

Discussion_12

Gabe Abreu

11/12/2020

Discussion 12

This data set was downloaded from Kaggle and sourced from the World Health Organization. I want to examine variables that are positively correlated with life expectancy.

```
LE_data <- read.csv('Life Expectancy Data.csv')
```

```
head(LE_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

```

```
summary(LE_data)
```

	Country	Year	Status	Life.expectancy
## Afghanistan	: 16	Min. :2000	Developed : 512	Min. :36.30
## Albania	: 16	1st Qu.:2004	Developing:2426	1st Qu.:63.10
## Algeria	: 16	Median :2008		Median :72.10
## Angola	: 16	Mean :2008		Mean :69.22
## Antigua and Barbuda	: 16	3rd Qu.:2012		3rd Qu.:75.70
## Argentina	: 16	Max. :2015		Max. :89.00
## (Other)	:2842			NA's :10
## Adult.Mortality	infant.deaths	Alcohol	percentage.expenditure	
## Min. : 1.0	Min. : 0.0	Min. : 0.0100	Min. : 0.000	
## 1st Qu.: 74.0	1st Qu.: 0.0	1st Qu.: 0.8775	1st Qu.: 4.685	
## Median :144.0	Median : 3.0	Median : 3.7550	Median : 64.913	
## Mean :164.8	Mean : 30.3	Mean : 4.6029	Mean : 738.251	
## 3rd Qu.:228.0	3rd Qu.: 22.0	3rd Qu.: 7.7025	3rd Qu.: 441.534	
## Max. :723.0	Max. :1800.0	Max. :17.8700	Max. :19479.912	
## NA's :10		NA's :194		
## Hepatitis.B	Measles	BMI	under.five.deaths	
## Min. : 1.00	Min. : 0.0	Min. : 1.00	Min. : 0.00	
## 1st Qu.:77.00	1st Qu.: 0.0	1st Qu.:19.30	1st Qu.: 0.00	
## Median :92.00	Median : 17.0	Median :43.50	Median : 4.00	
## Mean :80.94	Mean : 2419.6	Mean :38.32	Mean : 42.04	
## 3rd Qu.:97.00	3rd Qu.: 360.2	3rd Qu.:56.20	3rd Qu.: 28.00	
## Max. :99.00	Max. :212183.0	Max. :87.30	Max. :2500.00	
## NA's :553		NA's :34		
## Polio	Total.expenditure	Diphtheria	HIV.AIDS	
## Min. : 3.00	Min. : 0.370	Min. : 2.00	Min. : 0.100	
## 1st Qu.:78.00	1st Qu.: 4.260	1st Qu.:78.00	1st Qu.: 0.100	
## Median :93.00	Median : 5.755	Median :93.00	Median : 0.100	
## Mean :82.55	Mean : 5.938	Mean :82.32	Mean : 1.742	
## 3rd Qu.:97.00	3rd Qu.: 7.492	3rd Qu.:97.00	3rd Qu.: 0.800	
## Max. :99.00	Max. :17.600	Max. :99.00	Max. :50.600	
## NA's :19	NA's :226	NA's :19		
## GDP	Population	thinness..1.19.years		
## Min. : 1.68	Min. :3.400e+01	Min. : 0.10		
## 1st Qu.: 463.94	1st Qu.:1.958e+05	1st Qu.: 1.60		
## Median : 1766.95	Median :1.387e+06	Median : 3.30		
## Mean : 7483.16	Mean :1.275e+07	Mean : 4.84		
## 3rd Qu.: 5910.81	3rd Qu.:7.420e+06	3rd Qu.: 7.20		
## Max. :119172.74	Max. :1.294e+09	Max. :27.70		
## NA's :448	NA's :652	NA's :34		
## thinness.5.9.years	Income.composition.of.resources	Schooling		
## Min. : 0.10	Min. :0.0000	Min. : 0.00		
## 1st Qu.: 1.50	1st Qu.:0.4930	1st Qu.:10.10		
## Median : 3.30	Median :0.6770	Median :12.30		
## Mean : 4.87	Mean :0.6276	Mean :11.99		
## 3rd Qu.: 7.20	3rd Qu.:0.7790	3rd Qu.:14.30		
## Max. :28.60	Max. :0.9480	Max. :20.70		
## NA's :34	NA's :167	NA's :163		

```

flattenCorrMatrix <- function(cormat, pmat) {
  ut <- upper.tri(cormat)
  data.frame(
    row = rownames(cormat)[row(cormat)[ut]],
    column = rownames(cormat)[col(cormat)[ut]],
    cor = (cormat)[ut],
    p = pmat[ut]
  )
}

LE_data_df <- LE_data[, 4:length(LE_data)]

LE_data_df.cor = cor(LE_data_df, method = c("spearman"))

newCorr <- rcorr(as.matrix(LE_data_df))

flattenCorrMatrix(newCorr$r, newCorr$P)

