Case Study

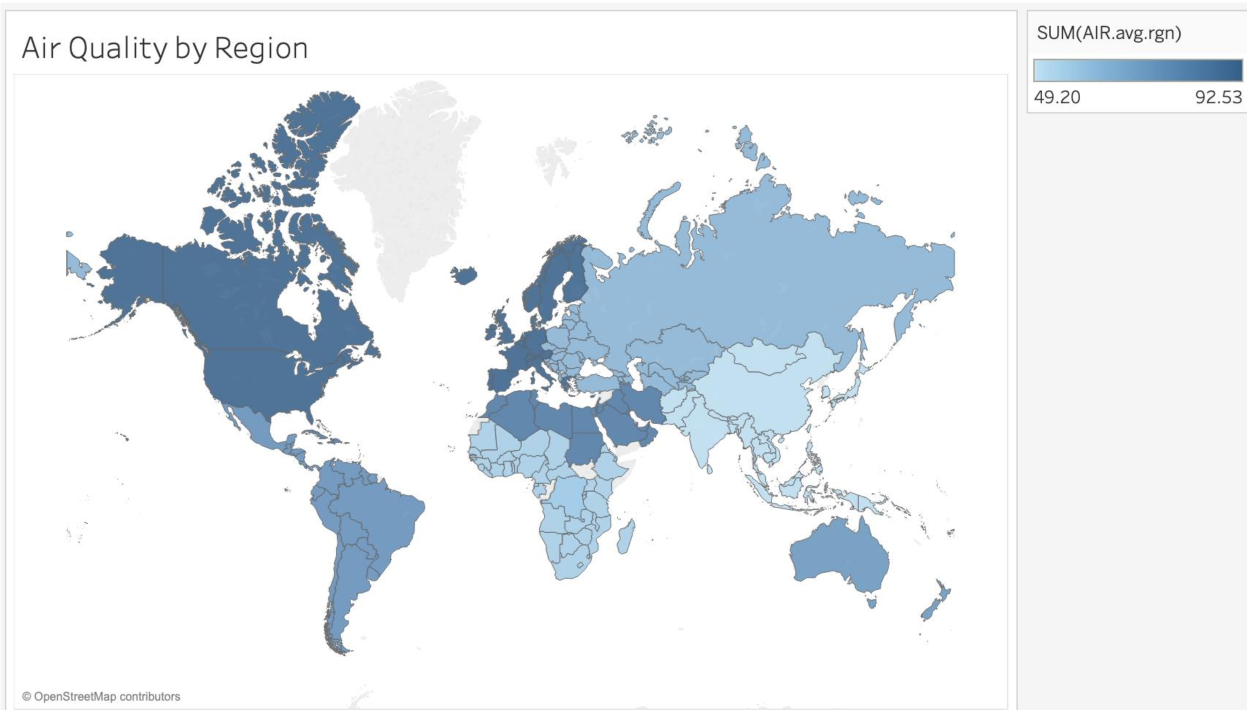
FW

Introduction

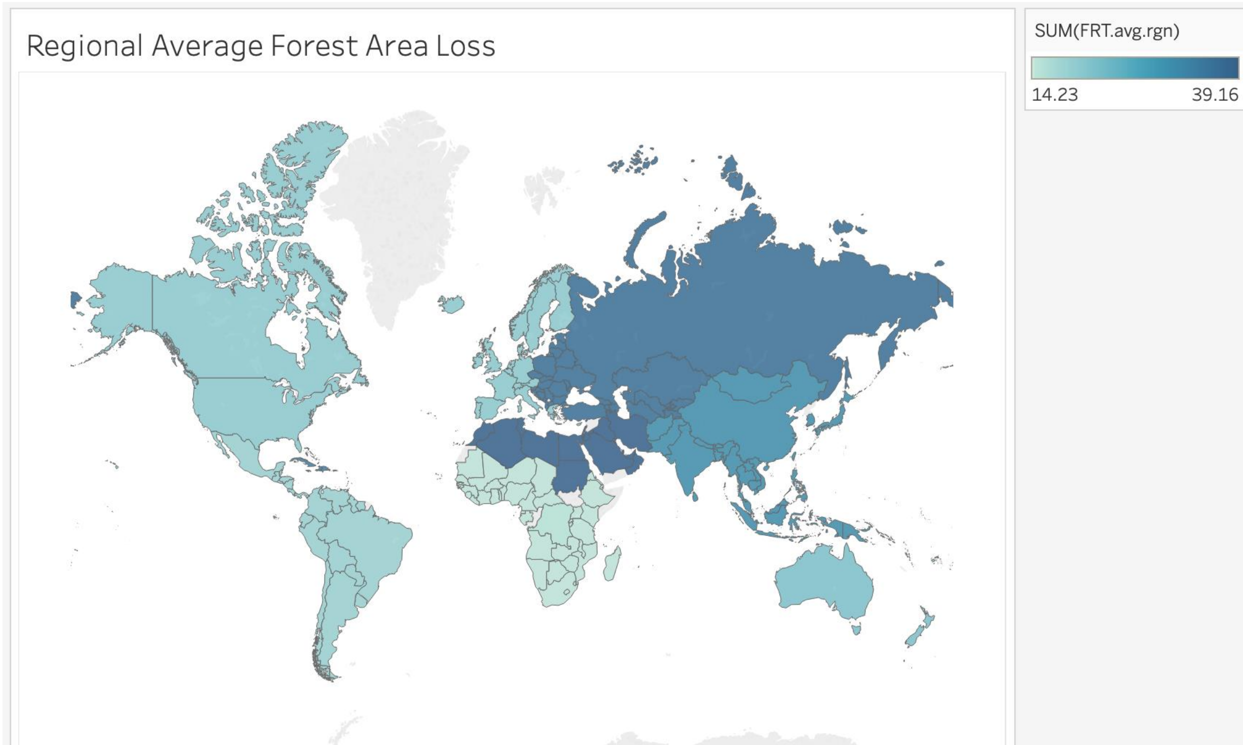
The initial inspiration for this Case Study came from reading about the Environmental Kuznet's Curve (EKC) proposed by the founder of modern economic science Simon Smith Kuznets.

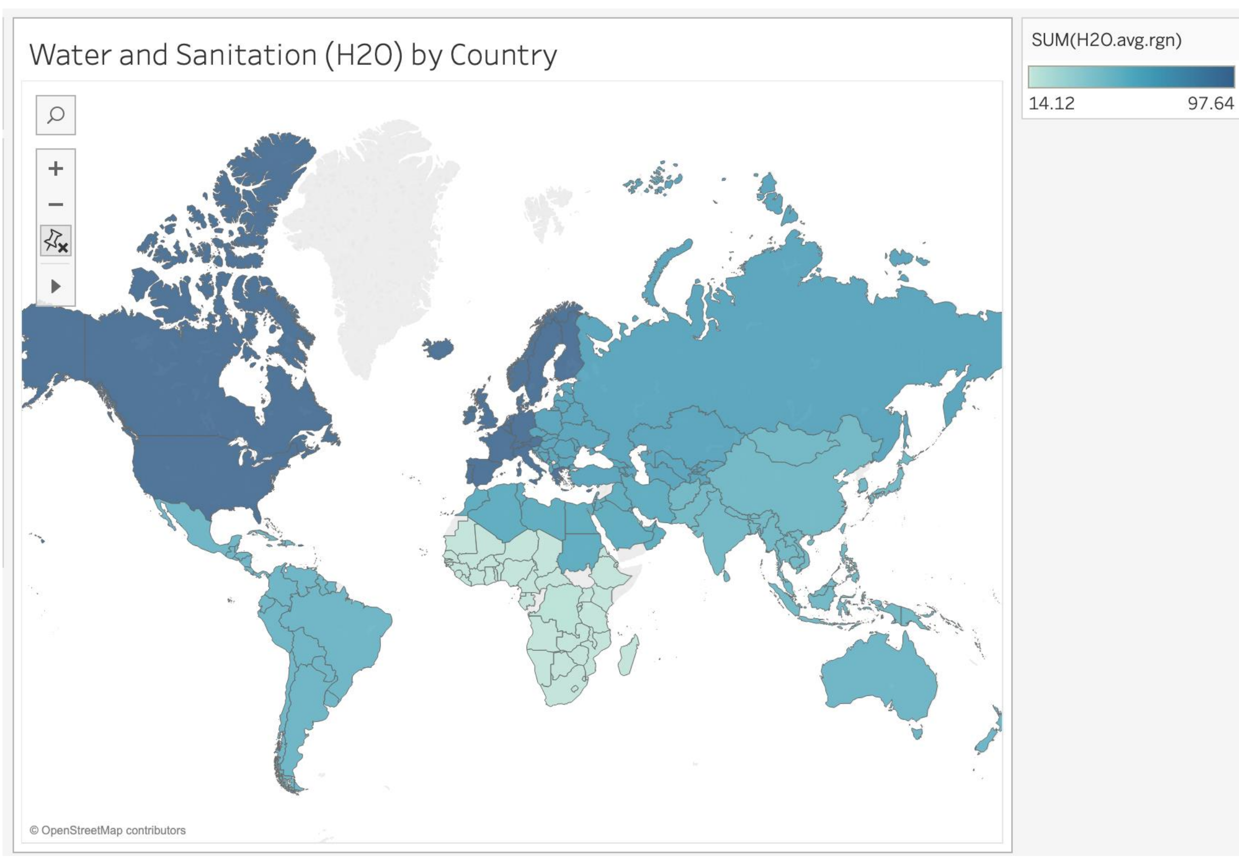
The EKC basically states that the "solution to pollution is economic growth." The EKC has been heavily criticized; however, there is also evidence for its support, especially in areas of air and water pollution, as well as deforestation.

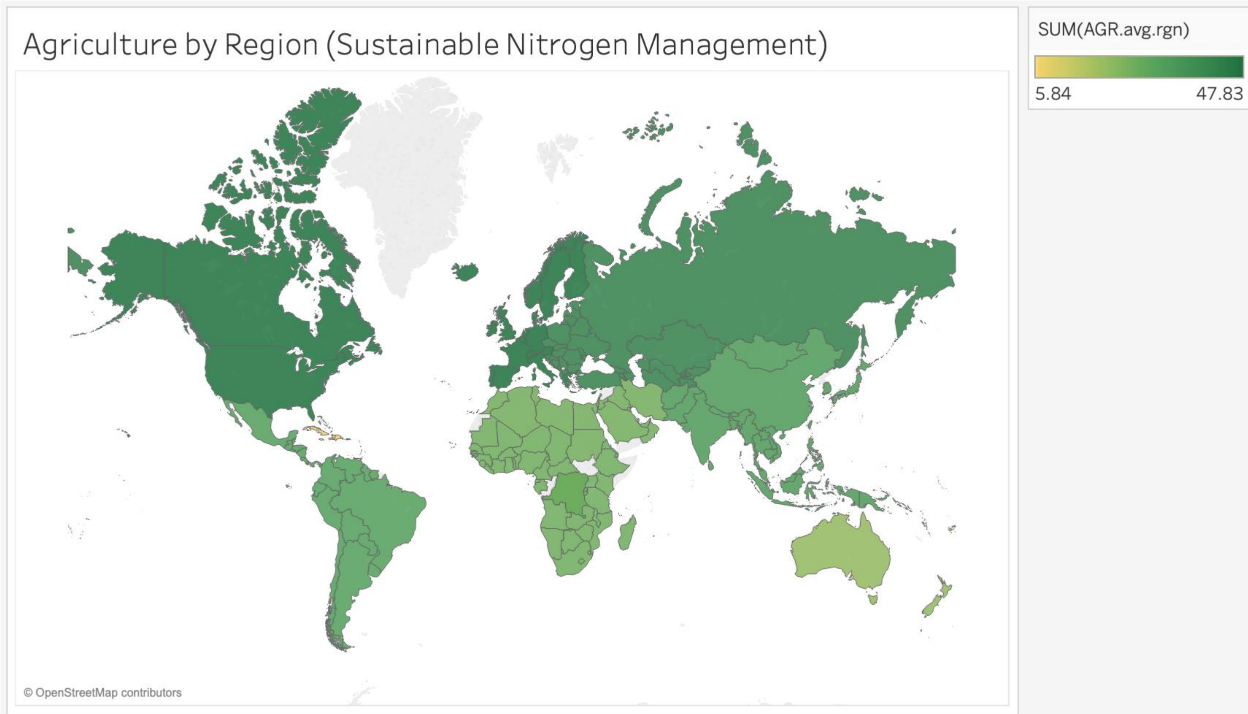
The data collected and produced for this Case Study also shows how GDP per capita is positively correlated with many of the environmental factors. Here are a few snapshots:



DO NOT SHARE THIS WITH ANYONE… COPYRIGHT PROTECTED







Data

The data used for the prediction and analysis of Environmental Quality Score based on GDPpc and other factors was obtained from Yale's Environmental Performance Index (EPI) produced jointly by Yale University and Columbia University in collaboration with the World Economic Forum.

