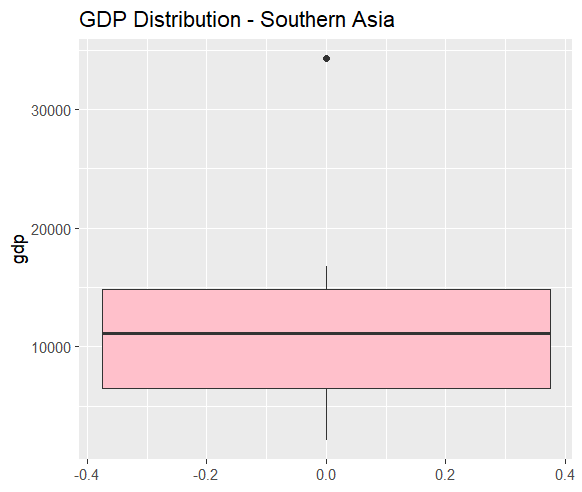
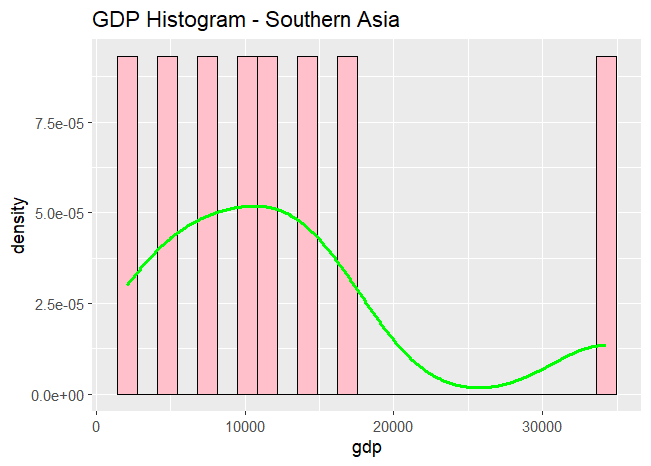
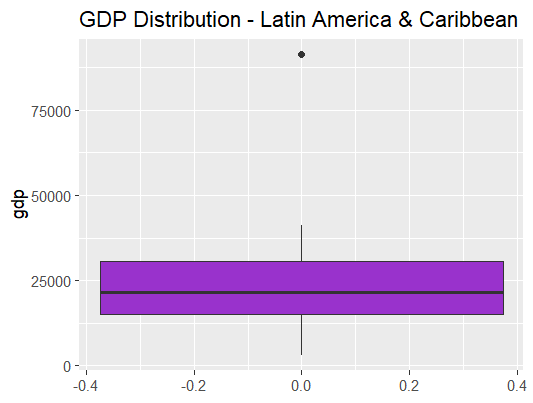
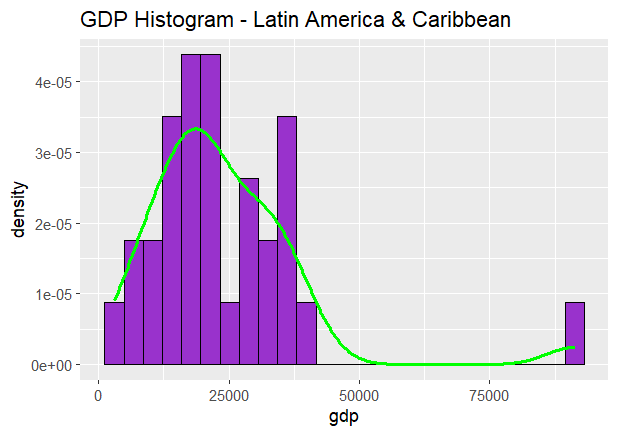
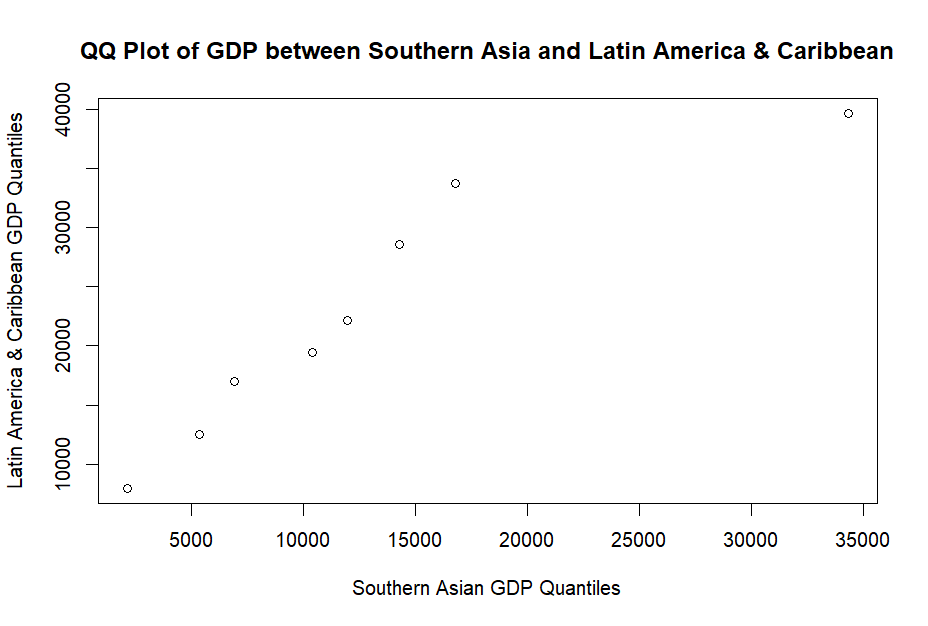
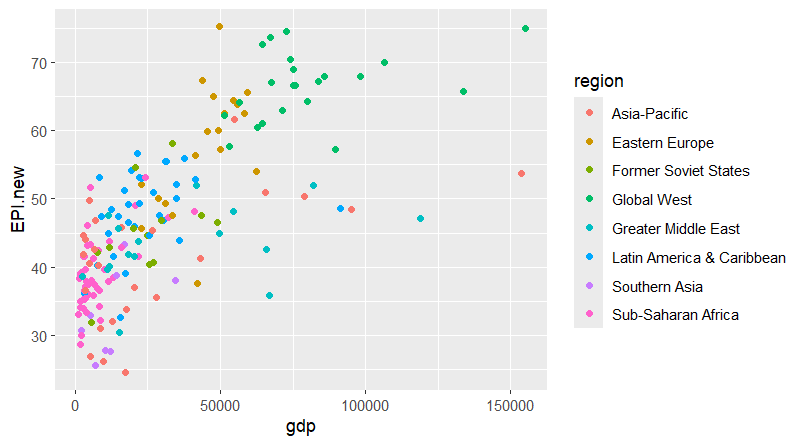
1. **Variable Distributions**
   1. **1.1**
      1. 
      2. 
      3. 
      4. 
   2. **1.2**
      1. 
2. **Linear Models**
   1. **2.1**
      1. 
      2. Call: lm(formula = EPI.new ~ log\_gdp, data = epi.data)

