

Life Expectancy Analysis

VMC

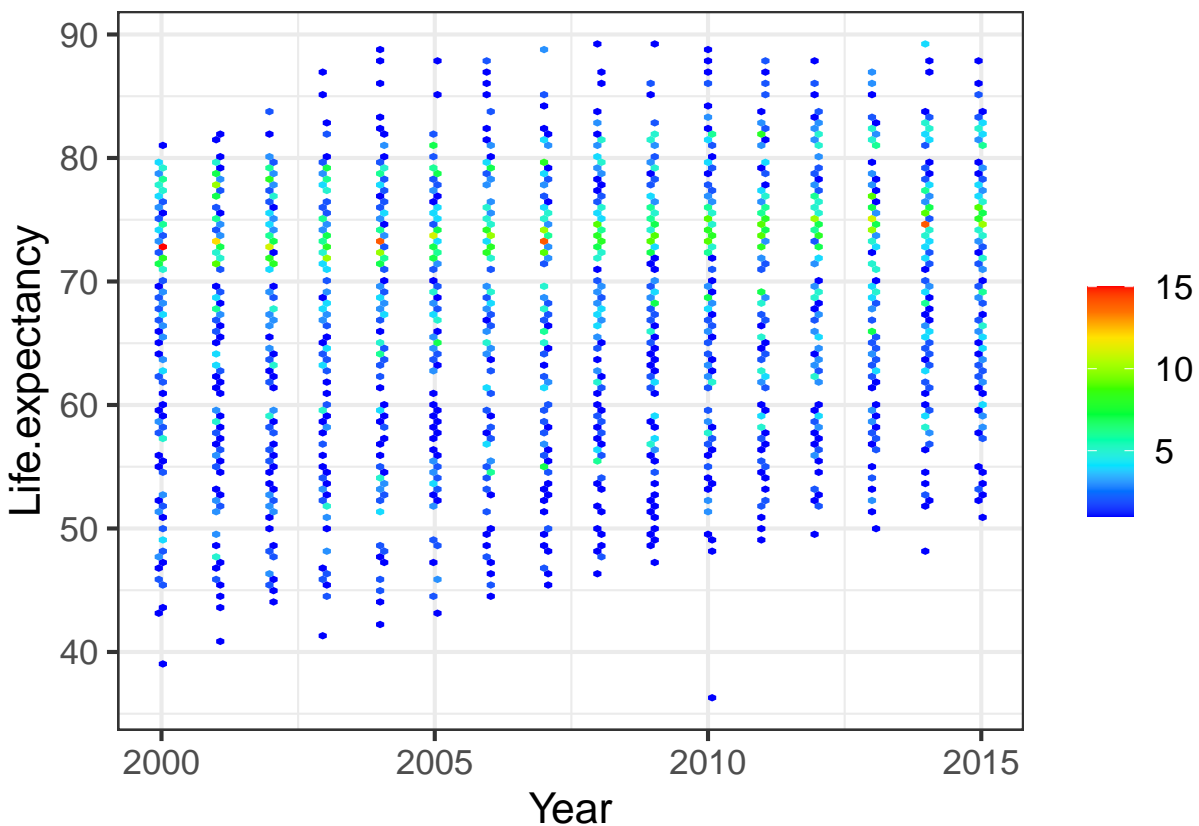
4/27/2021

```
library(rmarkdown)
library(ggplot2)
life_expectancy_data <- read.csv("Life Expectancy Data.csv", header = TRUE)
```

Scatter Plots

```
ggplot(life_expectancy_data) + geom_hex(aes(Year, Life.expectancy), bins = 100) +
  scale_fill_gradientn("", colours = rev(rainbow(10, end = 4/6)))
```

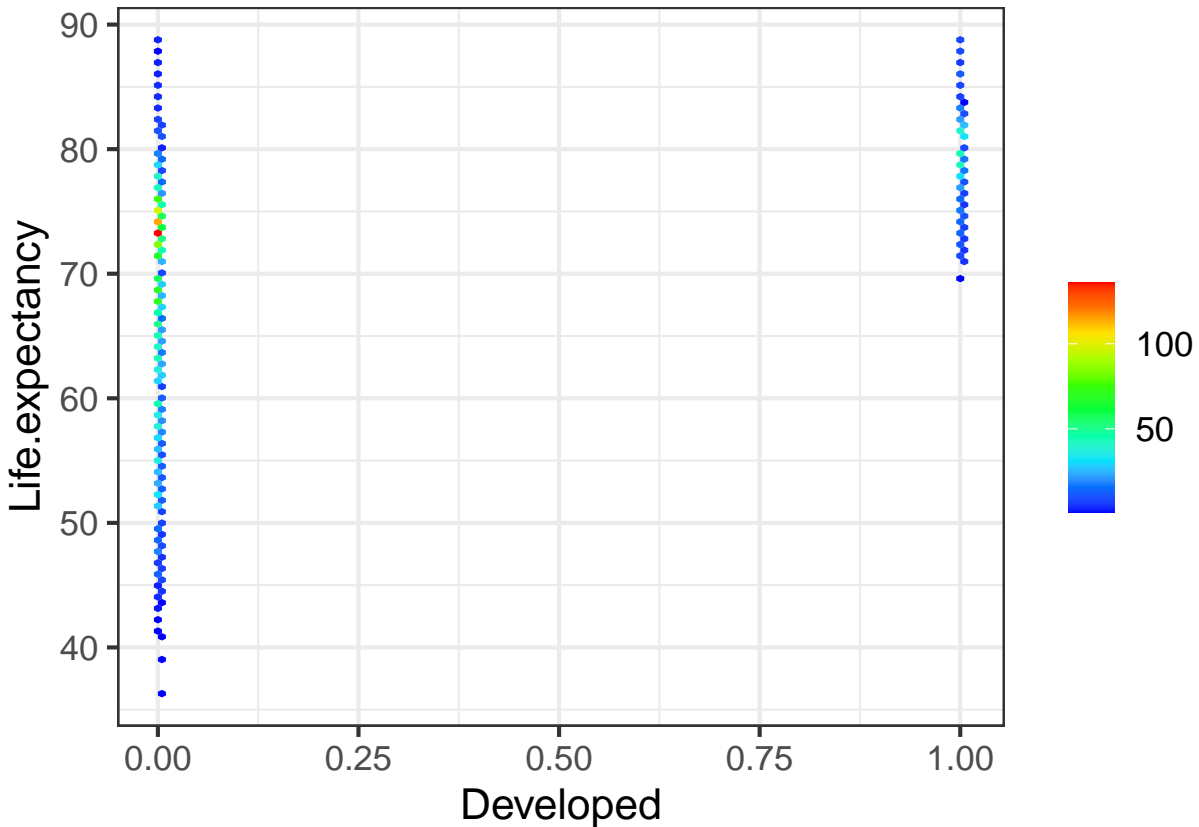
```
## Warning: Removed 10 rows containing non-finite values (stat_binhex).
```



```
life_expectancy_data$Developed <- as.integer (as.logical(life_expectancy_data$Status == "Developed"))

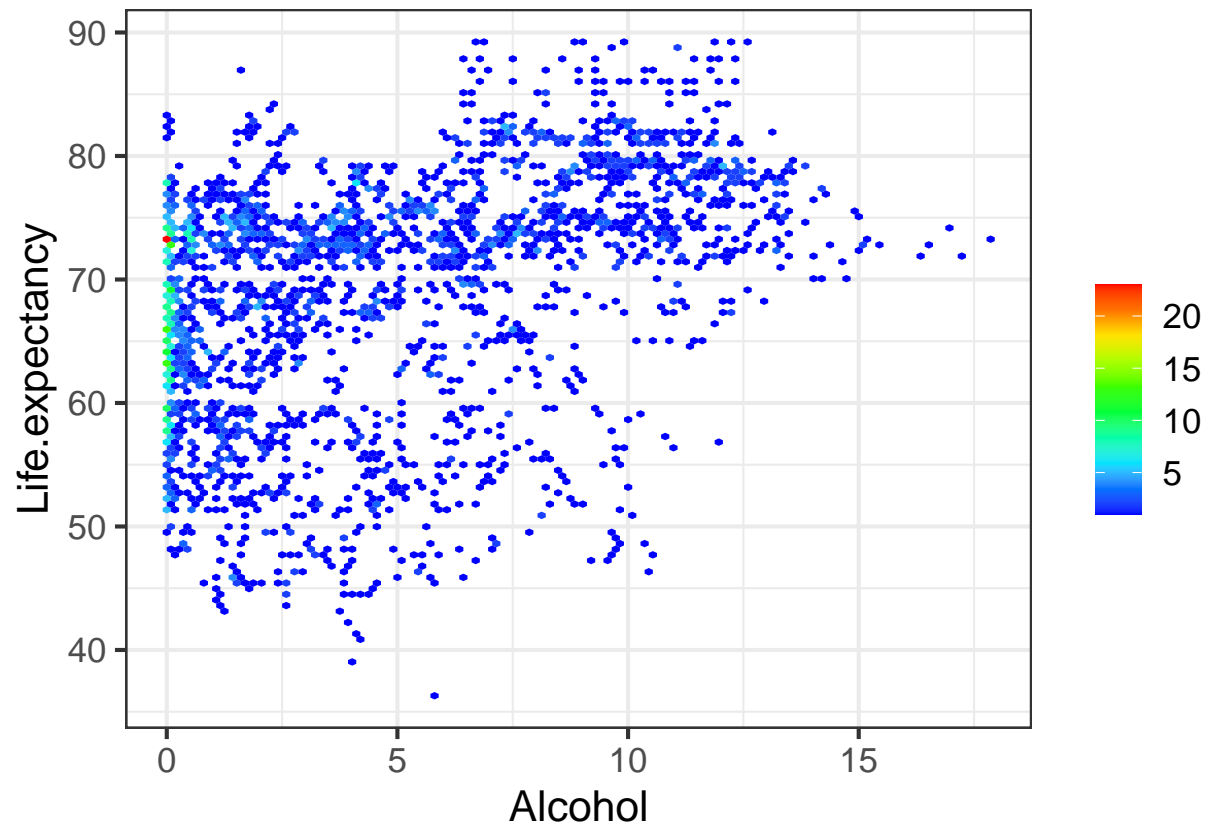
ggplot (life_expectancy_data) + geom_hex(aes(Developed, Life.expectancy), bins = 100) +
  scale_fill_gradientn("", colours = rev(rainbow(10, end = 4/6)))
```

