IS 445 Final Project Data Visualizations

Rachael Gu

2023-01-08

Link to Data Set: https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who (https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who)

```
setwd("C:/Users/Rachael/Desktop/IS 445")
data <- read.csv(file="Life Expectancy Data.csv", header=TRUE, sep=",")

## clean data

# replace blanks with NA
data[data == "" | data == " "] <- NA

# delete any rows containing missing data
data <- na.omit(data)

head(data)
```

```
##       Country Year     Status Life.expectancy Adult.Mortality infant.deaths
## 1 Afghanistan 2015 Developing            65.0             263            62
## 2 Afghanistan 2014 Developing            59.9             271            64
## 3 Afghanistan 2013 Developing            59.9             268            66
## 4 Afghanistan 2012 Developing            59.5             272            69
## 5 Afghanistan 2011 Developing            59.2             275            71
## 6 Afghanistan 2010 Developing            58.8             279            74
##   Alcohol percentage.expenditure Hepatitis.B Measles  BMI under.five.deaths
## 1    0.01              71.279624          65    1154 19.1                83
## 2    0.01              73.523582          62     492 18.6                86
## 3    0.01              73.219243          64     430 18.1                89
## 4    0.01              78.184215          67    2787 17.6                93
## 5    0.01               7.097109          68    3013 17.2                97
## 6    0.01              79.679367          66    1989 16.7               102
##   Polio Total.expenditure Diphtheria HIV.AIDS       GDP Population
## 1     6              8.16         65      0.1 584.25921   33736494
## 2    58              8.18         62      0.1 612.69651     327582
## 3    62              8.13         64      0.1 631.74498   31731688
## 4    67              8.52         67      0.1 669.95900    3696958
## 5    68              7.87         68      0.1  63.53723    2978599
## 6    66              9.20         66      0.1 553.32894    2883167
##   thinness..1.19.years thinness.5.9.years Income.composition.of.resources
## 1                 17.2                17.3                           0.479
## 2                 17.5                17.5                           0.476
## 3                 17.7                17.7                           0.470
## 4                 17.9                18.0                           0.463
## 5                 18.2                18.2                           0.454
## 6                 18.4                18.4                           0.448
##   Schooling
## 1      10.1
## 2      10.0
## 3       9.9
## 4       9.8
## 5       9.5
## 6       9.2
```

Stepwise Regression The final model includes Country, Year, HIV.AIDS, Schooling, thinness.5.9.years, Alcohol, Hepatitis.B, Income.composition.of.resources, under.five.deaths, and infant.deaths

```
library(alr4)
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
## Loading required package: effects
```

```
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

```
step(object=lm(Life.expectancy ~ 1, data=data), scope= ~ Country + Year + Status + Adult.Mor
tality + infant.deaths + Alcohol + percentage.expenditure + Hepatitis.B + Measles + BMI + un
der.five.deaths + Polio + Total.expenditure + Diphtheria + HIV.AIDS + GDP + Population + thi
nness..1.19.years + thinness.5.9.years + Income.composition.of.resources + Schooling, direct
ion="both")
```

```
## Start:  AIC=7172.14
## Life.expectancy ~ 1
##
##                                  Df Sum of Sq    RSS    AIC
## + Country                       132    119172   8357 2942.2
## + Schooling                       1     67520  60009 5931.1
## + Income.composition.of.resources   1     66310  61219 5964.0
## + Adult.Mortality                 1     62941  64589 6052.3
## + HIV.AIDS                        1     44730  82799 6461.9
## + BMI                             1     37469  90060 6600.5
## + thinness..1.19.years            1     26732 100797 6786.2
## + thinness.5.9.years              1     26694 100836 6786.9
## + Status                          1     25005 102525 6814.3
## + GDP                             1     24838 102691 6816.9
## + percentage.expenditure          1     21399 106130 6871.3
## + Alcohol                         1     20683 106846 6882.3
## + Diphtheria                      1     14858 112671 6969.9
## + Polio                           1     13661 113868 6987.3
## + Hepatitis.B                     1      5098 122431 7106.9
## + under.five.deaths               1      4714 122815 7112.0
## + Total.expenditure               1      3893 123636 7123.0
## + infant.deaths                   1      3646 123884 7126.3
## + Measles                         1       605 126924 7166.3
## + Year                            1       329 127201 7169.9
## <none>                                        127529 7172.1
## + Population                      1        63 127466 7173.3
##
## Step:  AIC=2942.18
## Life.expectancy ~ Country
##
##                                  Df Sum of Sq    RSS    AIC
## + Year                            1      2970   5386 2220.0
## + HIV.AIDS                        1      2265   6092 2422.8
## + Schooling                       1      1744   6612 2558.1
## + Income.composition.of.resources   1       719   7638 2795.8
## + Hepatitis.B                     1       289   8068 2886.2
## + Alcohol                         1       285   8071 2886.9
## + Adult.Mortality                 1       240   8117 2896.2
## + under.five.deaths               1       205   8152 2903.3
## + infant.deaths                   1       160   8197 2912.3
## + Diphtheria                      1       121   8235 2920.0
## + thinness..1.19.years            1        84   8273 2927.5
## + GDP                             1        84   8273 2927.5
## + thinness.5.9.years              1        55   8302 2933.3
## + percentage.expenditure          1        51   8306 2934.2
## + Polio                           1        37   8320 2936.9
## + Total.expenditure               1        35   8322 2937.3
## + Measles                         1        35   8322 2937.3
## + BMI                             1        33   8324 2937.7
## <none>                                         8357 2942.2
```

