LIFE EXPECTANCY

PRESENTED BY HECTOR PENA

Introduction

- Using global health and economic data, I wanted to look at what factors really shape a country's life expectancy, like healthcare, education, and income
  - We've all heard that where you live can affect how long you live, so I wanted to explore why that's true

About

- The data was collected from WHO (World Health organization) and United Nations with the help of Deeksha Russel and Duan Wang.
- This dataset was a collection of CDC records used to examine life expectancy from 2000 - 2015 across 193 countries
- Key predictors include economic indicators, average education level, general health (BMI), and other social factors

General Information

- The original data set was (2938 x 22)
- 22 original predictors
- This set of observations included 193 countries total,
- Variable of interest was Life.expectancy

- Data was examined for completeness, and found to have logical value ranges for variables
- Some variables were only good at predicting outcomes for low-income countries (vaccination rates)

Variables

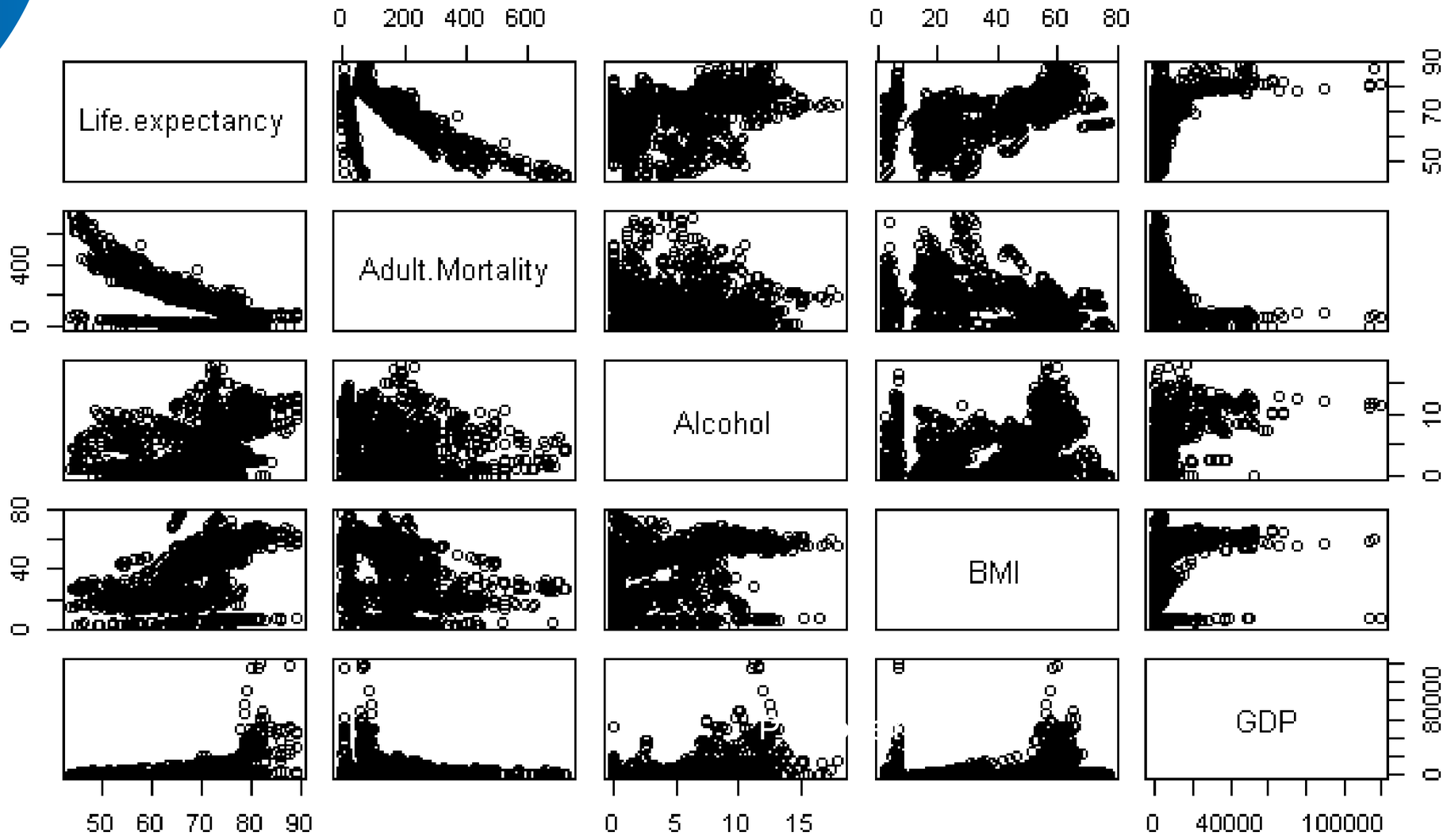
- Life.expectancy
- Adult.Mortality
- infant.deaths
- Alcohol
- percentage.expenditure
- Hepatitis.B
- Measles
- BMI
- under.five.deaths
- Polio
- Total.expenditure
- Diphtheria
- HIV.AIDS
- GDP
- Population
- thinness..1.19.years
- thinness.5.9.years
- Income.composition.of.resources
- Schooling

Research Questions

- How do structural factors like education and economic resource distribution shape a country's average life expectancy?