```

##	row	column
## 1	Life.expectancy	Adult.Mortality
## 2	Life.expectancy	infant.deaths
## 3	Adult.Mortality	infant.deaths
## 4	Life.expectancy	Alcohol
## 5	Adult.Mortality	Alcohol
## 6	infant.deaths	Alcohol
## 7	Life.expectancy	percentage.expenditure
## 8	Adult.Mortality	percentage.expenditure
## 9	infant.deaths	percentage.expenditure
## 10	Alcohol	percentage.expenditure
## 11	Life.expectancy	Hepatitis.B
## 12	Adult.Mortality	Hepatitis.B
## 13	infant.deaths	Hepatitis.B
## 14	Alcohol	Hepatitis.B
## 15	percentage.expenditure	Hepatitis.B
## 16	Life.expectancy	Measles
## 17	Adult.Mortality	Measles
## 18	infant.deaths	Measles
## 19	Alcohol	Measles
## 20	percentage.expenditure	Measles
## 21	Hepatitis.B	Measles
## 22	Life.expectancy	BMI
## 23	Adult.Mortality	BMI
## 24	infant.deaths	BMI
## 25	Alcohol	BMI
## 26	percentage.expenditure	BMI
## 27	Hepatitis.B	BMI
## 28	Measles	BMI
## 29	Life.expectancy	under.five.deaths
## 30	Adult.Mortality	under.five.deaths
## 31	infant.deaths	under.five.deaths
## 32	Alcohol	under.five.deaths
## 33	percentage.expenditure	under.five.deaths
## 34	Hepatitis.B	under.five.deaths

```

## 35               Measles           under.five.deaths
## 36                   BMI           under.five.deaths
## 37               Life.expectancy   Polio
## 38               Adult.Mortality  Polio
## 39               infant.deaths    Polio
## 40                   Alcohol      Polio
## 41               percentage.expenditure Polio
## 42                   Hepatitis.B  Polio
## 43                   Measles       Polio
## 44                   BMI          Polio
## 45               under.five.deaths Polio
## 46                   Life.expectancy Total.expenditure
## 47                   Adult.Mortality Total.expenditure
## 48                   infant.deaths  Total.expenditure
## 49                   Alcohol      Total.expenditure
## 50               percentage.expenditure Total.expenditure
## 51                   Hepatitis.B  Total.expenditure
## 52                   Measles       Total.expenditure
## 53                   BMI          Total.expenditure
## 54               under.five.deaths Total.expenditure
## 55                   Polio        Total.expenditure
## 56                   Life.expectancy Diphtheria
## 57                   Adult.Mortality Diphtheria
## 58                   infant.deaths  Diphtheria
## 59                   Alcohol      Diphtheria
## 60               percentage.expenditure Diphtheria
## 61                   Hepatitis.B  Diphtheria
## 62                   Measles       Diphtheria
## 63                   BMI          Diphtheria
## 64               under.five.deaths Diphtheria
## 65                   Polio        Diphtheria
## 66               Total.expenditure Diphtheria
## 67                   Life.expectancy HIV.AIDS
## 68                   Adult.Mortality HIV.AIDS
## 69                   infant.deaths  HIV.AIDS
## 70                   Alcohol      HIV.AIDS
## 71               percentage.expenditure HIV.AIDS
## 72                   Hepatitis.B  HIV.AIDS
## 73                   Measles       HIV.AIDS
## 74                   BMI          HIV.AIDS
## 75               under.five.deaths HIV.AIDS
## 76                   Polio        HIV.AIDS
## 77               Total.expenditure HIV.AIDS
## 78                   Diphtheria   HIV.AIDS
## 79                   Life.expectancy GDP
## 80                   Adult.Mortality GDP
## 81                   infant.deaths  GDP
## 82                   Alcohol      GDP
## 83               percentage.expenditure GDP
## 84                   Hepatitis.B  GDP
## 85                   Measles       GDP
## 86                   BMI          GDP
## 87               under.five.deaths GDP
## 88                   Polio        GDP