Residuals:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Min | 1Q | Median | 3Q | Max |
| -22.4327 | -4.9156 | 0.0425 | 6.2220 | 20.8989 |

Coefficients:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| Intercept | -22.338 | 5.007 | -4.462 | -1.44e-05 |
| log\_gdp | 16.340 | 1.174 | 13.913 | < 2e-16 |

---

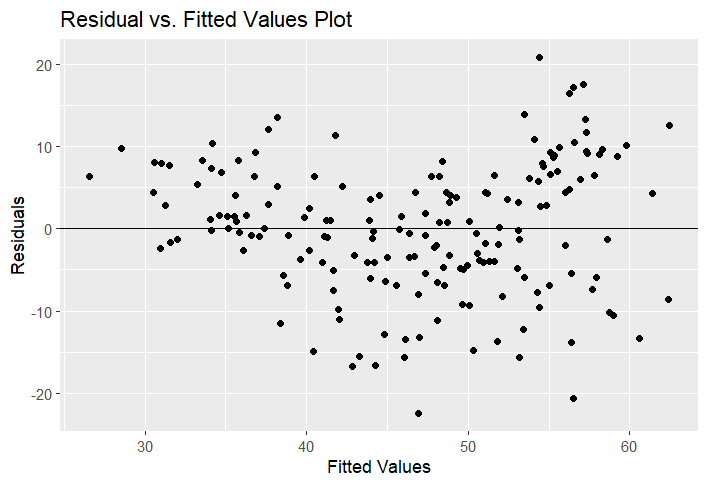
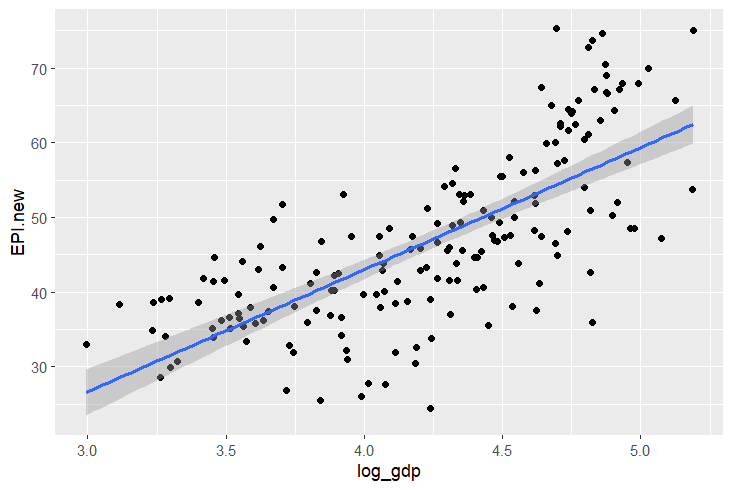
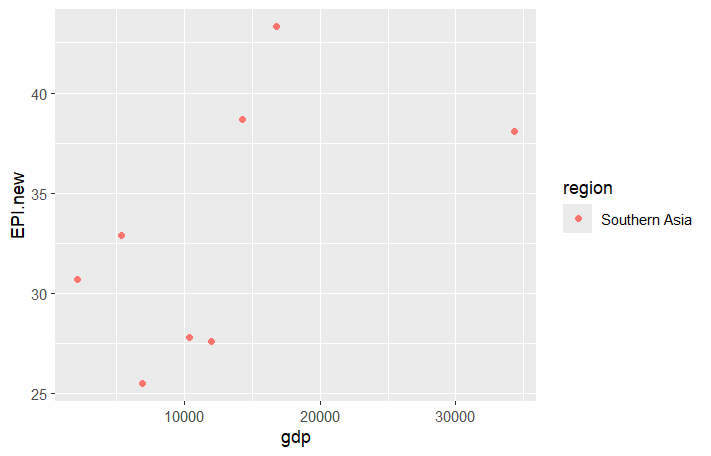
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.023 on 177 degrees of freedom

(1 observation deleted due to missingness)

Multiple R-squared: 0.5223, Adjusted R-squared: 0.5196

F-statistic: 193.6 on 1 and 177 DF, p-value: < 2.2e-16

* + 1. 
  1. **2.2**
     1. 
     2. Call: lm(formula = EPI.new ~ log\_gdp, data = region1)

Residuals:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Min | 1Q | Median | 3Q | Max |
| -6.273 | -5.710 | 1.138 | 3.682 | 8.115 |

Coefficients:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| Intercept | -2.483 | 24.631 | -0.101 | 0.923 |
| log\_gdp | 8.917 | 6.155 | 1.449 | 0.198 |

