Life Expentacy analysis and prediction

Laura Ahumada Chris Haub

2022-04-03

# EDA

#### **Looking into the data**

* There are 2938 data entries
* Out of 22 variables only 2 are categorical the rest numerical
* Country is repeated because that is recorded per yer year for each country but we will ignore that.
* No row is repeated

lifeExp=read.csv("~/Documents/Statistics 2/LifeExpectancyData.csv")  
  
# summary(data)  
print(head(lifeExp))

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

# 22 variables  
dim(lifeExp)[2]

[1] 22

# 2938 data size  
nrow(lifeExp)

[1] 2938

# no duplicates  
any(duplicated(lifeExp))

[1] FALSE

attach(lifeExp)

#### **Statistics**

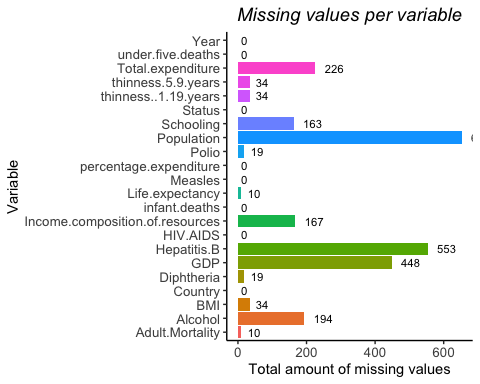
* There are a lot of NA’s and some skewness in most variables

library(Hmisc)  
print(describe(lifeExp))

lifeExp   
  
 22 Variables 2938 Observations  
--------------------------------------------------------------------------------  
Country   
 n missing distinct   
 2938 0 193   
  
lowest : Afghanistan Albania Algeria Angola Antigua and Barbuda   
highest: Venezuela (Bolivarian Republic of) Viet Nam Yemen Zambia Zimbabwe   
--------------------------------------------------------------------------------  
Year   
 n missing distinct Info Mean Gmd .05 .10   
 2938 0 16 0.996 2008 5.318 2000 2001   
 .25 .50 .75 .90 .95   
 2004 2008 2012 2014 2015   
  
lowest : 2000 2001 2002 2003 2004, highest: 2011 2012 2013 2014 2015  
   
Value 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010  
Frequency 183 183 183 183 183 183 183 183 183 183 183  
Proportion 0.062 0.062 0.062 0.062 0.062 0.062 0.062 0.062 0.062 0.062 0.062  
   
Value 2011 2012 2013 2014 2015  
Frequency 183 183 193 183 183  
Proportion 0.062 0.062 0.066 0.062 0.062  
--------------------------------------------------------------------------------  
Status   
 n missing distinct   
 2938 0 2   
   
Value Developed Developing  
Frequency 512 2426  
Proportion 0.174 0.826  
--------------------------------------------------------------------------------  
Life.expectancy   
 n missing distinct Info Mean Gmd .05 .10   
 2928 10 362 1 69.22 10.62 51.4 54.8   
 .25 .50 .75 .90 .95   
 63.1 72.1 75.7 79.7 82.0   
  
lowest : 36.3 39.0 41.0 41.5 42.3, highest: 85.0 86.0 87.0 88.0 89.0  
--------------------------------------------------------------------------------  
Adult.Mortality   
 n missing distinct Info Mean Gmd .05 .10   
 2928 10 425 1 164.8 134.1 13.0 19.0   
 .25 .50 .75 .90 .95   
 74.0 144.0 228.0 336.0 398.3   
  
lowest : 1 2 3 4 5, highest: 693 699 715 717 723  
--------------------------------------------------------------------------------  
infant.deaths   
 n missing distinct Info Mean Gmd .05 .10   
 2938 0 209 0.974 30.3 51.15 0.00 0.00   
 .25 .50 .75 .90 .95   
 0.00 3.00 22.00 58.00 94.15   
  
lowest : 0 1 2 3 4, highest: 1400 1500 1600 1700 1800  
--------------------------------------------------------------------------------  
Alcohol   
 n missing distinct Info Mean Gmd .05 .10   
 2744 194 1076 0.999 4.603 4.555 0.0100 0.0100   
 .25 .50 .75 .90 .95   
 0.8775 3.7550 7.7025 10.7570 11.9600   
  
lowest : 0.01 0.02 0.03 0.04 0.05, highest: 16.35 16.58 16.99 17.31 17.87  
--------------------------------------------------------------------------------  
percentage.expenditure   
 n missing distinct Info Mean Gmd .05 .10   
 2938 0 2328 0.991 738.3 1238 0.000 0.000   
 .25 .50 .75 .90 .95   
 4.685 64.913 441.534 1852.948 4506.638   
  
lowest : 0.000000e+00 9.987219e-02 1.080560e-01 2.756483e-01 3.284181e-01  
highest: 1.837933e+04 1.882287e+04 1.896135e+04 1.909905e+04 1.947991e+04  
--------------------------------------------------------------------------------  
Hepatitis.B   
 n missing distinct Info Mean Gmd .05 .10   
 2385 553 87 0.997 80.94 23.25 9 44   
 .25 .50 .75 .90 .95   
 77 92 97 99 99   
  
lowest : 1 2 4 5 6, highest: 95 96 97 98 99  
--------------------------------------------------------------------------------  
Measles   
 n missing distinct Info Mean Gmd .05 .10   
 2938 0 958 0.962 2420 4492 0.0 0.0   
 .25 .50 .75 .90 .95   
 0.0 17.0 360.2 3580.1 9985.6   
  
lowest : 0 1 2 3 4, highest: 133802 141258 168107 182485 212183  
--------------------------------------------------------------------------------  
BMI   
 n missing distinct Info Mean Gmd .05 .10   
 2904 34 608 1 38.32 22.87 5.20 11.90   
 .25 .50 .75 .90 .95   
 19.30 43.50 56.20 61.80 64.78   
  
lowest : 1.0 1.4 1.8 1.9 2.0, highest: 79.3 81.6 82.8 83.3 87.3  
--------------------------------------------------------------------------------  
under.five.deaths   
 n missing distinct Info Mean Gmd .05 .10   
 2938 0 252 0.979 42.04 71.19 0 0   
 .25 .50 .75 .90 .95   
 0 4 28 87 138   
  
lowest : 0 1 2 3 4, highest: 2100 2200 2300 2400 2500  
--------------------------------------------------------------------------------  
Polio   
 n missing distinct Info Mean Gmd .05 .10   
 2919 19 73 0.996 82.55 21.45 9 52   
 .25 .50 .75 .90 .95   
 78 93 97 99 99   
  
lowest : 3 4 5 6 7, highest: 95 96 97 98 99  
--------------------------------------------------------------------------------  
Total.expenditure   
 n missing distinct Info Mean Gmd .05 .10   
 2712 226 818 1 5.938 2.771 1.930 2.840   
 .25 .50 .75 .90 .95   
 4.260 5.755 7.492 9.120 9.760   
  
lowest : 0.37 0.65 0.74 0.76 0.92, highest: 17.00 17.14 17.20 17.24 17.60  
--------------------------------------------------------------------------------  
Diphtheria   
 n missing distinct Info Mean Gmd .05 .10   
 2919 19 81 0.996 82.32 21.69 9 49   
 .25 .50 .75 .90 .95   
 78 93 97 99 99   
  
lowest : 2 3 4 5 6, highest: 95 96 97 98 99  
--------------------------------------------------------------------------------  
HIV.AIDS   
 n missing distinct Info Mean Gmd .05 .10   
 2938 0 200 0.777 1.742 2.892 0.100 0.100   
 .25 .50 .75 .90 .95   
 0.100 0.100 0.800 4.400 8.515   
  
