

# DATA 607 Project 2 Part 2

Subhalaxmi Rout

03/08/2020

## UNICEF dataset on Under 5 Mortality

Discussion thread created by : Samuel Bellows

### 1. Introduction

UNICEF dataset that gives the under 5 mortality for many countries across the years 1950-2015. The problem is that the year variable is spread out into 65 different columns, 1 for each year, that need to be gathered into 1 column. In order to make this dataset tiny we would gather the year columns into one column until we had a 3 column dataset of Country name, Year, and Mortality.

### 2. Load library

```
#install.packages("dplyr")
#install.packages("tidyr")
#install.packages("ggplot2")
#install.packages("DT")
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(DT)
```

### 3. Data load and cleaning

Data is stored in the **Github** and loaded data from Github to Rstudio using **read.csv()** method.

```
# read csv file data in a variable using read.csv()
data <- read.csv('https://raw.githubusercontent.com/SubhalaxmiRout002/Data-607-Project-2-Dataset-1/master/data.csv')

# convert data to data frame
data <- data.frame(data)

# display data using datatable
datatable(data,options = list(scrollX = TRUE, paging=TRUE,fixedHeader=TRUE))
```

Show  entries Search:

	CountryName	U5MR.1950	U5MR.1951	U5MR.1952	U5MR.1953	U5MR.1954	U5MR.1955	U5MR.1956
1	Afghanistan							
2	Albania							
3	Algeria					251	249.9	
4	Andorra							
5	Angola							
6	Antigua & Barbuda							
7	Azerbaijan							
8	Argentina							
9	Australia	31.6	31	30.3	29.5	28.7	28	
10	Austria					55	53.4	

Showing 1 to 10 of 196 entries Previous  2 3 4 5 ... 20 Next

#### 3.1 Gather year from 1950 to 2015

This dataset year has given from 1950 to 2015. Each year mentioned as a column. Using **tidyr** convert these columns to Year column.

```
# using gather() convert column to row
data <- data %>% gather(Year, Mortality, U5MR.1950:U5MR.2015, na.rm = TRUE)

# Uremove "U5MR" from the name
data$Year <- sub('U5MR.', '', data$Year)

# arrange Mortality by desc order
data <- data %>% arrange(desc(Mortality))

# display data using datatable
datatable(data,options = list(scrollX = TRUE, paging=TRUE,fixedHeader=TRUE))
```

Show 10 entries

Search:

	CountryName	Year	Mortality
1	Mali	1963	443.5
2	Mali	1964	435.8
3	Mali	1965	428.5
4	Mali	1966	421.8
5	Mali	1967	415.4
6	Mali	1968	409.8
7	Yemen	1963	406.1
8	Mali	1969	404.6
9	Mali	1970	400
10	Yemen	1964	397.7

Showing 1 to 10 of 10,244 entries

Previous 1 2 3 4 5 ... 1025 Next

## 4. Analysis

Mortality based on year:

- Year: 1950, Highest/lowest mortality
- Year: 2015 Highest/lowest mortality

### 4.1 Highest mortality in the year 1950

```
# apply filter on year = 1950
data1 <- data %>% filter(Year == "1950")

# display data using datatable
datatable(data1,options = list(scrollX = TRUE, paging=TRUE,fixedHeader=TRUE))
```

Show 10 entries

Search:

	CountryName	Year	Mortality
1	Burkina Faso	1950	389.7
2	Iraq	1950	364.3
3	Benin	1950	348.2
4	Senegal	1950	346.2
5	Togo	1950	324.4
6	Mauritania	1950	316.4
7	Jordan	1950	275.1
8	Ghana	1950	256.7
9	Swaziland	1950	240.1
10	Thailand	1950	210

