A STATISTICAL DATA ANALYSIS ON THE WORLD HAPPINESS SCORES

Report Submitted To

Sister Nivedita University

For The Award Of The Degree Of

Master Of Science In Statistics

By

SATABDA

MAJUMDER

Enrollment Id: 2221209002003

Registration No: 220020035064

Under The Supervision Of

Prof. Biswajit Basak

Department Of Statistics

Sister Nivedita University

Newtown, Kolkata, West Bengal

Abstract

Human philosophy and psychology are both inter connected with happiness of mind. A nation's development and growth always depends on how happy its residents are in real life. This is the keystone behind the analysis of happiness data for the world. Happiness scores are subjected to reflect the nation's inner condition. The data have been presented by Gallup World Poll (GWP). Here the data are obtained from their survey from 2015-2023(initially started in 2012).

Theprimaryexploratorydataanalysisusinggraphsishelpfulin simple understanding of world happiness .The graphical analysis has been conducted for different continents for different years and highest and lowest happy countries around the world.

Thenathoroughmultivariateanalysishasbeenconductedtoknow how this world happiness is influenced by its six indicators(these indicators are globally accepted as the basis of happiness score) and how it ischanging between the continents over different years.Regression and regression diagnostics,principal component analysis have been done in the explanatory analysis.

CONTENTS

INTRODUCTION

In this world happiness report the main focus is given on the happy lives of people's lives. It is mainly based on the life evaluations of different people.These life evaluations help in determining the quality of human life.The World Happiness Report isa point of interest survey of the state of worldwide bliss.Howmodernscienceinfluencesnationalandindividualvarietiesisthemain motivation behind such happiness.The Gallup World Poll, which remainstheprincipalsourceofdatainthisreport,asksrespondentstoevaluate their current life as a whole using the image of a ladder, with the best possible life for them as a 10 and worst possible as a 0. Each respondent provides a numerical response on this scale, referred to as the Cantril ladder. Typically, around1,000responsesaregathered annuallyforeachcountry.Gallupweights are usedtoconstruct population-representative nationalaverages foreachyear in each country.

Theprimaryreportwasdistributedin2012byGallopWorldPoll,the second in 2013, the third in 2015, and it continued further.In the year 2014,no such report was published.This report proceeds to pick up worldwide acknowledgment as governments, organizations and respectful society progressively utilize joy pointers to educate their policy-making decisions.Note that the six key variables that contribute to explaining life evaluations are as follows:

1. **GDPper capita**
2. **Socialsupport**
3. **Healthylife expectancy**
4. **Freedom**
5. **Generosity**
6. **PerceptionOfCorruption(Government trust)**

As already noted ,our happiness rankings arenotbasedonanyindexofthesesix factors. Rather, scores are based on individuals' own assessments of their lives, in particular their answers to the single-item Cantril ladder life-evaluation question.We use observed data on the six variables and estimates of their associationswith life evaluations to help explain the variation of life evaluations across countries.In other words the variables give the estimate of the extent to which they influence the happiness scores.

The factor variables used in the world happiness report are determinantsthatexplainnational-leveldifferencesinlifeevaluations.However, certain variables, such as unemployment or inequality, are not considered because comparable data is not yet available across all countries.In makingthe reports, the expertise in fields including economics, psychology, survey analysis, and national statistics, describe how measurements of well-being can be used effectively to assess the progress of nations, and other topics.

Nowwegivetheexplanationofallthesixvariablesinfluencing the happiness score,

1. GDP per capita is in terms of Purchasing Power Parity (PPP) adjusted to constantinternational dollars (for each year's calculation there is a fixed base year,foreg.for2023,theyearhasbeentakentobe2017.),takenfromtheWorld Development Indicators (WDI) by the World Bank.
2. Thetimeseriesforhealthylifeexpectancyatbirthareconstructedbasedon data from the World Health Organization (WHO) Global Health Observatory data repository.
3.Socialsupportisthenationalaverageofthebinaryresponses(0=no,1=yes)to the Gallup World Poll (GWP) question "If you were in trouble, do you have relatives or friends you can count on to help you whenever you needthem,or not?"
4.Freedomtomakelifechoicesisthenationalaverage ofbinaryresponsestothe GWP question "Are you satisfied or dissatisfied with your freedom to choose what you do with your life?"
5.Generosityistheresidualofregressingthenationalaverageof GWPresponses to the donation question "Have you donated money to a charity in the past month?" on log GDP per capita.
6.Perceptions of corruption are the average of binary answers to two GWP questions: "Is corruption widespread throughout the government or not?" and "Is corruption widespread within businesses or not?" Where data for governmentcorruptionaremissing,theperceptionofbusinesscorruptionisused as the overall corruption-perception measure.

Objective:-Themainobjectives behindsuchdataanalysisare

1.Exploratorydataanalysisusinggraphstoknowthechangeofhappinessscores i.e,the situation of world's happiness from year to year
2.Toknowthevariationofhappinessfromcontinenttocontinent
3.Howtheindicators,globallyacceptedinfluencethehappiness score
4.Findingtheinfluentialpointsusingregressiondiagnostics
5.Findingthemainindicatorsofhappinessbymultivariateanalysisandtoverify whether the situation is same for the consecutive years.

In the first section the exploratory analysis has been done using graphical analysis.Graphsareusedtodisplaythehappinessscore'ssituationindifferent countries and continents and also how they vary over different years.

Inthesecond sectionwehavedonestepwiseregressioninordertofindoutwhich factors are actually influential for determination of world happiness for each year.

Inthethirdsection wehavedonetheregressiondiagnosticstofindoutthe influential data points and their variation over the years.

Inthefourth sectionwehavedoneathoroughprincipalcomponentanalysisto get an idea about the reduction of dimension of the data.

Viewof Data

Herewegraduallystartourstatistical analysisofthedataonhappinessscoresof different countries for the years 2015-2023.Note that the happiness report was started in 2012,but there was a drop in the year 2014,so for data analysis purpose the data have been taken from the year 2015.For our convenience we will do every statistical analysis for each year and thus we can judge how the situation is changing or whether it has been changing at all from year to year.

Thedataonhappinessscoreisaverylargedataset.Weare providing the first 20 rows of the data for understanding.

Country	happiness _score	gdp_per_ capita	Social support	health	freedom
Switzerland	7.587	1.39651	1.34951	0.94143	0.66557
Iceland	7.561	1.30232	1.40223	0.94784	0.62877
Denmark	7.527	1.32548	1.36058	0.87464	0.64938
Norway	7.522	1.459	1.33095	0.88521	0.66973
Canada	7.427	1.32629	1.32261	0.90563	0.63297
Finland	7.406	1.29025	1.31826	0.88911	0.64169
Netherlands	7.378	1.32944	1.28017	0.89284	0.61576
Sweden	7.364	1.33171	1.28907	0.91087	0.6598
NewZealand	7.286	1.25018	1.31967	0.90837	0.63938
Australia	7.284	1.33358	1.30923	0.93156	0.65124
Israel	7.278	1.22857	1.22393	0.91387	0.41319
CostaRica	7.226	0.95578	1.23788	0.86027	0.63376
Austria	7.2	1.33723	1.29704	0.89042	0.62433
Mexico	7.187	1.02054	0.91451	0.81444	0.48181
United States	7.119	1.39451	1.24711	0.86179	0.54604
Brazil	6.983	0.98124	1.23287	0.69702	0.49049
Luxembourg	6.946	1.56391	1.21963	0.91894	0.61583
Ireland	6.94	1.33596	1.36948	0.89533	0.61777
Belgium	6.937	1.30782	1.28566	0.89667	0.5845

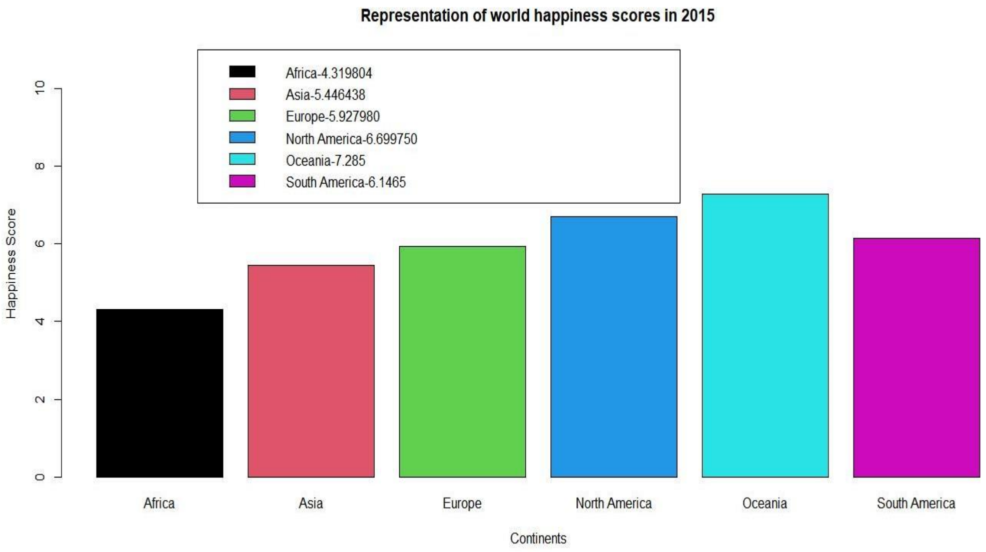
generosity	government_trust	continent	Year
0.29678	0.41978	Europe	2015
0.4363	0.14145	Europe	2015
0.34139	0.48357	Europe	2015
0.34699	0.36503	Europe	2015
0.45811	0.32957	NorthAmerica	2015
0.23351	0.41372	Europe	2015
0.4761	0.31814	Europe	2015
0.36262	0.43844	Europe	2015
0.47501	0.42922	Oceania	2015
0.43562	0.35637	Oceania	2015
0.33172	0.07785	Asia	2015
0.25497	0.10583	SouthAmerica	2015
0.33088	0.18676	Europe	2015
0.14074	0.21312	SouthAmerica	2015
0.40105	0.1589	NorthAmerica	2015
0.14574	0.17521	SouthAmerica	2015
0.28034	0.37798	Europe	2015
0.45901	0.28703	Europe	2015
0.2225	0.2254	Europe	2015

Data source:Kaggle.com

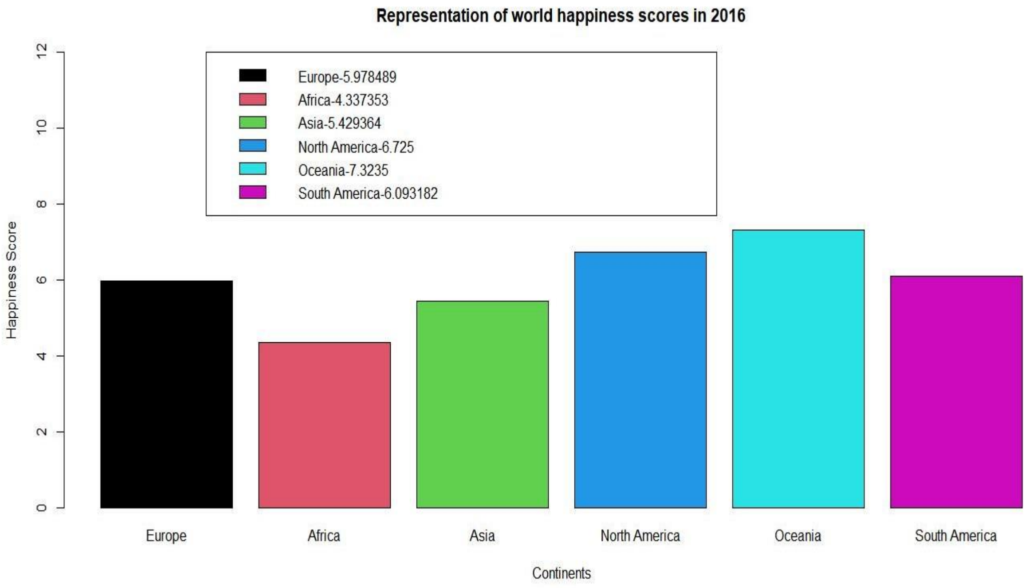
Exploratorydata analysis

Firstwegoforthegraphicalrepresentation,werepresentthehappinessscore for each continent as the mean of the happiness scores of different countries corresponding to each continent.Mean happiness score can be used as the representative value for the continents since there is well balance in the happiness scores for each continent.

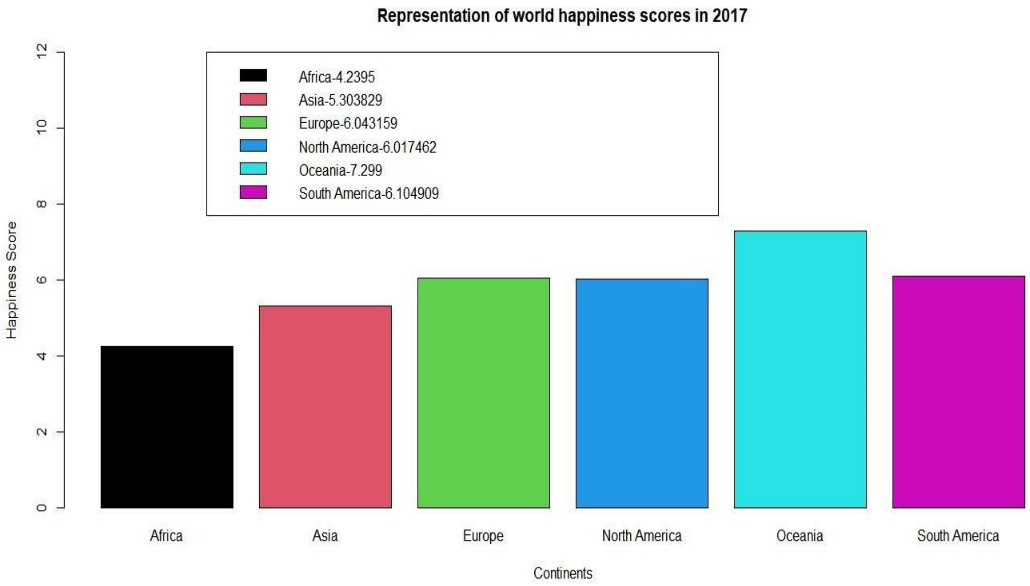
<u>Year2015</u>:



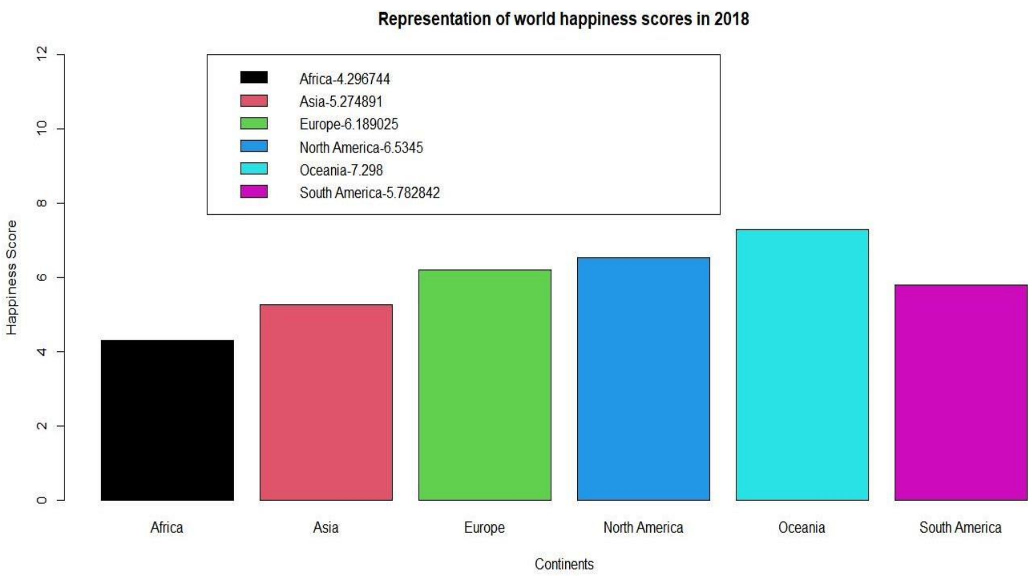
<u>Year2016:</u>



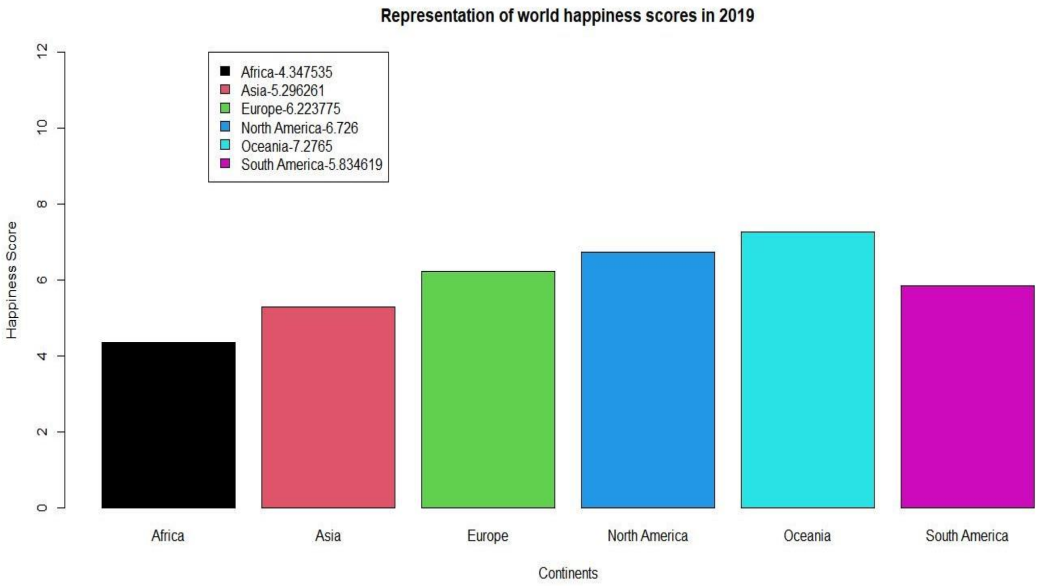
Year2017:



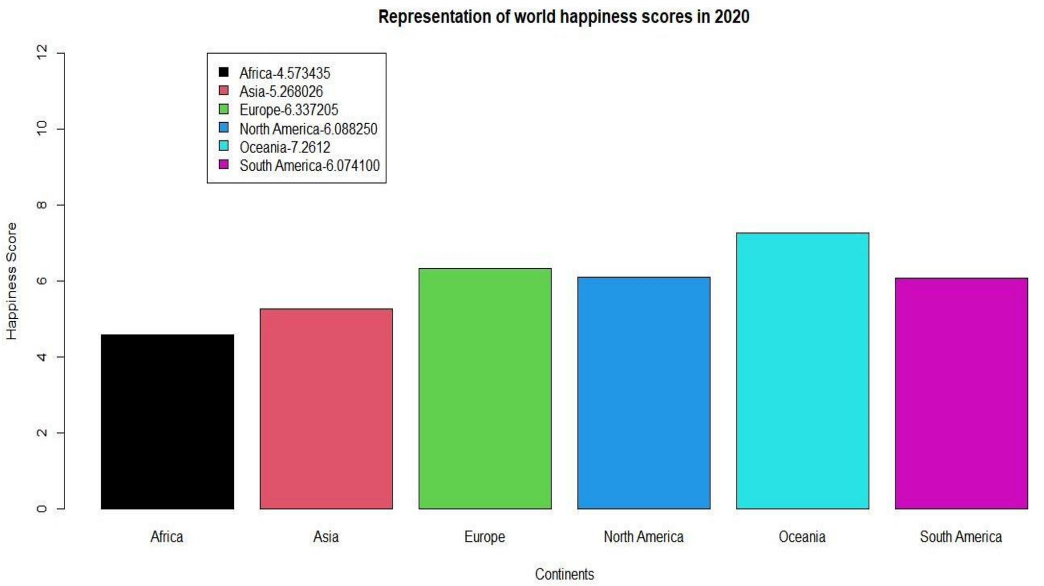
Year2018:



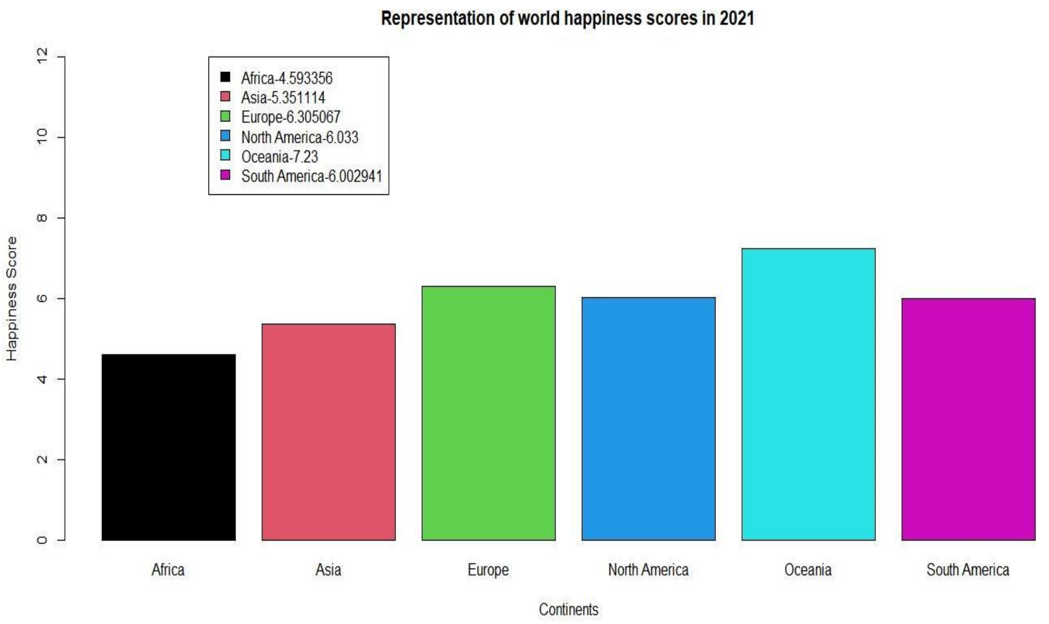
Year2019:



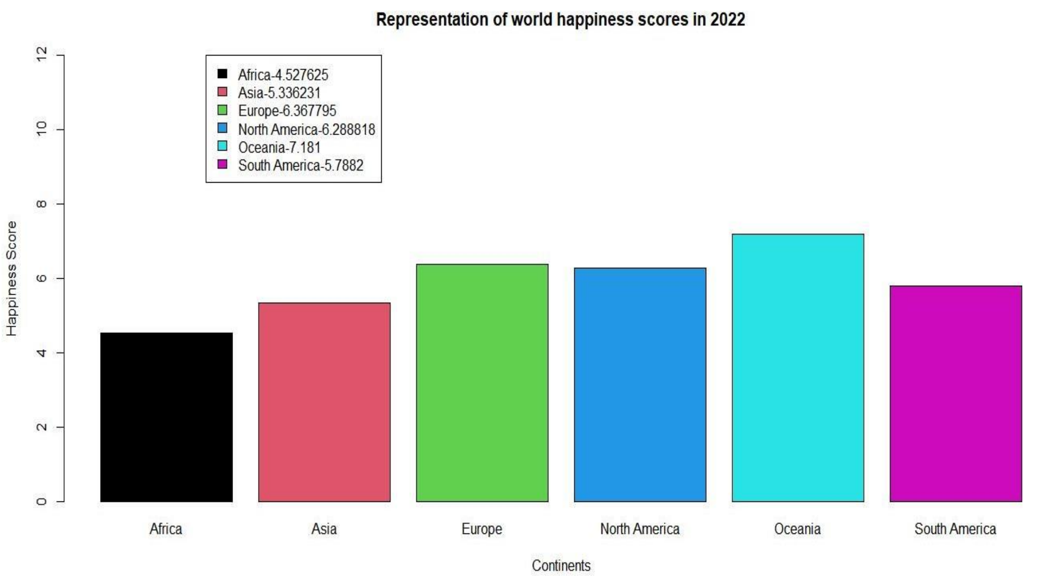
Year2020:



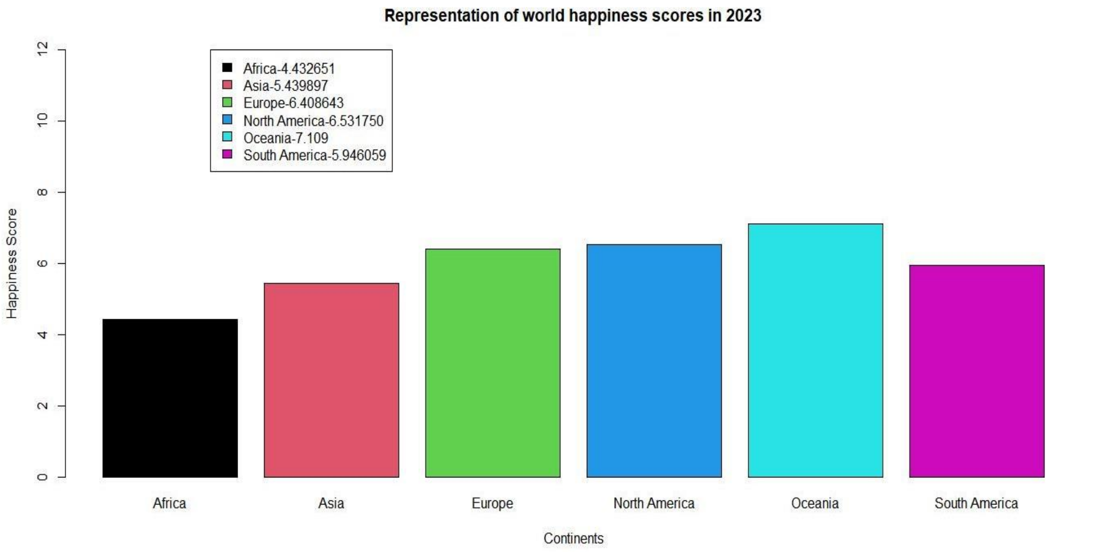
Year2021:



Year2022:



Year2023:



Interpretation:-Theresultsobtainedfromtheabovegraphicalrepresentations are listed below-

1. Ineach yearthecontinentOceaniahashighesthappinessscore,it'sbecause Oceania contains only two countries Australia and New Zealand,both having happiness score above 7
2. IneachyearAfricahaslowesthappinessscore,around4.3-4.5.Therehasbeen nosignificantchangeinhappiness score forAfrica,the highest level ofhappiness score for Africa is 4.59 in 2021,then 4.57 in 2020.
3. For Asia the happiness score ranges from 5.26-5.44 in different years.It has fallenfrom2015to2018,thenithasslightlyincreasedfrom2021to2023(5.35to 5.43)
4. Incomparisontoeastern world,thewesternworldhasbetterpictureinterms of happiness,which means people belonging to western countries are happier thanthoseofeasterncountries,whichmightbeclearduetothefactorsGDPper capita,social support,healthy life expectancy and freedom to make choices.
Nowinwesternworld,wecandrawacomparisonamong Europe,North and South America.
5. From2015to2019,thehappinessscorelevelishigherinNorthAmericathan inEuropeandSouthAmerica.From 2020 to 2022,Europehashigher score than the both Americas.And again,in 2023 North America has gone higher.
6. There is a great fluctuation in the score levels in both Americas.North Americahasexperiencedsuddenfallinhappinessscorefrom6.72to6.01(from 2016 to 2017) and again from 6.72 to 6.08 to 6.03 (from 2019 to 2021-2022).