lowest : 0.1 0.2 0.3 0.4 0.5, highest: 48.8 49.1 49.9 50.3 50.6  
--------------------------------------------------------------------------------  
GDP   
 n missing distinct Info Mean Gmd .05 .10   
 2490 448 2490 1 7483 11067 68.05 161.46   
 .25 .50 .75 .90 .95   
 463.94 1766.95 5910.81 23726.14 41606.85   
  
lowest : 1.681350e+00 3.685949e+00 4.613575e+00 5.668726e+00 8.376432e+00  
highest: 8.973971e+04 1.137519e+05 1.142938e+05 1.157616e+05 1.191727e+05  
--------------------------------------------------------------------------------  
Population   
 n missing distinct Info Mean Gmd .05 .10   
 2286 652 2278 1 12753375 21575245 9618 29382   
 .25 .50 .75 .90 .95   
 195793 1386542 7420359 25787136 47554416   
  
lowest : 34 36 41 43 123  
highest: 1126135777 1144118674 1161977719 1179681239 1293859294  
--------------------------------------------------------------------------------  
thinness..1.19.years   
 n missing distinct Info Mean Gmd .05 .10   
 2904 34 200 1 4.84 4.527 0.6 0.8   
 .25 .50 .75 .90 .95   
 1.6 3.3 7.2 9.8 13.8   
  
lowest : 0.1 0.2 0.3 0.4 0.5, highest: 27.2 27.3 27.4 27.5 27.7  
--------------------------------------------------------------------------------  
thinness.5.9.years   
 n missing distinct Info Mean Gmd .05 .10   
 2904 34 207 1 4.87 4.596 0.5 0.8   
 .25 .50 .75 .90 .95   
 1.5 3.3 7.2 9.7 13.8   
  
lowest : 0.1 0.2 0.3 0.4 0.5, highest: 28.2 28.3 28.4 28.5 28.6  
--------------------------------------------------------------------------------  
Income.composition.of.resources   
 n missing distinct Info Mean Gmd .05 .10   
 2771 167 625 1 0.6276 0.2271 0.277 0.387   
 .25 .50 .75 .90 .95   
 0.493 0.677 0.779 0.864 0.892   
  
lowest : 0.000 0.253 0.255 0.261 0.266, highest: 0.939 0.941 0.942 0.945 0.948  
--------------------------------------------------------------------------------  
Schooling   
 n missing distinct Info Mean Gmd .05 .10   
 2775 163 173 1 11.99 3.713 5.8 7.7   
 .25 .50 .75 .90 .95   
 10.1 12.3 14.3 15.9 16.8   
  
lowest : 0.0 2.8 2.9 3.0 3.1, highest: 20.3 20.4 20.5 20.6 20.7  
--------------------------------------------------------------------------------

#### **Display of missing Values**

* Most of the missing attributes make sense as they seem to be zero’s.
* The only ones that need to be looked at are Population and Life Expectancy.



#### **Dealing with Missing values**

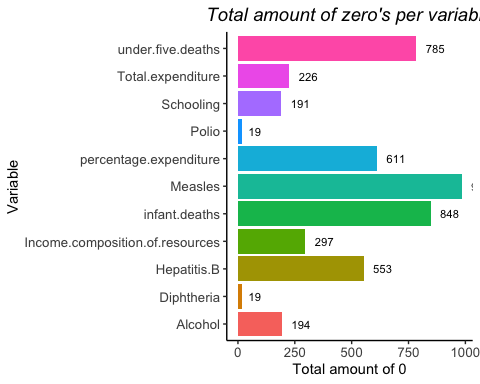
#### **Setting categorical variables as factors**

#### **Replacing some of the data that was set as NA but its really 0 based on analysis of the data**

* The response variable life expectancy had some missing values. After analysis it was imputed with the mean of the life Expectancy filtered by the year, diseases and other factors that they all had in common which was 73.55.
* The only variables that kept the missing attributes and were not imputed were Adult.Mortality, Population, BMI, GDP.
* The reasoning is on the code below but due to Adult.Mortality, BMI and GDP having very few data missing all rows with their missing values were deleted.
* Since population had too many missing variables and couldn’t be imputed, that variable was dropped from the data. We did run a linear regression to check if population was important before the deletion of this explanatory variable fnd found that it did not have a significant statistical value, therefore it confirm the idea of removing the variable.

#### **Checking 0’s and empty strings**

* The amount of 0 for each variable made sense
* There are no empty string or missing for the categorical columns



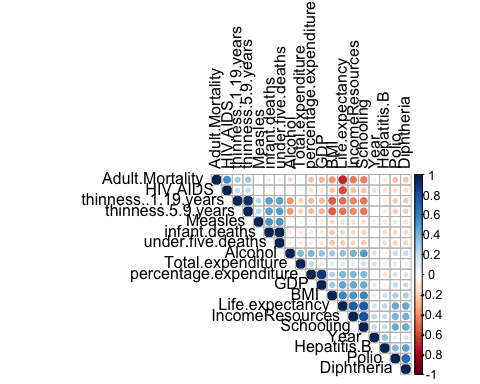
[1] "Empty strings"

Country Status   
 FALSE FALSE

#### **Relationship/correlations**

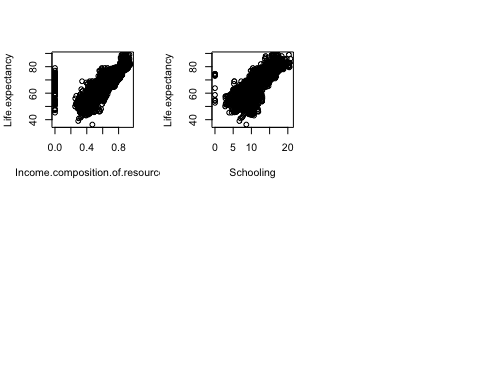
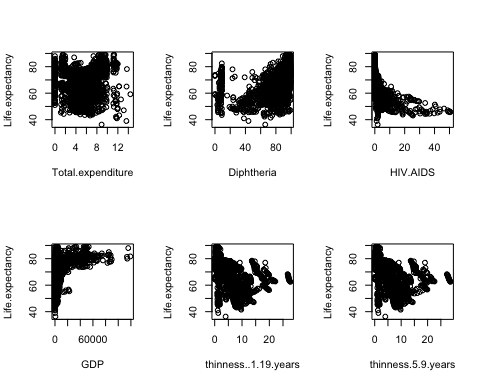
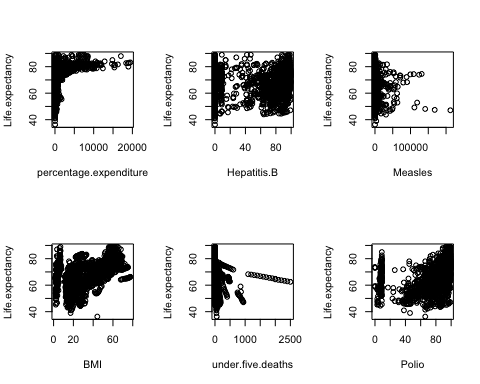
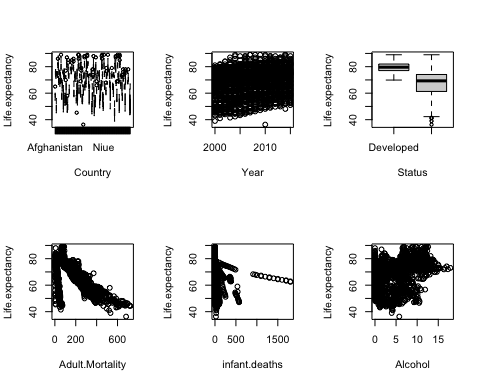
#### **Numeric variable correlation with our target variable**

* Adult mortality and life expectancy have high correlation as one would expect.
* BMI, GDP, Percentage Expenditure, Measles, Thinness, HIV.AIDS and Adult mortality have correlation with our response. We should expect these variables to be significant.
* However, we can see some multicolinearity therefore some may have to be dropped or not be as important since it may be another variable that is driving the correlation.