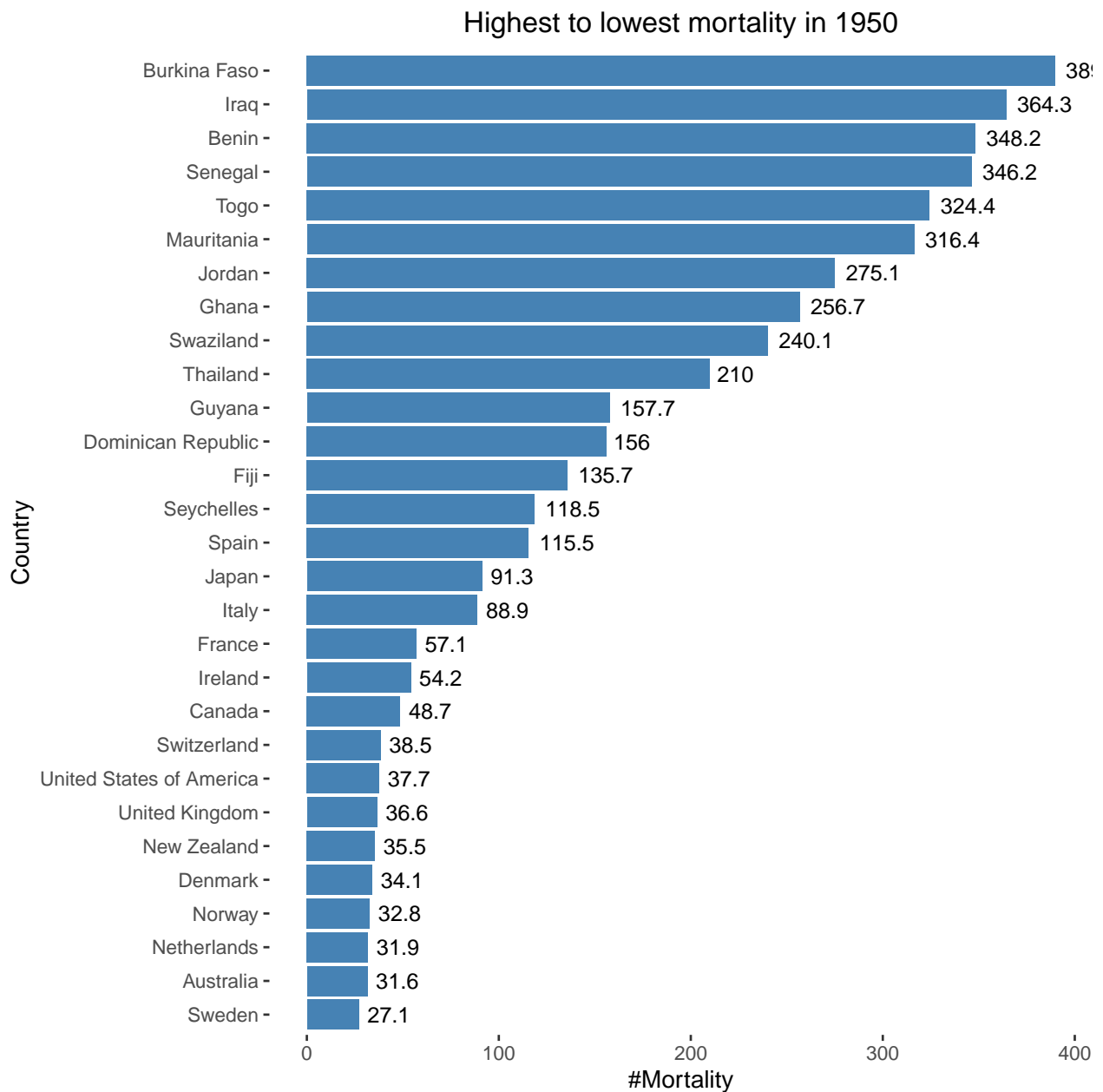
Showing 1 to 10 of 29 entries

Previous 1 2 3 Next

## 4.2 Highest to lowest mortality in 1950

```
data1 <- data.frame(data1)

ggplot(data = data1, mapping = aes(x = CountryName, y = Mortality)) +
  geom_bar(aes(reorder(CountryName,Mortality),Mortality),stat = "identity",fill = "steelblue") +
  coord_flip() + xlab("Country") + ylab("#Mortality") +
  ggtitle("Highest to lowest mortality in 1950") +
  theme(plot.title = element_text(hjust = 0.5),panel.background = element_rect(fill = "white", color = "black"),
  geom_text(aes( y = Mortality,label=Mortality), hjust = -0.20, color="black", size=3.5)
```



### 4.3 Highest mortality in 2015

```
# apply filter on year = 2015
data2 <- data %>% filter(Year == "2015")

# arrange Mortality by desc
data2 <- data2 %>% arrange(desc(Mortality))

# display data using datatable
datatable(data2,options = list(scrollX = TRUE, paging=TRUE,fixedHeader=TRUE))
```

Show  entries Search:

