

Visualizing Inequalities in Life Expectancy

Datacamp Project Solution

03 jul, 2021

1 United Nations life expectancy data

Life expectancy at birth is a measure of the average a living being is expected to live. It takes into account several demographic factors like gender, country, or year of birth.

Life expectancy at birth can vary along time or between countries because of many causes: the evolution of medicine, the degree of development of countries, or the effect of armed conflicts. Life expectancy varies between gender, as well. The data shows that women live longer than men. Why? Several potential factors, including biological reasons and the theory that women tend to be more health conscious.

Let's create some plots to explore the inequalities about life expectancy at birth around the world. We will use a dataset from the United Nations Statistics Division, which is available [here](#).

```
# Loading packages
```

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

```
# Loading data
```

```
life_expectancy <- tibble(read.csv("datasets/UNdata.csv"))
```

```
# Taking a look at the first few rows
```

```
head(life_expectancy)
```

```
## # A tibble: 6 x 7
```

##	Country.or.Area	Subgroup	Year	Source	Unit	Value	Value.Footnotes
##	<chr>	<chr>	<chr>	<chr>	<chr>	<int>	<int>
## 1	Afghanistan	Female	2000--	UNPD_World Popula~	Years	42	NA
## 2	Afghanistan	Female	1995--	UNPD_World Popula~	Years	42	NA
## 3	Afghanistan	Female	1990--	UNPD_World Popula~	Years	42	NA
## 4	Afghanistan	Female	1985--	UNPD_World Popula~	Years	41	NA
## 5	Afghanistan	Male	2000--	UNPD_World Popula~	Years	42	NA
## 6	Afghanistan	Male	1995--	UNPD_World Popula~	Years	42	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) %>%
  pivot_wider(names_from = Subgroup, values_from = Value)
```

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

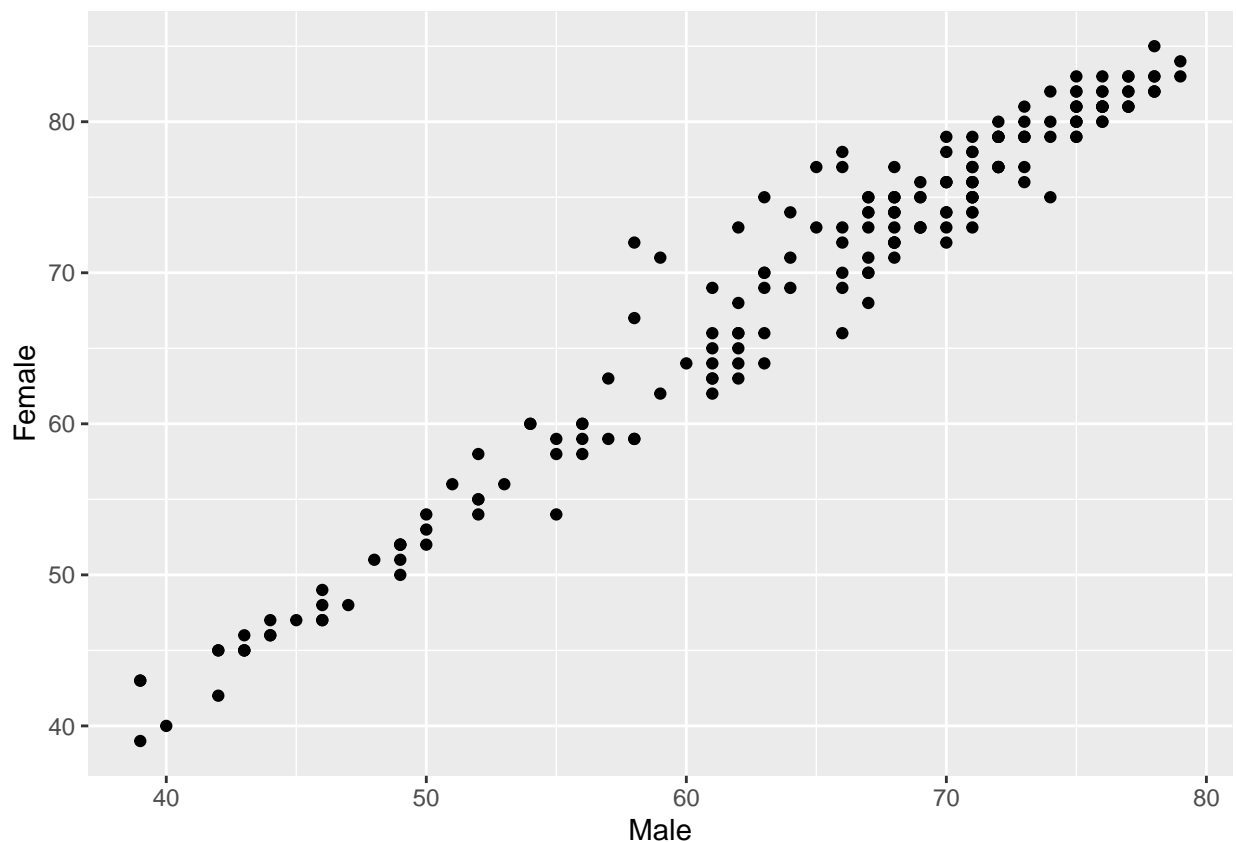
```
## # A tibble: 6 x 3  
##   Country.or.Area Female  Male  
##   <chr>          <int> <int>  
## 1 Afghanistan      42    42  
## 2 Albania          79    73  
## 3 Algeria          72    70  
## 4 Angola           43    39  
## 5 Argentina        78    71  
## 6 Armenia          75    68
```

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()
```



