

Visualizing Inequalities in Life Expectancy

Lucas Mendicino

3/4/2022

```
options(repr.plot.width = 6, repr.plot.height = 6)
library(readr)
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)
```

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.

```
life_expectancy <- read_csv("Life_Expectancy_UN.csv")

## Warning: One or more parsing issues, see `problems()` for details
## Rows: 1571 Columns: 7

## -- Column specification -----
## Delimiter: ","
## chr (5): Country or Area, Subgroup, Year, Source, Unit
## dbl (2): Value, Value Footnotes

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

life_expectancy <- life_expectancy %>%
  rename(Country.or.Area = `Country or Area`)
head(life_expectancy)
```

```
## # A tibble: 6 x 7
##   Country.or.Area Subgroup Year      Source      Unit  Value `Value Footnote~
##   <chr>           <chr>   <chr>   <chr>      <chr> <dbl>      <dbl>
## 1 Afghanistan    Female 2000-2005 UNPD_World Po~ Years    42          NA
## 2 Afghanistan    Female 1995-2000 UNPD_World Po~ Years    42          NA
## 3 Afghanistan    Female 1990-1995 UNPD_World Po~ Years    42          NA
## 4 Afghanistan    Female 1985-1990 UNPD_World Po~ Years    41          NA
## 5 Afghanistan    Male   2000-2005 UNPD_World Po~ Years    42          NA
## 6 Afghanistan    Male   1995-2000 UNPD_World Po~ Years    42          NA
```

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).

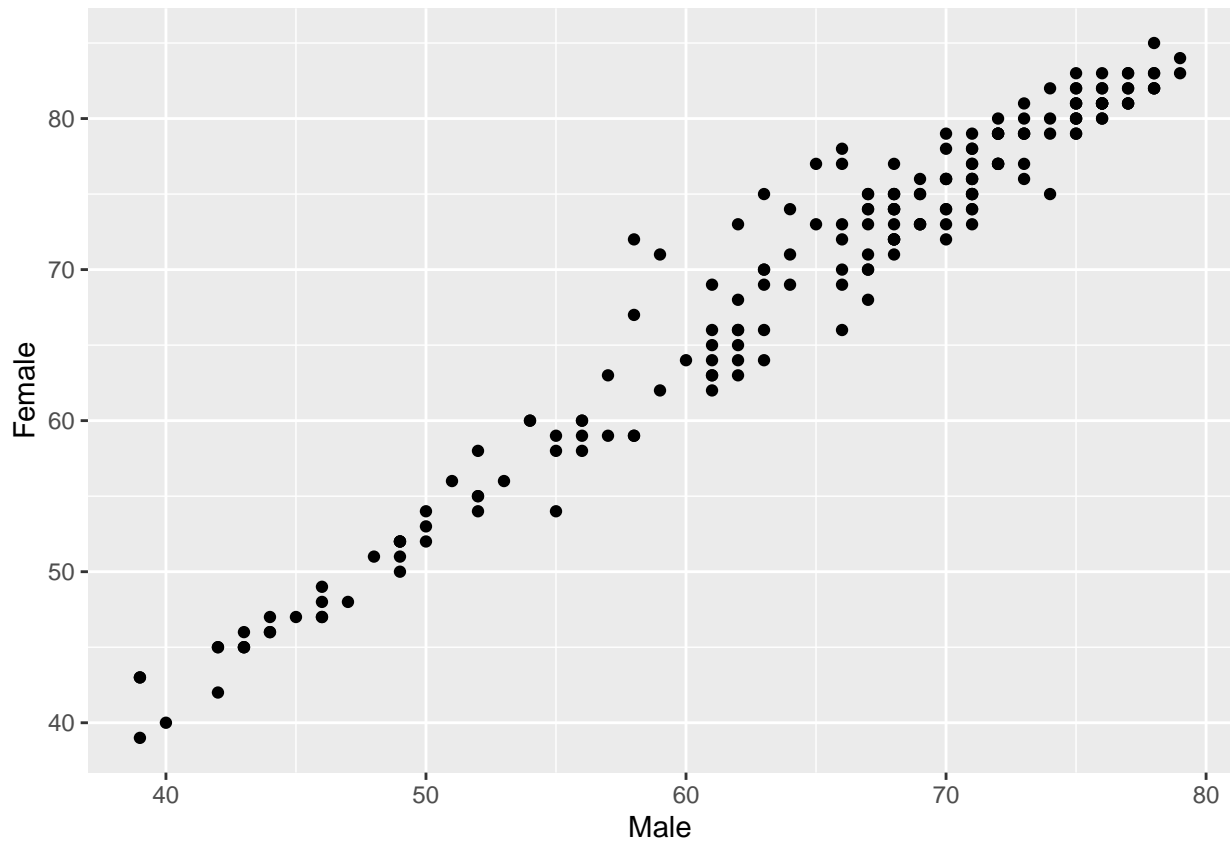
```
subdata <- life_expectancy %>%
  filter(Year=="2000-2005") %>%
  select(Country.or.Area, Subgroup, Value) %>%
  spread(Subgroup, Value)
```

