Introduction

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 over time or between countries because of many causes: the evolution of medicine, the degree of development of nations, 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 in life expectancy at birth around the world.

The dataset

We will use a dataset from the United Nations Statistics Division(Source: http://data.un.org/Data.aspx?d=GenderStat&f=inID:37&c=1,2,3,4,5,6&s=crEngName:asc,sgvEngName:asc,timeEngName:desc&v=1). It has the following columns:

- Country.or.Area = Where people live
- Subgroup = Gender
- Year = Research time range is taken.
- Source = Research Origin.
- Units = Time unit.
- Value = Age.

First we load the dataset, and packages. Then we take a look on a first few rows.

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

# Loading packages
library(dplyr)
library(tidyr)
library(ggplot2)

# Loading data
life_expectancy <- read.csv("UNdata.csv")

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

	Country.or.Area	Subgroup	Year	Source	Unit	Value
	<fctr>	<fctr>	<fctr>	<fctr>	<fctr>	<int>
1	Afghanistan	Female	2000-2005	UNPD_World Population Prospects_2006 (International estimate)	Years	42
2	Afghanistan	Female	1995-2000	UNPD_World Population Prospects_2006 (International estimate)	Years	42
3	Afghanistan	Female	1990-1995	UNPD_World Population Prospects_2006 (International estimate)	Years	42
4	Afghanistan	Female	1985-1990	UNPD_World Population Prospects_2006 (International estimate)	Years	41
5	Afghanistan	Male	2000-2005	UNPD_World Population Prospects_2006 (International estimate)	Years	42
6	Afghanistan	Male	1995-2000	UNPD_World Population Prospects_2006 (International estimate)	Years	42

Empirical Analysis

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).

```{r}
# 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)
```

	Country.or.Area	Female	Male
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

A scatter plot is a helpful way to visualize the relationship between two variables. It is a simple plot in which points are arranged on two axes, representing one of those variables. Let's create a scatter plot using ggplot2 to describe the 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.

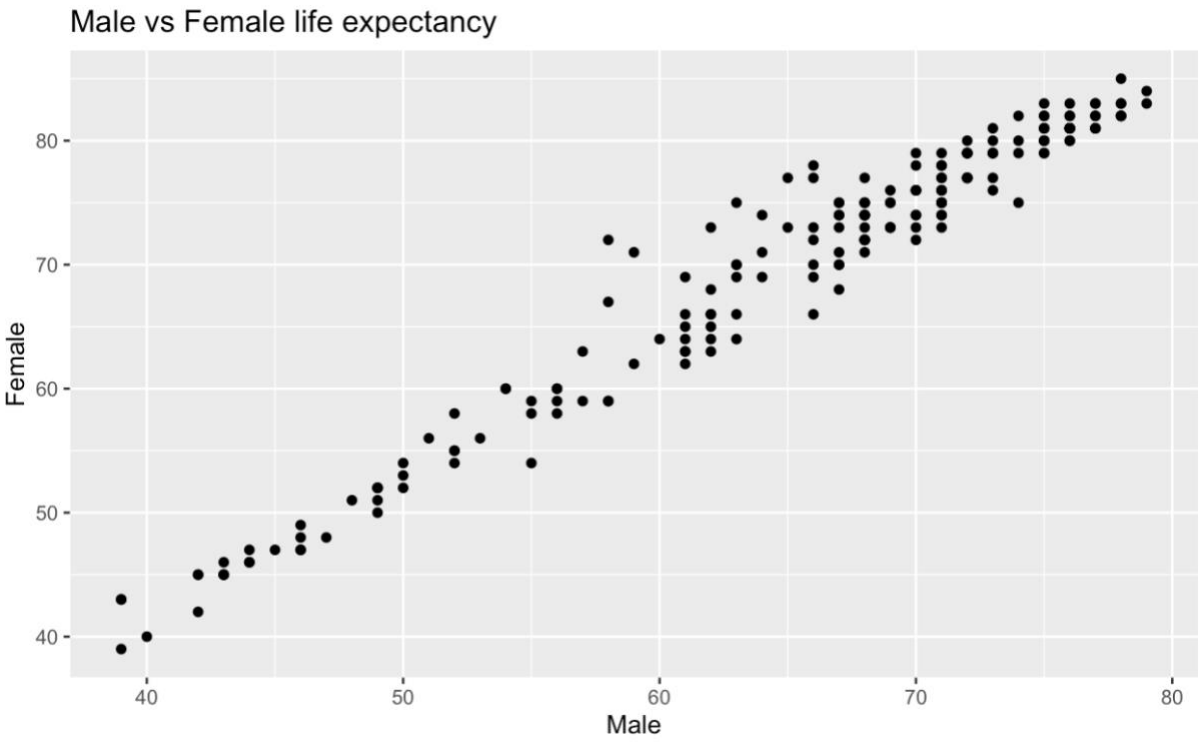
```{r}

# Plotting male and female life expectancy

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

  geom_point() + labs(title="Male vs Female life expectancy")