Warning: Removed 10 rows containing non-finite values (stat_binhex).



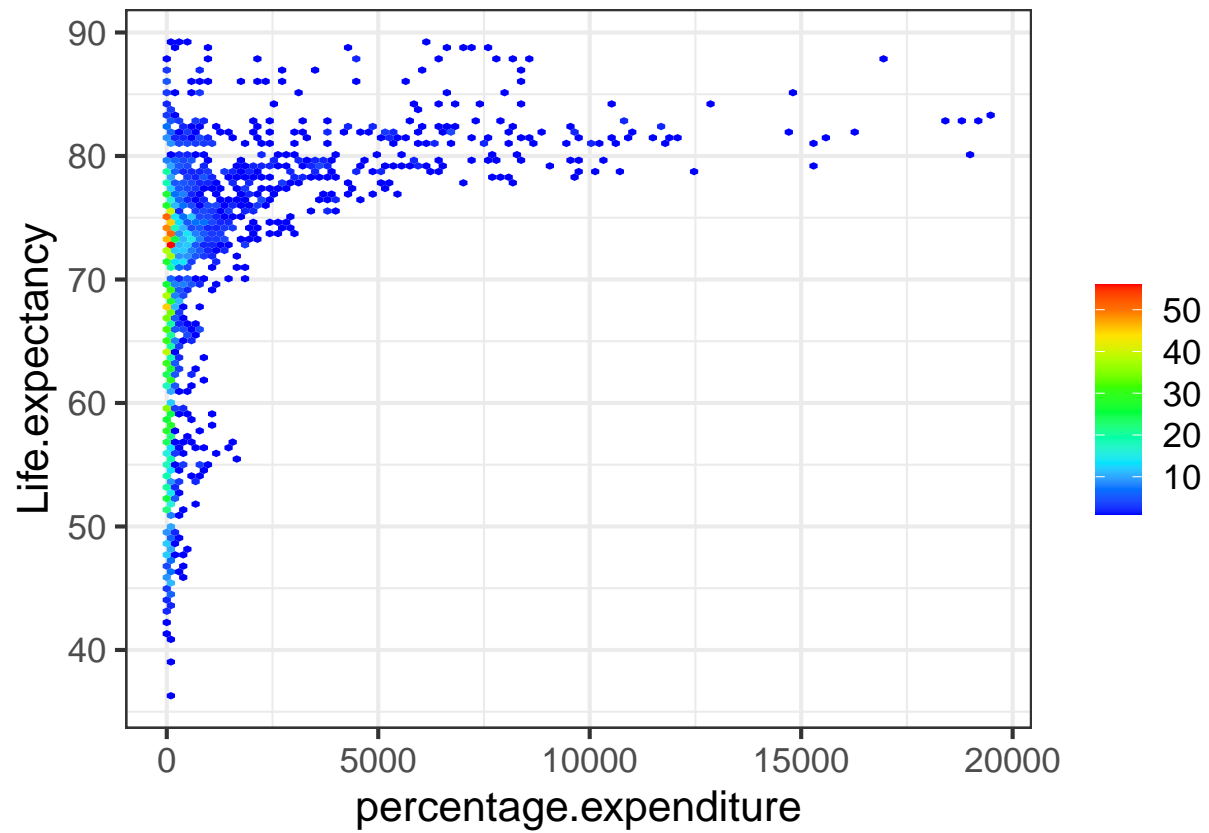
```
ggplot (life_expectancy_data) + geom_hex(aes(Alcohol, Life.expectancy), bins = 100) +
  scale_fill_gradientn("", colours = rev(rainbow(10, end = 4/6)))
```

Warning: Removed 203 rows containing non-finite values (stat_binhex).



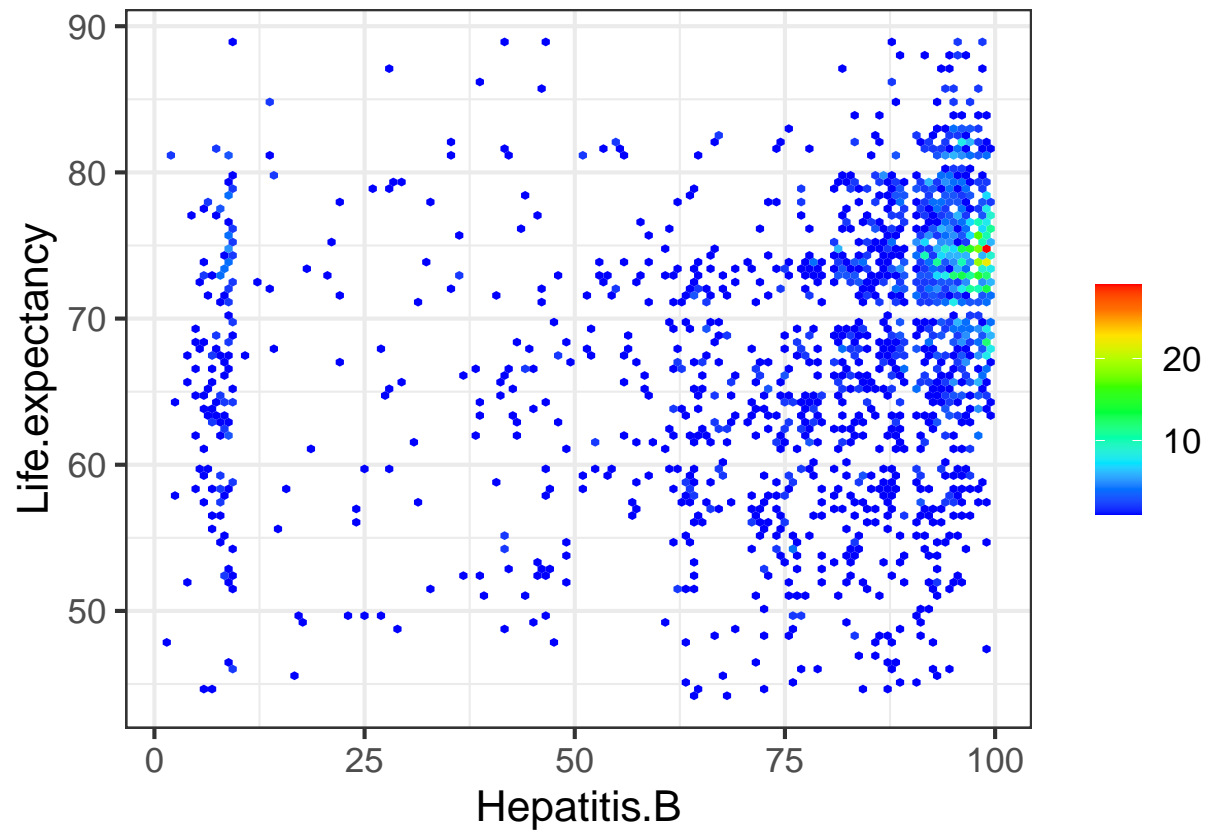
```
ggplot (life_expectancy_data) + geom_hex(aes(percentage.expenditure, Life.expectancy), bins = 100) +
  scale_fill_gradientn("", colours = rev(rainbow(10, end = 4/6)))
```

```
## Warning: Removed 10 rows containing non-finite values (stat_binhex).
```