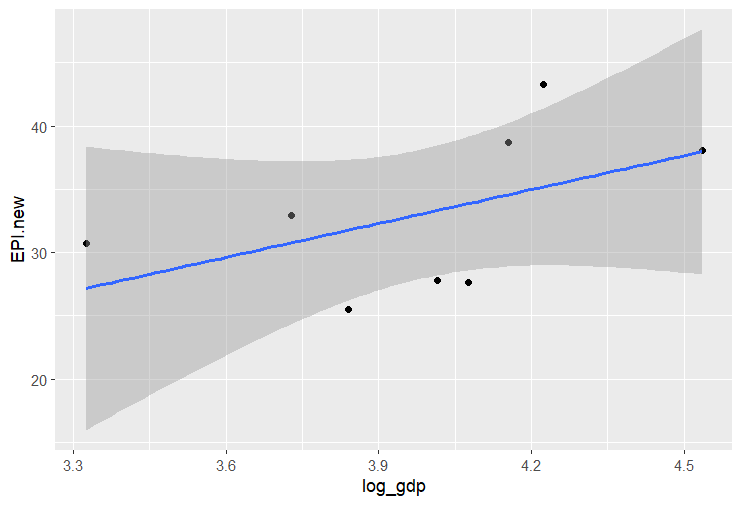
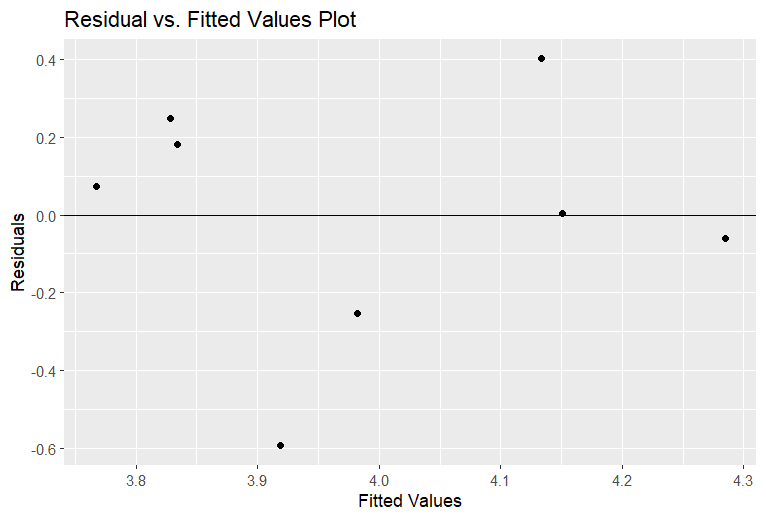
---