```



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.

```{r}

# Adding labels to previous plot

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

 geom_point(colour = "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",

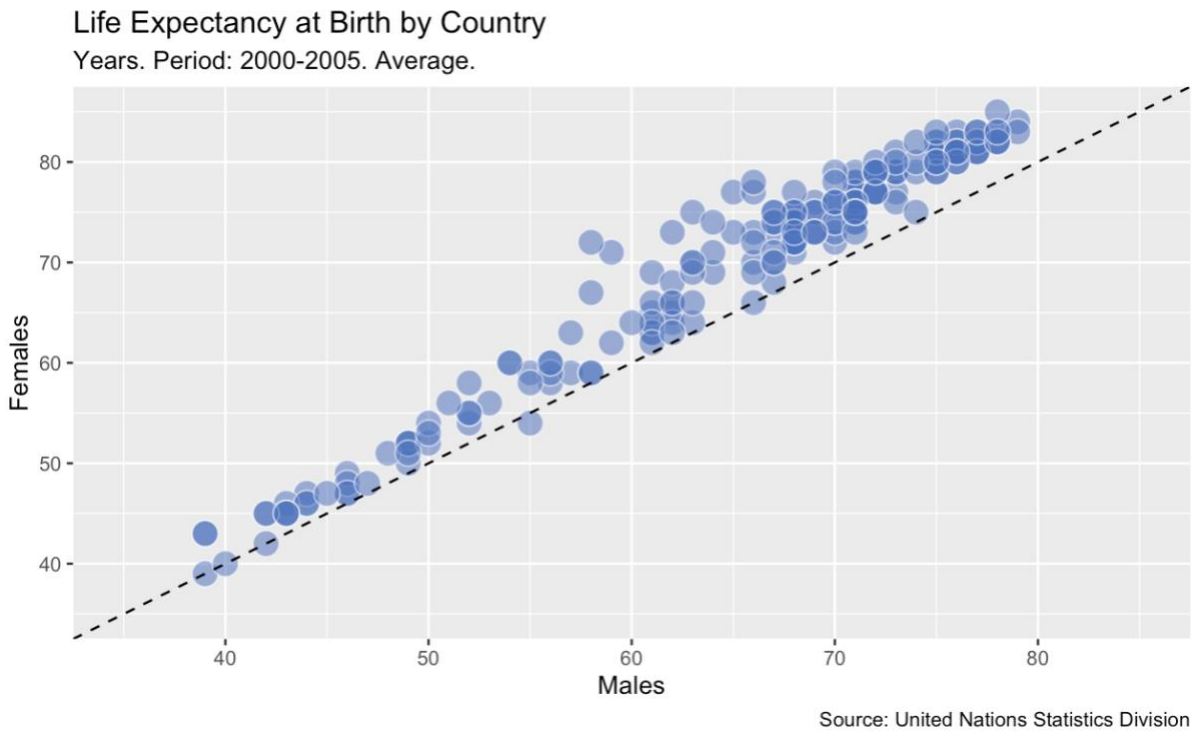
     subtitle = "Years. Period: 2000-2005. Average.",

     caption="Source: United Nations Statistics Division",

     x = "Males",

     y = "Females")
```