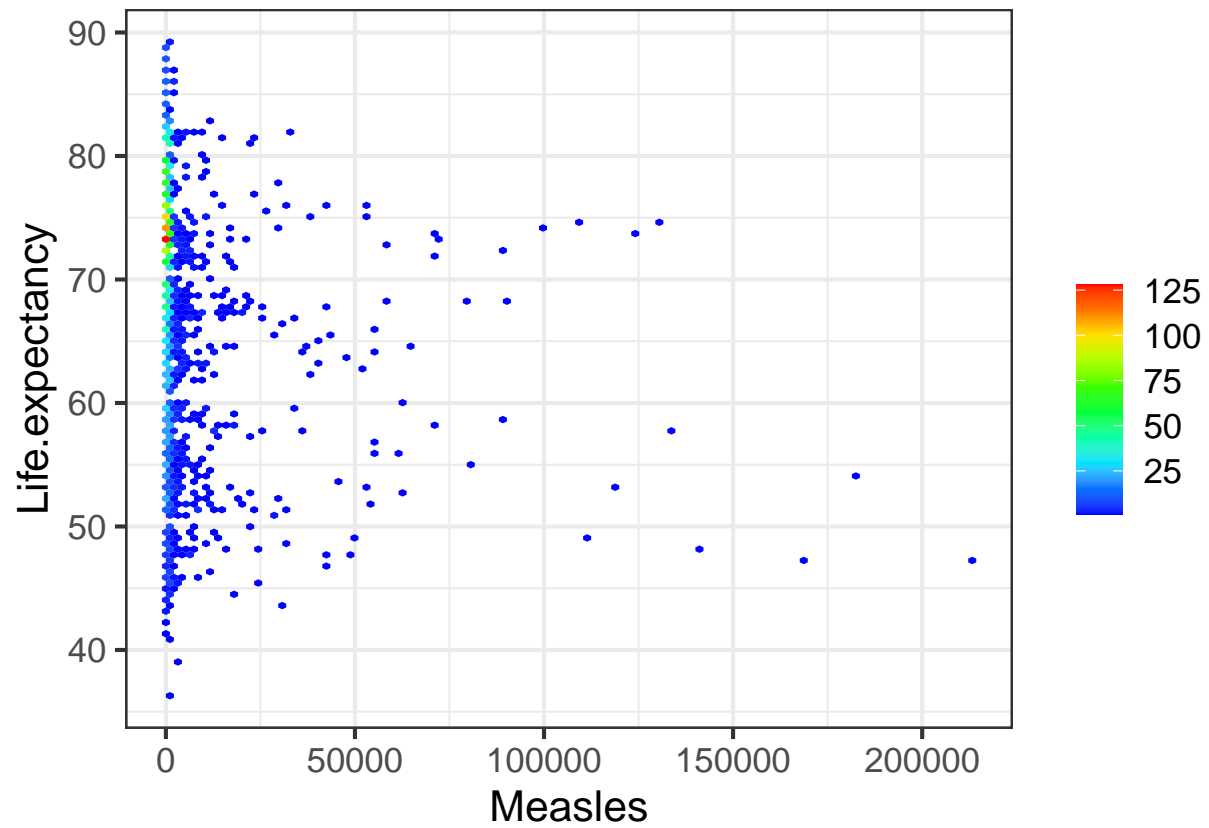
```
ggplot (life_expectancy_data) + geom_hex(aes(Hepatitis.B, Life.expectancy), bins = 100) +
  scale_fill_gradientn("", colours = rev(rainbow(10, end = 4/6)))
```

```
## Warning: Removed 563 rows containing non-finite values (stat_binhex).
```



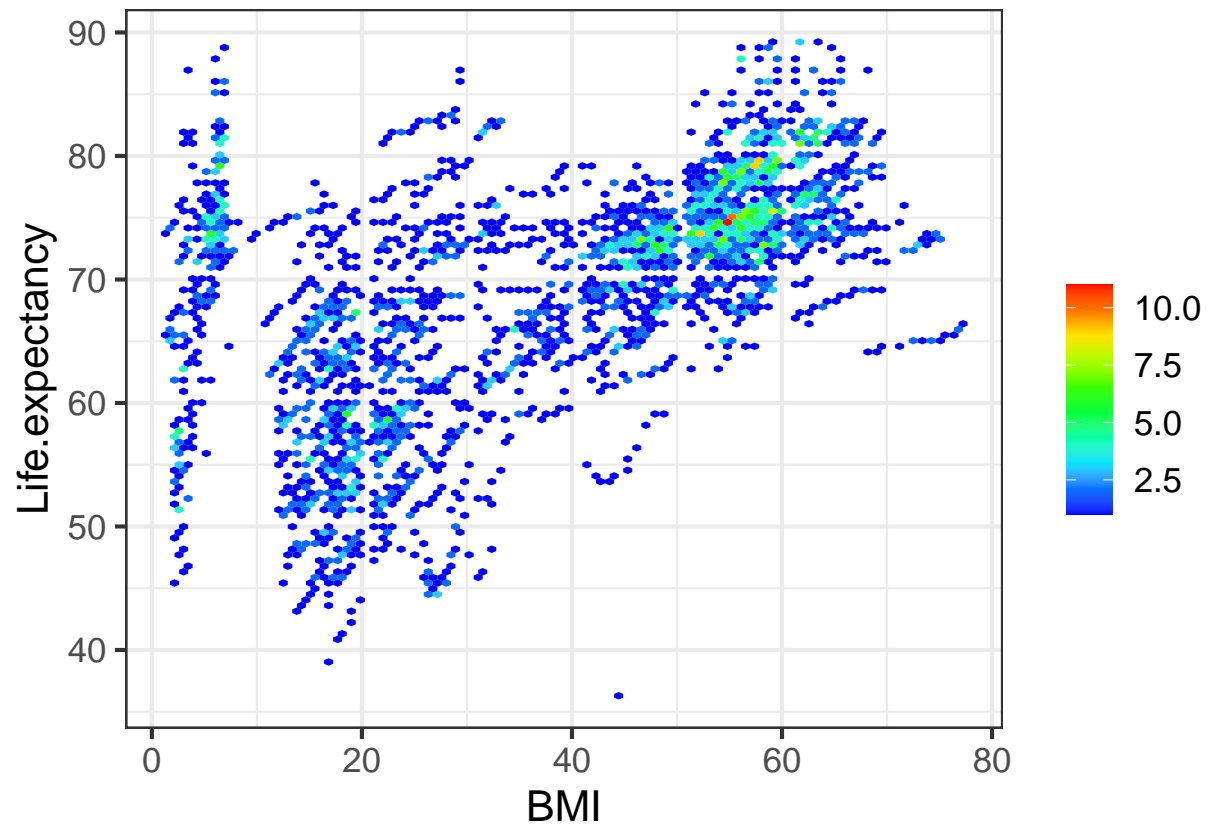
```
ggplot (life_expectancy_data) + geom_hex(aes(Measles, Life.expectancy), bins = 100) +
  scale_fill_gradientn("", colours = rev(rainbow(10, end = 4/6)))
```

```
## Warning: Removed 10 rows containing non-finite values (stat_binhex).
```