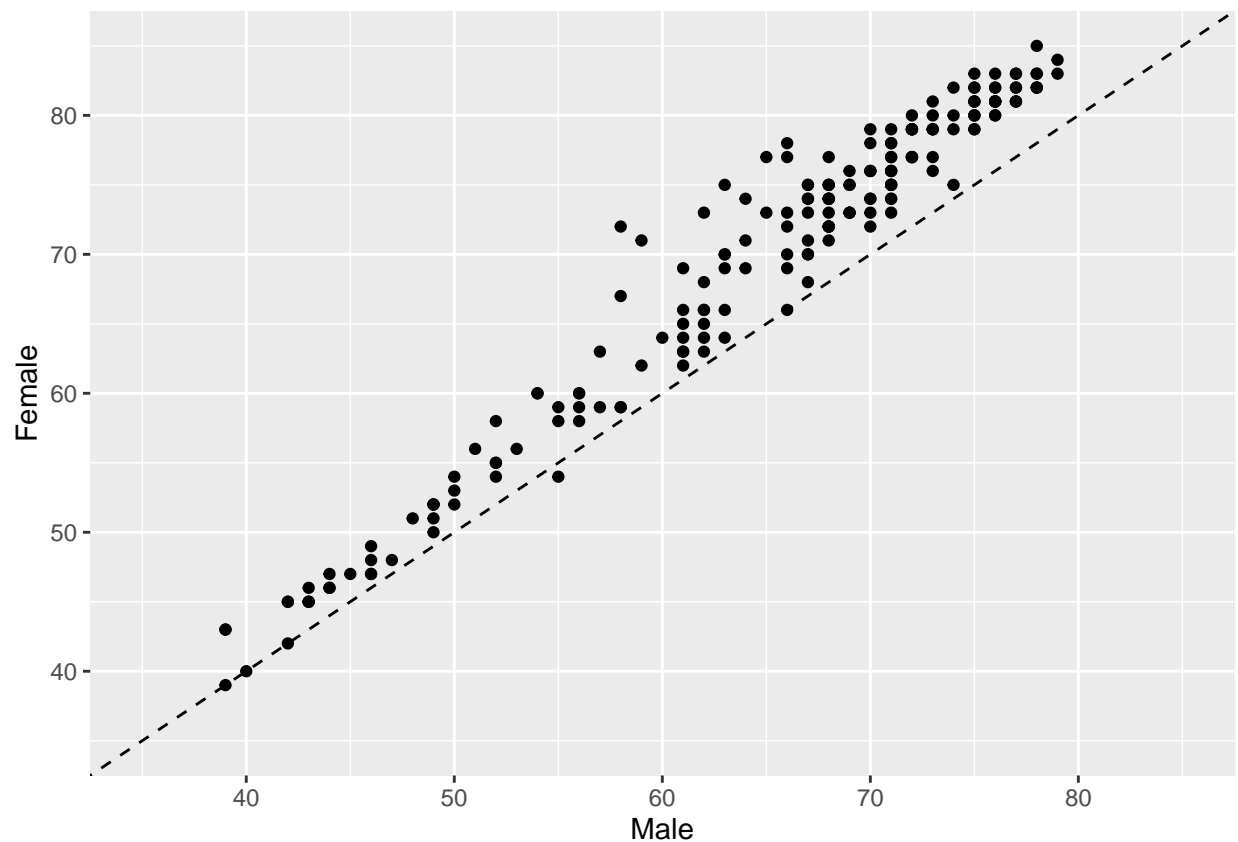
4 Reference lines I

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?

Adding an abline and changing the scale of axes of the previous plots

```
ggplot(subdata) +  
  aes(x = Male, y = Female) +  
