life expectancy by gender evolved?

Since our data contains historical information, let's see now how life expectancy has evolved in recent years. Our second plot will represent the difference between men and women across countries between two periods: 2000-2005 and 1985-1990. Let's start building a dataset called subdata2 for our second plot.

```
subdata2 <- life_expectancy %>%
  filter(Year %in% c("1985-1990","2000-2005")) %>%
  mutate(Sub_Year = paste(Subgroup, Year, sep = "_")) %>%
  mutate(Sub_Year = gsub("-","_", Sub_Year)) %>%
  select(-Subgroup,-Year) %>%
  spread(Sub_Year, Value) %>%
  mutate(diff_Female = Female_2000_2005 - Female_1985_1990,
         diff_Male = Male_2000_2005 - Male_1985_1990)

  head(subdata2)
```

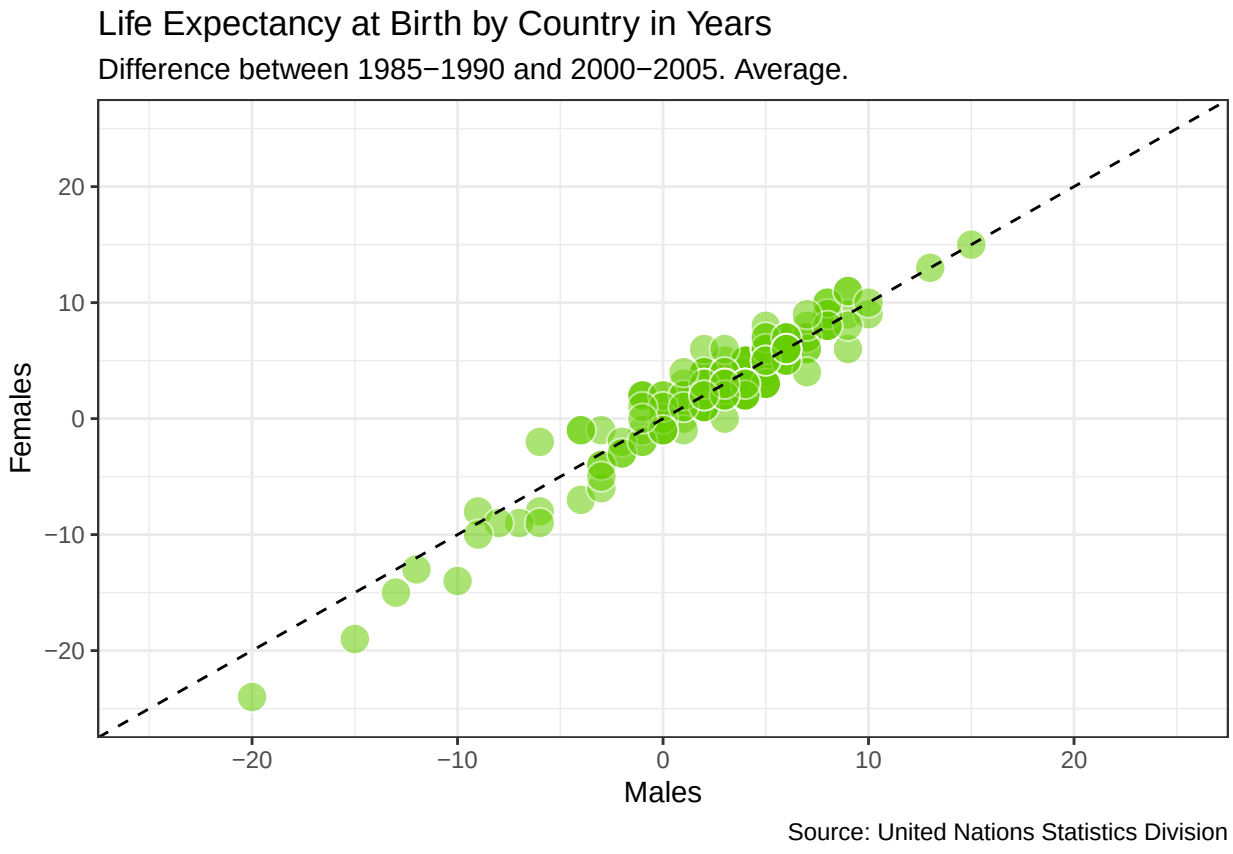
```
## # A tibble: 6 x 10
##   `Country or Area` Source              Unit  `Value Footnotes` Female_1985_1990
##   <chr>             <chr>               <chr>              <dbl>             <dbl>
## 1 Afghanistan       UNPD_World Populat~ Years                 NA                41
## 2 Albania           UNPD_World Populat~ Years                 NA                75
## 3 Algeria           UNPD_World Populat~ Years                 NA                67
```

```
## 4 Angola            UNPD_World Populat~ Years                NA              42
## 5 Argentina         UNPD_World Populat~ Years                NA              75
## 6 Armenia           UNPD_World Populat~ Years                NA              71
## # i 5 more variables: Female_2000_2005 <dbl>, Male_1985_1990 <dbl>,
## #   Male_2000_2005 <dbl>, diff_Female <dbl>, diff_Male <dbl>
```

