# Multiple Regression Model to select a Best Destination

# Problem Statement

- To Analyze Best tourist Destination based on the data acquired from an online tourism website which collects the ratings from users.

- The user will be able to pick the best Destination based on the nine independent variables.
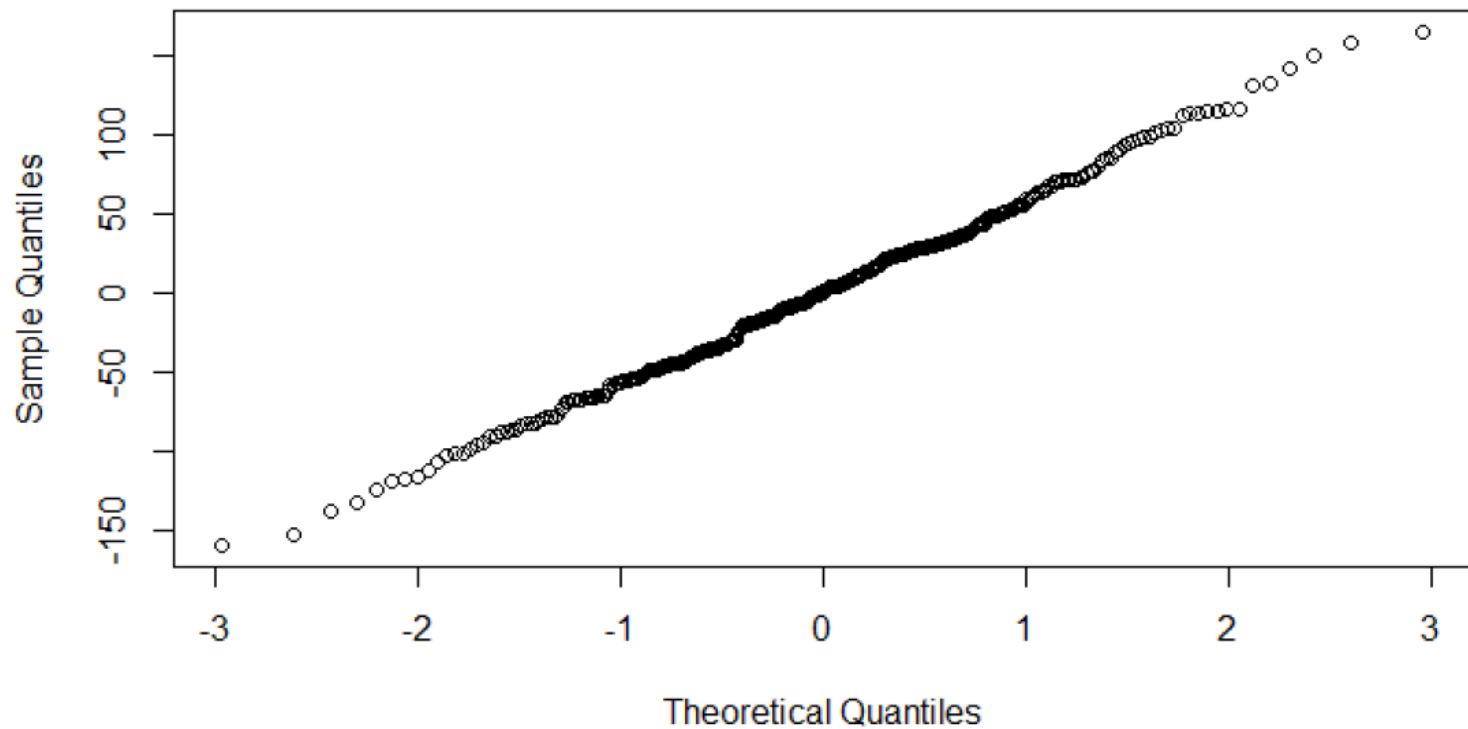
- Climate and Terrain

- Housing

- Health Care and Environment

- Crime

- Transportation

- Education

- Arts

- Recreation

- Economy

# Methodology

- There are more than one predictors.

- Multiple regression model will be used.

- Before that backward stepwise model will be used to select a best fit model.