```

```

## 89      Total.expenditure          GDP
## 90      Diphtheria                GDP
## 91      HIV.AIDS                 GDP
## 92      Life.expectancy          Population
## 93      Adult.Mortality          Population
## 94      infant.deaths           Population
## 95      Alcohol                  Population
## 96      percentage.expenditure Population
## 97      Hepatitis.B              Population
## 98      Measles                  Population
## 99      BMI                      Population
## 100     under.five.deaths        Population
## 101     Polio                     Population
## 102     Total.expenditure        Population
## 103     Diphtheria                Population
## 104     HIV.AIDS                 Population
## 105     GDP                      Population
## 106     Life.expectancy          thinness..1.19.years
## 107     Adult.Mortality          thinness..1.19.years
## 108     infant.deaths           thinness..1.19.years
## 109     Alcohol                  thinness..1.19.years
## 110     percentage.expenditure thinness..1.19.years
## 111     Hepatitis.B              thinness..1.19.years
## 112     Measles                  thinness..1.19.years
## 113     BMI                      thinness..1.19.years
## 114     under.five.deaths        thinness..1.19.years
## 115     Polio                     thinness..1.19.years
## 116     Total.expenditure        thinness..1.19.years
## 117     Diphtheria                thinness..1.19.years
## 118     HIV.AIDS                 thinness..1.19.years
## 119     GDP                      thinness..1.19.years
## 120     Population                thinness..1.19.years
## 121     Life.expectancy          thinness.5.9.years
## 122     Adult.Mortality          thinness.5.9.years
## 123     infant.deaths           thinness.5.9.years
## 124     Alcohol                  thinness.5.9.years
## 125     percentage.expenditure thinness.5.9.years
## 126     Hepatitis.B              thinness.5.9.years
## 127     Measles                  thinness.5.9.years
## 128     BMI                      thinness.5.9.years
## 129     under.five.deaths        thinness.5.9.years
## 130     Polio                     thinness.5.9.years
## 131     Total.expenditure        thinness.5.9.years
## 132     Diphtheria                thinness.5.9.years
## 133     HIV.AIDS                 thinness.5.9.years
## 134     GDP                      thinness.5.9.years
## 135     Population                thinness.5.9.years
## 136     thinness..1.19.years       thinness.5.9.years
## 137     Life.expectancy          Income.composition.of.resources
## 138     Adult.Mortality          Income.composition.of.resources
## 139     infant.deaths           Income.composition.of.resources
## 140     Alcohol                  Income.composition.of.resources
## 141     percentage.expenditure Income.composition.of.resources
## 142     Hepatitis.B              Income.composition.of.resources

```

```

## 143               Measles Income.composition.of.resources
## 144               BMI Income.composition.of.resources
## 145 under.five.deaths Income.composition.of.resources
## 146               Polio Income.composition.of.resources
## 147 Total.expenditure Income.composition.of.resources
## 148               Diphtheria Income.composition.of.resources
## 149               HIV.AIDS Income.composition.of.resources
## 150               GDP Income.composition.of.resources
## 151 Population Income.composition.of.resources
## 152 thinness..1.19.years Income.composition.of.resources
## 153 thinness.5.9.years Income.composition.of.resources
## 154 Life.expectancy Schooling
## 155 Adult.Mortality Schooling
## 156 infant.deaths Schooling
## 157 Alcohol Schooling
## 158 percentage.expenditure Schooling
## 159 Hepatitis.B Schooling
## 160 Measles Schooling
## 161 BMI Schooling
## 162 under.five.deaths Schooling
## 163 Polio Schooling
## 164 Total.expenditure Schooling
## 165 Diphtheria Schooling
## 166 HIV.AIDS Schooling
## 167 GDP Schooling
## 168 Population Schooling
## 169 thinness..1.19.years Schooling
## 170 thinness.5.9.years Schooling
## 171 Income.composition.of.resources Schooling

##          cor      p
## 1 -0.696359314 0.000000e+00
## 2 -0.196557177 0.000000e+00
## 3  0.078756012 1.986753e-05
## 4  0.404876761 0.000000e+00
## 5 -0.195848196 0.000000e+00
## 6 -0.115637677 1.239093e-09
## 7  0.381863503 0.000000e+00
## 8 -0.242859528 0.000000e+00
## 9 -0.085612222 3.367554e-06
## 10 0.341285313 0.000000e+00
## 11 0.256761948 0.000000e+00
## 12 -0.162476325 1.776357e-15
## 13 -0.223566281 0.000000e+00
## 14 0.087548711 3.706709e-05
## 15 0.016273693 4.269710e-01
## 16 -0.157585804 0.000000e+00
## 17 0.031176412 9.166560e-02
## 18 0.501128342 0.000000e+00
## 19 -0.051826674 6.618874e-03
## 20 -0.056595677 2.149003e-03
## 21 -0.120529372 3.533331e-09
## 22 0.567693548 0.000000e+00
## 23 -0.387016784 0.000000e+00
## 24 -0.227278888 0.000000e+00

