Tuan Linh Dao

Code

Project 2: Data visualization with life expectancy with Rstudio

A. Load library :

To install a package from R, type the command install.packages("name of the package") or click "Packages" then "Install", type the name of package that we want to install.

Hide

```
library(ggplot2)
library(gganimate)
```

```
No renderer backend detected. gganimate will default to writing frames to separate fi
les
Consider installing:
- the `gifski` package for gif output
- the `av` package for video output
and restarting the R session
```

Hide

```
library(dplyr)
library(ggthemes)
library(tidyverse)
```

B. Import dataset :

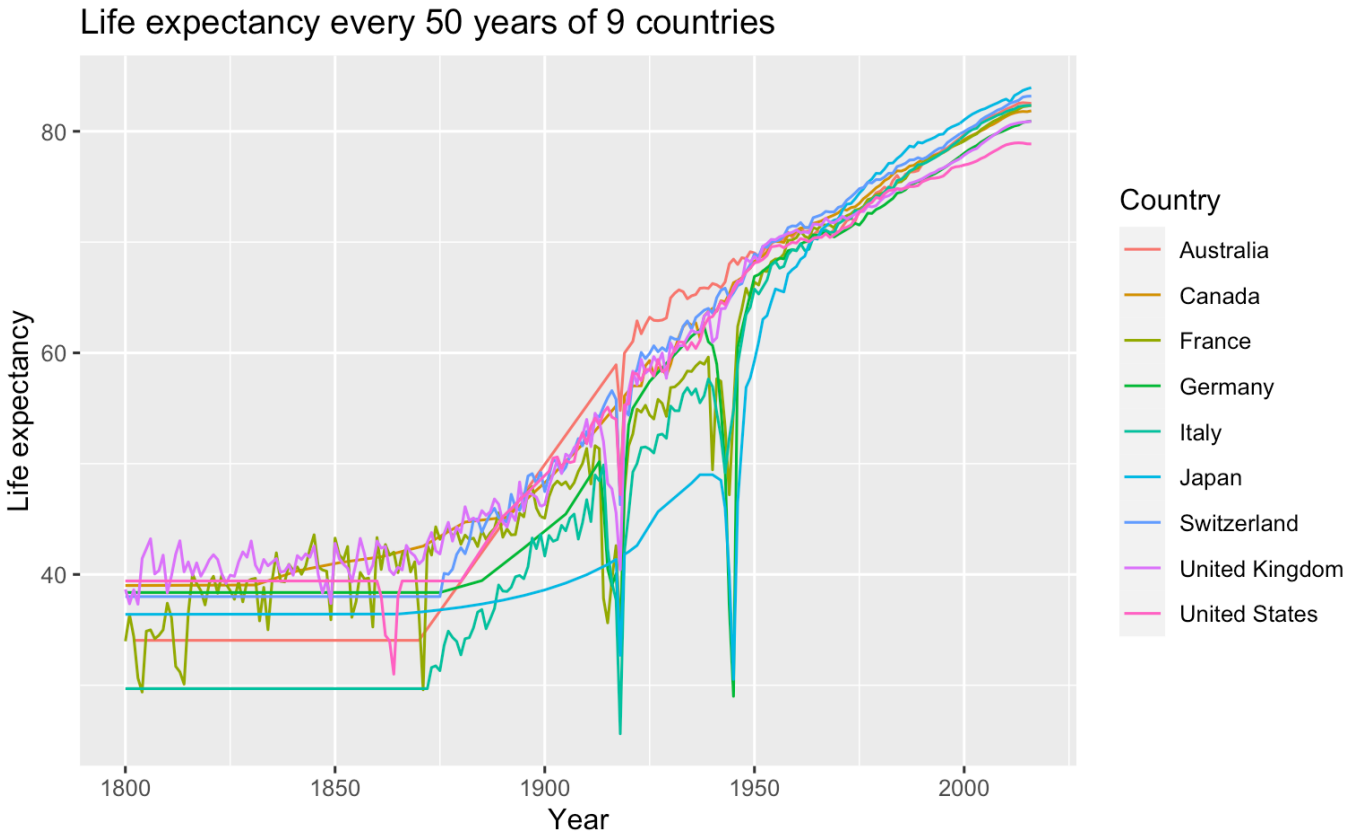
Hide

```
data <- read.csv(file.choose(), header = T)
data
```

C. Line chart with ggplot :

Hide

```
data %>%
  ggplot(aes(y=Life.expectancy,x=Year, col=Entity)) +
  geom_line() +
  labs(title = "Life expectancy every 50 years of 9 countries",
       y = "Life expectancy",
       x = "Year",
       color = "Country")
```



D. Countries :

1) Australia :

a) Dataframe :

Hide

```
aus <- data[data$Entity == "Australia", ]
aus
```

	Entity <chr>	Year <int>	Life.expectancy <dbl>
1	Australia	1802	34.05000
2	Australia	1803	34.05000
3	Australia	1804	34.05000
4	Australia	1805	34.05000
5	Australia	1806	34.05000
6	Australia	1807	34.05000
7	Australia	1808	34.05000
8	Australia	1809	34.05000
9	Australia	1810	34.05000
10	Australia	1811	34.05000

1-10 of 215 rows Previous **1** 2 3 4 5 6 … 22 Next

b) Summarize data :

Hide

```
summary(aus)
```

```
   Entity              Year      Life.expectancy
Length:215         Min.   :1802   Min.   :34.05
Class :character   1st Qu.:1856   1st Qu.:34.05
Mode  :character   Median :1909   Median :54.69
                   Mean   :1909   Mean   :54.29
                   3rd Qu.:1962   3rd Qu.:70.85
                   Max.   :2016   Max.   :82.58
```

Hide

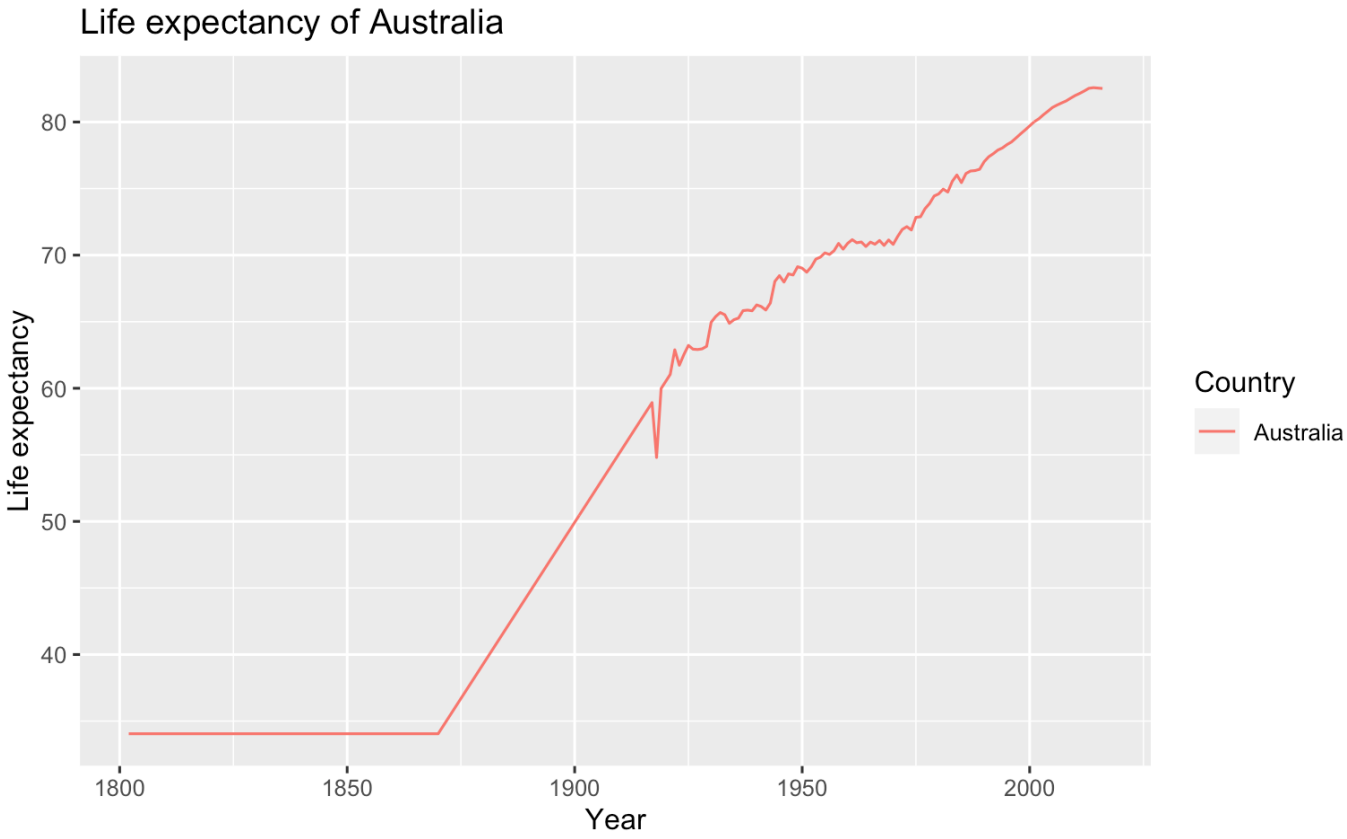
```
# Average life expectancy of Australia :
m_aus <- mean(aus$Life.expectancy)
m_aus
```

```
[1] 54.29
```

c) Plot :