#### **More detail plots of each explanatory variable vs response variable**

* Schooling , income of resources, DBMI have a strop positive linear trend with the response variable
* Diphtheria and Polio inmulization have a positive linear trend
* Thinness of both 10 to 19 and 5 to 9 years old, as well as HIV.AIDS and measles have a negative linear trend as expected although it may need to logged based on the plots
* Adult mortality has a strong negative trend
* Year seems to have light but constant linear increase
* GDP does seems to have somewhat a linear trend it may also need to be logged
* total expediture health doesn’t seems to have a trend which is interesting
* As for the categorical there does seems to be a difference in developed vs Developing country having Developed with higher life Expectancy and smaller range.



#### **Further updates to the data based on earlier findings**

* Created a new variable called LogYouthThinness, logging the summ of both thinness..1.19.years and thinness.5.9.years variables. Per graph the linear relationship it’s a little clearer.
* Doing the log of GDP , HIV.AIDS and Percentage Expenditure also helped visualize the relationship with the response variable
* Adding a quadratic term to Adult Mortality and Polio were indeed a better fit for the these variables
* If you want to see the graphs once these changes are applied check the

# Objective 1

#### **Multilinear Regression**

* Let’s confirm if the variables seen above influence life expectancy and obtain the explanatory variables to create a model to predict life expectancy

#### **We wil start with backwards and forward model to check variables**

* Varibles selected by Forward were fewer than those for backwards. Income Composition of Resources was the only variables in Forward that wasn’t in Backwards selection.
* Forward choose 11 variables while backwards 15 variables.
* The Adjusted R sqr was the same for both models, 96.

# Forward

## Selection Summary

Variable Adj.

## Step Entered R-Square R-Square C(p) AIC RMSE

1 Country 0.9426 0.9387 1560.5967 11422.0615 2.3870  
2 Adult.Mortality 0.9580 0.9551 488.0772 10656.7847 2.0429  
3 Income.composition.of.resources 0.9654 0.9630 -30.4895 10178.6672 1.8535  
4 Schooling 0.9660 0.9636 -66.0596 10142.1049 1.8394  
5 Year 0.9662 0.9638 -78.3661 10129.2603 1.8343  
6 HIV.AIDS 0.9663 0.9639 -85.7158 10121.5103 1.8310  
7 Hepatitis.B 0.9670 0.9647 -134.6072 10069.7517 1.8115  
8 Measles 0.9671 0.9648 -138.3927 10065.5653 1.8097  
9 under.five.deaths NA NA NA NA NA  
10 infant.deaths NA NA NA NA NA  
11 LogYouthThinness NA NA NA NA NA

# Backward

## Elimination Summary

Variable Adj.

## Step Removed R-Square R-Square C(p) AIC RMSE

1 Income.composition.of.resources 0.9673 0.9648 -136.9990 10068.1833 1.8078  
2 Polio 0.9673 0.9648 -138.9427 10066.2439 1.8075  
3 Diphtheria 0.9673 0.9649 -140.7314 10064.4712 1.8072  
4 BMI 0.9673 0.9649 -142.4299 10062.7954 1.8069  
5 Total.expenditure 0.9673 0.9649 -143.7202 10061.5585 1.8068  
—————————————————————————————————-

##### **Running models using earlier analysis and results**

* Since we saw a relationship between Developed and Developing with Life expectancy on the graphs on the EDA will keep it in the model
* under.five , infant\_death and measles seem to be highlighly correlated which we saw that in the heatmap in EDA. Also after doing models we found that having both in a model lead to some not being statically significant values and high VIF so we so we have decided to go with Measles and not use drop under.Five or infant mortality from model per selection method priority and EDA analysis
* After running models we did see that having Country affected a lot of explanatory variables and thus created a model without country using: **Status+ BMI+ Income.composition.of.resources+GDP+ Year+ Adult.Mortality + Hepatitis.B +HIV.AIDS+ Schooling + under.five.deathsIDS, schooling**
* Another one with country using **Country+ Year+ Adult.Mortality + Hepatitis.B + Measles +HIV.AIDS+ Schooling**
* An a iteration of each linear-linear and log-linear

#### **Check Multicolinearity for the models without new variable or modified explanatory variables**

* Youth and Schooling somewhat high VIF’s for school and logYouthThinness
* There was no multicolinearity for the models without country having the variables earlier filted out.

# Models 2 adding complexity: Logging HIV.AIDS and GDP and including the new variable logYouthThinness

* Models without Country and the updates, ended up with BMI no long being significant just like year. We dropped BMI but kept year due to practical significance
* Models with country: ++ Income.composition.of.resources , log(GDP) , Status needed to be dropped due to not being significant ++ Logging of HIV lead to very high VIF’s so we did not log HIV and only kept the new variable when adding country

### Checking multicolinearity

* Without Country all VIF coefficient are still good regardless of the new variable and logging GDP and HIV
* With Country VIF of school and the new variable were high (11,12) but that is better than when including log of HIV, at the point it was way higher.

# Models 2.2 adding complexity by adding adding square term to adult mortality as well as Polio and new variable

* Adding the square term to polio lead to very high VIF so that removed.
* Keeping HIV.AIDs with the square term of Adult mortality lead to very high VIF’s as well so HIV was removed
* We do see high VIF in the models produced but they aren’t too high and the high influence of adult\_mortality and education with Life expectance makes sense so we saved the models for the final comparison
* For the rest of the optional model adding a square term to the model made the VIF too high

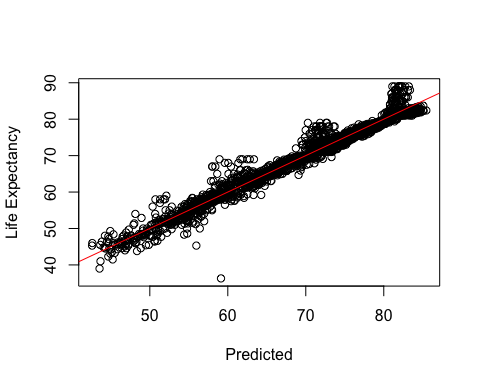
### Checking multicolinearity

* With Country included Square term the VIF’s of all models with country became to high for too many variables
* For those models without country, the square term of adult.mortality were still good and those

##### **We saved the results of every model and checked the assumption as each model was created and created a data frame with the results**

ModelType R\_Sqr HighVIF SimpleModel  
1 LogLinearcountryWith\_newVlogVariable 0.9613081 2 No  
2 reducedLogLinearWCountrySqr 0.9590872 3 No  
3 loglinearWithLogHIVGDP\_and\_newVlogVariable 0.8485022 0 No  
 Assumptions logLinearModel  
1 Yes Yes  
2 Yes Yes  
3 Yes Yes

ModelType R\_Sqr HighVIF SimpleModel Assumptions  
1 LogLinearWCountry 0.9611426 1 Yes Yes  
2 loglinearWithoutCountry 0.8149403 0 Yes Yes  
 logLinearModel  
1 Yes  
2 Yes



#### **Check Assumptions**

* While running each model the assumptions were checked but we will go through this check again for the top normal model and for the complex model
* When ploting the predicted results and the actual life expectancy values we found that it was necessary to only look at the log\_linear models, other wise it wasn’t giving a good prediction.