```

```

## 25  0.330408460 0.000000e+00
## 26  0.228699753 0.000000e+00
## 27  0.150379532 1.834088e-13
## 28  -0.175977063 0.000000e+00
## 29  -0.222529116 0.000000e+00
## 30  0.094146127 3.334215e-07
## 31  0.996628882 0.000000e+00
## 32  -0.112370397 3.583787e-09
## 33  -0.087852306 1.850661e-06
## 34  -0.233126251 0.000000e+00
## 35  0.507808707 0.000000e+00
## 36  -0.237668522 0.000000e+00
## 37  0.465555806 0.000000e+00
## 38  -0.274822815 0.000000e+00
## 39  -0.170688559 0.000000e+00
## 40  0.221733797 0.000000e+00
## 41  0.147259463 1.332268e-15
## 42  0.486170773 0.000000e+00
## 43  -0.136166014 1.489919e-13
## 44  0.284568764 0.000000e+00
## 45  -0.188720213 0.000000e+00
## 46  0.218086374 0.000000e+00
## 47  -0.115280689 1.859234e-09
## 48  -0.128616342 1.782818e-11
## 49  0.296941560 0.000000e+00
## 50  0.174419689 0.000000e+00
## 51  0.058280304 6.297607e-03
## 52  -0.106240588 2.925670e-08
## 53  0.242502604 0.000000e+00
## 54  -0.130148312 1.020739e-11
## 55  0.137330249 7.398526e-13
## 56  0.479494864 0.000000e+00
## 57  -0.275131358 0.000000e+00
## 58  -0.175171496 0.000000e+00
## 59  0.222020171 0.000000e+00
## 60  0.143624426 6.217249e-15
## 61  0.611494949 0.000000e+00
## 62  -0.141881938 1.332268e-14
## 63  0.283147336 0.000000e+00
## 64  -0.195668288 0.000000e+00
## 65  0.673553321 0.000000e+00
## 66  0.152753524 1.332268e-15
## 67  -0.556556253 0.000000e+00
## 68  0.523820508 0.000000e+00
## 69  0.025231318 1.715448e-01
## 70  -0.048844563 1.049749e-02
## 71  -0.097856819 1.068527e-07
## 72  -0.112675448 3.441238e-08
## 73  0.030898718 9.403233e-02
## 74  -0.243716531 0.000000e+00
## 75  0.038061512 3.911929e-02
## 76  -0.159559542 0.000000e+00
## 77  -0.001388836 9.423688e-01
## 78  -0.164860095 0.000000e+00

```

```

## 79   0.461455193 0.000000e+00
## 80   -0.296049318 0.000000e+00
## 81   -0.108427363 5.834695e-08
## 82    0.354712086 0.000000e+00
## 83    0.899372641 0.000000e+00
## 84    0.083903212 1.634528e-04
## 85   -0.076466053 1.338173e-04
## 86    0.301557394 0.000000e+00
## 87   -0.112081253 2.050190e-08
## 88    0.211975566 0.000000e+00
## 89    0.138364222 1.940448e-11
## 90    0.200665557 0.000000e+00
## 91   -0.136490819 7.947643e-12
## 92   -0.021538108 3.035322e-01
## 93   -0.013646972 5.144795e-01
## 94    0.556801332 0.000000e+00
## 95   -0.035252342 1.038404e-01
## 96   -0.025661888 2.200179e-01
## 97   -0.123320952 1.485790e-07
## 98    0.265966087 0.000000e+00
## 99   -0.072301023 5.921647e-04
## 100   0.544422649 0.000000e+00
## 101  -0.038540248 6.655204e-02
## 102  -0.079661838 2.322268e-04
## 103  -0.028443781 1.757934e-01
## 104  -0.027854290 1.830894e-01
## 105  -0.028269671 1.787417e-01
## 106  -0.477183192 0.000000e+00
## 107   0.302903787 0.000000e+00
## 108   0.465710883 0.000000e+00
## 109  -0.428795257 0.000000e+00
## 110  -0.251368631 0.000000e+00
## 111  -0.120429205 4.044164e-09
## 112   0.224808308 0.000000e+00
## 113  -0.532024750 0.000000e+00
## 114   0.467789051 0.000000e+00
## 115  -0.221823439 0.000000e+00
## 116  -0.277100607 0.000000e+00
## 117  -0.229518256 0.000000e+00
## 118   0.204063588 0.000000e+00
## 119  -0.285697168 0.000000e+00
## 120   0.253943752 0.000000e+00
## 121  -0.471583605 0.000000e+00
## 122   0.308457270 0.000000e+00
## 123   0.471350139 0.000000e+00
## 124  -0.417413629 0.000000e+00
## 125  -0.252904587 0.000000e+00
## 126  -0.124959929 1.025443e-09
## 127   0.221072200 0.000000e+00
## 128  -0.538910573 0.000000e+00
## 129   0.472262827 0.000000e+00
## 130  -0.222591763 0.000000e+00
## 131  -0.283773605 0.000000e+00
## 132  -0.222742797 0.000000e+00