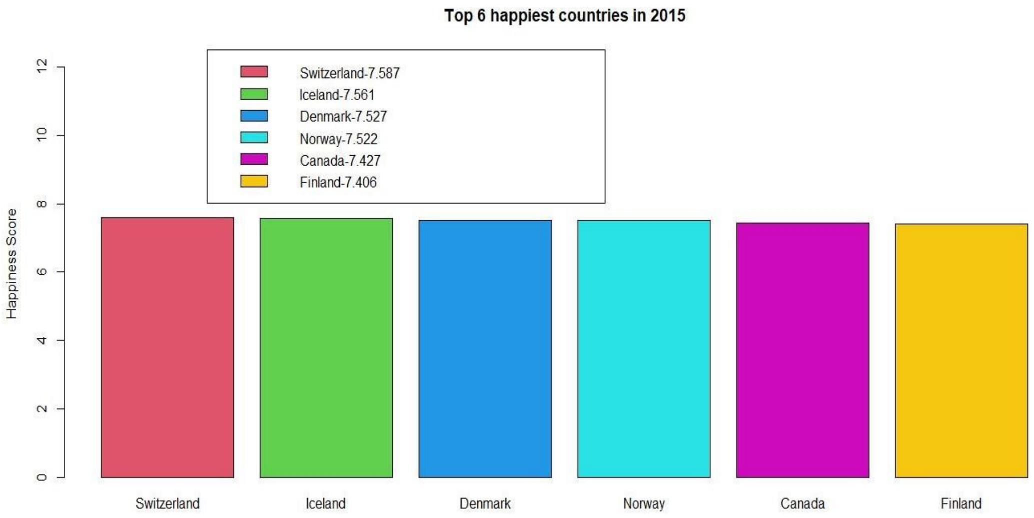
7. Therearealsotwosuddenchangesinscoreof SouthAmerica,from6.10to 5.78(from 2017 to 2018) ,and from 6.00 to 5.78(from 2021 to 2022).There is a sudden increase in happiness score from 5.83 to 6.07 (2019-2020)

8. OnlyEuropehasexperiencedasustainedincreaseinhappinessscorelevel from 2015 to 2023,from 5.92 to 6.40

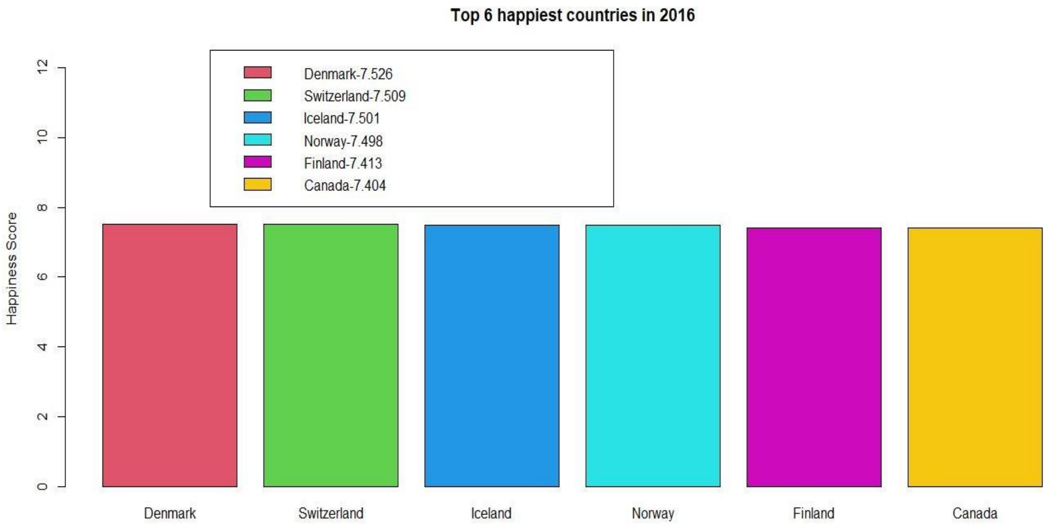
Nowwerepresentthe6tophappiestand6leasthappycountriesforeach yearas given below -

Firstwerepresentthetop6happiestcountries. Year

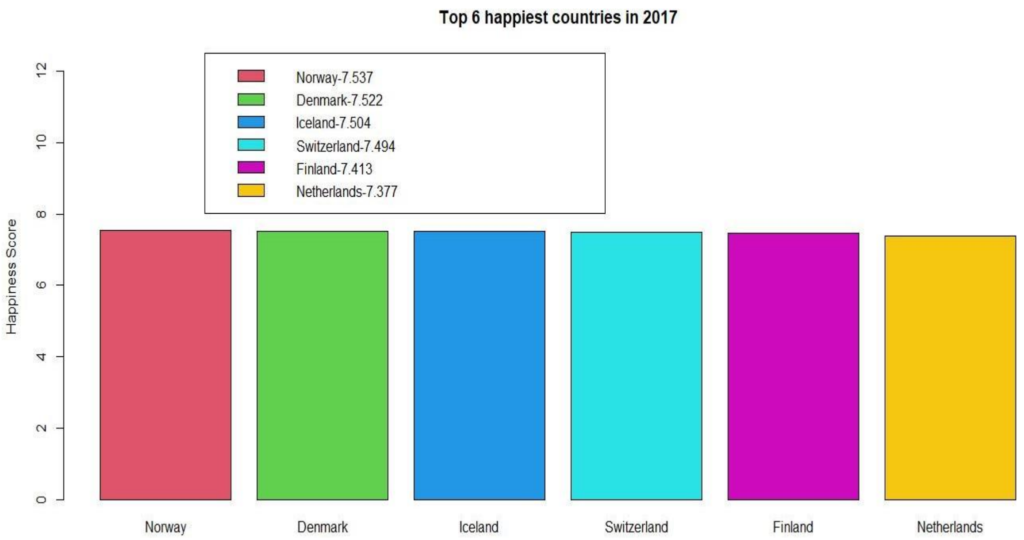
2015:



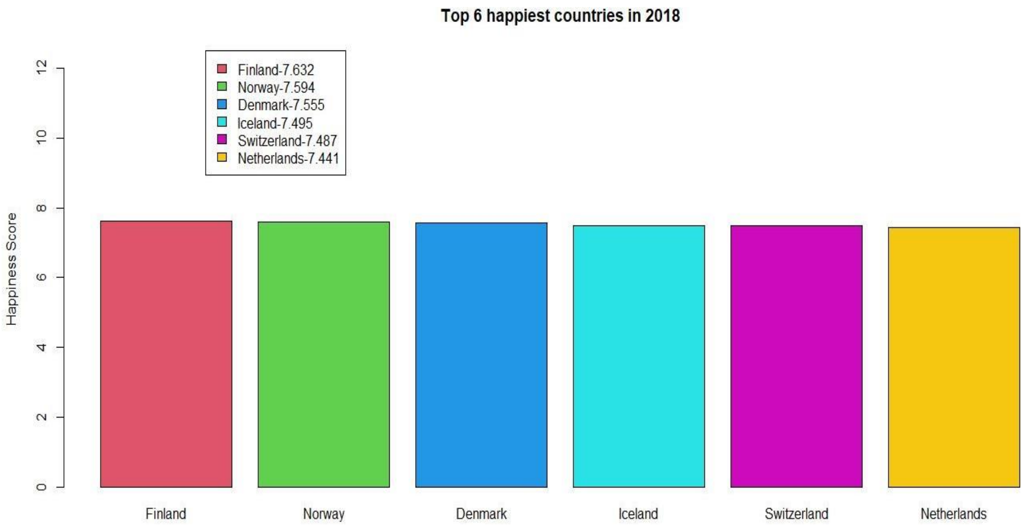
Year2016:



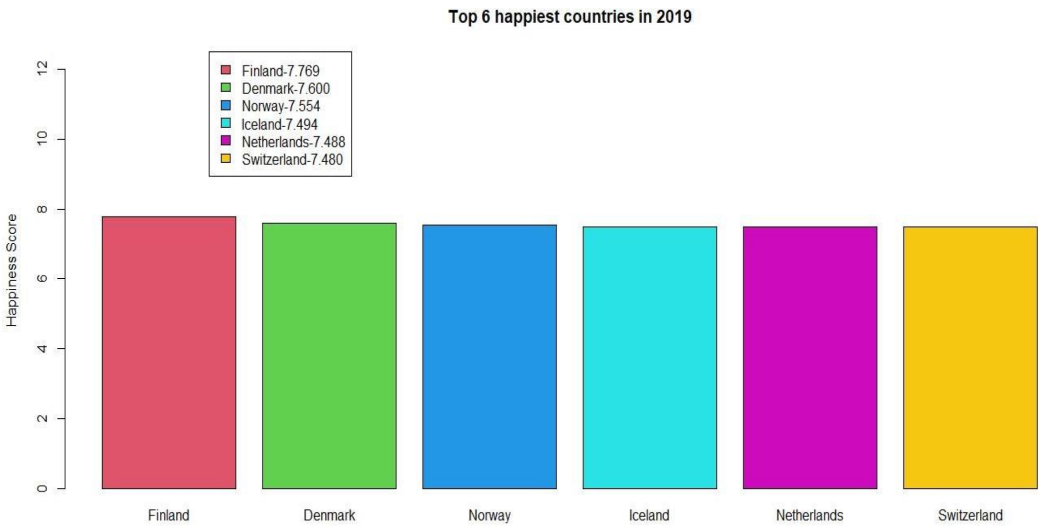
Year2017:



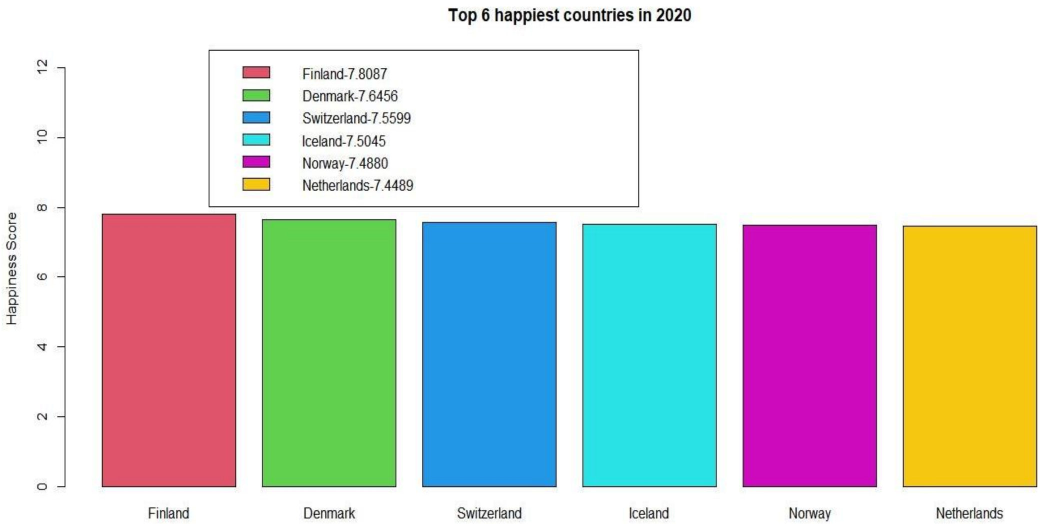
Year2018:



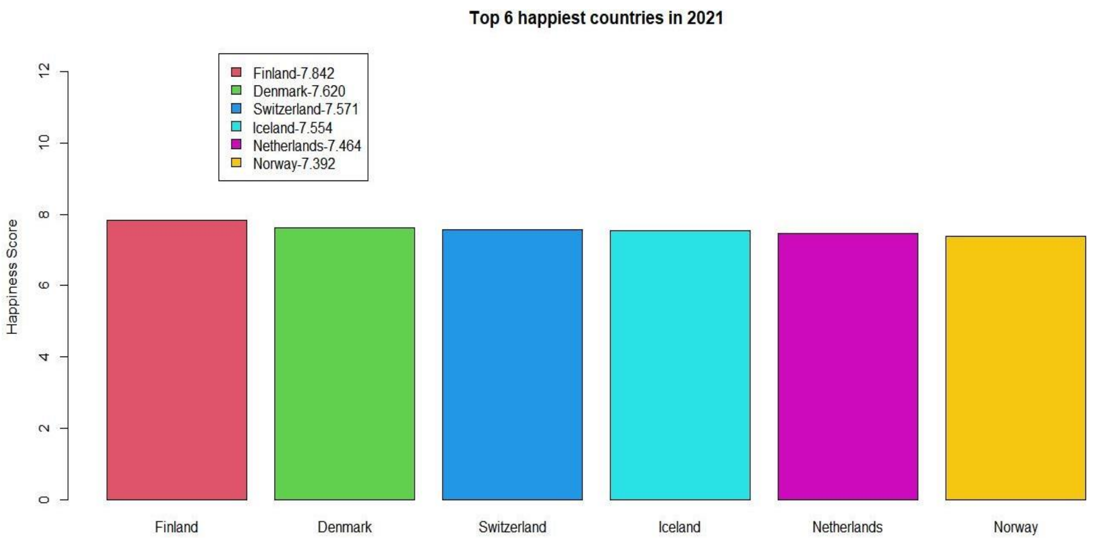
Year2019:



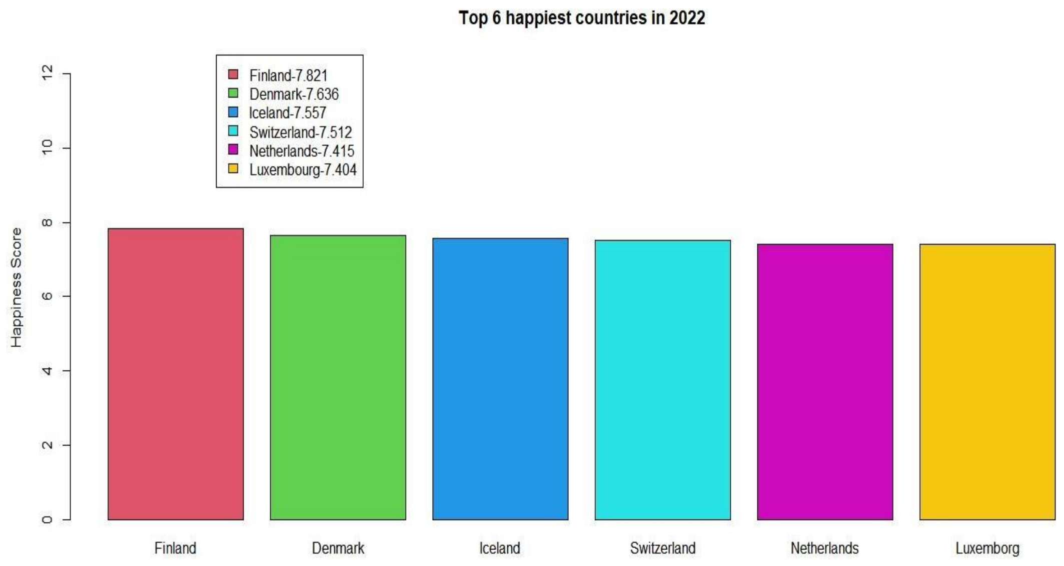
Year2020:



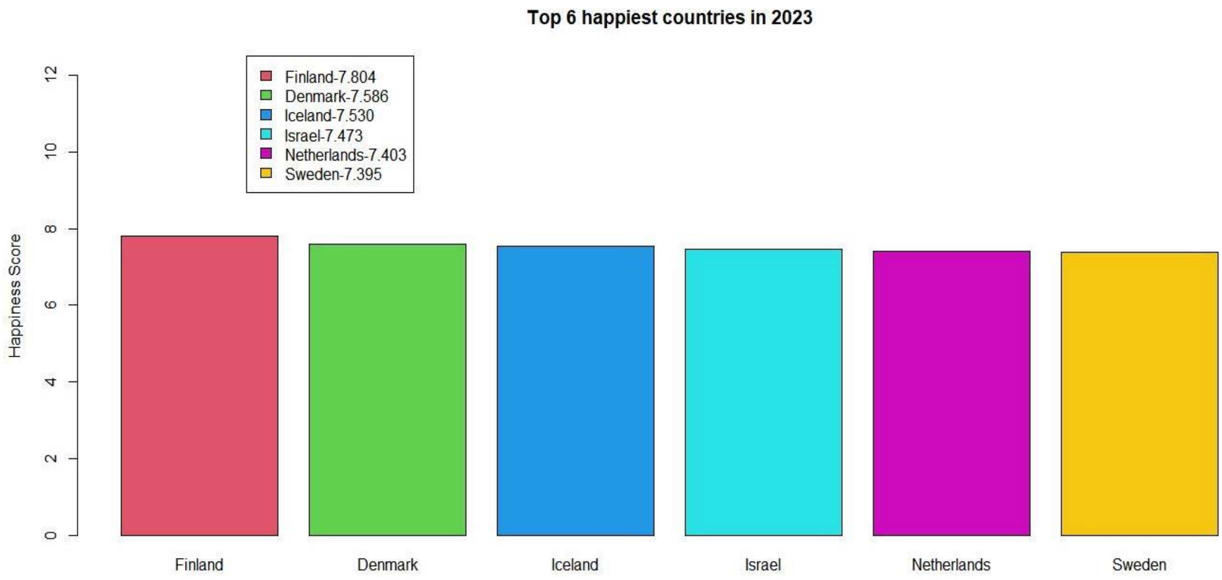
Year2021:



Year2022:



Year2023:

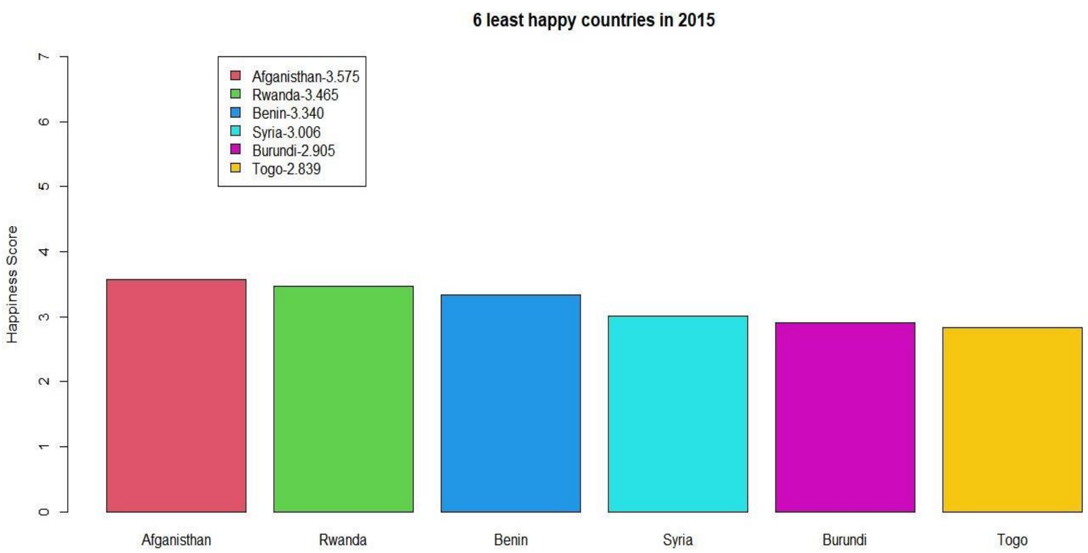


Interpretation:

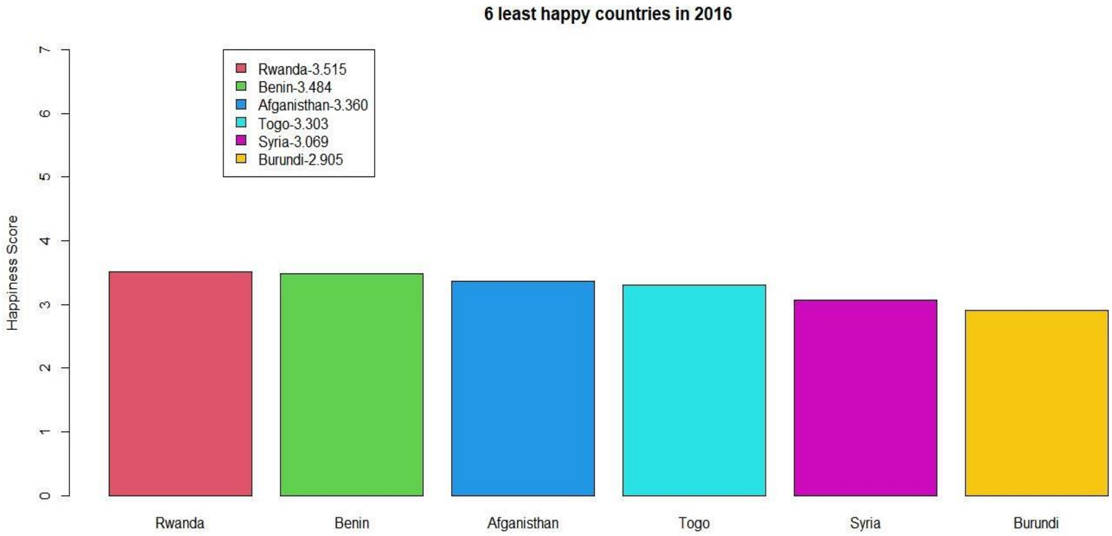
1. In2015and2016onlyCanadaasacountryofNorthAmericawaspresent,the remaining five were from Europe.
2. From2017to2022onlyEuropeancountriesremainedinthetop6positions and in 2023 Israel entered there as a member of Asia.
3. Also,Finland,Norway,Denmark,Iceland,Switzerland,Netherlands these six countrieswerealwaysinthetopsixpositions,onlyachangewasobserved when Luxembourg entered as a new European countryin 2022.
4. Finland holds the 1st position from 2018 to 2023,i.e. for 6 years continuously,andDenmarkholdsthe2nd positionfrom2019to2023,continuously.
5. All top countries for all years have scores above 7.Finland achieved the highestvalue7.842in2021,thenitfacedafallto7.804in2023.Butitmaintained the level 7.6 to 7.84,whereas before 2018(when Finland was not at the top),the top country has the score level around 7.5

Now,wegraphicallyrepresentthe6leasthappycountriesfortheconsecutive years.

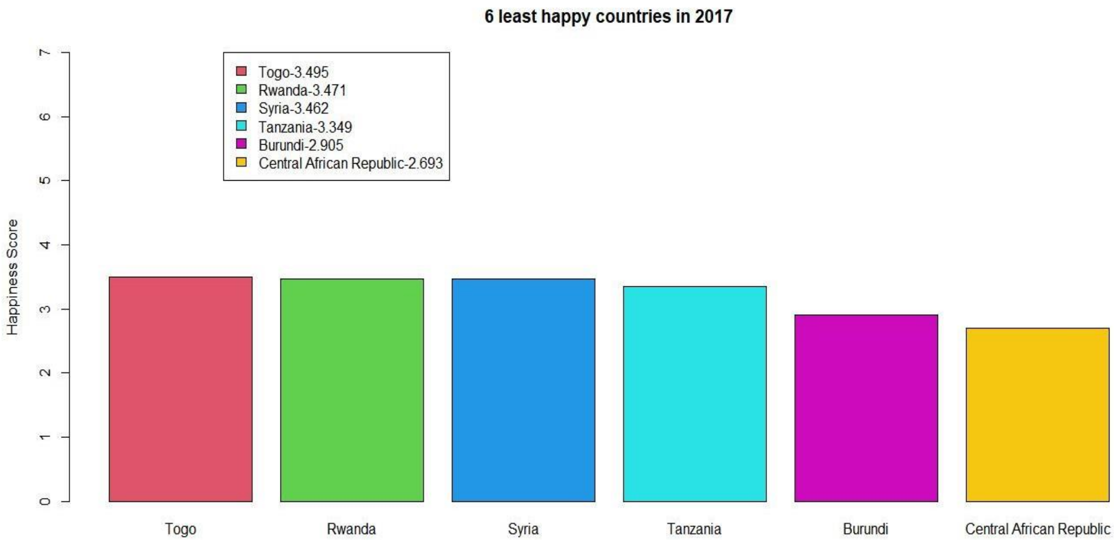
Year2015:



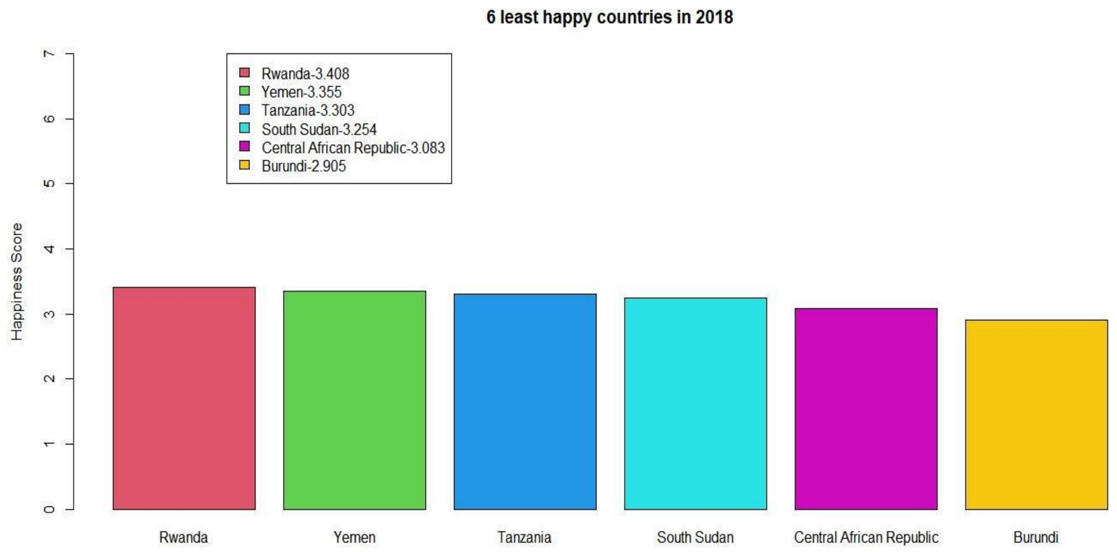
Year2016:



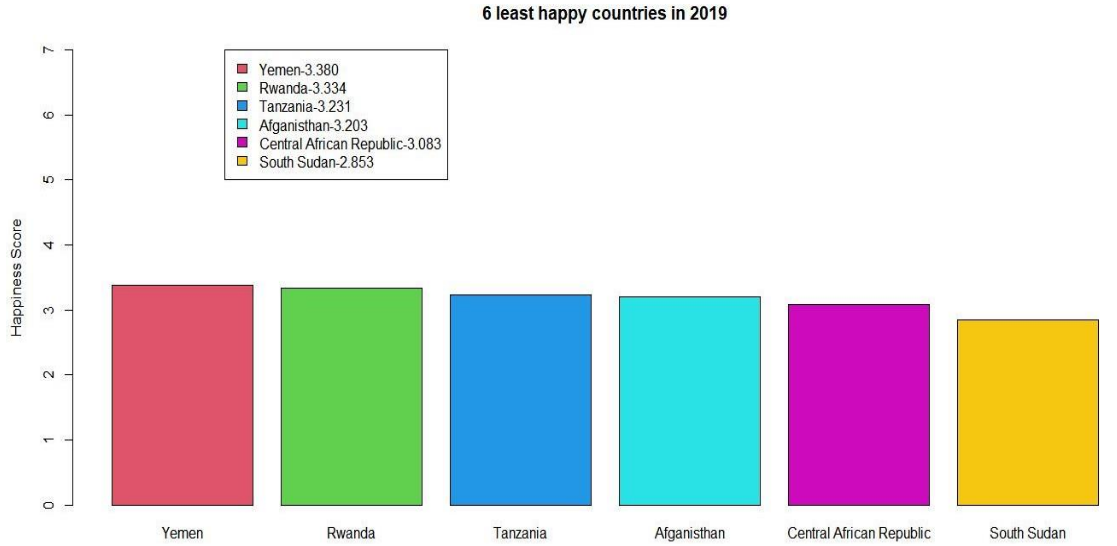
Year2017:



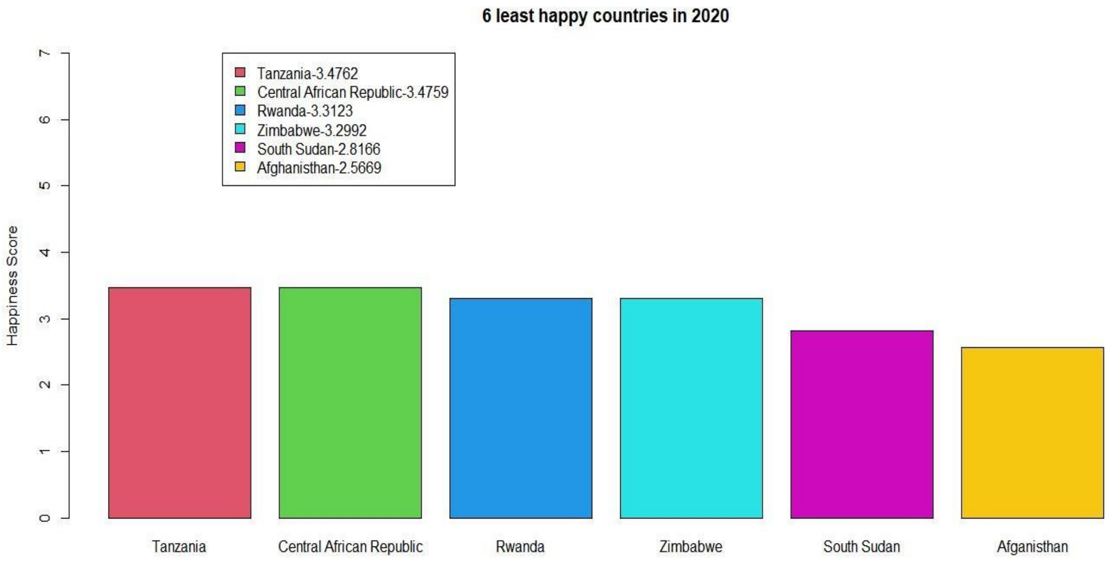
Year2018:



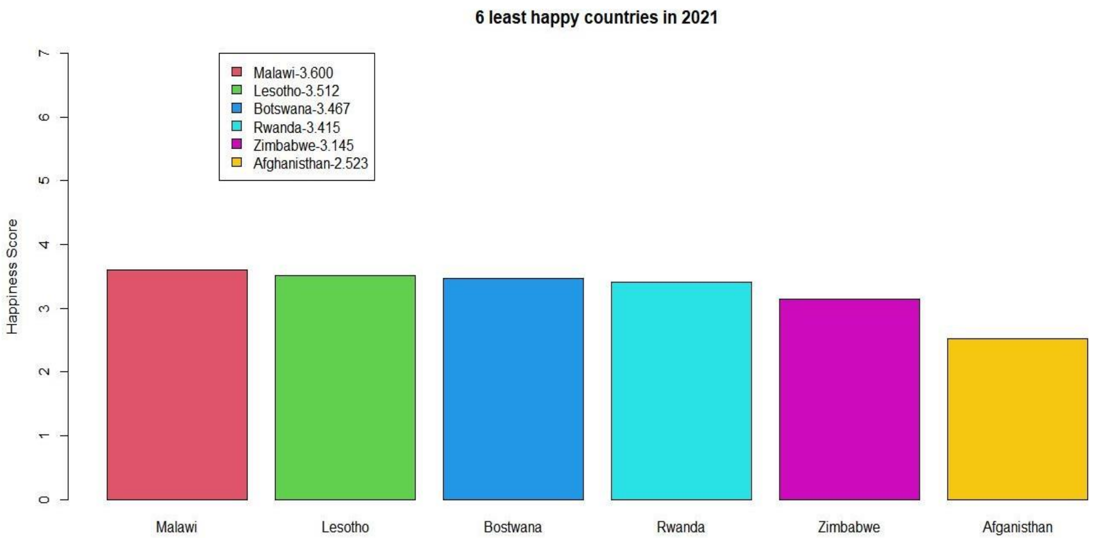
Year2019:



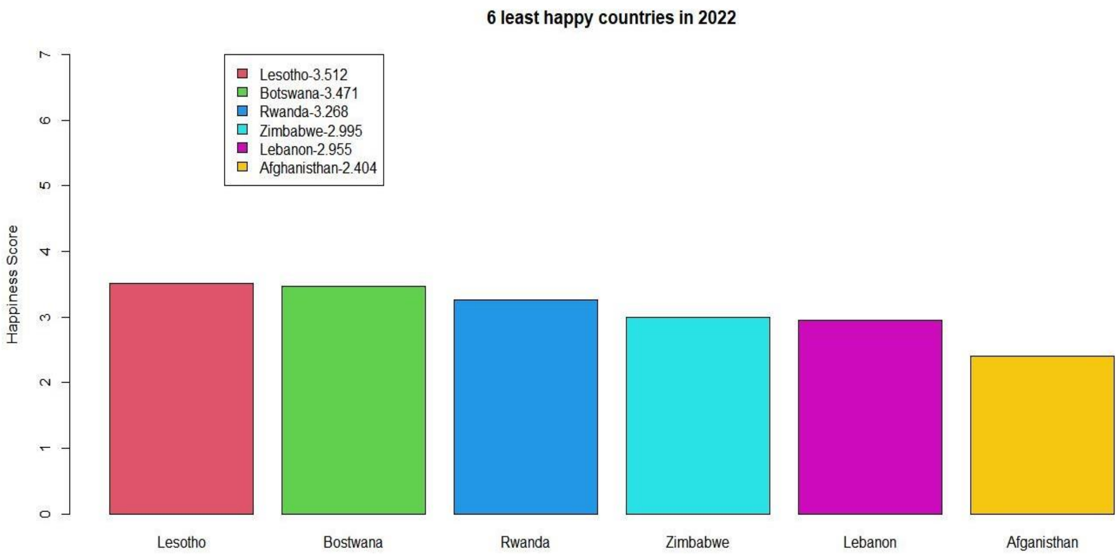
Year2020:



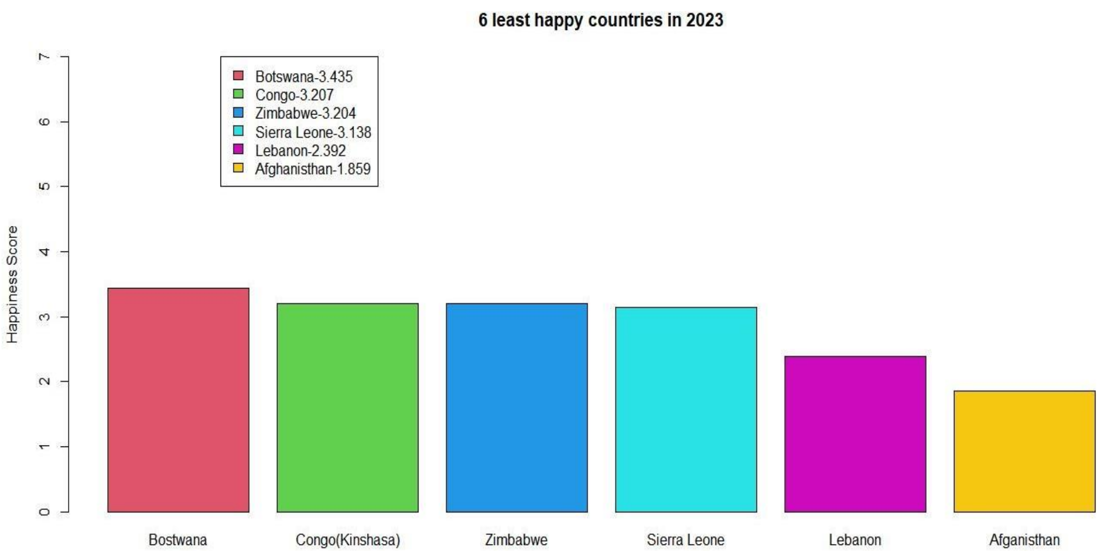
Year2021:



Year2022:



Year2023:



Interpretation:

1. ItisobservedthattheAfricanandAsiancountriesareinthelowestpositions for all the years.
2. AfghanistanandSyriawereinthe6leasthappycountriesinboth2015and 2016.Syria was only present in the list as an Asian country in 2017
3. In2018,noAsiancountrywasthereinthesixleasthappycountries' list.
4. From 2019 to 2023 only Afghanistan has been present in this list as from Asian continent.All other countries are from African territory.Mainly,Rwanda,Togo,Central African Republic,Burundi were present inthislistbefore2020,after2020,Malawi,Lesotho,Botswana,Lebanon,Zimbabwe took place in the list.
5. In 2023,thelowest happy country is Afghanisthan and it is the only country with happiness score 1.859,which is the lowest for any country in all these years. Itmightbebecauseofthegovernmentalchangeandthenewpoliciesadopted,the societal imbalance etc.