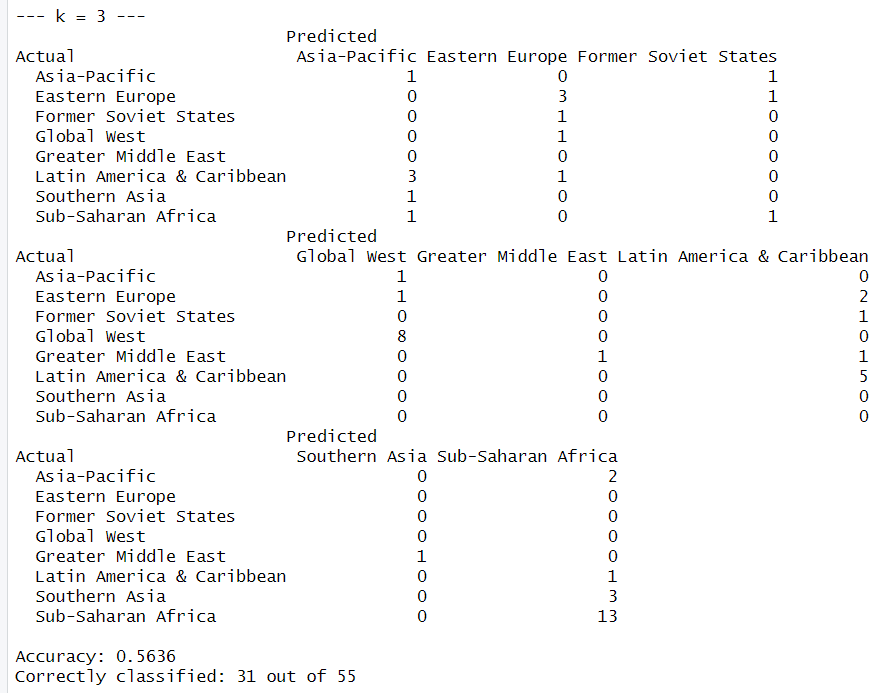
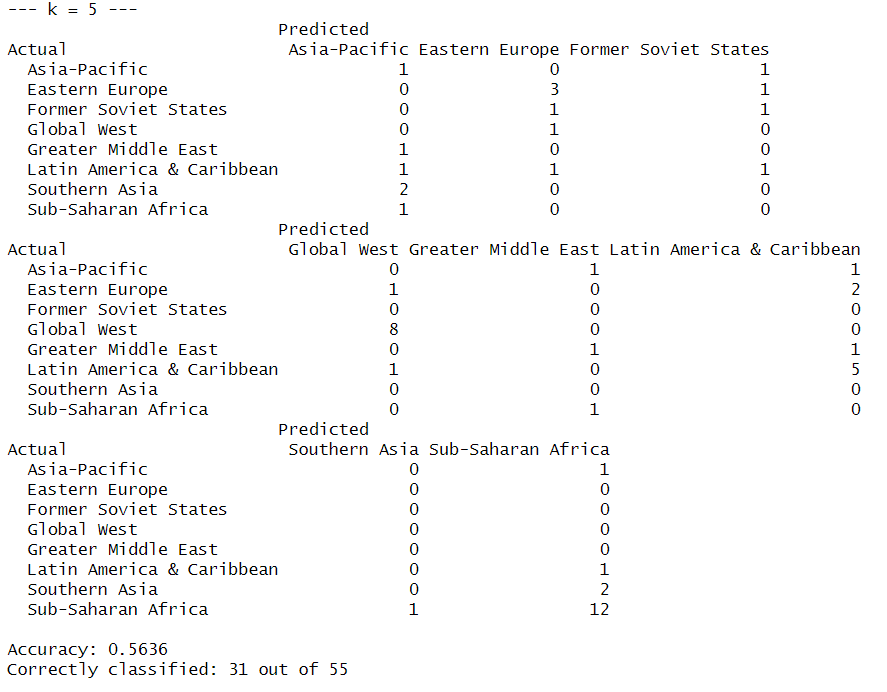
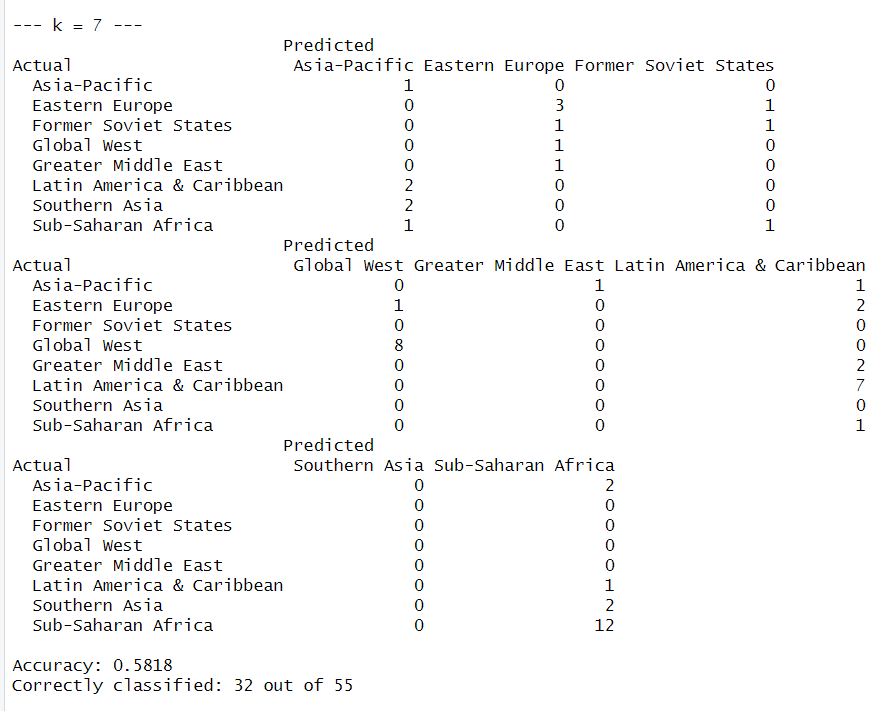