Hide

```
aus %>%
  ggplot(aes(y=Life.expectancy,x=Year, col=Entity)) +
  geom_line() +
  labs(title = "Life expectancy of Australia",
       y = "Life expectancy",
       x = "Year",
       color = "Country")
```



2) Canada :

a) Dataframe :

Hide

```
can <- data[data$Entity == "Canada", ]
can
```

	Entity <chr>	Year <int>	Life.expectancy <dbl>
216	Canada	1800	39.00000
217	Canada	1801	39.00149
218	Canada	1802	39.00299
219	Canada	1803	39.00449
220	Canada	1804	39.00599
221	Canada	1805	39.00748
222	Canada	1806	39.00898
223	Canada	1807	39.01048
224	Canada	1808	39.01197
225	Canada	1809	39.01347

1-10 of 217 rows Previous **1** 2 3 4 5 6 ... 22 Next

b) Summarize data :

Hide

```
summary(can)
```

```
   Entity              Year      Life.expectancy
 Length:217         Min.   :1800   Min.   :39.00
 Class :character   1st Qu.:1854   1st Qu.:41.22
 Mode  :character   Median :1908   Median :50.96
                    Mean   :1908   Mean   :55.78
                    3rd Qu.:1962   3rd Qu.:71.30
                    Max.   :2016   Max.   :81.85
```

Hide

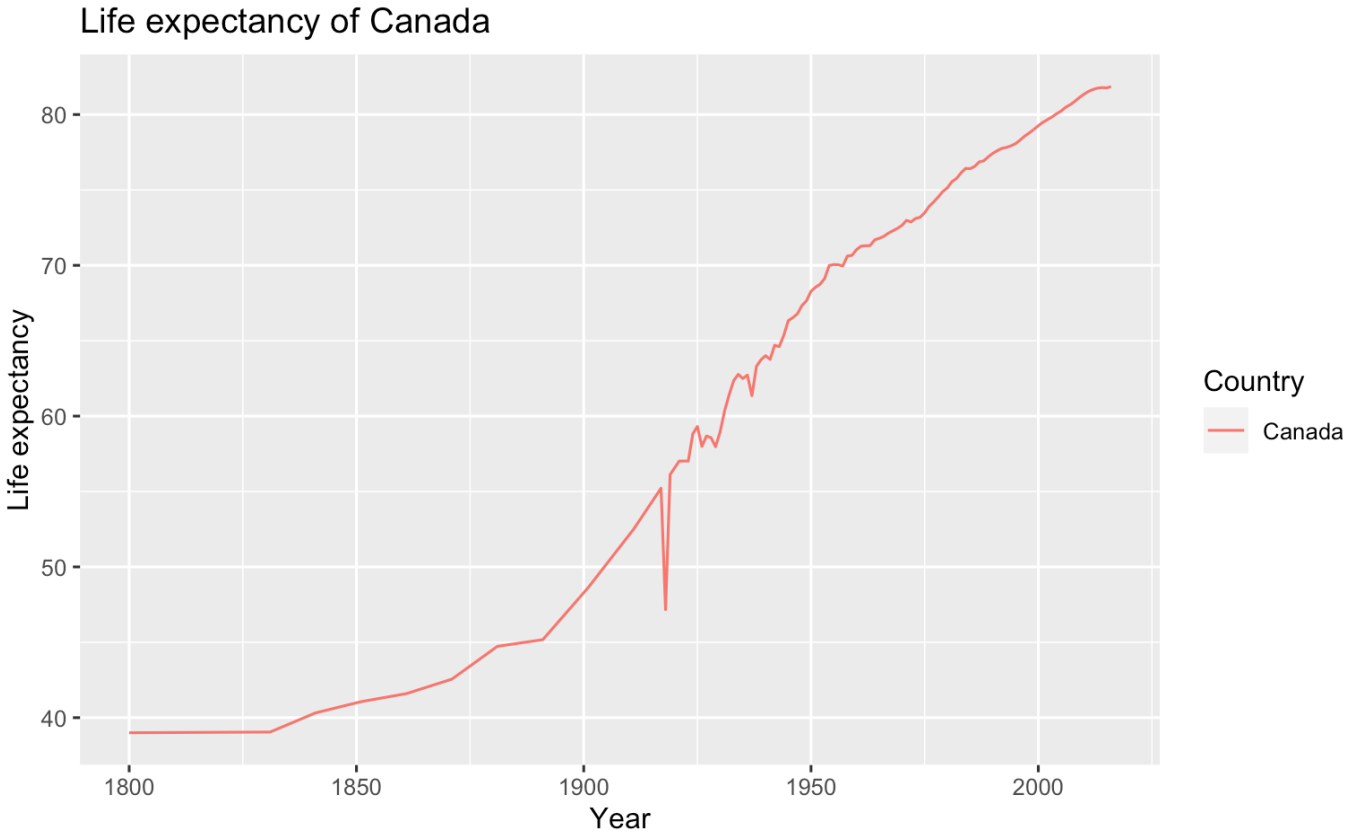
```
m_can <- mean(can$Life.expectancy)
m_can
```

```
[1] 55.77853
```

c) Plot :

Hide

```
can %>%
  ggplot(aes(y=Life.expectancy,x=Year, col=Entity)) +
  geom_line() +
  labs(title = "Life expectancy of Canada",
       y = "Life expectancy",
       x = "Year",
       color = "Country")
```



3) France :

a) Dataframe :

Hide

```
fra <- data[data$Entity == "France", ]
fra
```

	Entity <chr>	Year <int>	Life.expectancy <dbl>
433	France	1800	33.96717
434	France	1801	36.37465
435	France	1802	34.37999
436	France	1803	30.64262
437	France	1804	29.38176
438	France	1805	34.87632
439	France	1806	35.00000
440	France	1807	34.20000
441	France	1808	34.50000
442	France	1809	35.00000

1-10 of 217 rows Previous **1** 2 3 4 5 6 … 22 Next

b) Summarize data :

Hide

```
summary(fra)
```

```
   Entity              Year      Life.expectancy
 Length:217         Min.   :1800   Min.   :29.38
 Class :character   1st Qu.:1854   1st Qu.:40.29
 Mode  :character   Median :1908   Median :47.74
                    Mean   :1908   Mean   :53.69
                    3rd Qu.:1962   3rd Qu.:70.51
                    Max.   :2016   Max.   :82.34
```