### As for the best model on the first models obtained from EDA and selection models was LogLinearcountryWith\_newVlogVariable

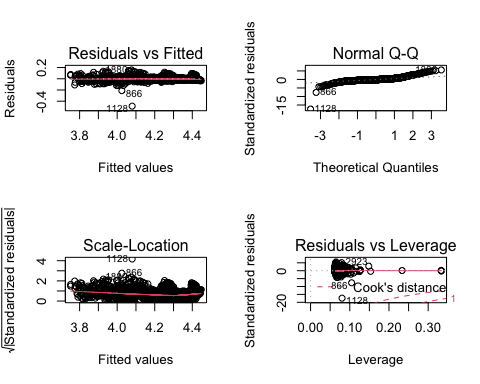
++ Explanatory variables used: Country + Year + Adult.Mortality + HIV.AIDS + Hepatitis.B + Measles + Schooling + LogYouthThinness

++ We do nave 2 high VIF variables which are School and LogYouthThinness which are variables that practically make sense. Actually all of the predictors practically make sense as well ++ After checking the residual plots we found normality, equal variance and we have assumed independence. There was o high VIF in this model and the cooks plot did not show any high leverage and high influence points. ++ All explanatory variables were significant meaning that all had an influence on Life expectancy

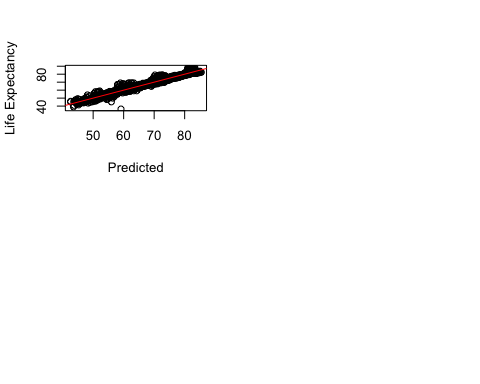
++ Not sure what he means by hypothesis testing

— Provide interpretation of the regression coefficients of your final model including hypothesis testing, interpretation of regression coefficients, and confidence intervals. It’s also good to mention the Practical vs Statistical significance of the predictors.