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

```
## # A tibble: 6 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
```

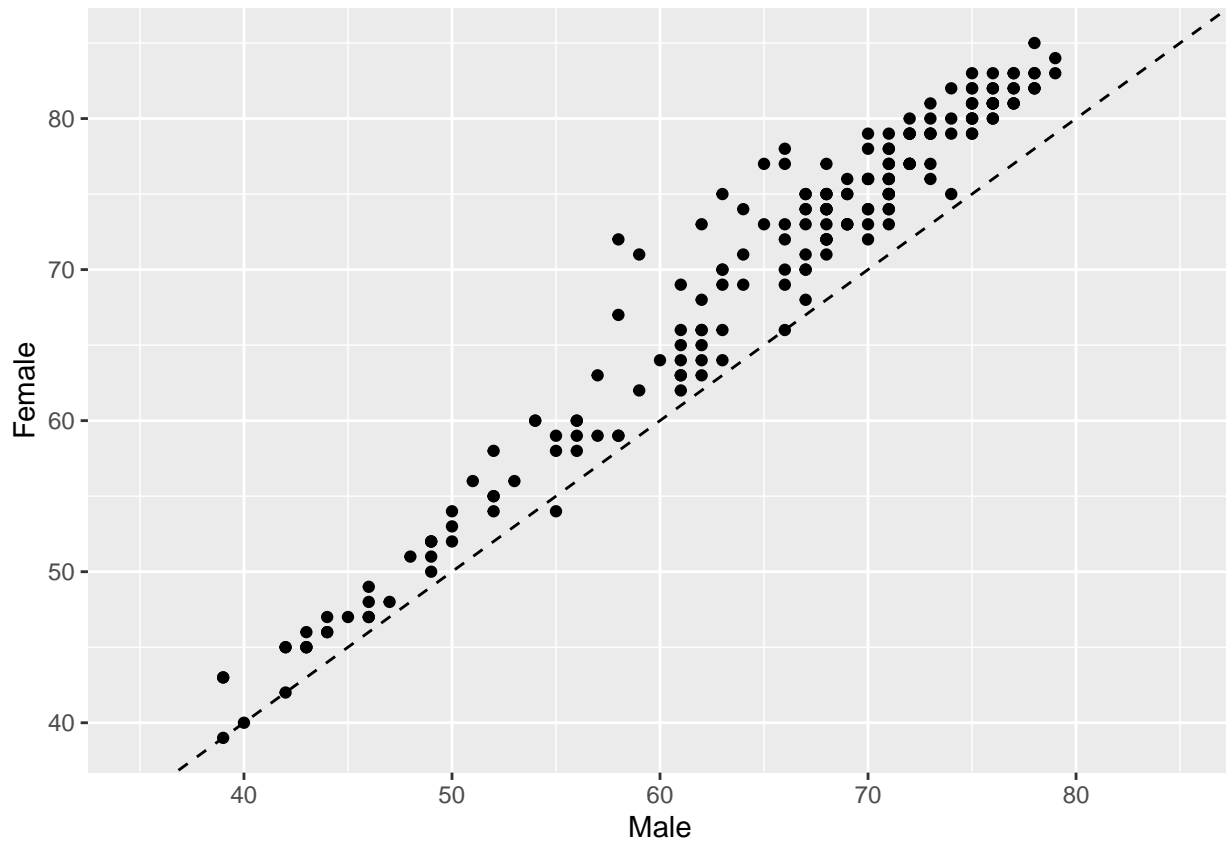
Let's create a scatter plot using ggplot2 to represent life expectancy of males (on the x-axis) against females (on the y-axis).

```
ggplot(subdata, aes(Male, Female)) + geom_point()
```