StepwiseRegression

In order tofind out which factors among the six factors viz, GDP per capita,social support,healthy life expectancy,freedom,generosity,perception of corruption actually influence the happiness score level,we will make use of stepwise regression.Our main purpose in this section is to check whether the samecovariatesareinfluentialfordifferentyears,i.e.wearedoingtheregression analysis for each year to check the presence of any change in the type of influential covariates.

Wedefinethevariablesasfollows-

x_1:Happinessscoreforanyrandomly selected country
x_2:Extend to which "GDP per capita" contributes to happiness score
x_3:Extend to which "social support" contributes to happiness score
x_4:Extendtowhich"healthylifeexpectancy"contributestohappinessscore
x_5:Extendtowhich"freedomtomakechoices"contributestohappinessscore x_6:Extend to which "generosity" contributes to happiness score
x_7:Extendtowhich"perceptionofcorruption"contributestohappinessscore All

these variables are measured in continuous scale.

Definethe samplepartialcorrelationcoefficientbetween$x_1$andanycovariatex_j, j=2(1)pisgivenby$r_{1j.23\ldots(j-1)(j+1)\ldots p}=(-1)^{1+1}$ $\sqrt{11}$,j=2(1)p.

HereRisthecorrelationmatrixoftherandom vector $\begin{bmatrix} 1 \\ \\ \\ \\ \end{bmatrix}$

AndR$_{ij}$isthecovariateofthe(i,j)thelementofR p is the no. of covariates ,here p=6

<u>Year2015</u>:

Firstwestartmodellingx$_1$ontheothercovariatesx$_2$to x$_7$,Afterfittingthemodel of multiple linear regression, we get the following result:

	Estimate	Std.Error	tvalue	Pr(>|t|)
(Intercept)	1.8602	0.1905	9.766	<2e-16***
x2	0.8607	0.2203	3.907	0.000141 ***
x3	1.4089	0.2227	6.327	2.69e-09***
x4	0.9753	0.3163	3.084	0.002433 **
x5	1.3334	0.3850	3.463	0.000694 ***
x6	0.3889	0.3910	0.995	0.321471
x7	0.7845	0.4365	1.797	0.074302 .

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.551 on 151 degrees of freedom
Multiple R-squared:0.7772,   AdjustedR-squared:0.7684
F-statistic: 87.81 on 6 and 151 DF,p-value: < 2.2e-16

Fromtheabovetableithasbeenclearthatx$_6$andx$_7$arethetwoinsignificant covariates.

Nowwegoforthepartialcorrelationcoefficientsandwillmarkthatcovariate insignificant first which will show minimum partial correlation with x$_1$in absolute terms and we will eliminate that variable immediately.Then we will strat afresh with the remaining variables.This process will go until there is a sudden drop in adjusted R-squared value which is here 0.7684.

Thepartialcorrelationcoefficientsaregivenbelow-

$r_{12.34567}$ **=0.3029689**
$r_{13.24567}$**=-0.4577754**
$r_{14.23567}$**=0.2433906**
$r_{15.23467}$**=-0.2712727**
$r_{16.23457}$**=0.08068409**
$r_{17.23456}$**=-0.1447148**

Itshowsthat$x_6$hasthelowestpartialcorrelationcoefficientwith$x_1$andhenceis eliminated as an predictor.

Aftereliminationofx_6,weagaindothemultipleregressionfor$x_1$onthe remaining predictors and thus we obtain the following table as-

	Estimate	Std.Error	tvalue	Pr(>|t|)
(Intercept)	1.8982	0.1866	10.173	<2e-16***
x2	0.8053	0.2132	3.778	0.000226 ***
x3	1.4164	0.2225	6.365	2.19e-09***
x4	1.0338	0.3108	3.327	0.001102 **
x5	1.4426	0.3690	3.909	0.000139 ***
x7	0.8540	0.4309	1.982	0.049281 *

--- Signif.codes:0 '*'0.001 '**'0.01 '*'0.05 '.'0.1 ''1**

Residual standard error: 0.551 on 152 degrees of freedom
Multiple R-squared:0.7758, AdjustedR-squared:0.7684
F-statistic: 105.2 on 5 and 152 DF,p-value: < 2.2e-16

Now$x_7$canbeseentobeinsignificantfromtheabovetable. Again

the partial correlation coefficients are obtained as :

$r_{12.3457}$**=0.2929837**
$r_{13.2457}$**=-0.4587258**
$r_{14.2357}$**=0.2605139**
$r_{15.2347}$**=-0.3022429**
$r_{17.2345}$**=0.1587246**

$x_7$hasminimumpartialcorrelationcoefficienttherewithx_1,hence$x_7$isdiscarded as a predictor.

Thetablecorrespondingtothemultipleregressionmodelof$x_1$onother remaining variables is given by-

	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	1.8963	0.1884	10.067	<2e-16***
x2	0.8826	0.2116	4.172	5.05e-05***
x3	1.3581	0.2227	6.099	8.37e-09***
x4	0.9959	0.3131	3.181	0.00178 **
x5	1.7707	0.3330	5.318	3.66e-07***

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residualstandarderror:0.5563on153degreesoffreedom
Multiple R-squared:0.77, Adjusted R-squared:0.764
F-statistic:128 on 4 and 153 DF,p-value: < 2.2e-16

Fromthetableitmightseemthatx4islesssignificant.Againwefindthepartial correlation coefficients as follows:-

$r_{12.345}$=0.3195656
$r_{13.245}$=-0.4422197
$r_{14.235}$=0.2490355
$r_{15.234}$=-0.3949868

$x_4$isdiscardedasanpredictor ,sinceithasthelowest partialcorrelationwithx_1

Thetablecorrespondingtothemultiplecorrelationcoefficientofx_1 onthe remaining covariates is given below-

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.0758	0.1850	11.223	<2e-16	***
x2	1.3717	0.1495	9.173	2.82e-16	***
x3	1.3422	0.2291	5.858	2.74e-08	***
x5	1.8875	0.3406	5.542	1.27e-07	***

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residualstandarderror:0.5725on154degreesoffreedom
Multiple R-squared:0.7548, Adjusted R-squared:0.75
F-statistic:158 on 3 and 154 DF,p-value: < 2.2e-16

It'sseenthatactuallynovariableisinsignificanthere,weagaincalculatethe partial correlation coefficients

$r_{12.35}$=0.5944385
$r_{13.25}$=-0.4268822
$r_{15.23}$=0.4077655

$x_5$iseliminated becauseof same reason,andfinallyifwemodel $x_1$on$x_2$and$x_3$then the summary table is obtained as-

	Estimate	Std. Error	tvalue	Pr(>|t|)	
(Intercept)	2.4383	0.1889	12.909	<2e-16	***
x2	1.4750	0.1620	9.107	4.03e-16	***
x3	1.7047	0.2397	7.111	3.97e-11	***

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.625 on 155 degrees of freedom
Multiple R-squared:0.7059, AdjustedR-squared:0.7021
F-statistic:186 on 2 and 155 DF,p-value: < 2.2e-16

Note thatthere is a great fallinadjusted R-squared ,hencewestop thestepwise regression here.It should be noted that $x_5$is an important predictor.One point is here we want more precision,so we do not want the adjusted R-squared to fall upto 0.75 also,i.e. we want shift inthe variation in adjusted R -squared strictly less than 2.3945%,w.r.t the adjusted R -squared in the full model (if it touches the value 0.75 as the adjusted R squared ,the regression will be stopped by us).This much precision istaken into account because $x_4$whichis the healthy life expectancyisthoughttobeanimportantpredictorforhappiness,soitshouldnot be excluded.

Assuch,here$x_4$isalsoincludedasanpredictor,thusx_2,x_3,x_4,$x_5$arethefourfinal predictors in predicting x_1.

Year2016:

Firstwestartmodelling$x_1$ontheothercovariates$x_2$to x_7,Afterfittingthemodel of multiple linear regression, we get the following result:

	Estimate	Std.Error	tvalue	Pr(>|t|)
(Intercept)	2.1903	0.1583	13.834	<2e-16***
x2	0.7214	0.2171	3.323	0.001120 **
x3	1.2298	0.2297	5.353	3.19e-07***
x4	1.4364	0.3489	4.117	6.32e-05***
x5	1.5139	0.3880	3.902	0.000144 ***
x6	0.1595	0.3621	0.440	0.660236
x7	0.9189	0.4648	1.977	0.049852 *

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residualstandarderror:0.5367on150degreesoffreedom
MultipleR-squared: 0.7875,AdjustedR-squared: 0.779F-statistic: 92.65 on 6 and 150 DF,p-value: < 2.2e-16

Fromtheabovetableithasbeenclearthat$x_6$and$x_7$arethetwoinsignificant covariates.

Nowwegoforthepartialcorrelationcoefficientsandwillmarkthatcovariate insignificant first which will show minimum partial correlation with $x_1$in absolute terms and we will eliminate that variable immediately.Then we will strat afresh with the remaining variables.This process will go until there is a sudden drop in adjusted R-squared value which is here 0.779.

Thepartialcorrelationcoefficientsaregivenbelow-

$r_{12.34567}$=-0.2618273
$r_{13.24567}$=-0.4004625
$r_{14.23567}$=-0.3186175
$r_{15.23467}$=-0.3035796
$r_{16.23457}$=-0.03594039
$r_{17.23456}$=-0.1593744

Itshowsthat$x_6$hasthelowestpartialcorrelationcoefficientwith$x_1$andhenceis eliminated as an predictor.

Aftereliminationofx_6,weagaindothemultipleregressionfor$x_1$onthe remaining predictors and thus we obtain the following table as-

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.2119	0.1502	14.731	<2e-16	***
x2	0.6971	0.2094	3.329	0.0011	**
x3	1.2344	0.2289	5.393	2.62e-07	***
x4	1.4623	0.3430	4.263	3.53e-05	***
x5	1.5588	0.3733	4.175	5.01e-05	***
x7	0.9590	0.4546	2.110	0.0365	*

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.5353 on 151 degrees of freedom
Multiple R-squared:0.7872, AdjustedR-squared:0.7802
F-statistic: 111.7 on 5 and 151 DF,p-value: < 2.2e-16

Now$x_7$canbeseentobeinsignificantfromtheabovetable. Again

the partial correlation coefficients are obtained as :

$r_{12.3457}$=-0.2614497
$r_{13.2457}$=-0.4018612
$r_{14.2357}$=-0.3277867
$r_{15.2347}$=-0.3217145
$r_{17.2345}$=-0.1692123

$x_7$hasminimumpartialcorrelationcoefficienttherewithx_1,hence$x_7$isdiscarded as a predictor.

Thetablecorrespondingtothemultipleregressionmodelof$x_1$onother remaining variables is given by-

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.2068	0.1518	14.535	<2e-16	***
x2	0.7615	0.2095	3.634	0.000381	***
x3	1.1730	0.2296	5.109	9.59e-07	***
x4	1.4451	0.3468	4.167	5.15e-05	***
x5	1.9200	0.3355	5.723	5.41e-08	***

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residualstandarderror: 0.5413 on 152 degrees offreedom
MultipleR-squared: 0.781,AdjustedR-squared: 0.7752F-statistic: 135.5 on 4 and 152 DF,p-value: < 2.2e-16

Fromthetableitmightseemthatnopredictorislesssignificant.Toverifyagain we find the partial correlation coefficients as follows:-

$r_{12.345}$=-0.2827324
$r_{13.245}$=-0.382809
$r_{14.235}$=-0.3202213
$r_{15.234}$=-0.4210281

$x_2$isdiscardedasanpredictor ,sinceithasthelowest partialcorrelationwithx_1

Thetablecorrespondingtothemultiplecorrelationcoefficientofx_1 onthe remaining covariates is given below-

	Estimate	Std.Error	tvalue	Pr(>|t|)
(Intercept)	2.1582	0.1572	13.733	<2e-16***
x3	1.4867	0.2211	6.725	3.29e-10***
x4	2.3717	0.2442	9.712	<2e-16***
x5	1.9451	0.3486	5.580	1.06e-07***

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.5625 on 153 degrees of freedom
Multiple R-squared:0.7619, AdjustedR-squared:0.7573
F-statistic: 163.2 on 3 and 153 DF,p-value: < 2.2e-16

Note thatthere is a great fall in adjusted R -squared(more than 2.395 % variation in the full model adjusted R-squared) ,hence we stop the stepwise regressionhere.Itshouldbenotedthat$x_2$isanimportantpredictorandshould not be eliminated.

Soherein2016also,forpredictingx_1,x_2,x_3,x_4,$x_5$aretheimportantpredictors. Year

2017:

Firstwestartmodelling$x_1$ontheothercovariates$x_2$to x_7,Afterfittingthemodel of multiple linear regression, we get the following result:

	Estimate	Std.Error	tvalue	Pr(>|t|)
(Intercept)	1.7430	0.1874	9.303	<2e-16***
x2	0.7844	0.2045	3.836	0.000185 ***
x3	1.1178	0.2021	5.532	1.40e-07***
x4	1.2889	0.3215	4.009	9.65e-05***
x5	1.4757	0.3425	4.309	2.98e-05***
x6	0.3807	0.3293	1.156	0.249524
x7	0.8266	0.4843	1.707	0.089975 .

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.4998 on 148 degrees of freedom
Multiple R-squared:0.8124, AdjustedR-squared:0.8048
F-statistic: 106.8 on 6 and 148 DF,p-value: < 2.2e-16

Fromtheabovetableithasbeenclearthat$x_6$and$x_7$arethetwoinsignificant covariates.

Nowwegoforthepartialcorrelationcoefficientsandwillmarkthatcovariate insignificant first which will show minimum partial correlation with $x_1$in absolute terms and we will eliminate that variable immediately.Then we will strat afresh with the remaining variables.This process will go until there is a sudden drop in adjusted R-squared value which is here 0.8048.

Thepartialcorrelationcoefficientsaregivenbelow-

$r_{12.34567}$ **=-0.3006944**
$r_{13.24567}$**=-0.4139314**
$r_{14.23567}$**=-0.3129562**
$r_{15.23467}$**=-0.3338415**
$r_{16.23457}$**=-0.09460045**
$r_{17.23456}$**=-0.1389294**

Itshowsthat$x_6$hasthelowestpartialcorrelationcoefficientwith$x_1$andhenceis eliminated as an predictor.

Aftereliminationofx_6,weagaindothemultipleregressionfor$x_1$onthe remaining predictors and thus we obtain the following table as-

	Estimate	Std.Error	tvalue	Pr(>|t|)
(Intercept)	1.8018	0.1805	9.980	<2e-16***
x2	0.7320	0.1997	3.667	0.000341 ***
x3	1.1231	0.2022	5.554	1.25e-07***
x4	1.3431	0.3184	4.218	4.26e-05***
x5	1.5627	0.3345	4.671	6.63e-06***
x7	0.9482	0.4733	2.003	0.046949 *

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.5003 on 149 degrees of freedom
Multiple R-squared:0.8107, AdjustedR-squared:0.8044
F-statistic: 127.6 on 5 and 149 DF,p-value: < 2.2e-16

Now x_7 can be seen to be insignificant from the above table. Again

the partial correlation coefficients are obtained as :

$r_{12.3457}$=-0.287679
$r_{13.2457}$=-0.4141251
$r_{14.2357}$=-0.3265893
$r_{15.2347}$=-0.3574213
$r_{17.2345}$=-0.161957

x_7 has minimum partial correlation coefficient there with x_1, hence x_7 is discarded as a predictor.

The table corresponding to the multiple regression model of x_1 on other remaining variables is given by-

	Estimate	Std.Error	tvalue	Pr(>|t|)
(Intercept)	1.7992	0.1823	9.867	<2e-16***
x2	0.8179	0.1970	4.153	5.50e-05***
x3	1.0685	0.2024	5.279	4.47e-07***
x4	1.3029	0.3210	4.059	7.90e-05***
x5	1.8612	0.3025	6.152	6.62e-09***

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.5053 on 150 degrees of freedom
Multiple R-squared:0.8056, AdjustedR-squared:0.8004
F-statistic: 155.4 on 4 and 150 DF,p-value: < 2.2e-16

From the table it might seem that no predictor is less significant. Again we find the partial correlation coefficients as follows:-

$r_{12.345}$=-0.3210976
$r_{13.245}$=-0.3958496
$r_{14.235}$=-0.3146008
$r_{15.234}$=-0.4488852

$x_4$isdiscardedasanpredictor ,sinceithasthelowest partialcorrelationwithx_1

Thetablecorrespondingtothemultiplecorrelationcoefficientofx_1 onthe remaining covariates is given below-

	Estimate	Std.Error	tvalue	Pr(>|t|)
(Intercept)	1.8499	0.1910	9.685	<2e-16***
x2	1.4025	0.1411	9.943	<2e-16***
x3	1.1215	0.2121	5.289	4.25e-07***
x5	1.9318	0.3171	6.092	8.87e-09***

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residualstandarderror:0.5306on151degreesoffreedom
Multiple R-squared:0.7843, Adjusted R-squared:0.78
F-statistic:183 on 3 and 151 DF,p-value: < 2.2e-16

NotethatthereisagreatfallinadjustedR-squared(morethan2.3945%),hence we stop the stepwise regression here.It should be noted that $x_4$is an important predictor.

Sofor2017,forpredicting x_1, x_2,x_3,x_4,$x_5$aretheimportantpredictors also.

<u>Year2018</u>:

Firstwestartmodelling$x_1$ontheothercovariates$x_2$to x_7,Afterfittingthemodel of multiple linear regression, we get the following result:

	Estimate	Std.Error	tvalue	Pr(>|t|)
(Intercept)	1.8856	0.1955	9.647	<2e-16***
x2	1.1114	0.2094	5.307	3.96e-07***
x3	0.9991	0.2015	4.958	1.91e-06***
x4	0.8085	0.3314	2.440	0.0159 *
x5	1.3831	0.3204	4.317	2.87e-05***
x6	0.6056	0.4721	1.283	0.2016
x7	0.5951	0.5251	1.133	0.2589

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.5249 on 149 degrees of freedom
Multiple R-squared:0.7886, AdjustedR-squared:0.7801
F-statistic: 92.66 on 6 and 149 DF,p-value: < 2.2e-16

Fromtheabovetableithasbeenclearthat$x_6$and$x_7$arethetwoinsignificant covariates.

Nowwegoforthepartialcorrelationcoefficientsandwillmarkthatcovariate insignificant first which will show minimum partial correlation with $x_1$in absolute terms and we will eliminate that variable immediately.Then we will strat afresh with the remaining variables.This process will go until there is a sudden drop in adjusted R-squared value which is here 0.7801.

Thepartialcorrelationcoefficientsaregivenbelow-

$r_{12.34567}$ **=-0.3987322**
$r_{13.24567}$**=-0.376326**
$r_{14.23567}$**=-0.1959947**
$r_{15.23467}$**=-0.3338415**
$r_{16.23457}$**=-0.10451**
$r_{17.23456}$**=-0.09243986**

Itshowsthat$x_7$hasthelowestpartialcorrelationcoefficientwith$x_1$andhenceis eliminated as an predictor.

Aftereliminationofx_7,weagaindothemultipleregressionfor$x_1$onthe remaining predictors and thus we obtain the following table as-

	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	1.8559	0.1939	9.573	<2e-16***
x2	1.1261	0.2092	5.383	2.77e-07***
x3	0.9813	0.2011	4.880	2.68e-06***
x4	0.8468	0.3300	2.566	0.0113 *
x5	1.5020	0.3030	4.957	1.92e-06***
x6	0.7572	0.4532	1.671	0.0969 .

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.5254 on 150 degrees of freedom
Multiple R-squared:0.7868, AdjustedR-squared:0.7797
F-statistic: 110.7 on 5 and 150 DF,p-value: < 2.2e-16

Now$x_6$canbeseentobeinsignificantfromtheabovetable. Again

the partial correlation coefficients are obtained as :

$r_{12.3456}$=-0.4023517
$r_{13.2456}$=-0.3701625
$r_{14.2356}$=-0.2050788
$r_{15.2346}$=-0.3751421
$r_{16.2345}$=-0.135161

$x_6$hasminimumpartialcorrelationcoefficientherewithx_1,hence$x_6$isdiscarded as a predictor.

Thetablecorrespondingtothemultipleregressionmodelof$x_1$onother remaining variables is given by-

	Estimate	Std. Error	tvalue	Pr(>|t|)
(Intercept)	1.9547	0.1857	10.524	<2e-16***
x2	1.1054	0.2101	5.262	4.81e-07***
x3	0.9613	0.2019	4.761	4.48e-06***
x4	0.8593	0.3319	2.589	0.0106 *
x5	1.6639	0.2888	5.762	4.53e-08***

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.5285 on 151 degrees of freedom
Multiple R-squared:0.7829, AdjustedR-squared:0.7771
F-statistic: 136.1 on 4 and 151 DF,p-value: < 2.2e-16

Fromthetableitmightseemthatnopredictorislesssignificant.Againwefind the partial correlation coefficients as follows:-

$r_{12.345}$=-0.3936361
$r_{13.245}$=-0.361263
$r_{14.235}$=-0.2062009
$r_{15.234}$=-0.4245274

$x_4$isdiscardedasanpredictor ,sinceithasthelowest partialcorrelationwithx_1

Thetablecorrespondingtothemultiplecorrelationcoefficientofx_1 onthe remaining covariates is given below-

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.9507	0.1892	10.311	<2e-16***
x2	1.4945	0.1495	9.995	<2e-16***
x3	1.0775	0.2005	5.374	2.84e-07***
x5	1.7286	0.2931	5.898	2.29e-08***

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.5384 on 152 degrees of freedom
Multiple R-squared:0.7732, AdjustedR-squared:0.7687
F-statistic: 172.7 on 3 and 152 DF,p-value: < 2.2e-16

Itseemsthatthereisnoinsignificantpredictor.Toverifywecancalculatethe partial correlation coefficients -

$r_{12.35}$=-0.6297524
$r_{13.25}$=-0.39956
$r_{15.23}$=-0.4315736

$x_3$isdiscardednow,andafter

elimination,when$x_1$ismodelledon$x_2$andx_5

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.6195	0.1549	16.908	<2e-16***
x2	1.9949	0.1272	15.684	<2e-16***
x5	2.1520	0.3069	7.012	7.08e-11 ***

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.5854 on 153 degrees of freedom
Multiple R-squared:0.7301, AdjustedR-squared:0.7266
F-statistic:207 on 2 and 153 DF,p-value: < 2.2e-16

NotethatthereisagreatfallinadjustedR-squared(morethan2.3945%),hence we stop the stepwise regression here.It should be noted that $x_3$is an important predictor,but $x_4$hasto be sacrificed as an predictor.Remember that if we didnot allow the variation $x_4$was already discarded in 2015,but for our precision it remained as a predictor,but in this year it is finally eliminated.

Sofor2018,forpredictingx_1,x_2,x_3,$x_5$aretheimportantpredictors also. **<u>Year</u>**

<u>2019</u>:

Firstwestartmodelling$x_1$ontheothercovariates$x_2$to x_7,Afterfittingthemodel of multiple linear regression, we get the following result:

	Estimate	Std. Error	tvalue	Pr(>|t|)
(Intercept)	1.7952	0.2111	8.505	1.77e-14***
x2	0.7754	0.2182	3.553	0.000510 ***
x3	1.1242	0.2369	4.745	4.83e-06***
x4	1.0781	0.3345	3.223	0.001560 **
x5	1.4548	0.3753	3.876	0.000159 ***

x6	**0.4898**	**0.4977**	**0.984**	**0.326709**
x7	**0.9723**	**0.5424**	**1.793**	**0.075053 .**

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.5335 on 149 degrees of freedom
Multiple R-squared:0.7792, AdjustedR-squared:0.7703
F-statistic: 87.62 on 6 and 149 DF,p-value: < 2.2e-16

Fromtheabovetableithasbeenclearthat$x_6$and$x_7$arethetwoinsignificant covariates.

Nowwegoforthepartialcorrelationcoefficientsandwillmarkthatcovariate insignificant first which will show minimum partial correlation with $x_1$in absolute terms and we will eliminate that variable immediately.Then we will strat afresh with the remaining variables.This process will go until there is a certain drop in adjusted R-squared value which is here 0.7703.

Thepartialcorrelationcoefficientsaregivenbelow-

$r_{12.34567}$ **=-0.2794804**
$r_{13.24567}$**=-0.3623422**
$r_{14.23567}$**=-0.2552729**
$r_{15.23467}$**=-0.3026475**
$r_{16.23457}$**=-0.08035206**
$r_{17.23456}$**=-0.1453036**

Itshowsthat$x_6$hasthelowestpartialcorrelationcoefficientwith$x_1$andhenceis eliminated as an predictor.

Aftereliminationofx_6,weagaindothemultipleregressionfor$x_1$onthe remaining predictors and thus we obtain the following table as-

	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	1.8689	0.1973	9.471	<2e-16***
x2	0.7455	0.2161	3.450	0.000728 ***
x3	1.1180	0.2368	4.722	5.33e-06***
x4	1.0840	0.3344	3.241	0.001467 **
x5	1.5340	0.3666	4.185	4.84e-05***
x7	1.1176	0.5218	2.142	0.033839 *

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.5335 on 150 degrees of freedom
Multiple R-squared:0.7777, AdjustedR-squared:0.7703
F-statistic:105 on 5 and 150 DF,p-value: < 2.2e-16

Now$x_7$can beseento beinsignificantfromtheabovetable.

Againthepartialcorrelationcoefficientsareobtainedas:

$r_{12.3457}$=-0.2711384
$r_{13.2457}$=-0.3597099
$r_{14.2357}$=-0.255835
$r_{15.2347}$=-0.3233292
$r_{17.2345}$=-0.1722492

$x_7$hasminimumpartialcorrelationcoefficientherewithx_1,hence$x_7$isdiscarded as a predictor.

Thetablecorrespondingtothemultipleregressionmodelof$x_1$onother remaining variables is given by-

	Estimate	Std.Error	tvalue	Pr(>|t|)
(Intercept)	1.8921	0.1994	9.491	<2e-16***
x2	0.8105	0.2165	3.745	0.000256 ***
x3	1.0166	0.2347	4.331	2.70e-05***
x4	1.1414	0.3373	3.384	0.000910 ***
x5	1.8458	0.3404	5.423	2.28e-07***

Signif.codes:0 ‘*’0.001 ‘**’0.01 ‘*’0.05‘.’0.1‘’1**

Residual standard error: 0.5398 on 151 degrees of freedom
Multiple R-squared:0.7709, AdjustedR-squared:0.7649
G-statistic:127 on 4 and 151 DF,p-value: < 2.2e-16

Fromthetableitmightseemthatnopredictorislesssignificant.Againwefind the partial correlation coefficients as follows:-

$r_{12.345}$=-0.2915062
$r_{13.245}$=-0.332386
$r_{14.235}$=-0.2655015
$r_{15.234}$=-0.4037249

$x_4$isdiscardedasanpredictor ,sinceithasthelowest partialcorrelationwithx_1

Thetablecorrespondingtothemultiplecorrelationcoefficientofx_1 onthe remaining covariates is given below-

	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	2.0492	0.2004	10.223	<2e-16***
x2	1.2792	0.1720	7.438	6.98e-12***
x3	1.1886	0.2369	5.017	1.45e-06***
x5	1.9442	0.3506	5.545	1.27e-07***

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.558 on 152 degrees of freedom
Multiple R-squared:0.7536, AdjustedR-squared:0.7487
F-statistic: 154.9 on 3 and 152 DF,p-value: < 2.2e-16

NotethatthereisagreatfallinadjustedR-squared(morethan2.3945%),hence we stop the stepwise regression here.It should be noted that $x_4$is an important predictor here as per our precision maintained.

Sofor2019,forpredictingx_1,x_2,x_3,x_4,$x_5$aretheimportantpredictorsalso.So$x_4$has again come back as an predictor.

<u>Year2020</u>:

Firstwestartmodelling$x_1$ontheothercovariates$x_2$to x_7,Afterfittingthemodel of multiple linear regression, we get the following result:

	Estimate	Std.Error	tvalue	Pr(>|t|)	
(Intercept)	1.8872	0.2272	8.307	6.14e-14	***
x2	0.7391	0.2648	2.791	0.005960	**
x3	1.1530	0.2799	4.119	6.35e-05	***
x4	0.9807	0.3604	2.721	0.007293	**
x5	1.4825	0.4151	3.571	0.000481	***
x6	0.6208	0.5096	1.218	0.225126	
x7	0.9729	0.4876	1.995	0.047857	*

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residualstandarderror:0.5693on146degreesoffreedom
MultipleR-squared: 0.7483,AdjustedR-squared: 0.738F-
statistic: 72.36 on 6 and 146 DF,p-value: < 2.2e-16

Fromtheabovetableithasbeenclearthat$x_6$and$x_7$arethetwoinsignificant covariates.

Nowwegoforthepartialcorrelationcoefficientsandwillmarkthatcovariate insignificant first which will show minimum partial correlation with $x_1$in absolute terms and we will eliminate that variable immediately.Then we will strat afresh with the remaining variables.This process will go until there is a sudden drop in adjusted R-squared value which is here 0.738.

Thepartialcorrelationcoefficientsaregivenbelow-

$r_{12.34567}$=-0.2250457
$r_{13.24567}$=-0.3226518
$r_{14.23567}$=-0.219717
$r_{15.23467}$=-0.2834439
$r_{16.23457}$=-0.1003077
$r_{17.23456}$=-0.1629358

Itshowsthat$x_6$hasthelowestpartialcorrelationcoefficientwith$x_1$andhenceis eliminated as an predictor.