8. Visualize II

Now let's create our second plot in which we will represent average life expectancy differences between "1985-1990" and "2000-2005" for men and women.

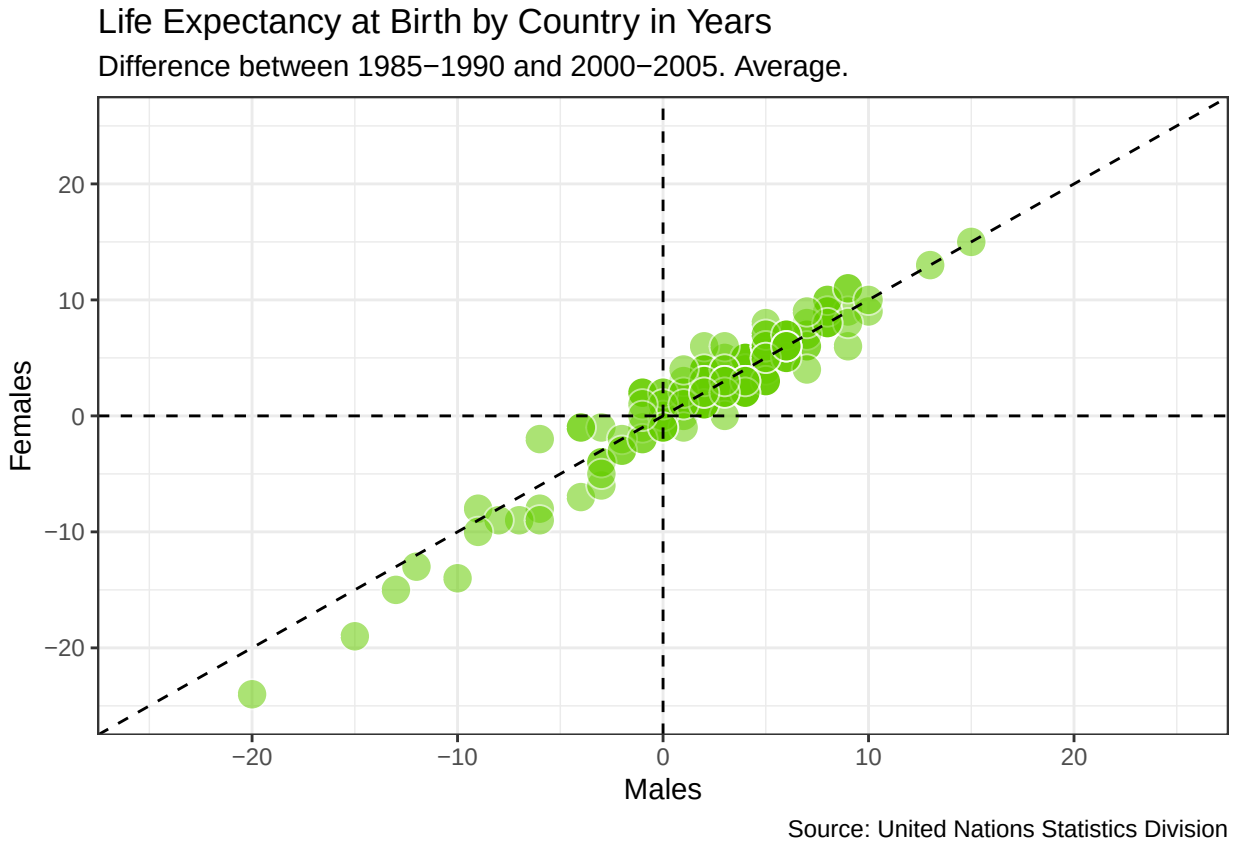
```
ggplot(subdata2, aes(x = diff_Male, y = diff_Female, label = "Country or Area")) +
  geom_point(color ="white", fill = "chartreuse3", shape = 21, alpha = 0.55, size =5) +
  geom_abline(intercept = 0, slope = 1, linetype = 2)+
  scale_x_continuous(limits = c(-25,25)) +
  scale_y_continuous(limits = c(-25,25))+
  labs(title = "Life Expectancy at Birth by Country in Years",
       subtitle = "Difference between 1985-1990 and 2000-2005. Average.",
       caption="Source: United Nations Statistics Division",
       x="Males",
       y="Females")+
  theme_bw()
```