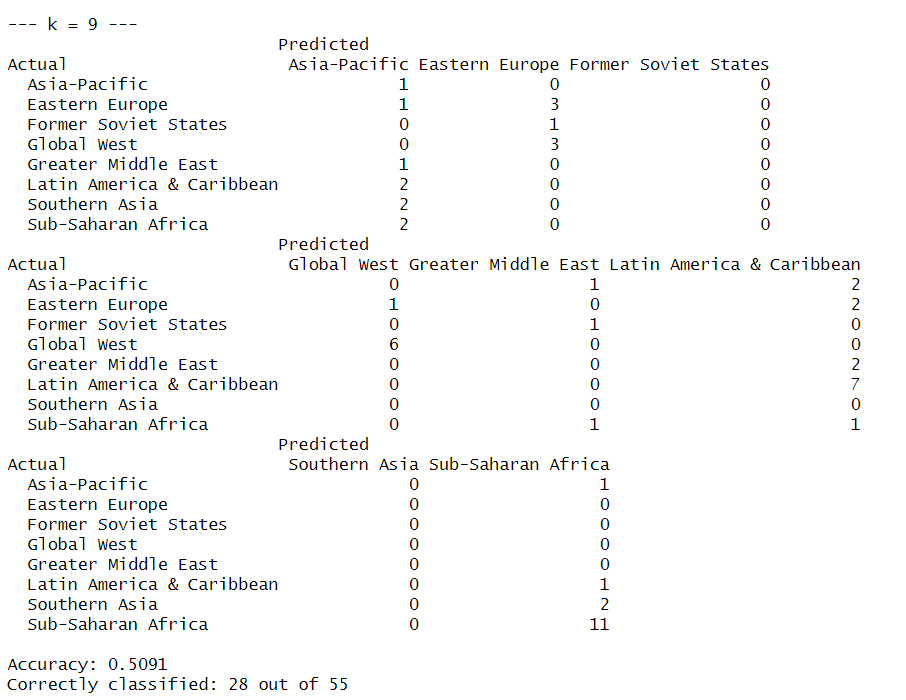
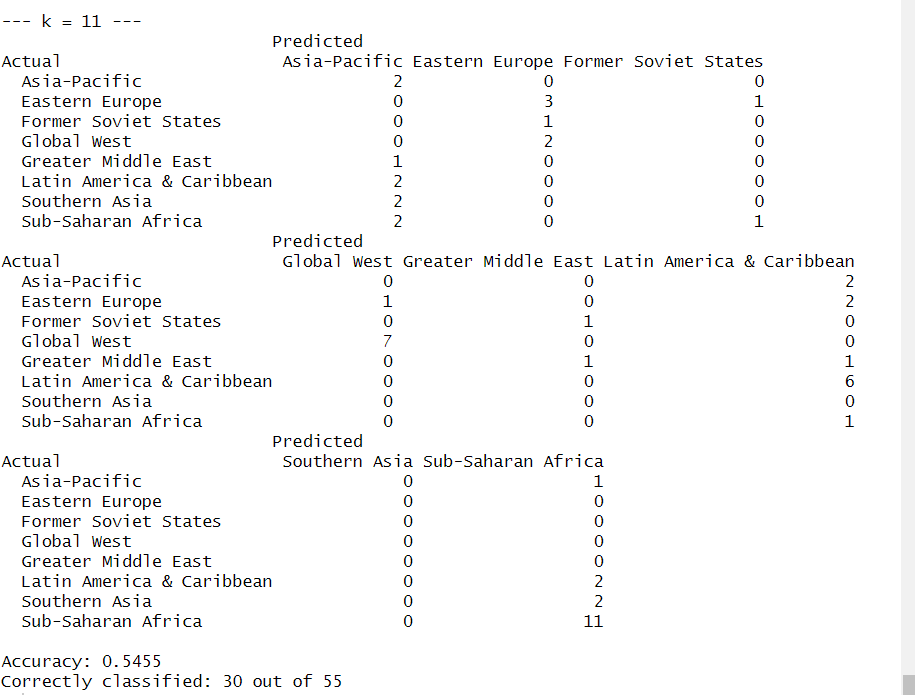
