DSC520 Final Project: Data Analysis of Life expectancy

Rohan Valder

August 3rd 2020

```
Life expectancy is a critical metric for assessing population health. It has increased rapidly sin

The population of many of the wealthiest countries in the world has life expectancies of over 80 y

The global inequalities in health that we see today also show that we can do much better. The almo
```

Data Cleaning

Below are the columns renamed:

Original Columns	New Columns
Average.Life.Expectancy..Years.	Avg_Life_Expectancy
Age.adjusted.Death.Rate	Age_Adjusted_Death_Rate
ï..Country.Name	Country_Name
Country.Code	Country_Code

Calculated the Average for multiple columns and created a new column. Dropped a few columns that were not required.

Life expectancy around the world

Summary

```
##  Country_Code       Country_Name            Avg           Region
##  Length:186         Length:186         Min.   :40.73   Length:186
##  Class :character   Class :character   1st Qu.:55.67   Class :character
##  Mode  :character   Mode  :character   Median :66.08   Mode  :character
##                                        Mean   :63.84
##                                        3rd Qu.:71.50
##                                        Max.   :78.01
##  IncomeGroup
##  Length:186
##  Class :character
##  Mode  :character
##
##
##
```

Data trend based on Income Groups

```
##
##                       40.7336101694915 43.617593220339 44.1101186440678
##   High income                        0               0                0
##   Low income                         1               1                1
##   Lower middle income                0               0                0
##   Upper middle income                0               0                0
##
##                       45.208813559322 45.6175423728814 45.9104237288136
##   High income                       0                0                0
##   Low income                        1                0                1
##   Lower middle income               0                1                0
##   Upper middle income               0                0                0
##
##                       46.0623050847458 46.3663728813559 46.5141694915254
##   High income                        0                 0                0
##   Low income                         1                 1                0
##   Lower middle income                0                 0                1
##   Upper middle income                0                 0                0
##
##                       46.5689491525424 46.8789491525424 46.9236440677966
##   High income                        0                 0                0
##   Low income                         1                 1                1
##   Lower middle income                0                 0                0
##   Upper middle income                0                 0                0
##
##                       47.1051016949153 47.876 47.9476101694915 48.1660847457627
##   High income                        0      0                0                0
##   Low income                         1      0                1                1
##   Lower middle income                0      0                0                0
##   Upper middle income                0      1                0                0
##
##                       48.5593898305085 48.9321525423729 49.0218305084746
##   High income                        0                0                0
##   Low income                         1                1                1
##   Lower middle income                0                0                0
##   Upper middle income                0                0                0
##
##                       49.1848305084746 49.39 49.4538474576271 49.4656949152542
##   High income                        0     0                0                0
##   Low income                         1     1                0                0
##   Lower middle income                0     0                1                1
##   Upper middle income                0     0                0                0
##
##                       49.5352881355932 49.9655254237288 50.0692881355932
##   High income                        0                0                0
##   Low income                         1                0                1
##   Lower middle income                0                1                0
##   Upper middle income                0                0                0
##
##                       50.8757966101695 50.8862372881356 51.1113559322034
##   High income                        0                0                0
##   Low income                         0                0                0
```

```
##   Lower middle income                1                 1                 1
##   Upper middle income                0                 0                 0
##
##                       51.1332881355932 51.7271694915254 51.7655593220339
##   High income                        0                 0                 0
##   Low income                         1                 0                 0
##   Lower middle income                0                 1                 1
##   Upper middle income                0                 0                 0
##
##                       52.3067627118644 52.4883389830508 52.7990338983051
##   High income                        0                 0                 0
##   Low income                         1                 1                 0
##   Lower middle income                0                 0                 1
##   Upper middle income                0                 0                 0
##
##                       52.9374406779661 52.9439661016949 53.3136779661017
##   High income                        0                 0                 0
##   Low income                         0                 1                 1
##   Lower middle income                1                 0                 0
##   Upper middle income                0                 0                 0
##
##                       53.6372711864407 54.1521355932203 54.2070169491525
##   High income                        0                 0                 0
##   Low income                         0                 0                 0
##   Lower middle income                1                 1                 1
##   Upper middle income                0                 0                 0
##
##                       54.2433389830508 54.3773220338983 54.4156949152542
##   High income                        0                 0                 0
##   Low income                         0                 0                 0
##   Lower middle income                1                 1                 1
##   Upper middle income                0                 0                 0
##
##                       54.975406779661 55.0251355932203 55.6413220338983
##   High income                       0                0                 0
##   Low income                        0                0                 0
##   Lower middle income               1                1                 0
##   Upper middle income               0                0                 1
##
##                       55.7409322033898 55.8148305084746 55.8159661016949
##   High income                        0                 0                 0
##   Low income                         0                 0                 0
##   Lower middle income                0                 1                 1
##   Upper middle income                1                 0                 0
##
##                       56.0489322033898 56.4149661016949 56.7625084745763
##   High income                        0                 0                 0
##   Low income                         0                 1                 0
##   Lower middle income                1                 0                 0
##   Upper middle income                0                 0                 1
##
##                       56.9352203389831 56.9814745762712 57.1260338983051
##   High income                        0                 0                 0
##   Low income                         0                 0                 0
```

```
##   Lower middle income                1                1                0
##   Upper middle income                0                0                1
##
##                       58.3235254237288 58.7894915254237 58.8077966101695
##   High income                        0                0                0
##   Low income                         0                0                0
##   Lower middle income                1                1                1
##   Upper middle income                0                0                0
##
##                       59.8879152542373 59.9267288135593 60.1074237288136
##   High income                        0                0                0
##   Low income                         0                1                0
##   Lower middle income                0                0                1
##   Upper middle income                1                0                0
##
##                       60.1817288135593 60.8587966101695 61.5923050847458
##   High income                        0                0                0
##   Low income                         0                0                0
##   Lower middle income                1                0                0
##   Upper middle income                0                1                1
##
##                       61.8388305084746 62.0376610169492 62.1482881355932
##   High income                        0                0                0
##   Low income                         0                0                0
##   Lower middle income                0                1                0
##   Upper middle income                1                0                1
##
##                       62.3024406779661 62.5627457627119 62.6912033898305
##   High income                        0                0                0
##   Low income                         0                0                0
##   Lower middle income                1                1                1
##   Upper middle income                0                0                0
##
##                       62.7920677966102 62.9257966101695 62.998406779661
##   High income                        0                0               0
##   Low income                         0                0               0
##   Lower middle income                1                1               1
##   Upper middle income                0                0               0
##
##                       63.1178983050847 63.337186440678 63.3553898305085
##   High income                        0               0                0
##   Low income                         0               0                0
##   Lower middle income                0               1                0
##   Upper middle income                1               0                1
##
##                       63.3564237288136 63.5925762711864 63.7957457627119
##   High income                        0                0                1
##   Low income                         0                0                0
##   Lower middle income                1                1                0
##   Upper middle income                0                0                0
##
##                       64.1699491525424 64.3806101694915 64.4952372881356
##   High income                        0                0                0
##   Low income                         0                0                0
```

```
##   Lower middle income                0                 0                 1
##   Upper middle income                1                 1                 0
##
##                       64.6622881355932 64.7421694915254 64.9621831330508
##   High income                        0                 0                 0
##   Low income                         0                 0                 0
##   Lower middle income                0                 0                 1
##   Upper middle income                1                 1                 0
##
##                       65.0588305084746 65.0935762711864 65.6481016949153
##   High income                        0                 0                 0
##   Low income                         0                 1                 0
##   Lower middle income                0                 0                 0
##   Upper middle income                1                 0                 1
##
##                       65.7381860276271 65.778220338983 65.8122033898305 65.95
##   High income                        0                0                 0     0
##   Low income                         0                0                 0     0
##   Lower middle income                0                1                 0     0
##   Upper middle income                1                0                 1     1
##
##                       66.2154406779661 66.3786101694915 66.5019661016949
##   High income                        0                 0                 0
##   Low income                         0                 0                 0
##   Lower middle income                1                 0                 0
##   Upper middle income                0                 1                 1
##
##                       66.7406101694915 66.8474745762712 66.9428983050847
##   High income                        0                 0                 0
##   Low income                         0                 1                 0
##   Lower middle income                1                 0                 0
##   Upper middle income                0                 0                 1
##
##                       67.1827966101695 67.2955593220339 67.3848983050847
##   High income                        0                 0                 0
##   Low income                         0                 0                 0
##   Lower middle income                0                 0                 0
##   Upper middle income                1                 1                 1
##
##                       67.7570508474576 67.7839057466102 67.9617966101695
##   High income                        1                 0                 1
##   Low income                         0                 0                 0
##   Lower middle income                0                 0                 0
##   Upper middle income                0                 1                 0
##
##                       68.2934237288136 68.3653799088136 68.6263898305085
##   High income                        0                 1                 0
##   Low income                         0                 0                 0
##   Lower middle income                0                 0                 1
##   Upper middle income                1                 0                 0
##
##                       68.6941186440678 68.7202203389831 68.8366779661017
##   High income                        1                 0                 0
##   Low income                         0                 0                 0
```

```
##   Lower middle income                0                 0                 0
##   Upper middle income                0                 1                 1
##
##                       68.8704745762712 68.9467457627119 69.1422033898305
##   High income                        0                 0                 0
##   Low income                         0                 0                 0
##   Lower middle income                0                 0                 0
##   Upper middle income                1                 1                 1
##
##                       69.2134576271186 69.2322711864407 69.2408644067797
##   High income                        0                 1                 0
##   Low income                         0                 0                 0
##   Lower middle income                1                 0                 0
##   Upper middle income                0                 0                 1
##
##                       69.3566949152542 69.3593985122034 69.4236949152542
##   High income                        1                 0                 0
##   Low income                         0                 0                 0
##   Lower middle income                0                 1                 0
##   Upper middle income                0                 0                 1
##
##                       69.4944406779661 69.6111525423729 69.6339423481356
##   High income                        0                 0                 1
##   Low income                         0                 0                 0
##   Lower middle income                0                 0                 0
##   Upper middle income                1                 1                 0
##
##                       69.9303898305085 70.0948495244068 70.2293220338983
##   High income                        1                 0                 0
##   Low income                         0                 0                 0
##   Lower middle income                0                 0                 0
##   Upper middle income                0                 1                 1
##
##                       70.2819156672881 70.289779661017 70.3760711038983
##   High income                        1                1                1
##   Low income                         0                0                0
##   Lower middle income                0                0                0
##   Upper middle income                0                0                0
##
##                       70.4440719301695 70.6048305084746 70.6924406779661
##   High income                        1                 0                 1
##   Low income                         0                 0                 0
##   Lower middle income                0                 0                 0
##   Upper middle income                0                 1                 0
##
##                       70.8783670935593 70.9623898305085 70.970165771017
##   High income                        1                 0                1
##   Low income                         0                 0                0
##   Lower middle income                0                 0                0
##   Upper middle income                0                 1                0
##
##                       71.1402203389831 71.1514576271186 71.4675609754237
##   High income                        1                 0                 1
##   Low income                         0                 0                 0
```

```
##   Lower middle income                0                0                0
##   Upper middle income                0                1                0
##
##                       71.478593220339 71.5070677966102 71.5284237288136
##   High income                       1                1                0
##   Low income                        0                0                0
##   Lower middle income               0                0                0
##   Upper middle income               0                0                1
##
##                       71.5634406779661 71.7858606859322 71.82352625
##   High income                        0                1           0
##   Low income                         0                0           0
##   Lower middle income                0                0           0
##   Upper middle income                1                0           1
##
##                       72.2716411737288 72.2728714347458 72.3406610169492
##   High income                        1                1                1
##   Low income                         0                0                0
##   Lower middle income                0                0                0
##   Upper middle income                0                0                0
##
##                       72.4812542372881 72.647813559322 73.0568292686441
##   High income                        1                0                1
##   Low income                         0                0                0
##   Lower middle income                0                0                0
##   Upper middle income                0                1                0
##
##                       73.1558644067797 73.1871947089831 73.5624018191525
##   High income                        0                1                1
##   Low income                         0                0                0
##   Lower middle income                0                0                0
##   Upper middle income                1                0                0
##
##                       73.6391777594915 73.7044745762712 73.8415254237288
##   High income                        1                1                1
##   Low income                         0                0                0
##   Lower middle income                0                0                0
##   Upper middle income                0                0                0
##
##                       73.9698305084746 74.6154596937288 74.6926316650847
##   High income                        0                1                1
##   Low income                         0                0                0
##   Lower middle income                0                0                0
##   Upper middle income                1                0                0
##
##                       74.7373708144068 74.9915663494915 75.0575403064407
##   High income                        1                1                1
##   Low income                         0                0                0
##   Lower middle income                0                0                0
##   Upper middle income                0                0                0
##
##                       75.0917784205085 75.1129350977966 75.1972302608475
##   High income                        1                1                1
##   Low income                         0                0                0
```

```
##   Lower middle income                0                 0                 0
##   Upper middle income                0                 0                 0
##
##                       75.5081934681356 75.638705250339 75.6664489454237
##   High income                        1                 1                 1
##   Low income                         0                 0                 0
##   Lower middle income                0                 0                 0
##   Upper middle income                0                 0                 0
##
##                       75.7520628350847 75.7975312110169 75.8039065720339
##   High income                        1                 1                 1
##   Low income                         0                 0                 0
##   Lower middle income                0                 0                 0
##   Upper middle income                0                 0                 0
##
##                       75.9679322033898 76.1323389830508 76.1422033898305
##   High income                        1                 1                 1
##   Low income                         0                 0                 0
##   Lower middle income                0                 0                 0
##   Upper middle income                0                 0                 0
##
##                       76.3660190162712 76.502769739661 76.5211657715254
##   High income                        1                 1                 1
##   Low income                         0                 0                 0
##   Lower middle income                0                 0                 0
##   Upper middle income                0                 0                 0
##
##                       76.6918396035593 76.8266391069492 76.9696188510169
##   High income                        1                 1                 1
##   Low income                         0                 0                 0
##   Lower middle income                0                 0                 0
##   Upper middle income                0                 0                 0
##
##                       76.9784291033898 77.2367589913559 77.4674121538983
##   High income                        1                 1                 1
##   Low income                         0                 0                 0
##   Lower middle income                0                 0                 0
##   Upper middle income                0                 0                 0
##
##                       77.6784497722034 77.7913972715254 78.0062133111864
##   High income                        1                 1                 1
##   Low income                         0                 0                 0
##   Lower middle income                0                 0                 0
##   Upper middle income                0                 0                 0
```