Hide

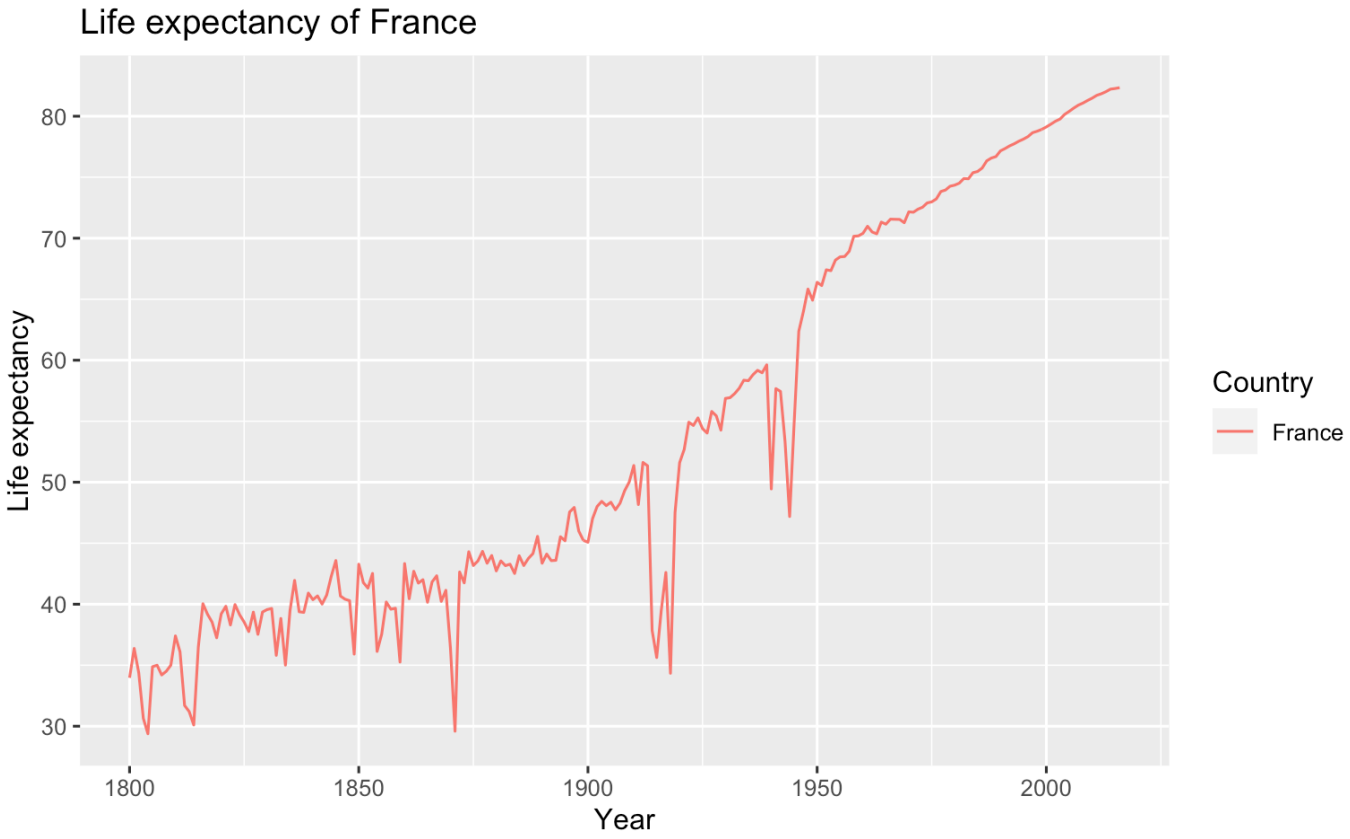
```
m_fra <- mean(fra$Life.expectancy)
m_fra
```

```
[1] 53.69172
```

c) Plot :

Hide

```
fra %>%
  ggplot(aes(y=Life.expectancy,x=Year, col=Entity)) +
  geom_line() +
  labs(title = "Life expectancy of France",
       y = "Life expectancy",
       x = "Year",
       color = "Country")
```



3) Germany :

a) Dataframe :

Hide

```
ger <- data[data$Entity == "Germany", ]
ger
```

	Entity <chr>	Year <int>	Life.expectancy <dbl>
650	Germany	1800	38.37000
651	Germany	1801	38.37000
652	Germany	1802	38.37000
653	Germany	1803	38.37000
654	Germany	1804	38.37000
655	Germany	1805	38.37000
656	Germany	1806	38.37000
657	Germany	1807	38.37000
658	Germany	1808	38.37000
659	Germany	1809	38.37000

1-10 of 217 rows Previous **1** 2 3 4 5 6 ... 22 Next

b) Summarize data :

Hide

```
summary(ger)
```

```
   Entity              Year      Life.expectancy
 Length:217         Min.   :1800   Min.   :29.00
 Class :character   1st Qu.:1854   1st Qu.:38.37
 Mode  :character   Median :1908   Median :44.53
                    Mean   :1908   Mean   :52.88
                    3rd Qu.:1962   3rd Qu.:70.00
                    Max.   :2016   Max.   :80.93
```

Hide

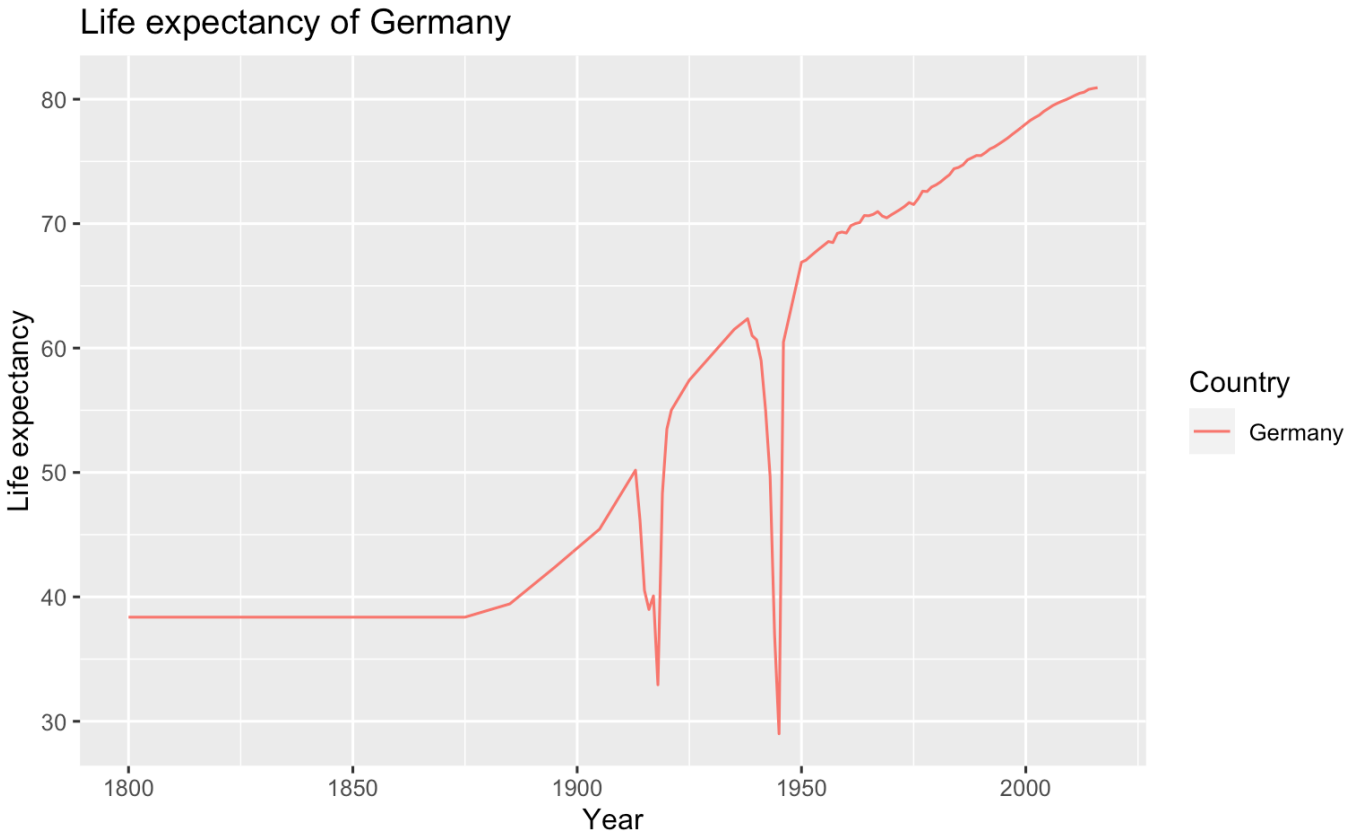
```
m_ger <- mean(ger$Life.expectancy)
m_ger
```

```
[1] 52.88094
```

c) Plot :