```
## + Population                         1         0   8357 2944.2
## - Country                          132    119172 127529 7172.1
##
## Step:  AIC=2219.96
## Life.expectancy ~ Country + Year
##
##                                  Df Sum of Sq    RSS    AIC
## + HIV.AIDS                        1      1100   4286 1845.2
## + Adult.Mortality                 1        55   5332 2205.1
## + Alcohol                         1        46   5341 2207.9
## + Schooling                       1        24   5362 2214.6
## + Diphtheria                      1        17   5369 2216.7
## + Hepatitis.B                     1        14   5373 2217.8
## + under.five.deaths               1         8   5378 2219.5
## <none>                                        5386 2220.0
## + BMI                             1         5   5382 2220.4
## + Polio                           1         3   5384 2221.1
## + infant.deaths                   1         3   5384 2221.2
## + GDP                             1         1   5385 2221.5
## + Measles                         1         1   5385 2221.6
## + percentage.expenditure          1         1   5385 2221.6
## + Income.composition.of.resources 1         1   5385 2221.6
## + thinness.5.9.years              1         0   5386 2221.9
## + thinness..1.19.years            1         0   5386 2221.9
## + Population                      1         0   5386 2222.0
## + Total.expenditure               1         0   5386 2222.0
## - Year                            1      2970   8357 2942.2
## - Country                       132    121814 127201 7169.9
##
## Step:  AIC=1845.25
## Life.expectancy ~ Country + Year + HIV.AIDS
##
##                                  Df Sum of Sq   RSS    AIC
## + Schooling                       1        42  4244 1830.8
## + thinness.5.9.years              1        21  4265 1839.1
## + Alcohol                         1        12  4275 1842.8
## + under.five.deaths               1        11  4275 1843.0
## + Hepatitis.B                     1        10  4277 1843.6
## + Income.composition.of.resources 1         7  4279 1844.5
## + thinness..1.19.years            1         7  4279 1844.5
## + infant.deaths                   1         6  4280 1844.8
## + Diphtheria                      1         6  4281 1845.1
## <none>                                        4286 1845.2
## + Measles                         1         4  4282 1845.6
## + Adult.Mortality                 1         4  4282 1845.7
## + Total.expenditure               1         2  4285 1846.6
## + Polio                           1         1  4285 1846.9
## + BMI                             1         0  4286 1847.1
## + Population                      1         0  4286 1847.2
## + percentage.expenditure          1         0  4286 1847.2
```

```
## + GDP                               1         0  4286 1847.2
## - HIV.AIDS                          1      1100  5386 2220.0
## - Year                              1      1805  6092 2422.8
## - Country                         132     78448 82735 6462.6
##
## Step:  AIC=1830.84
## Life.expectancy ~ Country + Year + HIV.AIDS + Schooling
##
##                                   Df Sum of Sq   RSS    AIC
## + thinness.5.9.years               1      20.9  4223 1824.7
## + Alcohol                          1      13.7  4230 1827.5
## + Hepatitis.B                      1       8.9  4235 1829.4
## + thinness..1.19.years             1       7.5  4236 1829.9
## + Income.composition.of.resources  1       7.3  4237 1830.0
## + under.five.deaths                1       6.6  4237 1830.3
## <none>                                         4244 1830.8
## + Diphtheria                       1       4.6  4239 1831.0
## + Adult.Mortality                  1       3.6  4240 1831.4
## + infant.deaths                    1       3.2  4241 1831.6
## + Measles                          1       2.7  4241 1831.8
## + Total.expenditure                1       1.4  4243 1832.3
## + Polio                            1       0.6  4243 1832.6
## + BMI                              1       0.3  4244 1832.7
## + GDP                              1       0.3  4244 1832.7
## + Population                       1       0.1  4244 1832.8
## + percentage.expenditure           1       0.0  4244 1832.8
## - Schooling                        1      42.4  4286 1845.2
## - Year                             1     684.3  4928 2075.4
## - HIV.AIDS                         1    1118.5  5362 2214.6
## - Country                        132   29642.9 33887 4992.7
##
## Step:  AIC=1824.7
## Life.expectancy ~ Country + Year + HIV.AIDS + Schooling + thinness.5.9.years
##
##                                   Df Sum of Sq   RSS    AIC
## + Alcohol                          1      13.6  4209 1821.4
## + Hepatitis.B                      1       8.0  4215 1823.6
## + Income.composition.of.resources  1       7.1  4216 1823.9
## + under.five.deaths                1       6.6  4216 1824.1
## <none>                                         4223 1824.7
## + Diphtheria                       1       4.6  4218 1824.9
## + Adult.Mortality                  1       3.5  4220 1825.3
## + infant.deaths                    1       3.1  4220 1825.5
## + Measles                          1       2.1  4221 1825.9
## + Total.expenditure                1       1.5  4222 1826.1
## + Polio                            1       0.4  4223 1826.5
## + GDP                              1       0.2  4223 1826.6
## + BMI                              1       0.2  4223 1826.6
## + percentage.expenditure           1       0.0  4223 1826.7
## + thinness..1.19.years             1       0.0  4223 1826.7
```