Plot based on Income Groups

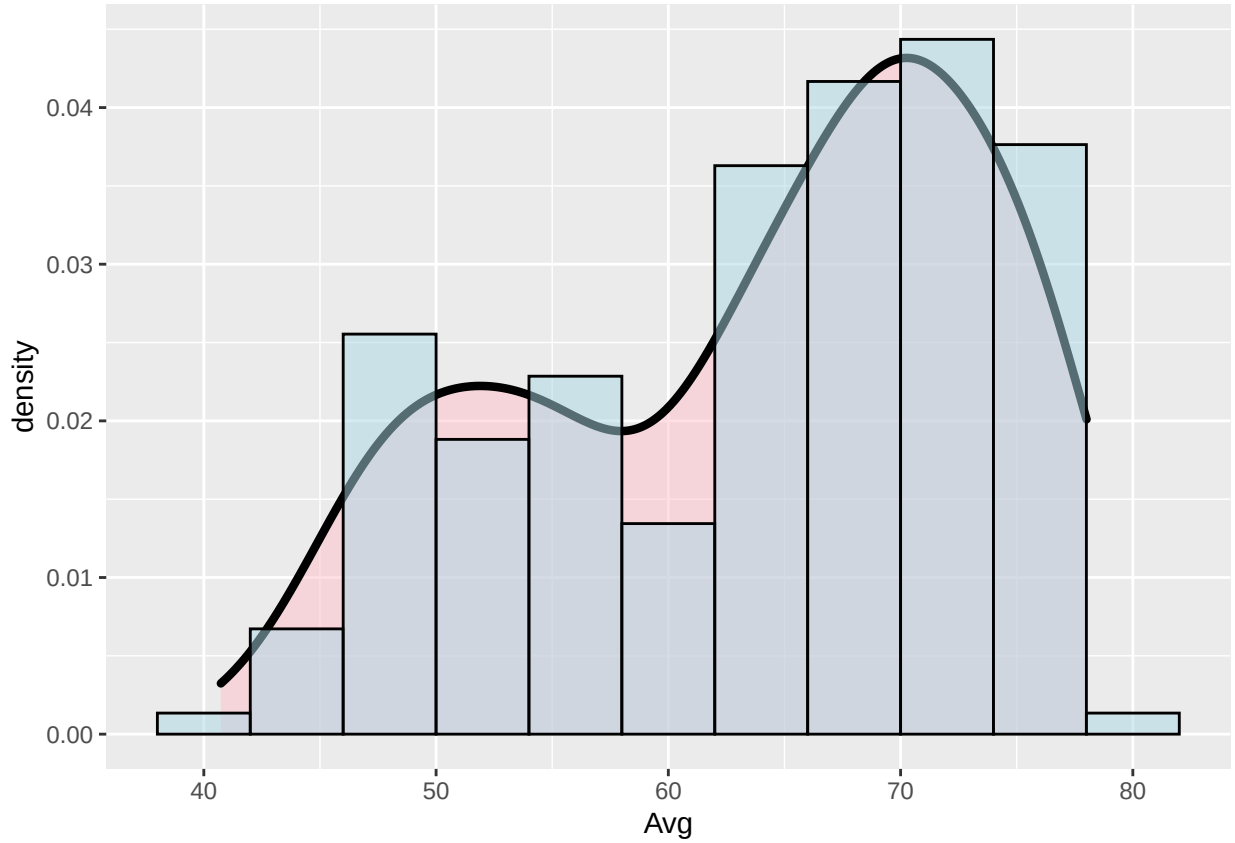
Density plot for average age

```
## Warning: package 'dplyr' was built under R version 4.0.2
```