```

```

## 133 0.207283246 0.000000e+00
## 134 -0.290539012 0.000000e+00
## 135 0.251402968 0.000000e+00
## 136 0.939101992 0.000000e+00
## 137 0.724775979 0.000000e+00
## 138 -0.457625638 0.000000e+00
## 139 -0.145139259 1.620926e-14
## 140 0.450039706 0.000000e+00
## 141 0.381952444 0.000000e+00
## 142 0.199549238 0.000000e+00
## 143 -0.129568176 7.578160e-12
## 144 0.508773746 0.000000e+00
## 145 -0.163304792 0.000000e+00
## 146 0.381077721 0.000000e+00
## 147 0.166682045 0.000000e+00
## 148 0.401455803 0.000000e+00
## 149 -0.249519497 0.000000e+00
## 150 0.460341479 0.000000e+00
## 151 -0.008734836 6.764449e-01
## 152 -0.422429087 0.000000e+00
## 153 -0.411053256 0.000000e+00
## 154 0.751975463 0.000000e+00
## 155 -0.454611932 0.000000e+00
## 156 -0.193719751 0.000000e+00
## 157 0.547378379 0.000000e+00
## 158 0.389687148 0.000000e+00
## 159 0.231116546 0.000000e+00
## 160 -0.137224530 3.874678e-13
## 161 0.546961042 0.000000e+00
## 162 -0.209373285 0.000000e+00
## 163 0.417866408 0.000000e+00
## 164 0.246384299 0.000000e+00
## 165 0.425332298 0.000000e+00
## 166 -0.220428716 0.000000e+00
## 167 0.448272829 0.000000e+00
## 168 -0.031667638 1.301149e-01
## 169 -0.471651570 0.000000e+00
## 170 -0.460631702 0.000000e+00
## 171 0.800092420 0.000000e+00

```

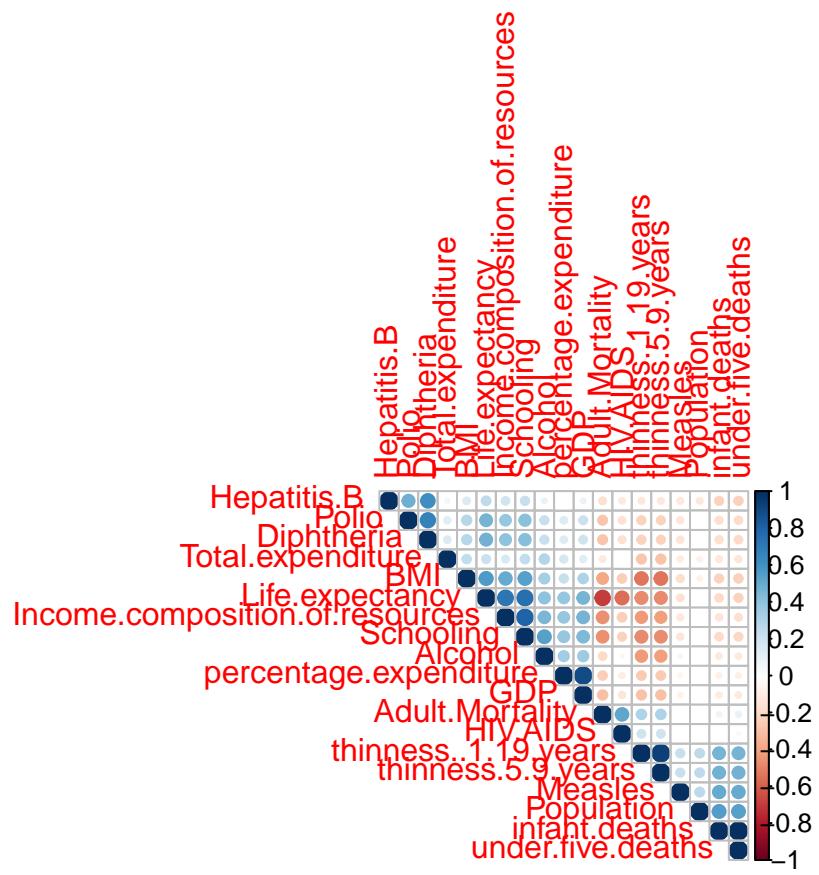
We can see that life expectancy and schooling have a high correlation and a statistically significant p value.