Two different datasets were used:

1. 2018 EPI Country Snapshot
2. 2018 EPI Regional Comparisons

DO NOT SHARE THIS WITH ANYONE… COPYRIGHT PROTECTED

epi2018country

code	iso	country	Pop	PopGrowthRate	PopDensity	GDP	LandArea	GDPpc	SDGI	EPI2018Score	EPI2018Rank
4	AFG	Afghanistan	34.656032	2.689163473	53.08340532	60.28705023	652860	1739.583177	46.8	37.74	168
8	ALB	Albania	2.876101	-0.159880412	104.9671898	32.66991645	27400	11359.09916	68.9	65.46	40
12	DZA	Algeria	40.606052	1.825463432	17.04889491	565.284	2381741	13921.17608	68.8	57.18	88
24	AGO	Angola	28.813463	3.367572131	23.11178551	172.438	1246700	5984.632947	50.2	37.44	170
28	ATG	Antigua and Barbuda	0.100963	1.03542237	229.4613636	2.121537299	440	21013.01763		59.18	76
32	ARG	Argentina	43.84743	0.984741906	16.0220668	810.714	2736690	18489.43028	72.5	59.3	74
51	ARM	Armenia	2.924816	0.269302288	102.7332631	23.95494292	28470	8190.239289	71.7	62.07	63
36	AUS	Australia	24.127159	1.410063834	3.140616612	1071.58	7682300	44413.84914	75.9	74.12	21
40	AUT	Austria	8.747358	1.314006681	105.9990306	388.015	82523	44357.965	81.4	78.97	8
31	AZE	Azerbaijan	9.762274	1.163574223	118.0972624	156.138	82663	15994.01943	70.8	62.33	59
44	BHS	Bahamas	0.391232	1.129473367	39.08411588	8.168481869	10010	20878.86949		54.99	98
48	BHR	Bahrain	1.425171	3.812796793	1848.470817	64.33731963	771	45143.579	64.6	55.15	96
50	BGD	Bangladesh	162.95156	1.080165239	1251.836521	540.894	130170	3319.354537	56.2	29.56	179
52	BRB	Barbados	0.284996	0.273711418	662.7813953	4.773832887	430	16750.52593	66	55.76	93
112	BLR	Belarus	9.50712	0.184284342	46.8538761	159.171	202910	16742.2942	77.1	64.98	44
56	BEL	Belgium	11.348159	0.653895349	374.7740753	477.288	30280	42058.62819	80	77.38	15
84	BLZ	Belize	0.366954	2.111220431	16.0874178	2.879126894	22810	7846.015833	66	57.79	81
204	BEN	Benin	10.872298	2.763534181	96.41981199	21.85289913	112760	2009.961384	49.5	38.17	167
64	BTN	Bhutan	0.797765	1.309546955	20.92937534	6.584188915	38117	8253.293783	65.5	47.22	131
68	BOL	Bolivia	10.887882	1.510046906	10.05066187	73.03543539	1083300	6707.956184	64.7	55.98	92
70	BIH	Bosnia and Herzegovina	3.516816	-0.54290805	68.6878125	39.8364244	51200	11327.41218	65.5	41.84	158
72	BWA	Botswana	2.25026	1.841666401	3.970603286	35.38137144	566730	15723.23707	58.3	51.7	113
76	BRA	Brazil	207.652865	0.817555711	24.84438703	2912.06	8358140	14023.69286	69.5	60.7	69

Region	EPI.new	EPI.rnk.new	EPI.rnk.rgn	EPI.avg.rgn	HLT.rnk.rgn	HLT.avg.rgn	ECO.rnk.rgn	ECO.avg.rgn	HMT.rnk.rgn	HMT.avg.rgn	AIR.rnk.rgn	AIR.avg.rgn	H2O.rnk.rgn	H2
Asia	37.74	168	22	50.13346154	18	49.11423077	24	50.81384615	26	48.35153846	15	49.20269231	21	
Eastern Europe & Eurasia	65.46	40	8	60.54827586	15	63.50758621	10	58.57655172	23	72.21275862	14	63.31413793	12	62.
Mid East & North Africa	57.18	88	12	58.70941176	6	73.89588235	14	48.58529412	10	39.02058824	6	83.98352941	10	57.
Sub-Saharan Africa	37.44	170	40	44.89	36	41.97956522	36	46.83108696	21	40.61608696	34	54.94152174	30	14.
Caribbean	59.18	76	6	57.55916667	3	70.7	7	48.79833333	3	40.15583333	3	83.33666667	5	48.
Latin America	59.3	74	9	58.173	2	66.851	19	52.387	5	48.3335	3	75.553	3	
Eastern Europe & Eurasia	62.07	63	17	60.54827586	25	63.50758621	9	58.57655172	16	72.21275862	24	63.31413793	17	62.
Pacific	74.12	21	2	57.89875	1	66.08375	4	52.445	4	59.56375	1	73.645	1	
Europe & North America	78.97	8	8	77.90318182	21	93.45409091	3	67.53545455	7	80.32090909	21	92.53227273	19	97.
Eastern Europe & Eurasia	62.33	59	16	60.54827586	28	63.50758621	2	58.57655172	22	72.21275862	27	63.31413793	26	62.
Caribbean	54.99	98	10	57.55916667	5	70.7	9	48.79833333	4	40.15583333	6	83.33666667	4	48.
Mid East & North Africa	55.15	96	13	58.70941176	15	73.89588235	9	48.58529412	5	39.02058824	16	83.98352941	6	57.
Asia	29.56	179	26	50.13346154	24	49.11423077	23	50.81384615	24	48.35153846	25	49.20269231	19	
Caribbean	55.76	93	9	57.55916667	1	70.7	11	48.79833333	2	40.15583333	1	83.33666667	3	48.
Eastern Europe & Eurasia	64.98	44	11	60.54827586	10	63.50758621	14	58.57655172	13	72.21275862	8	63.31413793	11	62.
Europe & North America	77.38	15	15	77.90318182	18	93.45409091	9	67.53545455	21	80.32090909	18	92.53227273	17	97.
Latin America	57.79	81	10	58.173	11	66.851	8	52.387	10	48.3335	12	75.553	10	
Sub-Saharan Africa	38.17	167	39	44.89	35	41.97956522	37	46.83108696	39	40.61608696	30	54.94152174	36	14.
Asia	47.22	131	15	50.13346154	20	49.11423077	7	50.81384615	21	48.35153846	19	49.20269231	16	
Latin America	55.98	92	13	58.173	15	66.851	9	52.387	12	48.3335	15	75.553	13	
Eastern Europe & Eurasia	41.84	158	29	60.54827586	17	63.50758621	29	58.57655172	21	72.21275862	17	63.31413793	2	62.
Sub-Saharan Africa	51.7	113	8	44.89	32	41.97956522	5	46.83108696	9	40.61608696	36	54.94152174	9	14.
Latin America	60.7	69	7	58.173	8	66.851	5	52.387	11	48.3335	7	75.553	11	
Asia	63.57	53	4	50.13346154	1	49.11423077	22	50.81384615	5	48.35153846	1	49.20269231	4	
Eastern Europe & Eurasia	67.85	30	3	60.54827586	9	63.50758621	8	58.57655172	12	72.21275862	10	63.31413793	3	62.
Sub-Saharan Africa	42.83	154	28	44.89	17	41.97956522	32	46.83108696	42	40.61608696	14	54.94152174	32	14.
Sub-Saharan Africa	27.43	180	46	44.89	44	41.97956522	45	46.83108696	43	40.61608696	44	54.94152174	43	14.
Sub-Saharan Africa	45.25	139	18	44.89	29	41.97956522	19	46.83108696	35	40.61608696	26	54.94152174	27	14.

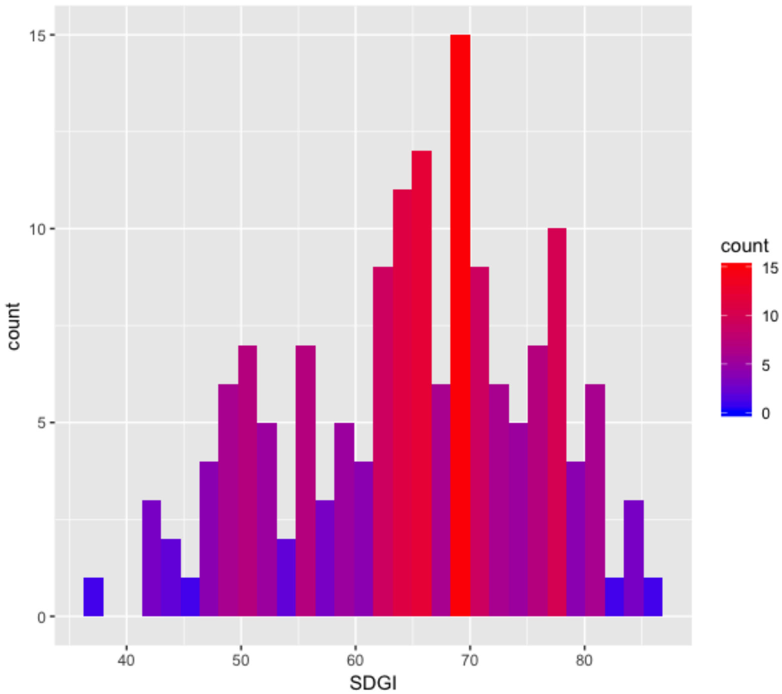
SDGI values for 37 missing countries were removed and the two datasets were merged.

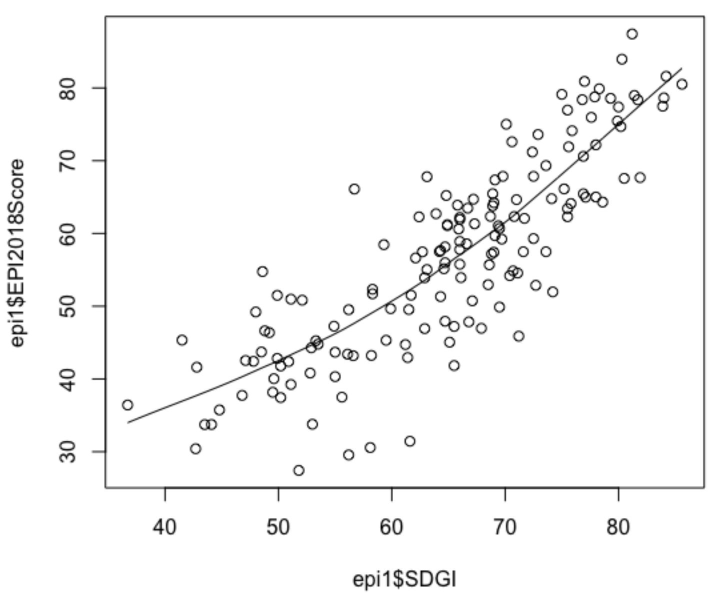
DO NOT SHARE THIS WITH ANYONE… COPYRIGHT PROTECTED