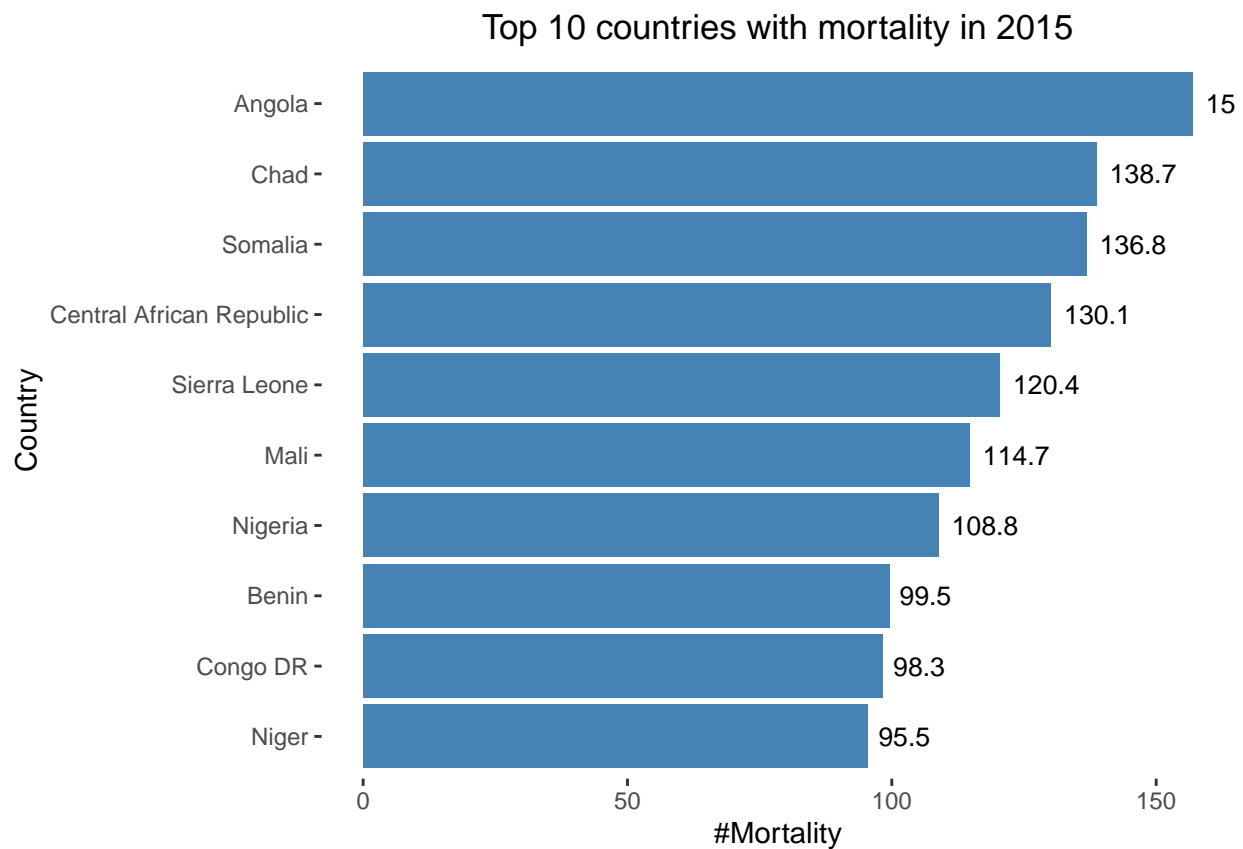
	CountryName	Year	Mortality
1	Angola	2015	156.9
2	Chad	2015	138.7
3	Somalia	2015	136.8
4	Central African Republic	2015	130.1
5	Sierra Leone	2015	120.4
6	Mali	2015	114.7
7	Nigeria	2015	108.8
8	Benin	2015	99.5
9	Congo DR	2015	98.3
10	Niger	2015	95.5

Showing 1 to 10 of 195 entries Previous  2 3 4 5 ... 20 Next

### 4.4 Top 10 countries with mortality in 2015

```
data3 <- data.frame(head(data2, 10))

ggplot(data = data3, mapping = aes(x = CountryName, y = Mortality)) +
  geom_bar(aes(reorder(CountryName,Mortality),Mortality),stat = "identity",fill = "steelblue") +
  coord_flip() + xlab("Country") + ylab("#Mortality") +
  ggtitle("Top 10 countries with mortality in 2015") +
  theme(plot.title = element_text(hjust = 0.5),panel.background = element_rect(fill = "white", color = "black"),
  geom_text(aes( y = Mortality,label=Mortality), hjust=-0.20, color="black", size=3.5))
```