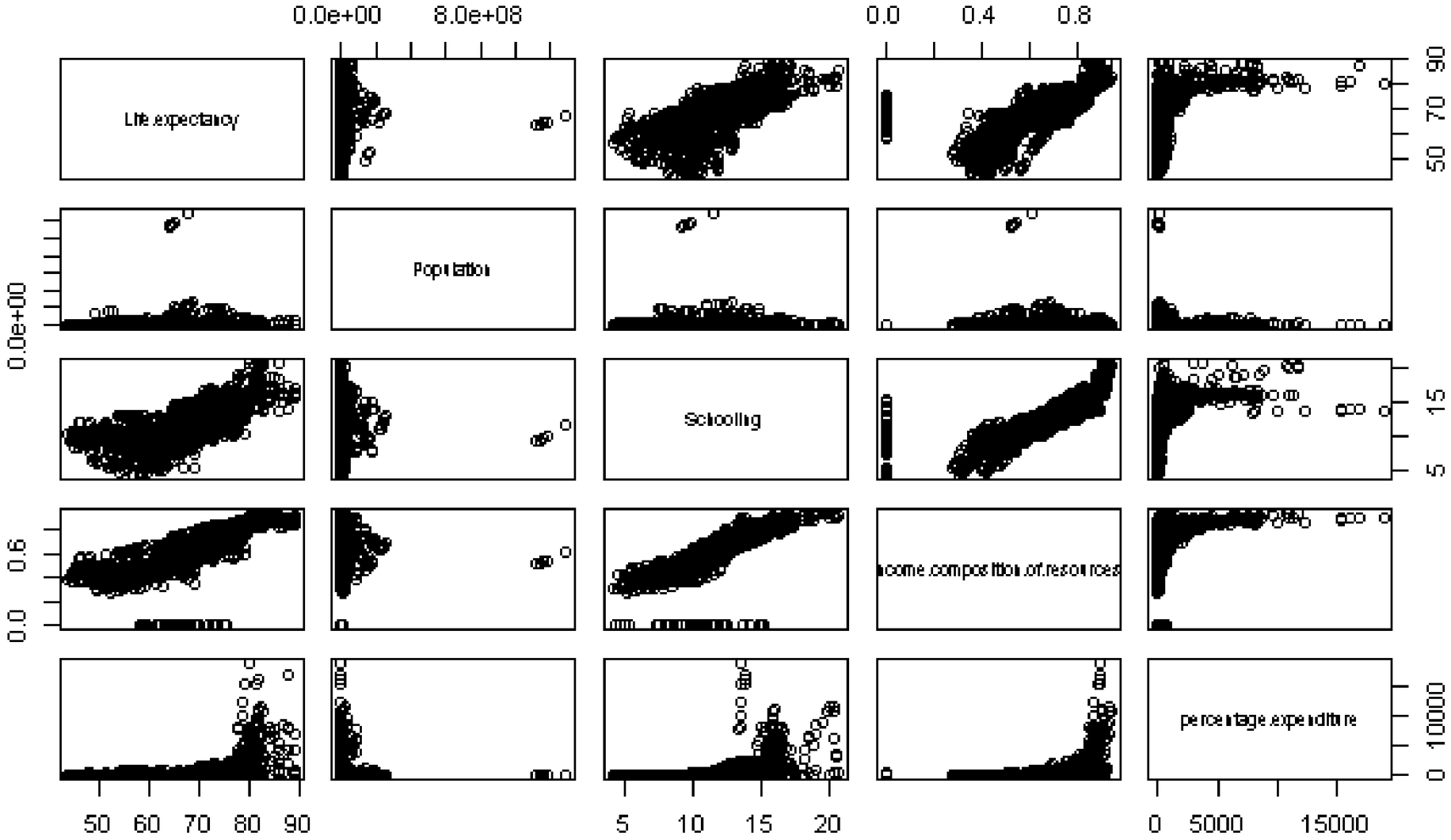
- What is the relationship between a country’s investment in healthcare and its average life expectancy?

Exploring Relationships



(Holding all other predictors constant)

Exploring Relationships (cont.)



(Holding all other predictors constant)

Model Selection

AIC Backward

```
Call:
lm(formula = Life.expectancy ~ Schooling + HIV.AIDS + Adult.Mortality +
    Income.composition.of.resources + percentage.expenditure +
    BMI + Diphtheria + under.five.deaths + infant.deaths + Status +
    Alcohol + thinness.5.9.years + Total.expenditure, data = df2)
```

AIC Forward

```
Call:
lm(formula = Life.expectancy ~ Status + Adult.Mortality + infant.deaths +
    Alcohol + percentage.expenditure + Hepatitis.B + BMI + under.five.deaths +
    Polio + Total.expenditure + Diphtheria + HIV.AIDS + thinness.5.9.years +
    Income.composition.of.resources + Schooling, data = df2)
```

BIC Backward

```
Call:
lm(formula = Life.expectancy ~ Adult.Mortality + infant.deaths +
    percentage.expenditure + BMI + under.five.deaths + Diphtheria +
    HIV.AIDS + Income.composition.of.resources + Schooling, data = df2)
```

BIC Forward

```
Call:
lm(formula = Life.expectancy ~ Schooling + HIV.AIDS + Adult.Mortality +
    Income.composition.of.resources + percentage.expenditure +
    BMI + Diphtheria + under.five.deaths + infant.deaths, data = df2)
```

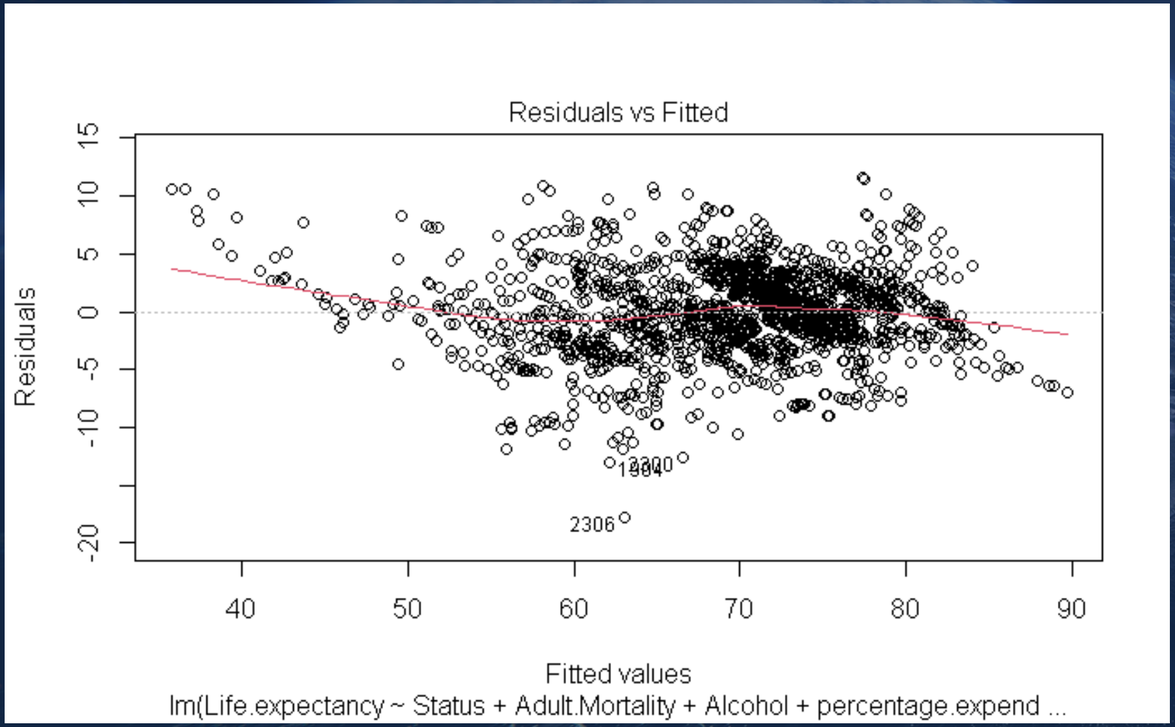
```
                 Adult.Mortality                    infant.deaths
percentage.expenditure
                        1.766417                       176.934458
1.248657
                             BMI                under.five.deaths
Diphtheria
                        1.549990                       178.045561
1.209337
                        HIV.AIDS Income.composition.of.resources
Schooling
                        1.444584                        2.849230
3.066677
```