Analysis

Variables Used for the Predication of the Environmental Quality Score

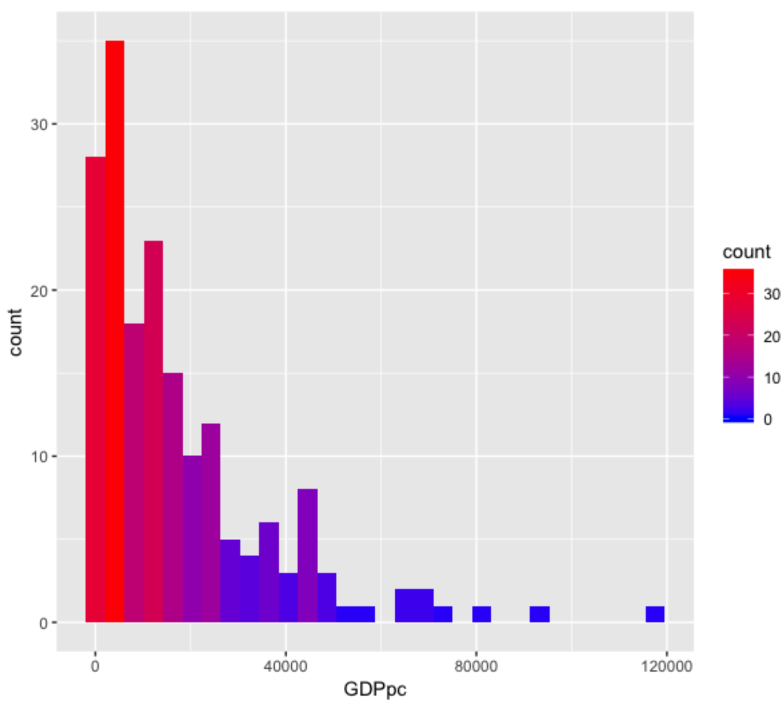
SDGI



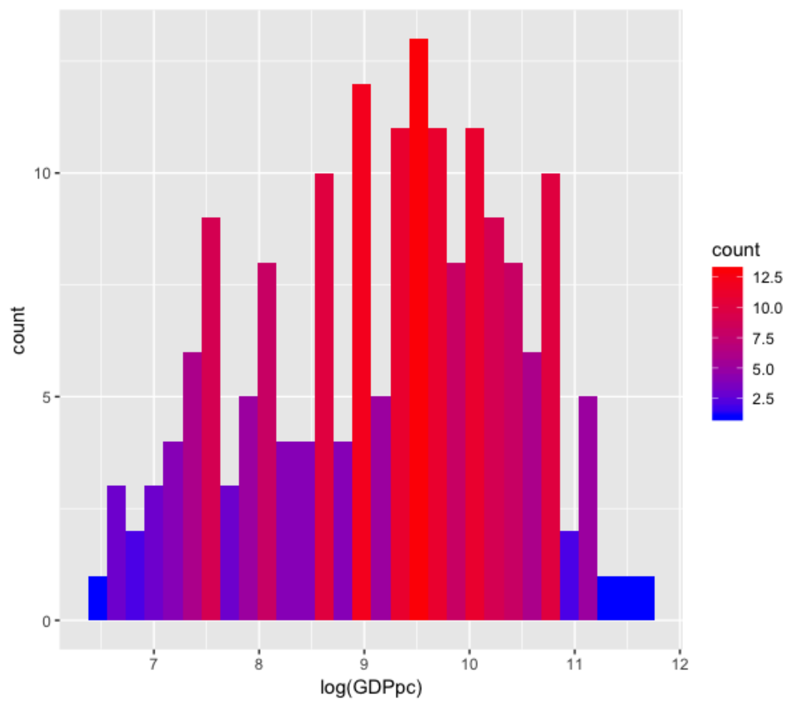


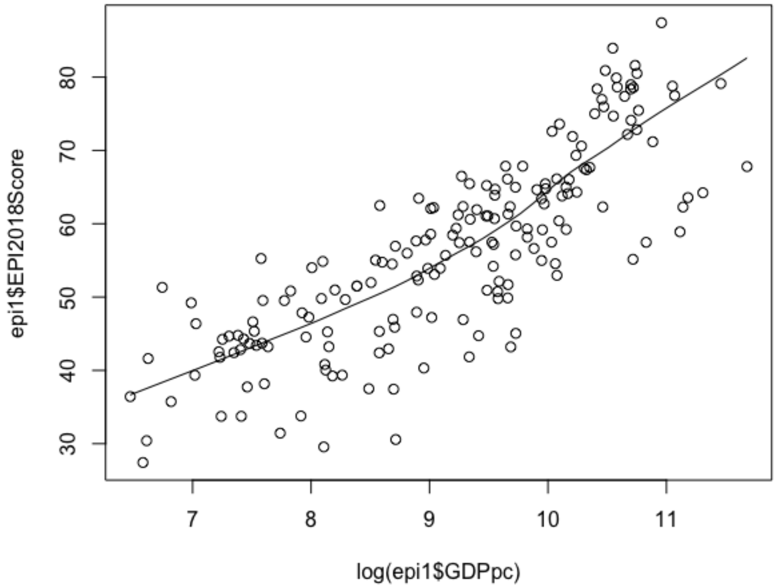
DO NOT SHARE THIS WITH ANYONE… COPYRIGHT PROTECTED

GDP per Capita



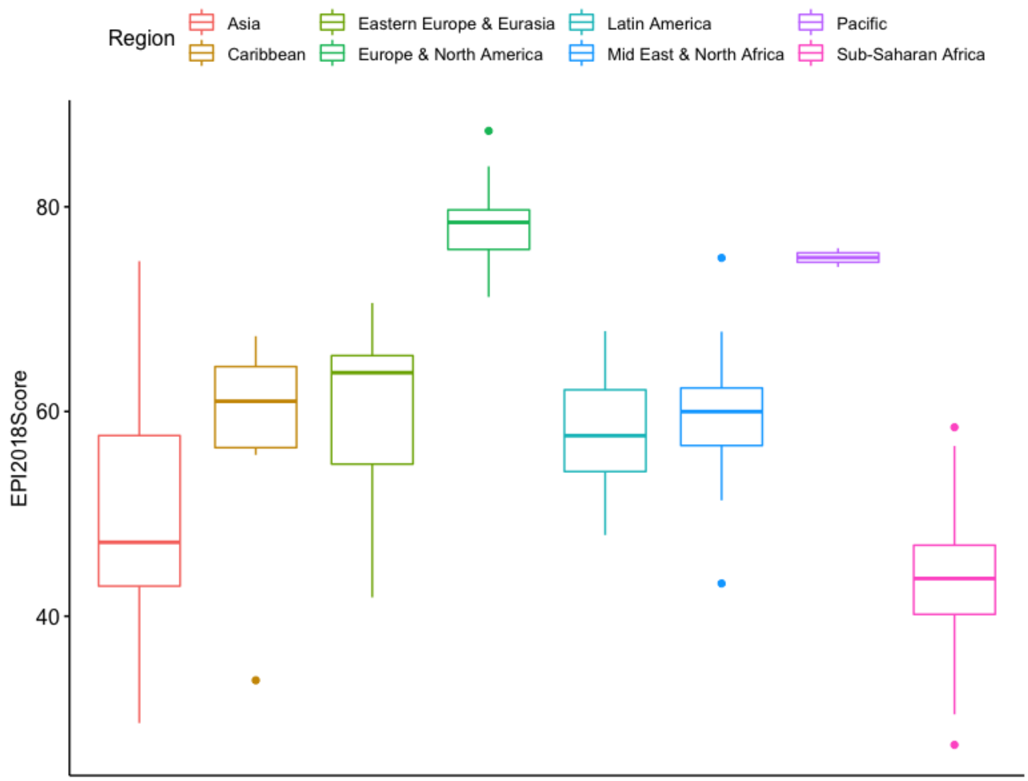
DO NOT SHARE THIS WITH ANYONE… COPYRIGHT PROTECTED



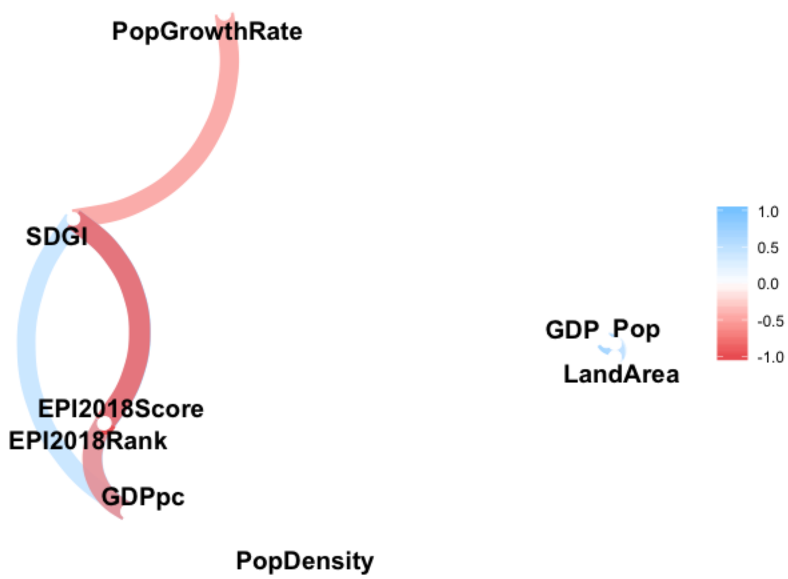


Region

The 7 Regions include in the regression analysis:



Correlation of Variables:



DO NOT SHARE THIS WITH ANYONE… COPYRIGHT PROTECTED

```
# A tibble: 9 x 10
  rowname           Pop PopGrowthRate PopDensity     GDP LandArea   GDPpc    SDGI EPI2018Score EPI2018Rank
  <chr>           <dbl>         <dbl>      <dbl>   <dbl>    <dbl>   <dbl>   <dbl>        <dbl>       <dbl>
1 Pop            1              0.106     0.0207  0.797    0.778  -0.0360 -0.102       -0.0889      0.0889
2 PopGrowthRate  0.106          1        -0.136  -0.194    0.168  -0.475  -0.704       -0.509       0.509
3 PopDensity     0.0207        -0.136     1       0.0793  -0.548   0.119   0.106        0.133      -0.133
4 GDP            0.797         -0.194     0.0793  1        0.594   0.536   0.442        0.398      -0.398
5 LandArea       0.778          0.168    -0.548   0.594    1      -0.0913 -0.180       -0.164       0.164
6 GDPpc         -0.0360        -0.475     0.119   0.536   -0.0913  1       0.811        0.836      -0.836
7 SDGI          -0.102         -0.704     0.106   0.442   -0.180   0.811   1            0.849      -0.849
8 EPI2018Score  -0.0889        -0.509     0.133   0.398   -0.164   0.836   0.849        1          -1
9 EPI2018Rank    0.0889         0.509    -0.133  -0.398    0.164  -0.836  -0.849       -1           1
```

Variables that determine the Environmental quality score:

	Policy Objective			Issue Category			Indicator		
	Title	TLA	Weight	Title	TLA	Weight	Title	TLA	Weight
EPI	Environmental Health	HLT	40%	Air Quality	AIR	65%	Household Solid Fuels	HAD	40%
							$PM_{2.5}$ Exposure	PME	30%
							$PM_{2.5}$ Exceedance	PMW	30%
				Water & Sanitation	H2O	30%	Drinking Water	UWD	50%
							Sanitation	USD	50%
				Heavy Metals	HMT	5%	Lead Exposure	PBD	100%
	Ecosystem Vitality	ECO	60%	Biodiversity & Habitat	BDH	25%	Marine Protected Areas	MPA	20%
							Biome Protection (National)	TBN	20%
							Biome Protection (Global)	TBG	20%
							Species Protection Index	SPI	20%
							Representativeness Index	PAR	10%
							Species Habitat Index	SHI	10%
				Forests	FOR	10%	Tree Cover Loss	TCL	100%
				Fisheries	FSH	10%	Fish Stock Status	FSS	50%
							Regional Marine Trophic Index	MTR	50%
				Climate & Energy	CCE	30%	CO_2 Emissions – Total	DCT	50%
							CO_2 Emissions – Power	DPT	20%
							Methane Emissions	DMT	20%
							N_2O Emissions	DNT	5%
							Black Carbon Emissions	DBT	5%
				Air Pollution	APE	10%	SO_2 Emissions	DST	50%
							NO_X Emissions	DXT	50%
				Water Resources	WRS	10%	Wastewater Treatment	WWT	100%
				Agriculture	AGR	5%	Sustainable Nitrogen Management	SNM	100%

Quick Overview:

FIGURE ES-1 THE 2018 EPI RANKINGS Rank, EPI Score, and Regional Standing (REG, shown in color) for 180 countries

