

# Data Analysis

Usama Ahmad

27/06/2021

A Quick look on the data.

```
## Rows: 149
## Columns: 9
## $ Country_name      <chr> "Finland", "Denmark", "Switzerland", "Ice~
## $ Regional_indicator <chr> "Western Europe", "Western Europe", "West~
## $ Log_GDP_Capita     <dbl> 10.775, 10.933, 11.117, 10.878, 10.932, 1~
## $ Ladder_score       <dbl> 7.842, 7.620, 7.571, 7.554, 7.464, 7.392,~
## $ Social_support     <dbl> 0.954, 0.954, 0.942, 0.983, 0.942, 0.954,~
## $ life_expectancy    <dbl> 72.000, 72.700, 74.400, 73.000, 72.400, 7~
## $ Freedom_to_make_life_choices <dbl> 0.949, 0.946, 0.919, 0.955, 0.913, 0.960,~
## $ Generosity         <dbl> -0.098, 0.030, 0.025, 0.160, 0.175, 0.093~
## $ Perceptions_of_corruption <dbl> 0.186, 0.179, 0.292, 0.673, 0.338, 0.270,~
```

## 1 GGpairs plot:

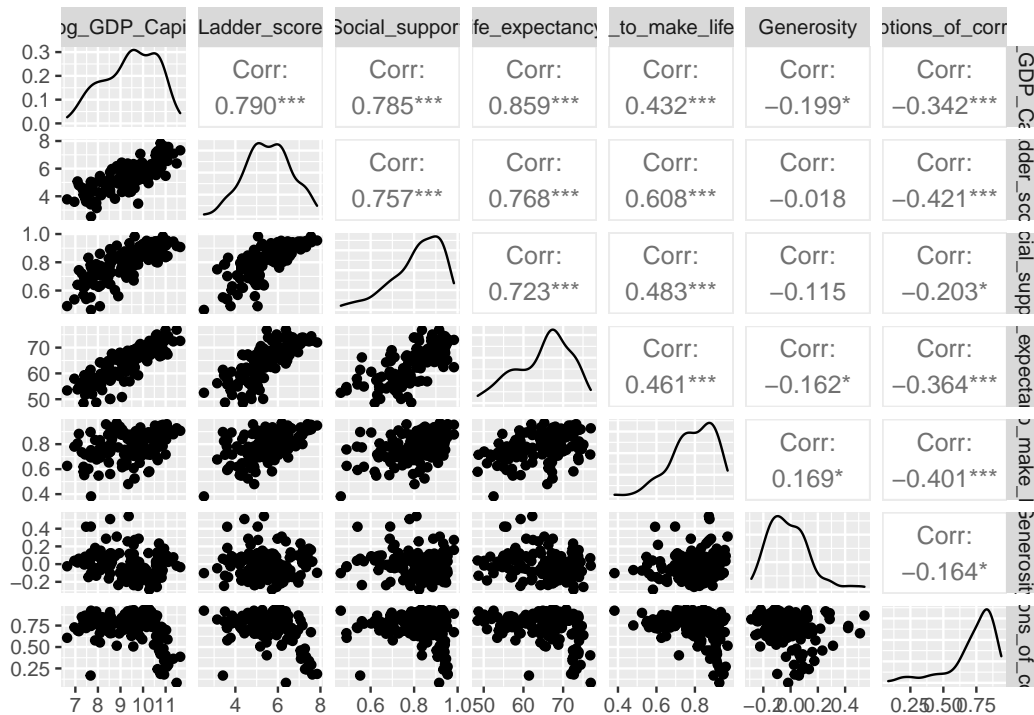


Figure 1: Correlation Plot

High correlation can be seen between variables from this graph, which explains multicollinearity in data. variable selection method will be helpful for fitting a good model.