# R-squared of .9637



Call:  
lm(formula = log(Life.expectancy) ~ Country + Year + Adult.Mortality +   
 HIV.AIDS + Hepatitis.B + Measles + Schooling + LogYouthThinness,   
 data = cleaned\_LifeExpec)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-0.48790 -0.01229 -0.00360 0.00797 0.15758   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -2.860e+00 3.556e-01 -8.044 1.38e-15 \*\*\*  
CountryAlbania 2.452e-01 1.155e-02 21.229 < 2e-16 \*\*\*  
CountryAlgeria 2.248e-01 1.089e-02 20.638 < 2e-16 \*\*\*  
CountryAngola -1.432e-01 1.060e-02 -13.512 < 2e-16 \*\*\*  
CountryAntigua and Barbuda 2.519e-01 1.079e-02 23.340 < 2e-16 \*\*\*  
CountryArgentina 2.443e-01 1.267e-02 19.283 < 2e-16 \*\*\*  
CountryArmenia 2.256e-01 1.128e-02 20.008 < 2e-16 \*\*\*  
CountryAustralia 3.183e-01 1.404e-02 22.680 < 2e-16 \*\*\*  
CountryAustria 3.203e-01 1.199e-02 26.707 < 2e-16 \*\*\*  
CountryAzerbaijan 1.934e-01 1.103e-02 17.537 < 2e-16 \*\*\*  
CountryBahrain 2.401e-01 1.117e-02 21.496 < 2e-16 \*\*\*  
CountryBangladesh 1.709e-01 1.041e-02 16.423 < 2e-16 \*\*\*  
CountryBarbados 2.276e-01 1.146e-02 19.859 < 2e-16 \*\*\*  
CountryBelarus 1.715e-01 1.167e-02 14.701 < 2e-16 \*\*\*  
CountryBelgium 3.130e-01 1.289e-02 24.283 < 2e-16 \*\*\*  
CountryBelize 1.646e-01 1.105e-02 14.903 < 2e-16 \*\*\*  
CountryBenin 3.298e-03 1.043e-02 0.316 0.751836   
CountryBhutan 1.156e-01 1.046e-02 11.055 < 2e-16 \*\*\*  
CountryBosnia and Herzegovina 2.615e-01 1.120e-02 23.346 < 2e-16 \*\*\*  
CountryBotswana 5.921e-02 1.140e-02 5.195 2.23e-07 \*\*\*  
CountryBrazil 2.162e-01 1.141e-02 18.957 < 2e-16 \*\*\*  
CountryBrunei Darussalam 2.503e-01 1.117e-02 22.404 < 2e-16 \*\*\*  
CountryBulgaria 2.124e-01 1.154e-02 18.404 < 2e-16 \*\*\*  
CountryBurkina Faso -1.864e-02 1.062e-02 -1.755 0.079403 .   
CountryBurundi -2.260e-02 1.045e-02 -2.162 0.030683 \*   
CountryCabo Verde 2.095e-01 1.074e-02 19.503 < 2e-16 \*\*\*  
CountryCambodia 1.038e-01 1.045e-02 9.931 < 2e-16 \*\*\*  
CountryCameroon -2.762e-02 1.058e-02 -2.611 0.009080 \*\*   
CountryCanada 3.412e-01 1.331e-02 25.644 < 2e-16 \*\*\*  
CountryCentral African Republic -1.038e-01 1.085e-02 -9.569 < 2e-16 \*\*\*  
CountryChad -9.613e-02 1.073e-02 -8.963 < 2e-16 \*\*\*  
CountryChile 3.072e-01 1.246e-02 24.658 < 2e-16 \*\*\*  
CountryChina 2.588e-01 1.192e-02 21.703 < 2e-16 \*\*\*  
CountryColombia 2.243e-01 1.129e-02 19.868 < 2e-16 \*\*\*  
CountryComoros 5.641e-02 1.049e-02 5.375 8.42e-08 \*\*\*  
CountryCosta Rica 2.916e-01 1.149e-02 25.369 < 2e-16 \*\*\*  
CountryCroatia 2.628e-01 1.174e-02 22.393 < 2e-16 \*\*\*  
CountryCuba 2.717e-01 1.152e-02 23.597 < 2e-16 \*\*\*  
CountryCyprus 3.048e-01 1.215e-02 25.079 < 2e-16 \*\*\*  
CountryDenmark 3.066e-01 1.300e-02 23.593 < 2e-16 \*\*\*  
CountryDjibouti 8.146e-02 1.074e-02 7.583 4.87e-14 \*\*\*  
CountryDominican Republic 2.190e-01 1.110e-02 19.718 < 2e-16 \*\*\*  
CountryEcuador 2.461e-01 1.175e-02 20.937 < 2e-16 \*\*\*  
CountryEl Salvador 2.041e-01 1.148e-02 17.781 < 2e-16 \*\*\*  
CountryEquatorial Guinea 3.877e-04 1.074e-02 0.036 0.971205   
CountryEritrea 4.904e-02 1.145e-02 4.282 1.93e-05 \*\*\*  
CountryEstonia 2.411e-01 1.199e-02 20.112 < 2e-16 \*\*\*  
CountryEthiopia 4.632e-02 1.059e-02 4.373 1.28e-05 \*\*\*  
CountryFiji 1.474e-01 1.130e-02 13.050 < 2e-16 \*\*\*  
CountryFinland 3.257e-01 1.318e-02 24.701 < 2e-16 \*\*\*  
CountryFrance 3.418e-01 1.300e-02 26.286 < 2e-16 \*\*\*  
CountryGabon 1.155e-01 1.094e-02 10.558 < 2e-16 \*\*\*  
CountryGeorgia 2.250e-01 1.120e-02 20.085 < 2e-16 \*\*\*  
CountryGermany 3.179e-01 1.263e-02 25.169 < 2e-16 \*\*\*  
CountryGhana 5.765e-02 1.049e-02 5.494 4.37e-08 \*\*\*  
CountryGreece 3.201e-01 1.278e-02 25.043 < 2e-16 \*\*\*  
CountryGrenada 2.117e-01 1.149e-02 18.433 < 2e-16 \*\*\*  
CountryGuatemala 2.168e-01 1.142e-02 18.989 < 2e-16 \*\*\*  
CountryGuinea -1.189e-02 1.051e-02 -1.131 0.258077   
CountryGuinea-Bissau -1.266e-02 1.060e-02 -1.194 0.232474   
CountryGuyana 1.202e-01 1.064e-02 11.294 < 2e-16 \*\*\*  
CountryHaiti 5.224e-02 1.076e-02 4.856 1.28e-06 \*\*\*  
CountryHonduras 2.256e-01 1.111e-02 20.317 < 2e-16 \*\*\*  
CountryHungary 2.402e-01 1.203e-02 19.973 < 2e-16 \*\*\*  
CountryIceland 3.439e-01 1.336e-02 25.731 < 2e-16 \*\*\*  
CountryIndia 1.294e-01 1.107e-02 11.685 < 2e-16 \*\*\*  
CountryIndonesia 1.530e-01 1.117e-02 13.694 < 2e-16 \*\*\*  
CountryIraq 1.780e-01 1.143e-02 15.574 < 2e-16 \*\*\*  
CountryIreland 3.199e-01 1.432e-02 22.339 < 2e-16 \*\*\*  
CountryIsrael 3.177e-01 1.243e-02 25.552 < 2e-16 \*\*\*  
CountryItaly 3.336e-01 1.317e-02 25.328 < 2e-16 \*\*\*  
CountryJamaica 2.492e-01 1.142e-02 21.828 < 2e-16 \*\*\*  
CountryJapan 3.527e-01 1.214e-02 29.048 < 2e-16 \*\*\*  
CountryJordan 2.103e-01 1.114e-02 18.872 < 2e-16 \*\*\*  
CountryKazakhstan 1.259e-01 1.147e-02 10.970 < 2e-16 \*\*\*  
CountryKenya 4.417e-02 1.066e-02 4.145 3.52e-05 \*\*\*  
CountryKiribati 1.267e-01 1.395e-02 9.085 < 2e-16 \*\*\*  
CountryKuwait 2.206e-01 1.134e-02 19.462 < 2e-16 \*\*\*  
CountryLatvia 2.202e-01 1.181e-02 18.637 < 2e-16 \*\*\*  
CountryLebanon 2.274e-01 1.117e-02 20.357 < 2e-16 \*\*\*  