RANK	COUNTRY	SCORE	REG
1	Switzerland	87.42	1
2	France	83.95	2
3	Denmark	81.60	3
4	Malta	80.90	4
5	Sweden	80.51	5
6	United Kingdom	79.89	6
7	Luxembourg	79.12	7
8	Austria	78.97	8
9	Ireland	78.77	9
10	Finland	78.64	10
11	Iceland	78.57	11
12	Spain	78.39	12
13	Germany	78.37	13
14	Norway	77.49	14
15	Belgium	77.38	15
16	Italy	76.96	18
17	New Zealand	75.96	1
18	Netherlands	75.46	17
19	Israel	75.01	1
20	Japan	74.69	1
21	Australia	74.12	2
22	Greece	73.60	18
23	Taiwan	72.84	2
24	Cyprus	72.60	19
25	Canada	72.18	20
26	Portugal	71.91	21
27	United States of America	71.19	22
28	Slovakia	70.60	1
29	Lithuania	69.33	2
30	Bulgaria	67.85	3
30	Costa Rica	67.85	1
32	Qatar	67.80	2
33	Czech Republic	67.68	4
34	Slovenia	67.57	5
35	Trinidad and Tobago	67.36	1
36	St. Vincent & Grenadines	66.48	2
37	Latvia	66.12	6
38	Turkmenistan	66.10	7
39	Seychelles	66.02	1
40	Albania	65.46	8
41	Croatia	65.45	9
42	Colombia	65.22	2
43	Hungary	65.01	10
44	Belarus	64.98	11
45	Romania	64.78	12
46	Dominican Republic	64.71	3
47	Uruguay	64.65	3
48	Estonia	64.31	13
49	Singapore	64.23	3
50	Poland	64.11	14
51	Venezuela	63.89	4
52	Russia	63.79	15
53	Brunei Darussalam	63.57	4
54	Morocco	63.47	3
55	Cuba	63.42	4
56	Panama	62.71	5
57	Tonga	62.49	3
58	Tunisia	62.35	4
59	Azerbaijan	62.33	16
60	South Korea	62.30	5
61	Kuwait	62.28	5
62	Jordan	62.20	6
63	Armenia	62.07	17
64	Peru	61.92	6
65	Montenegro	61.33	18
66	Egypt	61.21	7
67	Lebanon	61.08	8
68	Macedonia	61.06	19
69	Brazil	60.70	7
70	Sri Lanka	60.61	6
71	Equatorial Guinea	60.40	2
72	Mexico	59.69	8
73	Dominica	59.38	5
74	Argentina	59.30	9
75	Malaysia	59.22	7
76	Antigua and Barbuda	59.18	6
77	United Arab Emirates	58.90	9
78	Jamaica	58.58	7
79	Namibia	58.46	3
80	Iran	58.16	10
81	Belize	57.79	10
82	Philippines	57.65	8
83	Mongolia	57.51	9
84	Serbia	57.49	20
84	Chile	57.49	11
86	Saudi Arabia	57.47	11
87	Ecuador	57.42	12
88	Algeria	57.18	12
89	Cabo Verde	56.94	4
90	Mauritius	56.63	5
91	Saint Lucia	56.18	8
92	Bolivia	55.98	13
93	Barbados	55.76	9
94	Georgia	55.69	21
95	Kiribati	55.26	4
96	Bahrain	55.15	13
97	Nicaragua	55.04	14
98	Bahamas	54.99	10
99	Kyrgyzstan	54.86	22
100	Nigeria	54.76	6
101	Kazakhstan	54.56	23
102	Samoa	54.50	5
103	Suriname	54.20	15
104	São Tomé and Príncipe	54.01	7
105	Paraguay	53.93	16
106	El Salvador	53.91	17
107	Fiji	53.09	6
108	Turkey	52.96	24
109	Ukraine	52.87	25
110	Guatemala	52.33	18
111	Maldives	52.14	10
112	Moldova	51.97	26
113	Botswana	51.70	8
114	Honduras	51.51	19
115	Sudan	51.49	14
116	Oman	51.32	15
117	Zambia	50.97	9
118	Grenada	50.93	11
119	Tanzania	50.83	10
120	China	50.74	11
121	Thailand	49.88	12
122	Micronesia	49.80	13
123	Libya	49.79	16
124	Ghana	49.66	11
125	Timor-Leste	49.54	14
126	Senegal	49.52	12
127	Malawi	49.21	13
128	Guyana	47.93	20
129	Tajikistan	47.85	27
130	Kenya	47.25	14
131	Bhutan	47.22	15
132	Viet Nam	46.96	16
133	Indonesia	46.92	17
134	Guinea	46.62	15
135	Mozambique	46.37	16
136	Uzbekistan	45.88	28
137	Chad	45.34	17
138	Myanmar	45.32	18
139	Côte d'Ivoire	45.25	18
140	Gabon	45.05	19
141	Ethiopia	44.78	20
142	South Africa	44.73	21
143	Guinea-Bissau	44.67	22
144	Vanuatu	44.55	7
145	Uganda	44.28	23
146	Comoros	44.24	24
147	Mali	43.71	25
148	Rwanda	43.68	26
149	Zimbabwe	43.41	27
150	Cambodia	43.23	19
151	Solomon Islands	43.22	8
152	Iraq	43.20	17
153	Laos	42.94	20
154	Burkina Faso	42.83	28
155	Sierra Leone	42.54	29
156	Gambia	42.42	30
157	Republic of Congo	42.39	31
158	Bosnia and Herzegovina	41.84	29
159	Togo	41.78	32
160	Liberia	41.62	33
161	Cameroon	40.81	34
162	Swaziland	40.32	35
163	Djibouti	40.04	36
164	Papua New Guinea	39.35	21
165	Eritrea	39.34	37
166	Mauritania	39.24	38
167	Benin	38.17	39
168	Afghanistan	37.74	22
169	Pakistan	37.50	23
170	Angola	37.44	40
171	Central African Republic	36.42	41
172	Niger	35.74	42
173	Lesotho	33.78	43
174	Haiti	33.74	12
175	Madagascar	33.73	44
176	Nepal	31.44	24
177	India	30.57	25
178	Dem. Rep. Congo	30.41	45
179	Bangladesh	29.56	26
180	Burundi	27.43	46

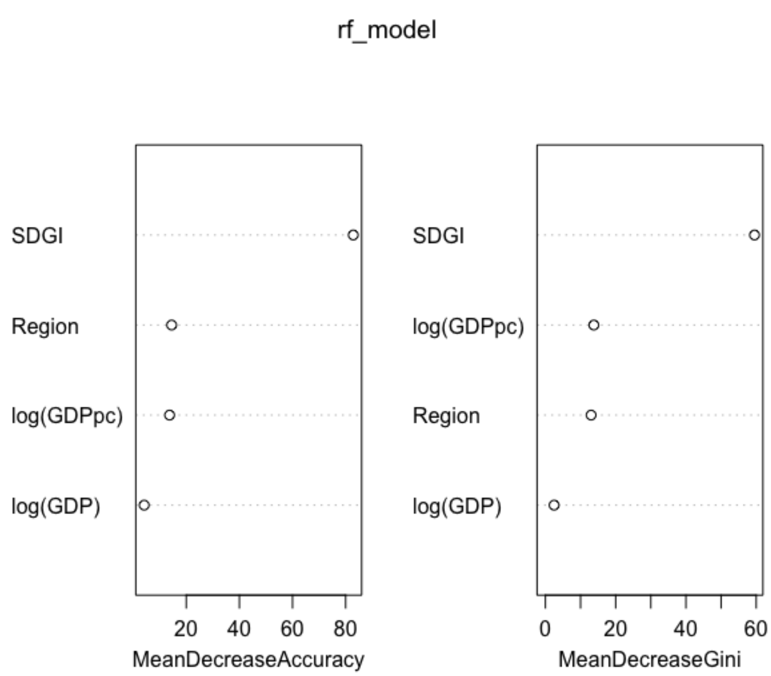
ASIA · CARIBBEAN · EASTERN EUROPE & EURASIA · EUROPE & NORTH AMERICA · LATIN AMERICA · MIDEAST & NORTH AFRICA · PACIFIC · SUB-SAHARAN AFRICA

Feature Selection

Embedded Feature Selection as Part of Training Using package randomForest in R

```
> rf_model <- randomForest(EcoQuality ~ log(GDP) + log(GDPpc) + SDGI + Region,  data = epiS, importance = TRUE)
> rf_model

Call:
 randomForest(formula = EcoQuality ~ log(GDP) + log(GDPpc) + SDGI +      Region, data = epiS, importance = TRUE
)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 2

        OOB estimate of  error rate: 1.29%
Confusion matrix:
       high low medium class.error
high     51   0      1  0.01923077
low       0  19      0  0.00000000
medium    0   1     83  0.01190476
```

```
> rf_model$importance
                  high        low      medium MeanDecreaseAccuracy MeanDecreaseGini
log(GDP)   0.001813651 0.02968824 0.002538592          0.005197546         2.280117
log(GDPpc) 0.042645347 0.14934271 0.003212928          0.033080449        13.823242
SDGI       0.590535509 0.76510617 0.351207290          0.474183200        61.013282
Region     0.041700208 0.10834696 0.010529920          0.032841199        12.036416
```



Interpretation:

Here, the “randomForest” function is first used to implement a random forest machine learning algorithm that has feature selection embedded in the dataset. The first output shows the results of the random forest, including the number of trees generated, the confusion matrix, and an error rate, as well as the class error rate for each species. The embeddedfeature selection can then be used to extract the most viable features by using the “importance” function, which shows the drop in Mean Accuracy for each of the four variables calculated as part of the random forest algorithm. By plotting the model, we further confirm that SDGI, GDPpc and Region are deemed as the most important features.

Checking for Multicollinearity

```
> car::vif(lm(EPI2018Score ~ SDGI + log(GDPpc) + Region, data=epiS))
               GVIF Df GVIF^(1/(2*Df))
SDGI       6.122059  1        2.474279
log(GDPpc) 3.702721  1        1.924246
Region     4.481649  7        1.113093
```

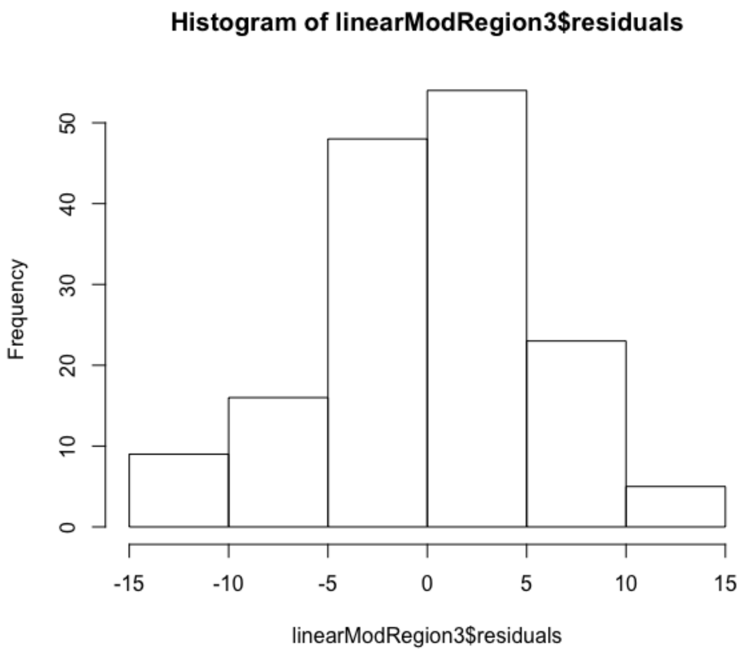
Linear Model Selection

(Stargazer comparison plot)