Hide

```
ger %>%
  ggplot(aes(y=Life.expectancy,x=Year, col=Entity)) +
  geom_line() +
  labs(title = "Life expectancy of Germany",
       y = "Life expectancy",
       x = "Year",
       color = "Country")
```



5) Italy :

a) Dataframe :

Hide

```
ita <- data[data$Entity == "Italy", ]
ita
```

	Entity <chr>	Year <int>	Life.expectancy <dbl>
867	Italy	1800	29.69
868	Italy	1801	29.69
869	Italy	1802	29.69
870	Italy	1803	29.69
871	Italy	1804	29.69
872	Italy	1805	29.69
873	Italy	1806	29.69
874	Italy	1807	29.69
875	Italy	1808	29.69
876	Italy	1809	29.69

1-10 of 217 rows Previous **1** 2 3 4 5 6 ... 22 Next

b) Summarize data :

Hide

```
summary(ita)
```

```
   Entity              Year      Life.expectancy
 Length:217         Min.   :1800   Min.   :25.62
 Class :character   1st Qu.:1854   1st Qu.:29.69
 Mode  :character   Median :1908   Median :43.52
                    Mean   :1908   Mean   :49.23
                    3rd Qu.:1962   3rd Qu.:69.35
                    Max.   :2016   Max.   :82.34
```

Hide

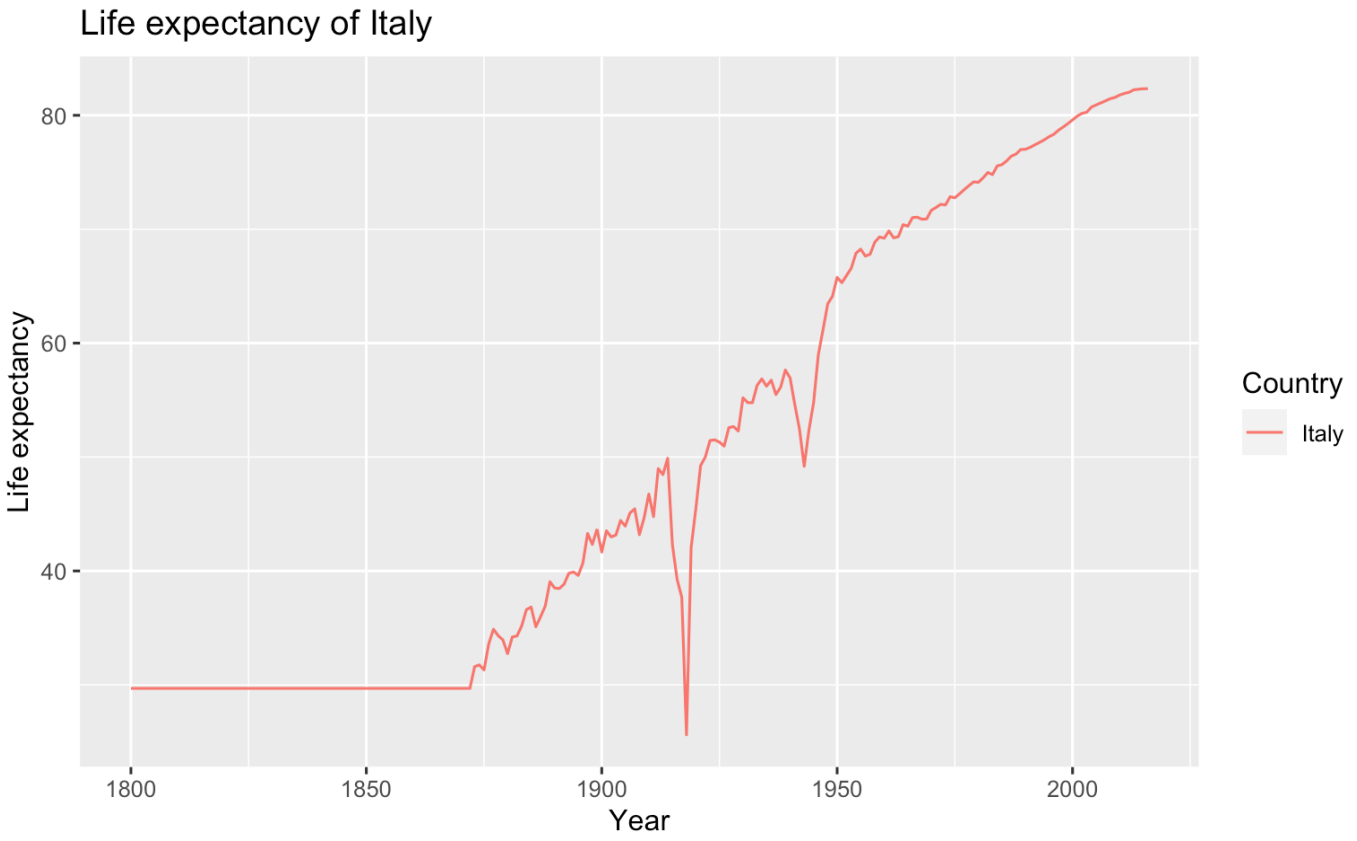
```
m_ita <- mean(ita$Life.expectancy)
m_ita
```

```
[1] 49.22636
```

c) Plot :

Hide

```
ita %>%
  ggplot(aes(y=Life.expectancy,x=Year, col=Entity)) +
  geom_line() +
  labs(title = "Life expectancy of Italy",
       y = "Life expectancy",
       x = "Year",
       color = "Country")
```



6) Japan :

a) Dataframe :

Hide

```
jpn <- data[data$Entity == "Japan", ]
jpn
```

	Entity	Year	Life.expectancy
	<chr>	<int>	<dbl>
1084	Japan	1800	36.40000

	Entity <chr>	Year <int>	Life.expectancy <dbl>
1085	Japan	1801	36.40040
1086	Japan	1802	36.40080
1087	Japan	1803	36.40119
1088	Japan	1804	36.40159
1089	Japan	1805	36.40199
1090	Japan	1806	36.40239
1091	Japan	1807	36.40278
1092	Japan	1808	36.40318
1093	Japan	1809	36.40358

1-10 of 217 rows Previous **1** 2 3 4 5 6 … 22 Next

b) Summarize data :

Hide

```
summary(jpn)
```

```
   Entity              Year      Life.expectancy
 Length:217         Min.   :1800   Min.   :30.54
 Class :character   1st Qu.:1854   1st Qu.:36.42
 Mode  :character   Median :1908   Median :39.37
                    Mean   :1908   Mean   :50.35
                    3rd Qu.:1962   3rd Qu.:68.73
                    Max.   :2016   Max.   :83.94
```

Hide

```
m_jpn <- mean(jpn$Life.expectancy)
m_jpn
```

```
[1] 50.34571
```

c) Plot :

Hide

```
jpn %>%
  ggplot(aes(y=Life.expectancy,x=Year, col=Entity)) +
  geom_line() +
  labs(title = "Life expectancy of Japan",
       y = "Life expectancy",
       x = "Year",
       color = "Country") +
  transition_time(Year)
```

```
Error in jpn %>% ggplot(aes(y = Life.expectancy, x = Year, col = Entity)) :
  could not find function "%>%"
```

7) Switzerland :

a) Dataframe :

Hide

```
swi <- data[data$Entity == "Switzerland", ]
swi
```

	Entity <chr>	Year <int>	Life.expectancy <dbl>
1301	Switzerland	1800	38.00
1302	Switzerland	1801	38.00
1303	Switzerland	1802	38.00
1304	Switzerland	1803	38.00
1305	Switzerland	1804	38.00
1306	Switzerland	1805	38.00
1307	Switzerland	1806	38.00
1308	Switzerland	1807	38.00
1309	Switzerland	1808	38.00
1310	Switzerland	1809	38.00

1-10 of 217 rows Previous **1** 2 3 4 5 6 … 22 Next

b) Summarize data :

Hide

```
summary(swi)
```

```
   Entity              Year      Life.expectancy
 Length:217         Min.   :1800   Min.   :38.00
 Class :character   1st Qu.:1854   1st Qu.:38.00
 Mode  :character   Median :1908   Median :51.19
                    Mean   :1908   Mean   :55.25
                    3rd Qu.:1962   3rd Qu.:71.46
                    Max.   :2016   Max.   :83.18
```