  geom_point() +  
  geom_abline(intercept = 0, slope = 1, linetype = "dashed") +  
  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.

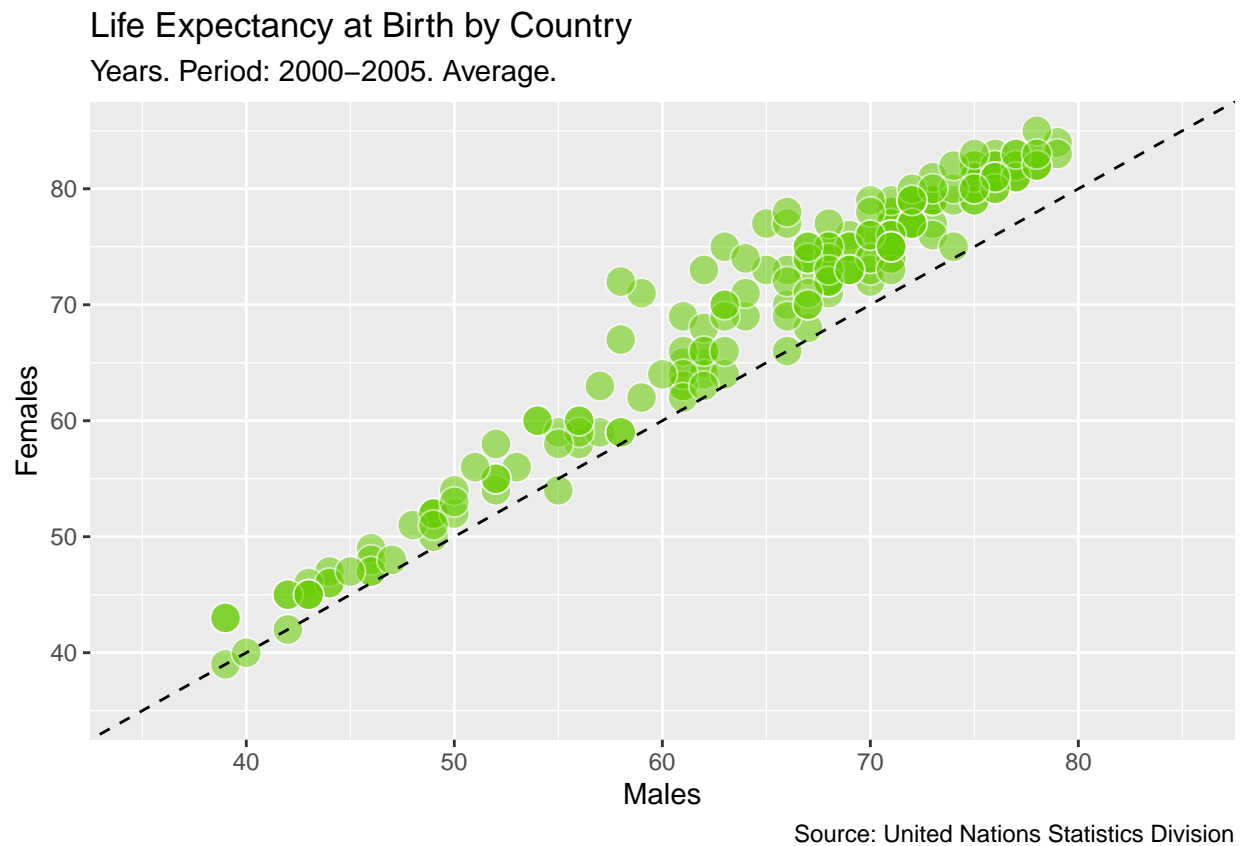
Adding labels to previous plot