```
## + Population                         1       0.0  4223 1826.7
## - thinness.5.9.years                 1      20.9  4244 1830.8
## - Schooling                          1      42.1  4265 1839.1
## - Year                               1     699.4  4922 2075.4
## - HIV.AIDS                           1    1139.0  5362 2216.5
## - Country                          132   28811.4 33034 4952.7
##
## Step:  AIC=1821.37
## Life.expectancy ~ Country + Year + HIV.AIDS + Schooling + thinness.5.9.years +
##     Alcohol
##
##                                   Df Sum of Sq   RSS    AIC
## + Hepatitis.B                      1       8.4  4201 1820.1
## + under.five.deaths                1       8.0  4201 1820.2
## + Income.composition.of.resources  1       7.7  4202 1820.4
## <none>                                         4209 1821.4
## + Diphtheria                       1       4.7  4205 1821.5
## + infant.deaths                    1       4.2  4205 1821.7
## + Adult.Mortality                  1       3.4  4206 1822.0
## + Measles                          1       2.3  4207 1822.5
## + Total.expenditure                1       1.6  4208 1822.7
## + Polio                            1       0.4  4209 1823.2
## + GDP                              1       0.4  4209 1823.2
## + BMI                              1       0.2  4209 1823.3
## + percentage.expenditure           1       0.1  4209 1823.3
## + thinness..1.19.years             1       0.1  4209 1823.3
## + Population                       1       0.0  4209 1823.4
## - Alcohol                          1      13.6  4223 1824.7
## - thinness.5.9.years               1      20.9  4230 1827.5
## - Schooling                        1      44.2  4254 1836.6
## - Year                             1     658.7  4868 2059.1
## - HIV.AIDS                         1    1103.4  5313 2203.2
## - Country                        132   28758.4 32968 4951.4
##
## Step:  AIC=1820.08
## Life.expectancy ~ Country + Year + HIV.AIDS + Schooling + thinness.5.9.years +
##     Alcohol + Hepatitis.B
##
##                                   Df Sum of Sq   RSS    AIC
## + Income.composition.of.resources  1       7.5  4194 1819.1
## + under.five.deaths                1       6.3  4195 1819.6
## <none>                                         4201 1820.1
## + Adult.Mortality                  1       3.7  4197 1820.6
## + infant.deaths                    1       3.0  4198 1820.9
## + Measles                          1       2.0  4199 1821.3
## + Total.expenditure                1       1.9  4199 1821.3
## - Hepatitis.B                      1       8.4  4209 1821.4
## + Diphtheria                       1       0.8  4200 1821.8
## + GDP                              1       0.5  4201 1821.9
## + BMI                              1       0.3  4201 1822.0
```

```
## + percentage.expenditure            1       0.1  4201 1822.0
## + thinness..1.19.years              1       0.0  4201 1822.1
## + Polio                             1       0.0  4201 1822.1
## + Population                        1       0.0  4201 1822.1
## - Alcohol                           1      14.0  4215 1823.6
## - thinness.5.9.years                1      19.9  4221 1825.9
## - Schooling                         1      43.7  4245 1835.1
## - Year                              1     617.6  4819 2044.2
## - HIV.AIDS                          1    1098.0  5299 2201.0
## - Country                         132   28694.1 32895 4949.7
##
## Step:  AIC=1819.15
## Life.expectancy ~ Country + Year + HIV.AIDS + Schooling + thinness.5.9.years +
##     Alcohol + Hepatitis.B + Income.composition.of.resources
##
##                                   Df Sum of Sq     RSS    AIC
## + under.five.deaths                1       6.6  4187.0 1818.6
## <none>                                         4193.6 1819.1
## + Adult.Mortality                  1       3.5  4190.1 1819.8
## + infant.deaths                    1       3.2  4190.4 1819.9
## - Income.composition.of.resources  1       7.5  4201.0 1820.1
## + Measles                          1       2.1  4191.4 1820.3
## - Hepatitis.B                      1       8.2  4201.7 1820.4
## + Total.expenditure                1       1.8  4191.8 1820.4
## + GDP                              1       0.6  4192.9 1820.9
## + BMI                              1       0.5  4193.0 1820.9
## + Diphtheria                       1       0.4  4193.1 1821.0
## + percentage.expenditure           1       0.2  4193.4 1821.1
## + Polio                            1       0.1  4193.5 1821.1
## + thinness..1.19.years             1       0.0  4193.5 1821.1
## + Population                       1       0.0  4193.6 1821.1
## - Alcohol                          1      14.6  4208.2 1822.9
## - thinness.5.9.years               1      19.7  4213.3 1824.9
## - Schooling                        1      43.9  4237.5 1834.3
## - Year                             1     495.7  4689.2 2001.4
## - HIV.AIDS                         1    1103.9  5297.4 2202.5
## - Country                        132   24254.4 28447.9 4712.2
##
## Step:  AIC=1818.55
## Life.expectancy ~ Country + Year + HIV.AIDS + Schooling + thinness.5.9.years +
##     Alcohol + Hepatitis.B + Income.composition.of.resources +
##     under.five.deaths
##
##                                   Df Sum of Sq     RSS    AIC
## + infant.deaths                    1      27.4  4159.6 1809.7
## <none>                                         4187.0 1818.6
## + Adult.Mortality                  1       3.8  4183.2 1819.0
## - Hepatitis.B                      1       6.5  4193.5 1819.1
## - under.five.deaths                1       6.6  4193.6 1819.1
## - Income.composition.of.resources  1       7.7  4194.7 1819.6
```