Aftereliminationofx_6,weagaindothemultipleregressionfor$x_1$onthe remaining predictors and thus we obtain the following table as-

	Estimate	Std. Error	tvalue	Pr(>|t|)
(Intercept)	1.9743	0.2160	9.140	4.66e-16***
x2	0.6895	0.2621	2.631	0.009434 **
x3	1.1609	0.2803	4.141	5.80e-05***
x4	0.9637	0.3607	2.672	0.008394 **
x5	1.6036	0.4037	3.972	0.000111 ***
x7	1.1269	0.4717	2.389	0.018165 *

Signif.codes:0 ‘*’0.001 ‘**’0.01 ‘*’0.05‘.’0.1‘’1**

Residual standard error: 0.5703 on 147 degrees of freedom
Multiple R-squared:0.7458, AdjustedR-squared:0.7371
F-statistic: 86.25 on 5 and 147 DF,p-value: < 2.2e-16

Now$x_7$canbeseentobeinsignificantfromtheabovetable. Again

the partial correlation coefficients are obtained as :

$r_{12.3457}$=-0.2120275
$r_{13.2457}$=-0.3232333
$r_{14.2357}$=-0.2152091
$r_{15.2347}$=-0.3113463
$r_{17.2345}$=-0.193321

$x_7$hasminimumpartialcorrelationcoefficienttherewithx_1,hence$x_7$isdiscarded as a predictor.

Thetablecorrespondingtothemultipleregressionmodelof$x_1$onother remaining variables is given by-

	Estimate	Std. Error	tvalue	Pr(>|t|)
(Intercept)	1.9818	0.2194	9.034	8.36e-16***
x2	0.7737	0.2638	2.933	0.00390 **
x3	1.0182	0.2782	3.660	0.00035 ***
x4	1.0720	0.3635	2.949	0.00371 **
x5	1.9412	0.3841	5.054	1.26e-06***

Signif.codes:0 ‘*’0.001 ‘**’0.01 ‘*’0.05‘.’0.1‘’1**

Residual standard error: 0.5793 on 148 degrees of freedom
Multiple R-squared:0.7359, AdjustedR-squared:0.7288
F-statistic: 103.1 on 4 and 148 DF,p-value: < 2.2e-16

Fromthetableitmightseemthatx_4 and$x_2$mightbelesssignificant.Againwe find the partial correlation coefficients as follows:-

$r_{12.345}$=-0.234342
$r_{13.245}$=-0.2880855
$r_{14.235}$=-0.2355906
$r_{15.234}$=-0.3836386

$x_2$isdiscardedasanpredictor ,sinceithasthelowest partialcorrelationwithx_1

Thetablecorrespondingtothemultiplecorrelationcoefficientofx_1 onthe remaining covariates is given below-

	Estimate	Std.Error	tvalue	Pr(>|t|)
(Intercept)	1.7981	0.2155	8.342	4.54e-14***
x3	1.3626	0.2585	5.271	4.69e-07***
x4	1.7527	0.2867	6.113	8.19e-09***
x5	1.9113	0.3936	4.855	3.00e-06***

Signif.codes:0 ‘*’0.001 ‘**’0.01 ‘*’0.05‘.’0.1‘’1**

Residual standard error: 0.5939 on 149 degrees of freedom
Multiple R-squared:0.7206, AdjustedR-squared:0.7149
F-statistic: 128.1 on 3 and 149 DF,p-value: < 2.2e-16

Itseems thatthereisnoinsignificant predictor.

NotethatthereisafallinadjustedR-squared(morethan2.3945%),hencewe stop the stepwise regression here.It should be noted that $x_2$is an important predictor.

Sofor2020 , forpredicting x_1,x_2,x_3,x_4,$x_5$aretheimportantpredictors also.

Year2021:

Firstwestartmodelling$x_1$ontheothercovariates$x_2$to x_7,Afterfittingthemodel of multiple linear regression, we get the following result:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.3264	0.1954	11.905	<2e-16***
x2	0.8002	0.2485	3.220	0.001591 **
x3	1.0983	0.2964	3.705	0.000302 ***
x4	0.9637	0.4232	2.277	0.024279 *
x5	1.6521	0.4066	4.063	7.97e-05***
x6	0.5615	0.4925	1.140	0.256133
x7	0.9477	0.4551	2.082	0.039118 *

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.5417 on 142 degrees of freedom
Multiple R-squared:0.7559, AdjustedR-squared:0.7456
F-statistic: 73.29 on 6 and 142 DF,p-value: < 2.2e-16

Fromtheabovetableithasbeenclearthat$x_6$and$x_7$arethetwoinsignificant covariates.

Nowwegoforthepartialcorrelationcoefficientsandwillmarkthatcovariate insignificant first which will show minimum partial correlation with $x_1$in absolute terms and we will eliminate that variable immediately.Then we will strat afresh with the remaining variables.This process will go until there is a sudden drop in adjusted R-squared value which is here 0.7456.

Thepartialcorrelationcoefficientsaregivenbelow-

$r_{12.34567}$ =-0.2608388
$r_{13.24567}$=-0.2969306
$r_{14.23567}$=-0.1876881
$r_{15.23467}$=-0.3227484
$r_{16.23457}$=-0.09524633
$r_{17.23456}$=-0.1721272

Itshowsthat$x_6$hasthelowestpartialcorrelationcoefficientwith$x_1$andhenceis eliminated as an predictor.

Aftereliminationofx_6,weagaindothemultipleregressionfor$x_1$onthe

remainingpredictorsandthusweobtainthefollowingtableas-

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.4114	0.1808	13.334	<2e-16	***
x2	0.7554	0.2457	3.075	0.002522	**
x3	1.1119	0.2965	3.751	0.000256	***
x4	0.9334	0.4228	2.207	0.028875	*
x5	1.7531	0.3972	4.413	1.99e-05	***
x7	1.0466	0.4473	2.340	0.020668	*

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.5423 on 143 degrees of freedom
Multiple R-squared:0.7537, AdjustedR-squared:0.7451
F-statistic:87.5 on 5 and 143 DF,p-value: < 2.2e-16

Now$x_7$and$x_4$canbeseento beinsignificantfromtheabovetable. Again the

partial correlation coefficients are obtained as :

$r_{12.3457}$**=-0.2490419**
$r_{13.2457}$**=-0.2992684**
$r_{14.2357}$**=-0.1815306**
$r_{15.2347}$**=-0.3462221**
$r_{17.2345}$**=-0.1920374**

x_4**hasminimumpartialcorrelationcoefficientherewith**x_1**,hence**x_4**isdiscarded as a predictor.**

Thetablecorrespondingtothemultipleregressionmodelofx_1**onother remaining variables is given by-**

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.4180	0.1832	13.196	<2e-16	***
x2	1.1084	0.1890	5.865	2.94e-08	***
x3	1.1976	0.2978	4.021	9.32e-05	***
x5	1.8535	0.3999	4.635	7.92e-06	***
x7	1.1644	0.4500	2.588	0.0107	*

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.5495 on 144 degrees of freedom
Multiple R-squared:0.7453, AdjustedR-squared:0.7382
F-statistic: 105.3 on 4 and 144 DF,p-value: < 2.2e-16

Fromthetableitmightseemthatx_7 mightbelesssignificant.Againwefindthe partial correlation coefficients as follows:-

$r_{12.357}$=-0.4391323
$r_{13.257}$=-0.3177126
$r_{15.237}$=-0.3603124
$r_{17.235}$=-0.2107832

$x_7$isdiscardedasanpredictor ,sinceithasthelowest partialcorrelationwithx_1

Thetablecorrespondingtothemultiplecorrelationcoefficientofx_1 onthe remaining covariates is given below-

	Estimate	Std. Error	tvalue	Pr(>|t|)
(Intercept)	2.3878	0.1864	12.809	<2e-16***
x2	1.2482	0.1846	6.762	3.10e-10***
x3	1.0398	0.2972	3.499	0.000621 ***
x5	2.2064	0.3832	5.757	4.91e-08***

Signif.codes:0 ‘*’0.001 ‘**’0.01 ‘*’0.05‘.’0.1‘’1**

Residual standard error: 0.5602 on 145 degrees of freedom
Multiple R-squared:0.7334, AdjustedR-squared:0.7279
F-statistic:133 on 3 and 145 DF,p-value: < 2.2e-16

Itseemsthatthereisnoinsignificantpredictor.Ifwecalculatethepartial correlations more,we have,

$r_{12.35}$=-0.4896347
$r_{13.25}$=-0.2790112
$r_{15.23}$=-0.431354

$x_3$willbeout andthetableonmodelling$x_1$on$x_2$and$x_5$willbegivenas- Estimate Std.

	Errort value	Pr(>|t|)		
(Intercept)	2.5802	0.1849	13.958	<2e-16***
x2	1.7197	0.1309	13.133	<2e-16***
x5	2.5510	0.3843	6.637	5.85e-10***

Signif.codes:0 ‘*’0.001 ‘**’0.01 ‘*’0.05‘.’0.1‘’1**

Residualstandarderror:0.5813on146degreesoffreedom
MultipleR-squared: 0.7109,AdjustedR-squared: 0.707F-statistic: 179.5 on 2 and 146 DF,p-value: < 2.2e-16

Notethatthereisahugefallinadjusted R-squared(more than2.3945%),hence we stop the stepwise regression here.It should be noted that $x_3$is an important predictor.

Sofor2021 ,forpredicting x_1,x_2,x_3,$x_5$arethe importantpredictorsalso.

Year2022:

Firstwestartmodelling$x_1$ontheothercovariates$x_2$to x_7,Afterfittingthemodel of multiple linear regression, we get the following result:

	Estimate	Std. Error	tvalue	Pr(>|t|)
(Intercept)	1.6677	0.2055	8.117	2.29e-13***
x2	0.5513	0.2072	2.661	0.00871 **
x3	1.4094	0.2424	5.815	3.98e-08***
x4	1.2735	0.4414	2.885	0.00454 **
x5	1.6032	0.3754	4.270	3.59e-05***
x6	0.9689	0.5681	1.705	0.09036 .
x7	0.7305	0.3964	1.843	0.06750 .

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.5281 on 139 degrees of freedom
Multiple R-squared:0.7737, AdjustedR-squared:0.7639
F-statistic:79.2 on 6 and 139 DF,p-value: < 2.2e-16

Fromtheabovetableithasbeenclearthat$x_6$and$x_7$arethetwoinsignificant covariates.

Nowwegoforthepartialcorrelationcoefficientsandwillmarkthatcovariate insignificant first which will show minimum partial correlation with $x_1$in absolute terms and we will eliminate that variable immediately.Then we will strat afresh with the remaining variables.This process will go until there is a sudden drop in adjusted R-squared value which is here 0.7639.

Thepartialcorrelationcoefficientsaregivenbelow-

$r_{12.34567}$ =-0.2201484
$r_{13.24567}$=-0.4423483
$r_{14.23567}$=-0.2377005
$r_{15.23467}$=-0.3405645
$r_{16.23457}$=-0.1431556
$r_{17.23456}$=-0.1544202

Itshowsthat$x_6$hasthelowestpartialcorrelationcoefficientwith$x_1$andhenceis eliminated as an predictor.

Aftereliminationofx_6,weagaindothemultipleregressionfor$x_1$onthe remaining predictors and thus we obtain the following table as-

	Estimate	Std. Error	tvalue	Pr(>|t|)
(Intercept)	1.8033	0.1908	9.454	<2e-16***
x2	0.4643	0.2022	2.297	0.02313 *
x3	1.4608	0.2421	6.033	1.36e-08***
x4	1.2739	0.4444	2.867	0.00479 **
x5	1.7397	0.3693	4.711	5.87e-06***
x7	0.8113	0.3963	2.047	0.04249 *

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.5317 on 140 degrees of freedom
Multiple R-squared:0.7689, AdjustedR-squared:0.7607
F-statistic: 93.18 on 5 and 140 DF,p-value: < 2.2e-16

Now$x_7$and$x_2$canbeseento beinsignificantfromtheabovetable. Again the

partial correlation coefficients are obtained as :

$r_{12.3457}$=-0.1905426
$r_{13.2457}$=-0.4542451
$r_{14.2357}$=-0.2354619
$r_{15.2347}$=-0.3699163
$r_{17.2345}$=-0.1705002

$x_7$hasminimumpartialcorrelationcoefficientherewithx_1,hence$x_7$isdiscarded as a predictor.

Thetablecorrespondingtothemultipleregressionmodelof$x_1$onother remaining variables is given by-

	Estimate	Std. Error	tvalue	Pr(>|t|)
(Intercept)	1.7408	0.1904	9.142	6.09e-16***
x2	0.5295	0.2019	2.623	0.00968 **
x3	1.3759	0.2412	5.704	6.63e-08***
x4	1.3676	0.4470	3.060	0.00265 **
x5	1.9679	0.3560	5.528	1.52e-07***

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residualstandarderror: 0.5377 on 141 degrees offreedom
MultipleR-squared: 0.762,AdjustedR-squared: 0.7553F-statistic: 112.9 on 4 and 141 DF,p-value: < 2.2e-16

Fromthetableitmightseemthatx_4 and$x_2$mightbelesssignificant.Againwe find the partial correlation coefficients as follows:-

$r_{12.357}$=-0.2156702
$r_{13.257}$=-0.4329692
$r_{15.237}$=-0.2495095
$r_{17.235}$=-0.4220331

$x_2$isdiscardedasanpredictor ,sinceithasthelowest partialcorrelationwithx_1

Thetablecorrespondingtothemultiplecorrelationcoefficientofx_1 onthe remaining covariates is given below-

	Estimate	Std. Error	tvalue	Pr(>|t|)
(Intercept)	1.7864	0.1935	9.232	3.44e-16***
x3	1.6164	0.2277	7.098	5.56e-11***
x4	2.1133	0.3520	6.004	1.53e-08***
x5	2.0576	0.3616	5.690	7.00e-08***

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.5487 on 142 degrees of freedom
Multiple R-squared:0.7504, AdjustedR-squared:0.7451
F-statistic: 142.3 on 3 and 142 DF,p-value: < 2.2e-16

Itseems thatthereisnoinsignificant predictor.

Note thatthere is a fall in adjusted R -squared as per our precision(more than 2.3945%),hencewestopthestepwiseregressionhere.Itshouldbenotedthat$x_2$is an important predictor and cannot be discarded.

Sofor2022 , forpredicting x_1,x_2,x_3,x_4,$x_5$aretheimportantpredictors also.

<u>Year2023</u>:

Firstwestartmodelling$x_1$ontheothercovariates$x_2$to x_7,Afterfittingthemodel of multiple linear regression, we get the following result:

	Estimate	Std. Error	tvalue	Pr(>|t|)
(Intercept)	1.5269	0.2014	7.581	5.76e-12***
x2	0.5623	0.1974	2.849	0.0051 **
x3	1.5415	0.2137	7.214	4.01e-11***
x4	0.7750	0.4845	1.599	0.1121
x5	1.7643	0.3490	5.055	1.43e-06***
x6	0.2915	0.5913	0.493	0.6229
x7	1.1080	0.3898	2.843	0.0052 **

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.483 on 130 degrees of freedom
Multiple R-squared:0.8284, AdjustedR-squared:0.8204
F-statistic: 104.6 on 6 and 130 DF,p-value: < 2.2e-16

Fromtheabovetableithasbeenclearthat$x_6$and$x_4$arethetwoinsignificant covariates.

Nowwegoforthepartialcorrelationcoefficientsandwillmarkthatcovariate insignificant first which will show minimum partial correlation with $x_1$in absolute terms and we will eliminate that variable immediately.Then we will strat afresh with the remaining variables.This process will go until there is a sudden drop in adjusted R-squared value which is here 0.8204.

Thepartialcorrelationcoefficientsaregivenbelow-

$r_{12.34567}$ **=-0.2423985**
$r_{13.24567}$**=-0.5346904**
$r_{14.23567}$**=-0.138925**
$r_{15.23467}$**=-0.4053155**
$r_{16.23457}$**=-0.04319266**
$r_{17.23456}$**=-0.2419218**

Itshowsthat$x_6$hasthelowestpartialcorrelationcoefficientwith$x_1$andhenceis eliminated as an predictor.

Aftereliminationofx_6,weagaindothemultipleregressionfor$x_1$onthe remaining predictors and thus we obtain the following table as-

	Estimate	Std. Error	tvalue	Pr(>|t|)
(Intercept)	1.5660	0.1846	8.483	4.06e-14***
x2	0.5380	0.1906	2.823	0.00550 **
x3	1.5613	0.2093	7.461	1.06e-11***
x4	0.7628	0.4825	1.581	0.11631
x5	1.7910	0.3438	5.210	7.16e-07***
x7	1.1459	0.3810	3.007	0.00316 **

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residualstandarderror: 0.4817 on 131 degrees offreedom
MultipleR-squared: 0.828,AdjustedR-squared: 0.8215F-statistic: 126.2 on 5 and 131 DF,p-value: < 2.2e-16

Now$x_4$canbeseen tobeinsignificantfromtheabovetable.

Again the partial correlation coefficients are obtained as :

$r_{12.3457}$**=-0.2394682**
$r_{13.2457}$**=-0.5460767**
$r_{14.2357}$**=-0.1368251**
$r_{15.2347}$**=-0.4142694**
$r_{17.2345}$**=-0.2541266**

$x_4$hasminimumpartialcorrelationcoefficientherewithx_1,hence$x_4$isdiscarded as a predictor.

Thetablecorrespondingtothemultipleregressionmodelof$x_1$onother remaining variables is given by-

	Estimate	Std. Error	tvalue	Pr(>|t|)
(Intercept)	1.4925	0.1797	8.307	1.03e-13***
x2	0.7203	0.1526	4.720	5.94e-06***
x3	1.6404	0.2044	8.027	4.82e-13***
x5	1.7783	0.3456	5.145	9.46e-07***
x7	1.2136	0.3808	3.187	0.00179 **

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.4844 on 132 degrees of freedom
Multiple R-squared:0.8248, AdjustedR-squared:0.8194
F-statistic: 155.3 on 4 and 132 DF,p-value: < 2.2e-16

Fromthetableitseems that$x_7$mightbelesssignificant.Againwefindthepartial correlation coefficients as follows:-

$r_{12.357}$**=-0.3799814**
$r_{13.257}$**=-0.5727104**
$r_{15.237}$**=-0.4087001**
$r_{17.235}$**=-0.2673271**

$x_7$isdiscardedasanpredictor ,sinceithasthelowest partialcorrelationwithx_1

Thetablecorrespondingtothemultiplecorrelationcoefficientofx_1 onthe remaining covariates is given below-

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.3960	0.1831	7.625	4.15e-12***
x2	0.8922	0.1476	6.045	1.41e-08***
x3	1.5241	0.2079	7.331	1.99e-11***
x5	2.0857	0.3432	6.078	1.21e-08***

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residualstandarderror:0.5008on133degreesoffreedom
MultipleR-squared: 0.8113,AdjustedR-squared: 0.807F-statistic: 190.6 on 3 and 133 DF,p-value: < 2.2e-16

Itseemsthatthereisnoinsignificantpredictor.Wefindthepartialcorrelation coefficients as-

$r_{12.35}$=-0.4642783
$r_{13.25}$=-0.5364827
$r_{15.23}$=-0.4662245

$x_2$isdiscardedandnowthetableformodelling$x_1$on$x_3$andx_5 isgivenby-

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.5935	0.2026	7.864	1.1e-12***
x3	2.3510	0.1761	13.349	< 2e-16***
x5	2.2742	0.3844	5.916	2.6e-08***

Signif.codes:0 '*'0.001 '**'0.01 '*'0.05'.'0.1''1**

Residual standard error: 0.5633 on 134 degrees of freedom
Multiple R-squared:0.7594, AdjustedR-squared:0.7558
F-statistic: 211.5 on 2 and 134 DF,p-value: < 2.2e-16

NotethatthereisagreatfallinadjustedR-squaredasperourprecision(more than 2.3945%) ,hence we stop the stepwise regression here.It should be noted that $x_2$is an important predictor and cannot be discarded.

Sofor2022 ,forpredicting x_1,x_2,x_3,$x_5$arethe importantpredictorsalso.

<u>Overall interpretation</u>:

In all the years $x_6$and $x_7$are not well predictors to predict x_1,in all the years x_2,x_3,x_4,$x_5$theinfluentialcovariatesforpredictingx_1,exceptforthreeyears,viz,for

2018,2021 and 2023,$x_4$has not been included as a predictor in the former list.But simply looking at the summary table for the full model,for all the years $x_6$and $x_7$have turned out to be insignificant but the remaining are significant.Thus $x_4$mightbetakenintoaccountinpredictingtheresponse,fromstatisticalviewpoint.

RegressionDiagnostics

Considerthefollowingregressionmodel:

$y_i=\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+\ldots\ldots\ldots+\beta_p x_{pi}+e_i$,i=1(1)n, y being the response variable,and$x_1,x_2\ldots\ldots,x_p$arethecovariatesinpredictingy,ebeingtheerror term,n is the total no of observations,p is the no of predictors.

Letthefittedvalueoftheresponsebey$_i$be^,andtheresidualfortheith observation is given by: $z_i=$ ———,$e_i=y_{i-}$^,i=1(1)n

Thestandardisedresidualplothelpsindeterminingthepotentialorsuspected outliers.Theresidualshavingvaluegreaterthan+2orlesserthan-2arecalled suspected outliers.

Inorder tofindtheinfluentialobservations,wemakeuseofWelschandKuh's "DFITS" measureand"Cook'sDistance"measure.

The Welsch and Kuh's measure is defined as-

DFITS$_i$= ,i=1,2,…,n

Thus,DFITS$_i$isthescaleddifferencebetweentheithfittedvalueobtainedfrom the full data and the ithfitted value obtained by deleting the ithobservation.,i=1(1)n,where^ istheunbiasedestimatorofσ^2,beingthe unknown common variance of e_i's and is given by-

^ =,whereSSE$_{(i)}$isthesumofsquaredresidualswhenwefitthemodelto the (n- 1) observations obtained by omitting the ithobservation.

Andthevalueiscalledtheleveragevaluefortheithobservation,beingweight (leverage)giventoindeterminingtheithfittedvalue^.

Pointswith|DFITS$_i$|largerthan2 $\sqrt{\frac{+1}{--1}}$ areusuallyclassifiedasinfluential points.

Cook'sdistancemeasurestheinfluenceoftheithobservationby-

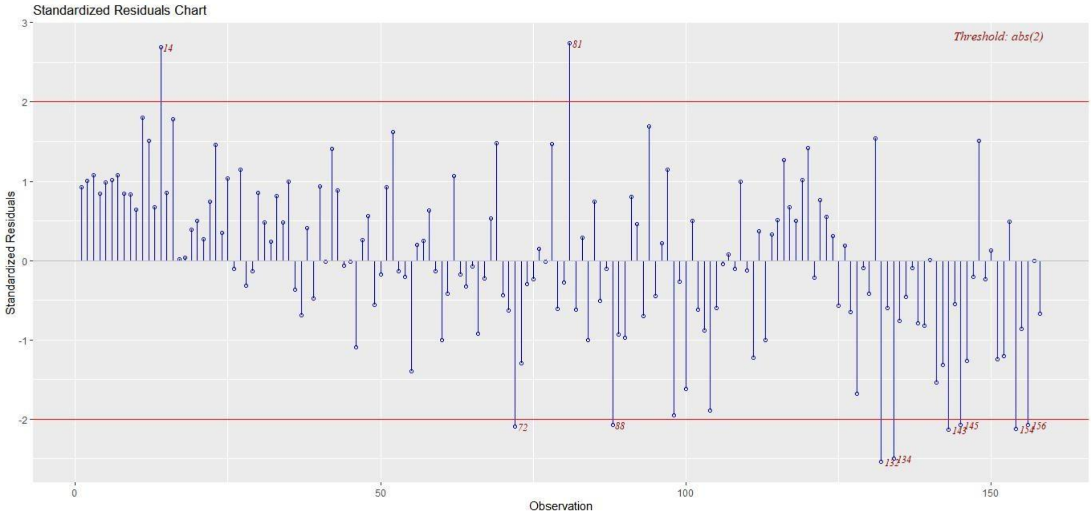
$$C_i = \frac{\sum_{=1}(\hat{} - \hat{}())}{\hat{}(+1)} \quad , i=1,2,\ldots\ldots,n$$

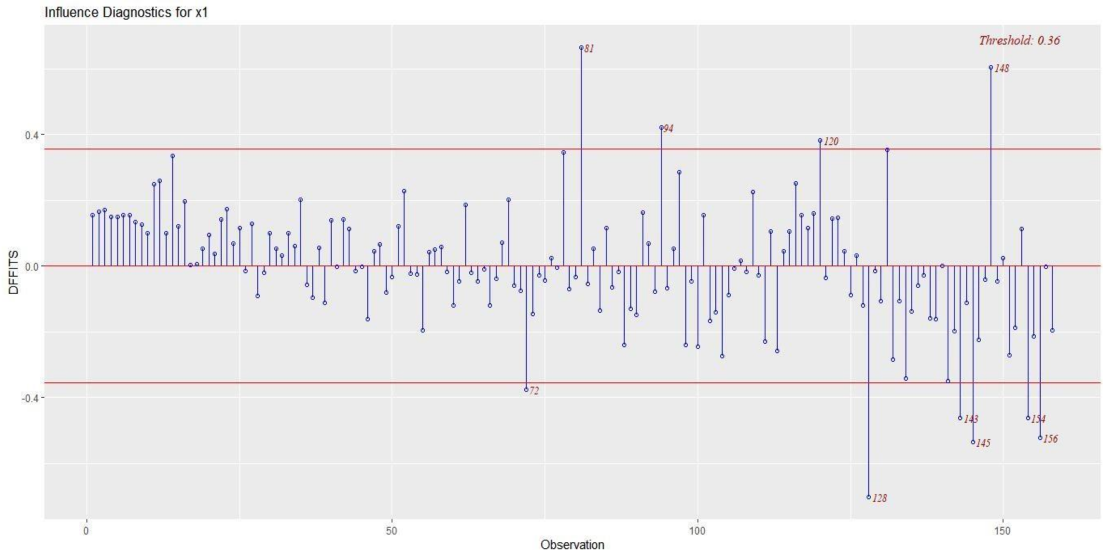
Itmeasuresthedifferencebetweenthefitted valuesobtainedfromthefulldata and the fitted values obtained by deleting the ithobservation.

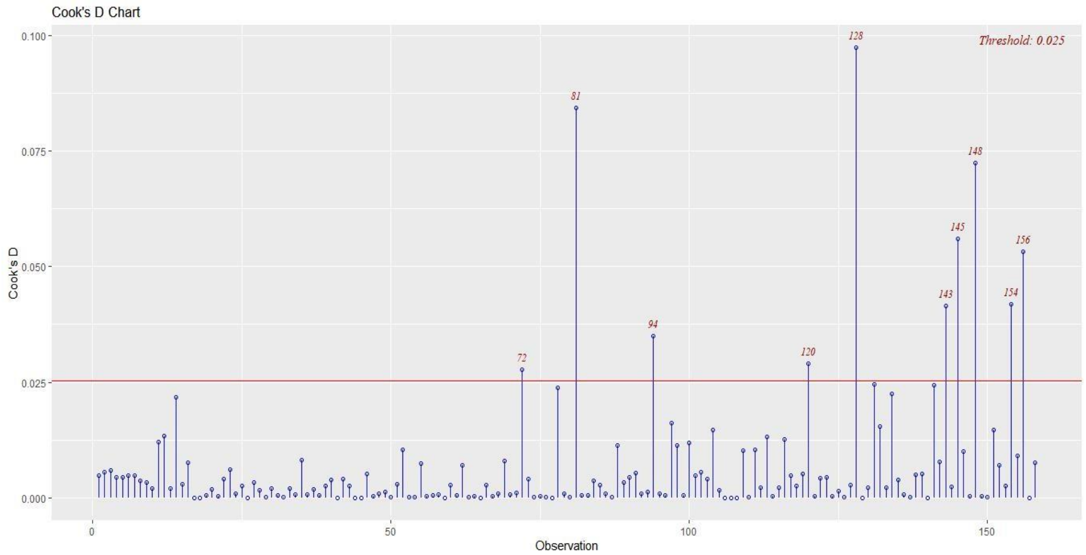
Therefore, a large value of $C_i$indicates that the point is influential. It has been suggested that points with $C_i$values greater than the 50% point of the F-distribution with (p+1) and (n- p- 1) degrees of freedom be classified as influentialpoints.Apracticaloperationalruleistoclassifypointswith$C_i$values greater than 1 as being influential.

Thestandardisedresidualchart,theDFITSchartandtheCook'sdistancecharts are presented for each year.