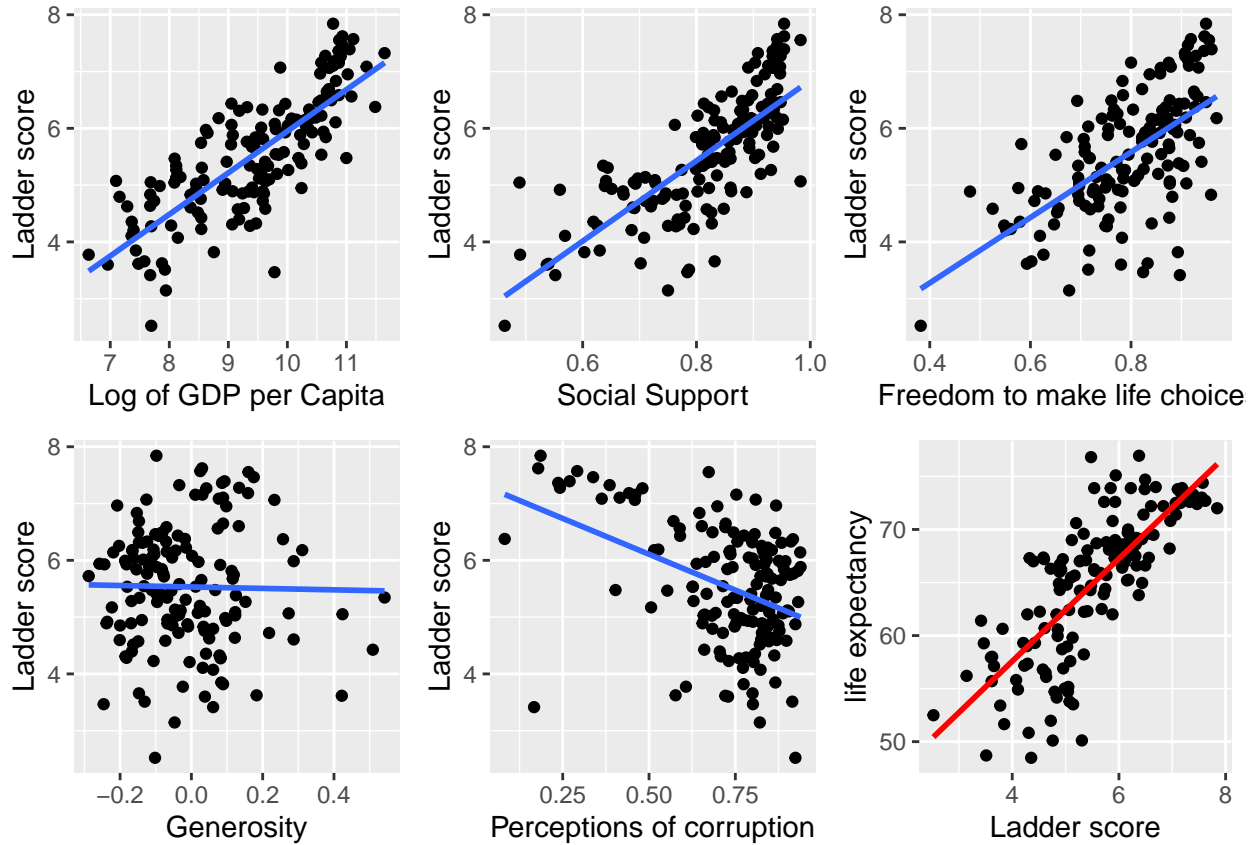
## 2 Summary:

Variable	n	Mean	SD	P25	P50	P75
Log_GDP_Capita	149	9.43	1.16	8.54	9.57	10.42
Ladder_score	149	5.53	1.07	4.85	5.53	6.26
Social_support	149	0.81	0.11	0.75	0.83	0.90
life_expectancy	149	64.99	6.76	59.80	66.60	69.60
Freedom_to_make_life_choices	149	0.79	0.11	0.72	0.80	0.88
Generosity	149	-0.02	0.15	-0.13	-0.04	0.08
Perceptions_of_corruption	149	0.73	0.18	0.67	0.78	0.84

Summary of data explains that values of all the variables are not much different (in good range) and good

## 3 Scatter Plots

Let's check plots, that we can make from this data.



These graph explains relation of explanatory variables with response variable individually, which explains that almost all graph shows some strong effect on response variable except generosity which explains almost no changes in ladder score. some of variables have influential points, if removed can create some good changes in graph. the last graph explains the ladder scores effect on life expectancy which is moderate positive.