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.

```
ggplot(subdata, aes(Male, Female)) + geom_point() +  
  geom_abline(intercept = 0, slope = 1, linetype = 2) +  
  scale_x_continuous(limits = c(35, 85))
```



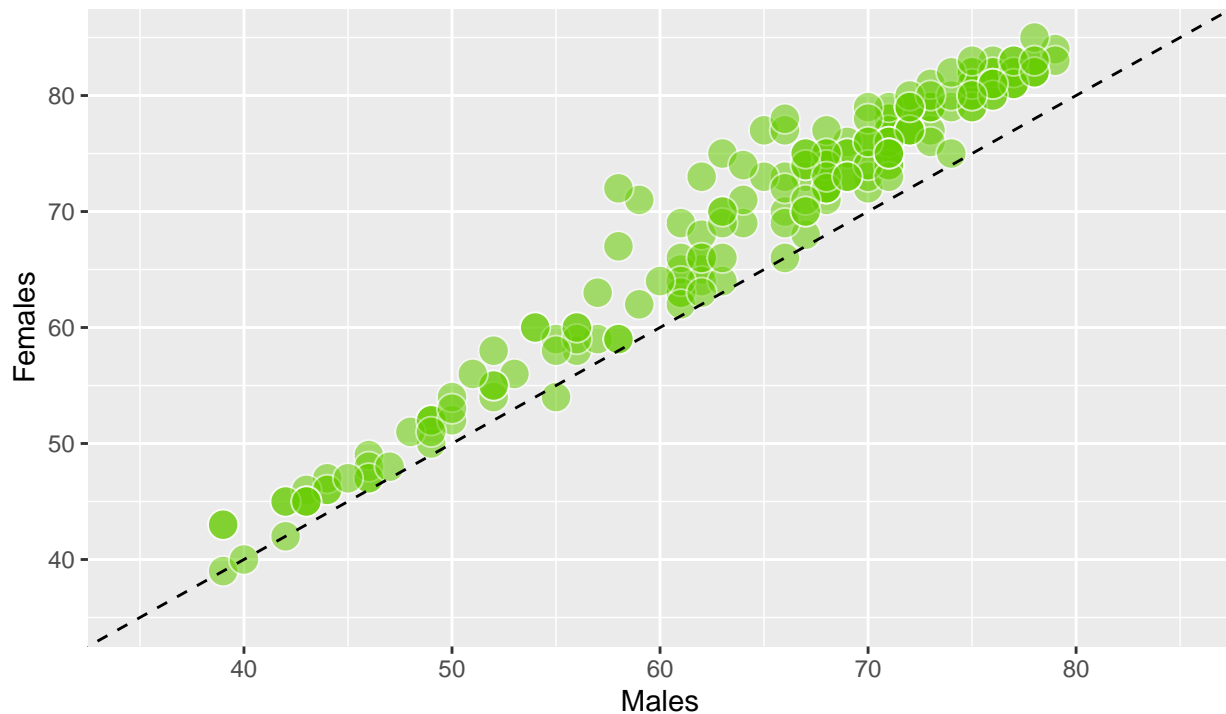
We see how most of the points are arranged above the diagonal and how there is a significant dispersion among them. This shows the females tend to have longer life expectancies than males.