CountryLesotho -1.789e-02 1.191e-02 -1.502 0.133222   
CountryLiberia 6.253e-03 1.054e-02 0.593 0.552951   
CountryLibya 2.028e-01 1.222e-02 16.599 < 2e-16 \*\*\*  
CountryLithuania 2.024e-01 1.185e-02 17.085 < 2e-16 \*\*\*  
CountryLuxembourg 3.186e-01 1.218e-02 26.152 < 2e-16 \*\*\*  
CountryMadagascar 8.178e-02 1.046e-02 7.815 8.31e-15 \*\*\*  
CountryMalawi -4.497e-02 1.123e-02 -4.003 6.45e-05 \*\*\*  
CountryMalaysia 2.183e-01 1.079e-02 20.238 < 2e-16 \*\*\*  
CountryMaldives 2.377e-01 1.070e-02 22.216 < 2e-16 \*\*\*  
CountryMali -3.670e-02 1.060e-02 -3.461 0.000547 \*\*\*  
CountryMalta 3.182e-01 1.251e-02 25.428 < 2e-16 \*\*\*  
CountryMauritania 9.223e-02 1.058e-02 8.713 < 2e-16 \*\*\*  
CountryMauritius 2.034e-01 1.094e-02 18.584 < 2e-16 \*\*\*  
CountryMexico 2.551e-01 1.149e-02 22.198 < 2e-16 \*\*\*  
CountryMongolia 1.168e-01 1.129e-02 10.340 < 2e-16 \*\*\*  
CountryMontenegro 2.526e-01 1.120e-02 22.557 < 2e-16 \*\*\*  
CountryMorocco 2.050e-01 1.062e-02 19.292 < 2e-16 \*\*\*  
CountryMozambique -3.403e-03 1.105e-02 -0.308 0.758064   
CountryMyanmar 9.856e-02 1.040e-02 9.478 < 2e-16 \*\*\*  
CountryNamibia 1.257e-01 1.105e-02 11.374 < 2e-16 \*\*\*  
CountryNepal 1.252e-01 1.050e-02 11.933 < 2e-16 \*\*\*  
CountryNetherlands 3.278e-01 1.295e-02 25.312 < 2e-16 \*\*\*  
CountryNew Zealand 3.234e-01 1.441e-02 22.446 < 2e-16 \*\*\*  
CountryNicaragua 2.300e-01 1.124e-02 20.453 < 2e-16 \*\*\*  
CountryNiger 9.203e-03 1.081e-02 0.852 0.394503   
CountryNigeria -6.907e-02 1.116e-02 -6.187 7.22e-10 \*\*\*  
CountryNorway 3.396e-01 1.336e-02 25.419 < 2e-16 \*\*\*  
CountryOman 2.322e-01 1.083e-02 21.440 < 2e-16 \*\*\*  
CountryPakistan 1.066e-01 1.044e-02 10.215 < 2e-16 \*\*\*  
CountryPanama 2.669e-01 1.141e-02 23.395 < 2e-16 \*\*\*  
CountryPapua New Guinea 8.179e-02 1.147e-02 7.128 1.36e-12 \*\*\*  
CountryParaguay 2.250e-01 1.128e-02 19.947 < 2e-16 \*\*\*  
CountryPeru 2.336e-01 1.187e-02 19.683 < 2e-16 \*\*\*  
CountryPhilippines 1.483e-01 1.074e-02 13.812 < 2e-16 \*\*\*  
CountryPoland 2.440e-01 1.181e-02 20.664 < 2e-16 \*\*\*  
CountryPortugal 3.058e-01 1.298e-02 23.557 < 2e-16 \*\*\*  
CountryQatar 2.629e-01 1.102e-02 23.850 < 2e-16 \*\*\*  
CountryRomania 2.266e-01 1.131e-02 20.037 < 2e-16 \*\*\*  
CountryRussian Federation 1.440e-01 1.141e-02 12.624 < 2e-16 \*\*\*  
CountryRwanda 3.970e-02 1.055e-02 3.765 0.000171 \*\*\*  
CountrySamoa 2.483e-01 1.372e-02 18.098 < 2e-16 \*\*\*  
CountrySao Tome and Principe 1.134e-01 1.074e-02 10.568 < 2e-16 \*\*\*  
CountrySaudi Arabia 2.108e-01 1.098e-02 19.190 < 2e-16 \*\*\*  
CountrySenegal 7.996e-02 1.052e-02 7.598 4.35e-14 \*\*\*  
CountrySerbia 2.317e-01 1.142e-02 20.280 < 2e-16 \*\*\*  
CountrySeychelles 2.011e-01 1.090e-02 18.454 < 2e-16 \*\*\*  
CountrySierra Leone -2.156e-01 1.050e-02 -20.537 < 2e-16 \*\*\*  
CountrySingapore 3.205e-01 1.164e-02 27.535 < 2e-16 \*\*\*  
CountrySlovenia 3.109e-01 1.244e-02 24.996 < 2e-16 \*\*\*  
CountrySolomon Islands 1.587e-01 1.143e-02 13.887 < 2e-16 \*\*\*  
CountrySomalia -4.907e-02 1.933e-02 -2.539 0.011189 \*   
CountrySouth Africa 1.027e-01 1.154e-02 8.899 < 2e-16 \*\*\*  
CountrySpain 3.309e-01 1.320e-02 25.076 < 2e-16 \*\*\*  
CountrySri Lanka 2.127e-01 1.082e-02 19.669 < 2e-16 \*\*\*  
CountrySuriname 1.867e-01 1.096e-02 17.040 < 2e-16 \*\*\*  
CountrySwaziland 8.695e-02 1.302e-02 6.676 3.07e-11 \*\*\*  
CountrySweden 3.452e-01 1.243e-02 27.776 < 2e-16 \*\*\*  
CountrySwitzerland 3.548e-01 1.331e-02 26.662 < 2e-16 \*\*\*  
CountrySyrian Arab Republic 2.367e-01 1.296e-02 18.258 < 2e-16 \*\*\*  
CountryTajikistan 1.334e-01 1.079e-02 12.368 < 2e-16 \*\*\*  
CountryThailand 2.132e-01 1.076e-02 19.812 < 2e-16 \*\*\*  
CountryTimor-Leste 1.021e-01 1.053e-02 9.701 < 2e-16 \*\*\*  
CountryTogo 1.395e-03 1.059e-02 0.132 0.895190   
CountryTonga 2.297e-01 1.470e-02 15.621 < 2e-16 \*\*\*  
CountryTrinidad and Tobago 1.964e-01 1.081e-02 18.170 < 2e-16 \*\*\*  
CountryTunisia 2.203e-01 1.119e-02 19.690 < 2e-16 \*\*\*  
CountryTurkey 2.259e-01 1.096e-02 20.603 < 2e-16 \*\*\*  
CountryTurkmenistan 1.038e-01 1.079e-02 9.622 < 2e-16 \*\*\*  
CountryUganda 7.630e-03 1.076e-02 0.709 0.478328   
CountryUkraine 1.813e-01 1.157e-02 15.673 < 2e-16 \*\*\*  
CountryUnited Arab Emirates 2.451e-01 1.100e-02 22.277 < 2e-16 \*\*\*  
CountryUruguay 2.531e-01 1.204e-02 21.024 < 2e-16 \*\*\*  
CountryUzbekistan 1.511e-01 1.098e-02 13.759 < 2e-16 \*\*\*  
CountryVanuatu 2.089e-01 1.136e-02 18.394 < 2e-16 \*\*\*  
CountryZambia -5.819e-04 1.097e-02 -0.053 0.957724   
CountryZimbabwe 9.731e-03 1.185e-02 0.821 0.411487   
Year 3.427e-03 1.783e-04 19.225 < 2e-16 \*\*\*  
Adult.Mortality -3.455e-05 7.445e-06 -4.641 3.67e-06 \*\*\*  
HIV.AIDS -6.604e-03 2.340e-04 -28.223 < 2e-16 \*\*\*  
Hepatitis.B 1.832e-04 2.504e-05 7.314 3.56e-13 \*\*\*  
Measles -3.892e-07 7.626e-08 -5.103 3.62e-07 \*\*\*  
Schooling 2.587e-03 6.055e-04 4.273 2.01e-05 \*\*\*  
LogYouthThinness 6.481e-03 1.969e-03 3.292 0.001011 \*\*   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.02929 on 2298 degrees of freedom  
Multiple R-squared: 0.9639, Adjusted R-squared: 0.9613   
F-statistic: 378.3 on 162 and 2298 DF, p-value: < 2.2e-16