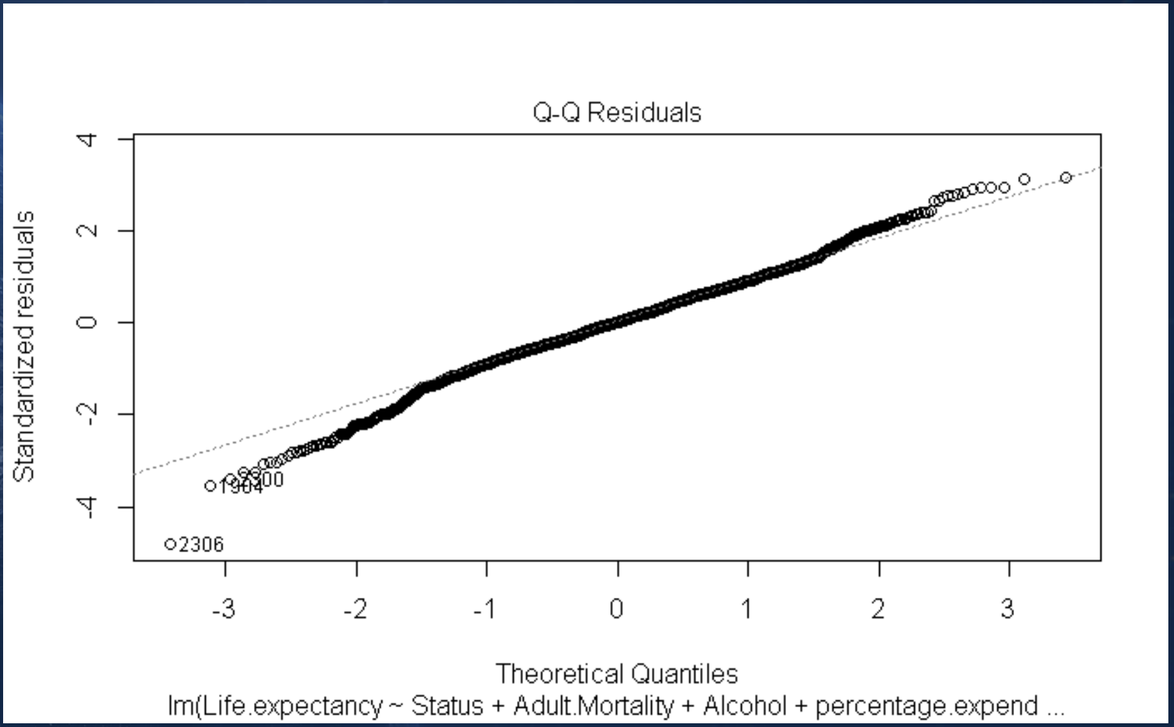
VIF / Multicollinearity

Fitting First Model

```
Call:
lm(formula = Life.expectancy ~ Status + Adult.Mortality + Alcohol +
    percentage.expenditure + BMI + HIV.AIDS + Income.composition.of.resources +
    Schooling + under.five.deaths, data = df2)

Residuals:
     Min       1Q   Median       3Q      Max
-17.7610  -2.1235   0.0492   2.3517  11.6131

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                      5.460e+01  7.210e-01  75.731  < 2e-16 ***
StatusDeveloping                -9.629e-01  3.467e-01  -2.777 0.005550 **
Adult.Mortality                 -1.806e-02  9.681e-04 -18.651  < 2e-16 ***
Alcohol                         -1.214e-01  3.314e-02  -3.664 0.000257 ***
percentage.expenditure           4.129e-04  6.091e-05   6.778 1.70e-11 ***
BMI                              3.678e-02  5.742e-03   6.406 1.95e-10 ***
HIV.AIDS                        -4.383e-01  1.825e-02 -24.011  < 2e-16 ***
Income.composition.of.resources  1.097e+01  8.453e-01  12.975  < 2e-16 ***
Schooling                        9.459e-01  5.999e-02  15.767  < 2e-16 ***
under.five.deaths               -2.317e-03  5.845e-04  -3.964 7.70e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.705 on 1639 degrees of freedom
Multiple R-squared:  0.8236,    Adjusted R-squared:  0.8226
F-statistic: 850.3 on 9 and 1639 DF,  p-value: < 2.2e-16
```