Including Plots

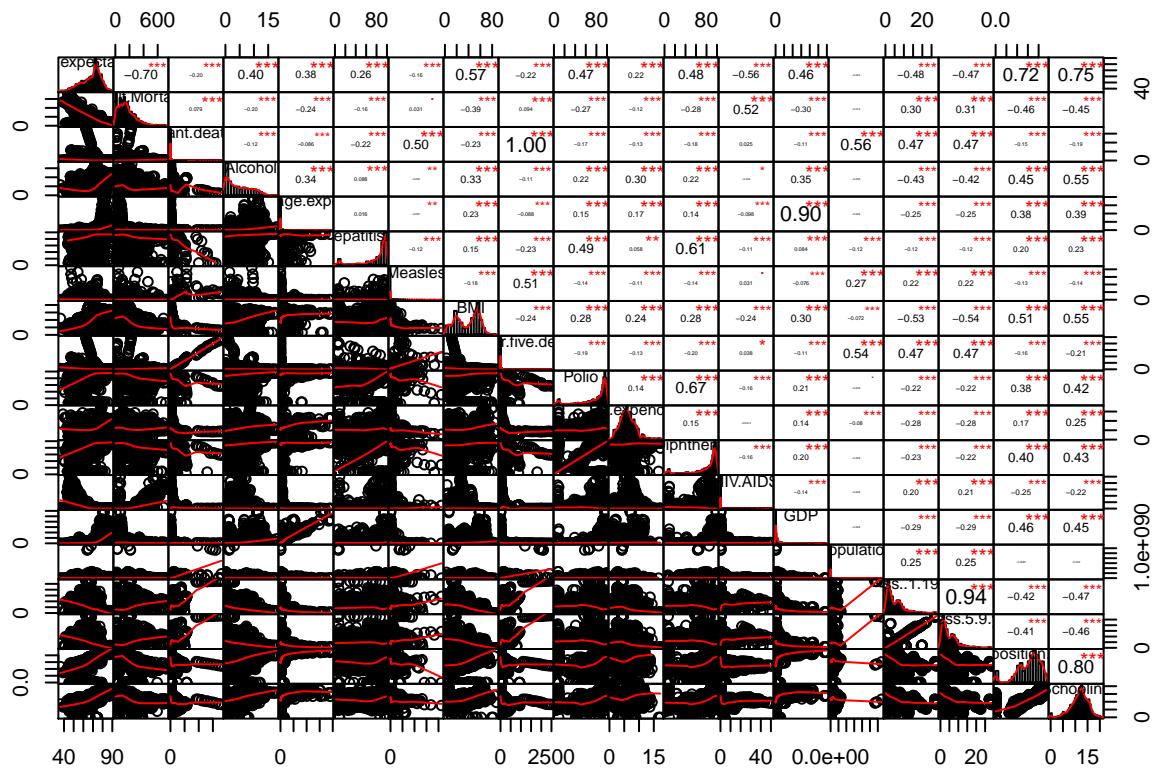
```

corrplot(newCorr$r, type="upper", order="hclust",
         p.mat = newCorr$P, sig.level = 0.05, insig = "blank")

```



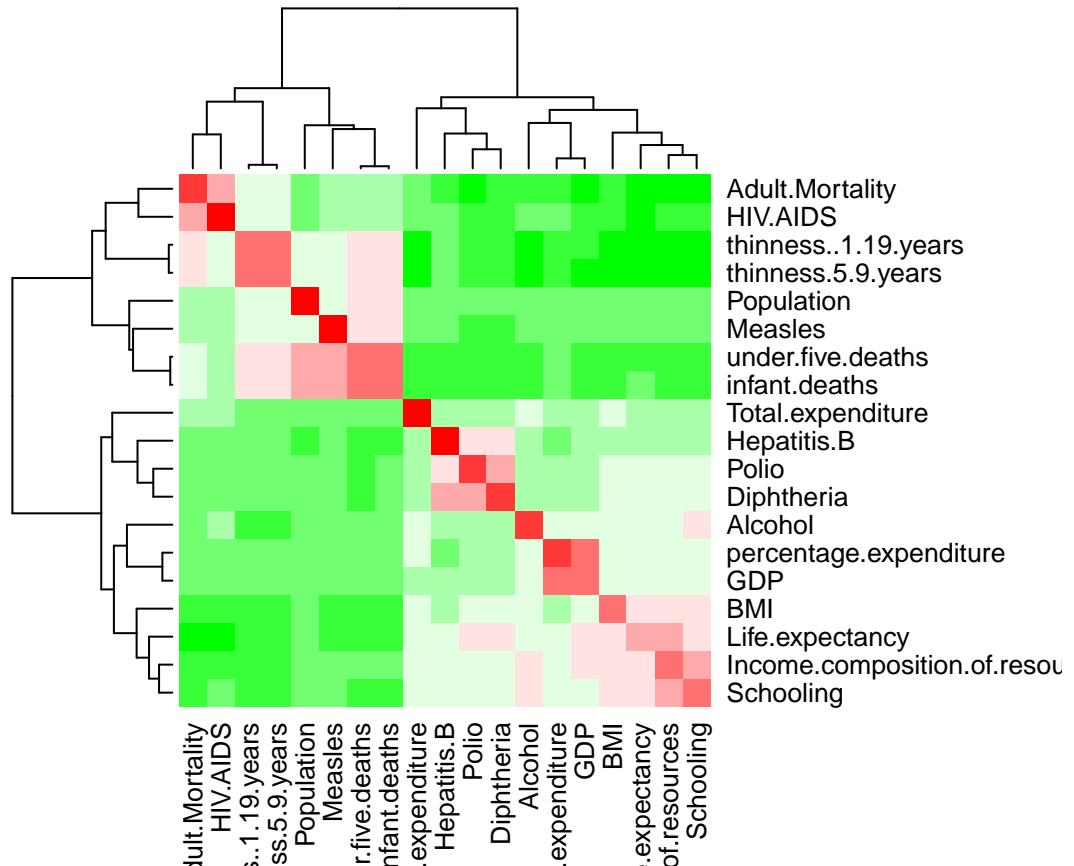
```
chart.Correlation(LE_data_df, histogramme=TRUE, pch=19)
```



```

palette = colorRampPalette(c("green", "white", "red")) (10)
heatmap(x = newCorr$r, scale="column", col = palette, symm = TRUE)