```
## + Total.expenditure                 1       1.8  4185.1 1819.8
## + Measles                           1       1.2  4185.8 1820.1
## + Population                        1       0.7  4186.3 1820.3
## + GDP                               1       0.7  4186.3 1820.3
## + BMI                               1       0.5  4186.5 1820.3
## + Diphtheria                        1       0.4  4186.5 1820.4
## + percentage.expenditure            1       0.2  4186.7 1820.5
## + Polio                             1       0.1  4186.9 1820.5
## + thinness..1.19.years              1       0.1  4186.9 1820.5
## - Alcohol                           1      15.9  4202.9 1822.8
## - thinness.5.9.years                1      19.8  4206.8 1824.4
## - Schooling                         1      39.5  4226.5 1832.1
## - Year                              1     486.6  4673.5 1997.8
## - HIV.AIDS                          1    1104.8  5291.8 2202.7
## - Country                         132   24235.1 28422.1 4712.7
##
## Step:  AIC=1809.74
## Life.expectancy ~ Country + Year + HIV.AIDS + Schooling + thinness.5.9.years +
##     Alcohol + Hepatitis.B + Income.composition.of.resources +
##     under.five.deaths + infant.deaths
##
##                                   Df Sum of Sq     RSS    AIC
## <none>                                         4159.6 1809.7
## + Adult.Mortality                  1       3.9  4155.7 1810.2
## - Hepatitis.B                      1       6.6  4166.2 1810.3
## + Measles                          1       2.8  4156.8 1810.6
## - Income.composition.of.resources  1       7.7  4167.3 1810.8
## + Total.expenditure                1       1.7  4157.9 1811.1
## + GDP                              1       0.8  4158.8 1811.4
## + BMI                              1       0.6  4159.1 1811.5
## + Population                       1       0.4  4159.2 1811.6
## + thinness..1.19.years             1       0.3  4159.3 1811.6
## + percentage.expenditure           1       0.2  4159.4 1811.6
## + Diphtheria                       1       0.1  4159.5 1811.7
## + Polio                            1       0.1  4159.6 1811.7
## - Alcohol                          1      12.3  4171.9 1812.6
## - thinness.5.9.years               1      22.9  4182.5 1816.8
## - infant.deaths                    1      27.4  4187.0 1818.6
## - under.five.deaths                1      30.8  4190.4 1819.9
## - Schooling                        1      38.6  4198.2 1823.0
## - Year                             1     489.2  4648.8 1991.1
## - HIV.AIDS                         1    1071.3  5230.9 2185.6
## - Country                        132   22622.6 26782.2 4616.7
```

```
##
## Call:
## lm(formula = Life.expectancy ~ Country + Year + HIV.AIDS + Schooling +
##     thinness.5.9.years + Alcohol + Hepatitis.B + Income.composition.of.resources +
##     under.five.deaths + infant.deaths, data = data)
##
## Coefficients:
##                     (Intercept)                  CountryAlbania
##                       -394.54234                        16.59783
##                   CountryAlgeria                   CountryAngola
##                        14.38306                        -6.48912
##                 CountryArgentina                   CountryArmenia
##                        15.42818                        14.91314
##                 CountryAustralia                   CountryAustria
##                        21.30262                        22.46395
##                CountryAzerbaijan                CountryBangladesh
##                        12.45992                        11.20687
##                   CountryBelarus                   CountryBelgium
##                        11.06819                        21.25262
##                    CountryBelize                    CountryBenin
##                        10.70959                         0.30509
##                    CountryBhutan    CountryBosnia and Herzegovina
##                         7.20742                        17.08574
##                  CountryBotswana                   CountryBrazil
##                         1.94032                        13.86648
##                  CountryBulgaria              CountryBurkina Faso
##                        14.16293                         0.60098
##                   CountryBurundi                CountryCabo Verde
##                        -1.02225                        13.65645
##                  CountryCambodia                  CountryCameroon
##                         7.15107                        -1.26651
##                    CountryCanada  CountryCentral African Republic
##                        22.70526                        -4.95028
##                     CountryChad                      CountryChile
##                        -3.97849                        20.06537
##                    CountryChina                   CountryColombia
##                        14.45727                        14.70889
##                   CountryComoros                CountryCosta Rica
##                         3.44643                        19.85546
##                   CountryCroatia                    CountryCyprus
##                        17.25206                        20.78216
##                  CountryDjibouti        CountryDominican Republic
##                         5.96773                        14.11573
##                   CountryEcuador               CountryEl Salvador
##                        15.98407                        13.04072
##         CountryEquatorial Guinea                   CountryEritrea
##                        -0.31239                         4.90053
##                   CountryEstonia                  CountryEthiopia
##                        14.92364                         6.27821
##                      CountryFiji                    CountryFrance
```