Let's add titles, axis labels, and a caption, and change the appearance.

```
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=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")
```

Life Expectancy at Birth by Country

Years. Period: 2000–2005. Average.



Source: United Nations Statistics Division

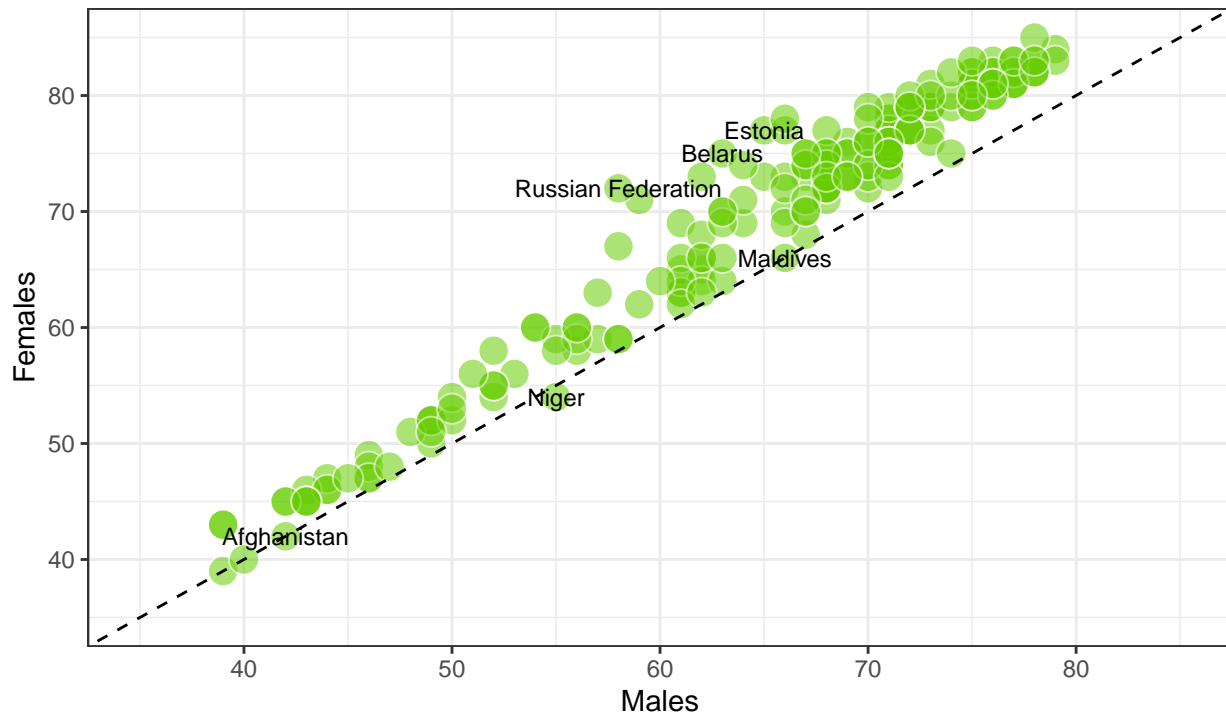
Let's further discriminate this visualization by drawing attention to countries where the gap in life expectancy between men and women is significantly high.

```
# Subsetting data to obtain countries of interest
top_male <- subdata %>% arrange(Male-Female) %>% head(3)
top_female <- subdata %>% arrange(Female-Male) %>% head(3)

# 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=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, size=3) +
  geom_text(data=top_female, size=3) +
  theme_bw()
```

Life Expectancy at Birth by Country

Years. Period: 2000–2005. Average.