```



Schooling and life expectancy are positively correlated. Let's create a linear regression model and later evaluate it.

```

lif <- LE_data_df$Life.expectancy
sch <- LE_data_df$Schooling

lif_sch_lm <- lm(sch ~ lif)

summary(lif_sch_lm)

##
## Call:
## lm(formula = sch ~ lif)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -13.4111  -1.1027   0.0164   1.2696   6.1774 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -6.643467  0.313559 -21.19   <2e-16 ***
## lif          0.268828  0.004481  59.99   <2e-16 ***
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
```

```

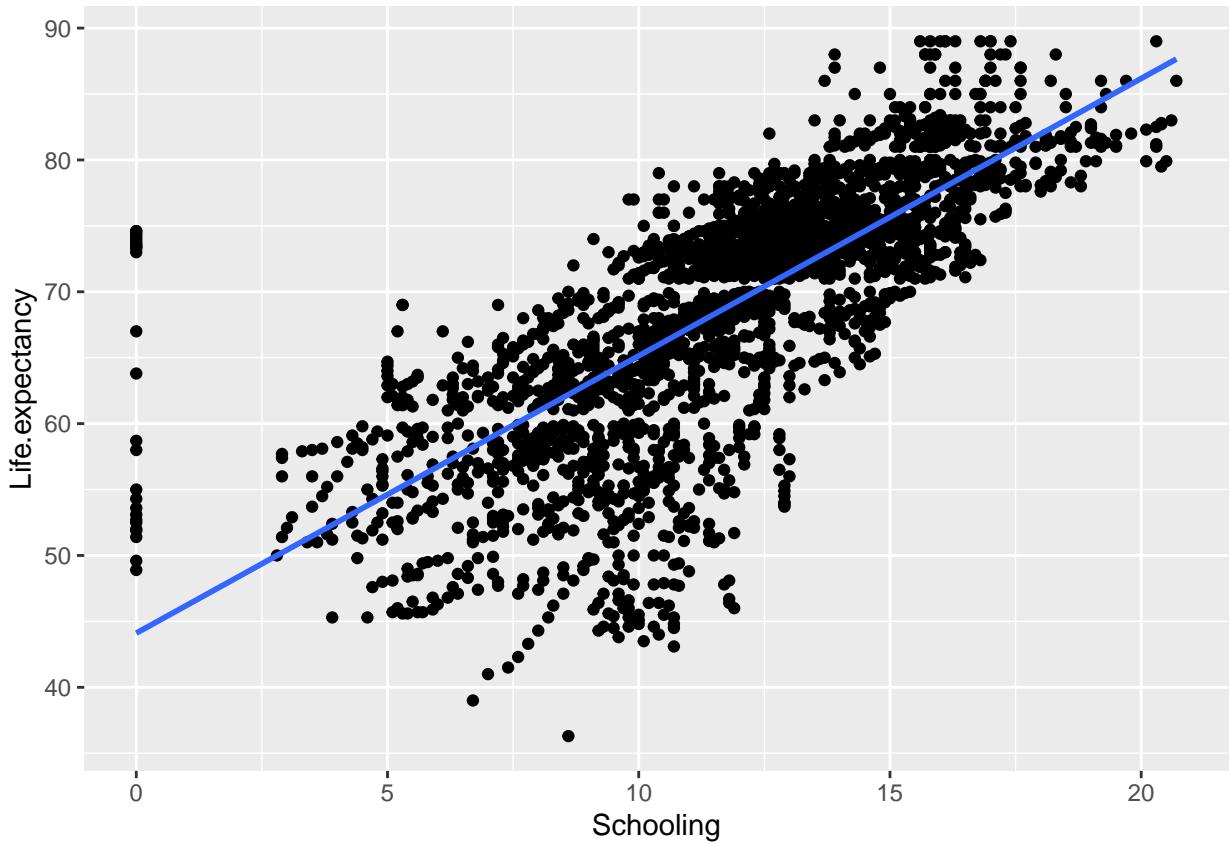
## Residual standard error: 2.206 on 2766 degrees of freedom
##   (170 observations deleted due to missingness)
## Multiple R-squared:  0.5655, Adjusted R-squared:  0.5653
## F-statistic:  3599 on 1 and 2766 DF,  p-value: < 2.2e-16

ggplot(data = LE_data_df, aes(x = Schooling, y = Life.expectancy)) +
  geom_point() +
  stat_smooth(method = "lm", se = FALSE)

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

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