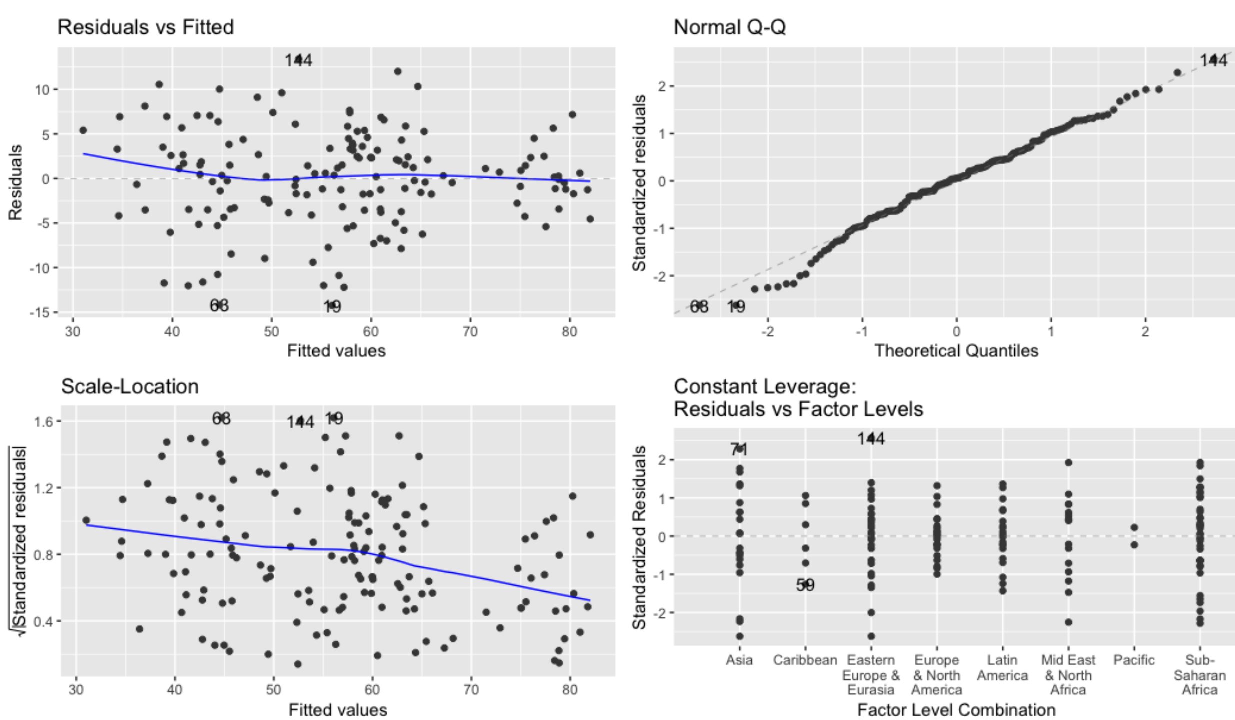
Final Model

```
> linearModRegion3 <- lm(EPI2018Score ~ SDGI + log(GDPpc) + Region, data=epiS)
> summary(linearModRegion3)

Call:
lm(formula = EPI2018Score ~ SDGI + log(GDPpc) + Region, data = epiS)

Residuals:
     Min       1Q   Median       3Q      Max 
-14.2327  -3.4417   0.3054   3.3395  13.3987 

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)    
(Intercept)                   -16.3095     5.8520  -2.787 0.006032 ** 
SDGI                            0.5144     0.1027   5.007 1.58e-06 ***
log(GDPpc)                      3.5762     0.7076   5.054 1.28e-06 ***
RegionCaribbean                 6.9012     2.5784   2.677 0.008296 ** 
RegionEastern Europe & Eurasia  5.3030     1.7414   3.045 0.002762 ** 
RegionEurope & North America   15.6033     2.0950   7.448 7.81e-12 ***
RegionLatin America             6.9091     1.7509   3.946 0.000123 ***
RegionMid East & North Africa   7.7672     1.9096   4.067 7.77e-05 ***
RegionPacific                  14.0086     4.2553   3.292 0.001249 ** 
RegionSub-Saharan Africa        5.2940     1.8139   2.919 0.004077 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.558 on 145 degrees of freedom
Multiple R-squared:  0.8379,	Adjusted R-squared:  0.8279 
F-statistic: 83.31 on 9 and 145 DF,  p-value: < 2.2e-16
```

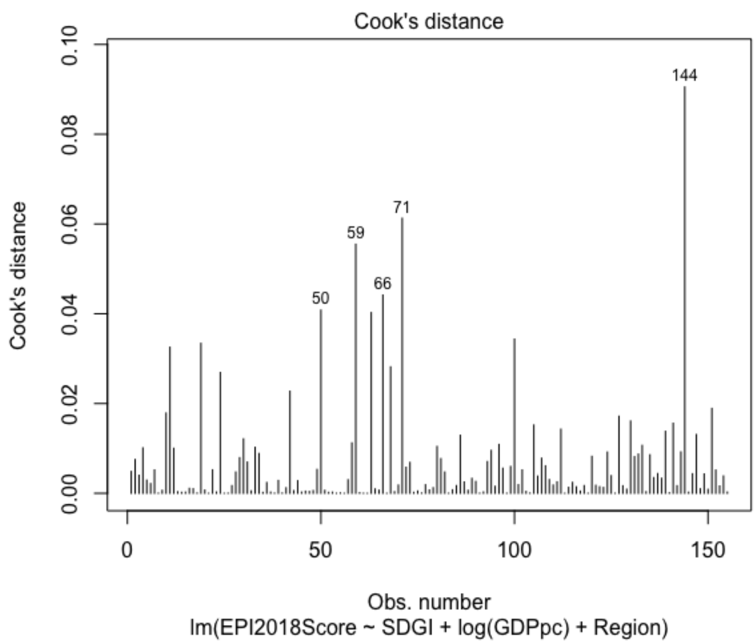
Checking Assumptions:



> ggplot2::autoplot(linearModRegion3)



plot(linearModRegion3, 4, id.n = 5)



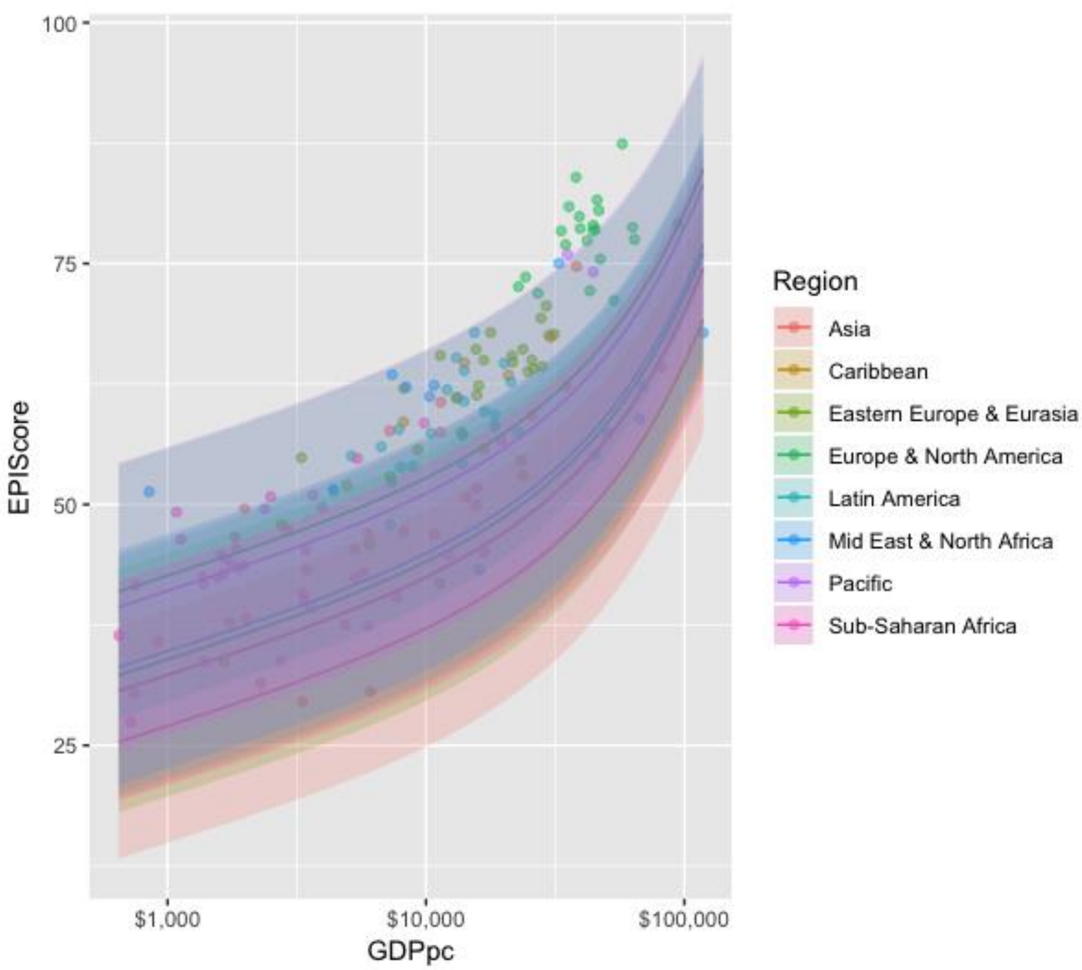
The 144th observation here seems to be the extreme outlier. This is the country of Turkmenistan. Another outlier seems to be Japan(71). I am guessing that this is due to regional differences and high SDGI and GDP per capita especially for Japan, as being an Asian country, it does not seem to fit in the average regional values with an EPI Score of 74.69 and a high SDGI of 80.2. I decided to keep the outliers since I did not find any reason to exclude them based on missing or incorrect data.

List of 37 countries not included in the 2018 SDG Index due to insufficient data availability:

Country	Missing Values	Percentage of Missing Values
Andorra	40	49%
Antigua and Barbuda	33	38%
Bahamas, The	27	31%
Barbados	19	22%
Brunei Darussalam	17	20%
Comoros	19	22%
Dominica	41	47%
Equatorial Guinea	25	29%
Eritrea	18	21%
Fiji	20	23%
Grenada	34	39%
Guinea-Bissau	19	22%
Kiribati	38	44%
Korea, Dem. Rep.	26	30%
Libya	18	21%
Liechtenstein	54	67%
Maldives	17	20%
Marshall Islands	45	52%
Micronesia, Fed. Sts.	42	48%
Monaco	52	60%
Nauru	55	63%
Palau	52	60%
Papua New Guinea	20	23%
Samoa	30	34%
San Marino	58	67%
Sao Tome and Principe	17	20%
Seychelles	29	33%
Solomon Islands	23	26%
Somalia	19	22%
South Sudan	23	28%
St. Kitts and Nevis	49	56%
St. Lucia	27	31%
St. Vincent and the Grenadines	31	36%
Timor-Leste	20	23%
Tonga	32	37%
Tuvalu	53	61%
Vanuatu	19	22%

