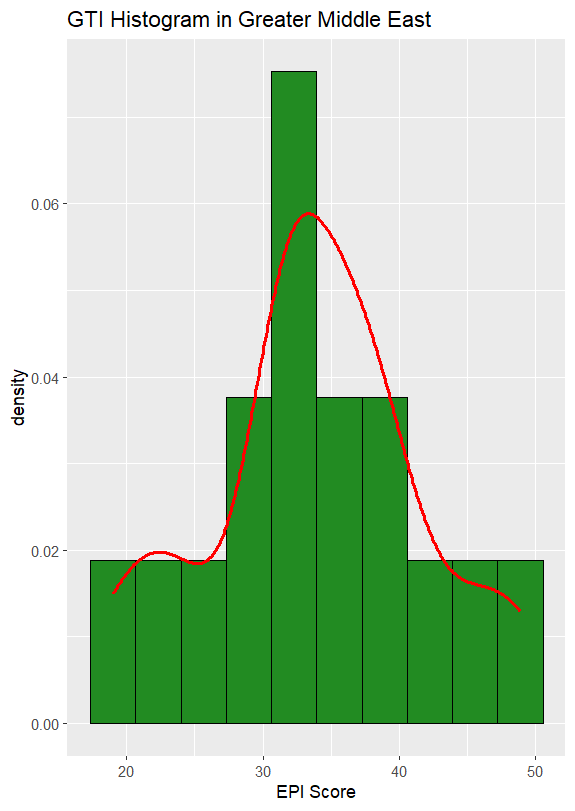
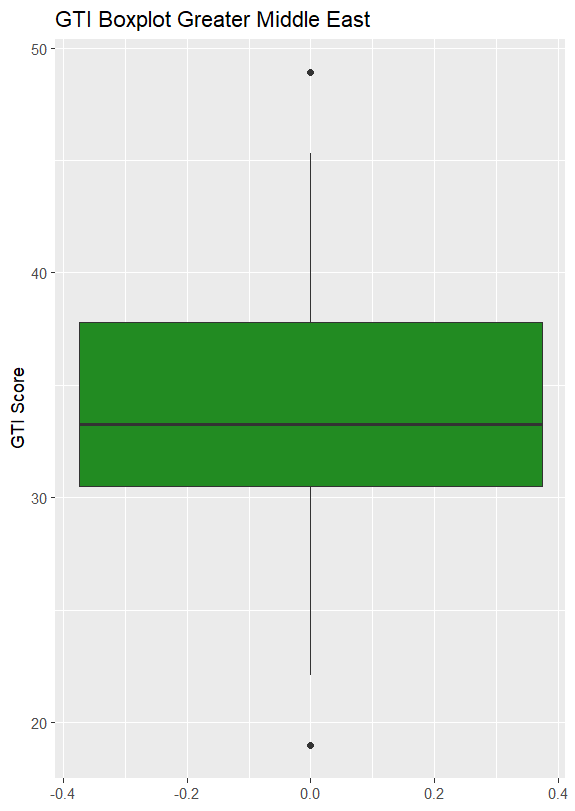
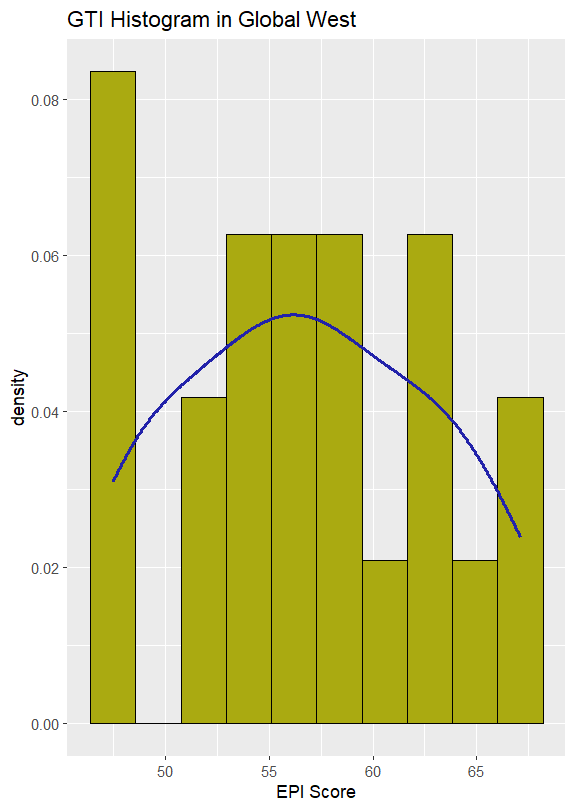
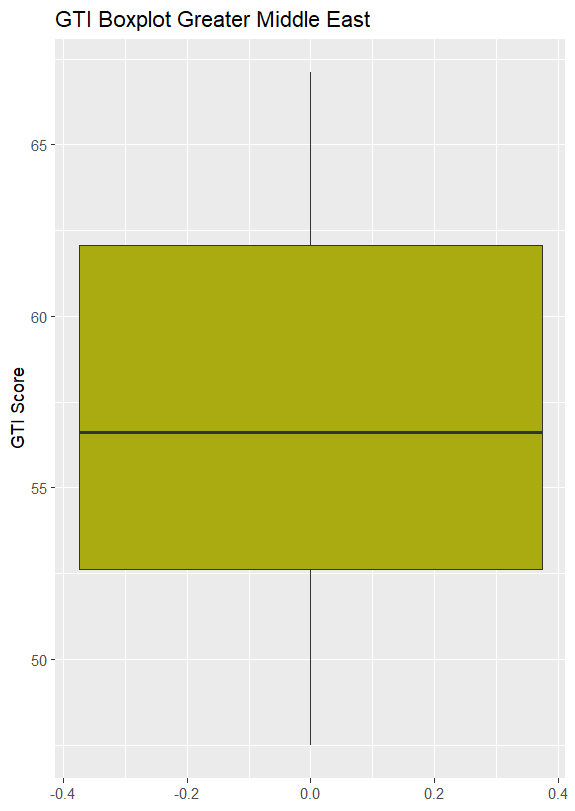
Variable Distributions:

1)

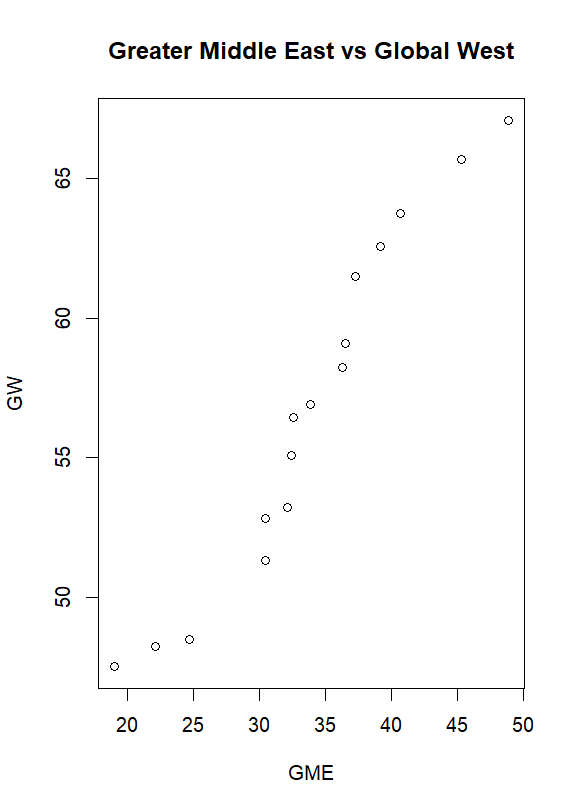
1.1:







1.2:



2)

2.1:

GTI vs GDP Model

Call:

lm(formula = GTI.new ~ gdp, data = Data)

Residuals:

Min 1Q Median 3Q Max

-36.204 -5.547 1.632 7.911 35.694

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.129e+01 1.312e+00 23.851 < 2e-16 \*\*\*

gdp 2.399e-04 3.064e-05 7.827 4.39e-13 \*\*\*

---

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

Residual standard error: 12.42 on 177 degrees of freedom

(1 observation deleted due to missingness)

Multiple R-squared: 0.2571, Adjusted R-squared: 0.2529

F-statistic: 61.26 on 1 and 177 DF, p-value: 4.389e-13

GTI vs Population Model

Call:

lm(formula = GTI.new ~ population, data = Data)

Residuals:

Min 1Q Median 3Q Max

-38.841 -8.373 0.742 9.209 40.050

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.886e+01 1.125e+00 34.544 <2e-16 \*\*\*

population -4.287e-09 6.941e-09 -0.618 0.538

---

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

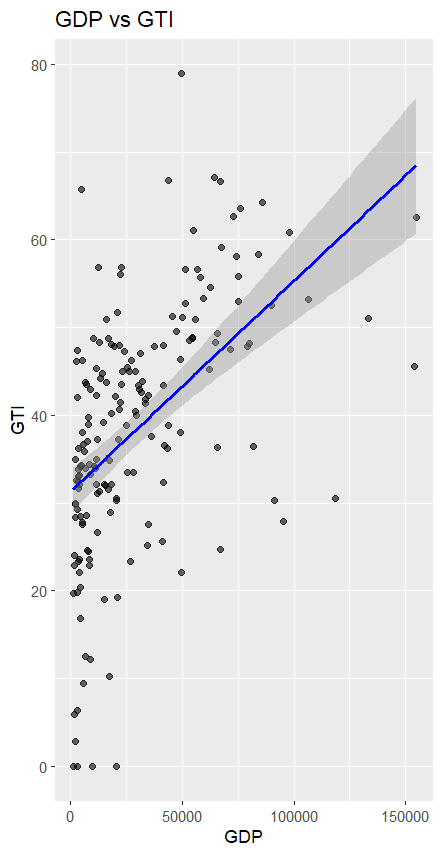
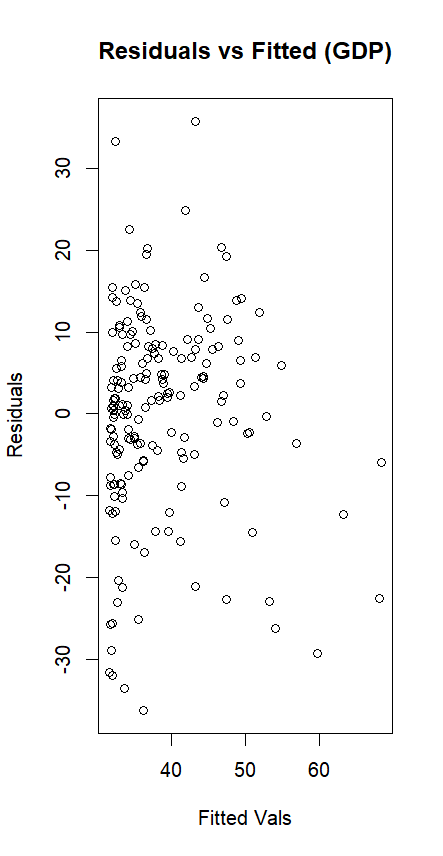
Residual standard error: 14.47 on 177 degrees of freedom

(1 observation deleted due to missingness)

Multiple R-squared: 0.00215, Adjusted R-squared: -0.003488

F-statistic: 0.3814 on 1 and 177 DF, p-value: 0.5377

GDP is the better predictor



2.2:

**GTI vs GDP (Log)**

Call:

lm(formula = GTI.new ~ log(gdp), data = Data)