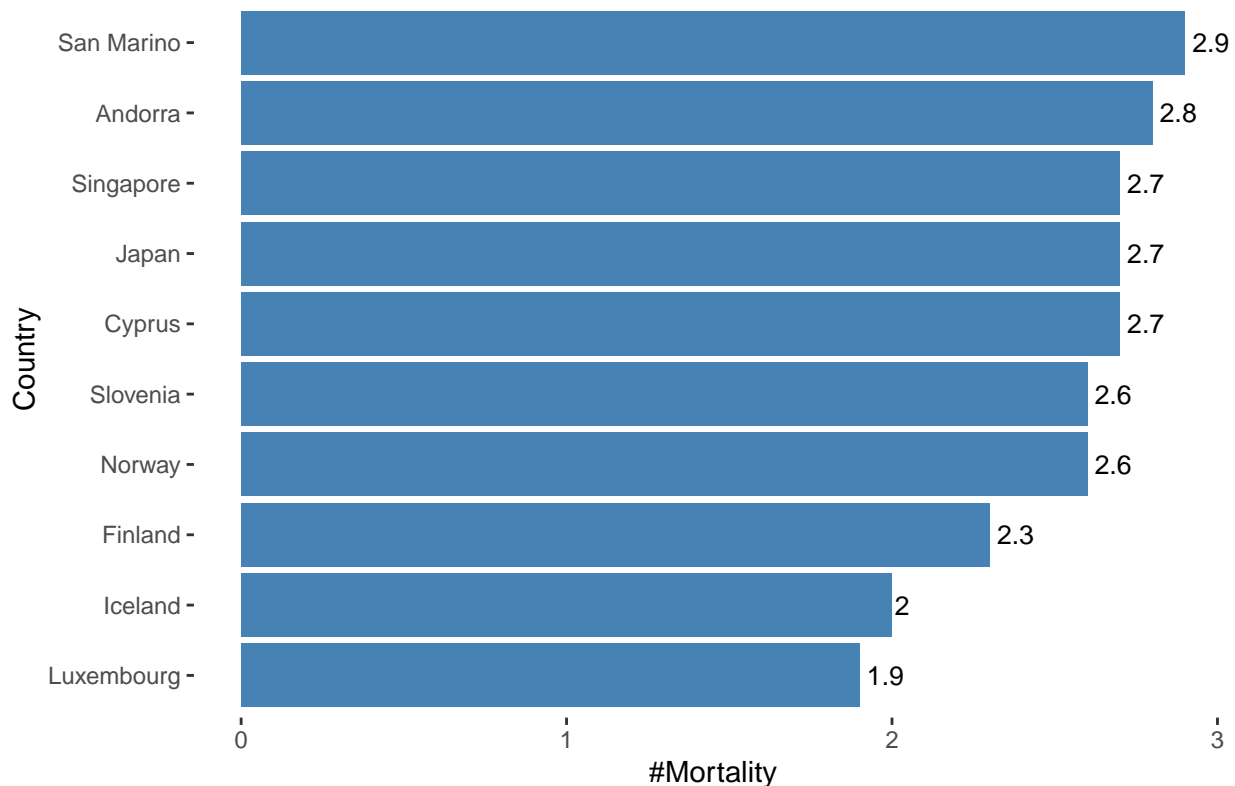


#### 4.5 Bottom 10 countries with mortality in 2015

```
# last 10 row data stored in the variable
data4 <- data.frame(tail(data2, 10))

ggplot(data = data4, mapping = aes(x = CountryName, y = Mortality)) +
  geom_bar(aes(reorder(CountryName,Mortality),Mortality),stat = "identity",fill = "steelblue") +
  coord_flip() + xlab("Country") + ylab("#Mortality") +
  ggtitle("Bottom 10 countries with mortality in 2015") +
  theme(plot.title = element_text(hjust = 0.5),panel.background = element_rect(fill = "white", color = "black")) +
  geom_text(aes( y = Mortality,label=Mortality), hjust=-0.20,color="black", size=3.5)
```