```
ggplot(subdata) +  
  aes(x = Male, y = Female) +
```

```

geom_point(
  colour = "white",
  fill = "chartreuse3",
  shape = 21,
  alpha = .55,
  size = 5
) +
geom_abline(intercept = 0, slope = 1, linetype = "dashed") +
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.

```
# Subsetting data to obtain countries of interest
```

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

```
top_male
```

```
## # A tibble: 3 x 3
##   Country.or.Area   Female   Male
##   <chr>             <int> <int>
## 1 Russian Federation    72    58
## 2 Belarus              75    63
## 3 Estonia              77    65
```

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

```
top_female
```

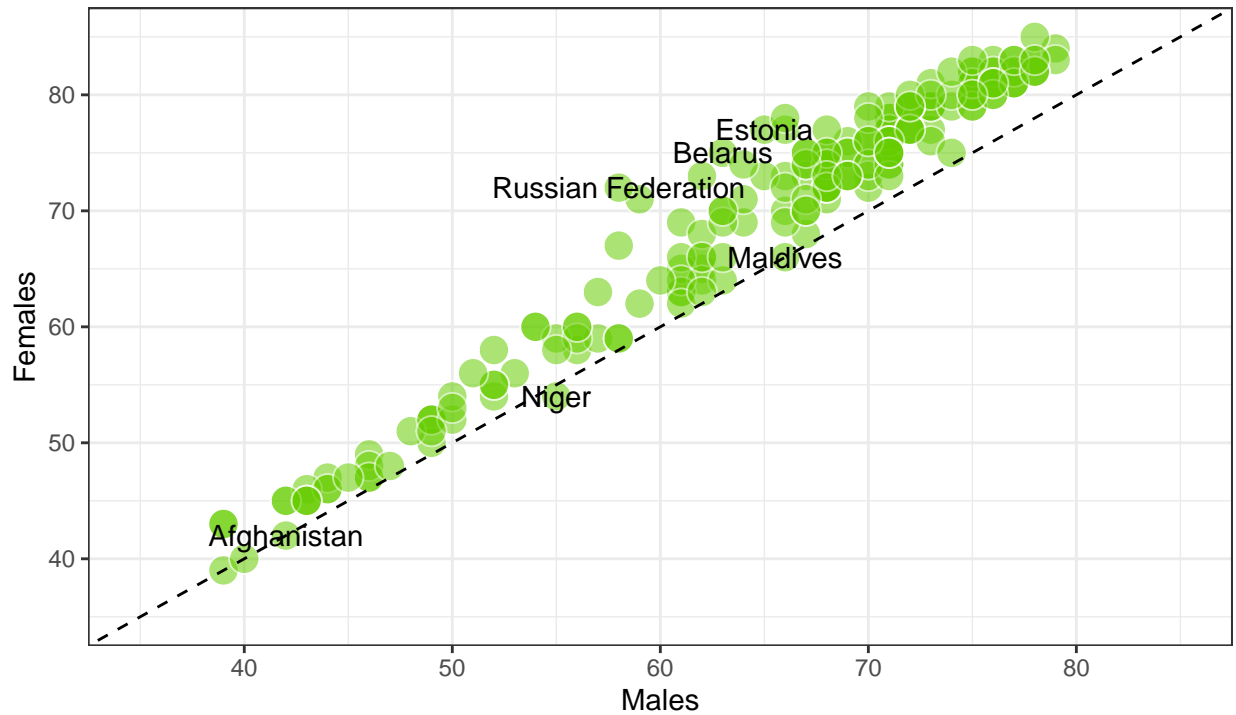
```
## # A tibble: 3 x 3
##   Country.or.Area   Female   Male
##   <chr>             <int> <int>
## 1 Niger             54     55
## 2 Afghanistan      42     42
## 3 Maldives         66     66
```

```
# Adding text to the previous plot to label countries of interest
```

```
ggplot(subdata) +
  aes(x = Male, y = Female, label = Country.or.Area) +
  geom_point(
    colour = "white",
    fill = "chartreuse3",
    shape = 21,
    alpha = .55,
    size = 5
  ) +
  geom_abline(intercept = 0, slope = 1, linetype = "dashed") +
  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(top_male, mapping = aes(label = Country.or.Area)) +
  geom_text(top_female, mapping = aes(label = Country.or.Area)) +
  theme_bw()
```