<u>Year2015</u>:

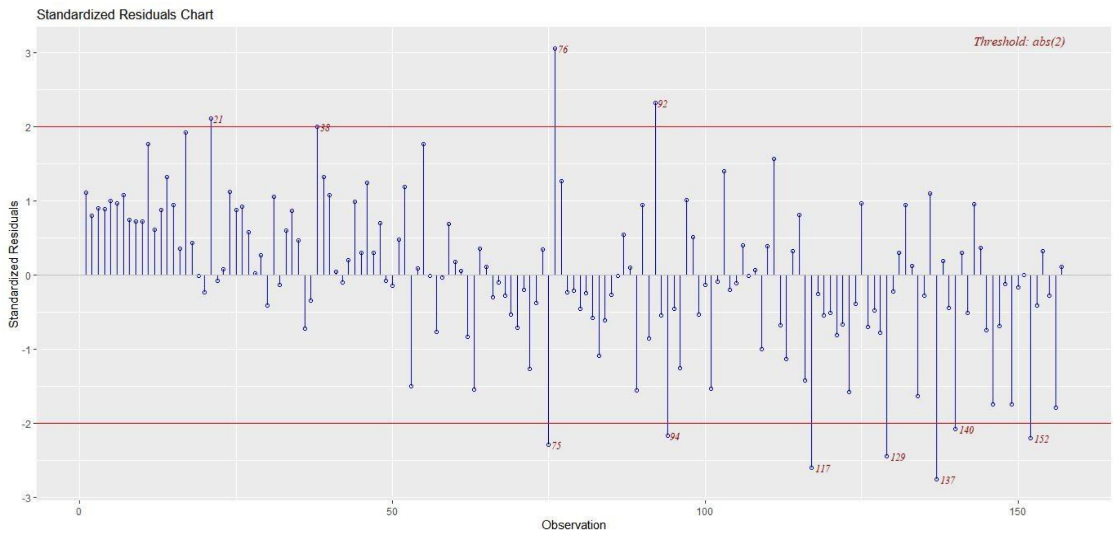


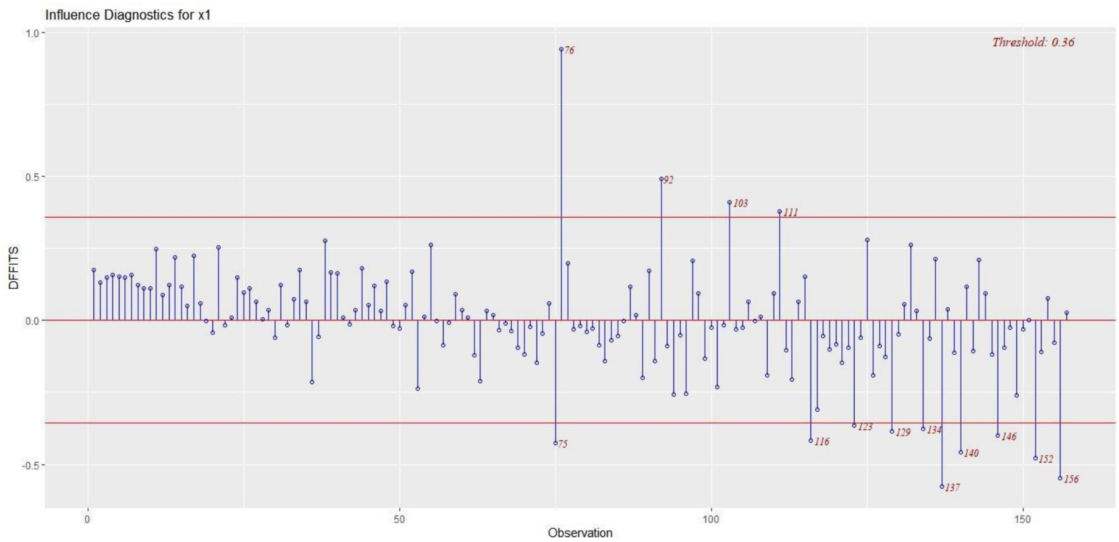


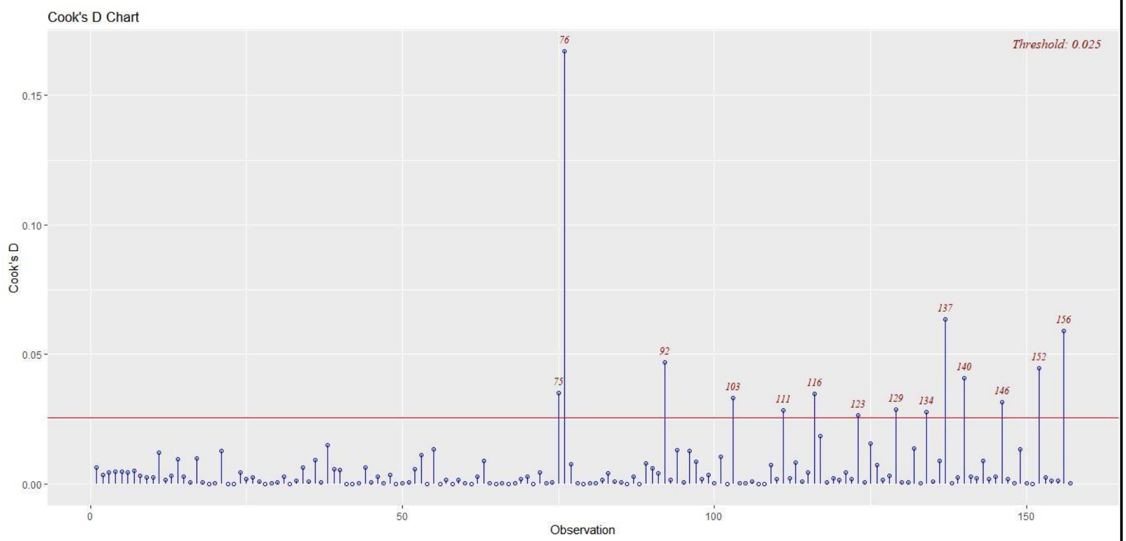


<u>Interpretation</u>:-The influential data points are corresponding to the countries Mauritus,Azerbaijan,Macedonia,Haiti,Armenia,Senegal,Niger,Madagascar,Afganisthan,Benin.Thelasttwo are in5 leasthappy countries also.Outofthe 10,5 are from Africa,2 are from Asia(Azerbaijan and Afganisthan), Macedonia,Haiti,Armenia are from Western territory.

<u>Year2016</u>:

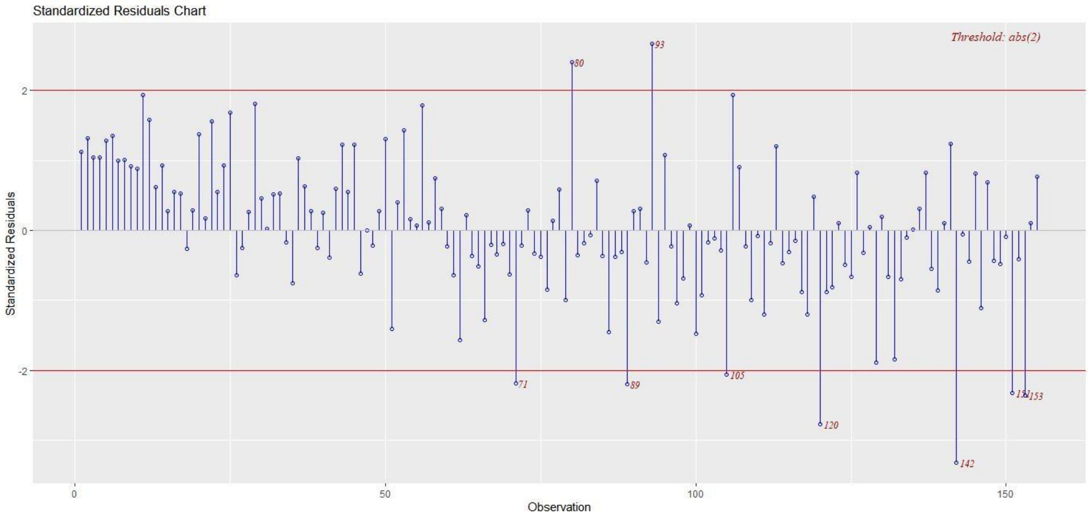

Standardized Residuals Chart
Threshold: abs(2)
Standardized Residuals
Observation
3
2
1
0
-1
-2
-3
0
50
100
150
76
92
21
38
75
94
117
129
137
140
152

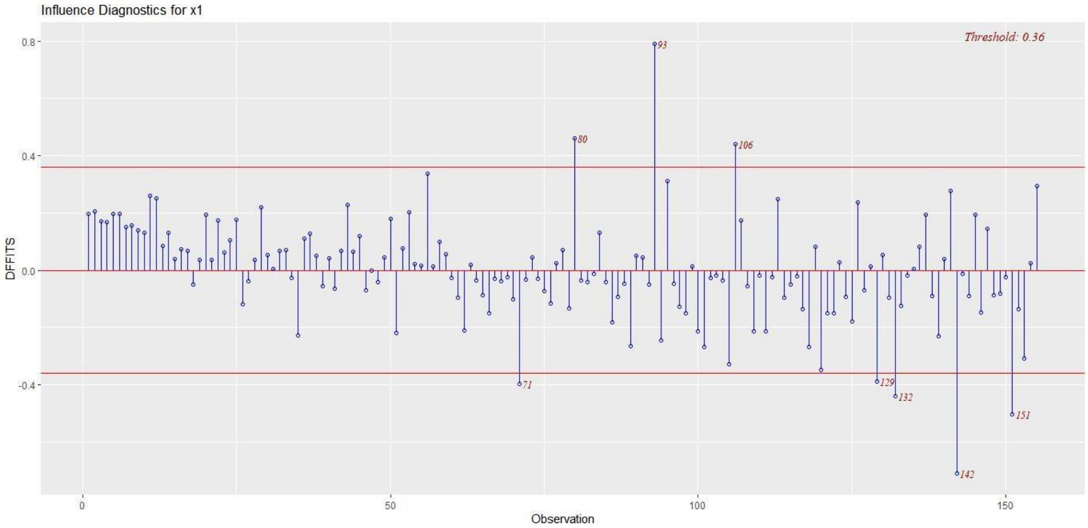

Influence Diagnostics for x1
Threshold: 0.36
DFFITS
Observation
1.0
0.5
0.0
-0.5
0
50
100
150
76
92
103
111
75
116
123
129
134
137
140
146
152
156

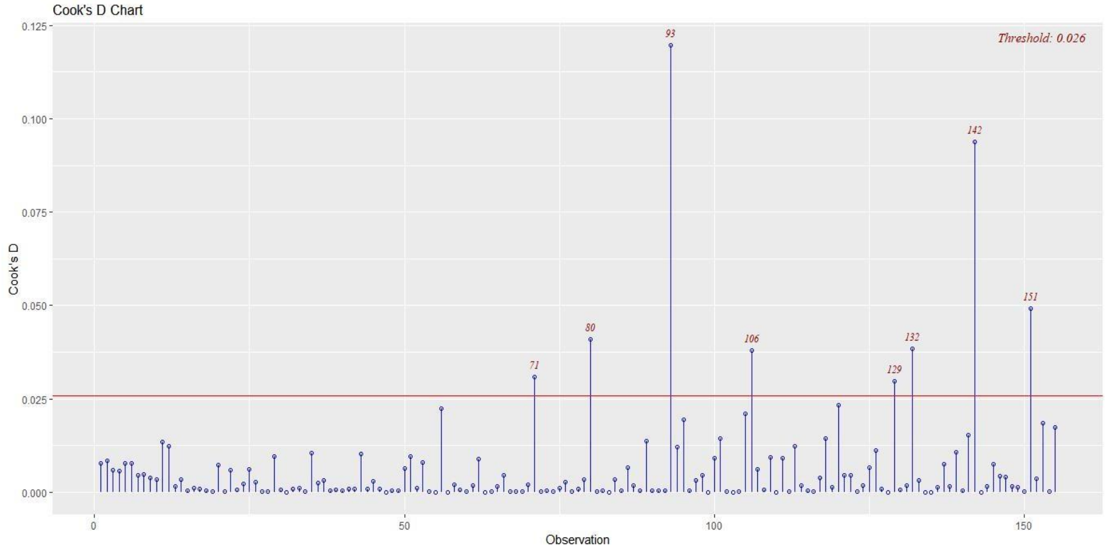

Cook's D Chart
Threshold: 0.025
Cook's D
Observation
0.15
0.10
0.05
0.00
0
50
100
150
76
75
92
103
111
116
123
129
134
137
140
146
152
156

<u>Interpretation</u>:-The influential points are data points corresponding to Croatia,Hong Kong,Hungary,Laos,Bangladesh,Ethiopia,Kenya,Senegal,Sudan, Haiti,IvoryCoast,BurkinaFaso,Guinea,Togo.Here,8countriesarefromAfrica,3 are from Asia,the remaining are from Western territory.

<u>Year2017</u>:

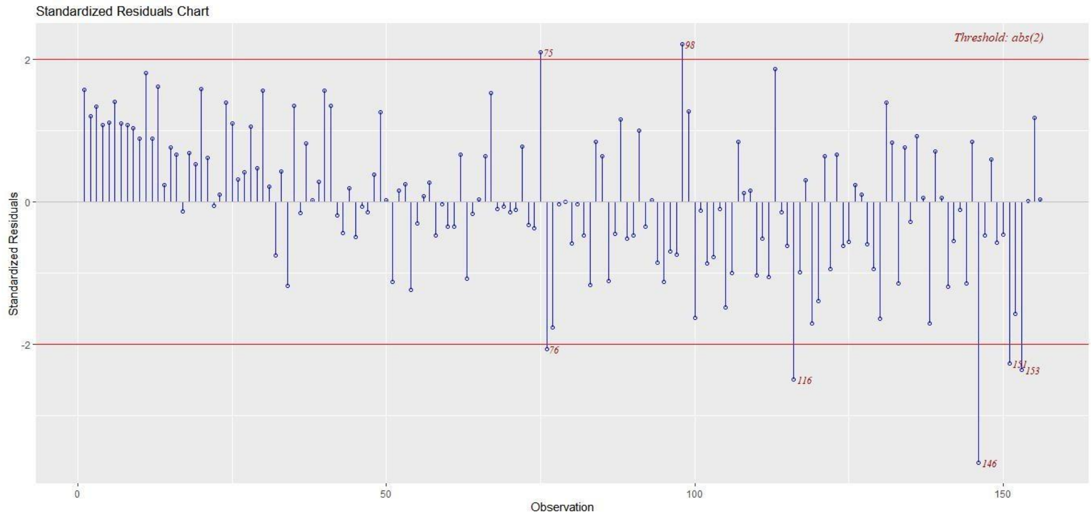


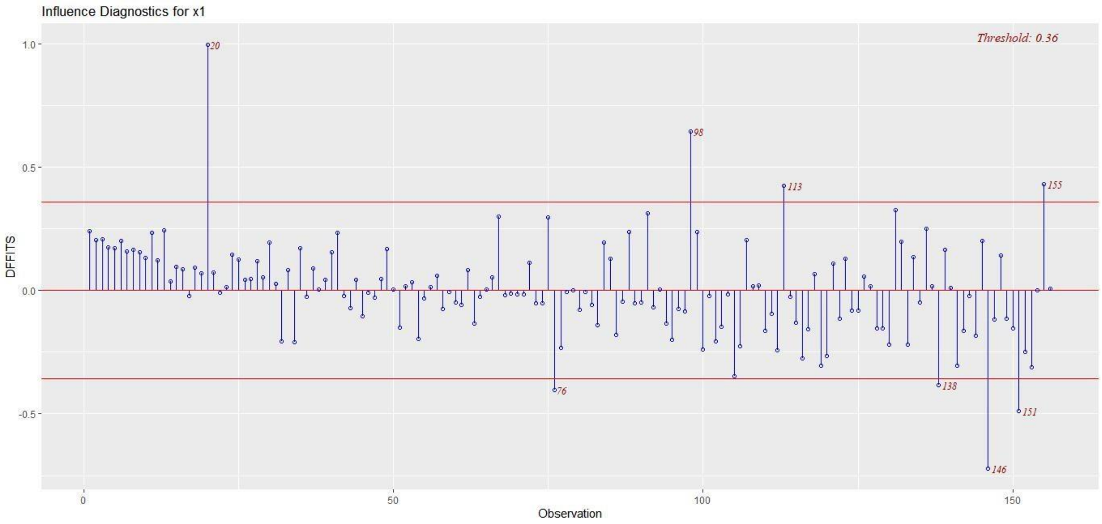


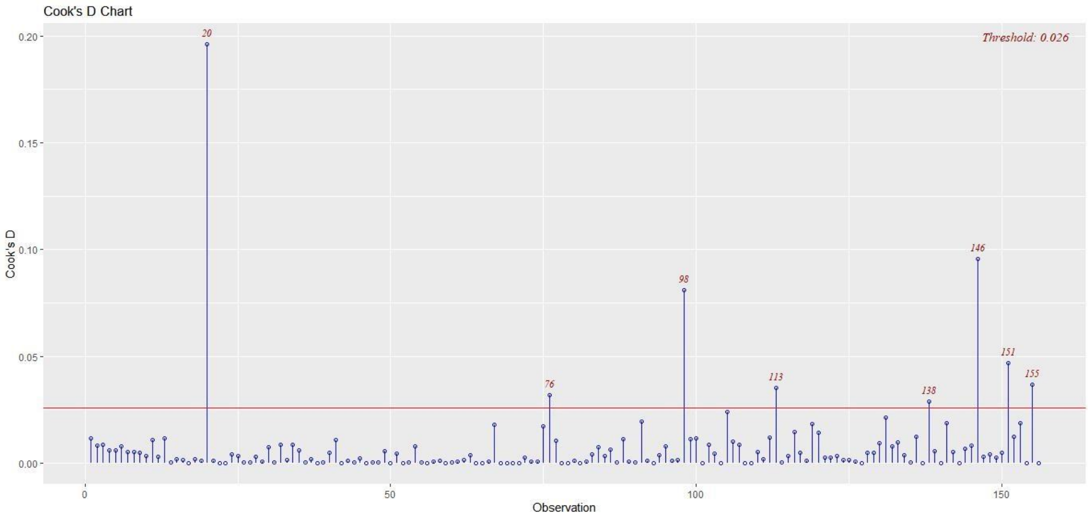


<u>Interpretation</u>:-The influential points are data points corresponding to Paraguay,China,Macedonia,Bulgaria,Ghana,Afghanisthan,Togo.Here2are from Africa,2 are from Asia.

<u>Year2018</u>:

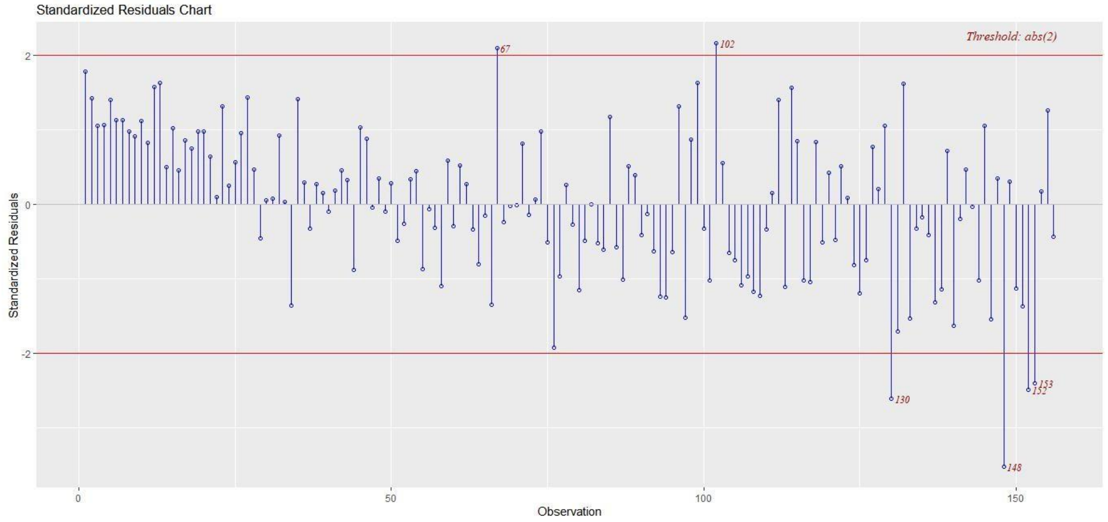


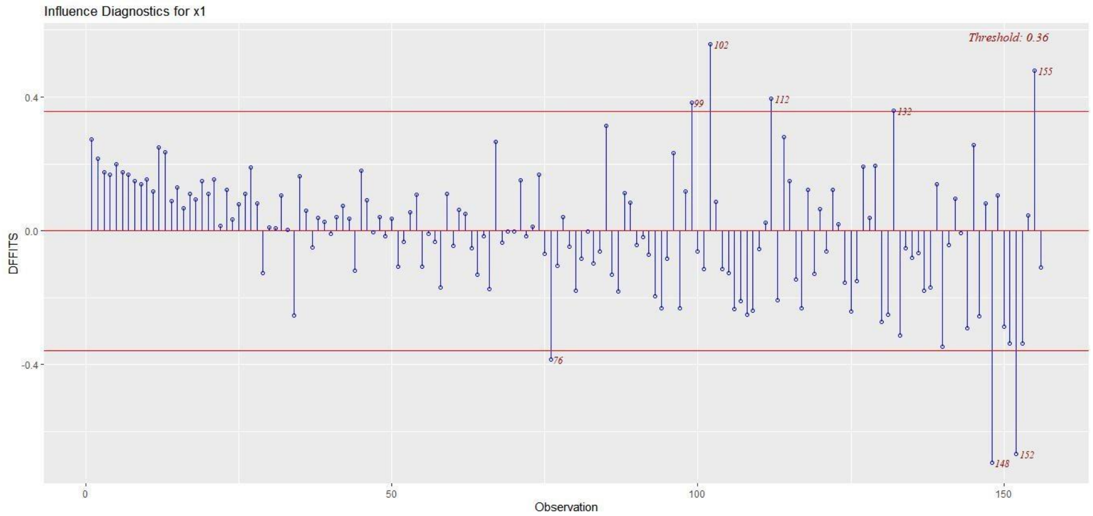


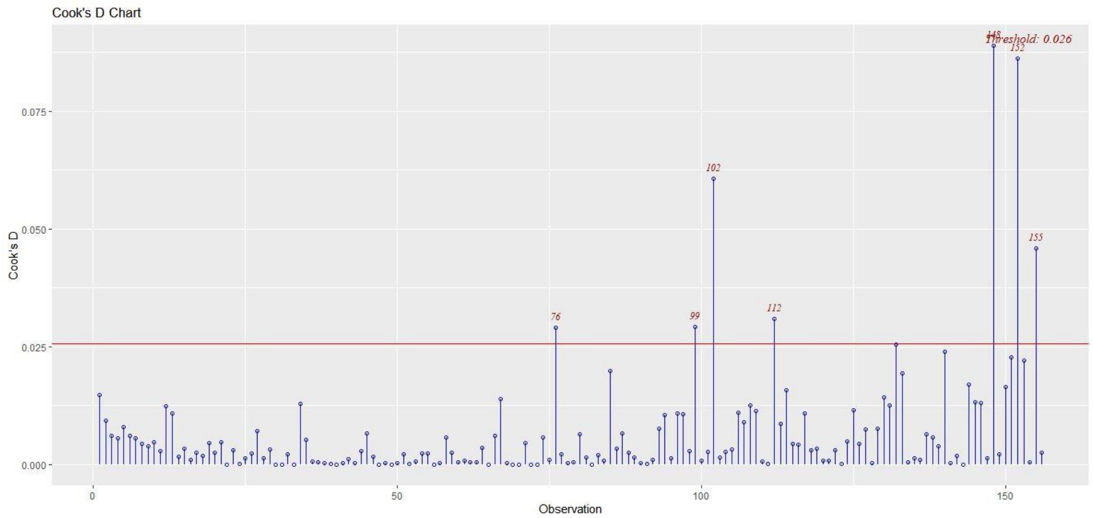


<u>Interpretation</u>:-The influential data points corresponding to the countries Israel,Pakistan,Bhutan,Albania,Sudan,Afghanisthan,Syria,SouthSudanamong which 3 are from Africa,remaining 5 are from Asia.

<u>Year2019</u>:

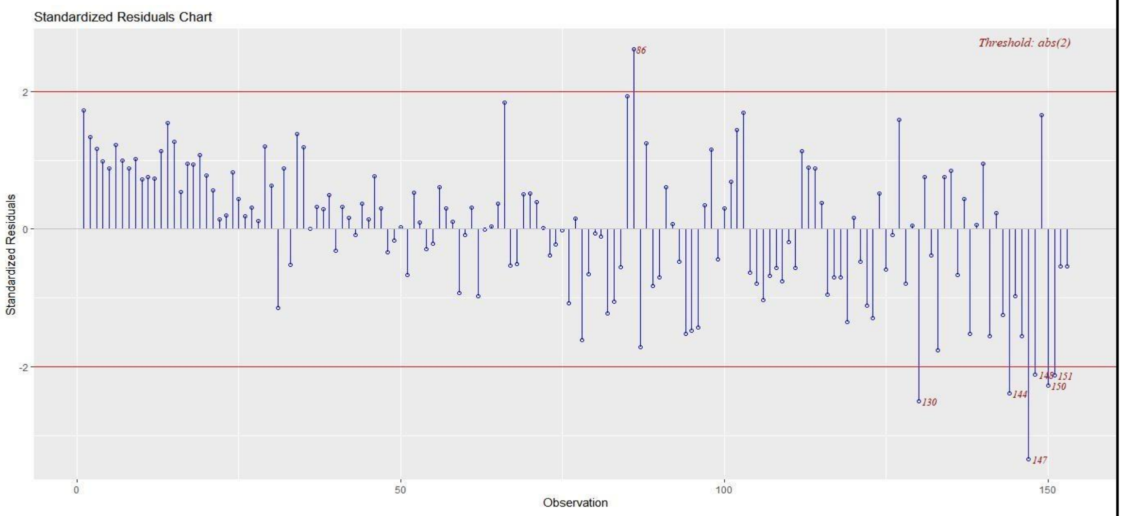


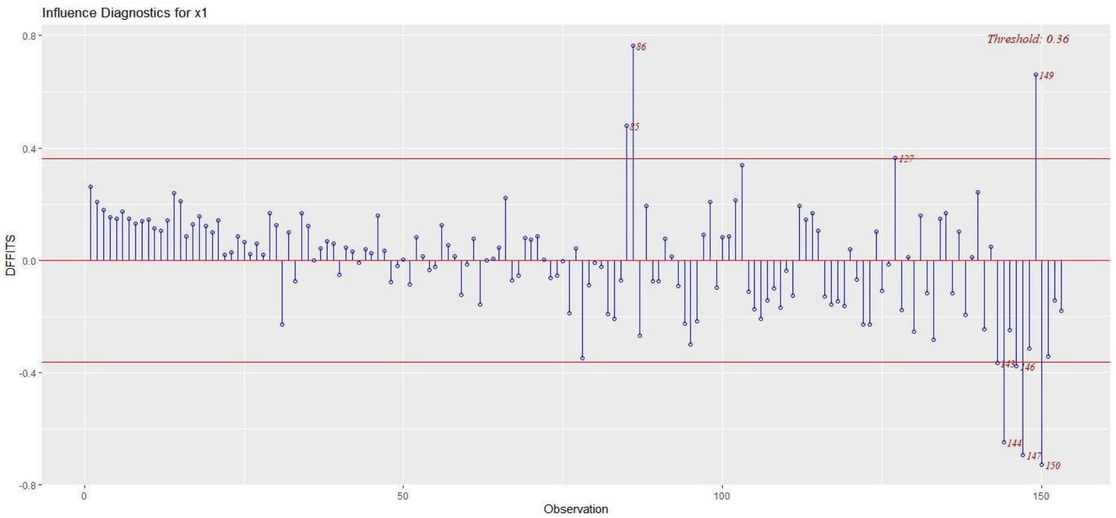


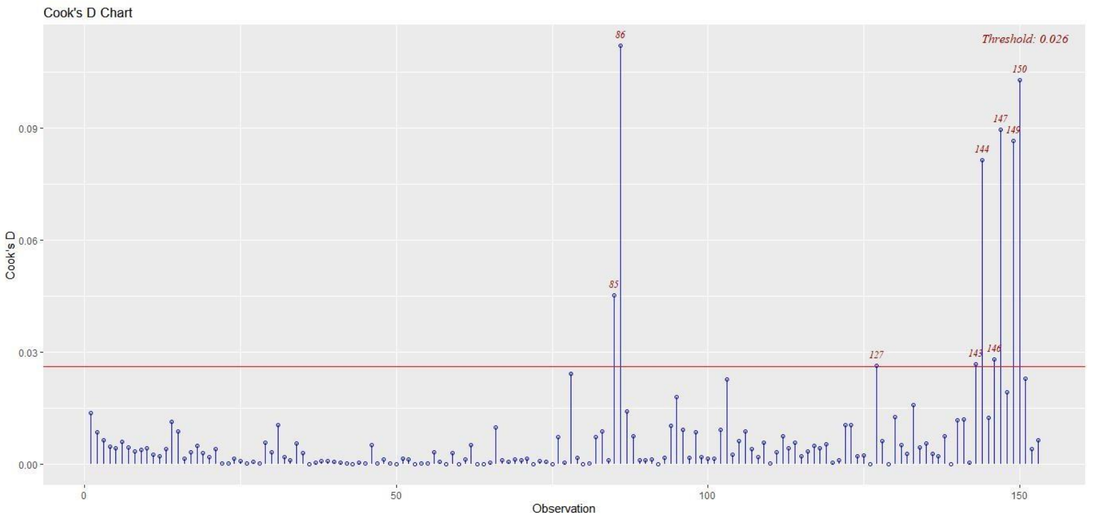


<u>Interpretation</u>:-Theinfluentialdatapointsarecorrespondingtothecountries Croatia,Ghana,Jordan,Senegal,Myanmar,Haiti,Yemen,Afghanisthan among

which3arefromAfrica,3arefromAsia,remainingarefromwesternterritory. **Year 2020:**

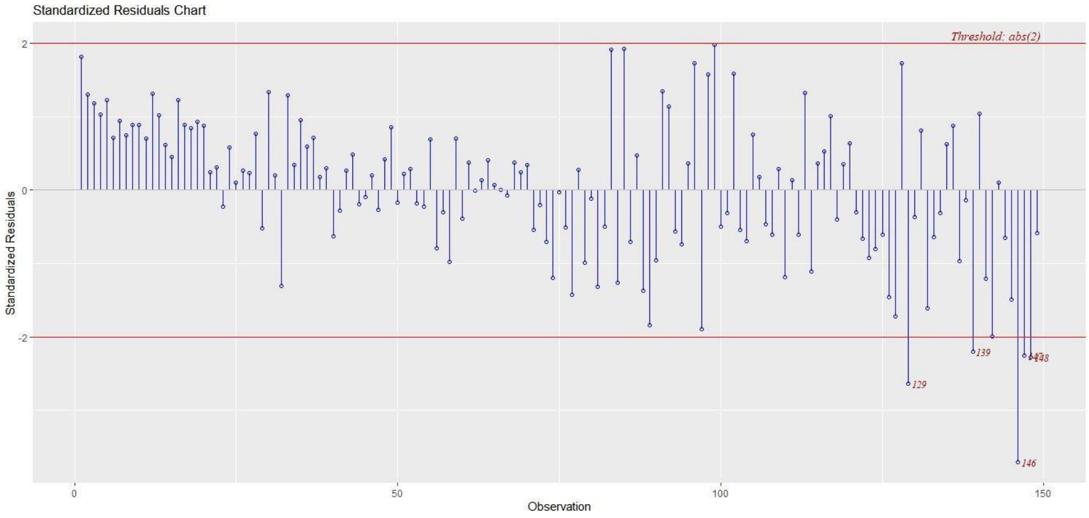


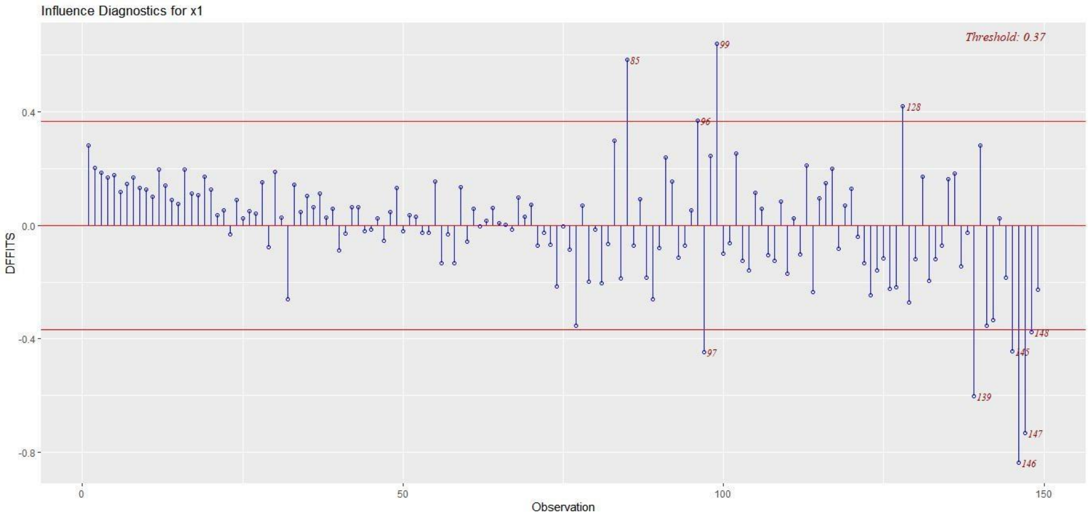


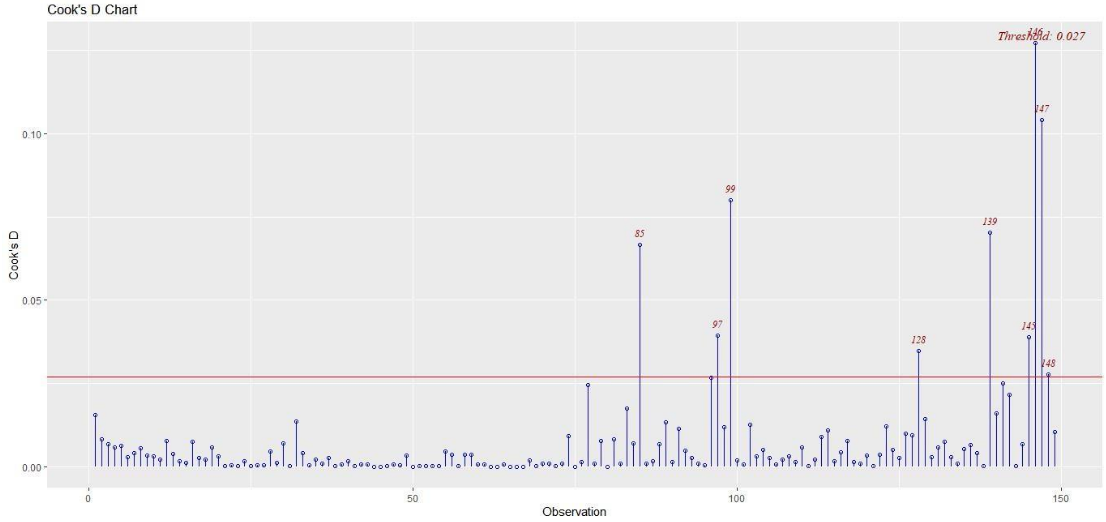


<u>Interpretation</u>:-The influential points are corresponding to data points Indonesia,Ivory Coast,Uganda,Haiti,Lesotho,Malawi,Yemen,Tanzania,Central AfricanRepublicamongwhich7arefromAfrica,1isfromAsia.Thelasttwoare in 6 least happy countries also.