Predictions Using the Created Model

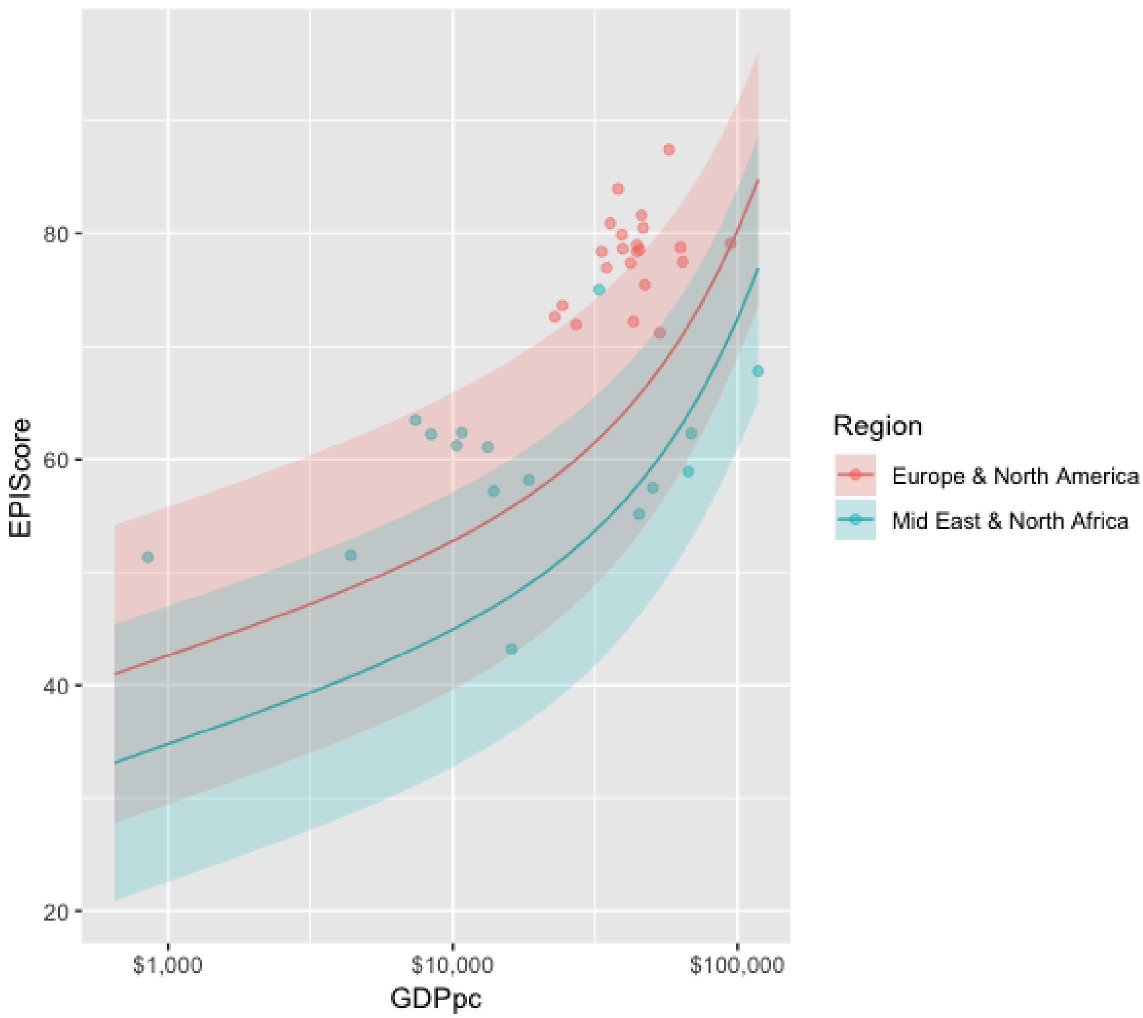
```
> newdata2$EPIScore <- predict(linearModRegion3, newdata = newdata2, type = "response")
> head(newdata2)
      SDGI    GDPpc Region EPIScore
1 36.00000  647.000   Asia 25.35663
2 36.49495 1834.465   Asia 29.33822
3 36.98990 3021.929   Asia 31.37787
4 37.48485 4209.394   Asia 32.81772
5 37.97980 5396.859   Asia 33.96102
6 38.47475 6584.323   Asia 34.92685
```

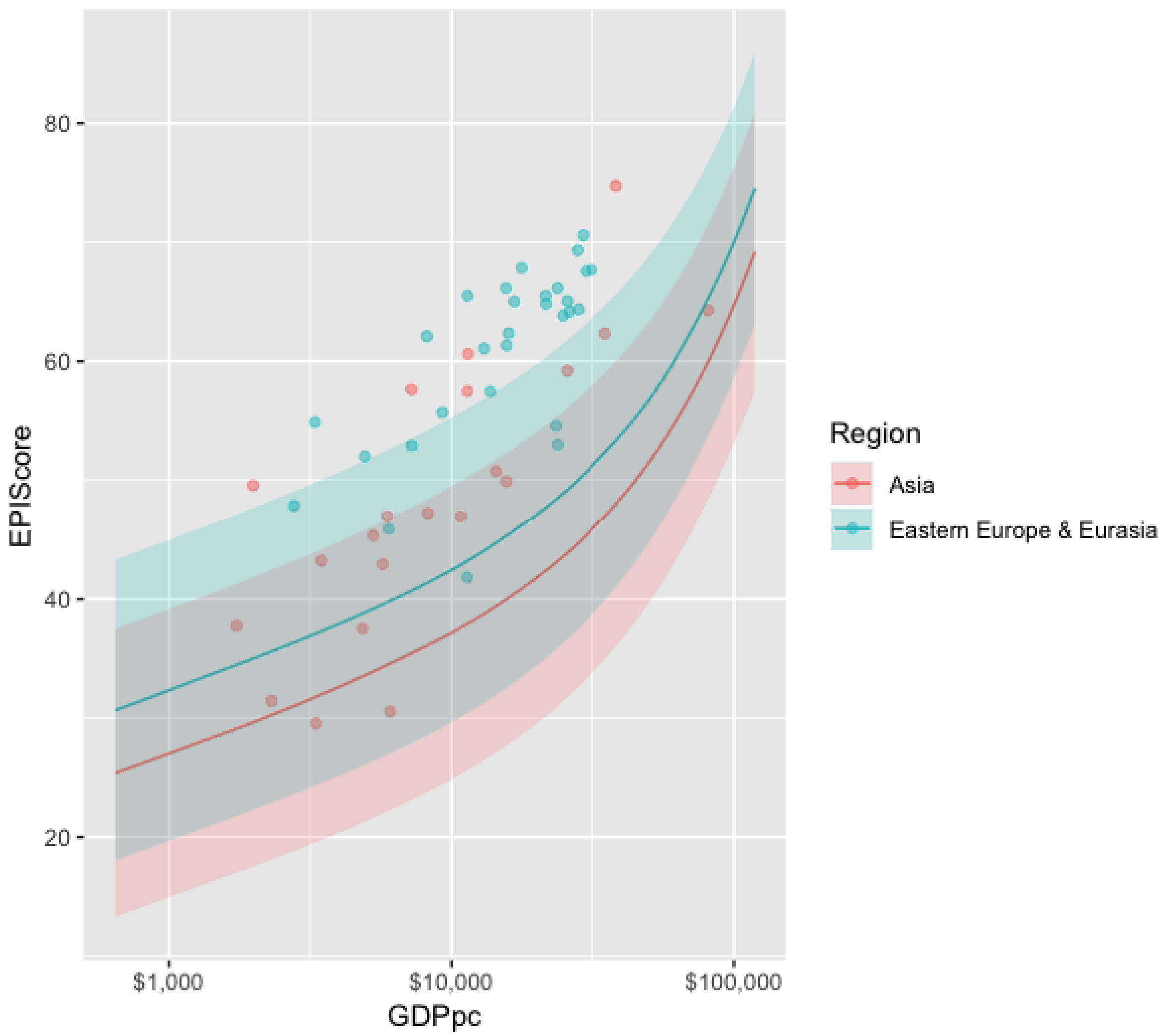


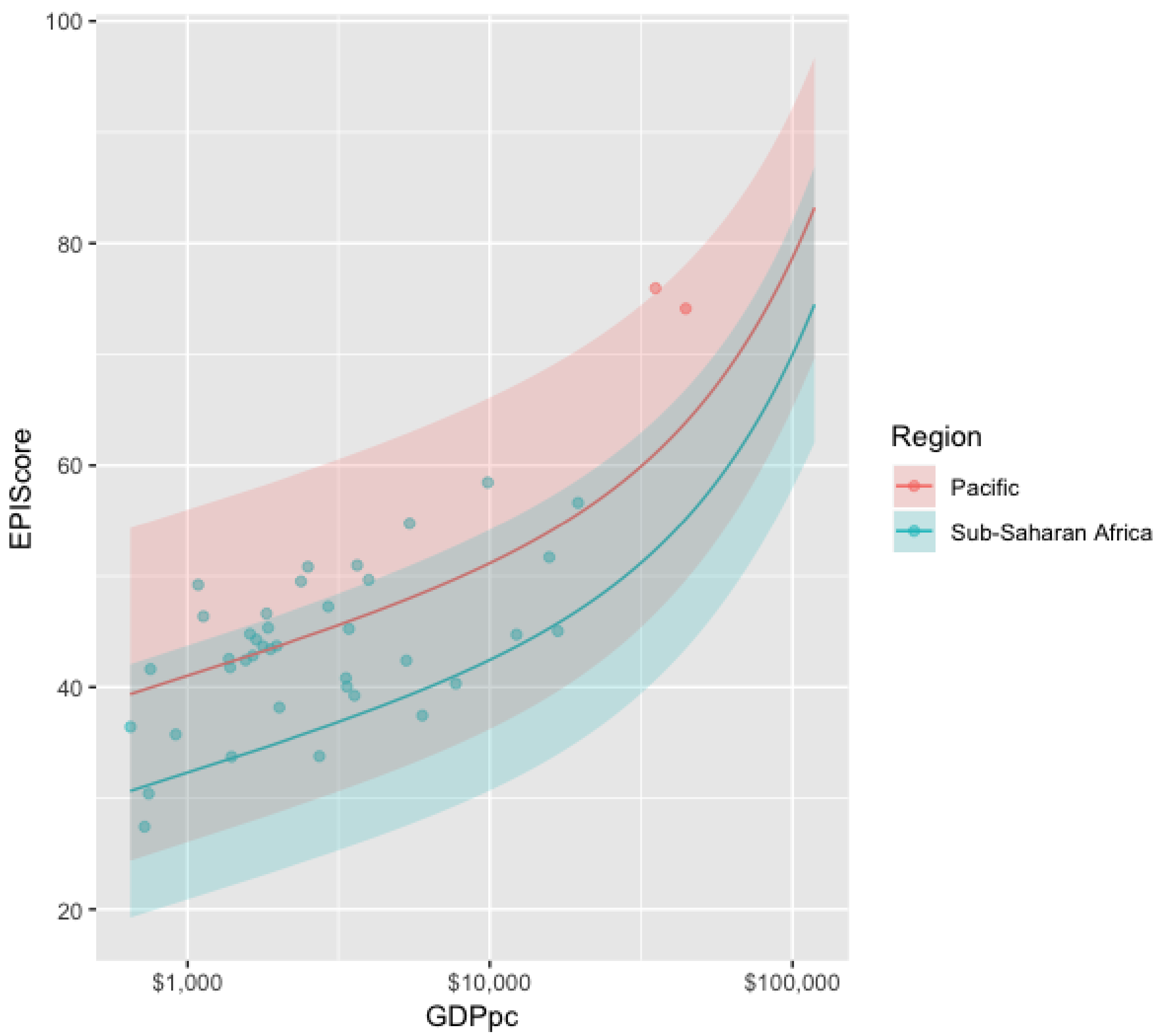
As we can see from the graph above, as the GDP per capita increases for each of the regions, the Environmental Quality Index Score also increase.

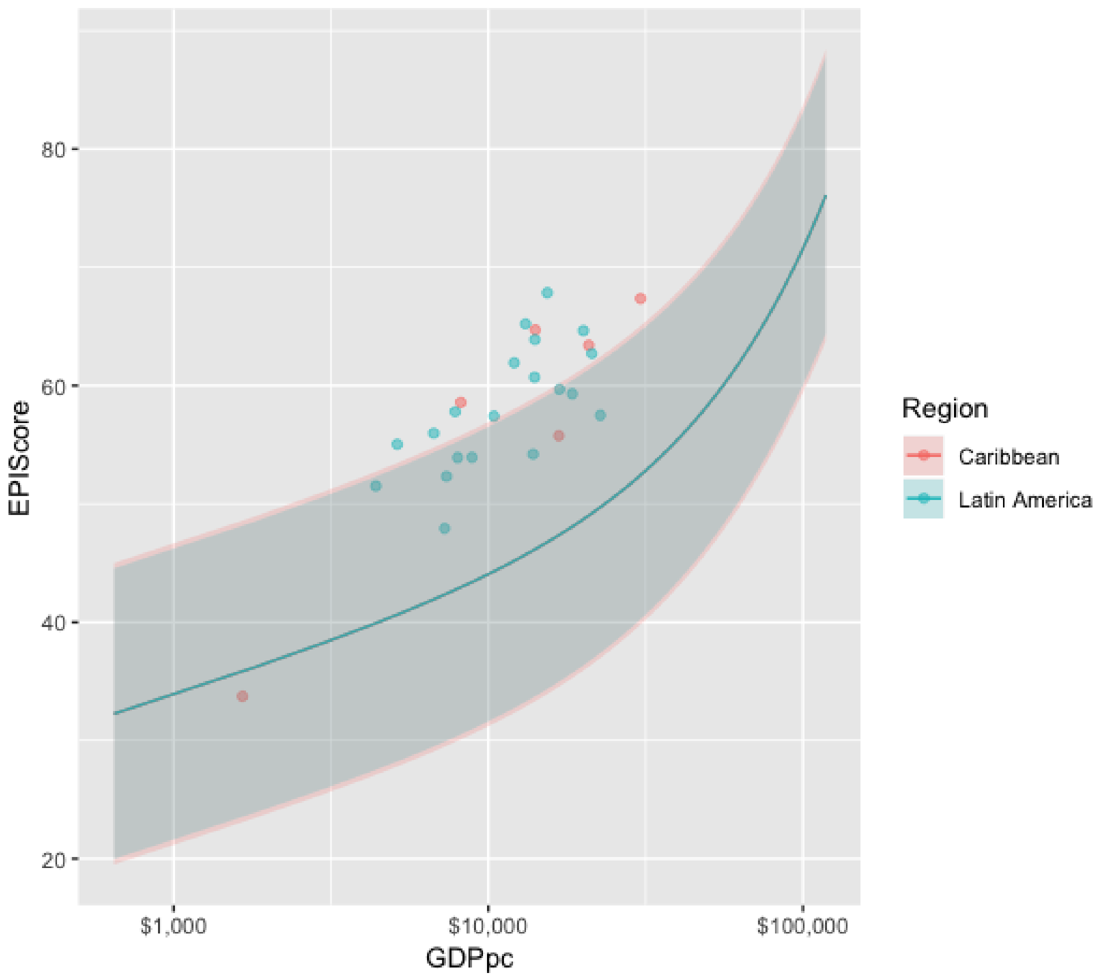
To make the CI graph clearer, I plotted confidence intervals of the predicted probabilities for only two regions on each graph.

