

# Life\_expectancy

```
library(rvest)
library(ggplot2)
library(dplyr)
library(maps)
library(scales)
```

## Import data from wikipedia

```
life_expect <- read_html("https://en.wikipedia.org/wiki/List_of_U.S._states_by_life_expectancy")
```

## Convert data into a table

```
life_expect <- life_expect %>%
  html_nodes(xpath = "//table") %>%
  .[[3]] %>%
  html_table(fill = T)
```

## Clean life\_expect data frame by selecting required columns

```
life_expect <- life_expect[, c(3:6)]
```

## Rename the columns

```
names(life_expect)[c(1, 2, 3, 4)] <- c("region", "life_expect_all", "life_expect_White", "life_expect_Black")
```

## Convert black and white expectancy to numeric variables:

```
life_expect <- life_expect %>%
  mutate(
    life_expect_White = as.numeric(life_expect_White),
    life_expect_Black = as.numeric(life_expect_Black),
    region = tolower(region)
  )
```

## Compute the difference between black and white life expect:

```
life_expect <- life_expect %>%
  mutate(life_expect_diff = life_expect_White - life_expect_Black)
```

How many regions(states) where African Americans live more than white Americans:

```
life_expect %>%
  filter(life_expect_diff < 0) %>%
  arrange(desc(life_expect_diff))
```

##	region	life_expect_all	life_expect_White	life_expect_Black
## 1	south dakota	79.5	80.4	81.7
## 2	minnesota	81.1	80.4	82.5
## 3	maine	79.2	79.1	81.8
## 4	idaho	79.5	79.4	83.3
## 5	vermont	80.5	80.4	84.4
## 6	north dakota	79.5	80.2	84.2
## 7	montana	78.5	79.1	83.4
## 8	wyoming	78.3	78.4	83.5
## 9	new hampshire	80.3	80.1	86.8
##	life_expect_diff			
## 1	-1.3			
## 2	-2.1			
## 3	-2.7			
## 4	-3.9			
## 5	-4.0			
## 6	-4.0			
## 7	-4.3			
## 8	-5.1			
## 9	-6.7			

There are 9 states where life\_expect is larger for African Americans with the highest value for New Hampshire.

How many regions(states) where life expectancy is larger for white Americans than African Americans:

```
life_expect %>%
  filter(life_expect_diff > 0) %>%
  arrange(desc(life_expect_diff))
```

##	region	life_expect_all	life_expect_White
## 1	district of columbia	76.5	84.3
## 2	rhode island	79.9	79.7
## 3	wisconsin	80.0	80.3
## 4	nebraska	79.8	80.0
## 5	utah	80.2	80.1
## 6	illinois	79.0	79.3
## 7	michigan	78.2	79.0
## 8	pennsylvania	78.5	78.9
## 9	kansas	78.7	78.8
## 10	new jersey	80.3	80.3
## 11	california	80.8	79.8
## 12	iowa	79.7	79.8
## 13	louisiana	75.7	76.7
## 14	ohio	77.8	78.1
## 15	virginia	79.0	79.4