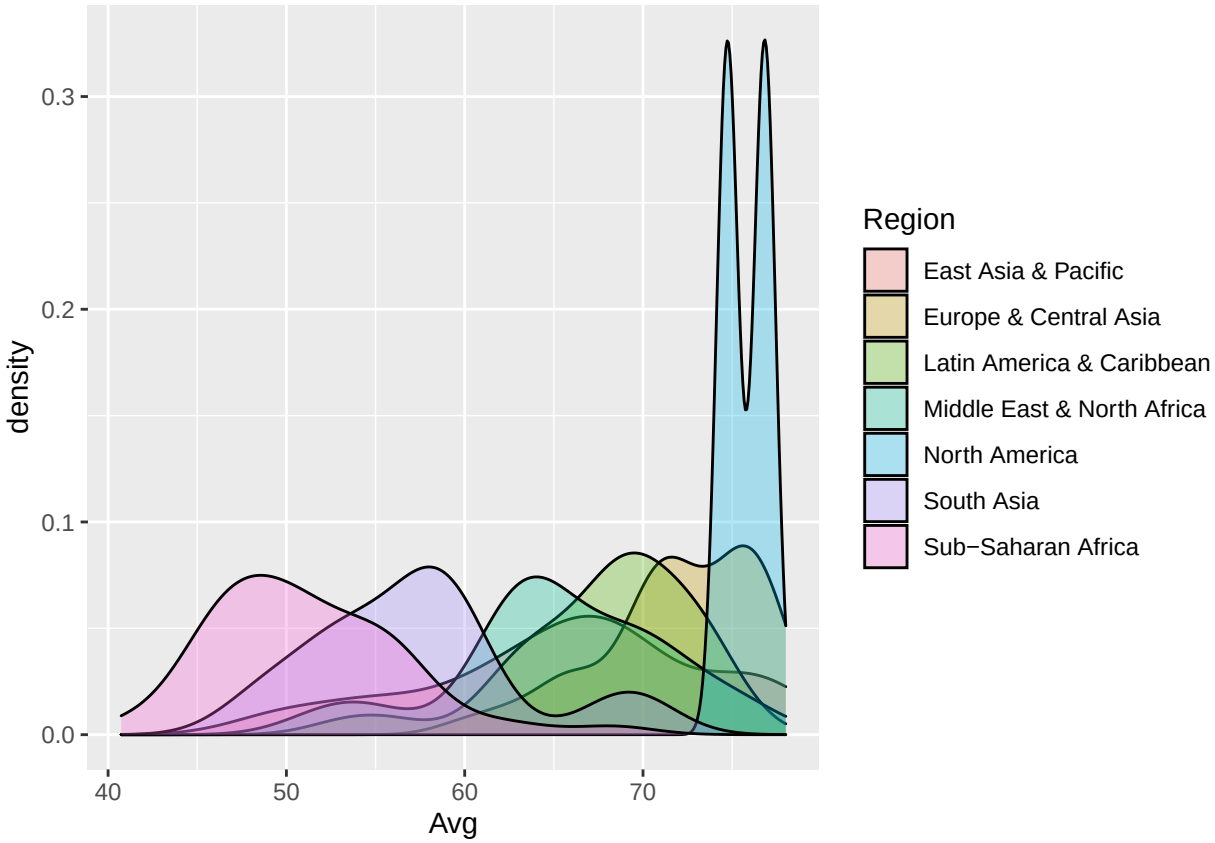
```
##
## Attaching package: 'dplyr'
```