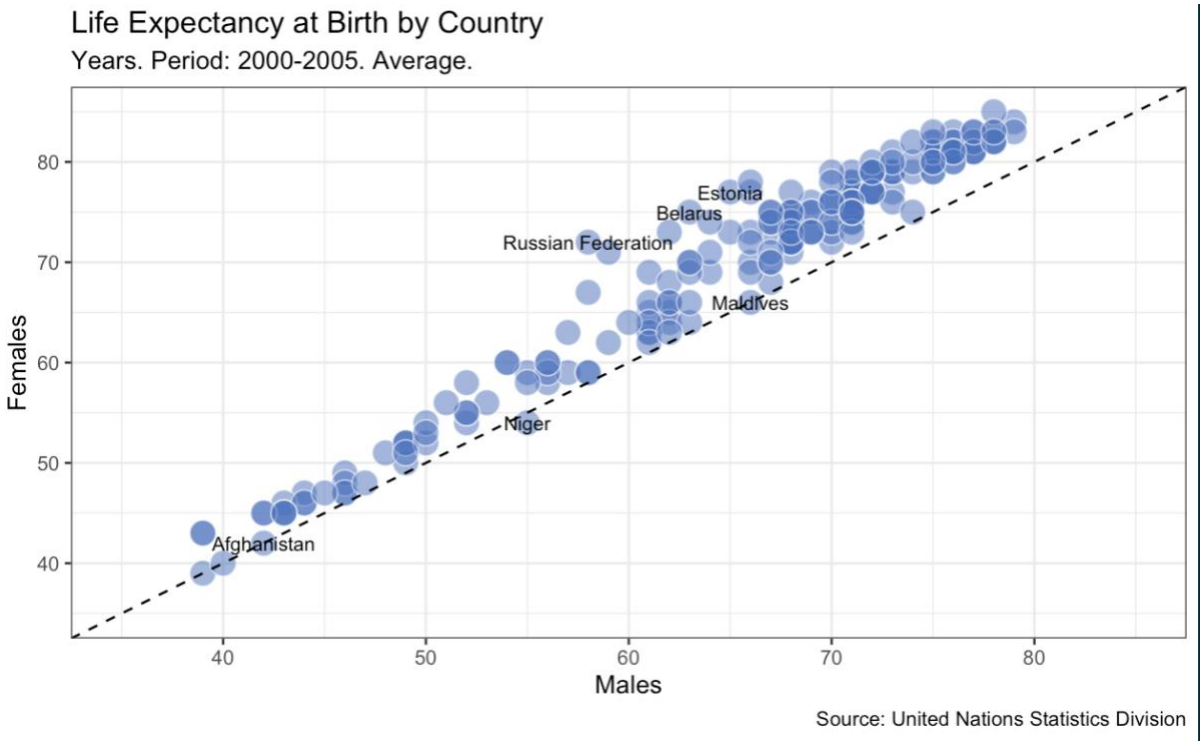
```




Now, we will label some points of our plot with the name of its corresponding country. We want to draw attention to some particular countries where the gap in life expectancy between men and women is significantly high.

```{r}
# Subseting 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="#4D73BE", 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()
```
```



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.

```{r}
# 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)
```