## 16	arkansas	76.0	76.3
## 17	maryland	78.8	79.4
## 18	indiana	77.6	77.7
## 19	new mexico	78.4	79.0
## 20	south carolina	77.0	77.8
## 21	tennessee	76.3	76.7
## 22	mississippi	75.0	76.1
## 23	texas	78.5	78.0
## 24	north carolina	77.8	78.3
## 25	colorado	80.0	80.2
## 26	missouri	77.5	77.7
## 27	arizona	79.6	79.8
## 28	florida	79.4	79.1
## 29	connecticut	80.8	81.0
## 30	oklahoma	75.9	76.0
## 31	delaware	78.4	78.6
## 32	new york	80.5	80.5
## 33	alabama	75.4	76.0
## 34	georgia	77.2	77.6
## 35	west virginia	75.4	75.4
## 36	kentucky	76.0	76.0
## 37	washington	79.9	79.7
## 38	oregon	79.5	79.2
## 39	massachusetts	80.5	80.4
## 40	hawaii	81.3	81.2
## 41	nevada	78.1	76.7
## 42	alaska	78.3	79.4
##	life_expect_Black	life_expect_diff	
## 1	71.6	12.7	
## 2	71.6	8.1	
## 3	74.0	6.3	
## 4	73.9	6.1	
## 5	74.3	5.8	
## 6	73.7	5.6	
## 7	73.4	5.6	
## 8	73.4	5.5	
## 9	73.6	5.2	
## 10	75.5	4.8	
## 11	75.1	4.7	
## 12	75.3	4.5	
## 13	72.4	4.3	
## 14	73.9	4.2	
## 15	75.3	4.1	
## 16	72.2	4.1	
## 17	75.5	3.9	
## 18	73.8	3.9	
## 19	75.2	3.8	
## 20	74.0	3.8	
## 21	72.9	3.8	
## 22	72.4	3.7	
## 23	74.4	3.6	
## 24	74.7	3.6	
## 25	76.7	3.5	
## 26	74.2	3.5	

```
## 27          76.5          3.3
## 28          75.8          3.3
## 29          77.8          3.2
## 30          72.8          3.2
## 31          75.4          3.2
## 32          77.4          3.1
## 33          72.9          3.1
## 34          74.7          2.9
## 35          72.8          2.6
## 36          73.5          2.5
## 37          77.5          2.2
## 38          77.2          2.0
## 39          78.8          1.6
## 40          79.7          1.5
## 41          75.9          0.8
## 42          79.3          0.1
```

There are 42 states where life\_expect is larger for white American with the highest value for District of Columbia.

## Load the map data:

```
states = map_data("state")
str(states)

## 'data.frame':  15537 obs. of  6 variables:
## $ long      : num  -87.5 -87.5 -87.5 -87.5 -87.6 ...
## $ lat       : num   30.4 30.4 30.4 30.3 30.3 ...
## $ group     : num   1 1 1 1 1 1 1 1 1 1 ...
## $ order     : int   1 2 3 4 5 6 7 8 9 10 ...
## $ region    : chr   "alabama" "alabama" "alabama" "alabama" ...
## $ subregion : chr   NA NA NA NA ...
```