### Bottom 10 countries with mortality in 2015



#### 4.6 Average mortality of the 10 developed countries

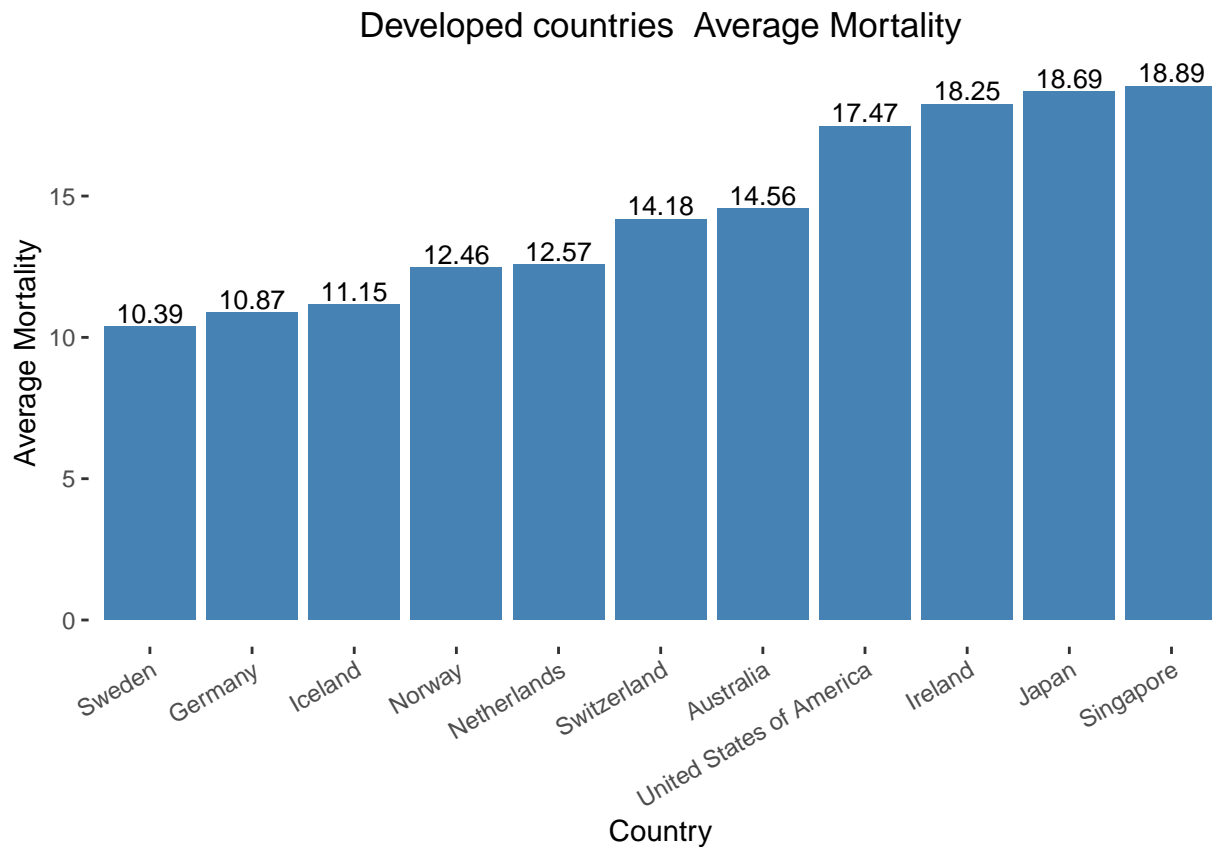
```
# stored data in data frame
data5 <- data.frame(data)

# calculate mean and group by using country
data5 <- data5 %>% group_by(CountryName) %>% summarise(round(mean(Mortality),2))
data5 <- mutate(data5,Avg_mortality = `round(mean(Mortality), 2)`)

# select country name and average mortality
data5 <- select(data5, CountryName, Avg_mortality)

#apply filter on country
data5 <- data5 %>% filter(grepl('United States of America|Japan|Norway|Switzerland|Australia|Ireland|Ge

# 10 developed countries average mortality
ggplot(data = data5, mapping = aes(x = CountryName, y = Avg_mortality)) +
  geom_bar(aes(reorder(CountryName,Avg_mortality),Avg_mortality),stat = "identity",fill = "steelblue") +
  xlab("Country") + ylab("Average Mortality") +
  ggtitle("Developed countries Average Mortality") +
  theme(plot.title = element_text(hjust = 0.5)) +
  geom_text(aes( y = Avg_mortality,label=Avg_mortality), vjust = - 0.2, color="black", size=3.5) +
  theme(axis.text.x=element_text(angle=30,hjust=1),panel.background = element_rect(fill = "white", color=
```