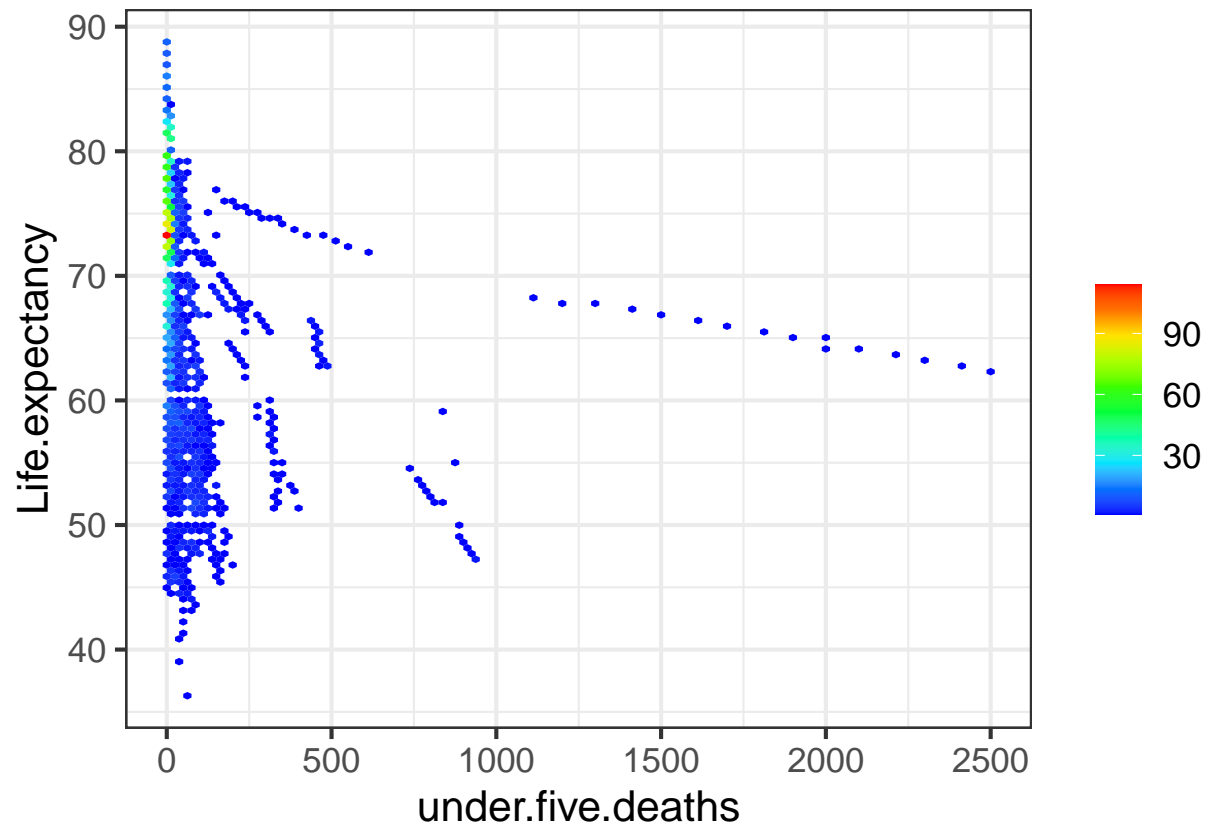
```
ggplot (life_expectancy_data) + geom_hex(aes(BMI, Life.expectancy), bins = 100) +  
  scale_fill_gradientn("", colours = rev(rainbow(10, end = 4/6)))
```

```
## Warning: Removed 42 rows containing non-finite values (stat_binhex).
```



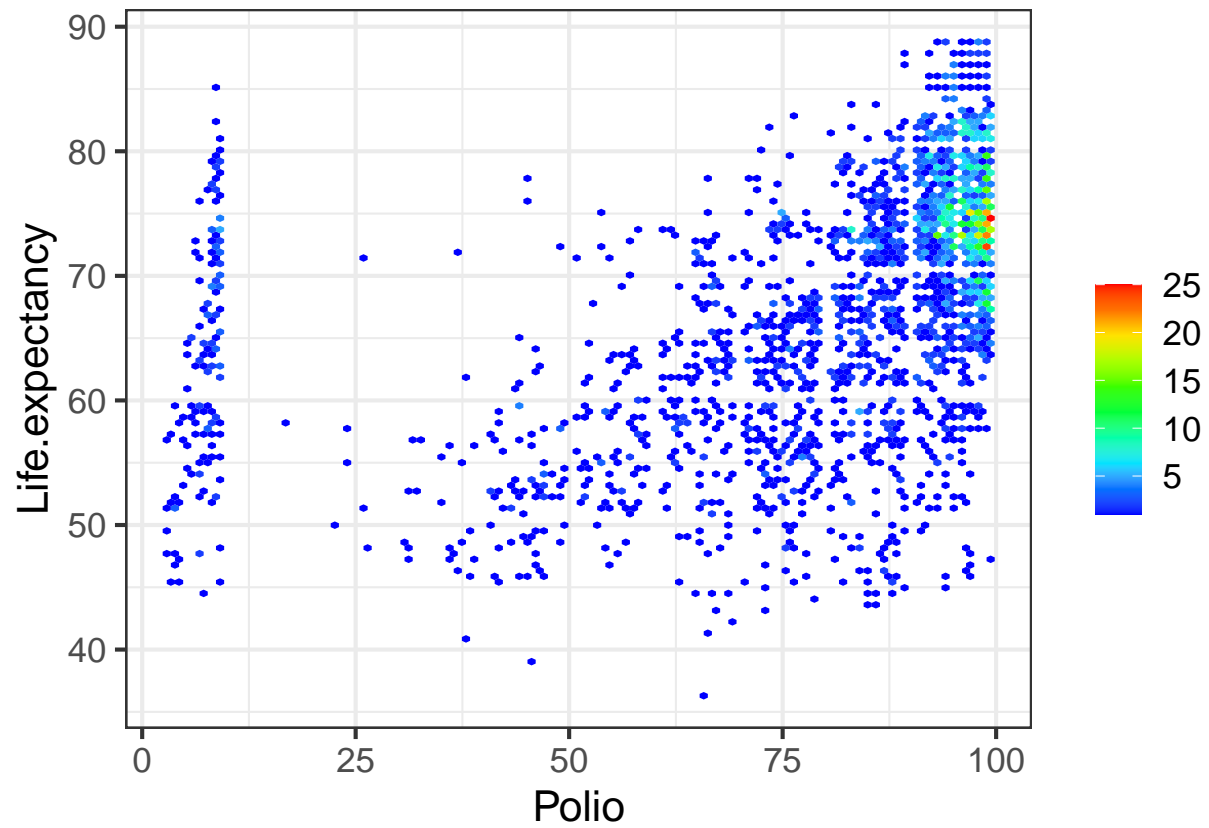
```
ggplot (life_expectancy_data) + geom_hex(aes(under.five.deaths, Life.expectancy), bins = 100) +  
  scale_fill_gradientn("", colours = rev(rainbow(10, end = 4/6)))
```

```
## Warning: Removed 10 rows containing non-finite values (stat_binhex).
```



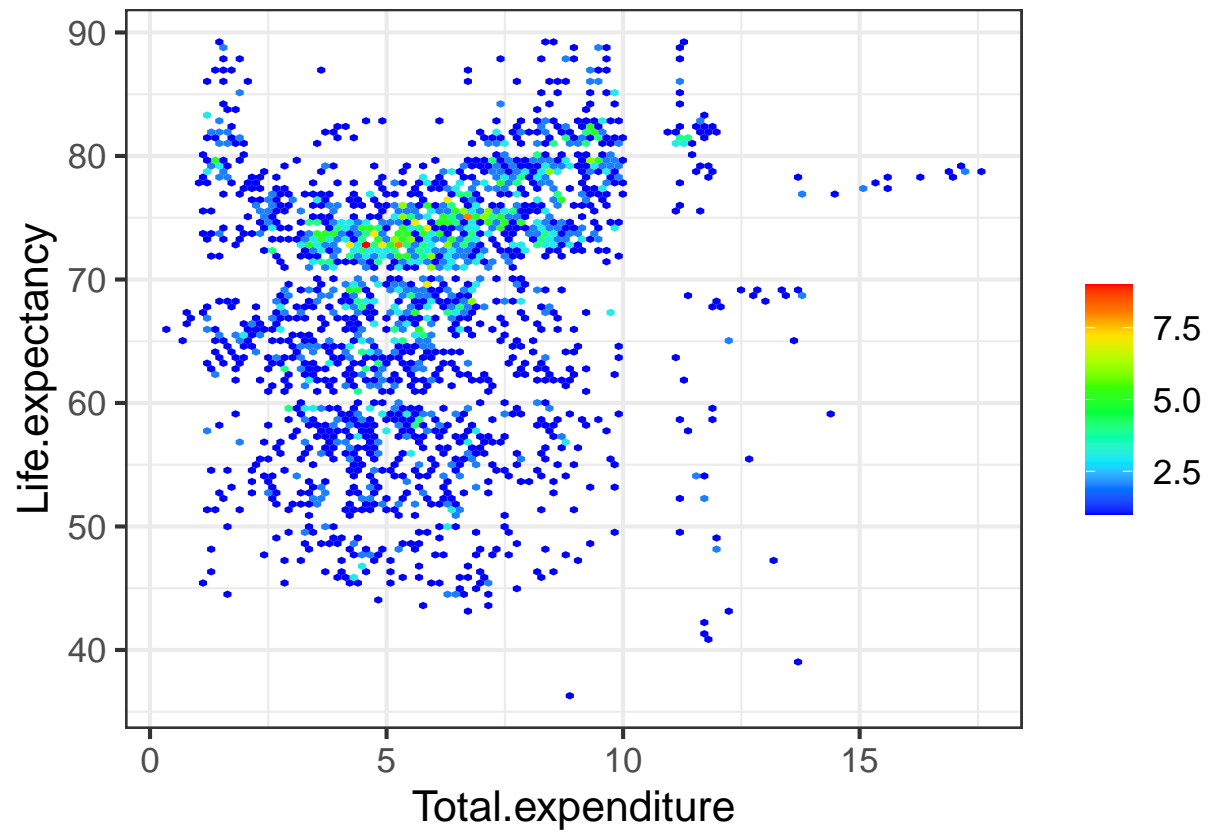
```
ggplot (life_expectancy_data) + geom_hex(aes(Polio, Life.expectancy), bins = 100) +
  scale_fill_gradientn("", colours = rev(rainbow(10, end = 4/6)))
```

```
## Warning: Removed 29 rows containing non-finite values (stat_binhex).
```