9. Reference lines II

Adding reference lines can make plots easier to understand. We already added a diagonal line to visualize differences between men and women more clearly. Now we will add two more lines to help to identify in which countries people increased or decreased their life expectancy in the period analyzed.

```
ggplot(subdata2, aes(x = diff_Male, y = diff_Female, label = "Country or Area")) +
  geom_point(color ="white", fill = "chartreuse3", shape = 21, alpha = 0.55, size =5) +
  geom_abline(intercept = 0, slope = 1, linetype = 2)+
  scale_x_continuous(limits = c(-25,25)) +
  scale_y_continuous(limits = c(-25,25)) +
  geom_hline(yintercept = 0, xintercept = 0, linetype = 2) +
  geom_vline(yintercept = 0, xintercept = 0, linetype = 2) +
  labs(title = "Life Expectancy at Birth by Country in Years",
       subtitle = "Difference between 1985-1990 and 2000-2005. Average.",
       caption="Source: United Nations Statistics Division",
       x="Males",
       y="Females") +
  theme_bw()
```



10. Highlighting remarkable countries II

As we did in the first plot, let's label some points. Concretely, we will point those three where the aggregated average life expectancy for men and women increased most and those three where decreased most in the period.

```
top <- subdata2 %>% arrange(diff_Male+diff_Female) %>% head(3)
bottom <- subdata2 %>% arrange(-(diff_Male+diff_Female)) %>% head (3)

ggplot(subdata2, aes(x = diff_Male, y = diff_Female, label = `Country or Area`)) +
  geom_point(color ="white", fill = "chartreuse3", shape = 21, alpha = 0.55, size =5) +
  geom_abline(intercept = 0, slope = 1, linetype = 2) +
  scale_x_continuous(limits = c(-25,25)) +
  scale_y_continuous(limits = c(-25,25)) +
  geom_hline(yintercept = 0, linetype = 2) +
  geom_vline(xintercept = 0, linetype = 2) +
  labs(title = "Life Expectancy at Birth by Country in Years",
       subtitle = "Difference between 1985-1990 and 2000-2005. Average.",
       caption="Source: United Nations Statistics Division",
       x="Males",
       y="Females") +
  geom_text(data = top, aes(label = `Country or Area`), size = 3) +
  geom_text(data = bottom, aes(label = `Country or Area`), size = 3) +
  theme_bw()
```