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

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

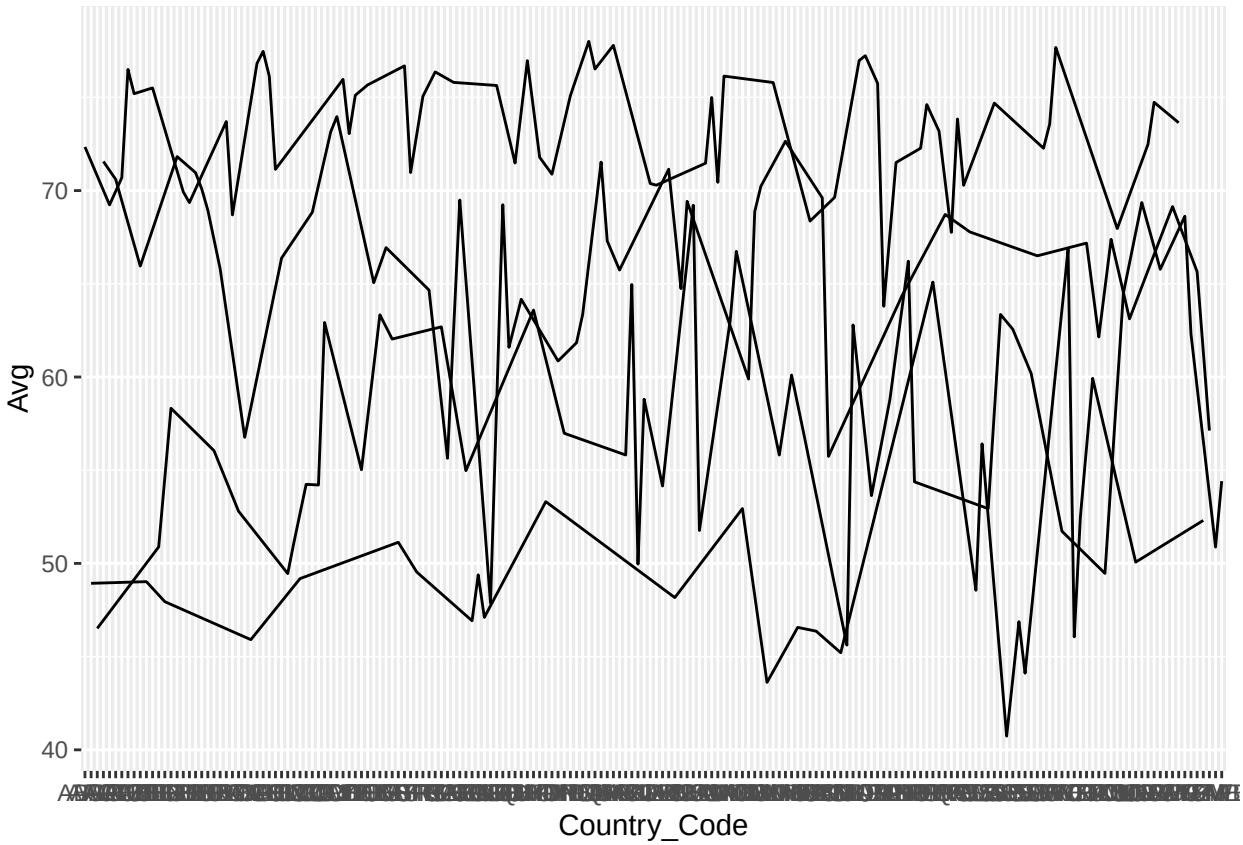

density
0.04
0.03
0.02
0.01
0.00
40
50
60
70
80
Avg

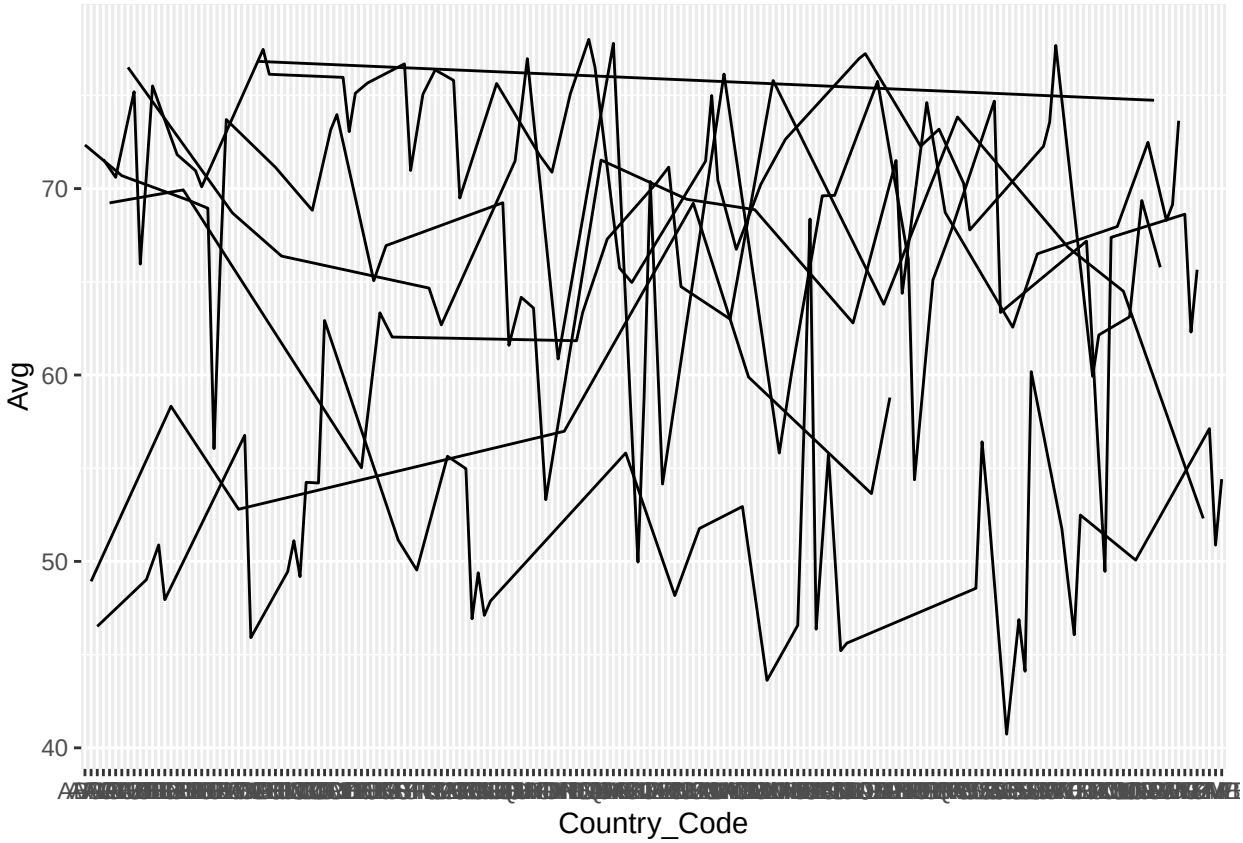
Average Age distribution across regions



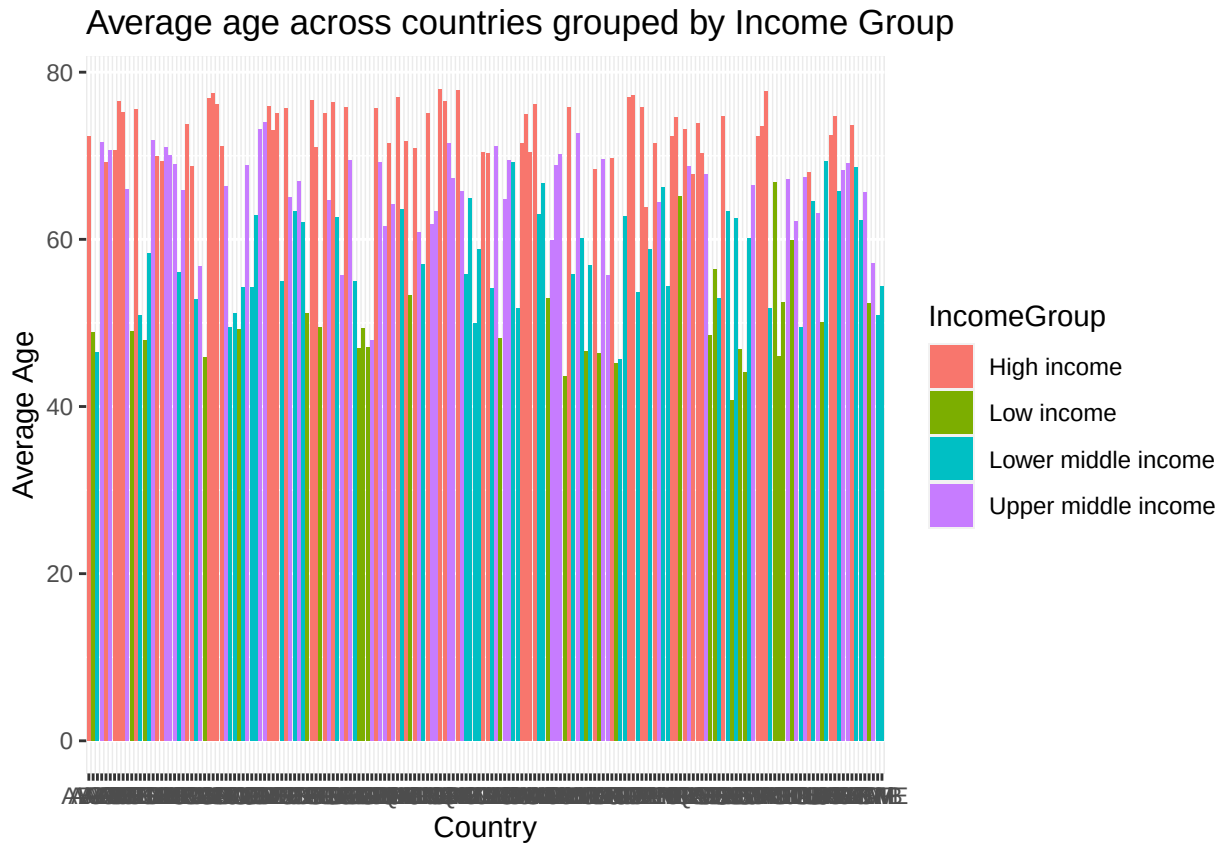
Country Codes Vs Average Age grouped by Income Groups and Regions

By the plots below you see Average Age depends more on the Income Group (1st plot) rather than the regions.


70
60
50
40
Avg
Country_Code


70
60
50
40
Avg
Country_Code

Country Vs Average Age



Linear model

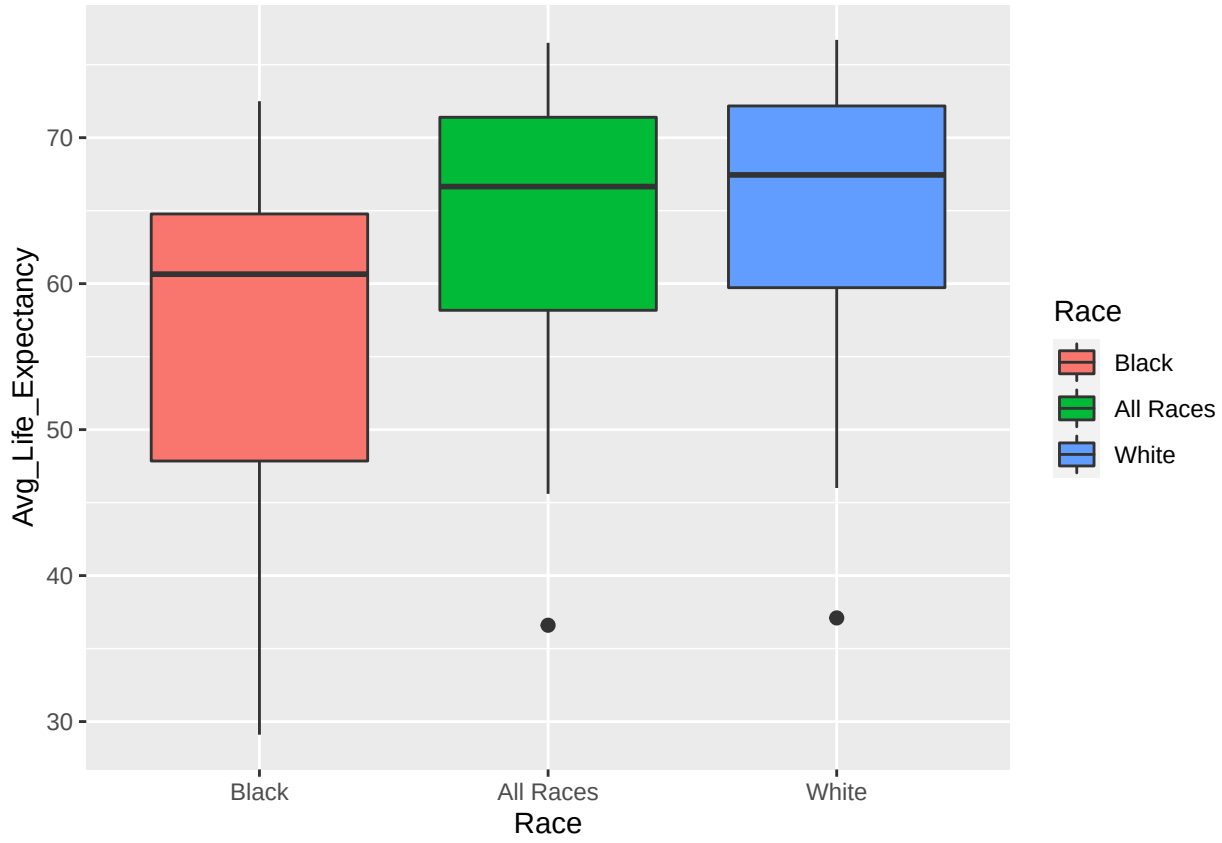
```
##
## Call:
## lm(formula = Avg ~ Region + IncomeGroup, data = project_countries_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.6157  -2.1516   0.3285   1.9358  11.9540
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       71.63232    0.83747  85.534  < 2e-16 ***
## RegionEurope & Central Asia        3.13446    0.92463   3.390 0.000863 ***
## RegionLatin America & Caribbean    1.15532    0.99523   1.161 0.247273
## RegionMiddle East & North Africa   0.01997    1.14442   0.017 0.986097
## RegionNorth America                4.14969    2.84768   1.457 0.146839
## RegionSouth Asia                  -3.82196    1.56775  -2.438 0.015769 *
## RegionSub-Saharan Africa          -8.60510    1.02252  -8.416 1.32e-14 ***
## IncomeGroupLow income            -14.79885    1.13589 -13.028  < 2e-16 ***
## IncomeGroupLower middle income   -10.55094    0.88396 -11.936  < 2e-16 ***
## IncomeGroupUpper middle income    -5.81406    0.77857  -7.468 3.64e-12 ***
## ---
```