Residuals:

Min 1Q Median 3Q Max

-39.846 -6.131 1.394 6.678 36.333

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -30.7403 7.3042 -4.209 4.08e-05 \*\*\*

log(gdp) 7.1093 0.7442 9.553 < 2e-16 \*\*\*

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.71 on 177 degrees of freedom

(1 observation deleted due to missingness)

Multiple R-squared: 0.3402, Adjusted R-squared: 0.3365

F-statistic: 91.27 on 1 and 177 DF, p-value: < 2.2e-16

**GTI vs Population (Log)**

Call:

lm(formula = GTI.new ~ log(population), data = Data)

Residuals:

Min 1Q Median 3Q Max

-39.171 -8.186 0.890 8.878 39.187

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 48.0307 8.7514 5.488 1.39e-07 \*\*\*

log(population) -0.5887 0.5460 -1.078 0.282

---

Signif. codes:

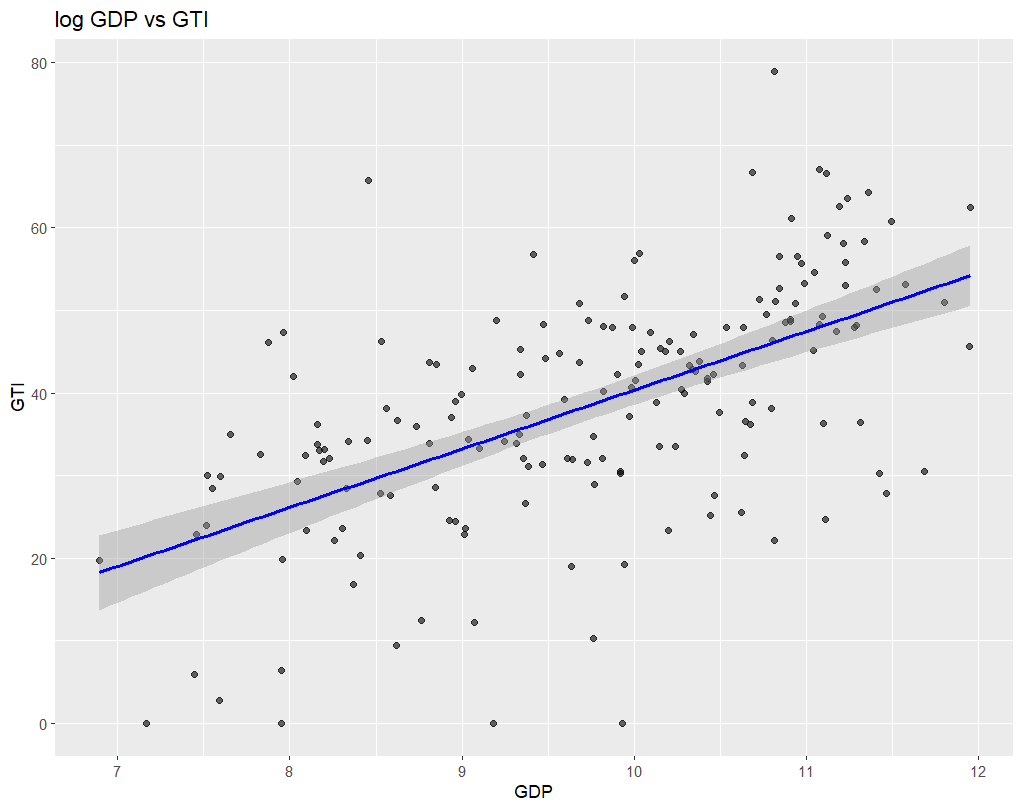
0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