Model Diagnostics



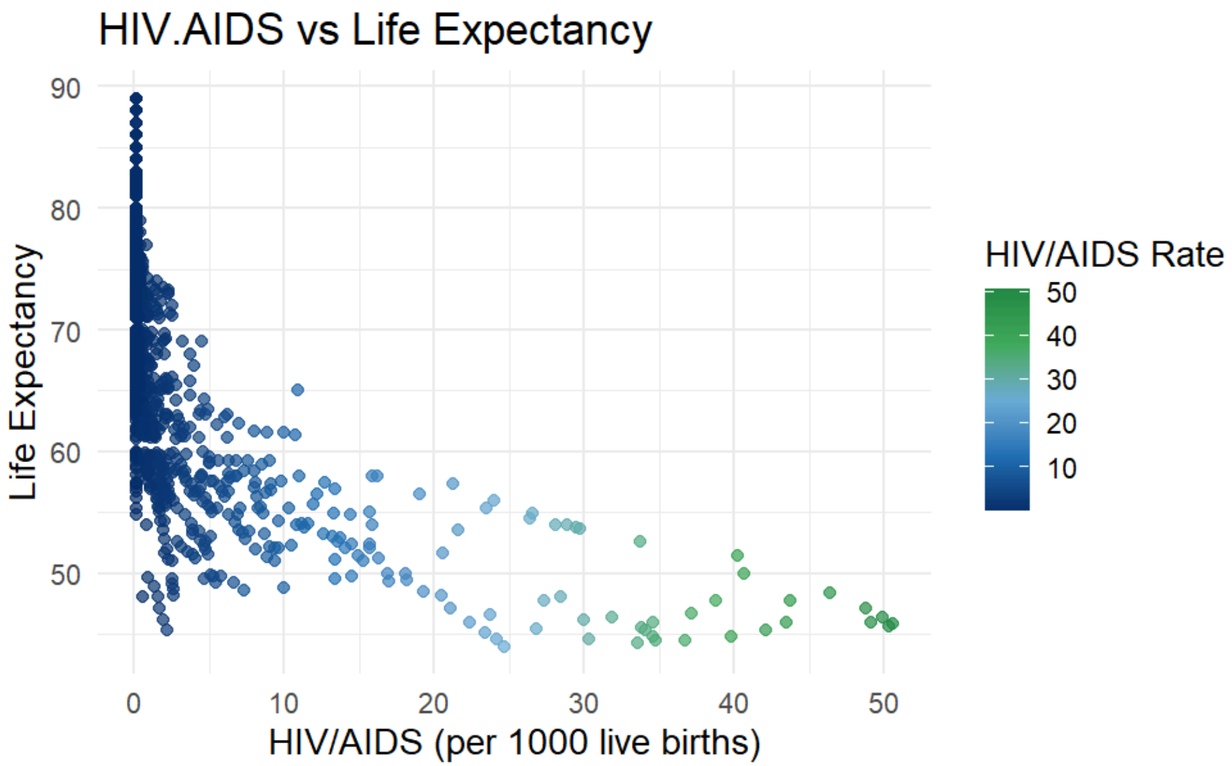


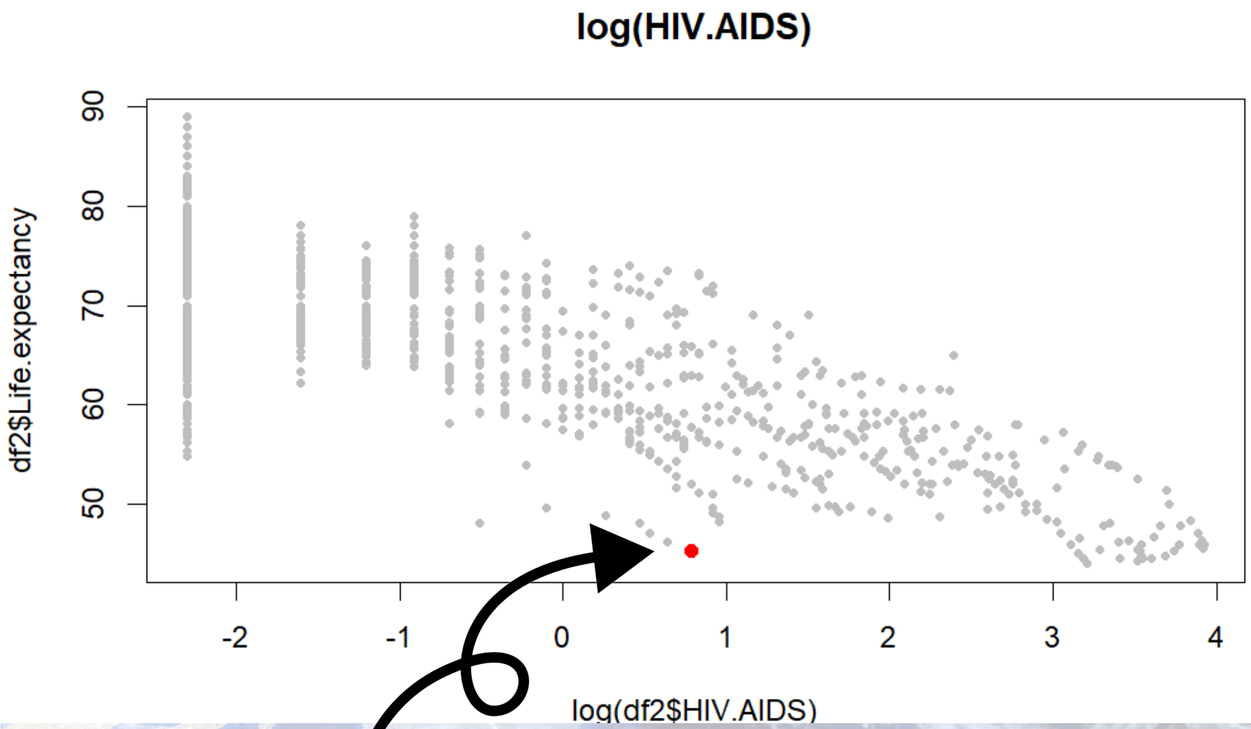
Transformed Model

```
Call:
lm(formula = Life.expectancy ~ Adult.Mortality + Alcohol + percentage.expenditure +
    BMI + log(HIV.AIDS) + Income.composition.of.resources + Schooling +
    under.five.deaths)

Residuals:
     Min       1Q   Median       3Q      Max
-15.4623  -1.8527  -0.0259   1.9161  12.9847

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                      5.486e+01  5.599e-01  97.978  < 2e-16 ***
Adult.Mortality                 -1.558e-02  9.235e-04 -16.875  < 2e-16 ***
Alcohol                          4.978e-03  2.895e-02   0.172  0.86350
percentage.expenditure           4.543e-04  5.567e-05   8.160 6.60e-16 ***
BMI                              1.658e-02  5.467e-03   3.033  0.00246 **
log(HIV.AIDS)                   -2.256e+00  7.630e-02 -29.571  < 2e-16 ***
Income.composition.of.resources  1.014e+01  7.971e-01  12.722  < 2e-16 ***
Schooling                        5.864e-01  5.754e-02  10.192  < 2e-16 ***
under.five.deaths               -2.612e-03  5.502e-04  -4.748 2.23e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.487 on 1640 degrees of freedom
Multiple R-squared:  0.8437,    Adjusted R-squared:  0.8429
F-statistic:  1106 on 8 and 1640 DF,  p-value: < 2.2e-16
```