DO NOT SHARE THIS WITH ANYONE… COPYRIGHT PROTECTED

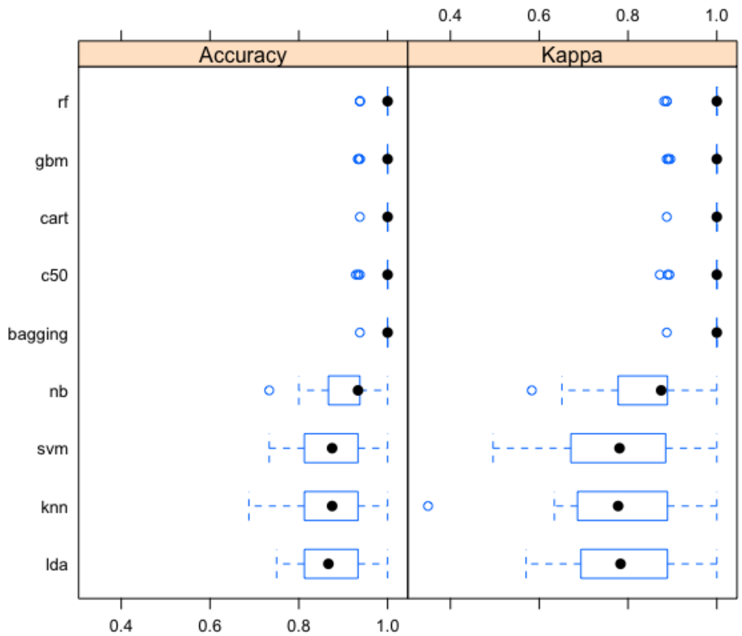


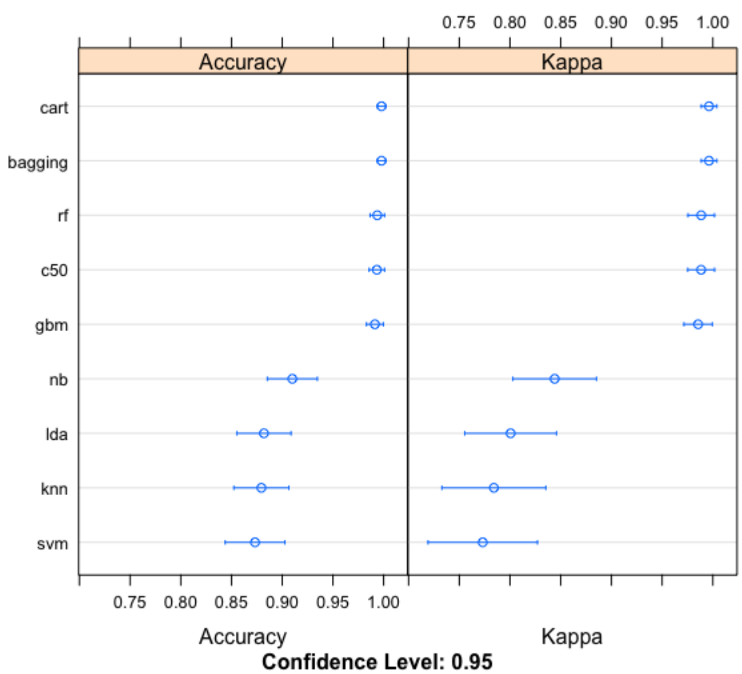

80
60
40
20
EPIScore
$1,000
$10,000
$100,000
GDPpc
Region
Asia
Eastern Europe & Eurasia

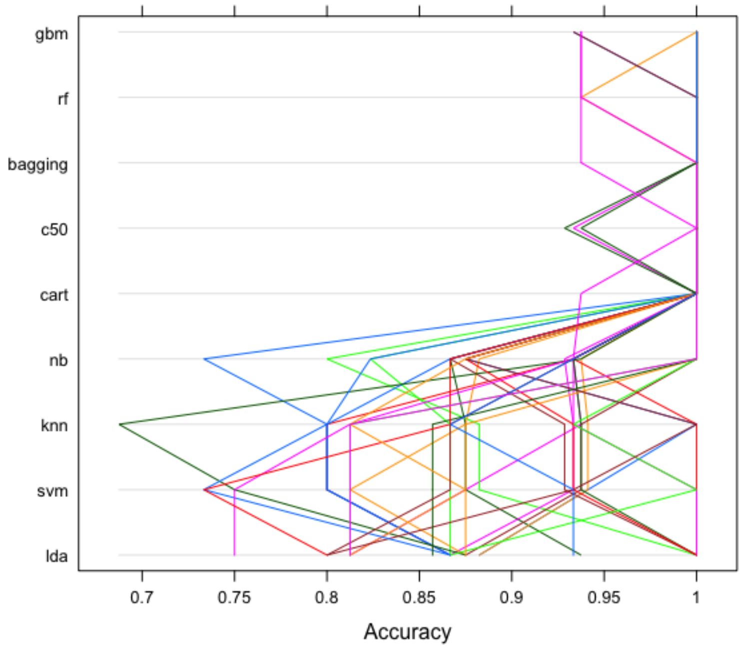

100
80
60
40
20
EPIScore
$1,000
$10,000
$100,000
GDPpc
Region
Pacific
Sub-Saharan Africa



Testing Other Algorithms for Prediction:


Accuracy
Kappa
rf
gbm
cart
c50
bagging
nb
svm
knn
lda
0.4
0.6
0.8
1.0


Accuracy
Kappa
cart
bagging
rf
c50
gbm
nb
lda
knn
svm
0.75
0.80
0.85
0.90
0.95
1.00
Confidence Level: 0.95



Conclusion & Analysis of Results

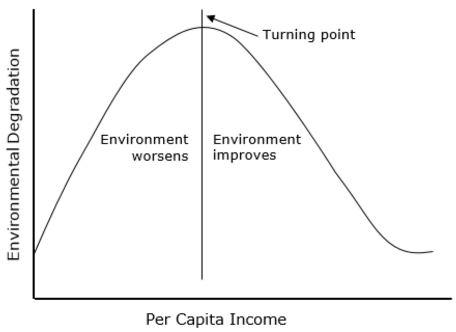
In conclusion, the measures I have used clearly show that the EPI Scores are positively correlated with GDP per capita and SDGI; however, the methodology of how the SDGI is determined as been criticized. Hence, I am left with a few questions that are beyond the scope of this project and would like to further research:

Questions

As the Environmental Performance Index is also directly linked to a higher GDP per capita as well as SDGI, does this mean that countries that have a higher Ecological Footprint per person (generally countries with a high GDPpc) are taking up resources to create a better environment? If so, wouldn't the higher Ecological Footprint per person counter the EPI Index by making the air quality, sanitation, forest conservation, marine life and other indicator of environmental health more sound?

Are the EPI indicators in richer countries being met at the expense of taking resources from poor nations. Thus, importing resources from poor nations to richer nations and depleting their resources to sustain a better environment? Or did the Kuznet's curve that I discussed in the beginning really is a true indicator of better environment, as some of the measures on the EPI Score such as Air Quality and Wastewater treatment really do improve with more wealth. If so, does the concerns raised by researcher from the Global Footprint Network still hold true: "Ignoring physical constraints imposed by planetary limits is anti-poor because with fewer resources to go around, the lowest-income people will lack the financial means to shield themselves from resource constraints, whether it is food-price shocks, weather calamities, or energy and water shortages."

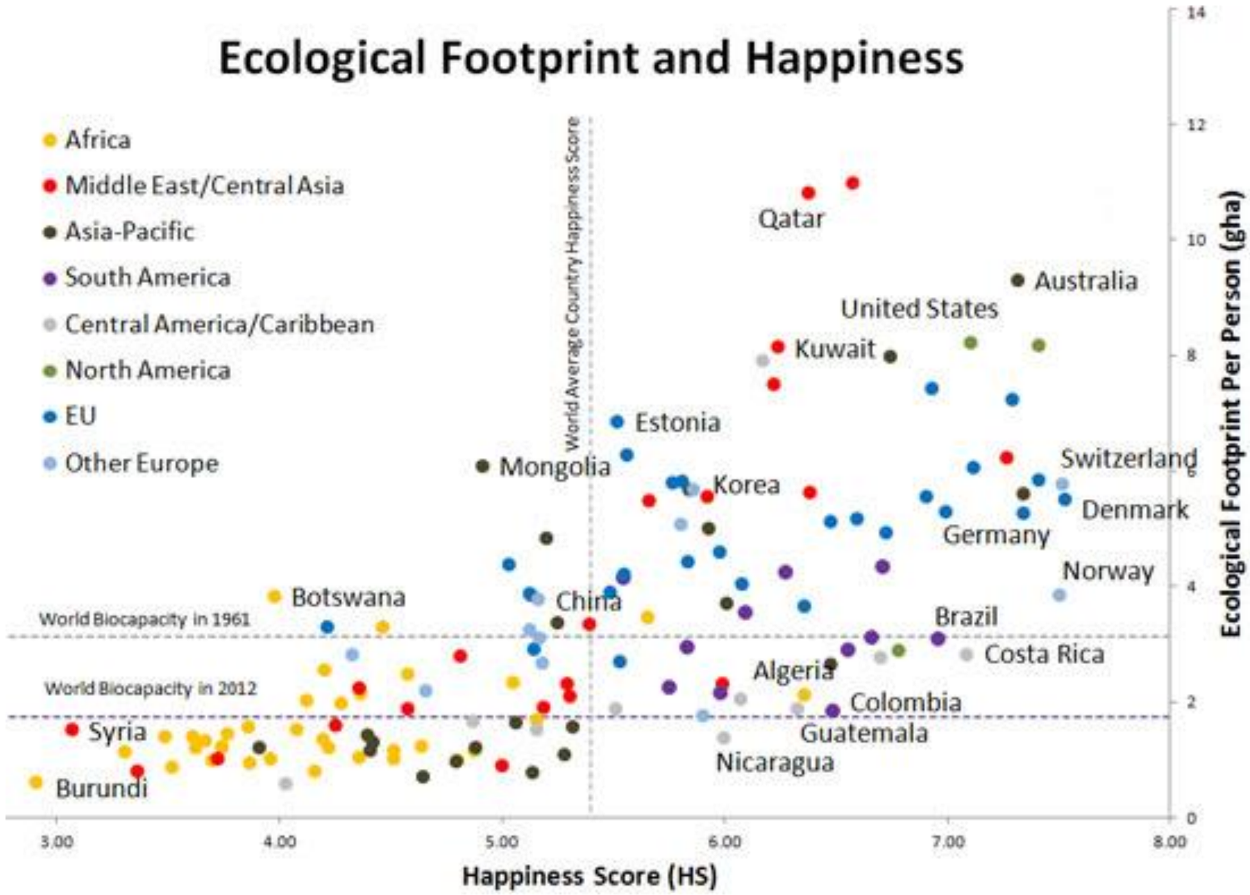
Or, should attention on U.N.'s Human Development Index and the fact that the SDGs are framed around the HDI is really the solution. In other words, should we concentrate on increasing nations' wealth at the expense of a higher Global Footprint per person in order to achieve higher EPI Scores. This also takes me back to the environmental Kuznet's curve; maybe we still haven't reached high enough levels of income increase or the turning point on the EKC where the Ecological Footprint starts decreasing.