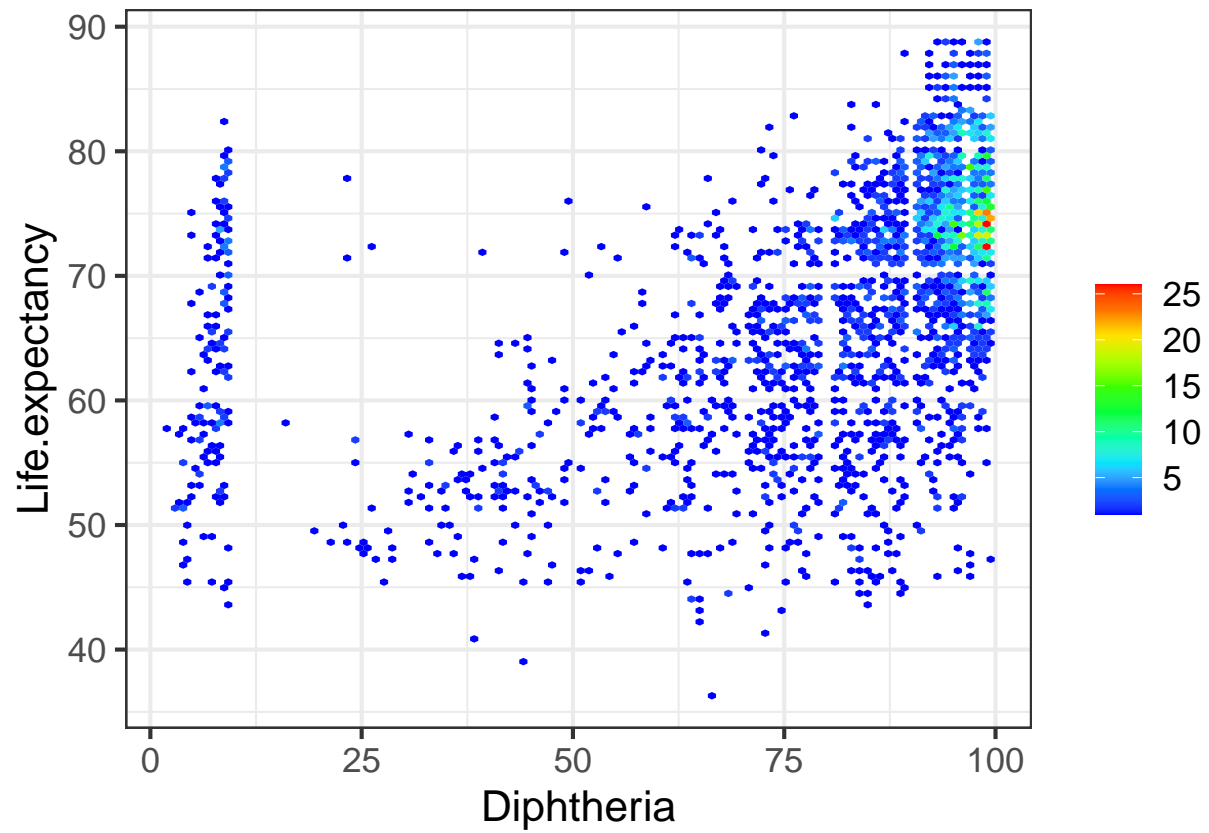
```
ggplot (life_expectancy_data) + geom_hex(aes(Total.expenditure, Life.expectancy), bins = 100) +
  scale_fill_gradientn("", colours = rev(rainbow(10, end = 4/6)))
```

```
## Warning: Removed 236 rows containing non-finite values (stat_binhex).
```



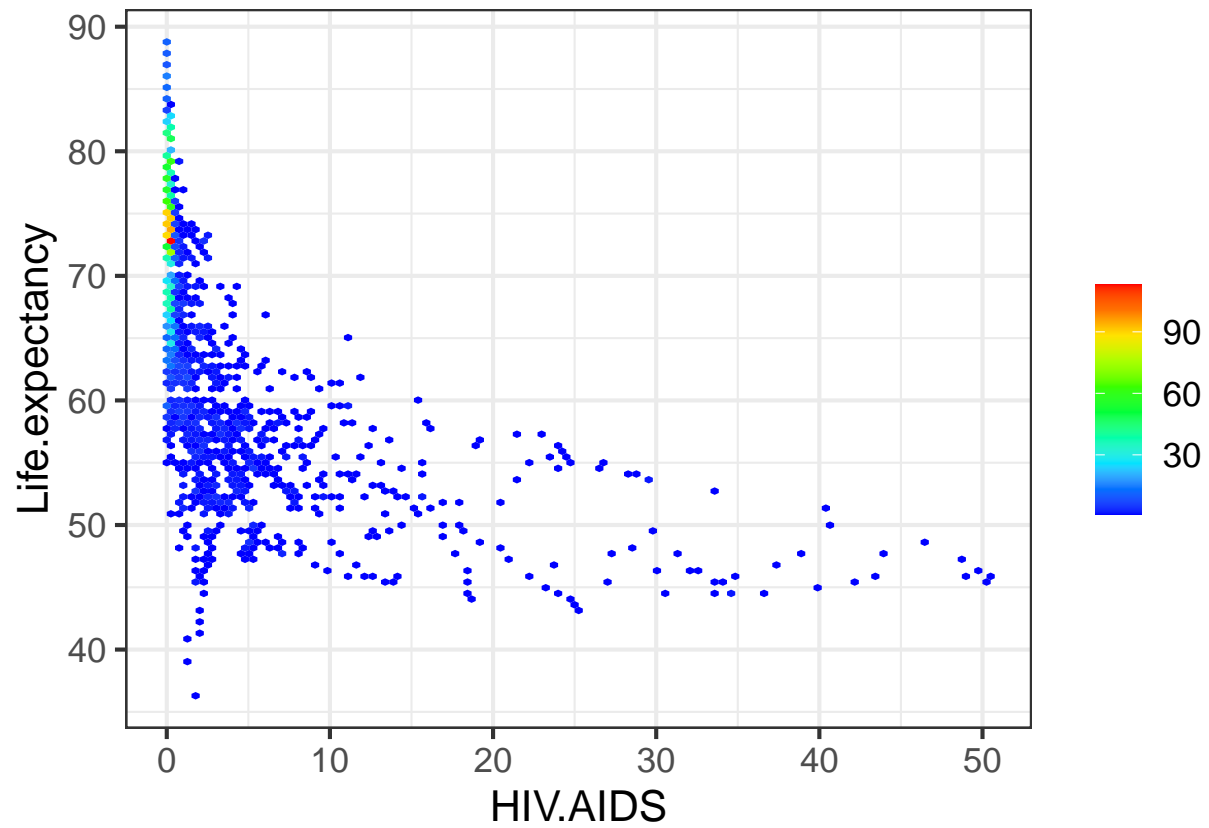
```
ggplot (life_expectancy_data) + geom_hex(aes(Diphtheria, Life.expectancy), bins = 100) +
  scale_fill_gradientn("", colours = rev(rainbow(10, end = 4/6)))
```

```
## Warning: Removed 29 rows containing non-finite values (stat_binhex).
```