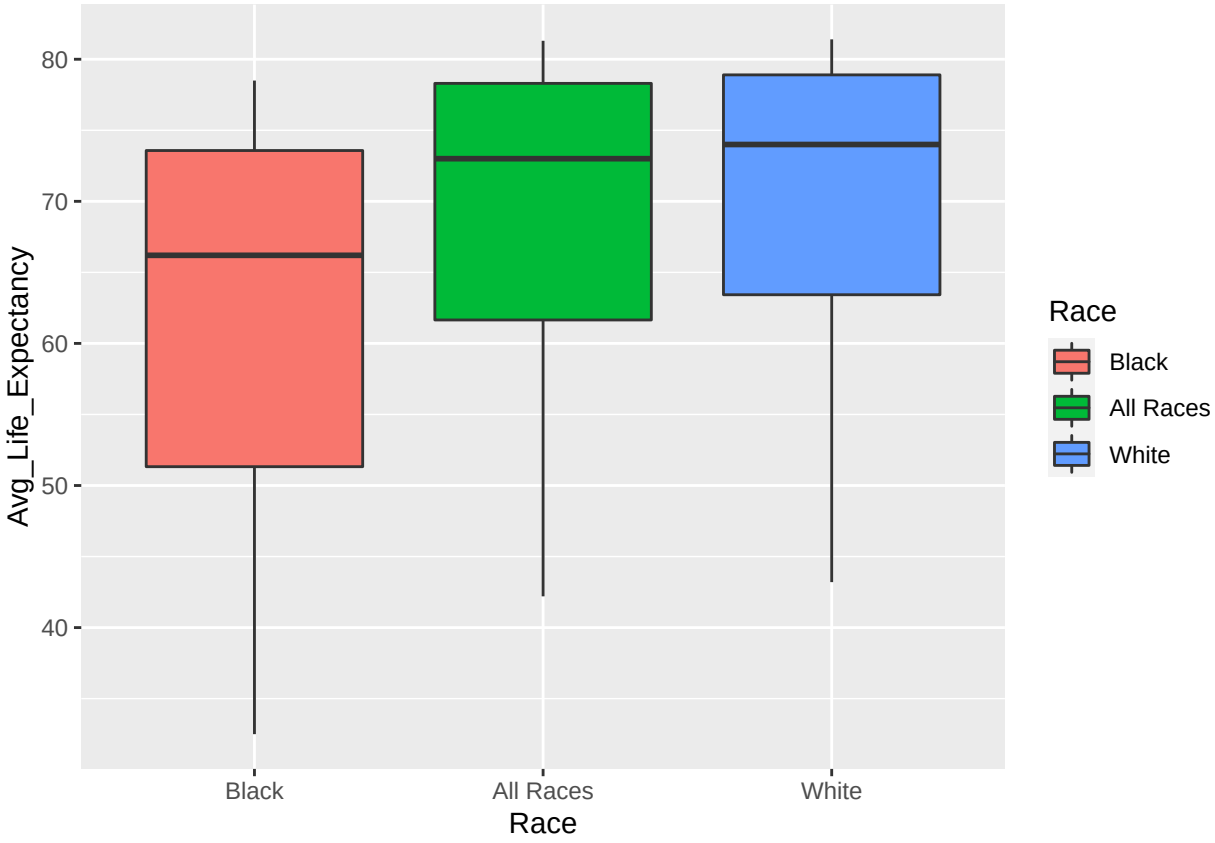
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.849 on 176 degrees of freedom
## Multiple R-squared:  0.8519, Adjusted R-squared:  0.8444
## F-statistic: 112.5 on 9 and 176 DF,  p-value: < 2.2e-16
```

Life expectancy in US

Male and Female Life expectancy based on races in US

By the plots, you see the female (2nd box plot) has a higher life expectancy compared to the male.

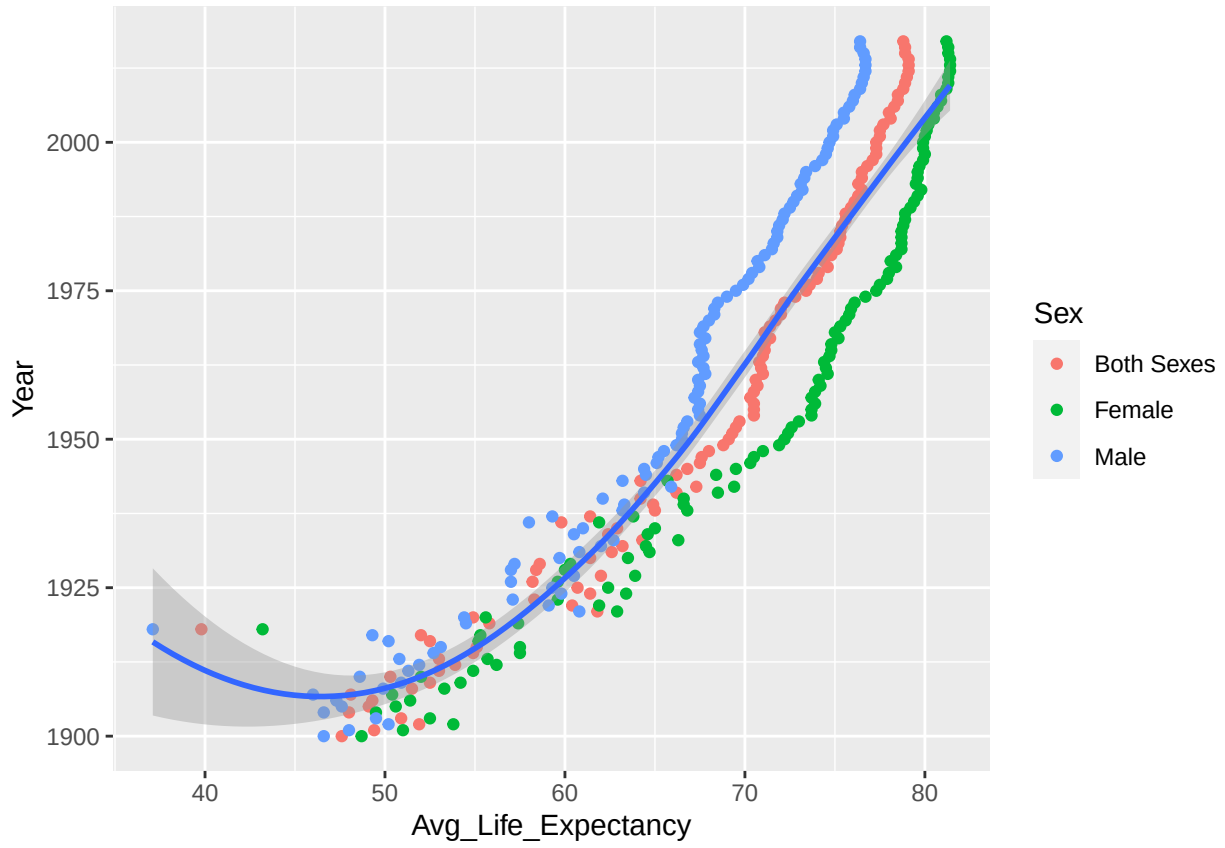