Hide

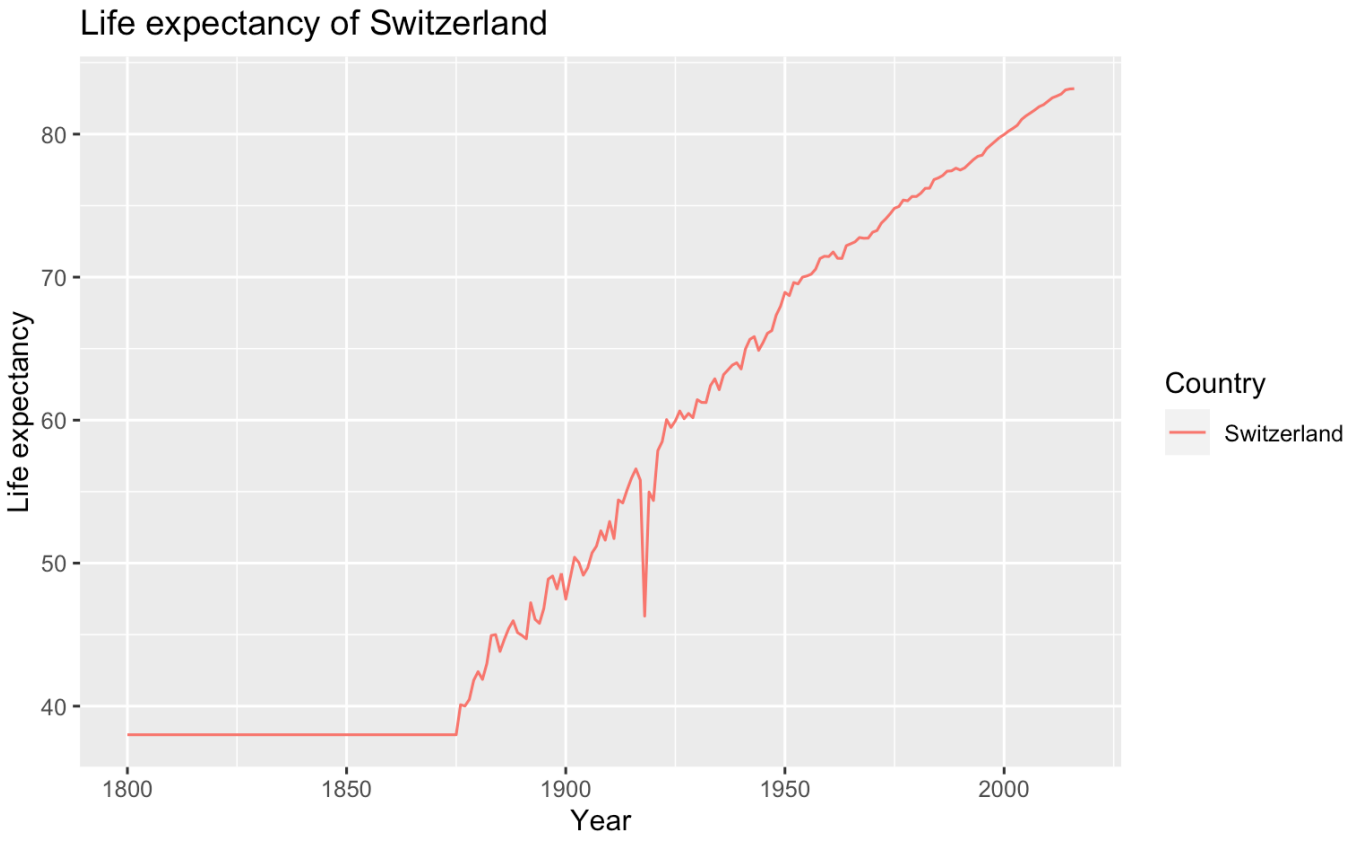
```
# Average life expectancy of Switzerland :
m_swi <- mean(swi$Life.expectancy)
m_swi
```

```
[1] 55.24604
```

c) Plot :

Hide

```
swi %>%
  ggplot(aes(y=Life.expectancy,x=Year, col=Entity)) +
  geom_line() +
  labs(title = "Life expectancy of Switzerland",
       y = "Life expectancy",
       x = "Year",
       color = "Country")
```



8) United Kingdom :

a) Dataframe :

Hide

```
uk <- data[data$Entity == "United Kingdom", ]
uk
```

	Entity <chr>	Year <int>	Life.expectancy <dbl>
1518	United Kingdom	1800	38.64976
1519	United Kingdom	1801	37.35082
1520	United Kingdom	1802	38.61560
1521	United Kingdom	1803	37.31685
1522	United Kingdom	1804	41.43692
1523	United Kingdom	1805	42.31726
1524	United Kingdom	1806	43.21611

	Entity <chr>	Year <int>	Life.expectancy <dbl>
1525	United Kingdom	1807	40.04594
1526	United Kingdom	1808	40.34779
1527	United Kingdom	1809	41.74220

1-10 of 217 rows Previous **1** 2 3 4 5 6 ... 22 Next

b) Summarize data :

Hide

```
summary(uk)
```

```
   Entity              Year      Life.expectancy
 Length:217         Min.   :1800   Min.   :37.32
 Class :character   1st Qu.:1854   1st Qu.:41.70
 Mode  :character   Median :1908   Median :50.43
                    Mean   :1908   Mean   :55.59
                    3rd Qu.:1962   3rd Qu.:70.92
                    Max.   :2016   Max.   :80.90
```

Hide

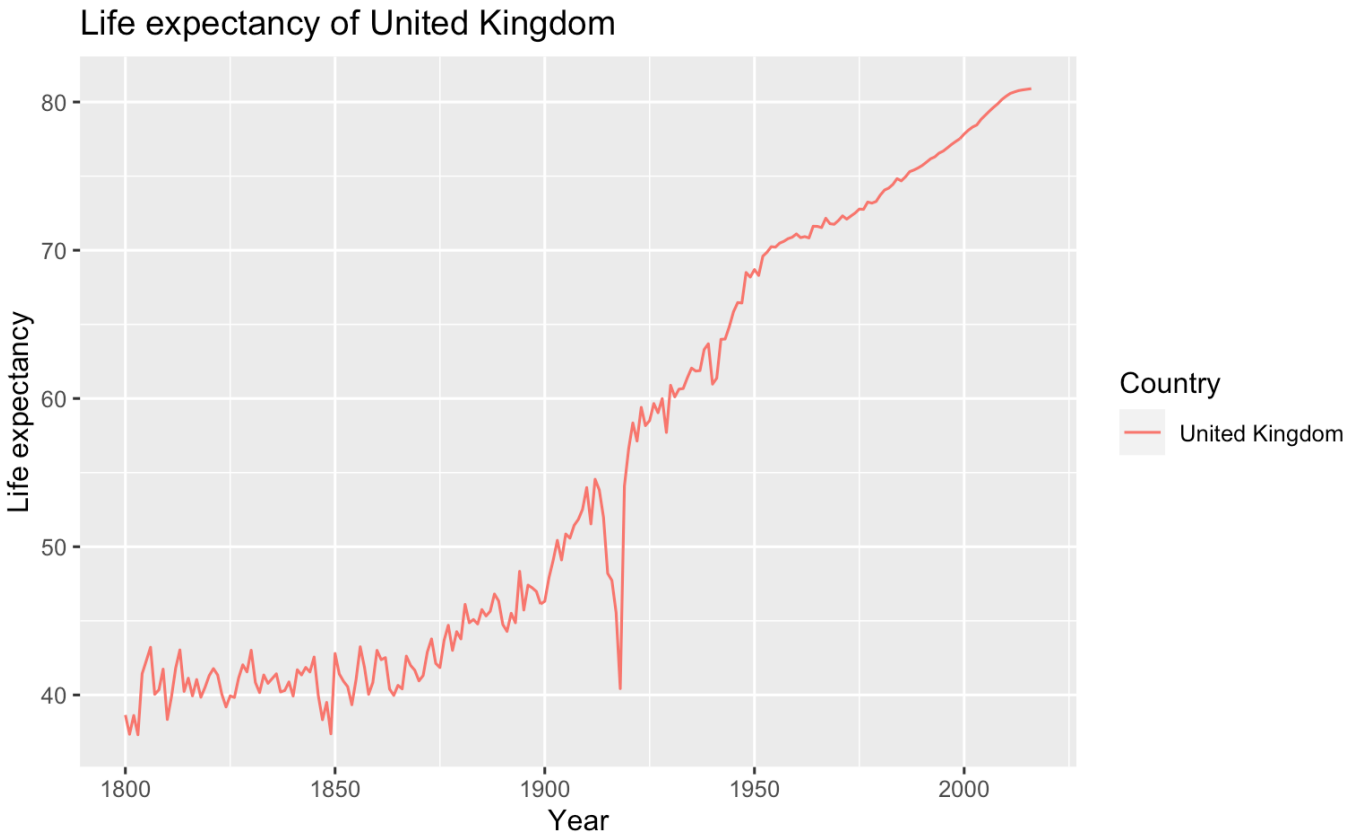
```
# Average life expectancy of UK :
m_uk <- mean(uk$Life.expectancy)
m_uk
```

```
[1] 55.59195
```

c) Plot :