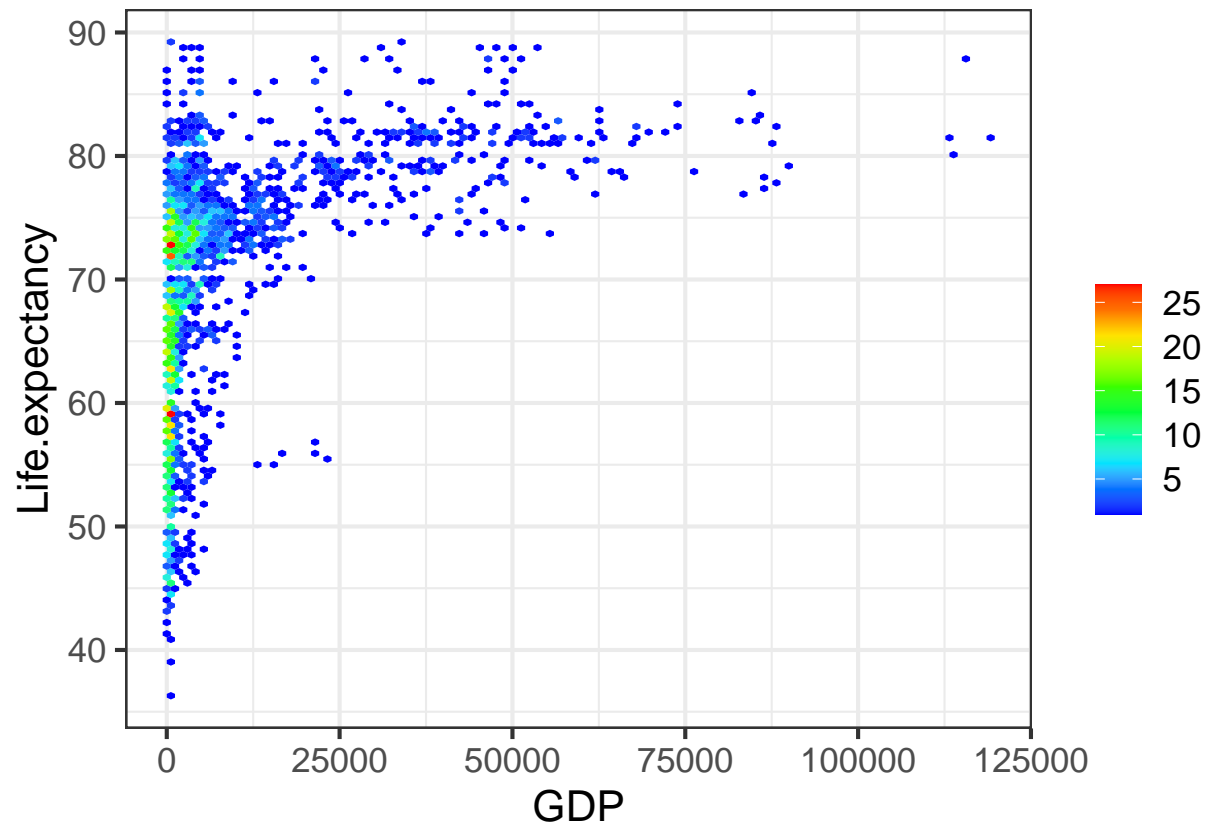
```
ggplot (life_expectancy_data) + geom_hex(aes(HIV.AIDS, Life.expectancy), bins = 100) +
  scale_fill_gradientn("", colours = rev(rainbow(10, end = 4/6)))
```

```
## Warning: Removed 10 rows containing non-finite values (stat_binhex).
```



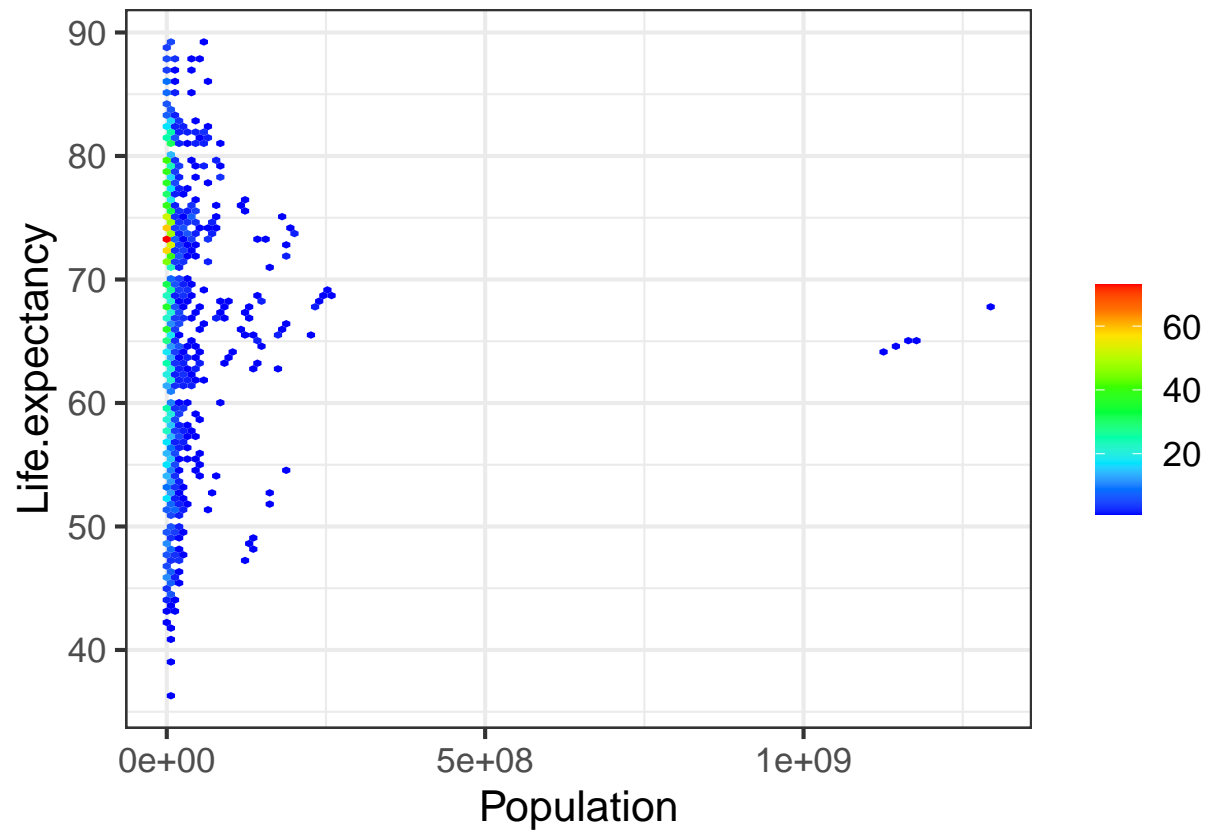
```
ggplot (life_expectancy_data) + geom_hex(aes(GDP, Life.expectancy), bins = 100) +  
  scale_fill_gradientn("", colours = rev(rainbow(10, end = 4/6)))
```

```
## Warning: Removed 453 rows containing non-finite values (stat_binhex).
```