```



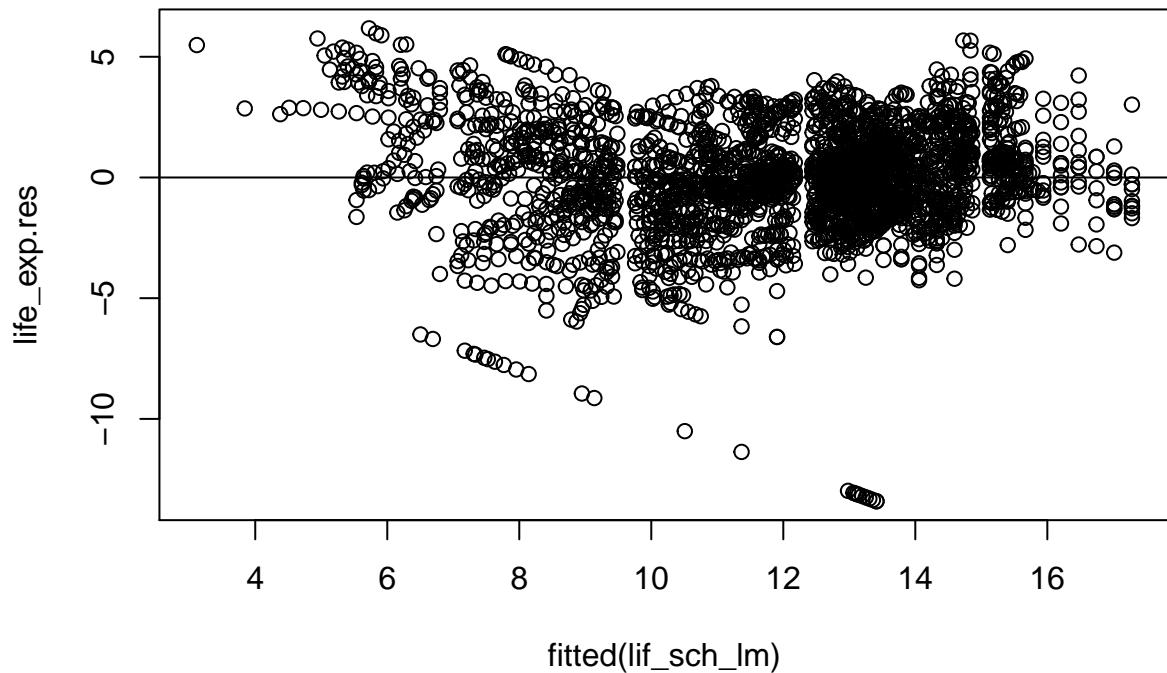
Examining the Residuals

```

life_exp.res = resid(lif_sch_lm)

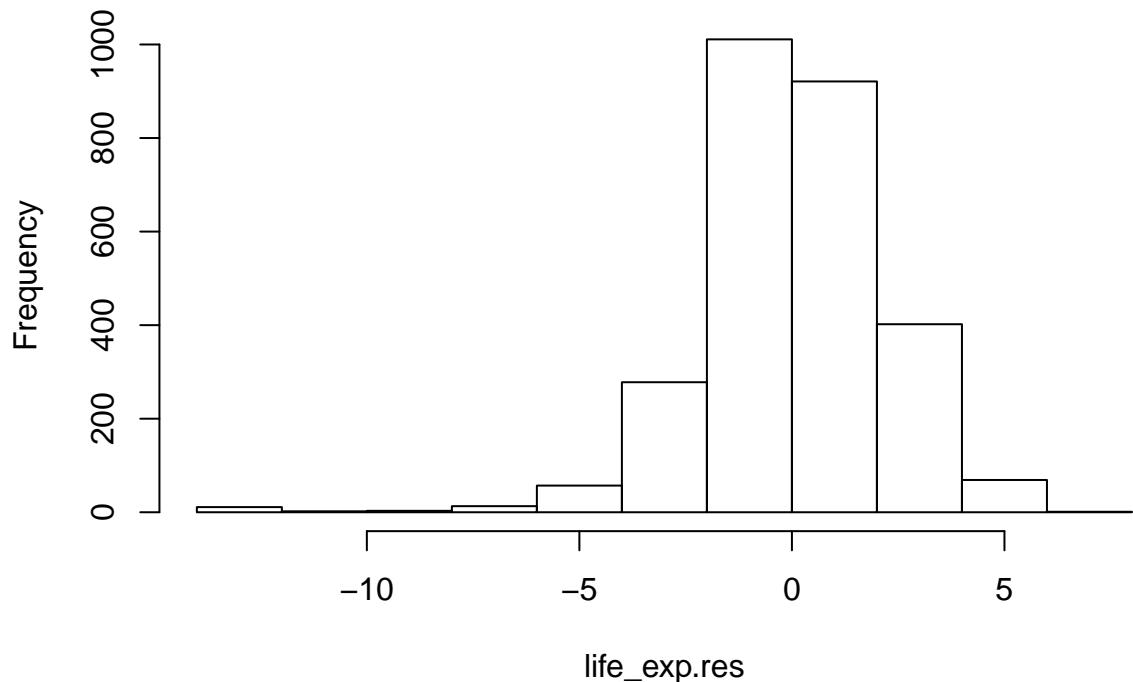
plot(fitted(lif_sch_lm), life_exp.res)
abline(0,0)

```



```
hist(life_exp.res)
```

Histogram of life_exp.res



```
qqnorm(resid(lif_sch_lm))
qqline(resid(lif_sch_lm))
```

Normal Q-Q Plot