<u>Year2021</u>:

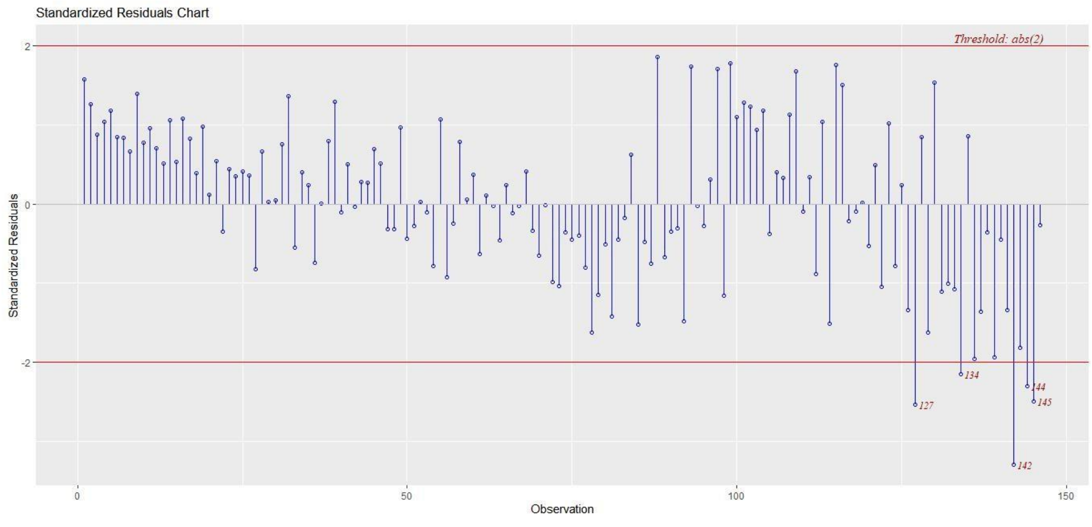


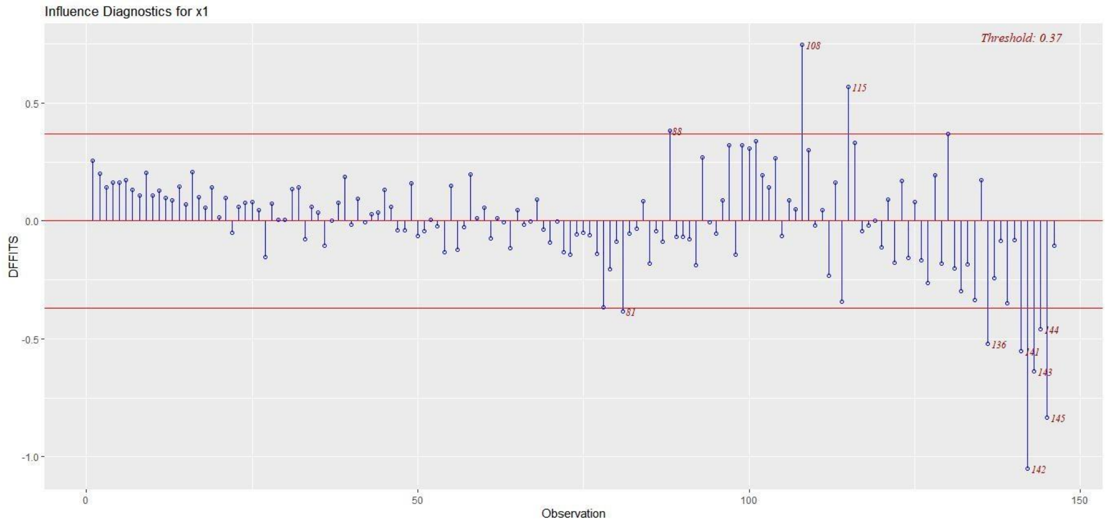


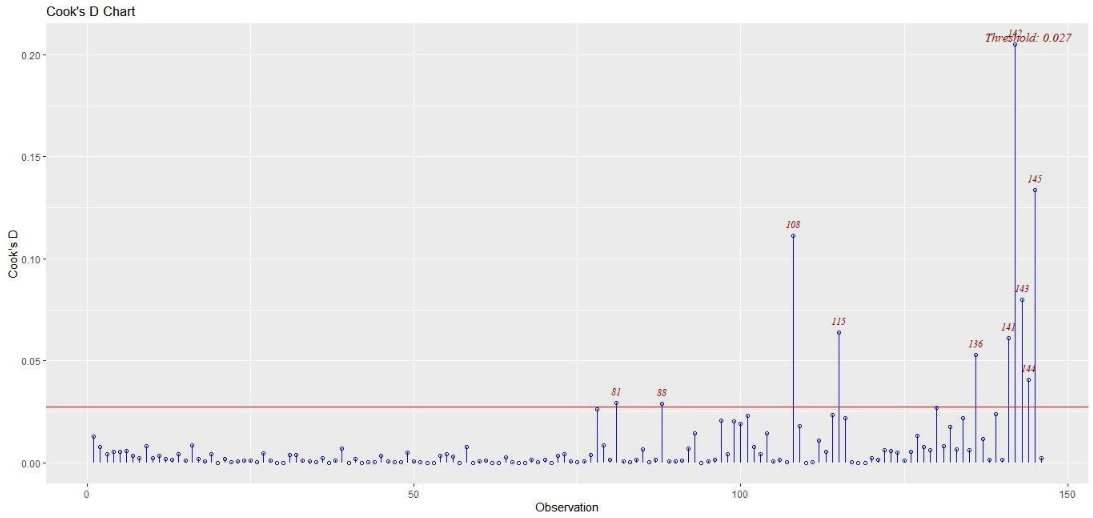


Interpretation:-Theinfluentialdatapointsarecorrespondingtothecountries China,Ghana,Niger,Gambia,Jordan,Sierra Leone,Malawi,Lesotho,Bostwana, Rwanda among which 8 are from Africa,remaining 2 are from Asia.The last fours are in 6 least happy countries also.

Year2022:

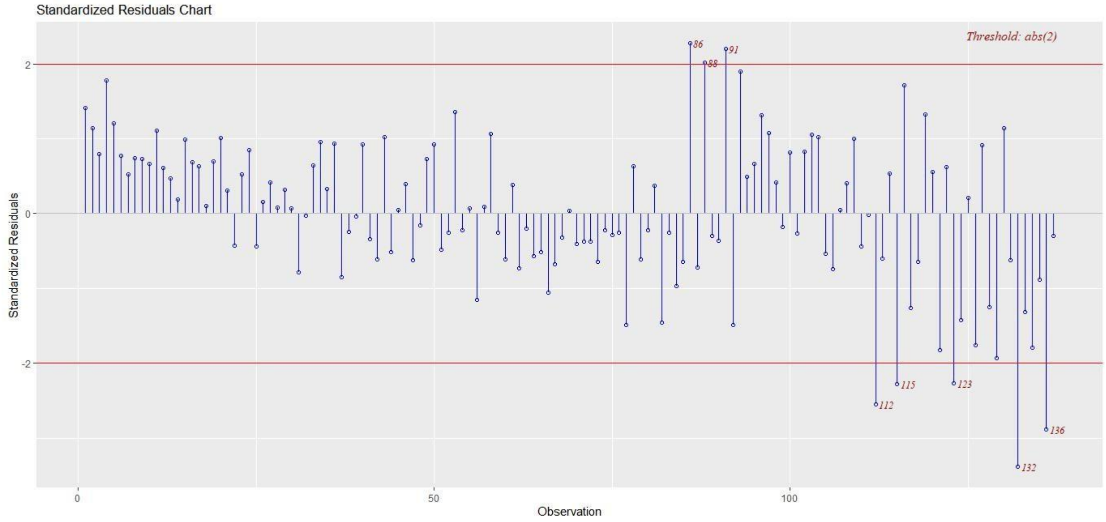


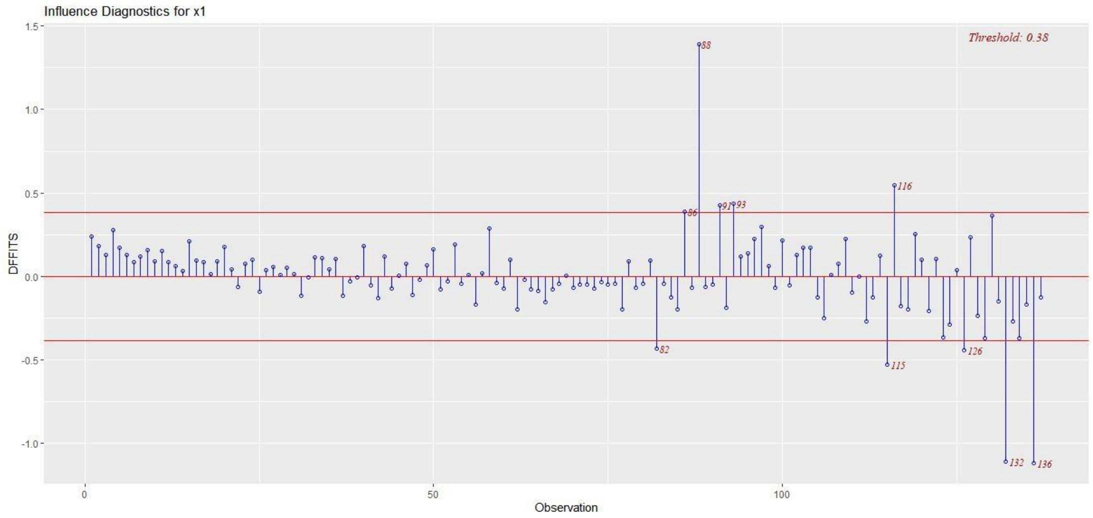


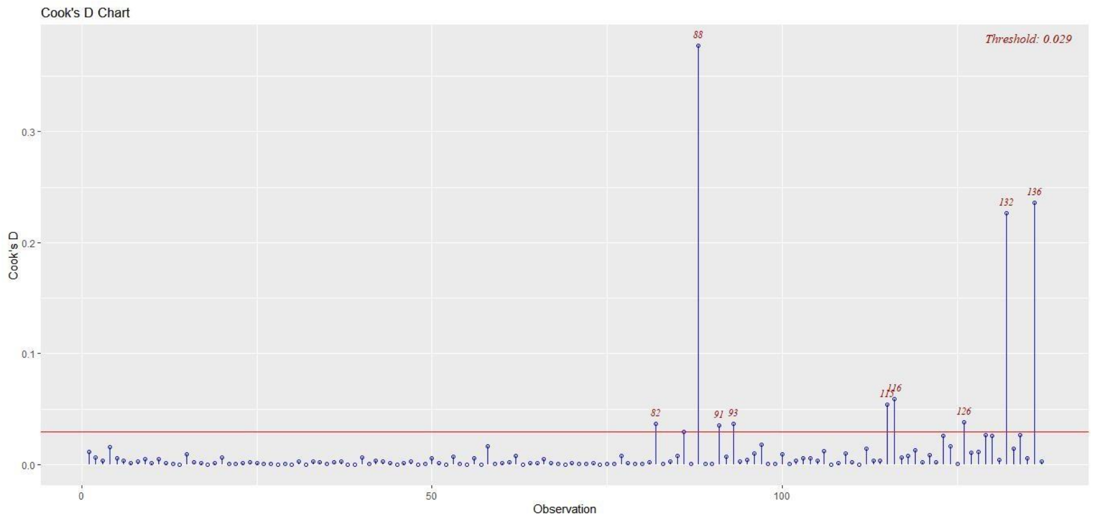


<u>Interpretation</u>:-The influential data points are corresponding to the countries Russia,Indonesia,Iraq,Cambodia,Togo,SierraLeone,Lesotho,Bostwana,Rwanda, Zimbabwe,among which 6 are African,3 are Asian.Also,the last fours are in 6 least happy countries.

<u>Year2023</u>:







<u>Interpretation</u>:-Theinfluentialdatapointsarecorrespondingtothecountries

Algeria,NorthMacedonia,Georgia,Ukraine,Chad,Cambodia,Malawi,Sierra Leone,among which 4 are African Cambodia comes from Asia,and the remaining are European.

Overall Interpretation:-It is to be noted that the influential countries in the data set vary from year to year.But most influential countries are from African continent,then Asia gets the second priority.Also,most of the suspected outliers,according to the standardised residual plot turn out to be highly influential and this happens for every year.It is an important note that what we thinkasoutlier,mightturnoutasinfluentialdatainreallife.Foreachyearonly2 or 3 data points are suspected outliers,but they do not create anyproblem in our statistical analysis,so we have kept them intact instead of avoiding them.

PrincipalComponentAnalysis

In order to get another view regarding the prediction of the response,we use principal component analysis (PCA).It has been used to find out if there can be anydimensionreductionfromthe6covariatestopredicttheresponseinabetter way.Let us consider the following random vector as follows-

$$X=\begin{pmatrix} 1 \\ \vdots \end{pmatrix}$$ **,withthedispersionmatrixhaving**

theeigenvalueeigenvector pairs(λ_1,e_1),(λ_2,e_2),…….(λ_m,e_m),where

$$\lambda_1 \geq \lambda_2 \geq \ldots\ldots \lambda_m \geq 0$$

ThejthprincipalcomponentisgivenbyY_j=,withVar(Y_j)=λ_j,j=1,2,….,m The

proportion of variation for the jth principal component Y_j is given by

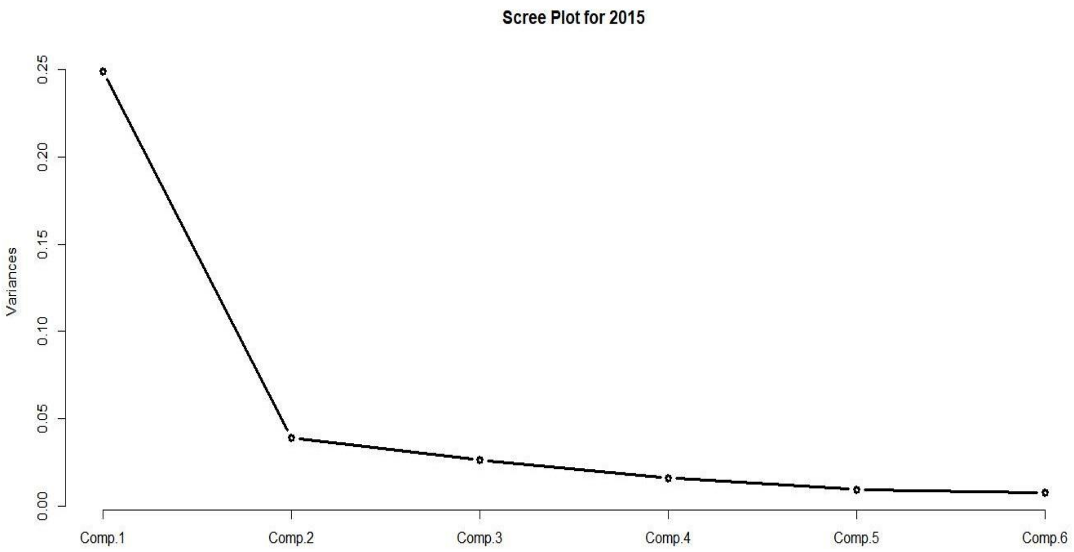
$\frac{}{1++\ldots\ldots+}$**, j=1,2,….,m**

Thecumulative proportionofvariationforthejthprincipalcomponentY_j is givenby$\frac{1\pm\pm\ldots\ldots\pm}{1++\ldots\ldots+}$ **, j=1,2,….,m**

Thepcaisdoneforyeartoyear.Thetableofcumulativeproportionofvariations and scree plot are given for year to year.

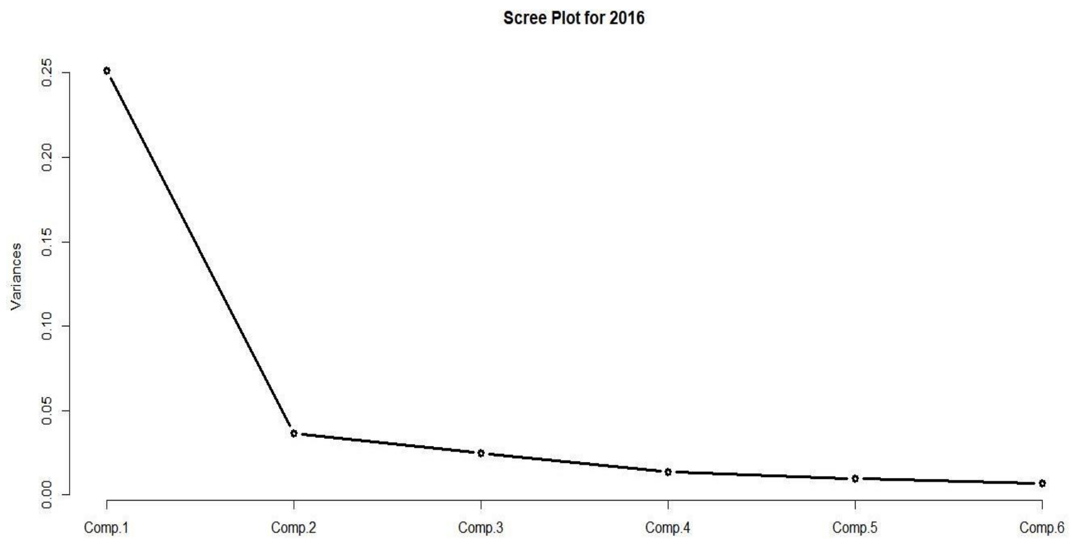
<u>Year2015</u>:

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
Standarddeviation 0.49889090.1983739 0.16328831 0.12729774 0.09863501
ProportionofVariance 0.71380550.11285940.076467870.04647401 0.02790172
CumulativeProportion0.71380550.82666490.903132790.949606810.97750852
Comp.6
Standard deviation 0.08855735
ProportionofVariance0.02249148
CumulativeProportion1.00000000



<u>Year2016</u>:

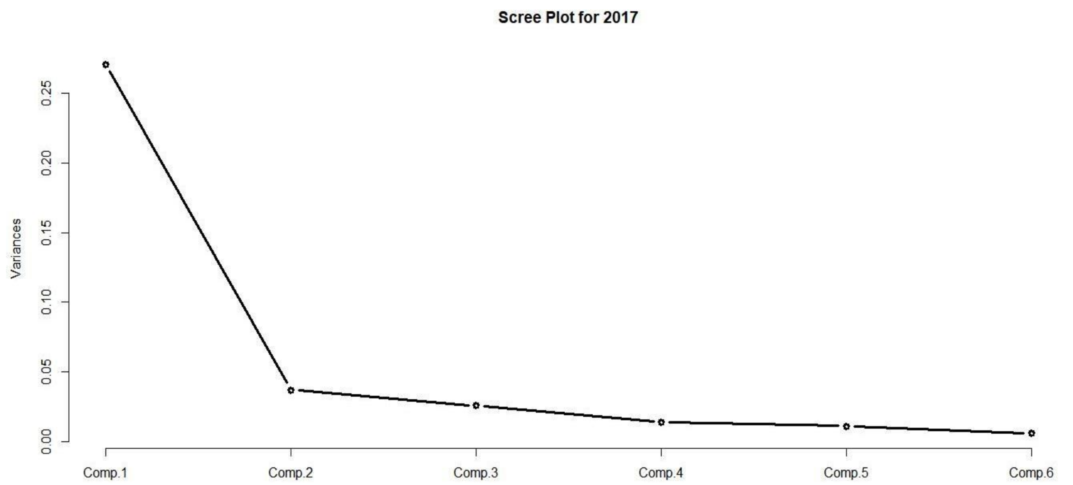
Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
Standarddeviation 0.5014984 0.1907981 0.1576398 0.11721940 0.09894869
ProportionofVariance 0.73289520.10608420.07241600.040040710.02853139
CumulativeProportion0.73289520.83897940.91139540.951436100.97996748
Comp.6
Standard deviation 0.08291181
ProportionofVariance0.02003252
CumulativeProportion1.00000000



<u>Year2017</u>:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standarddeviation	**0.5203477**	**0.1924917**	**0.16015820**	**0.11862870**	**0.10478154**
ProportionofVariance	**0.7429867**	**0.1016758**	**0.07038695**	**0.03861654**	**0.03012752**
CumulativeProportion	**0.7429867**	**0.8446625**	**0.91504948**	**0.95366602**	**0.98379353**

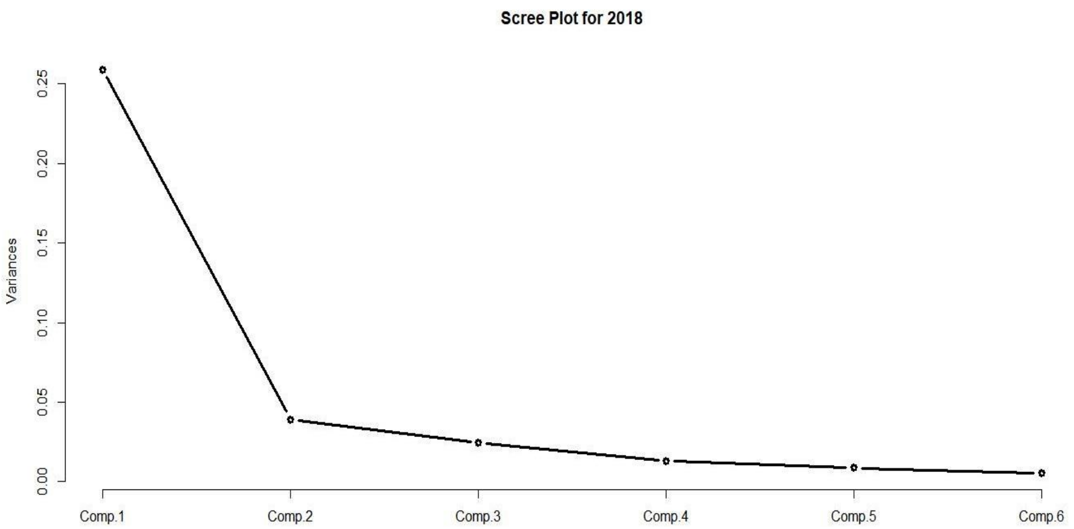
	Comp.6
Standard deviation	**0.07685060**
ProportionofVariance	**0.01620647**
CumulativeProportion	**1.00000000**



Year2018:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standarddeviation	0.5089955	0.1978356	0.15599224	0.11383938	0.09281554
ProportionofVariance	0.7413187	0.1119917	0.06962788	0.03708192	0.02465009
CumulativeProportion	0.7413187	0.8533105	0.92293834	0.96002026	0.98467035

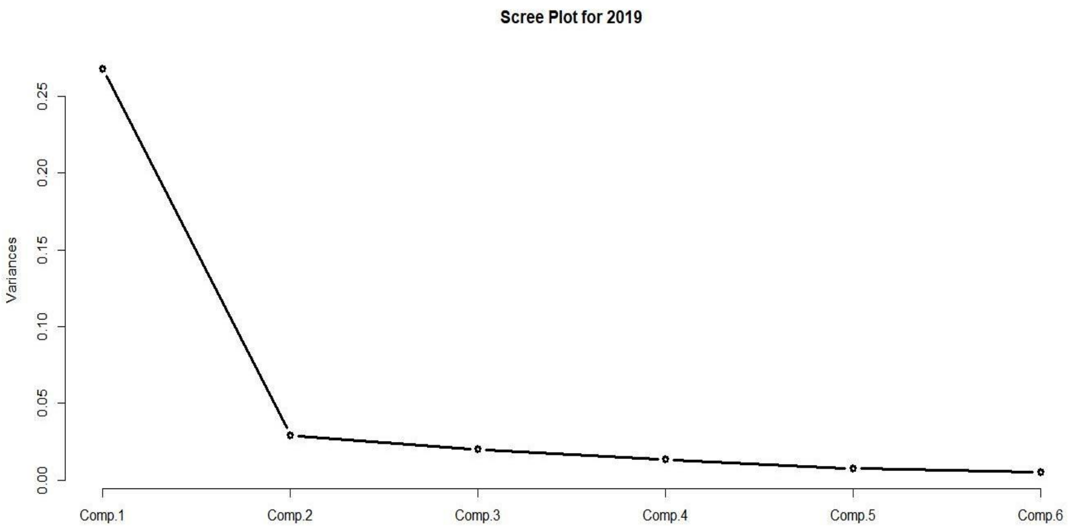
	Comp.6
Standard deviation	0.07319436
ProportionofVariance	0.01532965
CumulativeProportion	1.00000000



Year2019:

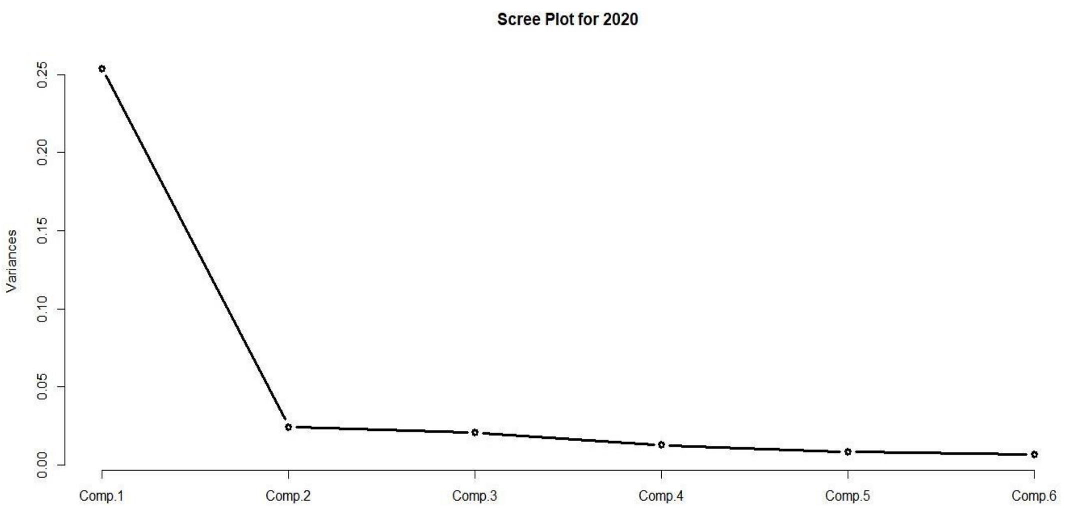
	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standarddeviation	0.5179183	0.17061518	0.14106621	0.11547628	0.08649536
ProportionofVariance	0.7816236	0.08482238	0.05798573	0.03885623	0.02180021
CumulativeProportion	0.7816236	0.86644600	0.92443173	0.96328796	0.98508817

	Comp.6
Standard deviation	0.07153653
ProportionofVariance	0.01491183
CumulativeProportion	1.00000000



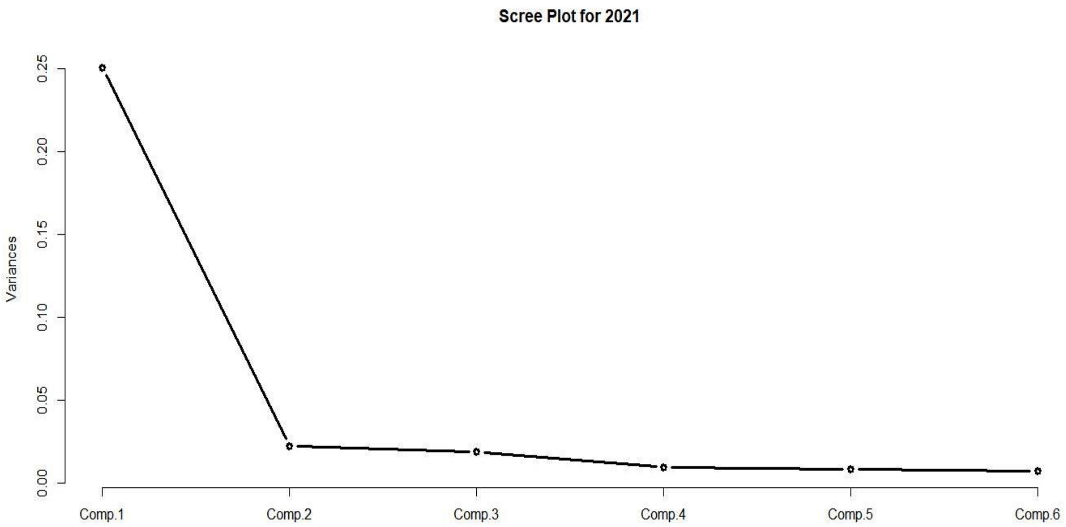
Year2020:

```
                      Comp.1     Comp.2      Comp.3      Comp.4     Comp.5
Standarddeviation        0.5039960 0.15565393 0.14366330 0.11214697 0.09033043
ProportionofVariance 0.77869140.074273070.063270730.038555490.02501378
CumulativeProportion0.77869140.852964500.916235230.954790720.97980450
                         Comp.6
Standard deviation       0.08116555
ProportionofVariance0.02019550
CumulativeProportion1.00000000
```