Male and Female Life expectancy since 1900 in US

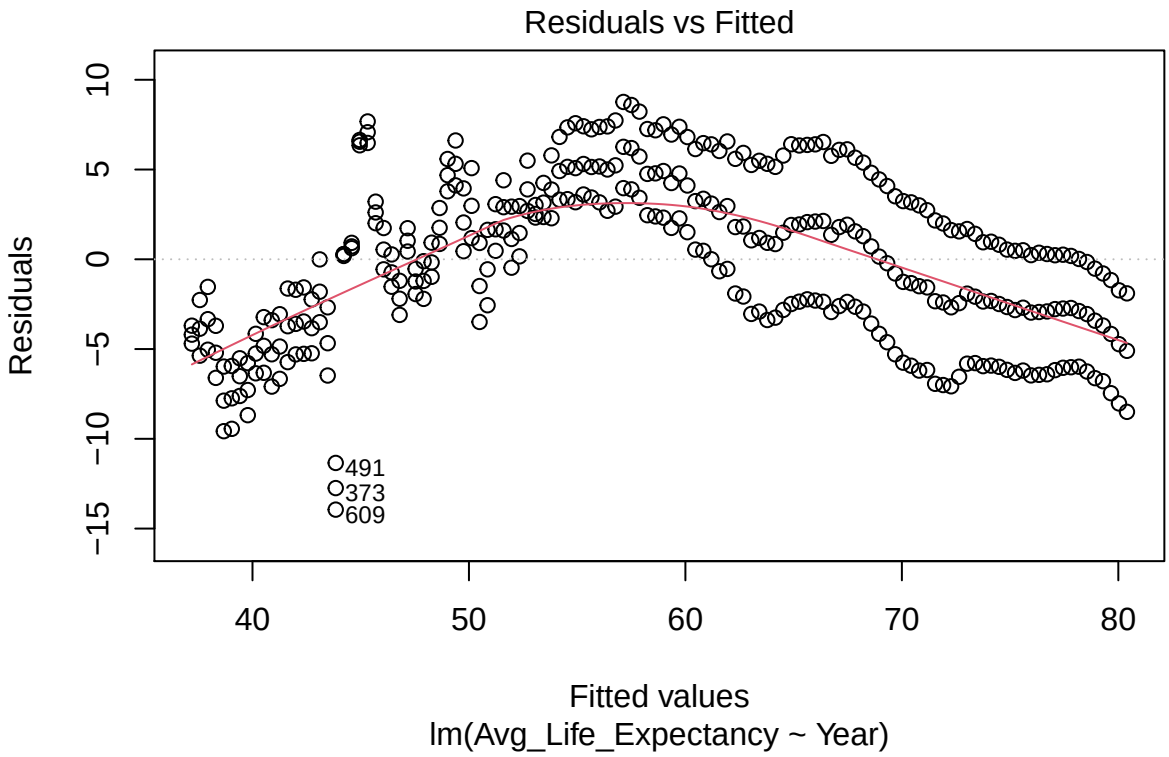
As per the pattern, it is consistent in recent years the male life is shorter compared to the female.

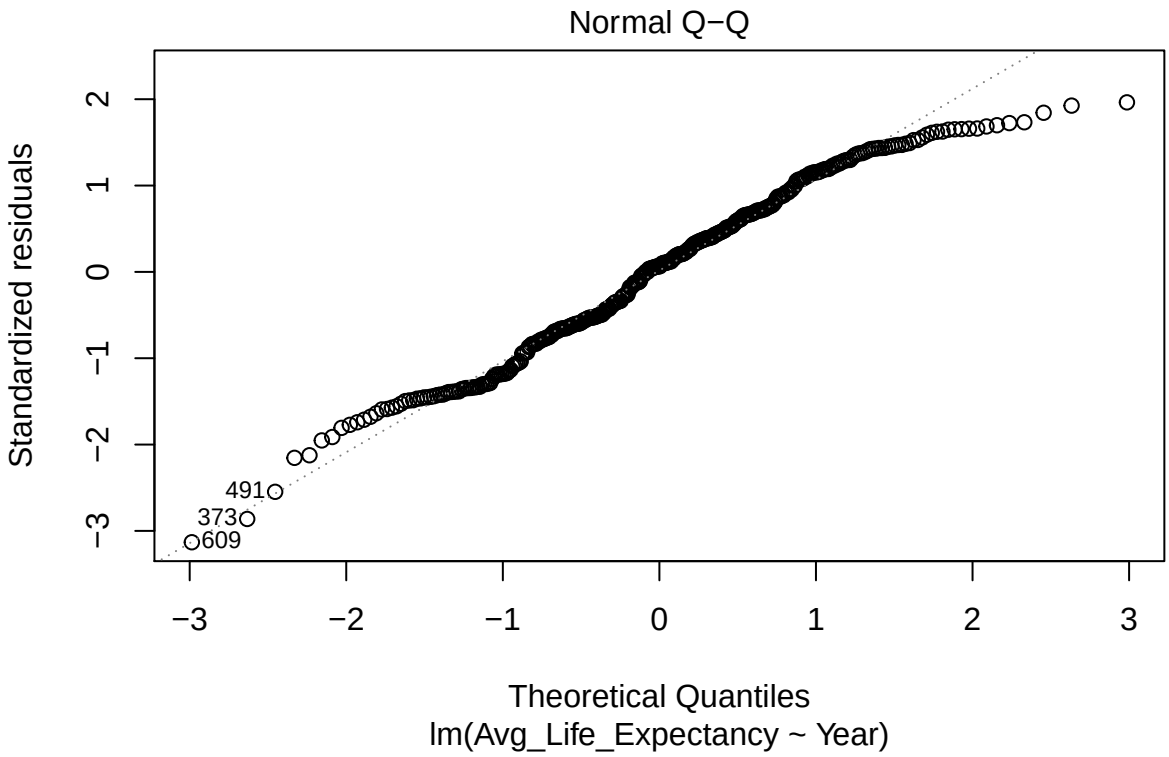
```
## `geom_smooth()` using formula 'y ~ x'
```

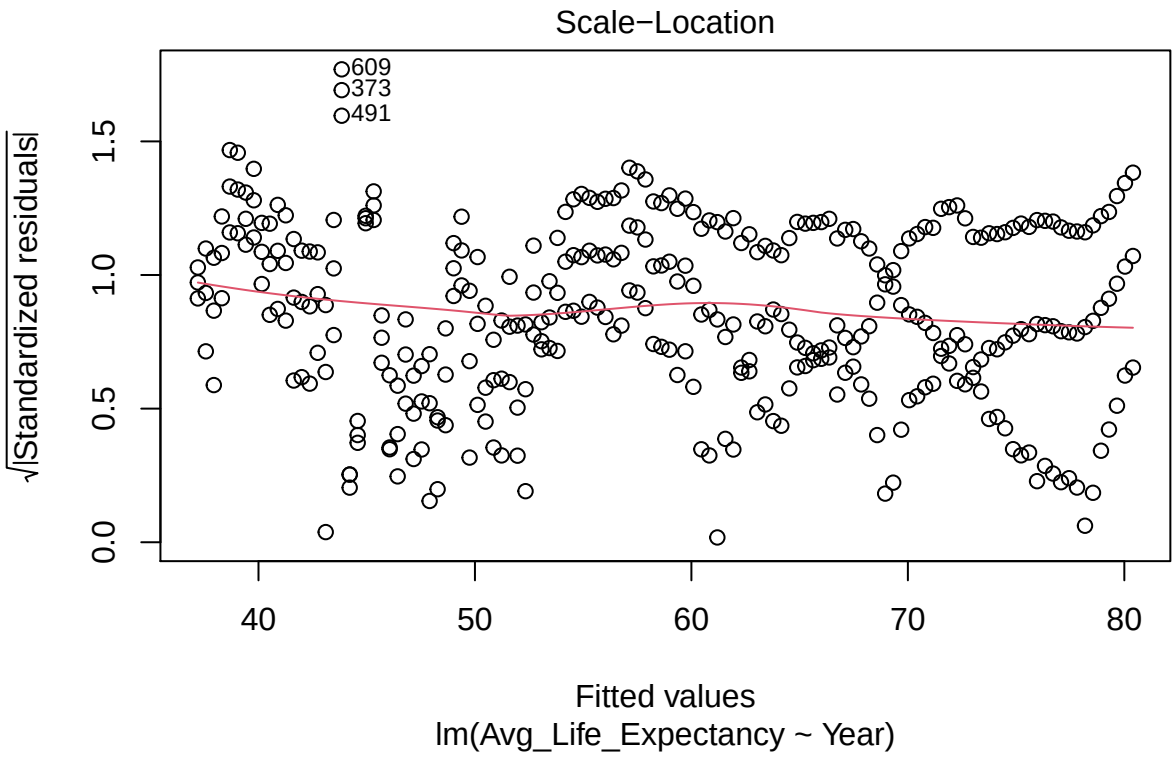


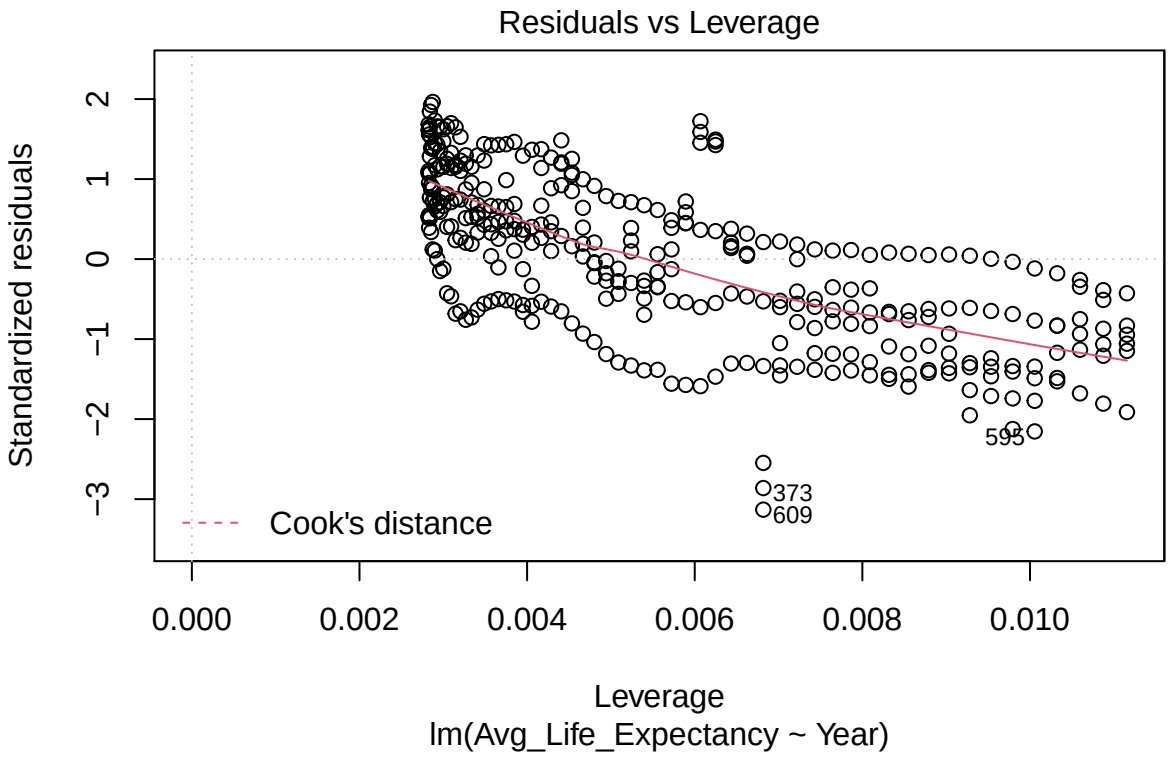
Linear model

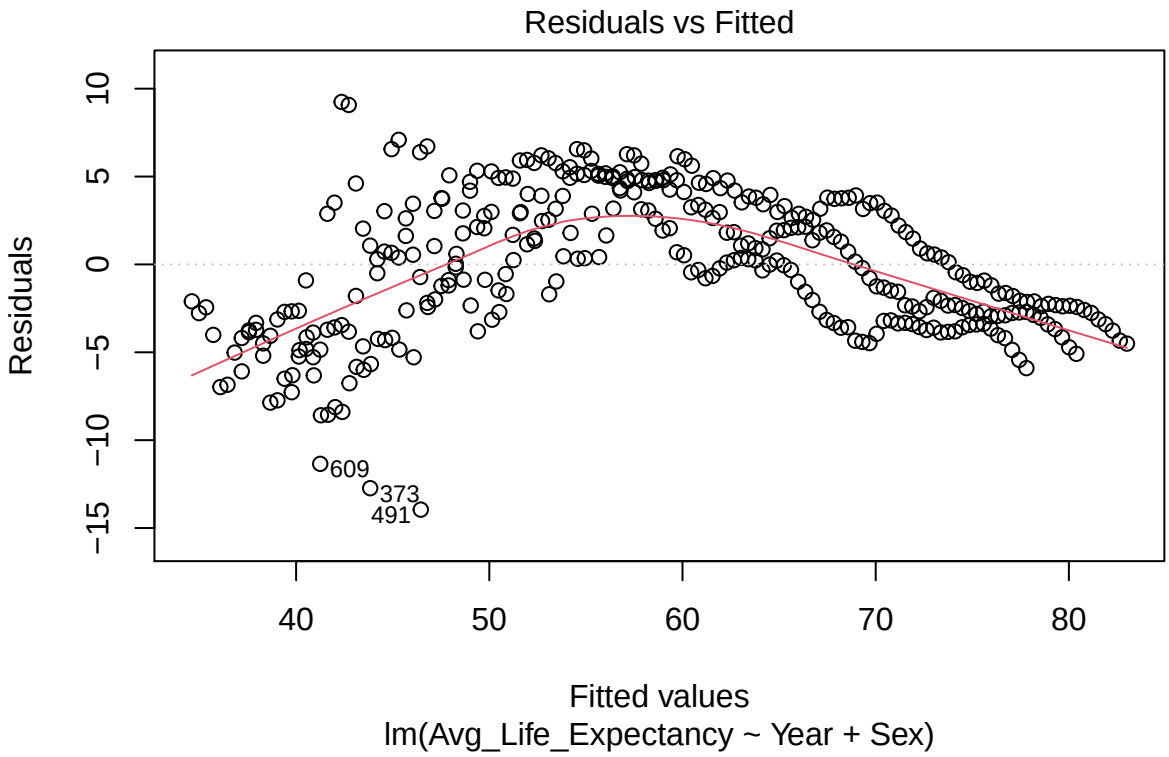
```
## Warning: package 'QuantPsyc' was built under R version 4.0.2