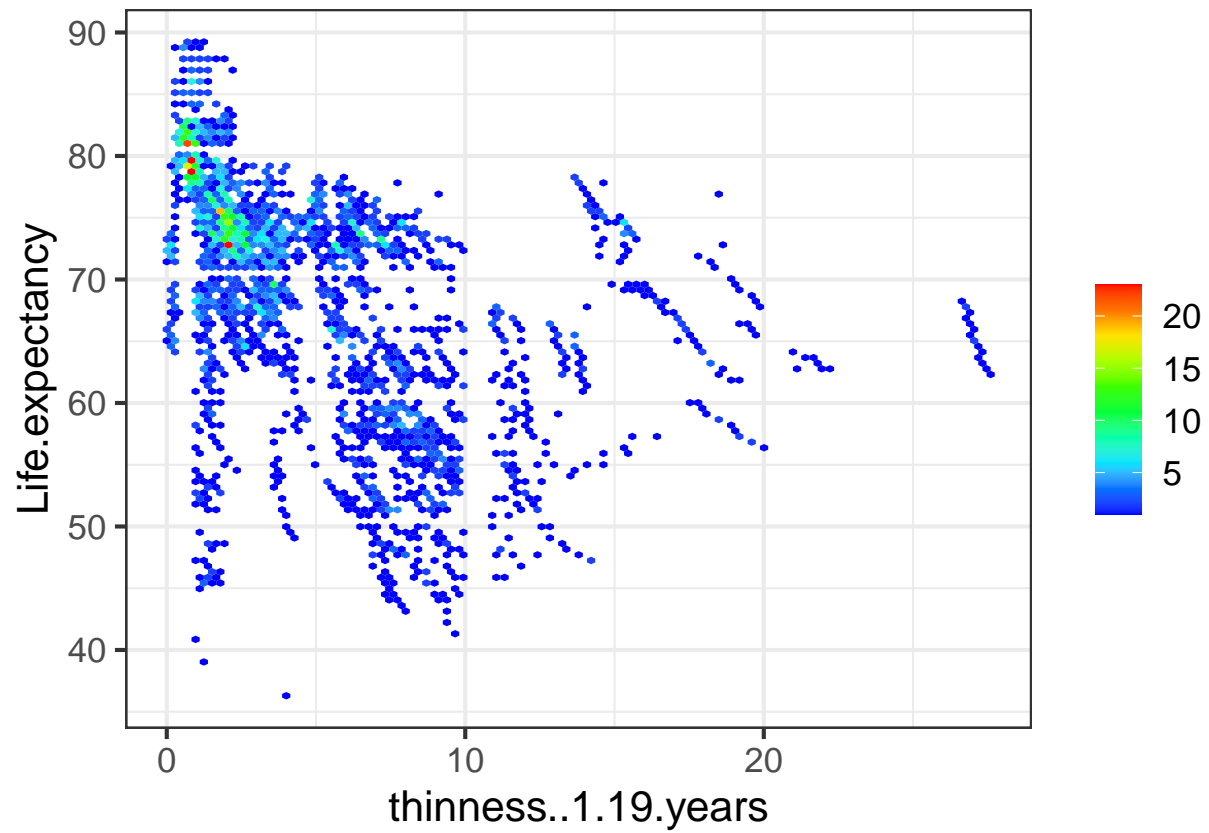
```
ggplot (life_expectancy_data) + geom_hex(aes(Population, Life.expectancy), bins = 100) +
  scale_fill_gradientn("", colours = rev(rainbow(10, end = 4/6)))
```

```
## Warning: Removed 654 rows containing non-finite values (stat_binhex).
```



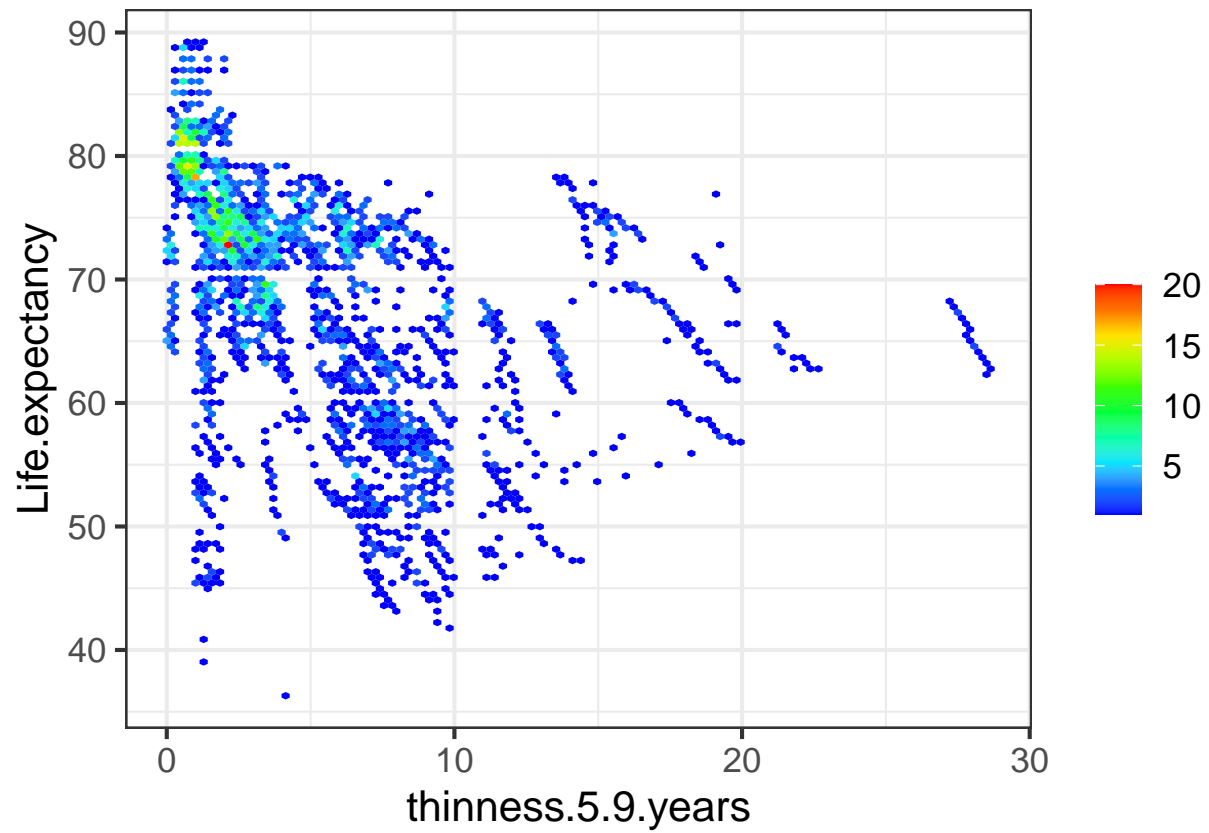
```
ggplot (life_expectancy_data) + geom_hex(aes(thinness..1.19.years, Life.expectancy), bins = 100) +
  scale_fill_gradientn("", colours = rev(rainbow(10, end = 4/6)))
```

```
## Warning: Removed 42 rows containing non-finite values (stat_binhex).
```



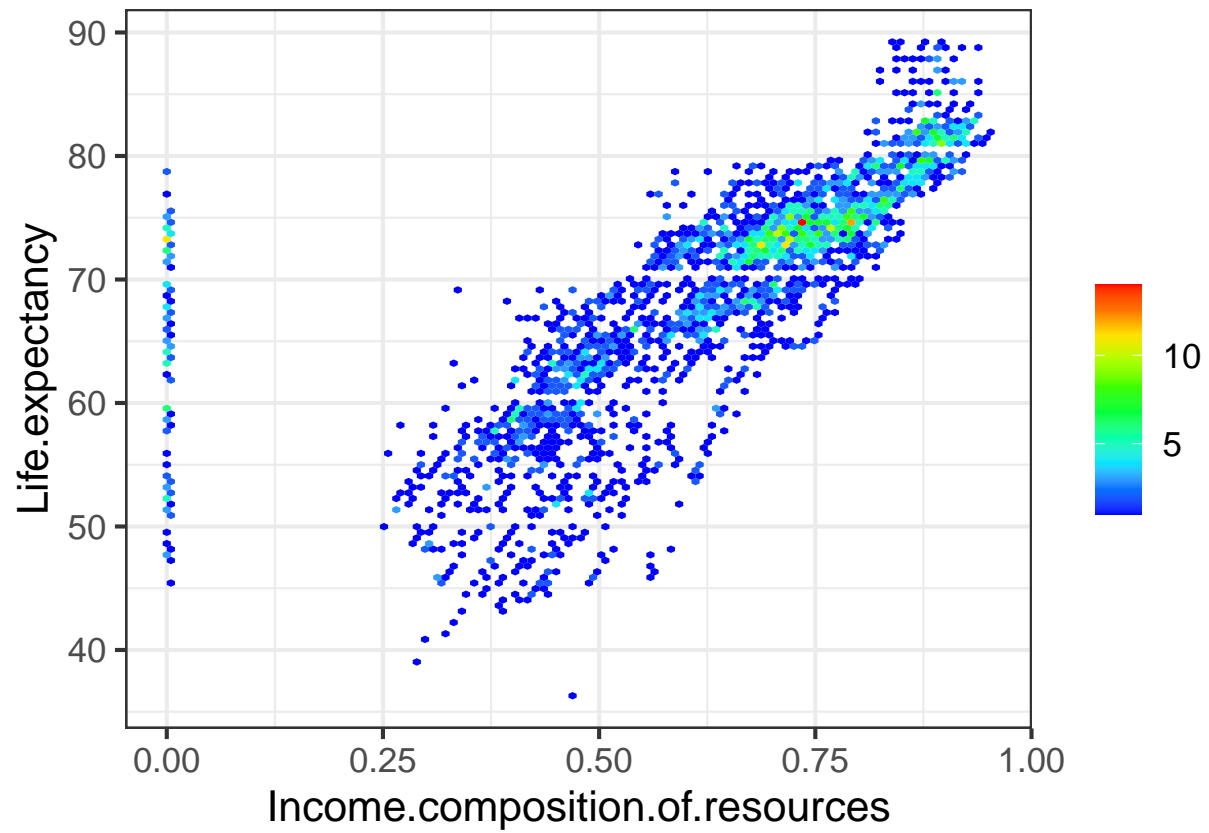
```
ggplot (life_expectancy_data) + geom_hex(aes(thinness.5.9.years, Life.expectancy), bins = 100) +
  scale_fill_gradientn("", colours = rev(rainbow(10, end = 4/6)))
```

```
## Warning: Removed 42 rows containing non-finite values (stat_binhex).
```