```
##                   9.25103                  23.22505
##               CountryGabon               CountryGeorgia
##                   6.10070                  14.92497
##              CountryGermany               CountryGhana
##                  21.81843                   3.86063
##               CountryGreece             CountryGuatemala
##                  21.94160                  14.71286
##               CountryGuinea          CountryGuinea-Bissau
##                   0.03363                   0.12242
##               CountryGuyana               CountryHaiti
##                   7.76764                   4.08298
##             CountryHonduras               CountryIndia
##                  14.98853                   6.36406
##            CountryIndonesia                CountryIraq
##                   8.44859                  11.55671
##             CountryIreland               CountryIsrael
##                  22.40642                  21.46772
##              CountryItaly               CountryJamaica
##                  22.76803                  15.88234
##              CountryJordan            CountryKazakhstan
##                  13.64986                   7.41140
##               CountryKenya              CountryKiribati
##                   2.17226                   7.02320
##               CountryLatvia              CountryLebanon
##                  14.39154                  15.15062
##              CountryLesotho              CountryLiberia
##                  -3.31379                   2.52618
##            CountryLithuania            CountryLuxembourg
##                  13.48344                  22.18566
##           CountryMadagascar               CountryMalawi
##                   5.24908                  -2.88955
##             CountryMalaysia              CountryMaldives
##                  14.29986                  15.86866
##                CountryMali               CountryMalta
##                  -0.83685                  21.63942
##           CountryMauritania            CountryMauritius
##                   5.40958                  13.21814
##               CountryMexico              CountryMongolia
##                  16.97907                   6.97281
##           CountryMontenegro              CountryMorocco
##                  15.27054                  13.56679
##           CountryMozambique              CountryMyanmar
##                  -0.01016                   6.10315
##              CountryNamibia                CountryNepal
##                   6.35474                   7.64320
##          CountryNetherlands             CountryNicaragua
##                  20.14263                  15.28472
##               CountryNiger               CountryNigeria
##                   5.16433                   0.94620
##             CountryPakistan               CountryPanama
```

```
##                        5.68184                        17.80430
##         CountryPapua New Guinea                CountryParaguay
##                        4.68449                        14.12907
##                     CountryPeru               CountryPhilippines
##                       14.86091                         8.73598
##                    CountryPoland                CountryPortugal
##                       16.37059                        20.88833
##                   CountryRomania        CountryRussian Federation
##                       15.09829                         8.67585
##                    CountryRwanda                  CountrySamoa
##                        3.88499                        15.15314
##    CountrySao Tome and Principe                 CountrySenegal
##                        7.34455                         6.22224
##                    CountrySerbia               CountrySeychelles
##                       15.07137                        13.44000
##              CountrySierra Leone           CountrySolomon Islands
##                       -9.57347                        10.26758
##              CountrySouth Africa                  CountrySpain
##                        4.25656                        22.66750
##                 CountrySri Lanka                 CountrySuriname
##                       13.15410                        12.14330
##                 CountrySwaziland                  CountrySweden
##                        3.31678                        21.27782
##     CountrySyrian Arab Republic               CountryTajikistan
##                       15.51036                         8.29392
##                  CountryThailand              CountryTimor-Leste
##                       14.02299                         6.90638
##                      CountryTogo                   CountryTonga
##                       -0.42030                        13.38678
##      CountryTrinidad and Tobago                  CountryTunisia
##                       12.59038                        14.73608
##                    CountryTurkey              CountryTurkmenistan
##                       14.69942                         6.62298
##                    CountryUganda                 CountryUkraine
##                        1.41688                        11.02748
##                  CountryUruguay                CountryUzbekistan
##                       16.72907                         9.16912
##                  CountryVanuatu                   CountryZambia
##                       13.60492                         1.21604
##                 CountryZimbabwe                             Year
##                       -0.80939                         0.22355
##                        HIV.AIDS                       Schooling
##                       -0.30520                         0.28972
##              thinness.5.9.years                         Alcohol
##                        0.07381                        -0.06372
##                     Hepatitis.B  Income.composition.of.resources
##                        0.00325                         0.98420
##               under.five.deaths                    infant.deaths
##                       -0.03528                         0.04642
```

The data story message is directed to all types of audience, both technical and non-technical. With the data visualization, readers will be able to see the change in life expectancy for the countries in the data set from 2000-2015. Huge developments in healthcare were made during this period which resulted in a longer life expectancy as seen in the graph. In 2000, most countries had a life expectancy of 70-75 years. However, in 2015, it can be seen that a lot of countries' life expectancy increased and were mostly 75-80 years. Researchers have been studying how to effectively increase life expectancy and which factors were the most influential. This visualization can be used to look at overall trends in life expectancy and make predictions on future changes.

```
data <- read.csv(file="Life Expectancy Data.csv", header=TRUE, sep=",")