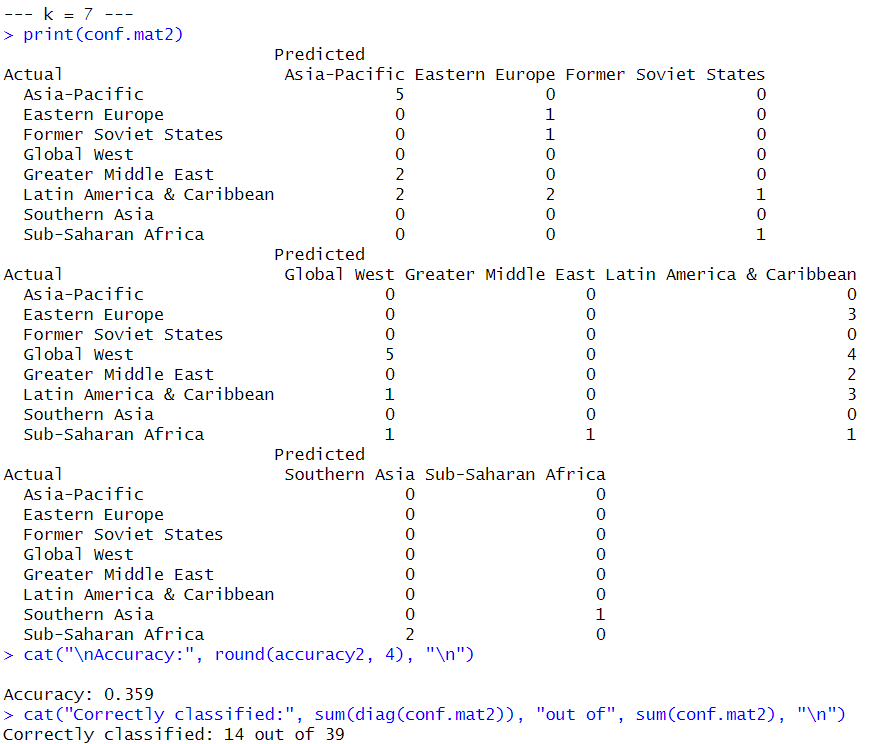
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.904 on 6 degrees of freedom