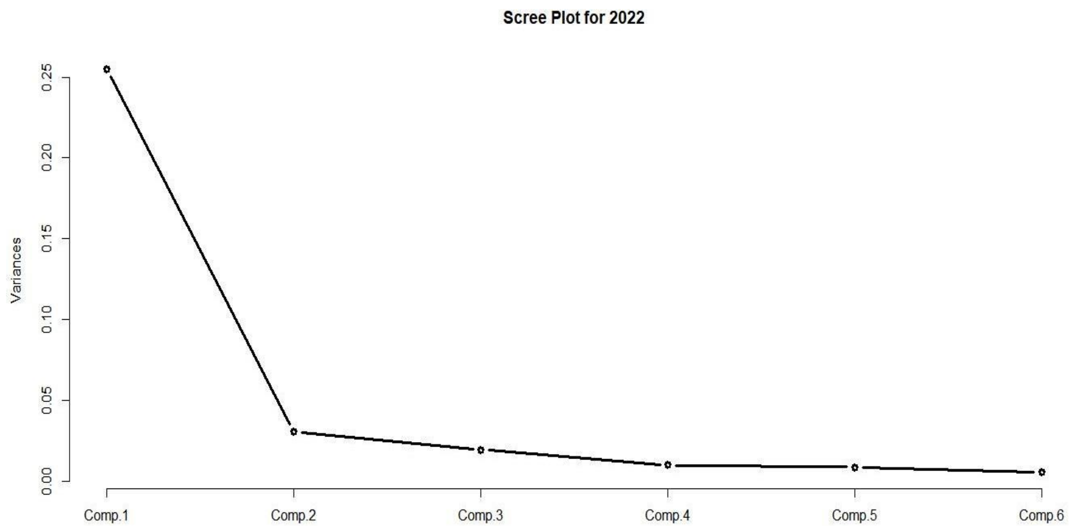
<u>Year2021</u>:

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
Standarddeviation 0.5008617 0.14815401 0.13690788 0.09678221 0.08898008
ProportionofVariance 0.79431890.069500220.059349380.029658580.02506946
CumulativeProportion0.79431890.863819130.923168510.952827100.97789655
Comp.6
Standard deviation 0.08355075
ProportionofVariance0.02210345
CumulativeProportion1.00000000



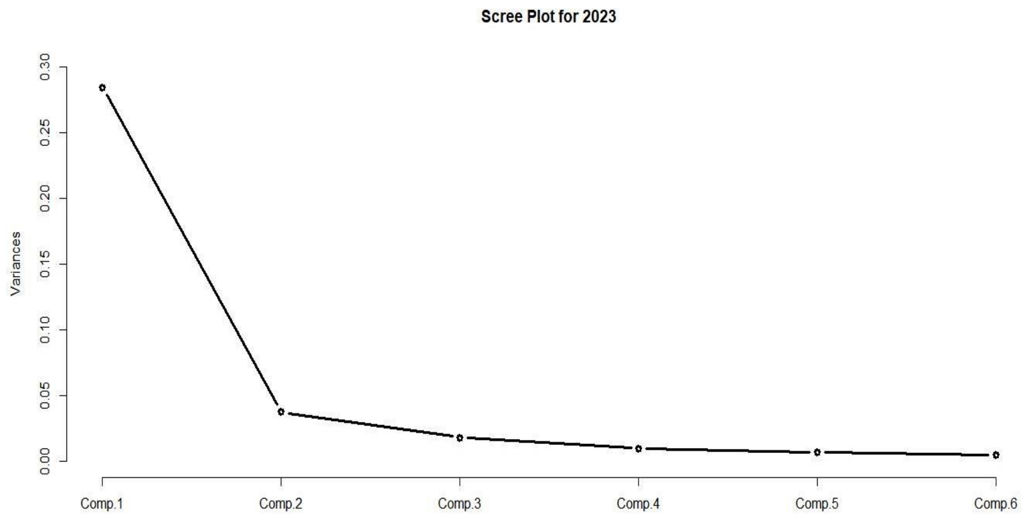
<u>Year2022</u>:

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
Standarddeviation 0.5049723 0.17515759 0.13950858 0.09980722 0.09356069
ProportionofVariance 0.77393680.093116860.059070720.030233910.02656790
CumulativeProportion0.77393680.867053680.926124400.956358310.98292622
Comp.6
Standard deviation 0.07500318
ProportionofVariance0.01707378
CumulativeProportion1.00000000



<u>Year2023</u>:

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
Standarddeviation 0.5331287 0.1933362 0.1346613 0.09882239 0.08196053
ProportionofVariance 0.78769310.10359030.05025490.027064730.01861669
CumulativeProportion0.78769310.89128340.94153830.968603020.98721971
Comp.6
Standard deviation 0.06790847
ProportionofVariance0.01278029
CumulativeProportion1.00000000



<u>Overall Interpretation</u>:-It is to be noted that for all the years,the first four principalcomponentsexplain95%ofthetotalvariation,nowthenoofpredictors finally obtained by stepwise regression is 4 ,and the no. of principal components is also 4,which means there is no significant dimension reduction.It might also indicate that the covariates are linearly independent,i.e multicollinearity is absent among the covariates.(as such Σ is diagonal matrix).

CONCLUSION

From the project report the main conclusions that can be drawn that GDP per capita,social support,healthy life expectancy and freedom to make life choices arethemain influentialfactorsforthehappinessscoreforanynation.Thesefour are thekeystonebehindtheprediction ofhappinessscore foranyupcomingyear in efficient way. These covariates that we have worked with are independent i.e. they have no control over each other.They independently work for any nation and make joint influence on the happiness score. Mainly the African countries and the Asian countries and a few European countries are providing as influential data points each year,but the countries are not the same i.e. they are varying over the years.The happiness score for the eastern countries is always lower than the western countries.The European countries are maintaining a sustained increase in the happiness score overall over the years,though the countries belonging to North America and Oceania are always around 6.5 and 7 respectively.African countries are always in the least happy countries,though in recent year 2023,Afghanisthan has the lowest happiness score level in all years among all the countries.Finland has maintained the top position as the happiest countryforsixyears continuously,andmaintainedthehappinesslevel morethan 7.6.Only European countries are still dominating as the maximum happy countries.In2015and2016,CanadawasthereandrecentlyIsraelispresent.This has been the first time that Israel has come into this list,it has made a remarkable progress in this happiness level over the years by crossing all other European countries and has taken 5thposition in 2023.

Acknowledgement

Inordertocompletethereport,Iamhighlyacknowledgedto my supervisor Professor Biswajit Basak, of Department of Statistics,Sister Nivedita University.He has helped me in completing my report work in efficient manner.He has also helped me with books and correction of my report errors.I have tried to follow his advice to make sure that my work is flawless and perfect in it’s own manner.However,any feedback regarding any inadequacy of the project work is always acceptable.

Reference

- **AnIntroductiontoMultivariateStatisticalAnalysis (T.W.Anderson)**
- **RegressionAnalysisbyExample(SampritChatterjeeand Ali S.Hadi)**
- **AnOutlineOfStatisticalTheory(VOL.1) (Goon,Gupta,Dasgupta)**
- **www.kaggle.com**