Source: United Nations Statistics Division

Since the data contains historical information, let's see now how life expectancy has evolved in recent years. The 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) %>%
  spread(Sub_Year, Value) %>%
  mutate(diff_Female = Female_2000_2005 - Female_1985_1990,
         diff_Male   = Male_2000_2005 - Male_1985_1990)

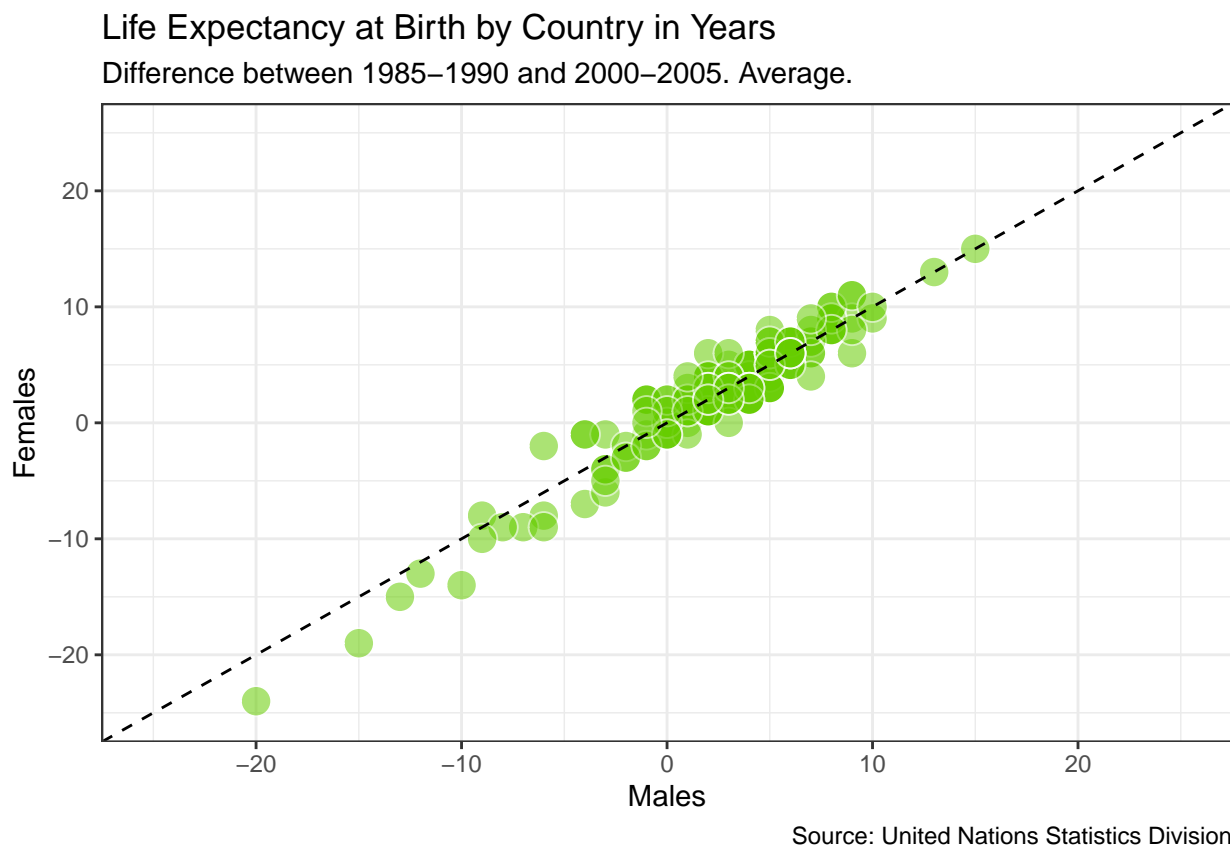
# Taking a look at the first few rows
head(subdata2)
```