library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.2
```

```
suppressWarnings({
p <- ggplot(data, aes(x = Life.expectancy)) +
  geom_histogram(fill = "cornflowerblue",
                 color = "white",
                 bins = 10) +
  facet_wrap(~Year) +
  labs(title = "Countries by Life Expectancy in Years",
       subtitle = "number of bins = 10",
       caption = "Figure 1: Histogram",
       x = "Life Expectancy (Years)",
       y = "Number of Countries")
p
})
```

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

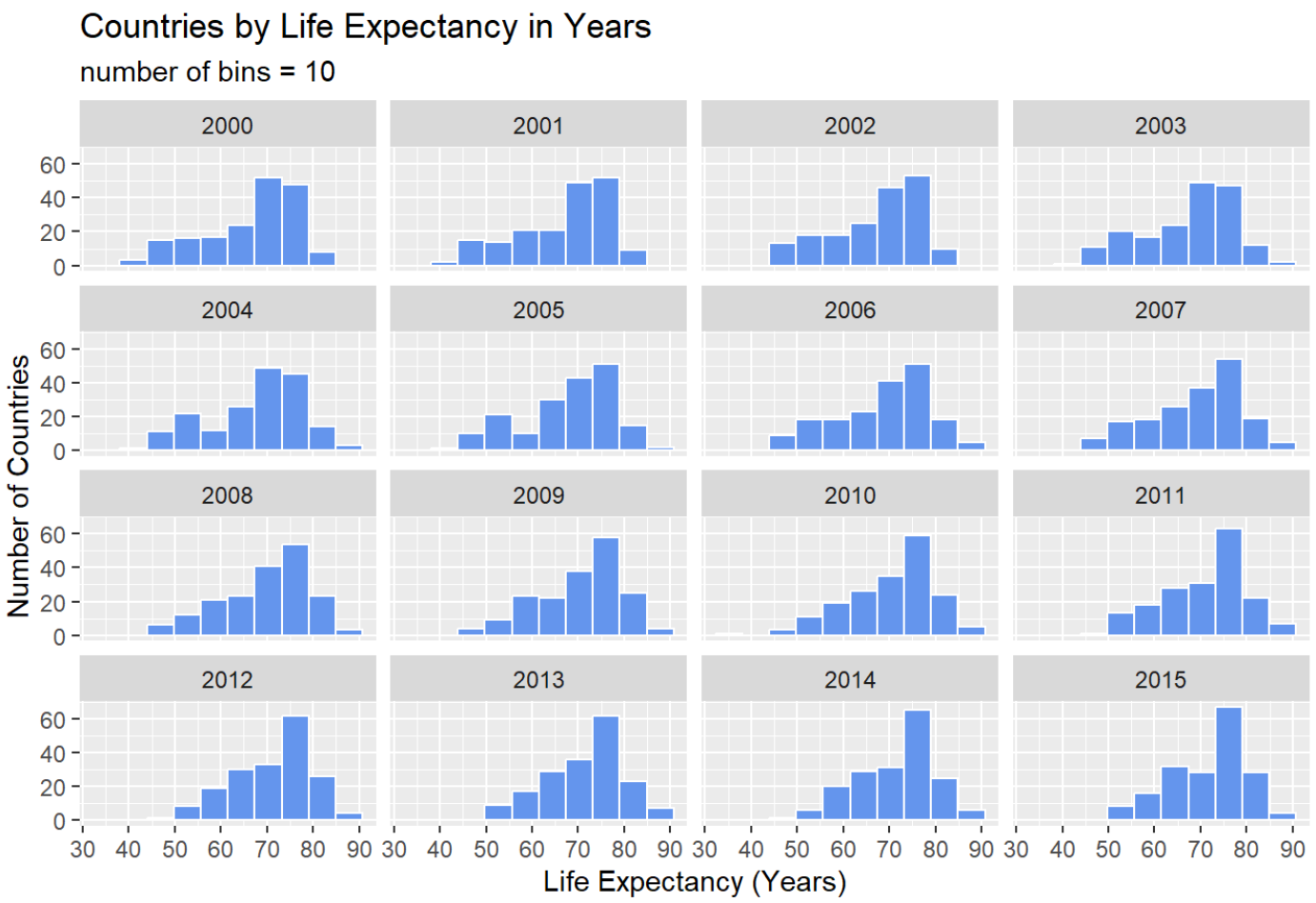

Figure 1: Histogram

The data story message is directed to all types of audience, both technical and non-technical. With the data visualization, readers will be able to see the correlations between life expectancy and different variables. Studies have shown that countries with higher levels of inequality, such as with income, education, and health, are the least efficient in their healthcare system. This graph compares life expectancy with factors that are related to immunizations, schooling, and economics. From this chart, it can be seen that there is a relatively strong positive correlation between life expectancy and schooling. There is some correlation between GDP and life expectancy (0.441) which indicates that both variables move in tandem. In other words, if GDP increases, life expectancy may also increase. HIV/AIDS and life expectancy have a negative correlation of -0.592 which means the variables move in opposite directions. This is expected because countries with lower HIV/AIDS cases tend to have a better healthcare system and longer life expectancy. Researchers have been studying how to effectively increase life expectancy and translate that into policies. This visualizations can be used to determine which factors are the most effective with improving life expectancy. In the future, policymakers may want to put more effort in increasing education opportunities to increase average life expectancy.

```
library(GGally)
```

```
## Warning: package 'GGally' was built under R version 4.2.2
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
# prepare data
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:car':
##
##     recode
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
suppressWarnings({
df <- data %>%
  mutate(gdp = as.numeric(GDP),
         measles = as.numeric(Measles)) %>%
  select(Life.expectancy, gdp, measles, HIV.AIDS, Schooling)