After taking log of HIV.AIDS, Alcohol was then insignificant in predicting our response

Transforming





Description: df [2 × 2]

	df <dbl>	AIC <dbl>
log_model	9	8830.216
first_model	10	9024.450

🤔 Take note of this

The Final Model

```
Call:
lm(formula = Life.expectancy ~ Adult.Mortality + percentage.expenditure +
    BMI + log(HIV.AIDS) + Income.composition.of.resources + Schooling +
    under.five.deaths)

Residuals:
    Min      1Q  Median      3Q     Max
-15.447  -1.859  -0.025   1.915  12.988

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                      5.482e+01  5.196e-01 105.507  < 2e-16 ***
Adult.Mortality                 -1.558e-02  9.220e-04 -16.894  < 2e-16 ***
percentage.expenditure           4.563e-04  5.444e-05   8.382  < 2e-16 ***
BMI                              1.663e-02  5.457e-03   3.047  0.00235 **
log(HIV.AIDS)                   -2.254e+00  7.501e-02 -30.049  < 2e-16 ***
Income.composition.of.resources  1.016e+01  7.852e-01  12.944  < 2e-16 ***
Schooling                        5.898e-01  5.396e-02  10.931  < 2e-16 ***
under.five.deaths               -2.607e-03  5.491e-04  -4.748 2.24e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.486 on 1641 degrees of freedom
Multiple R-squared:  0.8437,    Adjusted R-squared:  0.843
F-statistic:  1265 on 7 and 1641 DF,  p-value: < 2.2e-16
```

Interpretations

Adult Mortality = -0.0156

For every 1 death per 1000 adults aged 15 to 60, life expectancy decreases by approximately 0.016 years (about 5.8 days), assuming all other variables remain constant

% Expenditure (on health) = 0.000456

For each 1% increase in public health spending as a % of GDP, life expectancy increases by 0.00046 years (~4 hours). Although small per unit, it can add up

BMI = 0.0166

For each unit increase in average BMI percentage, life expectancy increases by about 0.017 years (around 6 days)

Interpretations (cont.)

log(HIV/AIDS deaths) = -2.254

For every 1% increase in HIV/AIDS deaths, life expectancy goes down by about 8 days

Income Composition of Resources = 10.16

A 1-unit increase in Income Composition of Resources (scaled 0–1) is associated with a 9.94-year increase in life expectancy. Smaller changes would yield proportionally smaller effects

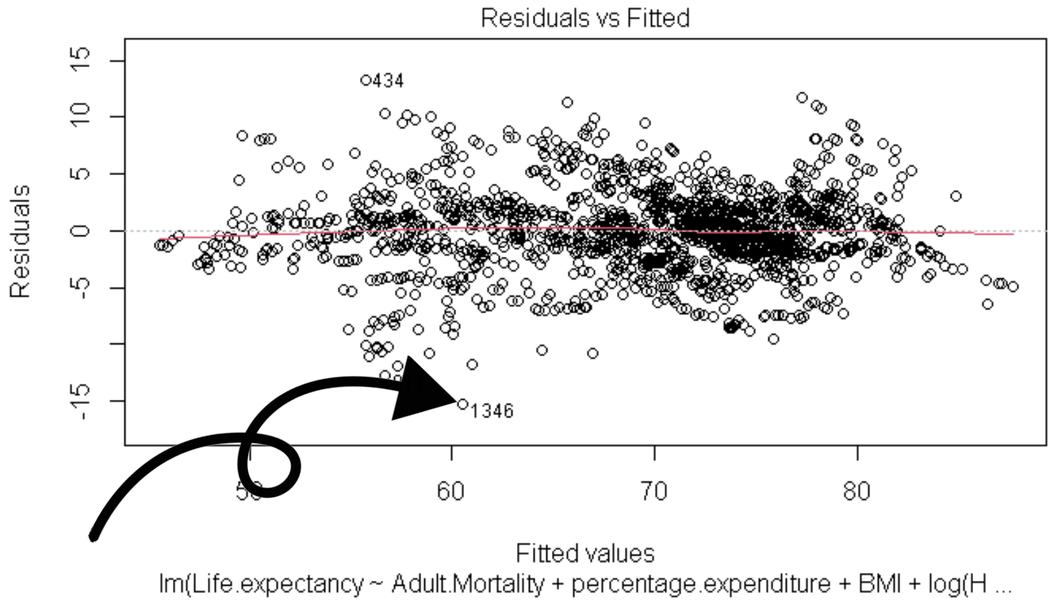
Schooling (years)
0.59

Each additional year of average schooling adds about 0.59 years (≈7 months) to life expectancy

