Class Activity 5

Courtney Hodge

2024-07-22

```
library(tidyverse)
```

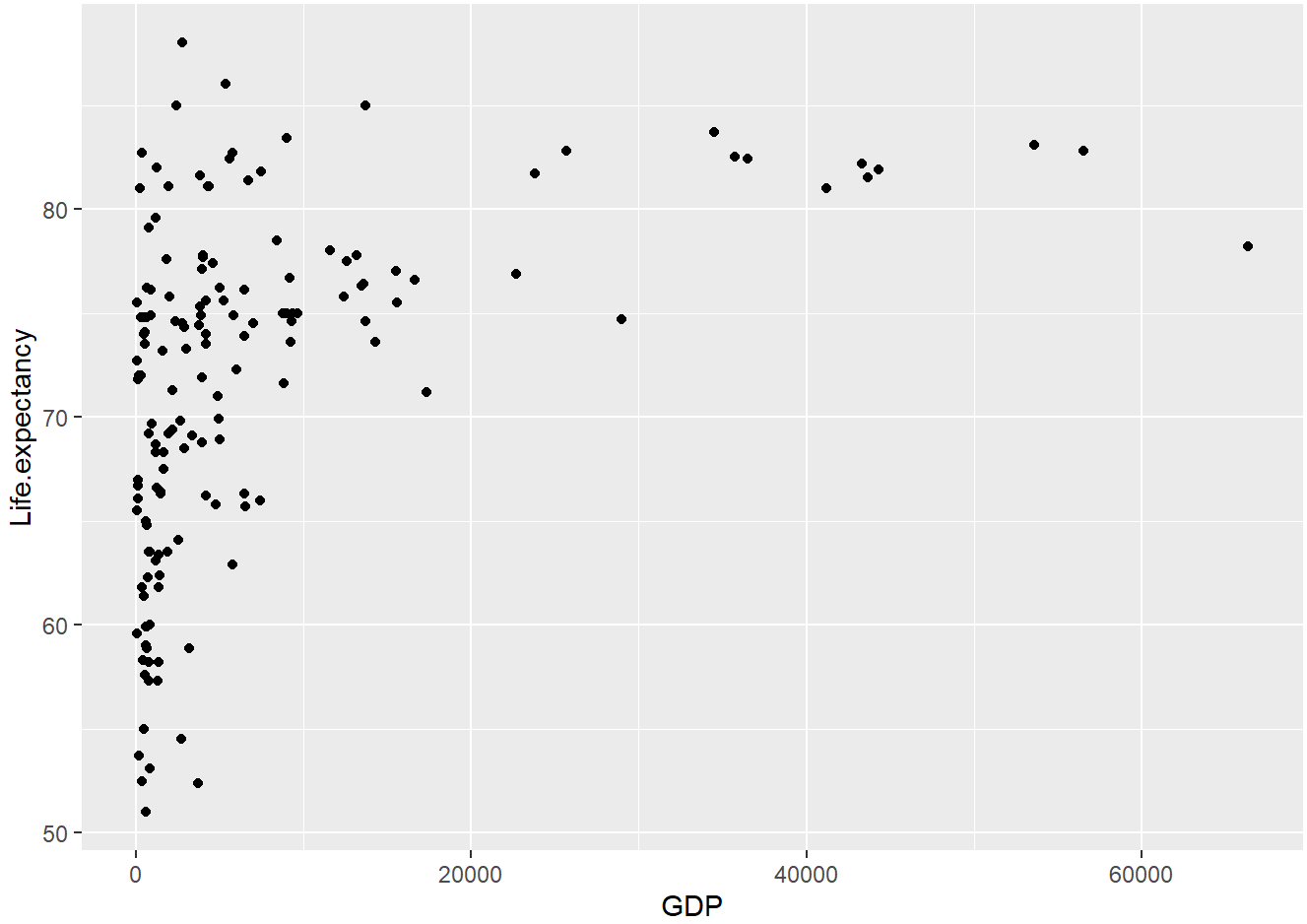
```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.3     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.4     ✓ tibble    3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr     1.3.0
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
e errors
```

```
life_data <- read.csv("C:\\Users\\hodge\\OneDrive - Baylor University\\Desktop\\UVA Coding Folde
r\\STAT 6021\\expectancy.csv")
```

```
life_data2<- select(life_data,Life.expectancy, GDP) |>
  na.omit()
```

1a

```
ggplot(life_data2, aes(x=GDP, y = Life.expectancy)) +
  geom_point()
```



1b

```
cor(life_data2$Life.expectancy, life_data2$GDP)
```

```
## [1] 0.454491
```

1c

> Based on the lienar model's correlation calculation of 0.454, and the non-linear points of the scatter plot, a linear model is not appropriate for predicting life.Expectancy using GDP.

2

```
beta_1 <- cor(life_data2$Life.expectancy, life_data2$GDP) * sd(life_data2$Life.expectancy) / sd
(life_data2$GDP)

beta_0 <- mean(life_data2$Life.expectancy) - beta_1 * mean(life_data2$GDP)

beta_0
```

```
## [1] 69.37846
```

```
beta_1
```

```
## [1] 0.000321739
```

3

```
model_1 <- lm(Life.expectancy~GDP, data = life_data2)
coef(model_1)
```

```
##  (Intercept)          GDP
## 69.378458568  0.000321739
```

4a

```
X <- cbind(1, life_data2$GDP)
```

4b

```
Y <- as.matrix(life_data2$Life.expectancy)
beta_hat <- solve((t(X) %*% (X)))%*%t(X)%*%Y

beta_hat
```

```
##              [,1]
## [1,] 69.378458568
## [2,]  0.000321739
```