```
## # A tibble: 6 x 10
##   Country.or.Area Source      Unit `Value Footnote~ Female_1985_1990
##   <chr>             <chr>      <chr>      <dbl>      <dbl>
## 1 Afghanistan      UNPD_World Population~ Years      NA      41
## 2 Albania           UNPD_World Population~ Years      NA      75
## 3 Algeria           UNPD_World Population~ Years      NA      67
## 4 Angola            UNPD_World Population~ Years      NA      42
## 5 Argentina         UNPD_World Population~ Years      NA      75
## 6 Armenia           UNPD_World Population~ Years      NA      71
```

```
## # ... with 5 more variables: Female_2000_2005 <dbl>, Male_1985_1990 <dbl>,  
## #   Male_2000_2005 <dbl>, diff_Female <dbl>, diff_Male <dbl>
```

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(colour="white", fill="chartreuse3", shape=21, alpha=.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()
```



Let's add in vertical and horizontal reference lines to identify which countries 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(colour="white", fill="chartreuse3", shape=21, alpha=.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",
```

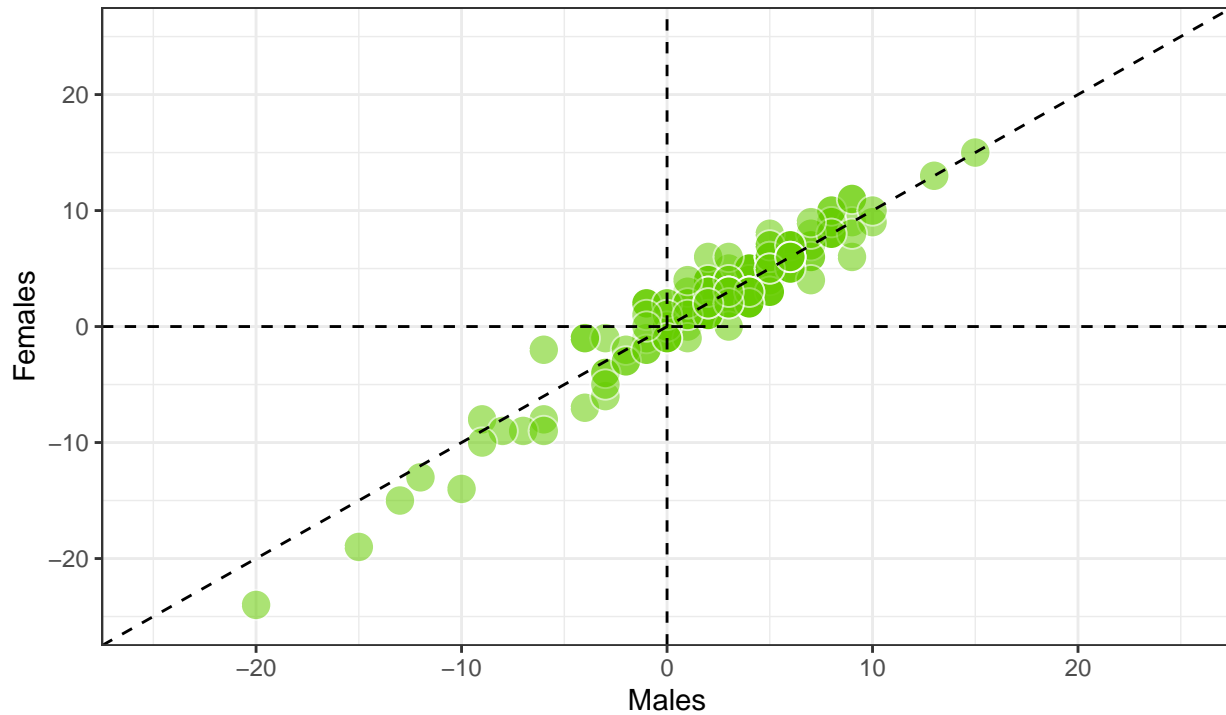
```

    subtitle="Years. 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

Years. Difference between 1985-1990 and 2000-2005. Average.