Appendix

Rcode:

```
#Year2015#
#----------

rm(list=ls())
data1=read.csv("C:/Users/USER/Documents/MSCDocs/RSProjectMsc/Year
2015.csv",header=TRUE)
#data1
attach(data1)
Data=matrix(as.numeric(unlist(data1[,-c(1,9,10)])),ncol=7)
colnames(Data)=c("x1","x2","x3","x4","x5","x6","x7");Data


#RegressionDiagnostics(ReducedModel)

x1=Data[,1];x1
x2=Data[,2];x2
x3=Data[,3];x3
x4=Data[,4];x4
x5=Data[,5];x5
model=lm(x1~x2+x3+x4+x5);summary(model)
library(olsrr)
ols_plot_resid_stand(model)
ols_plot_dffits(model)
ols_plot_cooksd_chart(model)




R=cor(Data);R
R11=det(R[-1,][,-1])
R12=-det(R[-1,][,-2])
R13=det(R[-1,][,-3])
R14=-det(R[-1,][,-4])
R15=det(R[-1,][,-5])
R16=-det(R[-1,][,-6])
R17=det(R[-1,][,-7])
R22=det(R[-2,][,-2])
```

```
R33=det(R[-3,][,-3])
R44=det(R[-4,][,-4])
R55=det(R[-5,][,-5])
R66=det(R[-6,][,-6])
R77=det(R[-7,][,-7])
Mult_corr=sqrt(1-(det(R)/R11));Mult_corr
PC12=-R12/(sqrt(R11*R22));PC12
PC13=R13/(sqrt(R11*R33));PC13
PC14=-R14/(sqrt(R11*R44));PC14
PC15=R15/(sqrt(R11*R55));PC15
PC16=-R16/(sqrt(R11*R66));PC16
PC17=R17/(sqrt(R11*R77));PC17

#x6discardedasa variable

summary(lm(x1~x2+x3+x4+x5+x7))
Data1=Data[,-6];Data1
R1=cor(Data1);R1
R1.11=det(R1[-1,][,-1])
R1.12=-det(R1[-1,][,-2])
R1.13=det(R1[-1,][,-3])
R1.14=-det(R1[-1,][,-4])
R1.15=det(R1[-1,][,-5])
R1.16=-det(R1[-1,][,-6])
R1.22=det(R1[-2,][,-2])
R1.33=det(R1[-3,][,-3])
R1.44=det(R1[-4,][,-4])
R1.55=det(R1[-5,][,-5])
R1.66=det(R1[-6,][,-6])
Mult_corr1=sqrt(1-(det(R1)/R1.11));Mult_corr1
PC1.12=-R1.12/(sqrt(R1.11*R1.22));PC1.12
PC1.13=R1.13/(sqrt(R1.11*R1.33));PC1.13
PC1.14=-R1.14/(sqrt(R1.11*R1.44));PC1.14
PC1.15=R1.15/(sqrt(R1.11*R1.55));PC1.15
PC1.16=-R1.16/(sqrt(R1.11*R1.66));PC1.16

#x7discardedasa variable

summary(lm(x1~x2+x3+x4+x5))
Data2=Data1[,-6];Data2
R2=cor(Data2);R2
R2.11=det(R2[-1,][,-1])
R2.12=-det(R2[-1,][,-2])
R2.13=det(R2[-1,][,-3])
R2.14=-det(R2[-1,][,-4])
R2.15=det(R2[-1,][,-5])
R2.22=det(R2[-2,][,-2])
R2.33=det(R2[-3,][,-3])
R2.44=det(R2[-4,][,-4])
```

```
R2.55=det(R2[-5,][,-5])
Mult_corr2=sqrt(1-(det(R2)/R2.11));Mult_corr2
PC2.12=-R2.12/(sqrt(R2.11*R2.22));PC2.12
PC2.13=R2.13/(sqrt(R2.11*R2.33));PC2.13
PC2.14=-R2.14/(sqrt(R2.11*R2.44));PC2.14
PC2.15=R2.15/(sqrt(R2.11*R2.55));PC2.15

#x4 discarded as a variable

summary(lm(x1~x2+x3+x5))

Data3=Data2[,-4];Data3
R3=cor(Data3);R3
R3.11=det(R3[-1,][,-1])
R3.12=-det(R3[-1,][,-2])
R3.13=det(R3[-1,][,-3])
R3.14=-det(R3[-1,][,-4])
R3.22=det(R3[-2,][,-2])
R3.33=det(R3[-3,][,-3])
R3.44=det(R3[-4,][,-4])
Mult_corr3=sqrt(1-(det(R3)/R3.11));Mult_corr3
PC3.12=-R3.12/(sqrt(R3.11*R3.22));PC3.12
PC3.13=R3.13/(sqrt(R3.11*R3.33));PC3.13
PC3.14=-R3.14/(sqrt(R3.11*R3.44));PC3.14

#x5isdiscardedasavariable

summary(lm(x1~x2+x3))

#End of step wise regression

#PCA
new_data=Data[,-1];new_data
pca=princomp(new_data,cor=FALSE)
summary(pca)
pca$loadings
plot(pca,type="lines",lwd=3,main="ScreePlotfor2015")

#Plotting

x=aggregate(happiness_score,by=list(continent),FUN=mean)
x
happy_index=as.vector(unlist(x[2]));happy_index
barplot(happy_index,names.arg=c("Africa","Asia","Europe","North
America","Oceania","South
America"),ylim=c(0,11),col=1:6,xlab="Continents",ylab="Happiness
Score",main="Representation of world happiness scores in 2015")
legend(11,c("Africa-4.319804","Asia-5.446438","Europe-5.927980","North
America-6.699750","Oceania-7.285","South America-6.1465"),fill=1:6)
```

```
top=head(data1);top
bottom=tail(data1);bottom
barplot(top[,2],main="Top 6 happiest countries in 2015",names.arg=c("Switzerland","Iceland","Denmark","Norway","Canada", "Finland"),ylab="Happiness Score",col=2:7,ylim=c(0,12.5))
legend(12.5,c("Switzerland-7.587","Iceland-7.561","Denmark-7.527","Norway-7.522","Canada-7.427","Finland-7.406"),fill=2:7)
barplot(bottom[,2],main="6 least happy countries in 2015",names.arg=c("Afganisthan","Rwanda","Benin","Syria","Burundi","Togo"),ylab="Happiness Score",col=2:7,ylim=c(0,7))
legend(7,c("Afganisthan-3.575","Rwanda-3.465","Benin-3.340","Syria-3.006","Burundi-2.905","Togo-2.839"),fill=2:7)



#Year2016#
#----------
rm(list=ls())
data1=read.csv("C:/Users/USER/Documents/MSCDocs/RSProjectMsc/Year2016.csv",header=TRUE)
#data1
attach(data1)
Data=matrix(as.numeric(unlist(data1[,-c(1,9,10)])),ncol=7)
colnames(Data)=c("x1","x2","x3","x4","x5","x6","x7");Data

x1=Data[,1];x1
x2=Data[,2];x2
x3=Data[,3];x3
x4=Data[,4];x4
x5=Data[,5];x5
x6=Data[,6];x6
x7=Data[,7];x7
model=lm(x1~x2+x3+x4+x5+x6+x7);summary(model)

R=cor(Data);R
R11=det(R[-1,][,-1])
R12=det(R[-1,][,-2])
R13=det(R[-1,][,-3])
R14=det(R[-1,][,-4])
R15=det(R[-1,][,-5])
R16=det(R[-1,][,-6])
R17=det(R[-1,][,-7])
R22=det(R[-2,][,-2])
R33=det(R[-3,][,-3])
R44=det(R[-4,][,-4])
R55=det(R[-5,][,-5])
R66=det(R[-6,][,-6])
R77=det(R[-7,][,-7])
Mult_corr=sqrt(1-(det(R)/R11));Mult_corr
```

```
PC12=-R12/(sqrt(R11*R22));PC12
PC13=R13/(sqrt(R11*R33));PC13
PC14=-R14/(sqrt(R11*R44));PC14
PC15=R15/(sqrt(R11*R55));PC15
PC16=-R16/(sqrt(R11*R66));PC16
PC17=R17/(sqrt(R11*R77));PC17

#x6 discarded as a variable

summary(lm(x1~x2+x3+x4+x5+x7))

Data1=Data[,-6];Data1
R1=cor(Data1);R1
R1.11=det(R1[-1,][,-1])
R1.12=det(R1[-1,][,-2])
R1.13=det(R1[-1,][,-3])
R1.14=det(R1[-1,][,-4])
R1.15=det(R1[-1,][,-5])
R1.16=det(R1[-1,][,-6])
R1.22=det(R1[-2,][,-2])
R1.33=det(R1[-3,][,-3])
R1.44=det(R1[-4,][,-4])
R1.55=det(R1[-5,][,-5])
R1.66=det(R1[-6,][,-6])
Mult_corr1=sqrt(1-(det(R1)/R1.11));Mult_corr1
PC1.12=-R1.12/(sqrt(R1.11*R1.22));PC1.12
PC1.13=R1.13/(sqrt(R1.11*R1.33));PC1.13
PC1.14=-R1.14/(sqrt(R1.11*R1.44));PC1.14
PC1.15=R1.15/(sqrt(R1.11*R1.55));PC1.15
PC1.16=-R1.16/(sqrt(R1.11*R1.66));PC1.16

#x7discardedasa variable

summary(lm(x1~x2+x3+x4+x5))
Data2=Data1[,-6];Data2
R2=cor(Data2);R2
R2.11=det(R2[-1,][,-1])
R2.12=det(R2[-1,][,-2])
R2.13=det(R2[-1,][,-3])
R2.14=det(R2[-1,][,-4])
R2.15=det(R2[-1,][,-5])
R2.22=det(R2[-2,][,-2])
R2.33=det(R2[-3,][,-3])
R2.44=det(R2[-4,][,-4])
R2.55=det(R2[-5,][,-5])
Mult_corr2=sqrt(1-(det(R2)/R2.11));Mult_corr2
PC2.12=-R2.12/(sqrt(R2.11*R2.22));PC2.12
PC2.13=R2.13/(sqrt(R2.11*R2.33));PC2.13
PC2.14=-R2.14/(sqrt(R2.11*R2.44));PC2.14
```

```
PC2.15=R2.15/(sqrt(R2.11*R2.55));PC2.15

#x2 discarded as a variable

summary(lm(x1~x3+x4+x5))

#Endofstepwiseregression

#PCA
new_data=Data[,-1];new_data
pca=princomp(new_data,cor=FALSE)
summary(pca)
pca$loadings
plot(pca,type="lines",lwd=3,main="ScreePlotfor2016")

#Plotting

x=aggregate(happiness_score,by=list(continent),FUN=mean)
x
happy_index=as.vector(unlist(x[2]));happy_index
barplot(happy_index,names.arg=c("Europe","Africa","Asia","North
America","Oceania","South
America"),ylim=c(0,12),col=1:6,xlab="Continents",ylab="Happiness
Score",main="Representation of world happiness scores in 2016")
legend(12,c("Europe-5.978489","Africa-4.337353","Asia-5.429364","North
America-6.725","Oceania-7.3235","South America-6.093182"),fill=1:6)


top=head(data1);top
bottom=tail(data1);bottom
barplot(top[,2],main="Top 6 happiest countries in
2016",names.arg=c("Denmark","Switzerland","Iceland","Norway","Finland",
"Canada"),ylab="Happiness Score",col=2:7,ylim=c(0,12.5))
legend(12.5,c("Denmark-7.526","Switzerland-7.509","Iceland-7.501","Norway-
7.498","Finland-7.413","Canada-7.404"),fill=2:7)
barplot(bottom[,2],main="6 least happy countries in
2016",names.arg=c("Rwanda","Benin","Afganisthan","Togo","Syria","Burun
di"),ylab="Happiness Score",col=2:7,ylim=c(0,7))
legend(7,c("Rwanda-3.515","Benin-3.484","Afganisthan-3.360","Togo-
3.303","Syria-3.069","Burundi-2.905"),fill=2:7)


#RegressionDiagnostics

model1=lm(x1~x2+x3+x4+x5)
library(olsrr)
ols_plot_resid_stand(model1)
ols_plot_dffits(model1)
ols_plot_cooksd_chart(model1)
```

```
#Year2017#
#----------
rm(list=ls())
data1=read.csv("C:/Users/USER/Documents/MSCDocs/RSProjectMsc/Year
2017.csv",header=TRUE)
#data1
attach(data1)
Data=matrix(as.numeric(unlist(data1[,-c(1,9,10)])),ncol=7)
colnames(Data)=c("x1","x2","x3","x4","x5","x6","x7");Data
x1=Data[,1];x1
x2=Data[,2];x2
x3=Data[,3];x3
x4=Data[,4];x4
x5=Data[,5];x5
x6=Data[,6];x6
x7=Data[,7];x7
model=lm(x1~x2+x3+x4+x5+x6+x7);summary(model)
R=cor(Data);R
R11=det(R[-1,][,-1])
R12=det(R[-1,][,-2])
R13=det(R[-1,][,-3])
R14=det(R[-1,][,-4])
R15=det(R[-1,][,-5])
R16=det(R[-1,][,-6])
R17=det(R[-1,][,-7])
R22=det(R[-2,][,-2])
R33=det(R[-3,][,-3])
R44=det(R[-4,][,-4])
R55=det(R[-5,][,-5])
R66=det(R[-6,][,-6])
R77=det(R[-7,][,-7])
Mult_corr=sqrt(1-(det(R)/R11));Mult_corr
PC12=-R12/(sqrt(R11*R22));PC12
PC13=R13/(sqrt(R11*R33));PC13
PC14=-R14/(sqrt(R11*R44));PC14
PC15=R15/(sqrt(R11*R55));PC15
PC16=-R16/(sqrt(R11*R66));PC16
PC17=R17/(sqrt(R11*R77));PC17

#x6discardedasa variable

summary(lm(x1~x2+x3+x4+x5+x7))
Data1=Data[,-6];Data1
R1=cor(Data1);R1
R1.11=det(R1[-1,][,-1])
R1.12=det(R1[-1,][,-2])
```

```
R1.13=det(R1[-1,][,-3])
R1.14=det(R1[-1,][,-4])
R1.15=det(R1[-1,][,-5])
R1.16=det(R1[-1,][,-6])
R1.22=det(R1[-2,][,-2])
R1.33=det(R1[-3,][,-3])
R1.44=det(R1[-4,][,-4])
R1.55=det(R1[-5,][,-5])
R1.66=det(R1[-6,][,-6])
Mult_corr1=sqrt(1-(det(R1)/R1.11));Mult_corr1
PC1.12=-R1.12/(sqrt(R1.11*R1.22));PC1.12
PC1.13=R1.13/(sqrt(R1.11*R1.33));PC1.13
PC1.14=-R1.14/(sqrt(R1.11*R1.44));PC1.14
PC1.15=R1.15/(sqrt(R1.11*R1.55));PC1.15
PC1.16=-R1.16/(sqrt(R1.11*R1.66));PC1.16

#x7discardedasa variable

summary(lm(x1~x2+x3+x4+x5))
Data2=Data1[,-6];Data2
R2=cor(Data2);R2
R2.11=det(R2[-1,][,-1])
R2.12=det(R2[-1,][,-2])
R2.13=det(R2[-1,][,-3])
R2.14=det(R2[-1,][,-4])
R2.15=det(R2[-1,][,-5])
R2.22=det(R2[-2,][,-2])
R2.33=det(R2[-3,][,-3])
R2.44=det(R2[-4,][,-4])
R2.55=det(R2[-5,][,-5])
Mult_corr2=sqrt(1-(det(R2)/R2.11));Mult_corr2
PC2.12=-R2.12/(sqrt(R2.11*R2.22));PC2.12
PC2.13=R2.13/(sqrt(R2.11*R2.33));PC2.13
PC2.14=-R2.14/(sqrt(R2.11*R2.44));PC2.14
PC2.15=R2.15/(sqrt(R2.11*R2.55));PC2.15

#x4 discarded as a variable

summary(lm(x1~x2+x3+x5))

#Endofstepwiseregression

#PCA
new_data=Data[,-1];new_data
pca=princomp(new_data,cor=FALSE)
summary(pca)
pca$loadings
plot(pca,type="lines",lwd=3,main="ScreePlotfor2017")
```

```
#Plotting

x=aggregate(happiness_score,by=list(Continent),FUN=mean)
x
happy_index=as.vector(unlist(x[2]));happy_index
barplot(happy_index,names.arg=c("Africa","Asia","Europe","North
America","Oceania","South
America"),ylim=c(0,12),col=1:6,xlab="Continents",ylab="Happiness
Score",main="Representation of world happiness scores in 2017")
legend(12,c("Africa-4.2395","Asia-5.303829","Europe-6.043159","North
America-6.017462","Oceania-7.299","SouthAmerica-6.104909"),fill=1:6)

top=head(data1);top
bottom=tail(data1);bottom
barplot(top[,2],main="Top 6 happiest countries in
2017",names.arg=c("Norway","Denmark","Iceland","Switzerland","Finland",
"Netherlands"),ylab="Happiness Score",col=2:7,ylim=c(0,12.5))
legend(12.5,c("Norway-7.537","Denmark-7.522","Iceland-7.504","Switzerland-
7.494","Finland-7.413","Netherlands-7.377"),fill=2:7)
barplot(bottom[,2],main="6 least happy countries in
2017",names.arg=c("Togo","Rwanda","Syria","Tanzania","Burundi","Central
African Republic"),ylab="Happiness Score",col=2:7,ylim=c(0,7))
legend(7,c("Togo-3.495","Rwanda-3.471","Syria-3.462","Tanzania-
3.349","Burundi-2.905","Central African Republic-2.693"),fill=2:7)


#RegressionDiagnostics

library(olsrr)
model1=lm(x1~x2+x3+x4+x5)
ols_plot_resid_stand(model1)
ols_plot_dffits(model1)
ols_plot_cooksd_chart(model1)


#Year2018#
#----------
rm(list=ls())
data1=read.csv("C:/Users/USER/Documents/MSCDocs/RSProjectMsc/Year
2018.csv",header=TRUE)
#data1
attach(data1)
Data=matrix(as.numeric(unlist(data1[,-c(1,9,10)])),ncol=7)
colnames(Data)=c("x1","x2","x3","x4","x5","x6","x7");Data
x1=Data[,1];x1
x2=Data[,2];x2
x3=Data[,3];x3
x4=Data[,4];x4
x5=Data[,5];x5
```

```
x6=Data[,6];x6
x7=Data[,7];x7
model=lm(x1~x2+x3+x4+x5+x6+x7);summary(model)
R=cor(Data);R
R11=det(R[-1,][,-1])
R12=det(R[-1,][,-2])
R13=det(R[-1,][,-3])
R14=det(R[-1,][,-4])
R15=det(R[-1,][,-5])
R16=det(R[-1,][,-6])
R17=det(R[-1,][,-7])
R22=det(R[-2,][,-2])
R33=det(R[-3,][,-3])
R44=det(R[-4,][,-4])
R55=det(R[-5,][,-5])
R66=det(R[-6,][,-6])
R77=det(R[-7,][,-7])
Mult_corr=sqrt(1-(det(R)/R11));Mult_corr
PC12=-R12/(sqrt(R11*R22));PC12
PC13=R13/(sqrt(R11*R33));PC13
PC14=-R14/(sqrt(R11*R44));PC14
PC15=R15/(sqrt(R11*R55));PC15
PC16=-R16/(sqrt(R11*R66));PC16
PC17=R17/(sqrt(R11*R77));PC17

#x7discardedasa variable

summary(lm(x1~x2+x3+x4+x5+x6))
Data1=Data[,-7];Data1
R1=cor(Data1);R1
R1.11=det(R1[-1,][,-1])
R1.12=det(R1[-1,][,-2])
R1.13=det(R1[-1,][,-3])
R1.14=det(R1[-1,][,-4])
R1.15=det(R1[-1,][,-5])
R1.16=det(R1[-1,][,-6])
R1.22=det(R1[-2,][,-2])
R1.33=det(R1[-3,][,-3])
R1.44=det(R1[-4,][,-4])
R1.55=det(R1[-5,][,-5])
R1.66=det(R1[-6,][,-6])
Mult_corr1=sqrt(1-(det(R1)/R1.11));Mult_corr1
PC1.12=-R1.12/(sqrt(R1.11*R1.22));PC1.12
PC1.13=R1.13/(sqrt(R1.11*R1.33));PC1.13
PC1.14=-R1.14/(sqrt(R1.11*R1.44));PC1.14
PC1.15=R1.15/(sqrt(R1.11*R1.55));PC1.15
PC1.16=-R1.16/(sqrt(R1.11*R1.66));PC1.16

#x6discardedasa variable
```

```
summary(lm(x1~x2+x3+x4+x5))
Data2=Data1[,-6];Data2
R2=cor(Data2);R2
R2.11=det(R2[-1,][,-1])
R2.12=det(R2[-1,][,-2])
R2.13=det(R2[-1,][,-3])
R2.14=det(R2[-1,][,-4])
R2.15=det(R2[-1,][,-5])
R2.22=det(R2[-2,][,-2])
R2.33=det(R2[-3,][,-3])
R2.44=det(R2[-4,][,-4])
R2.55=det(R2[-5,][,-5])
Mult_corr2=sqrt(1-(det(R2)/R2.11));Mult_corr2
PC2.12=-R2.12/(sqrt(R2.11*R2.22));PC2.12
PC2.13=R2.13/(sqrt(R2.11*R2.33));PC2.13
PC2.14=-R2.14/(sqrt(R2.11*R2.44));PC2.14
PC2.15=R2.15/(sqrt(R2.11*R2.55));PC2.15

#x4discardedasa variable

summary(lm(x1~x2+x3+x5))
Data3=Data2[,-4];Data3
R3=cor(Data3);R3
R3.11=det(R3[-1,][,-1])
R3.12=det(R3[-1,][,-2])
R3.13=det(R3[-1,][,-3])
R3.14=det(R3[-1,][,-4])
R3.22=det(R3[-2,][,-2])
R3.33=det(R3[-3,][,-3])
R3.44=det(R3[-4,][,-4])
Mult_corr3=sqrt(1-(det(R3)/R3.11));Mult_corr3
PC3.12=-R3.12/(sqrt(R3.11*R3.22));PC3.12
PC3.13=R3.13/(sqrt(R3.11*R3.33));PC3.13
PC3.14=-R3.14/(sqrt(R3.11*R3.44));PC3.14

#x3 is discarded

summary(lm(x1~x2+x5))

#Endofstepwiseregression

#PCA
new_data=Data[,-1];new_data
pca=princomp(new_data,cor=FALSE)
summary(pca)
pca$loadings
plot(pca,type="lines",lwd=3,main="ScreePlotfor2018")

#Plotting
```

```
x=aggregate(happiness_score,by=list(continent),FUN=mean)
x
happy_index=as.vector(unlist(x[2]));happy_index
barplot(happy_index,names.arg=c("Africa","Asia","Europe","North
America","Oceania","South
America"),ylim=c(0,12),col=1:6,xlab="Continents",ylab="Happiness
Score",main="Representation of world happiness scores in 2018")
legend(12,c("Africa-4.296744","Asia-5.274891","Europe-6.189025","North
America-6.5345","Oceania-7.298","South America-5.782842"),fill=1:6)

top=head(data1);top
bottom=tail(data1);bottom
barplot(top[,2],main="Top 6 happiest countries in
2018",names.arg=c("Finland","Norway","Denmark","Iceland","Switzerland",
"Netherlands"),ylab="Happiness Score",col=2:7,ylim=c(0,12.5))
legend(12.5,c("Finland-7.632","Norway-7.594","Denmark-7.555","Iceland-
7.495","Switzerland-7.487","Netherlands-7.441"),fill=2:7)
barplot(bottom[,2],main="6 least happy countries in
2018",names.arg=c("Rwanda","Yemen","Tanzania","South Sudan","Central
African Republic","Burundi"),ylab="Happiness Score",col=2:7,ylim=c(0,7))
legend(7,c("Rwanda-3.408","Yemen-3.355","Tanzania-3.303","South Sudan-
3.254","Central African Republic-3.083","Burundi-2.905"),fill=2:7)


#RegressionDiagnostics

model1=lm(x1~x2+x3+x4+x5)
library(olsrr)
ols_plot_resid_stand(model1)
ols_plot_dffits(model1)
ols_plot_cooksd_chart(model1)




#Year2019#
#----------
rm(list=ls())
data1=read.csv("C:/Users/USER/Documents/MSCDocs/RSProjectMsc/Year
2019.csv",header=TRUE)
#data1
attach(data1)
Data=matrix(as.numeric(unlist(data1[,-c(1,9,10)])),ncol=7)
colnames(Data)=c("x1","x2","x3","x4","x5","x6","x7");Data
x1=Data[,1];x1
x2=Data[,2];x2
x3=Data[,3];x3
x4=Data[,4];x4
x5=Data[,5];x5
```

```
x6=Data[,6];x6
x7=Data[,7];x7
model=lm(x1~x2+x3+x4+x5+x6+x7);summary(model)
R=cor(Data);R
R11=det(R[-1,][,-1])
R12=det(R[-1,][,-2])
R13=det(R[-1,][,-3])
R14=det(R[-1,][,-4])
R15=det(R[-1,][,-5])
R16=det(R[-1,][,-6])
R17=det(R[-1,][,-7])
R22=det(R[-2,][,-2])
R33=det(R[-3,][,-3])
R44=det(R[-4,][,-4])
R55=det(R[-5,][,-5])
R66=det(R[-6,][,-6])
R77=det(R[-7,][,-7])
Mult_corr=sqrt(1-(det(R)/R11));Mult_corr
PC12=-R12/(sqrt(R11*R22));PC12
PC13=R13/(sqrt(R11*R33));PC13
PC14=-R14/(sqrt(R11*R44));PC14
PC15=R15/(sqrt(R11*R55));PC15
PC16=-R16/(sqrt(R11*R66));PC16
PC17=R17/(sqrt(R11*R77));PC17

#x6discardedasa variable

summary(lm(x1~x2+x3+x4+x5+x7))
Data1=Data[,-6];Data1
R1=cor(Data1);R1
R1.11=det(R1[-1,][,-1])
R1.12=det(R1[-1,][,-2])
R1.13=det(R1[-1,][,-3])
R1.14=det(R1[-1,][,-4])
R1.15=det(R1[-1,][,-5])
R1.16=det(R1[-1,][,-6])
R1.22=det(R1[-2,][,-2])
R1.33=det(R1[-3,][,-3])
R1.44=det(R1[-4,][,-4])
R1.55=det(R1[-5,][,-5])
R1.66=det(R1[-6,][,-6])
Mult_corr1=sqrt(1-(det(R1)/R1.11));Mult_corr1
PC1.12=-R1.12/(sqrt(R1.11*R1.22));PC1.12
PC1.13=R1.13/(sqrt(R1.11*R1.33));PC1.13
PC1.14=-R1.14/(sqrt(R1.11*R1.44));PC1.14
PC1.15=R1.15/(sqrt(R1.11*R1.55));PC1.15
PC1.16=-R1.16/(sqrt(R1.11*R1.66));PC1.16

#x7discardedasa variable
```

```
summary(lm(x1~x2+x3+x4+x5))
Data2=Data1[,-6];Data2
R2=cor(Data2);R2
R2.11=det(R2[-1,][,-1])
R2.12=det(R2[-1,][,-2])
R2.13=det(R2[-1,][,-3])
R2.14=det(R2[-1,][,-4])
R2.15=det(R2[-1,][,-5])
R2.22=det(R2[-2,][,-2])
R2.33=det(R2[-3,][,-3])
R2.44=det(R2[-4,][,-4])
R2.55=det(R2[-5,][,-5])
Mult_corr2=sqrt(1-(det(R2)/R2.11));Mult_corr2
PC2.12=-R2.12/(sqrt(R2.11*R2.22));PC2.12
PC2.13=R2.13/(sqrt(R2.11*R2.33));PC2.13
PC2.14=-R2.14/(sqrt(R2.11*R2.44));PC2.14
PC2.15=R2.15/(sqrt(R2.11*R2.55));PC2.15

#x4 discarded as a variable

summary(lm(x1~x2+x3+x5))

#Endofstepwiseregression

#PCA
new_data=Data[,-1];new_data
pca=princomp(new_data,cor=FALSE)
summary(pca)
pca$loadings
plot(pca,type="lines",lwd=3,main="ScreePlotfor2019")

#Plotting

x=aggregate(happiness_score,by=list(continent),FUN=mean)
x
happy_index=as.vector(unlist(x[2]));happy_index
barplot(happy_index,names.arg=c("Africa","Asia","Europe","North
America","Oceania","South
America"),ylim=c(0,12),col=1:6,xlab="Continents",ylab="Happiness
Score",main="Representation of world happiness scores in 2019")
legend(12,c("Africa-4.347535","Asia-5.296261","Europe-6.223775","North
America-6.726","Oceania-7.2765","South America-5.834619"),fill=1:6)

top=head(data1);top
bottom=tail(data1);bottom
barplot(top[,2],main="Top 6 happiest countries in
2019",names.arg=c("Finland","Denmark","Norway","Iceland","Netherlands",
"Switzerland"),ylab="Happiness Score",col=2:7,ylim=c(0,12.5))
legend(12.5,c("Finland-7.769","Denmark-7.600","Norway-7.554","Iceland-
```

```
7.494","Netherlands-7.488","Switzerland-7.480"),fill=2:7)
barplot(bottom[,2],main="6 least happy countries in
2019",names.arg=c("Yemen","Rwanda","Tanzania","Afganisthan","Central
AfricanRepublic","SouthSudan"),ylab="HappinessScore",col=2:7,ylim=c(0,7))
legend(7,c("Yemen-3.380","Rwanda-3.334","Tanzania-3.231","Afganisthan-
3.203","Central African Republic-3.083","South Sudan-2.853"),fill=2:7)


#RegressionDiagnostics

model1=lm(x1~x2+x3+x4+x5)
library(olsrr)
ols_plot_resid_stand(model1)
ols_plot_dffits(model1)
ols_plot_cooksd_chart(model1)




#Year2020#
#----------
rm(list=ls())
data1=read.csv("C:/Users/USER/Documents/MSCDocs/RSProjectMsc/Year
2020.csv",header=TRUE)
#data1
attach(data1)
Data=matrix(as.numeric(unlist(data1[,-c(1,9,10)])),ncol=7)
colnames(Data)=c("x1","x2","x3","x4","x5","x6","x7");Data
x1=Data[,1];x1
x2=Data[,2];x2
x3=Data[,3];x3
x4=Data[,4];x4
x5=Data[,5];x5
x6=Data[,6];x6
x7=Data[,7];x7
model=lm(x1~x2+x3+x4+x5+x6+x7);summary(model)

R=cor(Data);R
R11=det(R[-1,][,-1])
R12=det(R[-1,][,-2])
R13=det(R[-1,][,-3])
R14=det(R[-1,][,-4])
R15=det(R[-1,][,-5])
R16=det(R[-1,][,-6])
R17=det(R[-1,][,-7])
R22=det(R[-2,][,-2])
R33=det(R[-3,][,-3])
R44=det(R[-4,][,-4])
R55=det(R[-5,][,-5])
```

```
R66=det(R[-6,][,-6])
R77=det(R[-7,][,-7])
Mult_corr=sqrt(1-(det(R)/R11));Mult_corr
PC12=-R12/(sqrt(R11*R22));PC12
PC13=R13/(sqrt(R11*R33));PC13
PC14=-R14/(sqrt(R11*R44));PC14
PC15=R15/(sqrt(R11*R55));PC15
PC16=-R16/(sqrt(R11*R66));PC16
PC17=R17/(sqrt(R11*R77));PC17

#x6isdiscardedasavariable

summary(lm(x1~x2+x3+x4+x5+x7))
Data1=Data[,-6];Data1
R1=cor(Data1);R1
R1.11=det(R1[-1,][,-1])
R1.12=det(R1[-1,][,-2])
R1.13=det(R1[-1,][,-3])
R1.14=det(R1[-1,][,-4])
R1.15=det(R1[-1,][,-5])
R1.16=det(R1[-1,][,-6])
R1.22=det(R1[-2,][,-2])
R1.33=det(R1[-3,][,-3])
R1.44=det(R1[-4,][,-4])
R1.55=det(R1[-5,][,-5])
R1.66=det(R1[-6,][,-6])
Mult_corr1=sqrt(1-(det(R1)/R1.11));Mult_corr1
PC1.12=-R1.12/(sqrt(R1.11*R1.22));PC1.12
PC1.13=R1.13/(sqrt(R1.11*R1.33));PC1.13
PC1.14=-R1.14/(sqrt(R1.11*R1.44));PC1.14
PC1.15=R1.15/(sqrt(R1.11*R1.55));PC1.15
PC1.16=-R1.16/(sqrt(R1.11*R1.66));PC1.16

#x7discardedasa variable

summary(lm(x1~x2+x3+x4+x5))
Data2=Data1[,-6];Data2
R2=cor(Data2);R2
R2.11=det(R2[-1,][,-1])
R2.12=det(R2[-1,][,-2])
R2.13=det(R2[-1,][,-3])
R2.14=det(R2[-1,][,-4])
R2.15=det(R2[-1,][,-5])
R2.22=det(R2[-2,][,-2])
R2.33=det(R2[-3,][,-3])
R2.44=det(R2[-4,][,-4])
R2.55=det(R2[-5,][,-5])
Mult_corr2=sqrt(1-(det(R2)/R2.11));Mult_corr2
PC2.12=-R2.12/(sqrt(R2.11*R2.22));PC2.12
```

```
PC2.13=R2.13/(sqrt(R2.11*R2.33));PC2.13
PC2.14=-R2.14/(sqrt(R2.11*R2.44));PC2.14
PC2.15=R2.15/(sqrt(R2.11*R2.55));PC2.15

#x2 discarded as a variable

summary(lm(x1~x3+x4+x5))

#Endofstepwiseregression

#PCA
new_data=Data[,-1];new_data
pca=princomp(new_data,cor=FALSE)
summary(pca)
pca$loadings
plot(pca,type="lines",lwd=3,main="ScreePlotfor2020")

#Plotting

x=aggregate(happiness_score,by=list(continent),FUN=mean)
x
happy_index=as.vector(unlist(x[2]));happy_index
barplot(happy_index,names.arg=c("Africa","Asia","Europe","North
America","Oceania","South
America"),ylim=c(0,12),col=1:6,xlab="Continents",ylab="Happiness
Score",main="Representation of world happiness scores in 2020")
legend(12,c("Africa-4.573435","Asia-5.268026","Europe-6.337205","North
America-6.088250","Oceania-7.2612","South America-6.074100"),fill=1:6)

top=head(data1);top
bottom=tail(data1);bottom
barplot(top[,2],main="Top 6 happiest countries in
2020",names.arg=c("Finland","Denmark","Switzerland","Iceland","Norway",
"Netherlands"),ylab="Happiness Score",col=2:7,ylim=c(0,12.5))
legend(12.5,c("Finland-7.8087","Denmark-7.6456","Switzerland-
7.5599","Iceland-7.5045","Norway-7.4880","Netherlands-7.4489"),fill=2:7)
barplot(bottom[,2],main="6 least happy countries in
2020",names.arg=c("Tanzania","Central African
Republic","Rwanda","Zimbabwe","South
Sudan","Afganisthan"),ylab="Happiness Score",col=2:7,ylim=c(0,7))
legend(7,c("Tanzania-3.4762","Central African Republic-3.4759","Rwanda-
3.3123","Zimbabwe-3.2992","South Sudan-2.8166","Afghanisthan-
2.5669"),fill=2:7)


#RegressionDiagnostics

model1=lm(x1~x2+x3+x4+x5)
library(olsrr)
```

```
ols_plot_resid_stand(model1)
ols_plot_dffits(model1)
ols_plot_cooksd_chart(model1)




#Year2021#
#----------
rm(list=ls())
data1=read.csv("C:/Users/USER/Documents/MSCDocs/RSProjectMsc/Year
2021.csv",header=TRUE)
#data1
attach(data1)
Data=matrix(as.numeric(unlist(data1[,-c(1,9,10)])),ncol=7)
colnames(Data)=c("x1","x2","x3","x4","x5","x6","x7");Data
x1=Data[,1];x1
x2=Data[,2];x2
x3=Data[,3];x3
x4=Data[,4];x4
x5=Data[,5];x5
x6=Data[,6];x6
x7=Data[,7];x7
model=lm(x1~x2+x3+x4+x5+x6+x7);summary(model)
R=cor(Data);R
R11=det(R[-1,][,-1])
R12=det(R[-1,][,-2])
R13=det(R[-1,][,-3])
R14=det(R[-1,][,-4])
R15=det(R[-1,][,-5])
R16=det(R[-1,][,-6])
R17=det(R[-1,][,-7])
R22=det(R[-2,][,-2])
R33=det(R[-3,][,-3])
R44=det(R[-4,][,-4])
R55=det(R[-5,][,-5])
R66=det(R[-6,][,-6])
R77=det(R[-7,][,-7])
Mult_corr=sqrt(1-(det(R)/R11));Mult_corr
PC12=-R12/(sqrt(R11*R22));PC12
PC13=R13/(sqrt(R11*R33));PC13
PC14=-R14/(sqrt(R11*R44));PC14
PC15=R15/(sqrt(R11*R55));PC15
PC16=-R16/(sqrt(R11*R66));PC16
PC17=R17/(sqrt(R11*R77));PC17

#x6 discarded as a variable

summary(lm(x1~x2+x3+x4+x5+x7))
```

```
Data1=Data[,-6];Data1
R1=cor(Data1);R1
R1.11=det(R1[-1,][,-1])
R1.12=det(R1[-1,][,-2])
R1.13=det(R1[-1,][,-3])
R1.14=det(R1[-1,][,-4])
R1.15=det(R1[-1,][,-5])
R1.16=det(R1[-1,][,-6])
R1.22=det(R1[-2,][,-2])
R1.33=det(R1[-3,][,-3])
R1.44=det(R1[-4,][,-4])
R1.55=det(R1[-5,][,-5])
R1.66=det(R1[-6,][,-6])
Mult_corr1=sqrt(1-(det(R1)/R1.11));Mult_corr1
PC1.12=-R1.12/(sqrt(R1.11*R1.22));PC1.12
PC1.13=R1.13/(sqrt(R1.11*R1.33));PC1.13
PC1.14=-R1.14/(sqrt(R1.11*R1.44));PC1.14
PC1.15=R1.15/(sqrt(R1.11*R1.55));PC1.15
PC1.16=-R1.16/(sqrt(R1.11*R1.66));PC1.16

#x4discardedasa variable

summary(lm(x1~x2+x3+x5+x7))
Data2=Data1[,-4];Data2
R2=cor(Data2);R2
R2.11=det(R2[-1,][,-1])
R2.12=det(R2[-1,][,-2])
R2.13=det(R2[-1,][,-3])
R2.14=det(R2[-1,][,-4])
R2.15=det(R2[-1,][,-5])
R2.22=det(R2[-2,][,-2])
R2.33=det(R2[-3,][,-3])
R2.44=det(R2[-4,][,-4])
R2.55=det(R2[-5,][,-5])
Mult_corr2=sqrt(1-(det(R2)/R2.11));Mult_corr2
PC2.12=-R2.12/(sqrt(R2.11*R2.22));PC2.12
PC2.13=R2.13/(sqrt(R2.11*R2.33));PC2.13
PC2.14=-R2.14/(sqrt(R2.11*R2.44));PC2.14
PC2.15=R2.15/(sqrt(R2.11*R2.55));PC2.15

#x7discardedasa variable

summary(lm(x1~x2+x3+x5))
Data3=Data2[,-5];Data3
R3=cor(Data3);R3
R3.11=det(R3[-1,][,-1])
R3.12=det(R3[-1,][,-2])
R3.13=det(R3[-1,][,-3])
R3.14=det(R3[-1,][,-4])
```

```
R3.22=det(R3[-2,][,-2])
R3.33=det(R3[-3,][,-3])
R3.44=det(R3[-4,][,-4])
Mult_corr3=sqrt(1-(det(R3)/R3.11));Mult_corr3
PC3.12=-R3.12/(sqrt(R3.11*R3.22));PC3.12
PC3.13=R3.13/(sqrt(R3.11*R3.33));PC3.13
PC3.14=-R3.14/(sqrt(R3.11*R3.44));PC3.14

#x3isdiscardedasavariable

summary(lm(x1~x2+x5))

#End of step wise regression

#PCA
new_data=Data[,-1];new_data
pca=princomp(new_data,cor=FALSE)
summary(pca)
pca$loadings
plot(pca,type="lines",lwd=3,main="ScreePlotfor2021")

#Plotting

x=aggregate(happiness_score,by=list(continent),FUN=mean)
x
happy_index=as.vector(unlist(x[2]));happy_index
barplot(happy_index,names.arg=c("Africa","Asia","Europe","North
America","Oceania","South
America"),ylim=c(0,12),col=1:6,xlab="Continents",ylab="Happiness
Score",main="Representation of world happiness scores in 2021")
legend(12,c("Africa-4.593356","Asia-5.351114","Europe-6.305067","North
America-6.033","Oceania-7.23","South America-6.002941"),fill=1:6)

top=head(data1);top
bottom=tail(data1);bottom
barplot(top[,2],main="Top 6 happiest countries in
2021",names.arg=c("Finland","Denmark","Switzerland","Iceland","Netherlan
ds","Norway"),ylab="Happiness Score",col=2:7,ylim=c(0,12.5))
legend(12.5,c("Finland-7.842","Denmark-7.620","Switzerland-7.571","Iceland-
7.554","Netherlands-7.464","Norway-7.392"),fill=2:7)
barplot(bottom[,2],main="6 least happy countries in
2021",names.arg=c("Malawi","Lesotho","Bostwana","Rwanda","Zimbabwe",
"Afganisthan"),ylab="Happiness Score",col=2:7,ylim=c(0,7))
legend(7,c("Malawi-3.600","Lesotho-3.512","Botswana-3.467","Rwanda-
3.415","Zimbabwe-3.145","Afghanisthan-2.523"),fill=2:7)


#RegressionDiagnostics
```

```
model1=lm(x1~x2+x3+x4+x5)
library(olsrr)
ols_plot_resid_stand(model1)
ols_plot_dffits(model1)
ols_plot_cooksd_chart(model1)



#Year2022#
#----------
rm(list=ls())
data1=read.csv("C:/Users/USER/Documents/MSCDocs/RSProjectMsc/Year
2022.csv",header=TRUE)
#data1
attach(data1)
Data=matrix(as.numeric(unlist(data1[,-c(1,9,10)])),ncol=7)
colnames(Data)=c("x1","x2","x3","x4","x5","x6","x7");Data
x1=Data[,1];x1
x2=Data[,2];x2
x3=Data[,3];x3
x4=Data[,4];x4
x5=Data[,5];x5
x6=Data[,6];x6
x7=Data[,7];x7
model=lm(x1~x2+x3+x4+x5+x6+x7);summary(model)

R=cor(Data);R
R11=det(R[-1,][,-1])
R12=det(R[-1,][,-2])
R13=det(R[-1,][,-3])
R14=det(R[-1,][,-4])
R15=det(R[-1,][,-5])
R16=det(R[-1,][,-6])
R17=det(R[-1,][,-7])
R22=det(R[-2,][,-2])
R33=det(R[-3,][,-3])
R44=det(R[-4,][,-4])
R55=det(R[-5,][,-5])
R66=det(R[-6,][,-6])
R77=det(R[-7,][,-7])
Mult_corr=sqrt(1-(det(R)/R11));Mult_corr
PC12=-R12/(sqrt(R11*R22));PC12
PC13=R13/(sqrt(R11*R33));PC13
PC14=-R14/(sqrt(R11*R44));PC14
PC15=R15/(sqrt(R11*R55));PC15
PC16=-R16/(sqrt(R11*R66));PC16
PC17=R17/(sqrt(R11*R77));PC17

#x6discardedasa variable
```

```
summary(lm(x1~x2+x3+x4+x5+x7))
Data1=Data[,-6];Data1
R1=cor(Data1);R1
R1.11=det(R1[-1,][,-1])
R1.12=det(R1[-1,][,-2])
R1.13=det(R1[-1,][,-3])
R1.14=det(R1[-1,][,-4])
R1.15=det(R1[-1,][,-5])
R1.16=det(R1[-1,][,-6])
R1.22=det(R1[-2,][,-2])
R1.33=det(R1[-3,][,-3])
R1.44=det(R1[-4,][,-4])
R1.55=det(R1[-5,][,-5])
R1.66=det(R1[-6,][,-6])
Mult_corr1=sqrt(1-(det(R1)/R1.11));Mult_corr1
PC1.12=-R1.12/(sqrt(R1.11*R1.22));PC1.12
PC1.13=R1.13/(sqrt(R1.11*R1.33));PC1.13
PC1.14=-R1.14/(sqrt(R1.11*R1.44));PC1.14
PC1.15=R1.15/(sqrt(R1.11*R1.55));PC1.15
PC1.16=-R1.16/(sqrt(R1.11*R1.66));PC1.16

#x7discardedasa variable

summary(lm(x1~x2+x3+x4+x5))
Data2=Data1[,-6];Data2
R2=cor(Data2);R2
R2.11=det(R2[-1,][,-1])
R2.12=det(R2[-1,][,-2])
R2.13=det(R2[-1,][,-3])
R2.14=det(R2[-1,][,-4])
R2.15=det(R2[-1,][,-5])
R2.22=det(R2[-2,][,-2])
R2.33=det(R2[-3,][,-3])
R2.44=det(R2[-4,][,-4])
R2.55=det(R2[-5,][,-5])
Mult_corr2=sqrt(1-(det(R2)/R2.11));Mult_corr2
PC2.12=-R2.12/(sqrt(R2.11*R2.22));PC2.12
PC2.13=R2.13/(sqrt(R2.11*R2.33));PC2.13
PC2.14=-R2.14/(sqrt(R2.11*R2.44));PC2.14
PC2.15=R2.15/(sqrt(R2.11*R2.55));PC2.15

#x2 discarded as a variable

summary(lm(x1~x3+x4+x5))

#Endofstepwiseregression

#PCA
new_data=Data[,-1];new_data
```

```
pca=princomp(new_data,cor=FALSE)
summary(pca)
pca$loadings
plot(pca,type="lines",lwd=3,main="ScreePlotfor2022")

#Plotting

x=aggregate(happiness_score,by=list(continent),FUN=mean)
x
happy_index=as.vector(unlist(x[2]));happy_index
barplot(happy_index,names.arg=c("Africa","Asia","Europe","North
America","Oceania","South
America"),ylim=c(0,12),col=1:6,xlab="Continents",ylab="Happiness
Score",main="Representation of world happiness scores in 2022")
legend(12,c("Africa-4.527625","Asia-5.336231","Europe-6.367795","North
America-6.288818","Oceania-7.181","South America-5.7882"),fill=1:6)

top=head(data1);top
bottom=tail(data1);bottom
barplot(top[,2],main="Top 6 happiest countries in
2022",names.arg=c("Finland","Denmark","Iceland","Switzerland","Netherlan
ds","Luxemborg"),ylab="Happiness Score",col=2:7,ylim=c(0,12.5))
legend(12.5,c("Finland-7.821","Denmark-7.636","Iceland-7.557","Switzerland-
7.512","Netherlands-7.415","Luxembourg-7.404"),fill=2:7)
barplot(bottom[,2],main="6 least happy countries in
2022",names.arg=c("Lesotho","Bostwana","Rwanda","Zimbabwe","Lebanon",
"Afganisthan"),ylab="Happiness Score",col=2:7,ylim=c(0,7))
legend(7,c("Lesotho-3.512","Botswana-3.471","Rwanda-3.268","Zimbabwe-
2.995","Lebanon-2.955","Afghanisthan-2.404"),fill=2:7)


#RegressionDiagnostics

model1=lm(x1~x2+x3+x4+x5)
library(olsrr)
ols_plot_resid_stand(model1)
ols_plot_dffits(model1)
ols_plot_cooksd_chart(model1)




#Year2023#
#----------
rm(list=ls())
data1=read.csv("C:/Users/USER/Documents/MSCDocs/RSProjectMsc/Year
2023.csv",header=TRUE)
#data1
attach(data1)
```

```
Data=matrix(as.numeric(unlist(data1[,-c(1,9,10)])),ncol=7)
colnames(Data)=c("x1","x2","x3","x4","x5","x6","x7");Data
x1=Data[,1];x1
x2=Data[,2];x2
x3=Data[,3];x3
x4=Data[,4];x4
x5=Data[,5];x5
x6=Data[,6];x6
x7=Data[,7];x7
model=lm(x1~x2+x3+x4+x5+x6+x7);summary(model)
R=cor(Data);R
R11=det(R[-1,][,-1])
R12=det(R[-1,][,-2])
R13=det(R[-1,][,-3])
R14=det(R[-1,][,-4])
R15=det(R[-1,][,-5])
R16=det(R[-1,][,-6])
R17=det(R[-1,][,-7])
R22=det(R[-2,][,-2])
R33=det(R[-3,][,-3])
R44=det(R[-4,][,-4])
R55=det(R[-5,][,-5])
R66=det(R[-6,][,-6])
R77=det(R[-7,][,-7])
Mult_corr=sqrt(1-(det(R)/R11));Mult_corr
PC12=-R12/(sqrt(R11*R22));PC12
PC13=R13/(sqrt(R11*R33));PC13
PC14=-R14/(sqrt(R11*R44));PC14
PC15=R15/(sqrt(R11*R55));PC15
PC16=-R16/(sqrt(R11*R66));PC16
PC17=R17/(sqrt(R11*R77));PC17

#x6discardedasa variable

summary(lm(x1~x2+x3+x4+x5+x7))
Data1=Data[,-6];+xData1
R1=cor(Data1);R1
R1.11=det(R1[-1,][,-1])
R1.12=det(R1[-1,][,-2])
R1.13=det(R1[-1,][,-3])
R1.14=det(R1[-1,][,-4])
R1.15=det(R1[-1,][,-5])
R1.16=det(R1[-1,][,-6])
R1.22=det(R1[-2,][,-2])
R1.33=det(R1[-3,][,-3])
R1.44=det(R1[-4,][,-4])
R1.55=det(R1[-5,][,-5])
R1.66=det(R1[-6,][,-6])
Mult_corr1=sqrt(1-(det(R1)/R1.11));Mult_corr1
```

```
PC1.12=-R1.12/(sqrt(R1.11*R1.22));PC1.12
PC1.13=R1.13/(sqrt(R1.11*R1.33));PC1.13
PC1.14=-R1.14/(sqrt(R1.11*R1.44));PC1.14
PC1.15=R1.15/(sqrt(R1.11*R1.55));PC1.15
PC1.16=-R1.16/(sqrt(R1.11*R1.66));PC1.16

#x4discardedasa variable

summary(lm(x1~x2+x3+x5+x7))
Data2=Data1[,-4];Data2
R2=cor(Data2);R2
R2.11=det(R2[-1,][,-1])
R2.12=det(R2[-1,][,-2])
R2.13=det(R2[-1,][,-3])
R2.14=det(R2[-1,][,-4])
R2.15=det(R2[-1,][,-5])
R2.22=det(R2[-2,][,-2])
R2.33=det(R2[-3,][,-3])
R2.44=det(R2[-4,][,-4])
R2.55=det(R2[-5,][,-5])
Mult_corr2=sqrt(1-(det(R2)/R2.11));Mult_corr2
PC2.12=-R2.12/(sqrt(R2.11*R2.22));PC2.12
PC2.13=R2.13/(sqrt(R2.11*R2.33));PC2.13
PC2.14=-R2.14/(sqrt(R2.11*R2.44));PC2.14
PC2.15=R2.15/(sqrt(R2.11*R2.55));PC2.15

#x7discardedasa variable

summary(lm(x1~x2+x3+x5))
Data3=Data2[,-5];Data3
R3=cor(Data3);R3
R3.11=det(R3[-1,][,-1])
R3.12=det(R3[-1,][,-2])
R3.13=det(R3[-1,][,-3])
R3.14=det(R3[-1,][,-4])
R3.22=det(R3[-2,][,-2])
R3.33=det(R3[-3,][,-3])
R3.44=det(R3[-4,][,-4])
Mult_corr3=sqrt(1-(det(R3)/R3.11));Mult_corr3
PC3.12=-R3.12/(sqrt(R3.11*R3.22));PC3.12
PC3.13=R3.13/(sqrt(R3.11*R3.33));PC3.13
PC3.14=-R3.14/(sqrt(R3.11*R3.44));PC3.14

#x2isdiscardedasavariable

summary(lm(x1~x3+x5))
Data4=Data3[,-2];Data4
R4=cor(Data4);R4
R4.11=det(R4[-1,][,-1])
```

```
R4.12=det(R4[-1,][,-2])
R4.13=det(R4[-1,][,-3])
R4.22=det(R4[-2,][,-2])
R4.33=det(R4[-3,][,-3])
Mult_corr4=sqrt(1-(det(R4)/R4.11));Mult_corr4
PC4.12=-R4.12/(sqrt(R4.11*R4.22));PC4.12
PC4.13=R4.13/(sqrt(R4.11*R4.33));PC4.13

#Endofstepwiseregression

#PCA
new_data=Data[,-1];new_data
pca=princomp(new_data,cor=FALSE)
summary(pca)
pca$loadings
plot(pca,type="lines",lwd=3,ylim=c(0,0.3),main="ScreePlotfor2023")


#Plotting

x=aggregate(happiness_score,by=list(continent),FUN=mean)
x
happy_index=as.vector(unlist(x[2]));happy_index
barplot(happy_index,names.arg=c("Africa","Asia","Europe","North
America","Oceania","South
America"),ylim=c(0,12),col=1:6,xlab="Continents",ylab="Happiness
Score",main="Representation of world happiness scores in 2023")
legend(12,c("Africa-4.432651","Asia-5.439897","Europe-6.408643","North
America-6.531750","Oceania-7.109","South America-5.946059"),fill=1:6)

top=head(data1);top
bottom=tail(data1);bottom
barplot(top[,2],main="Top 6 happiest countries in
2023",names.arg=c("Finland","Denmark","Iceland","Israel","Netherlands","Sw
eden"),ylab="Happiness Score",col=2:7,ylim=c(0,12.5)) legend(12.5,c("Finland-
7.804","Denmark-7.586","Iceland-7.530","Israel- 7.473","Netherlands-
7.403","Sweden-7.395"),fill=2:7)barplot(bottom[,2],main="6 least happy
countries in
2023",names.arg=c("Bostwana","Congo(Kinshasa)","Zimbabwe","Sierra
Leone","Lebanon","Afganisthan"),ylab="HappinessScore",col=2:7,ylim=c(0,6))
legend(7,c("Botswana-3.435","Congo-3.207","Zimbabwe-3.204","Sierra Leone-
3.138","Lebanon-2.392","Afghanisthan-1.859"),fill=2:7)
#RegressionDiagnostics

model1=lm(x1~x2+x3+x4+x5)
library(olsrr)
ols_plot_resid_stand(model1)
ols_plot_dffits(model1)
ols_plot_cooksd_chart(model1)
```