Life Expectancy at Birth by Country

Years. Period: 2000–2005. Average.



Source: United Nations Statistics Division

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.

```
# Subsetting, mutating and reshaping the life expectancy data
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) %>%
  pivot_wider(names_from = Sub_Year, values_from = Value) %>%
  mutate(
    diff_Female = Female_2000_2005 - Female_1985_1990,
    diff_Male = Male_2000_2005 - Male_1985_1990
  )

# Taking a look at the first few rows
head(subdata2)
```

```
## # A tibble: 6 x 10
##   Country.or.Area Source Unit   Value.Footnotes Female_2000_2005 Female_1985_1990
##   <chr>           <chr> <chr>         <int>         <int>         <int>
## 1 Afghanistan    UNPD_~ Years          NA           42           41
```

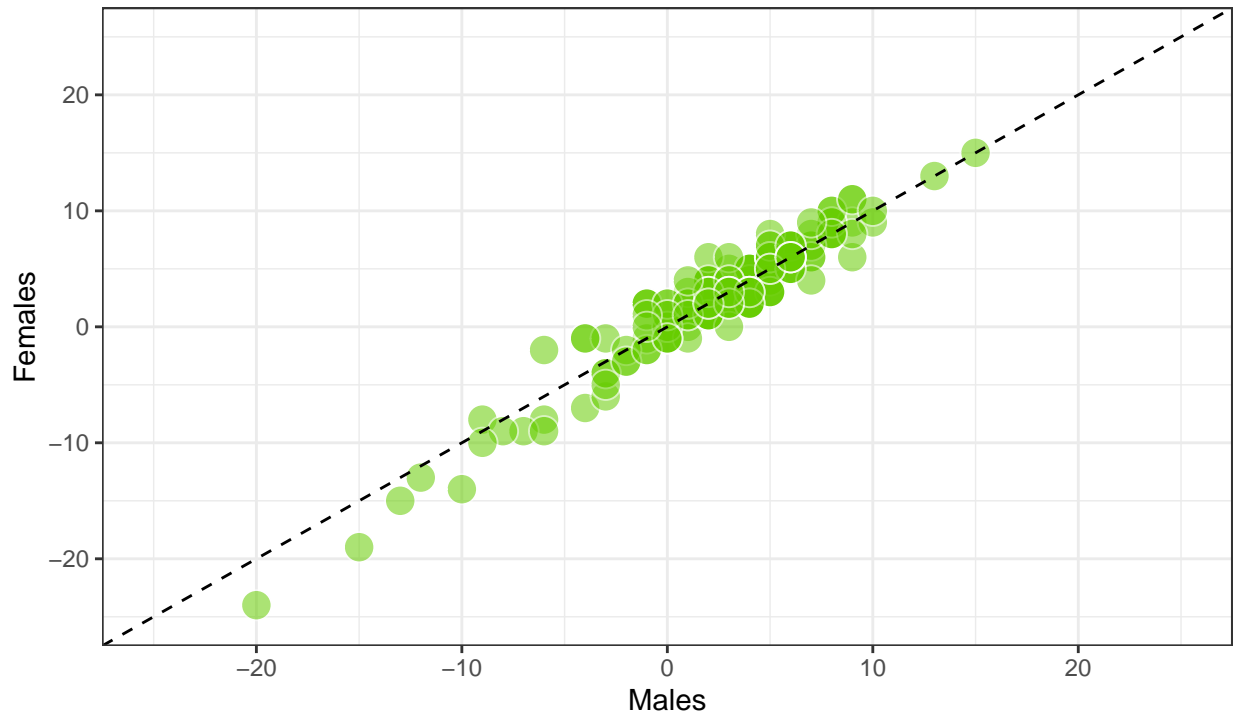
```
## 2 Albania      UNPD_~ Years      NA      79      75
## 3 Algeria      UNPD_~ Years      NA      72      67
## 4 Angola       UNPD_~ Years      NA      43      42
## 5 Argentina    UNPD_~ Years      NA      78      75
## 6 Armenia      UNPD_~ Years      NA      75      71
## # ... with 4 more variables: Male_2000_2005 <int>, Male_1985_1990 <int>,
## #   diff_Female <int>, diff_Male <int>
```

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.

```
# Doing a nice first version of the plot with abline, scaling axis and adding labels
ggplot(subdata2) +
  aes(x = diff_Male, y = diff_Female, label = Country.or.Area) +
  geom_point(
    colour = "white",
    fill = "chartreuse3",
    shape = 21,
    alpha = .55,
    size = 5
  ) +
  geom_abline(intercept = 0, slope = 1, linetype = "dashed") +
  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()
```

Life Expectancy at Birth by Country in Years
 Difference between 1985–1990 and 2000–2005. Average.