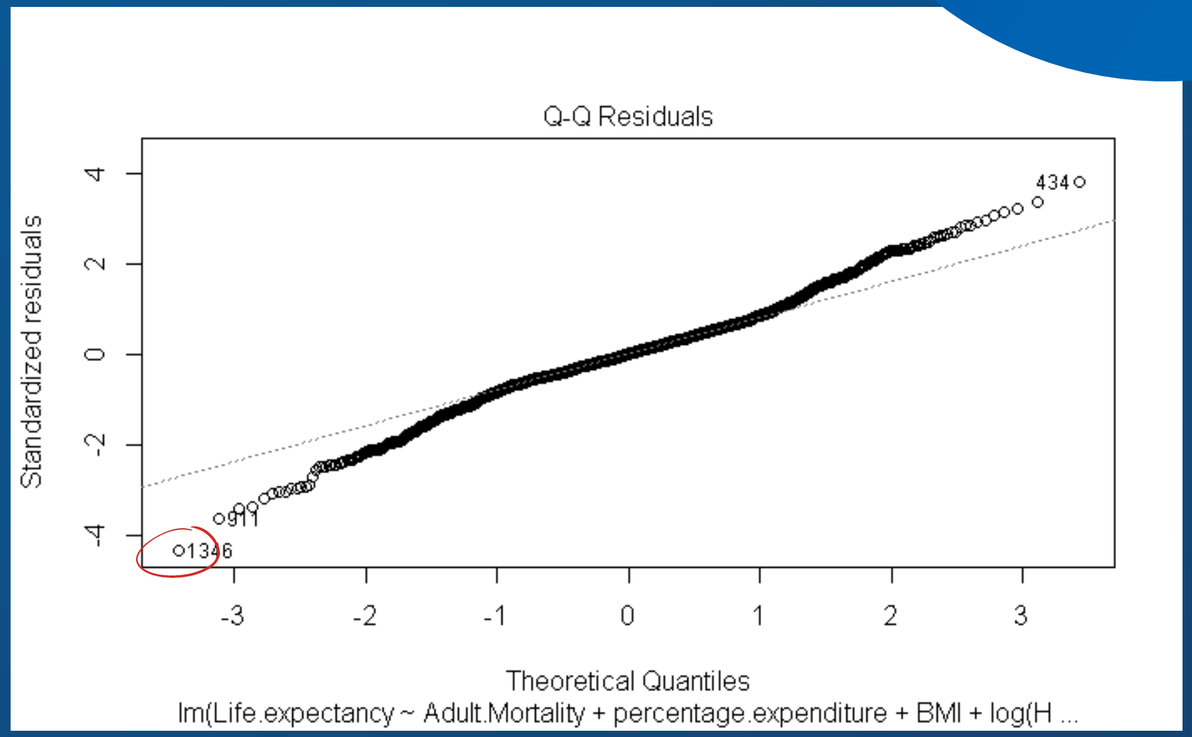
Under-Five Deaths = -0.00261

For each additional under-five death per 1000 people, life expectancy drops by ~0.0026 years (≈1 day)