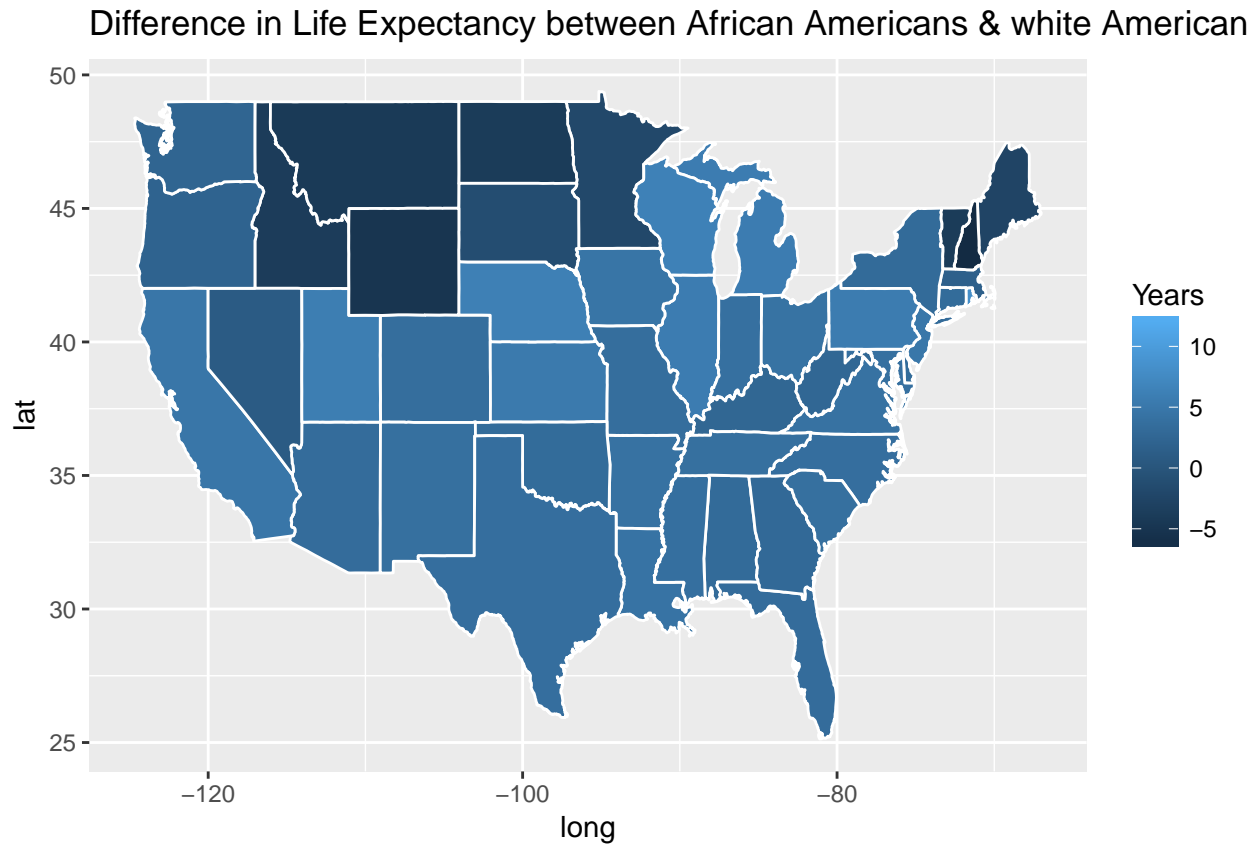
## Merge the two datasets:

```
states <- merge(states, life_expect, by = "region", all.x = T)
str(states)

## 'data.frame':  15537 obs. of  10 variables:
## $ region      : chr   "alabama" "alabama" "alabama" "alabama" ...
## $ long        : num  -87.5 -87.5 -87.5 -87.5 -87.6 ...
## $ lat         : num   30.4 30.4 30.4 30.3 30.3 ...
## $ group       : num   1 1 1 1 1 1 1 1 1 1 ...
## $ order       : int   1 2 3 4 5 6 7 8 9 10 ...
## $ subregion   : chr   NA NA NA NA ...
## $ life_expect_all : num  75.4 75.4 75.4 75.4 75.4 75.4 75.4 75.4 75.4 75.4 ...
## $ life_expect_White: num  76 76 76 76 76 76 76 76 76 76 ...
## $ life_expect_Black: num  72.9 72.9 72.9 72.9 72.9 72.9 72.9 72.9 72.9 72.9 ...
## $ life_expect_diff : num   3.1 3.1 3.1 3.1 3.1 ...
```