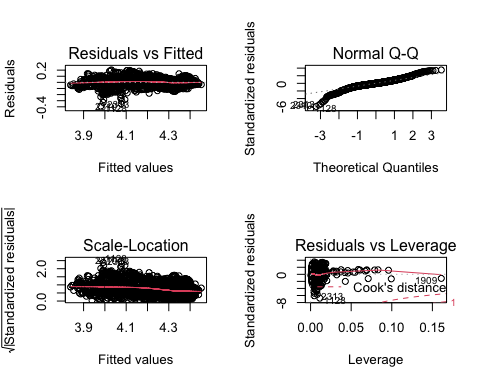
### Here is the confidence interval

\*\*\*\*\*we can add something here\*\*\*\*\*\*

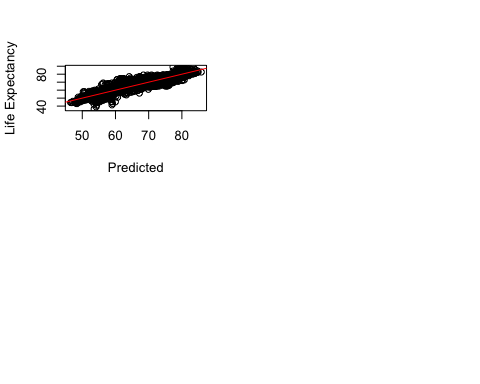
2.5 % 97.5 %  
(Intercept) -4.809252e+02 -3.936552e+02  
CountryAlbania 1.471285e+01 1.754758e+01  
CountryAlgeria 1.341143e+01 1.608474e+01  
CountryAngola -9.112295e+00 -6.512126e+00  
CountryAntigua and Barbuda 1.520198e+01 1.785102e+01  
CountryArgentina 1.446826e+01 1.757701e+01  
CountryArmenia 1.322446e+01 1.599210e+01  
CountryAustralia 2.035401e+01 2.379859e+01  
CountryAustria 2.072350e+01 2.366675e+01  
CountryAzerbaijan 1.090043e+01 1.360634e+01  
CountryBahrain 1.481384e+01 1.755475e+01  
CountryBangladesh 9.601962e+00 1.215580e+01  
CountryBarbados 1.365360e+01 1.646594e+01  
CountryBelarus 9.450519e+00 1.231328e+01  
CountryBelgium 1.986583e+01 2.302940e+01  
CountryBelize 9.026499e+00 1.173779e+01  
CountryBenin -1.276201e+00 1.283012e+00  
CountryBhutan 6.060269e+00 8.626665e+00  
CountryBosnia and Herzegovina 1.589283e+01 1.864133e+01  
CountryBotswana 1.580471e+00 4.377361e+00  
CountryBrazil 1.277302e+01 1.557212e+01  
CountryBrunei Darussalam 1.558631e+01 1.832775e+01  
CountryBulgaria 1.235343e+01 1.518553e+01  
CountryBurkina Faso -2.591496e+00 1.487450e-02  
CountryBurundi -2.765321e+00 -2.000006e-01  
CountryCabo Verde 1.234720e+01 1.498318e+01  
CountryCambodia 5.012625e+00 7.577676e+00  
CountryCameroon -3.224757e+00 -6.287673e-01  
CountryCanada 2.153641e+01 2.480206e+01  
CountryCentral African Republic -7.116732e+00 -4.454380e+00  
CountryChad -6.816230e+00 -4.184121e+00  
CountryChile 1.915626e+01 2.221392e+01  
CountryChina 1.507088e+01 1.799714e+01  
CountryColombia 1.312798e+01 1.589844e+01  
CountryComoros 1.998910e+00 4.574242e+00  
CountryCosta Rica 1.824705e+01 2.106749e+01  
CountryCroatia 1.589882e+01 1.877889e+01  
CountryCuba 1.707625e+01 1.990220e+01  
CountryCyprus 1.916011e+01 2.214306e+01  
CountryDenmark 1.900758e+01 2.219684e+01  
CountryDjibouti 3.102547e+00 5.738827e+00  
CountryDominican Republic 1.263725e+01 1.536237e+01  
CountryEcuador 1.457163e+01 1.745568e+01  
CountryEl Salvador 1.160814e+01 1.442581e+01  
CountryEquatorial Guinea -1.765340e+00 8.700907e-01  
CountryEritrea 1.493097e+00 4.303990e+00  
CountryEstonia 1.443175e+01 1.737326e+01  
CountryEthiopia 1.198119e+00 3.797431e+00  
CountryFiji 7.980640e+00 1.075280e+01  
CountryFinland 2.047843e+01 2.371405e+01  
CountryFrance 2.184930e+01 2.504050e+01  
CountryGabon 4.996194e+00 7.680952e+00  
CountryGeorgia 1.326317e+01 1.601233e+01  
CountryGermany 2.035899e+01 2.345897e+01  
CountryGhana 1.957663e+00 4.532734e+00  
CountryGreece 2.042513e+01 2.356213e+01  
CountryGrenada 1.254506e+01 1.536399e+01  
CountryGuatemala 1.225985e+01 1.506227e+01  
CountryGuinea -2.226200e+00 3.537312e-01  
CountryGuinea-Bissau -2.377944e+00 2.234501e-01  
CountryGuyana 5.976652e+00 8.588998e+00  
CountryHaiti 1.739169e+00 4.379228e+00  
CountryHonduras 1.313939e+01 1.586495e+01  
CountryHungary 1.399526e+01 1.694701e+01  
CountryIceland 2.201652e+01 2.529625e+01  
CountryIndia 6.450468e+00 9.167422e+00  
CountryIndonesia 7.924339e+00 1.066655e+01  
CountryIraq 9.907804e+00 1.271324e+01  
CountryIreland 1.970291e+01 2.321724e+01  
CountryIsrael 2.039194e+01 2.344346e+01  
CountryItaly 2.137251e+01 2.460485e+01  
CountryJamaica 1.466574e+01 1.746771e+01  
CountryJapan 2.269223e+01 2.567174e+01  
CountryJordan 1.239952e+01 1.513407e+01  
CountryKazakhstan 6.376021e+00 9.191816e+00  
CountryKenya 8.527718e-01 3.467834e+00  
CountryKiribati 5.390613e+00 8.813898e+00  
CountryKuwait 1.313726e+01 1.591930e+01  
CountryLatvia 1.302336e+01 1.592268e+01  
CountryLebanon 1.365087e+01 1.639239e+01  
CountryLesotho -2.969506e+00 -4.653762e-02  
CountryLiberia -1.041510e+00 1.544539e+00  
CountryLibya 1.190706e+01 1.490505e+01  
CountryLithuania 1.183903e+01 1.474703e+01  
CountryLuxembourg 2.028074e+01 2.327080e+01  
CountryMadagascar 3.521746e+00 6.089896e+00  
CountryMalawi -4.091149e+00 -1.334293e+00  
CountryMalaysia 1.317122e+01 1.581860e+01  
CountryMaldives 1.481320e+01 1.743897e+01  
CountryMali -3.609813e+00 -1.007414e+00  
CountryMalta 2.002216e+01 2.309277e+01  
CountryMauritania 3.982359e+00 6.579995e+00  
CountryMauritius 1.207134e+01 1.475691e+01  
CountryMexico 1.540986e+01 1.823029e+01  
CountryMongolia 5.712445e+00 8.483932e+00  
CountryMontenegro 1.494432e+01 1.769206e+01  
CountryMorocco 1.201256e+01 1.461970e+01  
CountryMozambique -2.166780e+00 5.439397e-01  
CountryMyanmar 4.703419e+00 7.255413e+00  
CountryNamibia 5.291670e+00 8.004067e+00  
CountryNepal 6.584138e+00 9.159923e+00  
CountryNetherlands 2.077519e+01 2.395347e+01  
CountryNew Zealand 2.033089e+01 2.386679e+01  
CountryNicaragua 1.346944e+01 1.622883e+01  
CountryNiger -9.671218e-01 1.684959e+00  
CountryNigeria -5.475706e+00 -2.736105e+00  
CountryNorway 2.154399e+01 2.482313e+01  
CountryOman 1.419210e+01 1.685024e+01  
CountryPakistan 5.147977e+00 7.709304e+00  
CountryPanama 1.629567e+01 1.909582e+01  
CountryPapua New Guinea 3.036580e+00 5.852678e+00  
CountryParaguay 1.312102e+01 1.588940e+01  
CountryPeru 1.358086e+01 1.649350e+01  
CountryPhilippines 7.843599e+00 1.047831e+01  
CountryPoland 1.481551e+01 1.771337e+01  
CountryPortugal 1.920506e+01 2.239039e+01  
CountryQatar 1.640681e+01 1.911212e+01  
CountryRomania 1.353914e+01 1.631469e+01  
CountryRussian Federation 7.563685e+00 1.036328e+01  
CountryRwanda 1.007409e+00 3.595473e+00  
CountrySamoa 1.389705e+01 1.726390e+01  
CountrySao Tome and Principe 5.538881e+00 8.173392e+00  
CountrySaudi Arabia 1.264563e+01 1.534093e+01  
CountrySenegal 3.379984e+00 5.962743e+00  
CountrySerbia 1.370115e+01 1.650471e+01  
CountrySeychelles 1.183547e+01 1.450993e+01  
CountrySierra Leone -1.243358e+01 -9.856656e+00  
CountrySingapore 2.077372e+01 2.363022e+01  
CountrySlovenia 1.955177e+01 2.260393e+01  
CountrySolomon Islands 8.170904e+00 1.097522e+01  
CountrySomalia -5.804925e+00 -1.061227e+00  
CountrySouth Africa 3.634409e+00 6.466784e+00  
CountrySpain 2.119931e+01 2.443771e+01  
CountrySri Lanka 1.284863e+01 1.550314e+01  
CountrySuriname 1.039847e+01 1.308794e+01  
CountrySwaziland 2.178063e+00 5.374245e+00  
CountrySweden 2.227919e+01 2.532913e+01  
CountrySwitzerland 2.245374e+01 2.571978e+01  
CountrySyrian Arab Republic 1.391445e+01 1.709571e+01  
CountryTajikistan 6.845304e+00 9.492437e+00  
CountryThailand 1.271611e+01 1.535750e+01  
CountryTimor-Leste 5.004989e+00 7.589159e+00  
CountryTogo -1.500724e+00 1.098357e+00  
CountryTonga 1.241641e+01 1.602436e+01  
CountryTrinidad and Tobago 1.125276e+01 1.390492e+01  
CountryTunisia 1.336052e+01 1.610602e+01  
CountryTurkey 1.352437e+01 1.621488e+01  
CountryTurkmenistan 4.863587e+00 7.511594e+00  
CountryUganda -1.144790e+00 1.495936e+00  
CountryUkraine 1.001299e+01 1.285251e+01  
CountryUnited Arab Emirates 1.506398e+01 1.776381e+01  
CountryUruguay 1.534857e+01 1.830311e+01  
CountryUzbekistan 8.018657e+00 1.071299e+01  
CountryVanuatu 1.175722e+01 1.454429e+01  
CountryZambia -1.601059e+00 1.092336e+00  
CountryZimbabwe -1.308282e+00 1.598951e+00  
Year 2.240337e-01 2.677807e-01  
Adult.Mortality -2.856943e-03 -1.029942e-03  
Hepatitis.B 6.158122e-03 1.230382e-02  
HIV.AIDS -3.565161e-01 -2.990898e-01  
Schooling 5.846601e-02 2.070515e-01  
Measles -2.759797e-05 -8.882334e-06  
LogYouthThinness -1.952378e-02 4.637055e-01