Source: United Nations Statistics Division

Let's label some points where the aggregated life expectancy for men and women increased most and decreased most in the period.

```

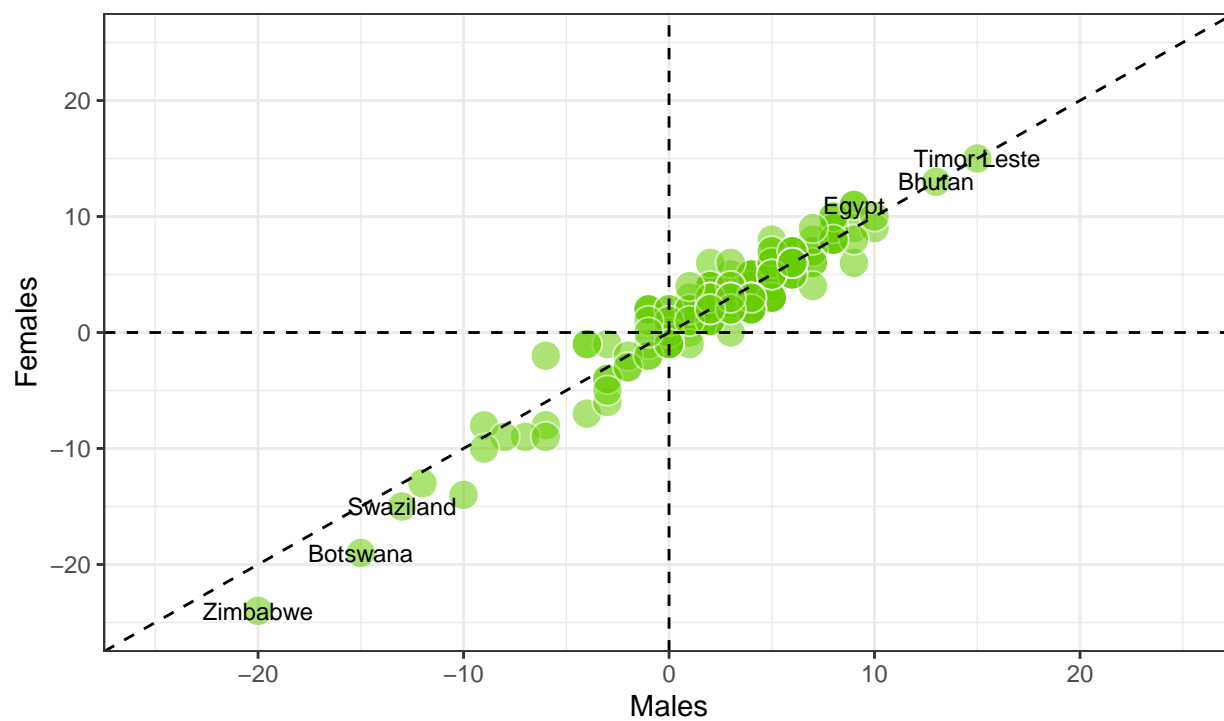
top <- subdata2 %>% arrange(diff_Male+diff_Female) %>% head(3)
bottom <- subdata2 %>% arrange(-(diff_Male+diff_Female)) %>% head(3)

ggplot(subdata2, aes(x=diff_Male, y=diff_Female, label=Country.or.Area), guide=FALSE)+
  geom_point(colour="white", fill="chartreuse3", shape=21, alpha=.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",
       subtitle="Years. Difference between 1985-1990 and 2000-2005. Average.",
       caption="Source: United Nations Statistics Division",
       x="Males",
       y="Females")+
  geom_text(data=top, size=3)+
  geom_text(data=bottom, size=3)+
  theme_bw()

```


Life Expectancy at Birth by Country

Years. Difference between 1985–1990 and 2000–2005. Average.



Source: United Nations Statistics Division