Hide

```
uk %>%
  ggplot(aes(y=Life.expectancy,x=Year, col=Entity)) +
  geom_line() +
  labs(title = "Life expectancy of United Kingdom",
       y = "Life expectancy",
       x = "Year",
       color = "Country")
```



9) United States :

a) Dataframe :

Hide

```
usa <- data[data$Entity == "United States", ]
usa
```

	Entity <chr>	Year <int>	Life.expectancy <dbl>
1735	United States	1800	39.41000
1736	United States	1801	39.41000
1737	United States	1802	39.41000
1738	United States	1803	39.41000
1739	United States	1804	39.41000
1740	United States	1805	39.41000
1741	United States	1806	39.41000

	Entity <chr>	Year <int>	Life.expectancy <dbl>
1742	United States	1807	39.41000
1743	United States	1808	39.41000
1744	United States	1809	39.41000

1-10 of 217 rows Previous **1** 2 3 4 5 6 ... 22 Next

b) Summarize data :

Hide

```
summary(usa)
```

```
   Entity              Year      Life.expectancy
 Length:217         Min.   :1800   Min.   :31.00
 Class :character   1st Qu.:1854   1st Qu.:39.41
 Mode  :character   Median :1908   Median :50.60
                    Mean   :1908   Mean   :54.65
                    3rd Qu.:1962   3rd Qu.:70.21
                    Max.   :2016   Max.   :78.96
```

Hide

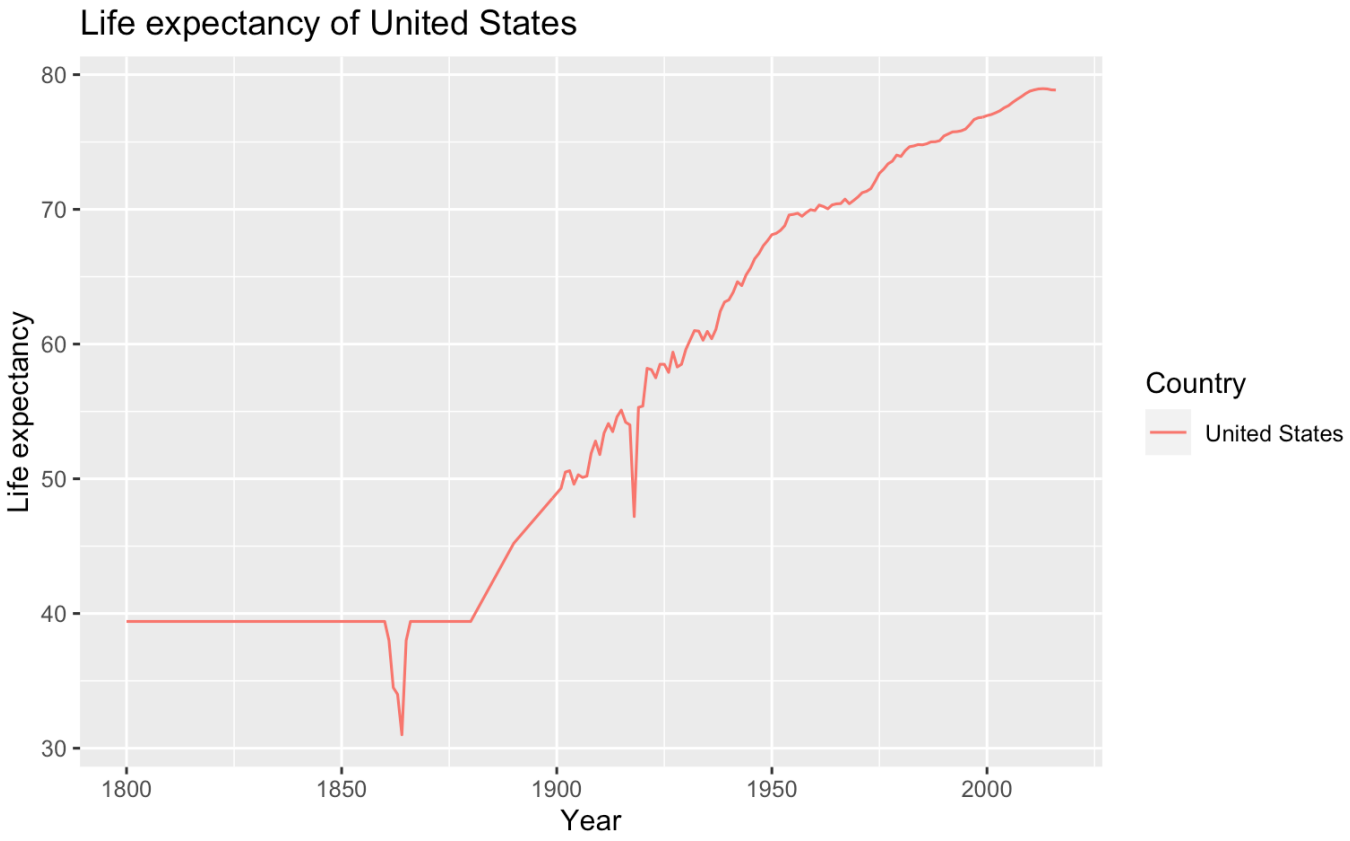
```
# Average life expectancy of USA :
m_usa <- mean(usa$Life.expectancy)
m_usa
```

```
[1] 54.65106
```

c) Plot :

Hide

```
usa %>%
  ggplot(aes(y=Life.expectancy,x=Year, col=Entity)) +
  geom_line() +
  labs(title = "Life expectancy of United States",
       y = "Life expectancy",
       x = "Year",
       color = "Country")
```



E. Summary :

a) Dataframe with all countries :

Hide

```
# Column of countries :
Country <- c("Australia","Canada","France","Germany","Italy","Japan","Swiss","UK","US
A")
n
```

```
[1] "Australia" "Canada"    "France"    "Germany"   "Italy"
[6] "Japan"     "Swiss"     "UK"        "USA"
```

Hide

```
df1 <- rbind(m_aus,m_can,m_fra,m_ger,m_ita,m_jpn,m_swi,m_uk,m_usa)
df1
```

```
          [,1]
m_aus 54.29000
m_can 55.77853
m_fra 53.69172
m_ger 52.88094
m_ita 49.22636
m_jpn 50.34571
m_swi 55.24604
m_uk  55.59195
m_usa 54.65106
```

Hide

```
df2 <- data.frame(Country,df1)
df2
```

	Country <chr>	**df1** <dbl>
m_aus	Australia	54.29000
m_can	Canada	55.77853
m_fra	France	53.69172
m_ger	Germany	52.88094
m_ita	Italy	49.22636
m_jpn	Japan	50.34571
m_swi	Swiss	55.24604
m_uk	UK	55.59195
m_usa	USA	54.65106