Source: United Nations Statistics Division

9 Reference lines II

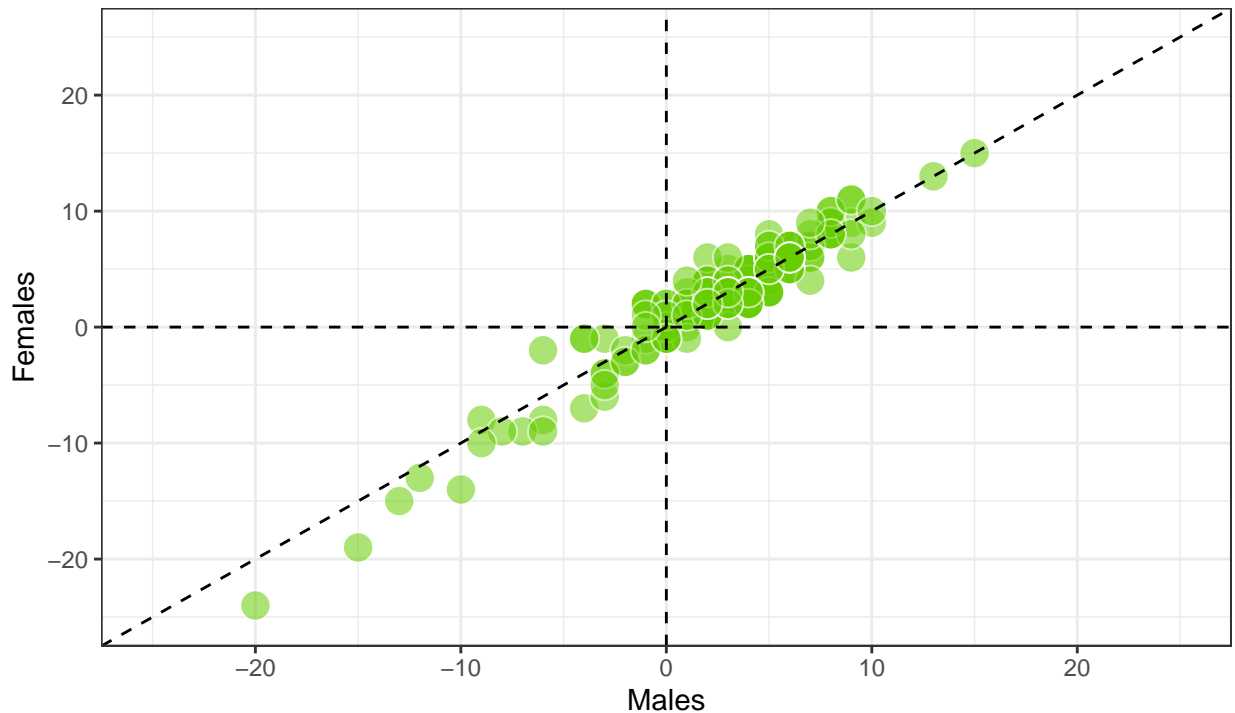
Adding reference lines can make plots easier to understand. We already added a diagonal line to visualize differences between men and women more clearly. Now we will add two more lines to help to identify in which countries people increased or decreased their life expectancy in the period analyzed.

```
# Adding an hline and vline to previous plots
ggplot(subdata2) +
  aes(x = diff_Male, y = diff_Female, label = Country.or.Area) +
  geom_point(
    colour = "white",
    fill = "chartreuse3",
    shape = 21,
    alpha = .55,
    size = 5
  ) +
  geom_abline(intercept = 0, slope = 1, linetype = "dashed") +
  scale_x_continuous(limits = c(-25, 25)) +
  scale_y_continuous(limits = c(-25, 25)) +
  geom_hline(yintercept = 0, linetype = "dashed") +
  geom_vline(xintercept = 0, linetype = "dashed") +
  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()
```

Life Expectancy at Birth by Country in Years
Difference between 1985–1990 and 2000–2005. Average.