# create scatterplot matrix
ggpairs(df) +
  labs(title = "Correlation Between Life Expectancy (Years), GDP, Measles, HIV/AIDS, Schooli
ng",
       caption = "Figure 3: Scatterplot Matrix")
})
```

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

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 453 rows containing missing values
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 10 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 10 rows containing missing values
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 170 rows containing missing values
```

```
## Warning: Removed 453 rows containing missing values (`geom_point()`).
```

```
## Warning: Removed 448 rows containing non-finite values (`stat_density()`).
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 448 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 448 rows containing missing values
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 451 rows containing missing values
```

```
## Warning: Removed 10 rows containing missing values (`geom_point()`).
```

```
## Warning: Removed 448 rows containing missing values (`geom_point()`).
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 163 rows containing missing values
```

```
## Warning: Removed 10 rows containing missing values (`geom_point()`).
```

```
## Warning: Removed 448 rows containing missing values (`geom_point()`).
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 163 rows containing missing values
```

```
## Warning: Removed 170 rows containing missing values (`geom_point()`).
```

```
## Warning: Removed 451 rows containing missing values (`geom_point()`).
```

```
## Warning: Removed 163 rows containing missing values (`geom_point()`).
## Removed 163 rows containing missing values (`geom_point()`).
```

```
## Warning: Removed 163 rows containing non-finite values (`stat_density()`).
```