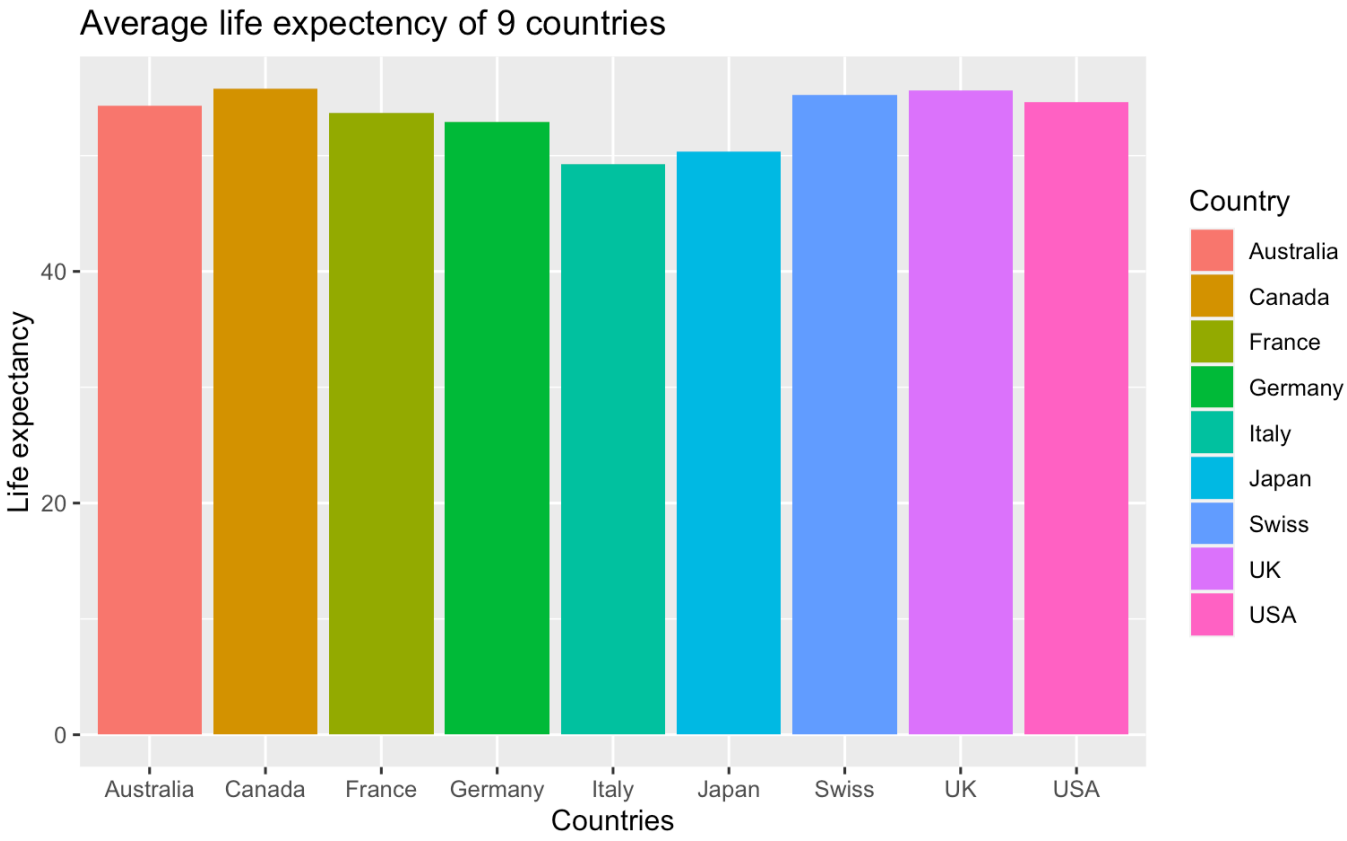
9 rows

b) Plot :

b.1) Bar plot by defautlt :

Hide

```
df2 %>%
  ggplot(aes(Country,df1,fill=Country)) +
  geom_bar(stat = "identity") +
  labs(title = "Average life expectency of 9 countries",
       y="Life expectancy",
       x="Countries")
```



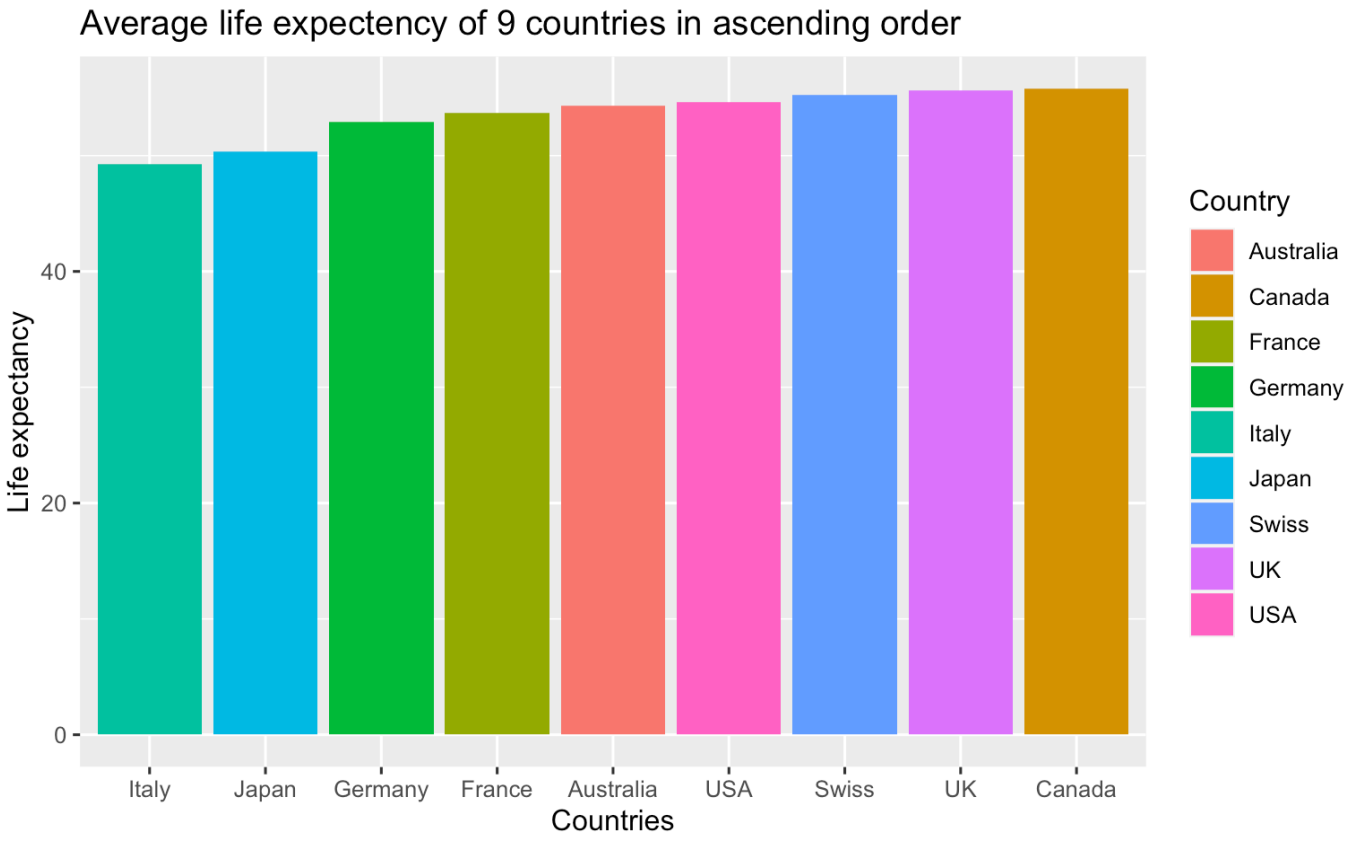
Hide

```
  # Customize the color of the bar :
  #+ scale_fill_manual(values = c("Australia" = "salmon",
                                  #"Canada" = "firebrick",
                                  #"France" = "blue",
                                  #"Germany" = "yellow",
                                  #"Italy" = "green",
                                  #"Japan" = "white",
                                  #"Swiss" = "brown",
                                  #"UK" = "red",
                                  #"USA" = "orange"))
```

b.2) Bar plot in ascending order :