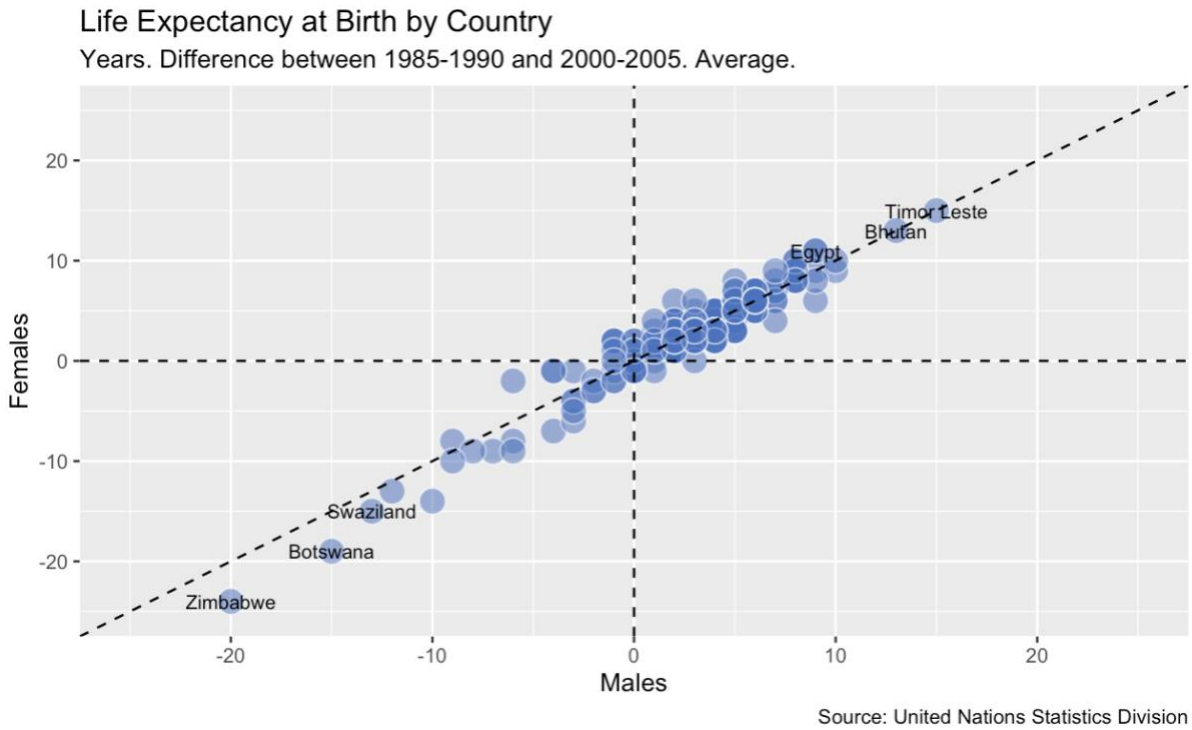
	Country.or.Area	Unit	Female_1985_1990	Female_2000_2005	Male_1985_1990	Male_2000_2005	diff_Female	diff_Male
1	Afghanistan	Years	41	42	41	42	1	1
2	Albania	Years	75	79	69	73	4	4
3	Algeria	Years	67	72	65	70	5	5
4	Angola	Years	42	43	38	39	1	1
5	Argentina	Years	75	78	68	71	3	3
6	Armenia	Years	71	75	66	68	4	2
7	Aruba	Years	78	76	71	70	-2	-1
8	Australia	Years	79	83	73	78	4	5
9	Austria	Years	78	82	72	76	4	4
10	Azerbaijan	Years	70	70	62	63	0	1
11	Bahamas	Years	73	74	66	68	1	2
12	Bahrain	Years	73	77	69	73	4	4
13	Bangladesh	Years	53	63	53	61	10	8
14	Barbados	Years	77	79	72	73	2	1

Conclusion

Now let's create our second plot to represent average life expectancy differences between "1985-1990" and "2000-2005" for men and women. 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 identify in which countries people increased or decreased their life expectancy in the period analyzed.

```{r}
# Subseting data to obtain countries of interest
top <- subdata2 %>% arrange(diff_Male+diff_Female) %>% head(3)
bottom <- subdata2 %>% arrange(-(diff_Male+diff_Female)) %>% head(3)

# Adding text to the previous plot to label countries of interest
ggplot(subdata2, aes(x=diff_Male, y=diff_Female, label=Country.or.Area), guide=FALSE)+
  geom_point(colour="white", fill="#4D73BE", 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()
```



From these plots, we can conclude females have a longer life expectancy than males. From the plot above, we can conclude Zimbabwe, Botswana, and Swaziland are countries where people decreased their life expectancy in the period analyzed. Egypt, Bhutan, and Timor Leste are countries where people increased their life expectancy in the period studied.

Source

- http://data.un.org/Data.aspx?d=GenderStat&f=inID:37&c=1,2,3,4,5,6&s=crEngName:asc,sgvEngName:asc,timeEngName:desc&v=1
- https://www.voanews.com/science-health/new-study-looks-why-females-live-longer-males
- https://ourworldindata.org/why-do-women-live-longer-than-men