Source: United Nations Statistics Division

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.

```
# Subsetting data to obtain countries of interest
top <- subdata2 %>%
  arrange(diff_Male + diff_Female) %>%
  select(-Source, -Unit, - Value.Footnotes) %>%
  head(3)
```

```
top
```

```
## # A tibble: 3 x 7
##   Country.or.Area Female_2000_2005 Female_1985_1990 Male_2000_2005
##   <chr>                <int>                <int>                <int>
## 1 Zimbabwe             40                  64                  40
## 2 Botswana             47                  66                  46
## 3 Swaziland            45                  60                  43
## # ... with 3 more variables: Male_1985_1990 <int>, diff_Female <int>,
## #   diff_Male <int>
```

```

bottom <- subdata2 %>%
  arrange(diff_Male + diff_Female) %>%
  select(-Source, -Unit, - Value.Footnotes) %>%
  tail(3)

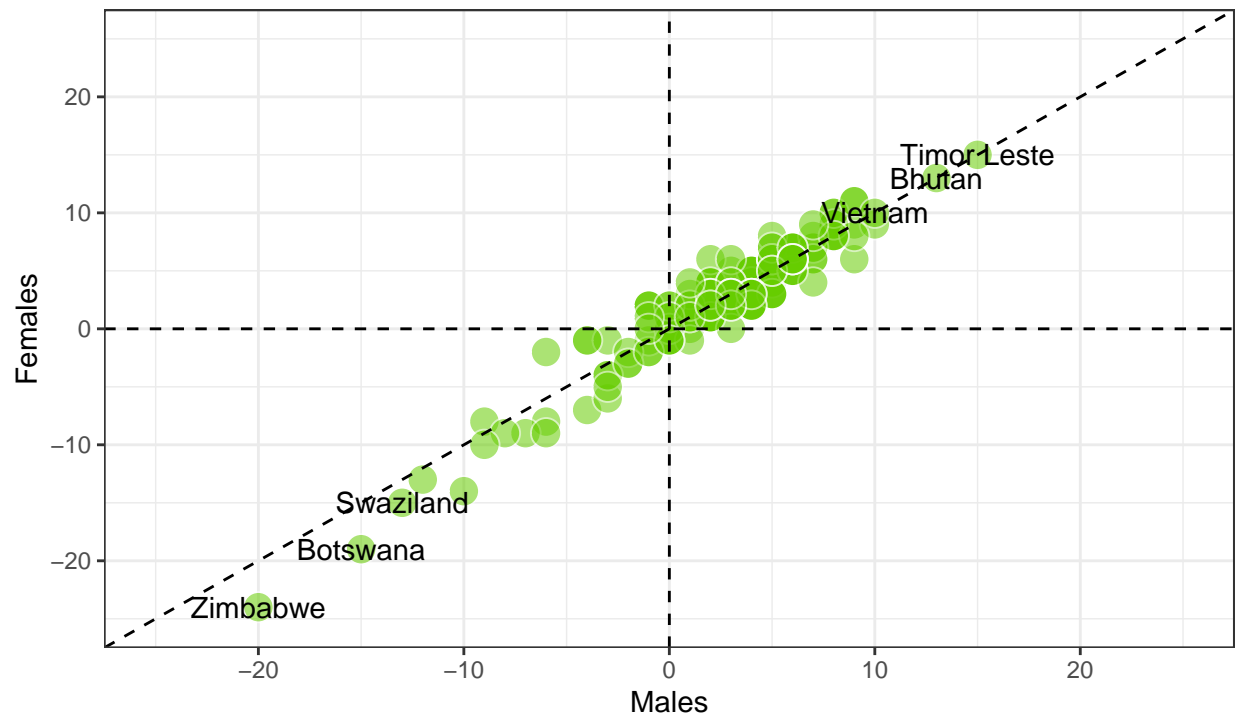
bottom

## # A tibble: 3 x 7
##   Country.or.Area Female_2000_2005 Female_1985_1990 Male_2000_2005
##   <chr>                <int>                <int>                <int>
## 1 Vietnam              75                  65                  71
## 2 Bhutan               65                  52                  62
## 3 Timor Leste          59                  44                  58
## # ... with 3 more variables: Male_1985_1990 <int>, diff_Female <int>,
## #   diff_Male <int>

# Adding text to the previous plot to label countries of interest
ggplot(subdata2, guide = FALSE) +
  aes(x = diff_Male, y = diff_Female, label = Country.or.Area) +
  geom_point(
    colour = "white",
    fill = "chartreuse3",
    shape = 21,
    alpha = .55,
    size = 5
  ) +
  geom_abline(intercept = 0, slope = 1, linetype = "dashed") +
  scale_x_continuous(limits = c(-25, 25)) +
  scale_y_continuous(limits = c(-25, 25)) +
  geom_hline(yintercept = 0, linetype = "dashed") +
  geom_vline(xintercept = 0, linetype = "dashed") +
  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(top, mapping = aes(label = Country.or.Area)) +
  geom_text(bottom, mapping = aes(label = Country.or.Area)) +
  theme_bw()

```

Life Expectancy at Birth by Country in Years
Difference between 1985–1990 and 2000–2005. Average.



Source: United Nations Statistics Division