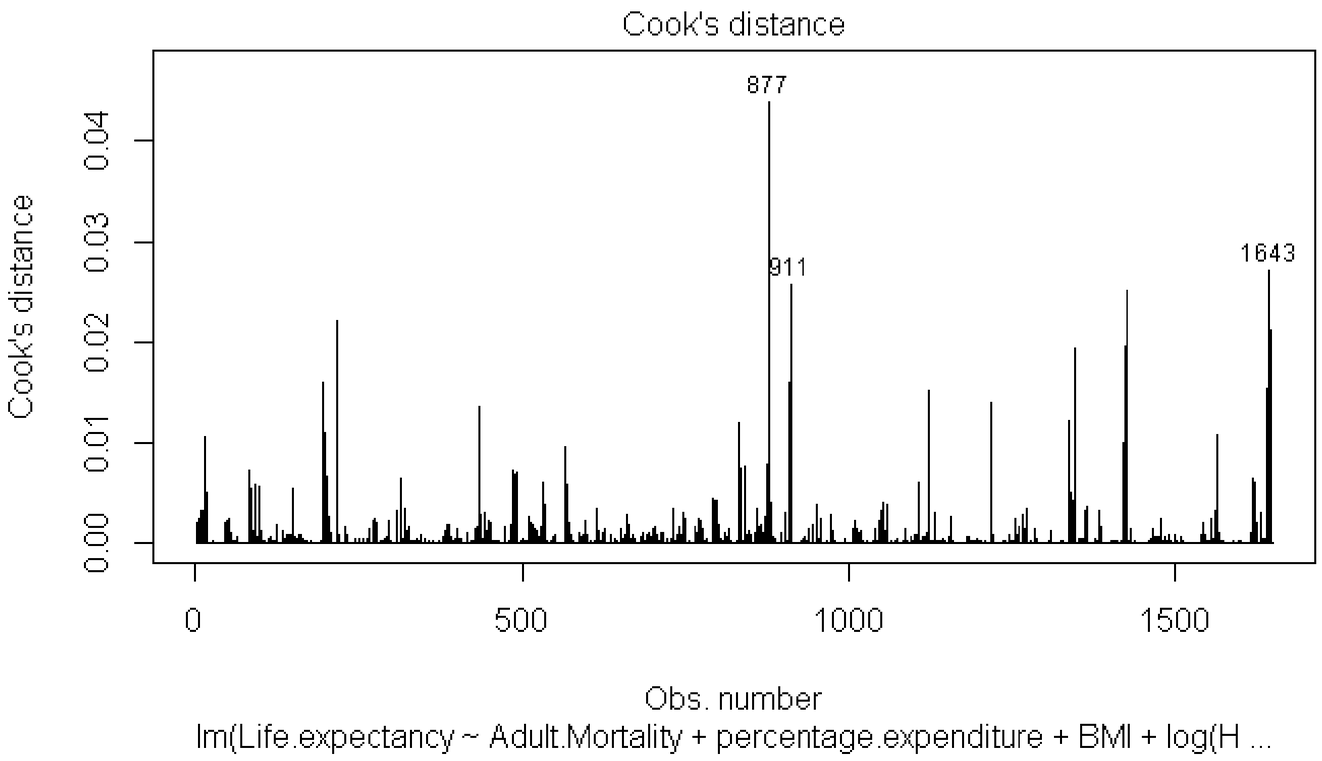
Final Model Diagnostics



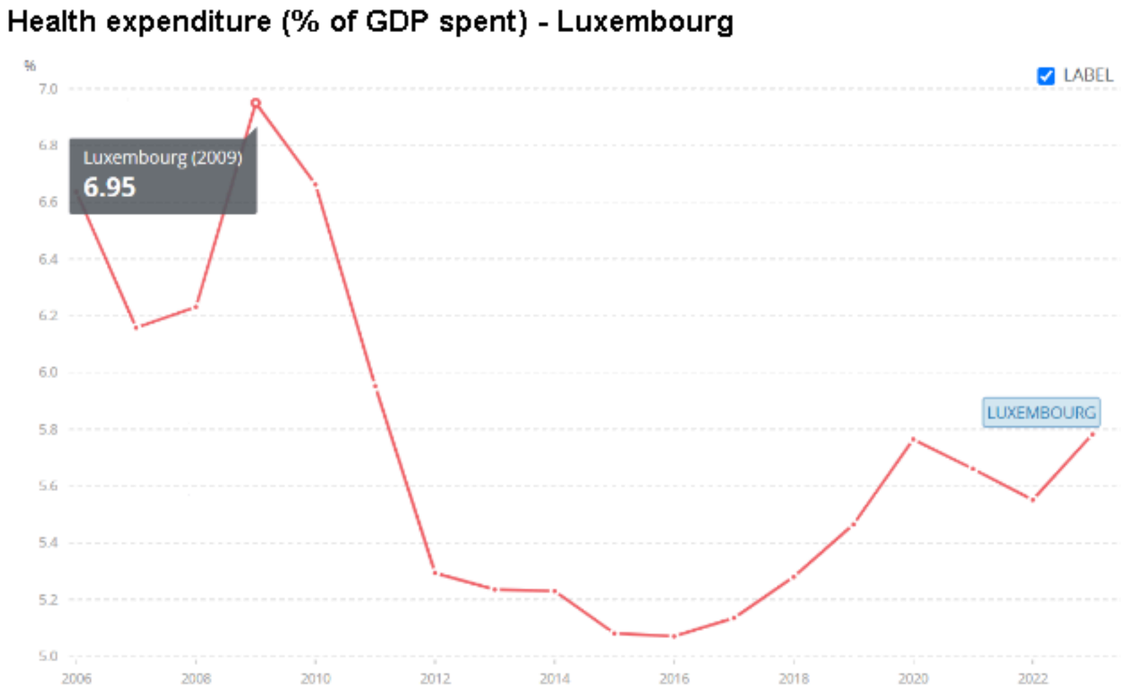
Remember this?