#### 4.7 Mortality of United States from 1950 to 2015

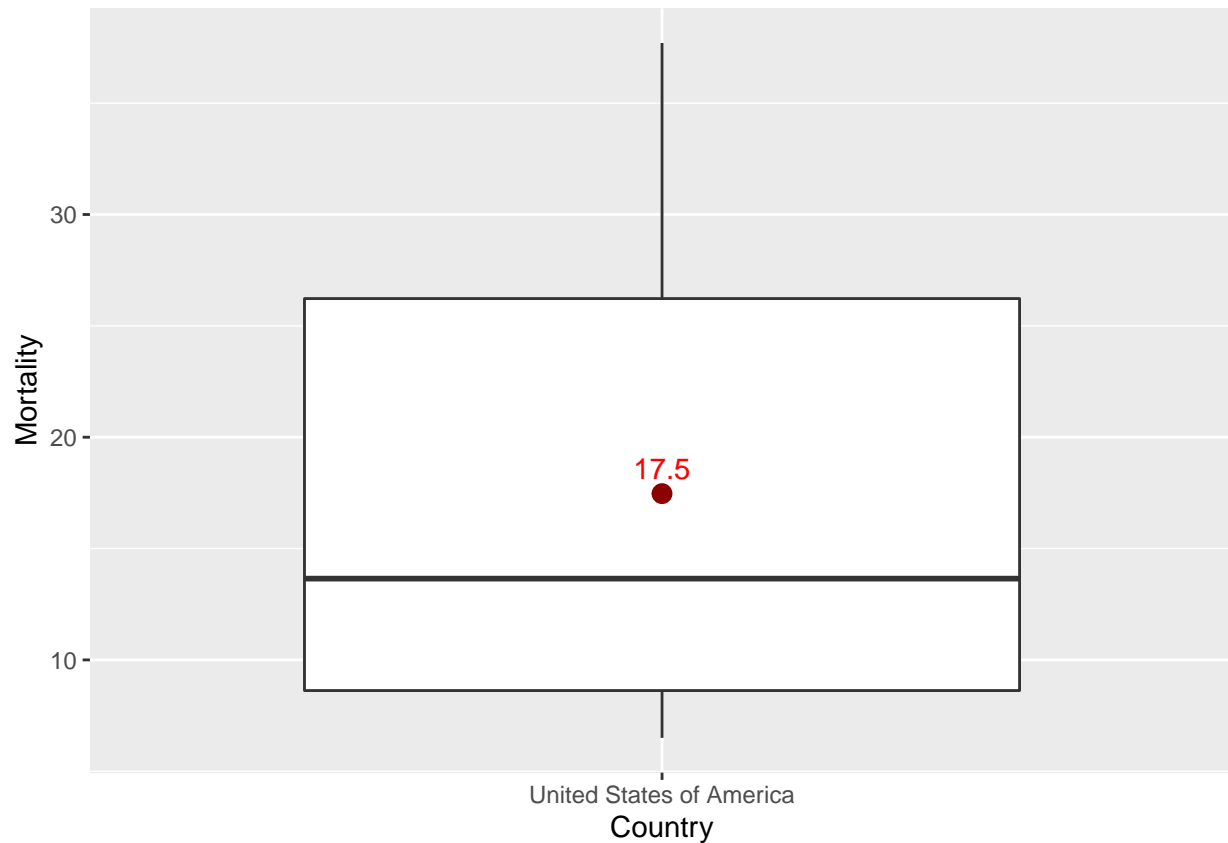
```
# store data in data frame
data6 <- data.frame(data)

# apply filter on Contry name = US
data6 <- data6 %>% filter(CountryName == "United States of America")

ggplot(data6, aes(x = CountryName, y = Mortality)) +
  geom_boxplot() + xlab("Country") +

stat_summary(fun=mean, colour="darkred", geom="point", size=3, show.legend = FALSE) +
stat_summary(fun=mean, colour="red", geom="text", show.legend = FALSE,
  vjust=-0.7, aes( label=round(..y.., digits=1)))
```





```
# display mean, median, upper hinge, lower hindge
summary(data6$Mortality)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      6.500   8.625  13.650  17.467  26.225  37.700
```

## 5. Conclusion

The plot 4.2, plot 4.4, and plot 4.5 shows mortality goes down year over year.

- In the year 1950, the African continent countries have more mortality.
- In the year 2015, developed countries have fewer mortality as compared to undeveloped or developing countries.
- The average mortality for 10 developed countries: Sweden has the least average and Singapore has the highest average.
- The boxplot (plot 4.7) shows the mortality of the United States from 1950 to 2015.
- The mortality of the United States has gone down year over year.