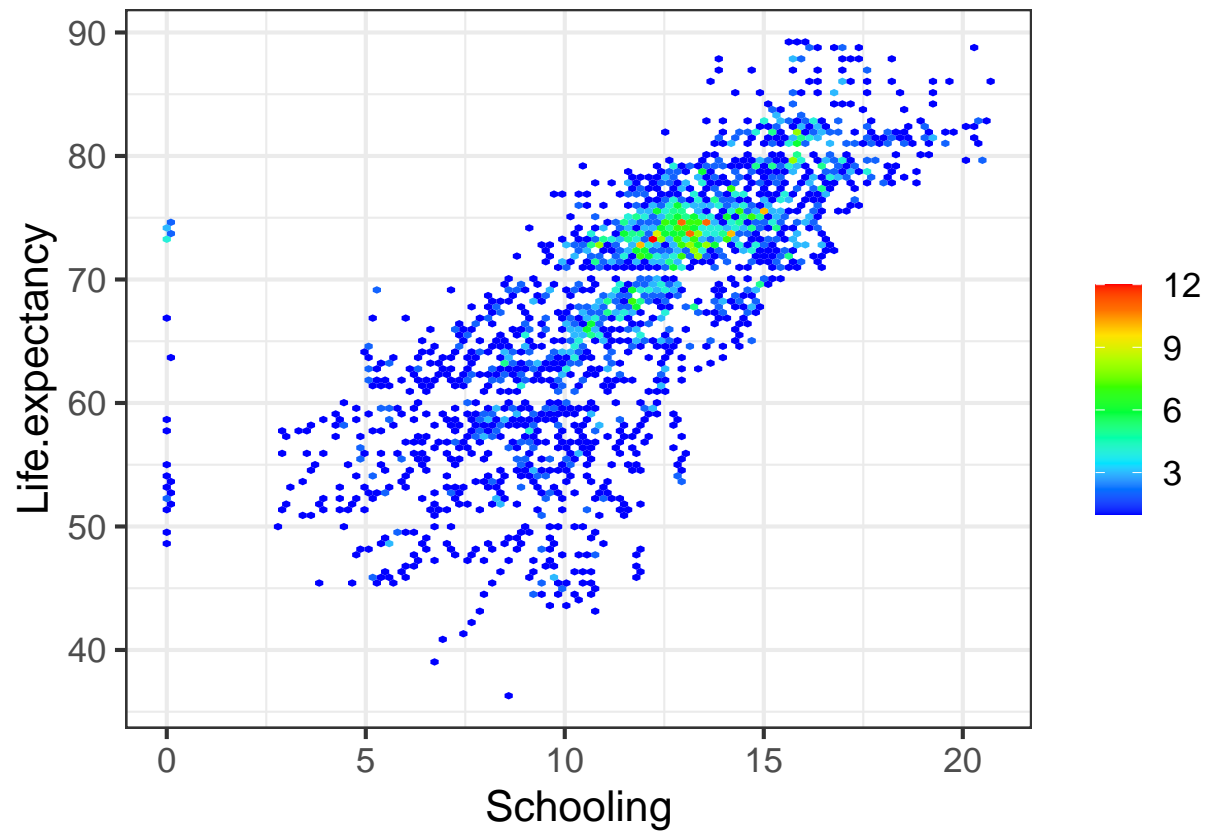
```
ggplot (life_expectancy_data) + geom_hex(aes(Income.composition.of.resources, Life.expectancy), bins = 1000) +
  scale_fill_gradientn("", colours = rev(rainbow(10, end = 4/6)))
```

```
## Warning: Removed 170 rows containing non-finite values (stat_binhex).
```

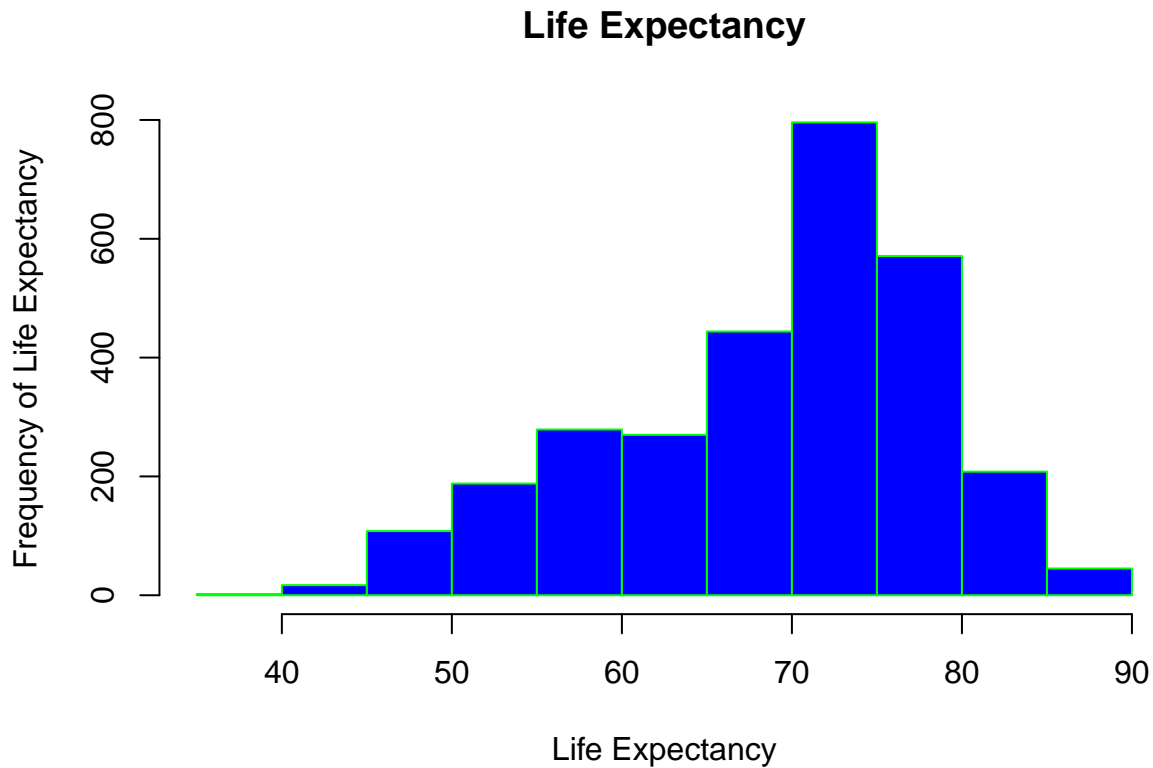
```
ggplot (life_expectancy_data) + geom_hex(aes(Schooling, Life.expectancy), bins = 100) +
  scale_fill_gradientn("", colours = rev(rainbow(10, end = 4/6)))
```

```
## Warning: Removed 170 rows containing non-finite values (stat_binhex).
```



Histograms

```
hist (life_expectancy_data$Life.expectancy, xlab="Life Expectancy", ylab="Frequency of Life Expectancy"
```



Fitting the model

In the following step, we will perform hypothesis testing to figure out variables that have a high linear correlation with life expectancy to arrive at a fitted multiple linear regression model

```
# render("life_expectancy_analysis.Rmd", "pdf_document")
```

Residual Analysis

In the following step, we will perform residual analysis, to figure out outliers and variables that contribute for and against the accuracy of our model.

Response to Feedback

How many data points do you have? Is each country a data point or a combination of country and year a data point? (-1)

Describe members' responsibility on the planned data analysis, not about the proposal. (-1)

Exploratory data analysis described in GOAL is better placed in PLAN.

this is updated text