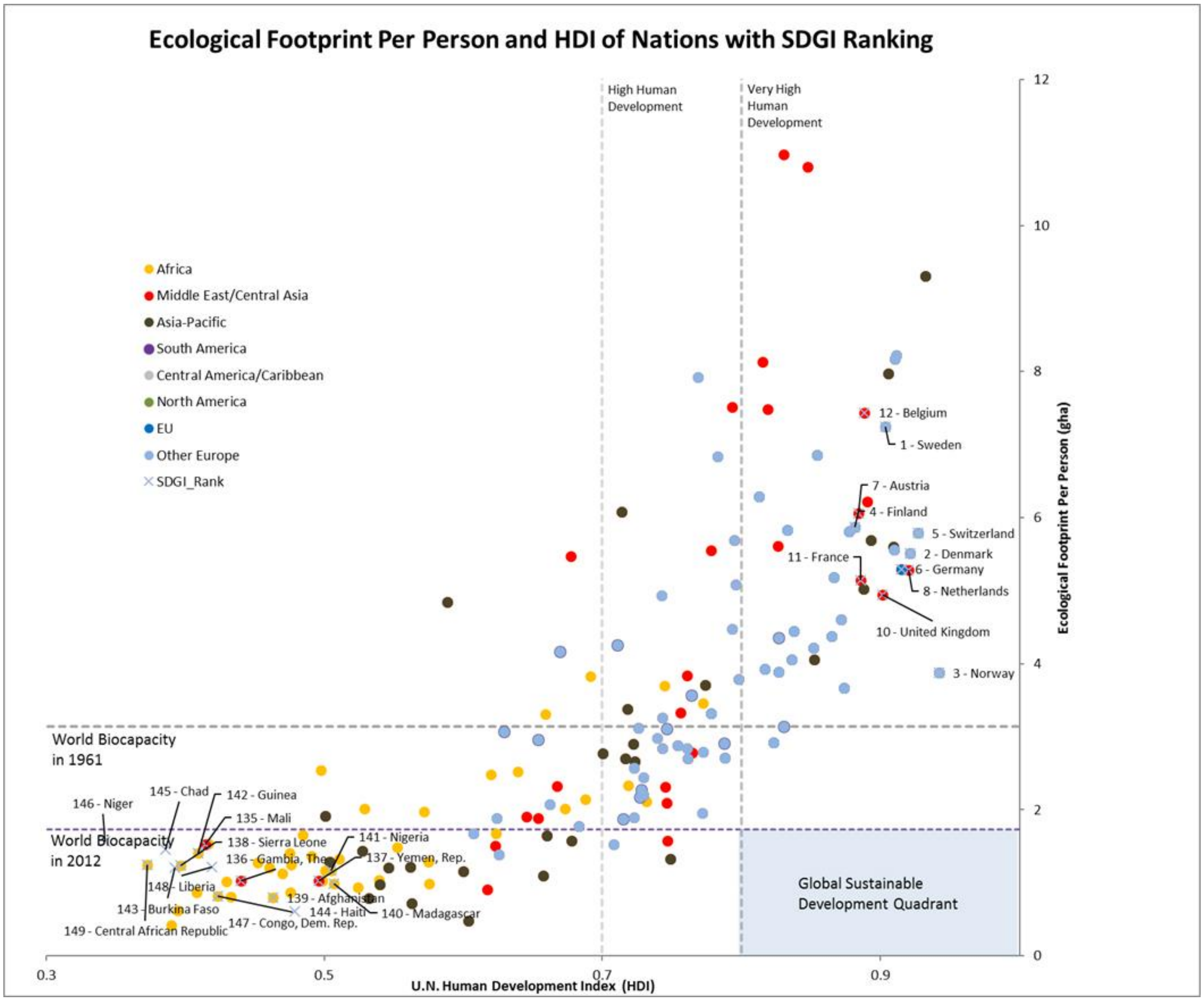


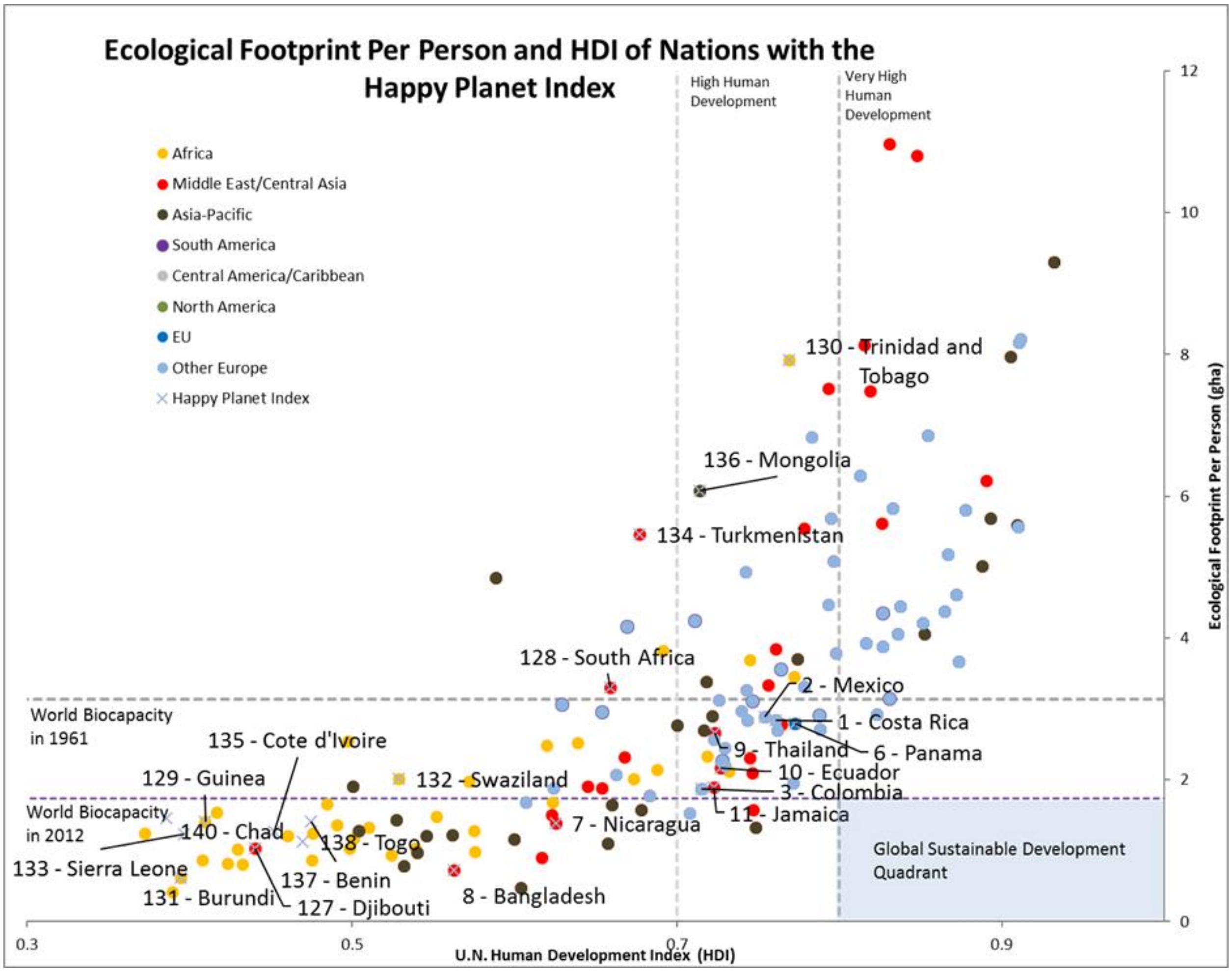
Additional Insights

Another factor that is not considered in the Environmental Score rankings is the Ecological Footprint per Person.

DO NOT SHARE THIS WITH ANYONE… COPYRIGHT PROTECTED




Ecological Footprint Per Person and HDI of Nations with SDGI Ranking
High Human Development
Very High Human Development
Africa
Middle East/Central Asia
Asia-Pacific
South America
Central America/Caribbean
North America
EU
Other Europe
SDGI_Rank
12 - Belgium
1 - Sweden
7 - Austria
4 - Finland
5 - Switzerland
2 - Denmark
11 - France
6 - Germany
8 - Netherlands
10 - United Kingdom
3 - Norway
World Biocapacity in 1961
World Biocapacity in 2012
145 - Chad
142 - Guinea
146 - Niger
135 - Mali
138 - Sierra Leone
136 - Gambia, The
141 - Nigeria
137 - Yemen, Rep.
148 - Liberia
139 - Afghanistan
144 - Haiti
140 - Madagascar
143 - Burkina Faso
147 - Congo, Dem. Rep.
149 - Central African Republic
Global Sustainable Development Quadrant
Ecological Footprint Per Person (gha)
U.N. Human Development Index (HDI)
0.3
0.5
0.7
0.9
0
2
4
6
8
10
12



Appendix

R Libraries Used

```
> library(ggplot2)
> library(broom)
> library(margins)
> library(tidyverse)
> library(dplyr)
> library(ggfortify)
> library(mlbench)
> library(caret)
```

Source code

Predictions and Confidence Intervals

First, holding SDGI and GDP per capita at their means, the probability of the EPI Score is calculated for each region:

> newdata1 <- with(epiS, data.frame(SDGI = mean(SDGI), GDPpc = mean(log(GDPpc)), Region = levels(Region)))

> newdata1$EPIScore <- predict(linearModRegion3, newdata = newdata1, type = "response")

```
> newdata1
      SDGI    GDPpc                  Region EPIScore
1 64.98129 9.221074                    Asia 25.06385
2 64.98129 9.221074               Caribbean 31.96505
3 64.98129 9.221074 Eastern Europe & Eurasia 30.36682
4 64.98129 9.221074   Europe & North America 40.66716
5 64.98129 9.221074          Latin America 31.97299
6 64.98129 9.221074  Mid East & North Africa 32.83108
7 64.98129 9.221074                 Pacific 39.07242
8 64.98129 9.221074      Sub-Saharan Africa 30.35785
```

newdata2 <- with(epiS, data.frame(SDGI = rep(seq(from = 36, to = 85, length.out = 100), 8), GDPpc = rep(seq(from = 647, to = 118206, length.out = 100), 8), Region = factor(rep(levels(Region), each = 100))))

Algorithm Comparison

> library(mlbench)
> library(caret)

> control <- trainControl(method="repeatedcv", number=10, repeats=3)
> seed <- 7
> metric <- "Accuracy"
> set.seed(seed)

> fit.lda <- train(EcoQuality~ log(GDPpc) + SDGI + Region, data=epiS, method="lda", metric=metric, preProc=c("center", "scale"), trControl=control)

> fit.svmRadial <- train(EcoQuality~ log(GDPpc) + SDGI + Region, data=epiS, method="svmRadial", metric=metric, preProc=c("center", "scale"), trControl=control, fit=FALSE)

```
> fit.knn <- train(EcoQuality~ log(GDPpc) + SDGI + Region, data=epiS, method="knn",
metric=metric, preProc=c("center", "scale"), trControl=control)

> fit.nb <- train(EcoQuality~ log(GDPpc) + SDGI, data=epiS, method="nb", metric=metric,
trControl=control)

> fit.cart <- train(EcoQuality~ log(GDPpc) + SDGI + Region, data=epiS, method="rpart",
metric=metric, trControl=control)

> fit.treebag <- train(EcoQuality~ log(GDPpc) + SDGI + Region, data=epiS, method="treebag",
metric=metric, trControl=control)

> set.seed(seed)
> fit.rf <- train(EcoQuality~ log(GDPpc) + SDGI + Region, data=epiS, method="rf", metric=metric,
trControl=control)

> set.seed(seed)
> fit.gbm <- train(EcoQuality~ log(GDPpc) + SDGI + Region, data=epiS, method="gbm",
metric=metric, trControl=control, verbose=FALSE)

> results <- resamples(list(lda=fit.lda,
+         svm=fit.svmRadial, knn=fit.knn, nb=fit.nb, cart=fit.cart, c50=fit.c50,
+         bagging=fit.treebag, rf=fit.rf, gbm=fit.gbm))
> summary(results)
```