Hide

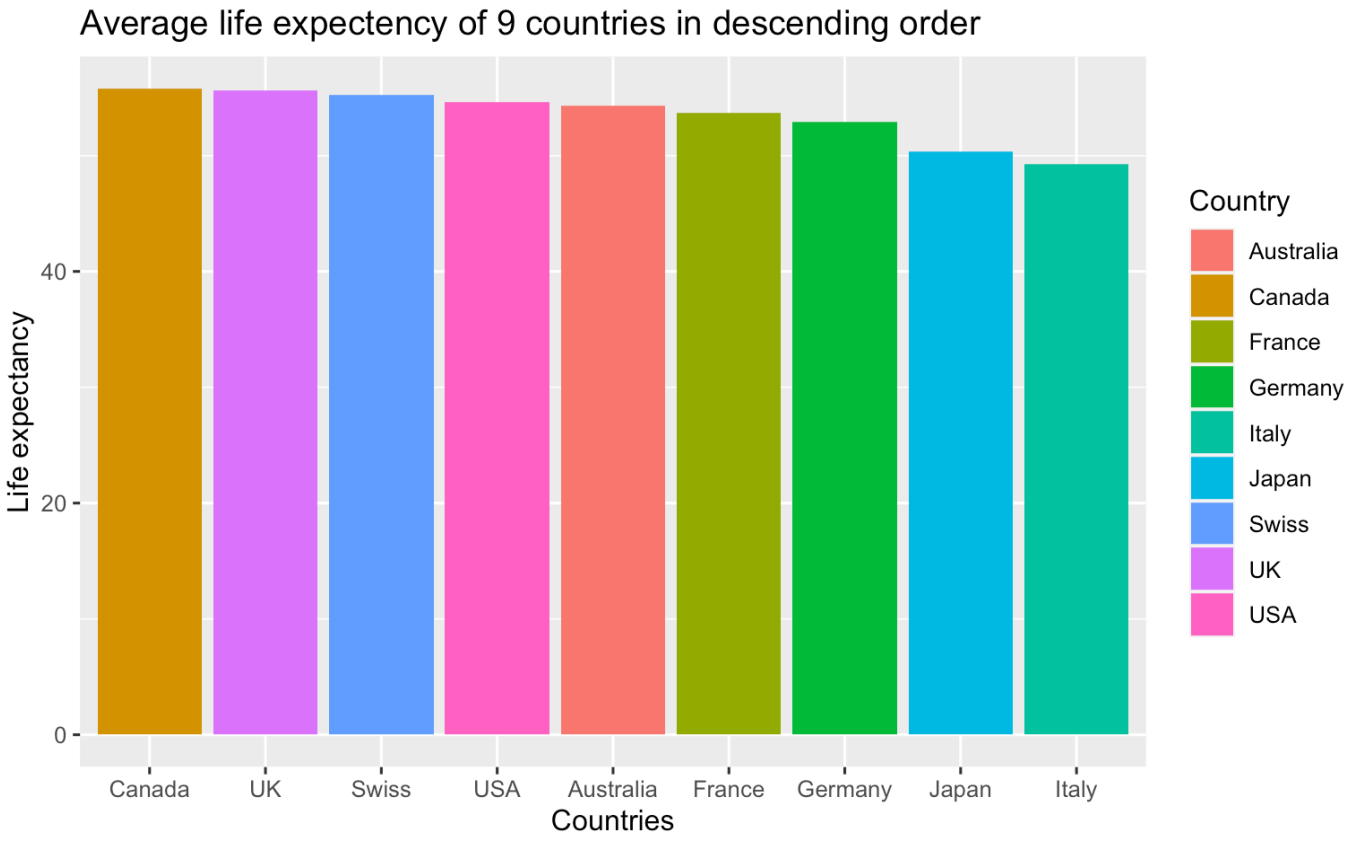
```
df2 %>%
  ggplot(aes(reorder(Country,df1),df1,fill=Country)) +
  geom_bar(stat = "identity") +
  labs(title = "Average life expectency of 9 countries in ascending order",
       y="Life expectancy",
       x="Countries")
```



b.3) Bar plot in descending order :

Hide

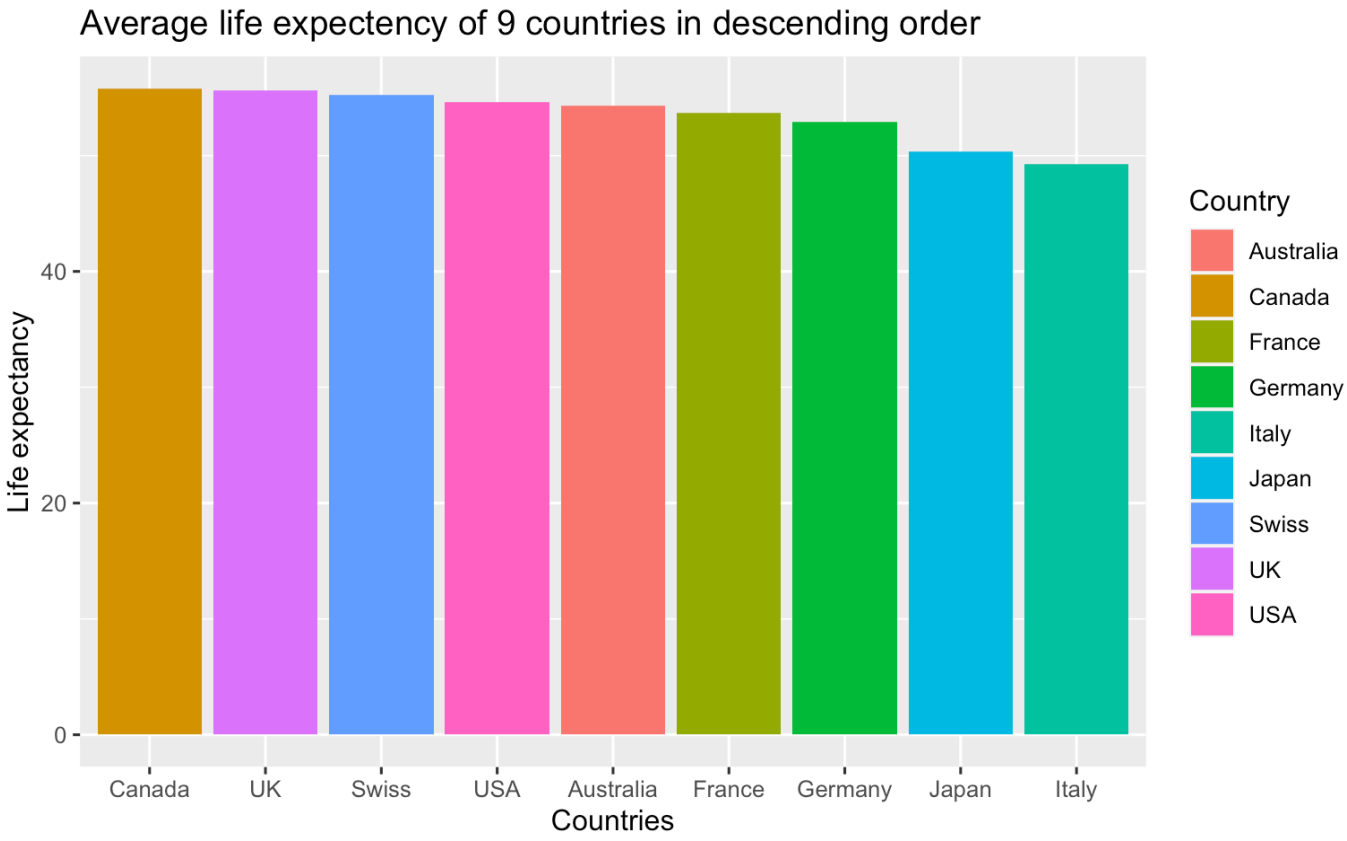
```
 # type geom_col() in stead of geom_bar(stat="identity), they are the same function.
df2 %>%
  ggplot(aes(reorder(Country,-df1),df1,fill=Country)) +
  geom_bar(stat = "identity") +
  labs(title = "Average life expectency of 9 countries in descending order",
       y="Life expectancy",
       x="Countries")
```



b.4) We can type geom_col() instead of geom_bar(stat="identity) to reduce the amount of code, they are the same function.

Hide

```
df2 %>%
  ggplot(aes(reorder(Country,-df1),df1,fill=Country)) +
  geom_col() +
  labs(title = "Average life expectency of 9 countries in descending order",
       y="Life expectancy",
       x="Countries")
```


Average life expectency of 9 countries in descending order
Life expectancy
0
20
40
Canada
UK
Swiss
USA
Australia
France
Germany
Japan
Italy
Countries
Country
Australia
Canada
France
Germany
Italy
Japan
Swiss
UK
USA