* From the second set of models the best one that had a good prediction was loglinearWithLogHIVGDP\_and\_newVlogVariable which excluded country ++ Explanatory variables used: Status + Income.composition.of.resources + log(GDP) + Year + Adult.Mortality + Hepatitis.B + log(HIV.AIDS) + Schooling + Measles + under.five.deaths + LogYouthThinness ++ After checking the residual plots we found normality, equal variance and we have assumed independence. There was o high VIF in this model and the cooks plot did not show any high leverage and high influence points. ++ Due to it not having Country the R’2 is not as high but having dropped 152 variables and still have a r^2 is pretty good is not bad. ++ All explanatory variables were significant meaning that all had an influence on Life expectancy

loglinearWithLogHIVGDP\_and\_newVlogVariable



Call:  
lm(formula = log(Life.expectancy) ~ Status + Income.composition.of.resources +   
 log(GDP) + Year + Adult.Mortality + Hepatitis.B + log(HIV.AIDS) +   
 Schooling + Measles + under.five.deaths + LogYouthThinness,   
 data = cleaned\_LifeExpec)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-0.38923 -0.02762 -0.00043 0.03082 0.20350   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 3.228e+00 5.549e-01 5.817 6.77e-09 \*\*\*  
StatusDeveloping -1.016e-02 3.870e-03 -2.625 0.00872 \*\*   
Income.composition.of.resources 1.118e-01 9.700e-03 11.520 < 2e-16 \*\*\*  
log(GDP) 6.064e-03 8.182e-04 7.411 1.72e-13 \*\*\*  
Year 3.977e-04 2.775e-04 1.433 0.15199   
Adult.Mortality -2.164e-04 1.244e-05 -17.393 < 2e-16 \*\*\*  
Hepatitis.B 1.091e-04 3.442e-05 3.170 0.00154 \*\*   
log(HIV.AIDS) -4.230e-02 1.040e-03 -40.663 < 2e-16 \*\*\*  
Schooling 7.227e-03 6.638e-04 10.888 < 2e-16 \*\*\*  
Measles -3.283e-07 1.245e-07 -2.638 0.00839 \*\*   
under.five.deaths -2.226e-05 8.278e-06 -2.689 0.00722 \*\*   
LogYouthThinness -6.845e-03 1.446e-03 -4.733 2.34e-06 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.05796 on 2449 degrees of freedom  
Multiple R-squared: 0.8492, Adjusted R-squared: 0.8485   
F-statistic: 1254 on 11 and 2449 DF, p-value: < 2.2e-16



#### **KNN**

* The non parametric model KNN lead to higher R^2

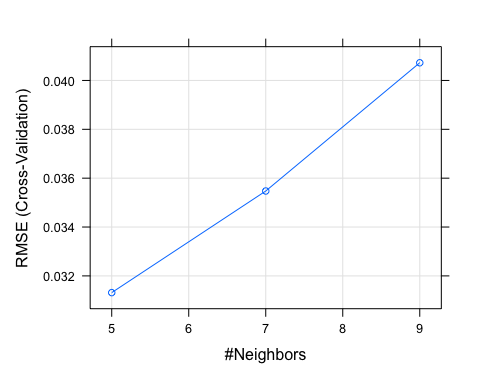
k-Nearest Neighbors   
  
2092 samples  
 7 predictor  
  
Pre-processing: centered (198), scaled (198)   
Resampling: Bootstrapped (25 reps)   
Summary of sample sizes: 2092, 2092, 2092, 2092, 2092, 2092, ...   
Resampling results across tuning parameters:  
  
 k RMSE Rsquared MAE   
 5 0.03650946 0.9385230 0.02104288  
 7 0.04123823 0.9215958 0.02485914  
 9 0.04804861 0.8949759 0.03053419  
  
RMSE was used to select the optimal model using the smallest value.  
The final value used for the model was k = 5.

[1] 65.701

[1] 0.9570129

[1] 65.701

[1] 0.9570129

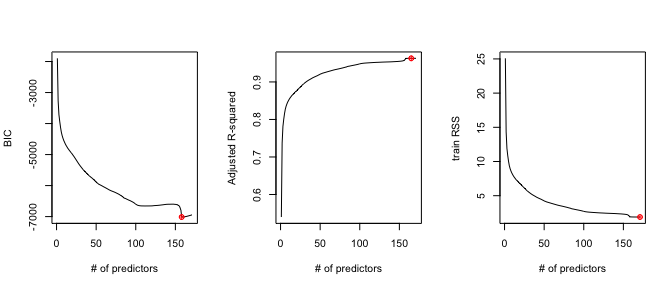


### Further small test on finding the best amount of predictors and checking if it alliance with our best models

* The best BIC states is at 165 predictors according to graph. We have 156 unique countries so we have 155 dummies so if country is included it will require around 10 other predictors to have the best BIC
* Our best model did include country and had 7 extra predictors which is very similar and according to this leads to the best metrics

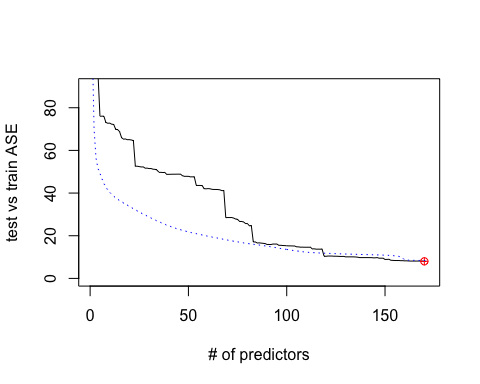
Reordering variables and trying again:

[1] 171

 ### **Checking the ASE at 160, the model seems to be almost perfect no bias nor variance**

Reordering variables and trying again:

testASE<-c()  
#note my index is to 20 since that what I set it in regsubsets  
for (i in 1:171){  
 predictions<-predict.regsubsets(object=reg.fwd,newdata=testLife,id=i)   
 testASE[i]<-mean(((testLife$Life.expectancy)-predictions)^2)  
}  
par(mfrow=c(1,1))  
plot(1:171,testASE,type="l",xlab="# of predictors",ylab="test vs train ASE",ylim=c(0,90))  
index<-which(testASE==min(testASE))  
points(index,testASE[index],col="red",pch=10)  
rss<-summary(reg.fwd)$rss  
lines(1:171,rss/738,lty=3,col="blue") #Dividing by 177 since ASE=RSS/sample size



# With and without Cross Validation same results

## RMSE is 65.9119

## Test R^2 is 0.96015

### <https://koalatea.io/r-knn-regression/>

#it is not runnin on my version so add ```{r} for it to woek model\_tuned <- tuneRF( x=trainLife[,-4], #define predictor variables y=trainLife$Life.expectancy, #define response variable ntreeTry=500, mtryStart=4, stepFactor=1.5, improve=0.01, trace=FALSE #don’t show real-time progress )

randomForestModel <- randomForest( log(Life.expectancy) ~ Country+ Year+ Adult.Mortality + Hepatitis.B + Measles +HIV.AIDS+ Schooling, data=trainLife) predict(randomForestModel, newdata=testLife[-4])