## Loading required package: boot

## Loading required package: MASS

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select

##
## Attaching package: 'QuantPsyc'

## The following object is masked from 'package:base':
##
##     norm
```

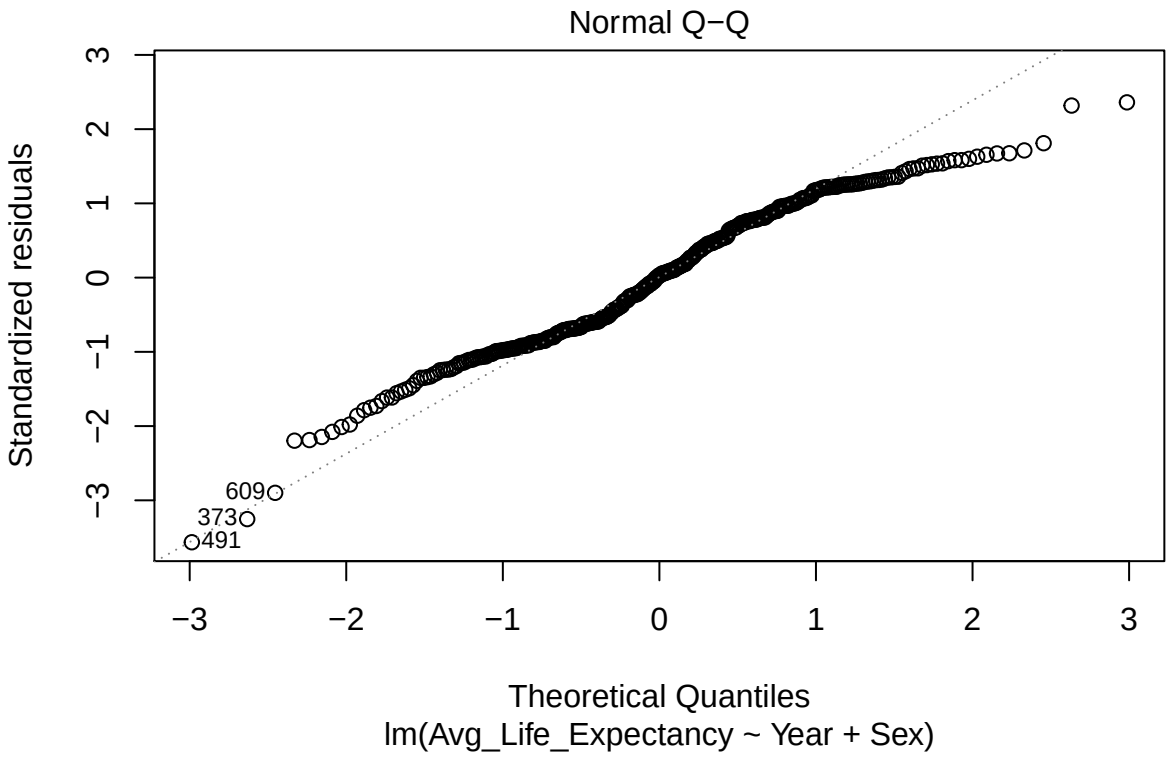

Residuals vs Fitted
Residuals
10
5
0
−5
−10
−15
491
373
609
40
50
60
70
80
Fitted values
lm(Avg_Life_Expectancy ~ Year)

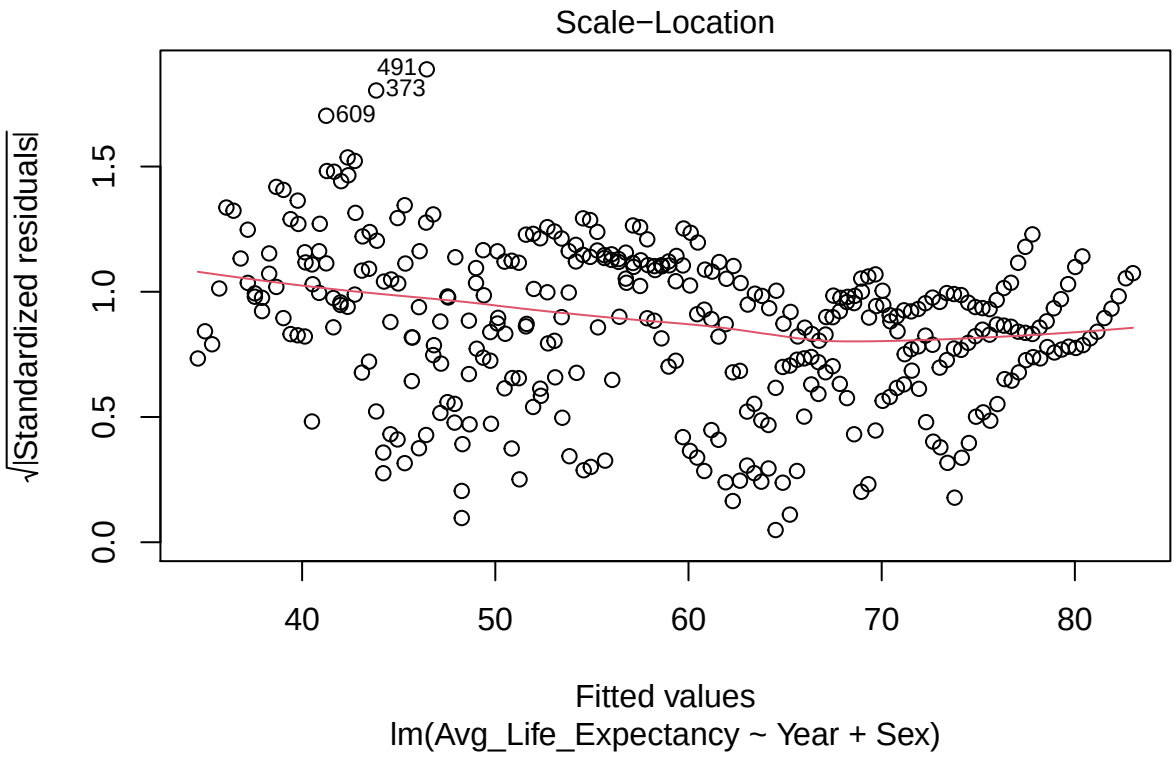

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(Avg_Life_Expectancy ~ Year)
491
373
609
−3
−2
−1
0
1
2
3

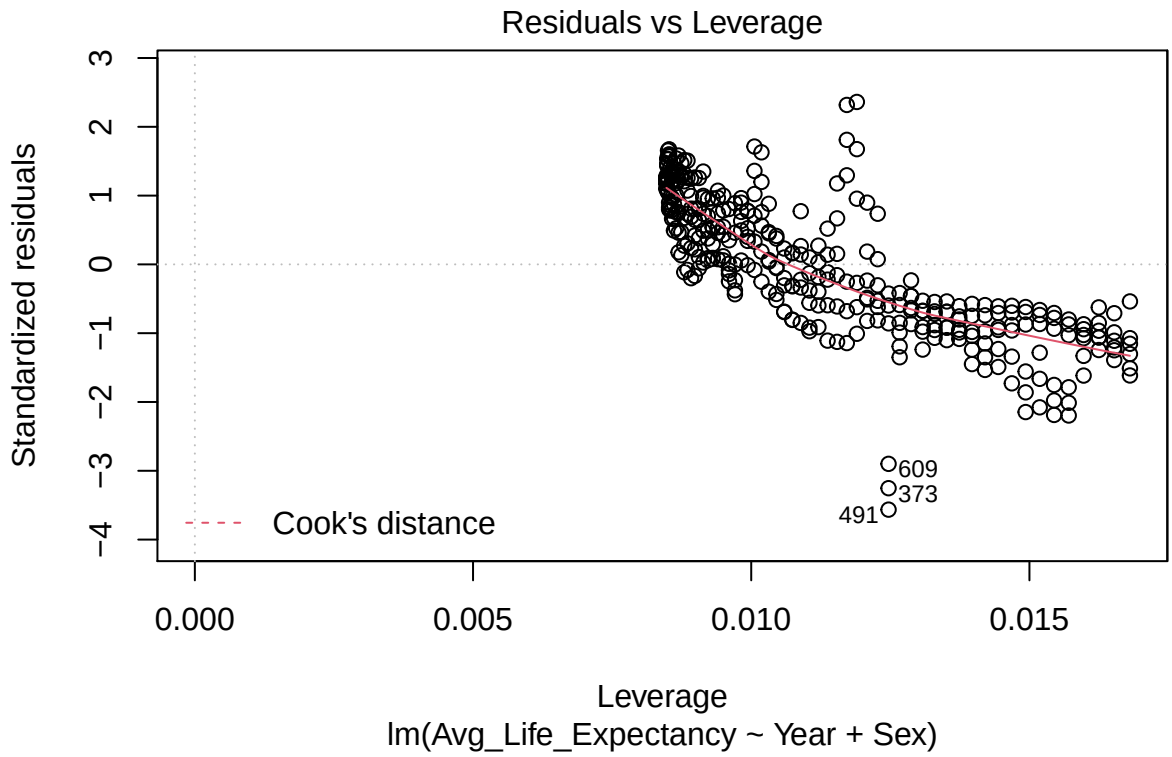

Scale−Location
609
373
491
√|Standardized residuals|
0.0
0.5
1.0
1.5
40
50
60
70
80
Fitted values
lm(Avg_Life_Expectancy ~ Year)


Residuals vs Leverage
Standardized residuals
2
1
0
−1
−2
−3
595
373
609
Cook's distance
0.000
0.002
0.004
0.006
0.008
0.010
Leverage
lm(Avg_Life_Expectancy ~ Year)


Residuals vs Fitted
Residuals
10
5
0
−5
−10
−15
609
373
491
40
50
60
70
80
Fitted values
lm(Avg_Life_Expectancy ~ Year + Sex)


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(Avg_Life_Expectancy ~ Year + Sex)
609
373
491
−3
−2
−1
0
1
2
3


Scale−Location
491
373
609
√|Standardized residuals|
0.0
0.5
1.0
1.5
40
50
60
70
80
Fitted values
lm(Avg_Life_Expectancy ~ Year + Sex)




```
##
## Call:
## lm(formula = Avg_Life_Expectancy ~ Year, data = project_Black_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.9449  -3.0891   0.2899   3.2400   8.7629
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.643e+02  1.366e+01  -48.64   <2e-16 ***
## Year         3.692e-01  6.973e-03   52.95   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.469 on 352 degrees of freedom
## Multiple R-squared:  0.8885, Adjusted R-squared:  0.8882
## F-statistic:  2804 on 1 and 352 DF,  p-value: < 2.2e-16


##
## Call:
## lm(formula = Avg_Life_Expectancy ~ Year + Sex, data = project_Black_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -13.9531  -3.1075   0.1151   3.1759   9.2443
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.643e+02  1.204e+01 -55.166  < 2e-16 ***
## Year         3.692e-01  6.146e-03  60.075  < 2e-16 ***
## SexFemale    2.619e+00  5.128e-01   5.108 5.36e-07 ***
## SexMale     -2.586e+00  5.128e-01  -5.042 7.41e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.939 on 350 degrees of freedom
## Multiple R-squared:  0.9138, Adjusted R-squared:  0.9131
## F-statistic:  1237 on 3 and 350 DF,  p-value: < 2.2e-16
```

```
## Warning in var(if (is.vector(x) || is.factor(x)) x else as.double(x), na.rm =
## na.rm): NAs introduced by coercion
```

```
## Warning in b * sx: longer object length is not a multiple of shorter object
## length
```

```
##       Year  SexFemale    SexMale
##  0.9425881         NA -6.6006715
```

References:

1. https://healthdata.gov/dataset/nchs-death-rates-and-life-expectancy-birth
2. http://euclid.psych.yorku.ca/www/psy6135/tutorials/gapminder.html
3. https://ourworldindata.org/life-expectancy
4. https://rmarkdown.rstudio.com/authoring_basics.html