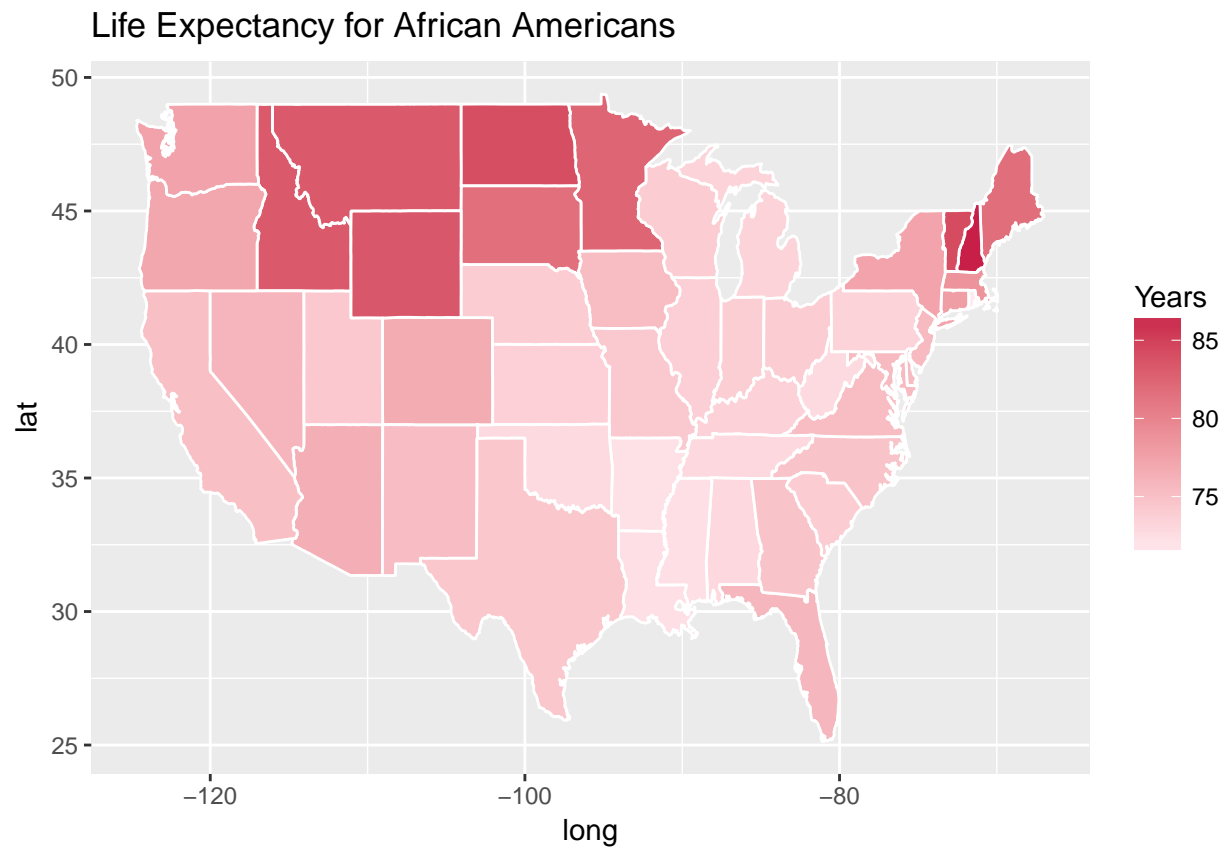
## Map of difference in life expectancy between White Americans & African Americans

```
ggplot(data = states, aes(x = long, y = lat, group = group, fill = life_expect_diff)) + geom_polygon() +
  scale_fill_gradient(name = "Years", low = "#132B43", high = "#56B1F7", guide = "colorbar", na.value = NA) +
  labs(title = "Difference in Life Expectancy between African Americans & white Americans")
```



## Map of African Americans life expectancy

```
ggplot(data = states, aes(x = long, y = lat, group = group, fill = life_expect_Black)) + geom_polygon() +
  scale_fill_gradient(name = "Years", low = "#ffe8ee", high = "#c81f49", guide = "colorbar", na.value = NA) +
  labs(title = "Life Expectancy for African Americans")
```



Map of White Americans life expectancy

```
ggplot(data = states, aes(x = long, y = lat, group = group, fill = life_expect_White)) + geom_polygon(c
  scale_fill_gradient(name = "Years", low = "#ffe8ee", high = "#c81f49", guide = "colorbar", na.value =
  labs(title = "Life Expectancy for White Americans")
```

A choropleth map of the contiguous United States showing the number of years since the last fire. The map is color-coded using a red gradient, where darker shades represent a longer time since the last fire. A legend on the right, titled "Years", shows a gradient from light pink (76) to dark red (84). The map includes latitude and longitude axes, with latitude ranging from 25 to 50 and longitude from -120 to -80.

```
ggplot(data = states, aes(x = long, y = lat, group = group, fill = life_expect_all)) + geom_polygon(col = "black",  
  scale_fill_gradient(name = "Years", low = "#132B43", high = "#56B1F7", guide = "colorbar", na.value = "white"),  
  labs(title = "Life Expectancy in the USA")
```