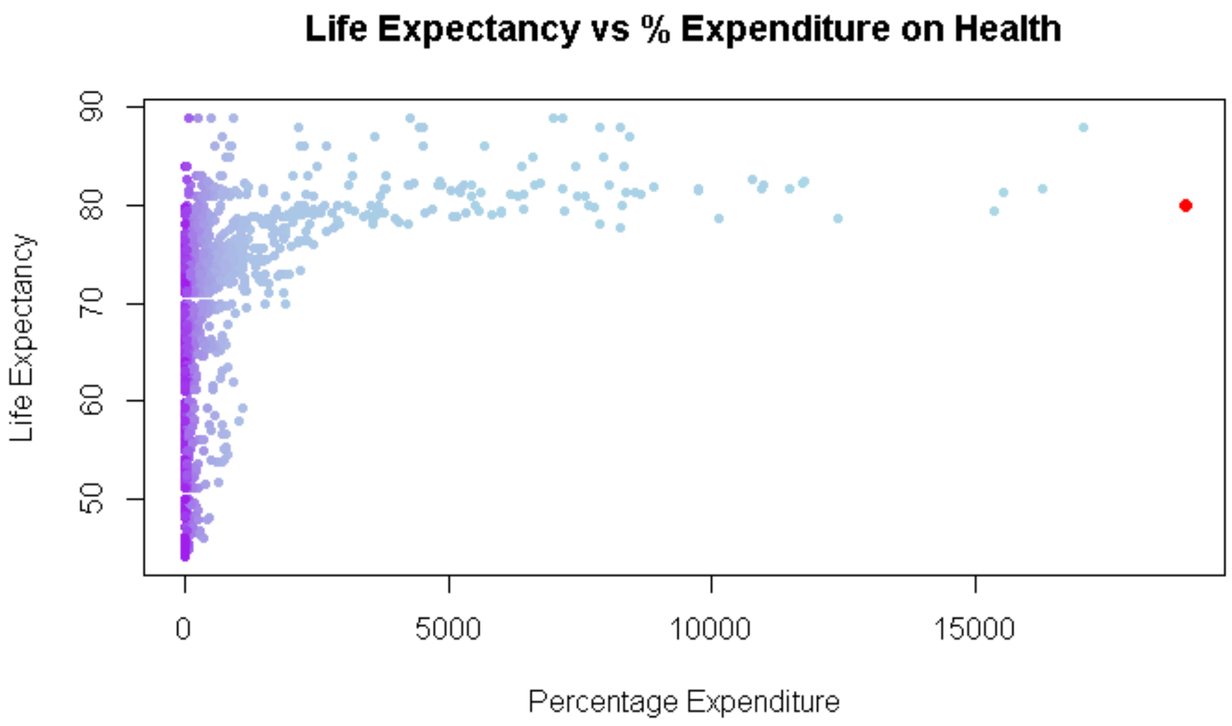


Exploring Outliers



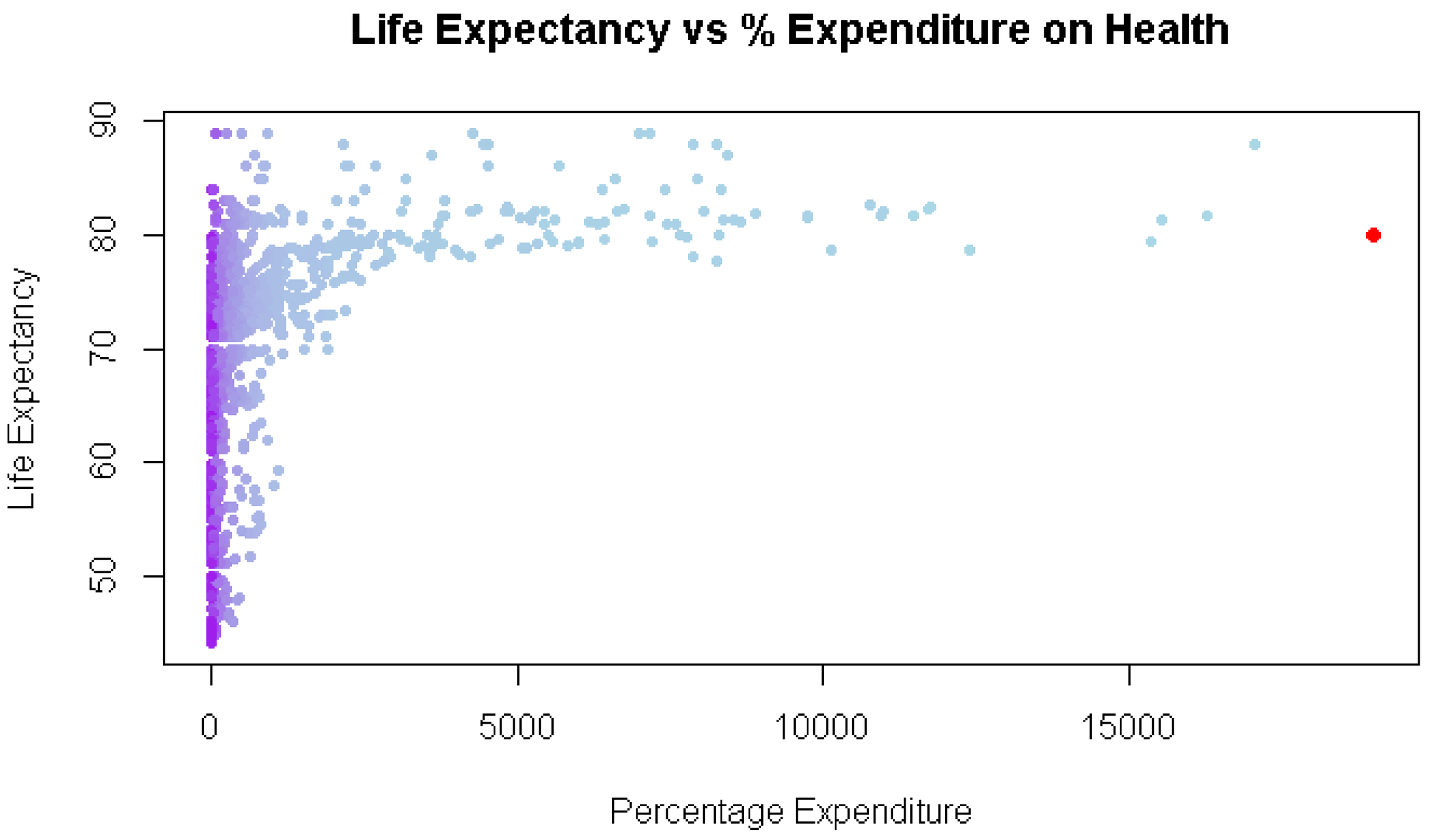
Luxembourg's High Spending in 2009





https://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS?end=2023&locations=LU&start=2000

Reminder: Outliers can emerge not from bad data, but from unusual but real conditions


Life Expectancy vs % Expenditure on Health
Life Expectancy
90
80
70
60
50
0
5000
10000
15000
Percentage Expenditure

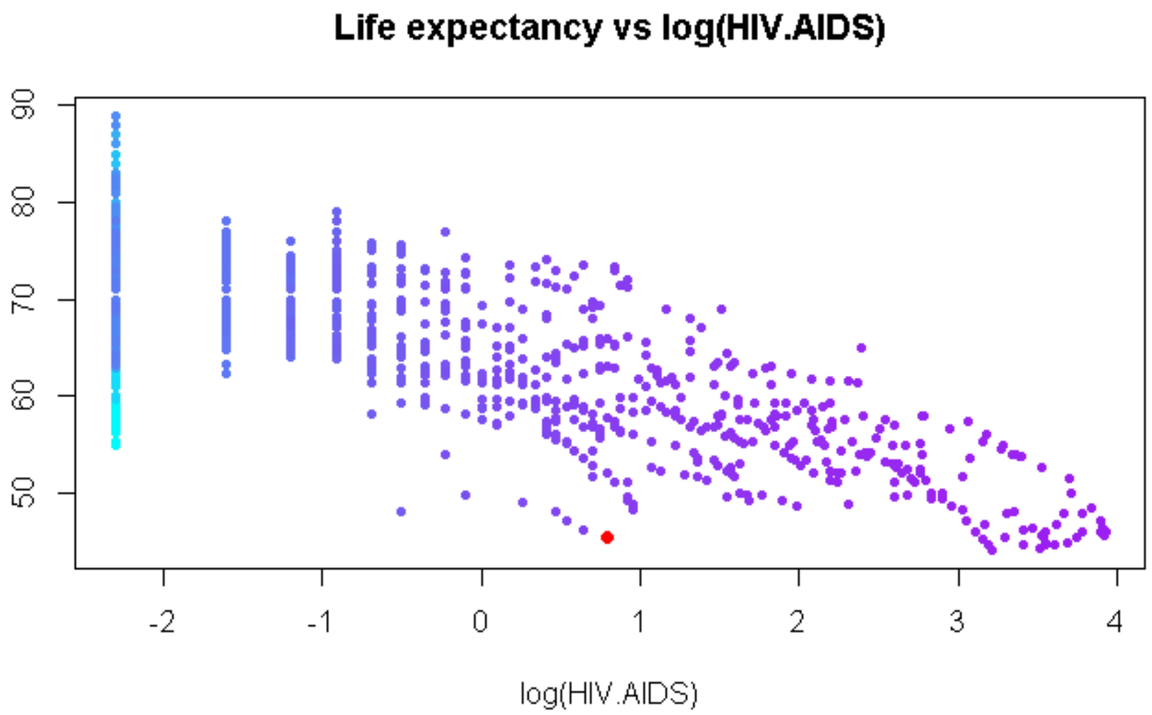
HIV births in Sierra Leone in 2008

SIERRA LEONE 2023

HIV and AIDS Estimates

Adults and children living with HIV	**77 000 [64 000 - 92 000]**
Adults aged 15 and over living with HIV	71 000 [59 000 - 86 000]
Women aged 15 and over living with HIV	45 000 [37 000 - 55 000]
Men aged 15 and over living with HIV	26 000 [21 000 - 31 000]
Children aged 0 to 14 living with HIV	6200 [4600 - 8000]
Adult aged 15 to 49 HIV prevalence rate	**1.4 [1.1 - 1.7]**
Women aged 15 to 49 HIV prevalence rate	1.8 [1.5 - 2.2]
Men aged 15 to 49 HIV prevalence rate	0.9 [0.7 - 1.1]

https://www.unaids.org/en/regionscountries/countries/sierraleone



Findings

1. Countries that invest more in public health tend to have slightly higher life expectancy, suggesting that even small increases in healthcare spending can contribute to longer lives

2. Economic resource distribution has one of the strongest effects which means countries with more equal access to income and resources see significantly higher life expectancy

3. Education levels are closely linked to life expectancy. On average, each additional year of schooling is associated with a meaningful increase in how long people live

Sources

https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-whoph text

https://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS?end=2023&locations=LU&start=2006

https://www.unaids.org/en/regionscountries/countries/sierraleone

THANK YOU