# Normality Assumption
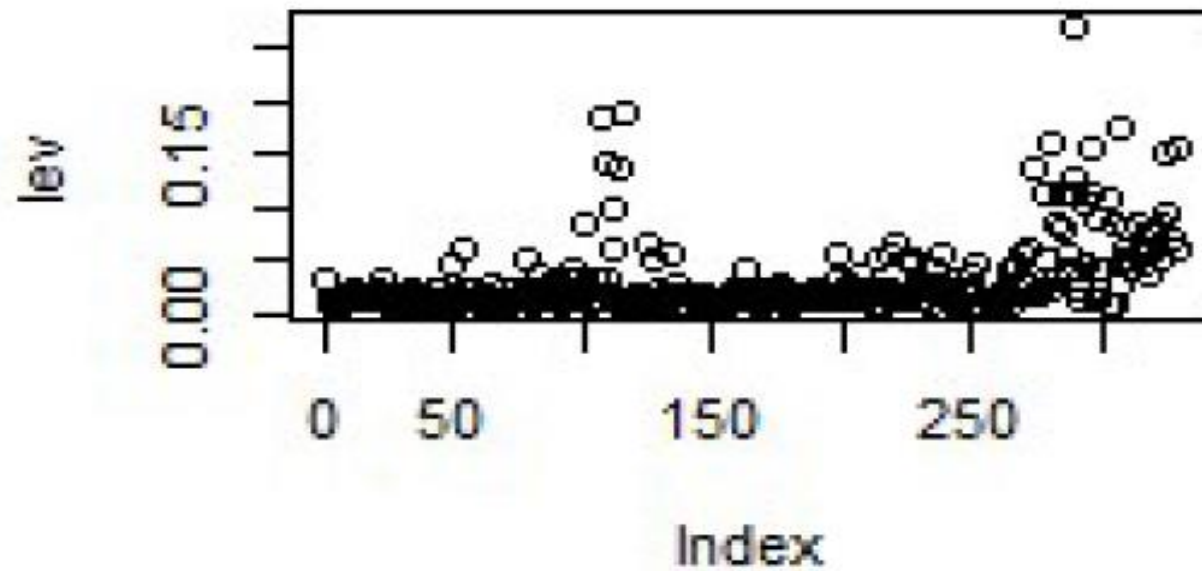


Normal Q-Q Plot

# Outliers Test



Figure 5: Leverage plot

```
> summary(fit)

Call: lm(formula = best_destination ~ Climate + Housing + HealthCare +Crime + Transportation + Education + Arts + Recreation +economy, data = tourism)

Min 1Q Median 3Q Max

-159.50 -42.86 0.31 34.80 164.45

Coefficients:

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

(Intercept) -50.062015 13.823384 -3.622 0.000341 ***

Climate 0.007481 0.001609 4.649 4.89e-06 ***

Housing 0.003007 0.001151 2.612 0.009427 **

HealthCare 0.011871 0.001206 9.846 < 2e-16 ***

Crime 0.001125 0.003250 0.346 0.729547

Transportation 0.011002 0.001873 5.875 1.06e-08 ***

Education 0.006675 0.001965 3.396 0.000769 ***

Arts 0.006951 0.001009 6.891 2.96e-11 ***

Recreation 0.001860 0.001941 0.958 0.338621

economy 0.007159 0.001126 6.355 7.18e-10 ***

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

Residual standard error: 59.09 on 319 degrees of freedom

Multiple R-squared: 0.6246,Adjusted R-squared: 0.614

F-statistic: 58.98 on 9 and 319 DF, p-value: < 2.2e-16
```

# Backward Stepwise Model

Start: AIC=2693.91

best_destination ~ Climate + Housing + HealthCare + Crime + Transportation +

Education + Arts + Recreation + economy

Df Sum of Sq RSS AIC

- Crime 1 418 1114406 2692.0

- Recreation 1 3207 1117195 2692.9

\<none\> 1113987 2693.9

- Housing 1 23826 1137813 2698.9

- Education 1 40282 1154269 2703.6

- Climate 1 75483 1189470 2713.5

- Transportation 1 120537 1234525 2725.7

- economy 1 141043 1255030 2731.1

- Arts 1 165845 1279833 2737.6

 HealthCare 1 338561 1452549 2779.2

Step: AIC=2692.04

best_destination ~ Climate + Housing + HealthCare + Transportation +

Education + Arts + Recreation + economy

Df Sum of Sq RSS AIC

- Recreation 1 3234 1117640 2691.0

<none> 1114406 2692.0

- Housing 1 23708 1138113 2697.0

- Education 1 40149 1154555 2701.7

- Climate 1 75268 1189674 2711.5

- Transportation 1 123271 1237676 2724.6

- economy 1 141240 1255646 2729.3

- Arts 1 165557 1279963 2735.6

- HealthCare 1 340438 1454844 2777.7

Step: AIC=2690.99

best_destination ~ Climate + Housing + HealthCare + Transportation + Education + Arts + economy

Df Sum of Sq RSS AIC

<none> 1117640 2691.0

- Housing 1 24344 1141983 2696.1

- Education 1 43136 1160776 2701.4

- Climate 1 73928 1191568 2710.1

- Transportation 1 123159 1240799 2723.4

- economy 1 155594 1273234 2731.9

- Arts 1 163873 1281513 2734.0

- HealthCare 1 351404 1469044 2778.9

Call:

lm(formula = best_destination ~ Climate + Housing + HealthCare +

Transportation + Education + Arts + economy, data = tourism)

Coefficients:

(Intercept) Climate Housing HealthCare Transportation Education

-47.344681 0.007393 0.003037 0.011944 0.011063 0.006867

Arts economy

0.006902 0.007375

**Least square regression Line**

y= -47.344681 + 0.007393 (Climate)+ 0.003037 (Housing )+ 0.011944 (HealthCare )+ 0.011063 (Transportation) + 0.006867 (Education) + 0.006902( Arts) + 0.007375 (economy).

The reduced model has been predicted and sorted to obtain the best destinations. The top 5 Best Destinations.

s-no Destination

329 - Yuba City, CA

322 - Williamsport, PA

314 - Washington, DC-MD-VA

319 - Wheeling, WV-OH

293 - Stockton, CA

```
predicted<-predict(m2, data=tourism)
sort(predicted)
```