## 4 Model Selection

let's have a look at different models ro select a good model to select

### 4.1 Model 1:

$\text{Ladder\_score} \sim \text{Log\_GDP\_Capita}$

Table 1: Model 1: Summary

Variable	Estimate	Std Error	T	P value	Lower CI	Upper CI
intercept	-1.37	0.45	-3.08	0	-2.25	-0.49
Log_GDP_Capita	0.73	0.05	15.61	0	0.64	0.82

### 4.2 Model 2:

$\text{Ladder\_score} \sim \text{Log\_GDP\_Capita} + \text{Freedom\_to\_make\_life\_choices}$

Table 2: Model 2: Summary

Variable	Estimate	Std Error	T	P value	Lower CI	Upper CI
intercept	-1.37	0.45	-3.08	0	-2.25	-0.49
Log_GDP_Capita	0.73	0.05	15.61	0	0.64	0.82

### 4.3 Model 3:

Ladder\_score ~ Log\_GDP\_Capita + Freedom\_to\_make\_life\_choices + Social\_support

Table 3: Model 3: Summary

Variable	Estimate	Std Error	T	P value	Lower CI	Upper CI
intercept	-2.61	0.42	-6.27	0	-3.44	-1.79
Log_GDP_Capita	0.44	0.06	6.76	0	0.31	0.56
Freedom_to_make_life_choices	2.68	0.47	5.76	0	1.76	3.61
Social_support	2.34	0.67	3.50	0	1.02	3.67

### 4.4 Model 4:

Ladder\_score ~ Log\_GDP\_Capita + Freedom\_to\_make\_life\_choices + Social\_support + Perceptions\_of\_corruption

Table 4: Model 4: Summary

Variable	Estimate	Std Error	T	P value	Lower CI	Upper CI
intercept	-1.56	0.58	-2.71	0.01	-2.70	-0.42
Log_GDP_Capita	0.39	0.07	5.86	0.00	0.26	0.52
Freedom_to_make_life_choices	2.26	0.49	4.63	0.00	1.29	3.22
Social_support	2.70	0.67	4.02	0.00	1.37	4.03
Perceptions_of_corruption	-0.74	0.29	-2.59	0.01	-1.31	-0.17

### 4.5 Model 5:

Ladder\_score ~ Log\_GDP\_Capita + Freedom\_to\_make\_life\_choices + Social\_support + Perceptions\_of\_corruption + Generosity

Table 5: Model 5: Summary

Variable	Estimate	Std Error	T	P value	Lower CI	Upper CI
intercept	-1.66	0.58	-2.83	0.00	-2.82	-0.50
Log_GDP_Capita	0.40	0.07	5.92	0.00	0.27	0.54
Freedom_to_make_life_choices	2.15	0.50	4.32	0.00	1.17	3.14
Social_support	2.68	0.67	3.99	0.00	1.35	4.01
Perceptions_of_corruption	-0.69	0.29	-2.36	0.02	-1.27	-0.11
Generosity	0.32	0.32	0.98	0.33	-0.32	0.96

After fitting models with the addition of next variable, we can see that the Model 4 seems promising as it has good R square value and all the p values are explains significant effect on response variable. While in

Model 5, the p-value of *Generosity* is very high and even the scatter plot above showed insignificant effect on Response variable.