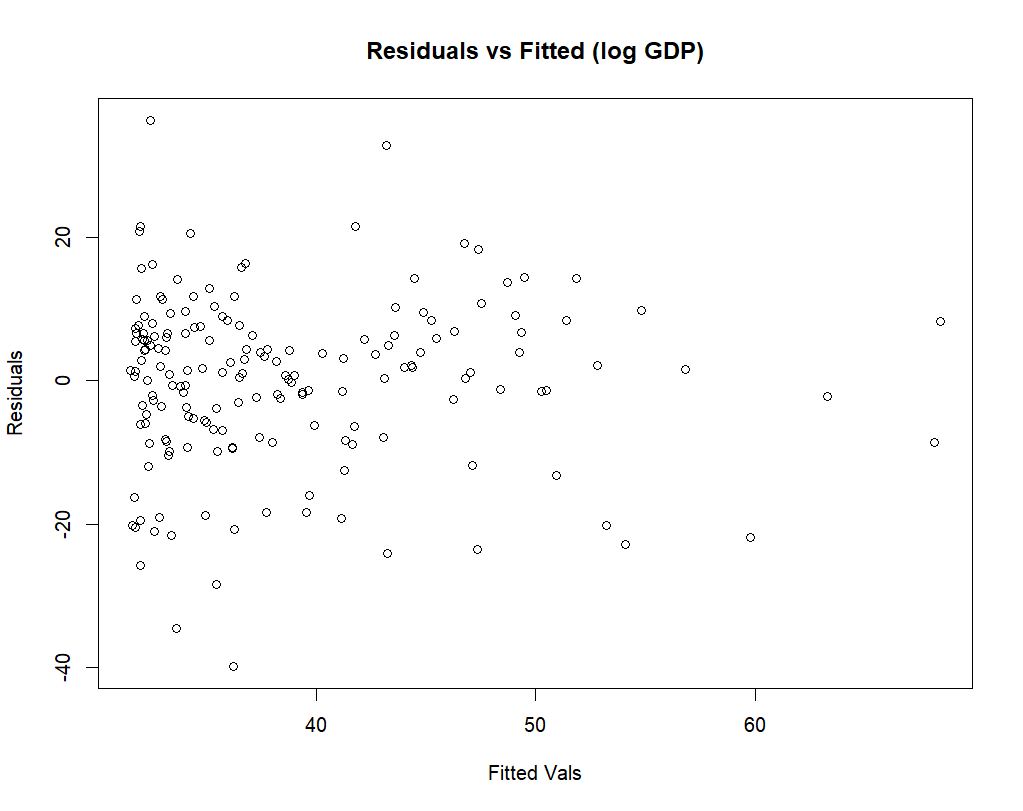
Residual standard error: 14.44 on 177 degrees of freedom

(1 observation deleted due to missingness)

Multiple R-squared: 0.006526, Adjusted R-squared: 0.0009136

F-statistic: 1.163 on 1 and 177 DF, p-value: 0.2824





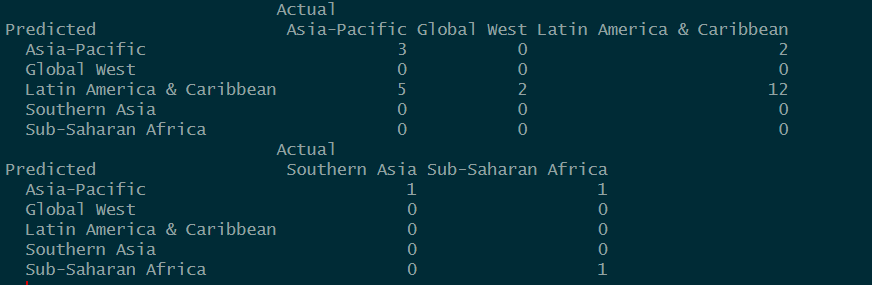
The log transformation is better than the original model. This is because p remained highly significant while decreasing residual standard error.

3)

3.1:

k=5

Confusion Matrix:



Accuracy:



3.2:

Confusion Matrix:



Accuracy:



Model 2 is a better model as it achieved ~70% accuracy where as the first reached ~59%. Model 2 also was more correct across more regions.