Multiple R-squared: 0.2592, Adjusted R-squared: 0.1357

F-statistic: 2.099 on 1 and 6 DF, p-value: 0.1976

* + 1. 
    2. 
  1. I think the region-specific models were a better fit because they display the information on a country-by-country basis. With the larger data set, much of the data is generalized. However, with the regions, we can break down the statistics for each country.

1. **Classification (kNN)**
   1. **3.1**
      1. 
      2. 
      3. 
      4. 
      5. 
   2. **3.2**
      1. 
   3. Model 1 (EPI.new, ECO.new, BDH.new) Accuracy: 0.5818
   4. Model 2 (MKP.new, MHP.new, MPE.new) Accuracy: 0.359
   5. CONCLUSION: Model 1 performs better with 22.28 percentage points higher accuracy, suggesting that overall environmental performance indicators (EPI, ECO, BDH) are more effective at distinguishing regions than marine and habitat protection metrics.

**Sources**

<https://www.google.com/url?sa=i&url=https%3A%2F%2Fr-graph-gallery.com%2Fcolors.html&psig=AOvVaw2dwnaqzstYmAHbtW_GQY2M&ust=1760464699659000&source=images&cd=vfe&opi=89978449&ved=0CBYQjhxqFwoTCKjmsJrgoZADFQAAAAAdAAAAABAE>