## 4.6 Criteria for Model Selection:

Now let's have a look at the AIC value of every model to access if our selected model is good or not.

```
model_selection <- ols_step_both_aic(lm5, details = T)
```

```
## Stepwise Selection Method
## -----
##
## Candidate Terms:
##
## 1 . Log_GDP_Capita
## 2 . Freedom_to_make_life_choices
## 3 . Social_support
## 4 . Perceptions_of_corruption
## 5 . Generosity
##
## Step 0: AIC = 447.0933
## Ladder_score ~ 1
##
## Variables Entered/Removed:
##
##                               Enter New Variables
## -----
## Variable                      DF      AIC      Sum Sq      RSS      R-Sq      Adj. R-Sq
## -----
## Log_GDP_Capita                1    303.457    106.463     64.227     0.624     0.621
## Social_support                1    322.341     97.785     72.905     0.573     0.570
## Freedom_to_make_life_choices  1    380.400     63.047    107.643     0.369     0.365
## Perceptions_of_corruption     1    420.003     30.273    140.417     0.177     0.172
## Generosity                    1    449.046      0.054    170.636     0.000    -0.006
## -----
##
## - Log_GDP_Capita added
##
## Step 1 : AIC = 303.4573
## Ladder_score ~ Log_GDP_Capita
##
##                               Enter New Variables
## -----
## Variable                      DF      AIC      Sum Sq      RSS      R-Sq      Adj. R-Sq
## -----
## Freedom_to_make_life_choices  1    266.161    121.352     49.338     0.711     0.707
## Social_support                1    284.785    114.783     55.907     0.672     0.668
## Perceptions_of_corruption     1    294.895    110.858     59.832     0.649     0.645
## Generosity                    1    297.198    109.926     60.764     0.644     0.639
## -----
##
```

```

## - Freedom_to_make_life_choices added
##
##
## Step 2 : AIC = 266.1605
## Ladder_score ~ Log_GDP_Capita + Freedom_to_make_life_choices
##
## Remove Existing Variables
## -----
## Variable          DF      AIC      Sum Sq      RSS      R-Sq      Adj. R-Sq
## -----
## Freedom_to_make_life_choices    1    303.457    106.463    64.227    0.624    0.621
## Log_GDP_Capita                  1    380.400     63.047   107.643    0.369    0.365
## -----
##
## Enter New Variables
## -----
## Variable          DF      AIC      Sum Sq      RSS      R-Sq      Adj. R-Sq
## -----
## Social_support          1    256.096    125.190    45.500    0.733    0.728
## Perceptions_of_corruption    1    265.184    122.328    48.362    0.717    0.711
## Generosity              1    266.320    121.958    48.732    0.714    0.709
## -----
##
## - Social_support added
##
##
## Step 3 : AIC = 256.0965
## Ladder_score ~ Log_GDP_Capita + Freedom_to_make_life_choices + Social_support
##
## Remove Existing Variables
## -----
## Variable          DF      AIC      Sum Sq      RSS      R-Sq      Adj. R-Sq
## -----
## Social_support          1    266.161    121.352    49.338    0.711    0.707
## Freedom_to_make_life_choices    1    284.785    114.783    55.907    0.672    0.668
## Log_GDP_Capita          1    294.925    110.846    59.844    0.649    0.645
## -----
##
## Enter New Variables
## -----
## Variable          DF      AIC      Sum Sq      RSS      R-Sq      Adj. R-Sq
## -----
## Perceptions_of_corruption    1    251.332    127.209    43.481    0.745    0.738
## Generosity              1    256.036    125.815    44.876    0.737    0.730
## -----
##
## - Perceptions_of_corruption added
##
##
## Step 4 : AIC = 251.3325
## Ladder_score ~ Log_GDP_Capita + Freedom_to_make_life_choices + Social_support + Perceptions_of_corr
##
## Remove Existing Variables
## -----

```

```

## Variable                DF      AIC      Sum Sq      RSS      R-Sq      Adj. R-Sq
## -----
## Perceptions_of_corruption    1    256.096    125.190    45.500    0.733      0.728
## Social_support                1    265.184    122.328    48.362    0.717      0.711
## Freedom_to_make_life_choices  1    270.044    120.725    49.965    0.707      0.701
## Log_GDP_Capita               1    281.215    116.835    53.856    0.684      0.678
## -----
##
##                               Enter New Variables
## -----
## Variable      DF      AIC      Sum Sq      RSS      R-Sq      Adj. R-Sq
## -----
## Generosity    1    252.337    127.499    43.191    0.747      0.738
## -----
##
##
## No more variables to be added or removed.
##
## Final Model Output
## -----
##
##                               Model Summary
## -----
## R                0.863      RMSE                0.550
## R-Squared        0.745      Coef. Var          9.932
## Adj. R-Squared   0.738      MSE                0.302
## Pred R-Squared   0.724      MAE                0.424
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                               ANOVA
## -----
##                               Sum of
##                               Squares      DF      Mean Square      F      Sig.
## -----
## Regression    127.209      4      31.802    105.322    0.0000
## Residual      43.481     144      0.302
## Total        170.690     148
## -----
##
##                               Parameter Estimates
## -----
##                               model      Beta      Std. Error      Std. Beta      t      Sig      lower      up
## -----
## (Intercept)    -1.563      0.577      -2.709    0.008    -2.703    -0.
## Log_GDP_Capita  0.387      0.066      0.418    0.000    0.257    0
## Freedom_to_make_life_choices  2.255      0.487      0.238    0.000    1.293    3
## Social_support  2.699      0.671      0.289    0.000    1.372    4
## Perceptions_of_corruption -0.743      0.287     -0.124    0.011    -1.311    -0
## -----