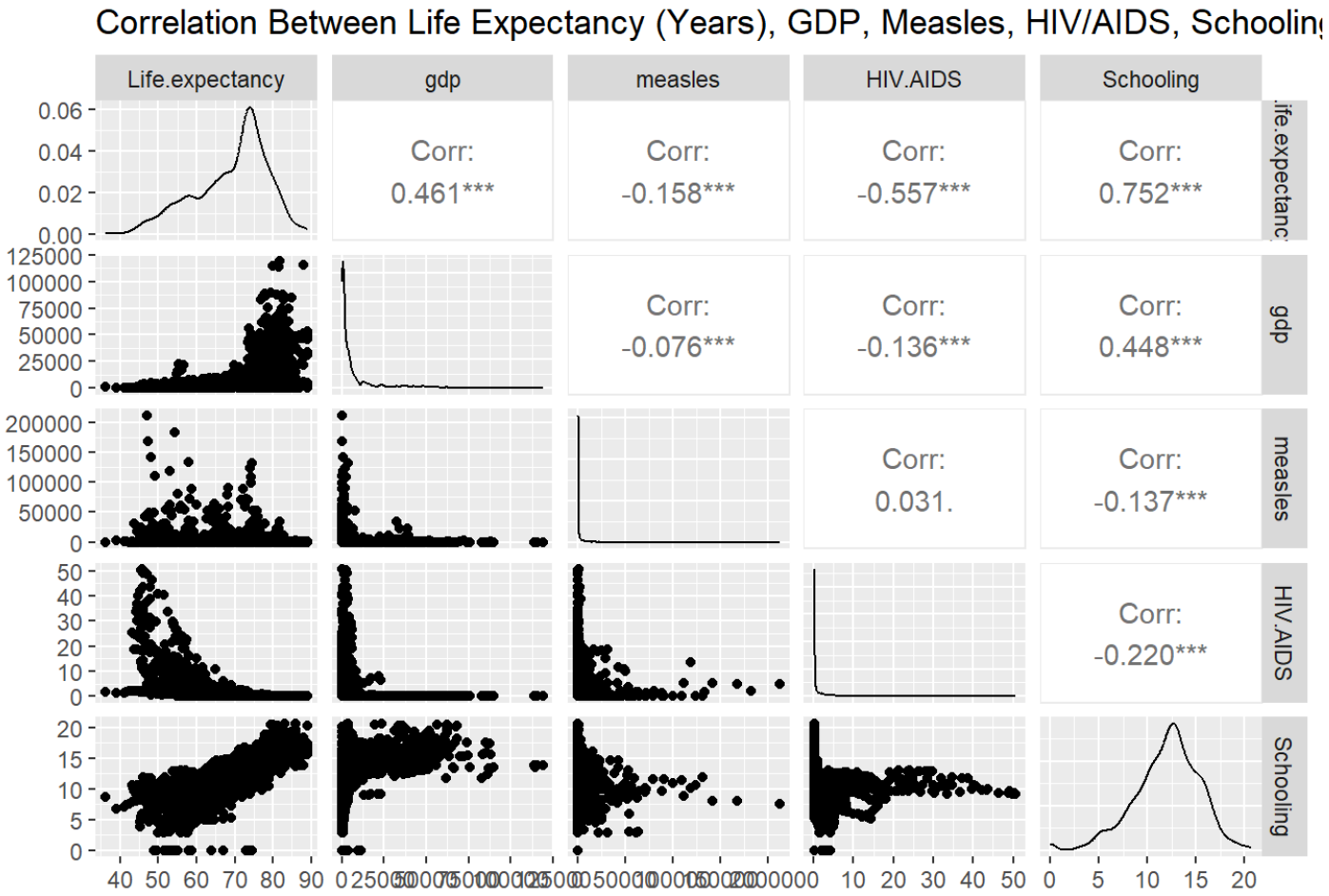

Figure 3: Scatterplot Matrix

The data story message is directed to all types of audience, both technical and non-technical. With the data visualization, readers will be able to compare the life expectancy for different countries based on population and status for 2015. Although life expectancy has increased in both developing and developed countries, it can be seen that there are global health disparities. There are still large variations in life expectancy between countries based on status. From the chart, it can be seen that all developed countries in 2015 had a life expectancy that was longer than the average of all 193 countries in the data set. Researchers have been studying global health disparities between countries based on socio-economic development. This visualization can be used to analyze how much variation there is in life expectancy for developed and developing countries. In the future, policymakers may want to put more effort in improving life expectancy for countries that are below the dashed line to close the gap in life expectancy between developed and developing countries.

```
# filter through data with the year 2015
mydf <- filter(data, Year == "2015")

# select columns
newdata <- select(mydf, Status, Life.expectancy, Population)
head(newdata)
```

```
##       Status Life.expectancy Population
## 1 Developing            65.0   33736494
## 2 Developing            77.8      28873
## 3 Developing            75.6   39871528
## 4 Developing            52.4    2785935
## 5 Developing            76.4         NA
## 6 Developing            76.3   43417765
```

```
# add annotation line and text label
suppressWarnings({
mean_life_expectancy <- mean(newdata$Life.expectancy)

ggplot(newdata,
       aes(x = log(as.numeric(Population)), y = as.numeric(Life.expectancy), color = factor
(Status))) +
  geom_point(size = 3) +
  geom_hline(yintercept = mean_life_expectancy,
             color = "darkred",
             linetype = "dashed") +
  ggplot2::annotate("text",
                    9.5,
                    mean_life_expectancy - 1,
                    label = "Mean Life Expectancy",
                    color = "darkred") +
  labs(title = "Life Expectancy by Population and Status for 2015",
       caption = "Figure 2: Graph with line annotation",
       x = "Log Population",
       y = "Life Expectancy in Years",
       color = "Status")
})
```

```
## Warning: Removed 41 rows containing missing values (`geom_point()`).
```

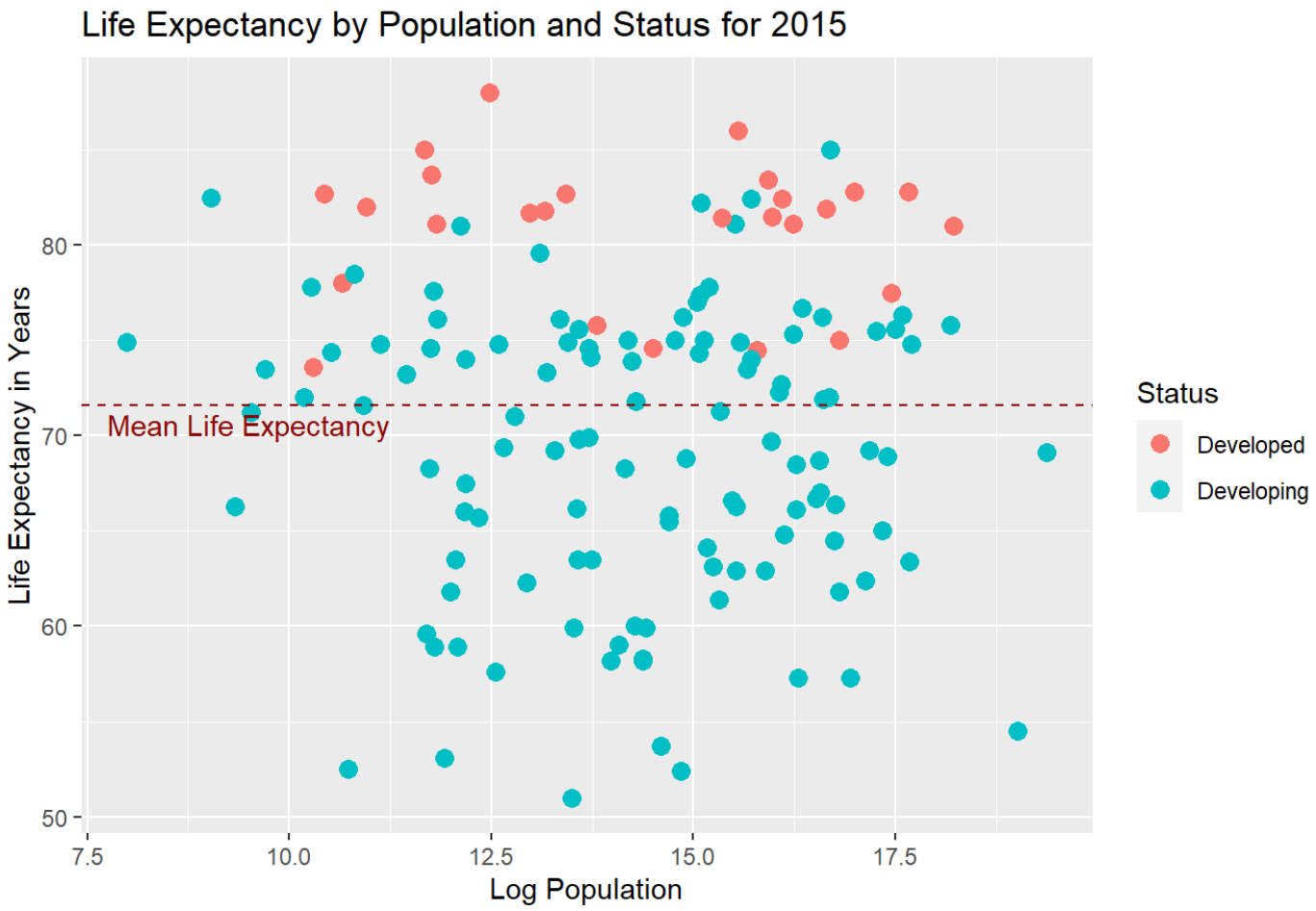

Figure 2: Graph with line annotation