```

Stepwise selection gives us the same model which we have selected on the basis of p-values, by dropping

*Generosity.*

A brief summary of different criterion of all the fitted models:

Model	R2	Adj_R2	AIC	BIC
1	0.62	0.62	303.46	312.47
2	0.71	0.71	266.16	278.18
3	0.73	0.73	256.10	271.12
4	0.75	0.74	251.33	269.36
5	0.75	0.74	252.34	273.36

According to criterion table, Model 4 and 5 have same R2 and adjusted R2, but if we look at the AIC and BIC values then model 4 has the lowest which is best model for our Data. So model 4 is the same model which we used in stepwise selection to get it.

## 5 Assumption check

Model Equation:

$$\widehat{LS} = \widehat{\alpha} + \widehat{\beta} \text{ LGDP} + \widehat{\gamma} \text{ Freedom} + \widehat{\delta} \text{ SS} + \widehat{\xi} \text{ Corruption}$$

Have to write details here



## 5.1 Residual Plots

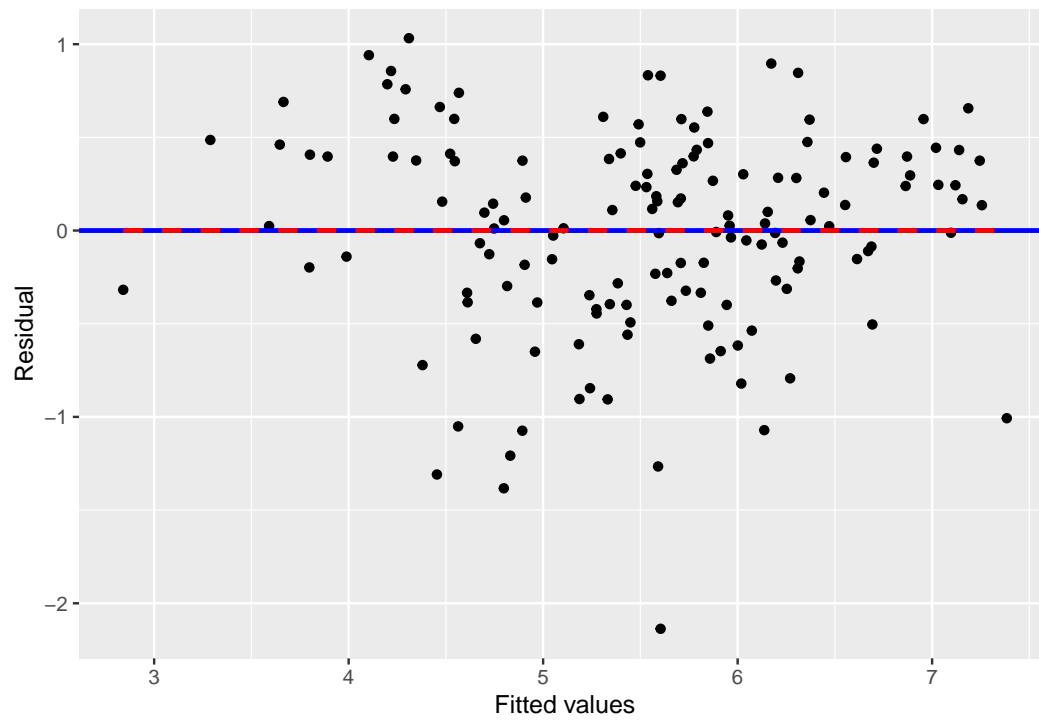


Figure 2: Residual vs fitted values

Residual vs fitted line plot shows equal spread of points around the horizontal line, which shows that mean of residuals are zero.

