EDA FA4 BUENAFE

Karl Isaiah Buenafe

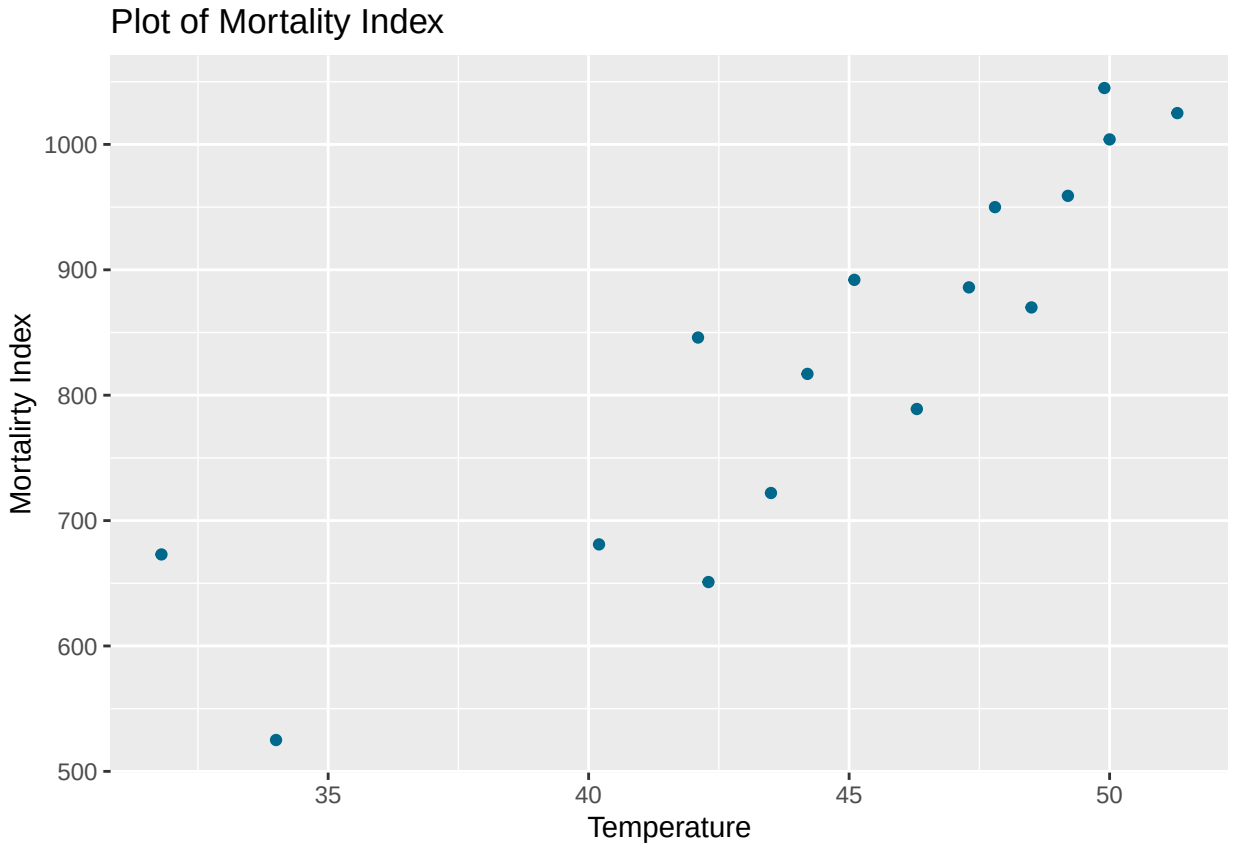
2025-03-02

MORTALITY BY LATITUDE DATASET

```
##    latitude mortality_index temperature
## 1        50            1025        51.3
## 2        51            1045        49.9
## 3        52            1004        50.0
## 4        53             959        49.2
## 5        54             870        48.5
## 6        55             950        47.8
## 7        56             886        47.3
## 8        57             892        45.1
## 9        58             789        46.3
## 10       59             846        42.1
```

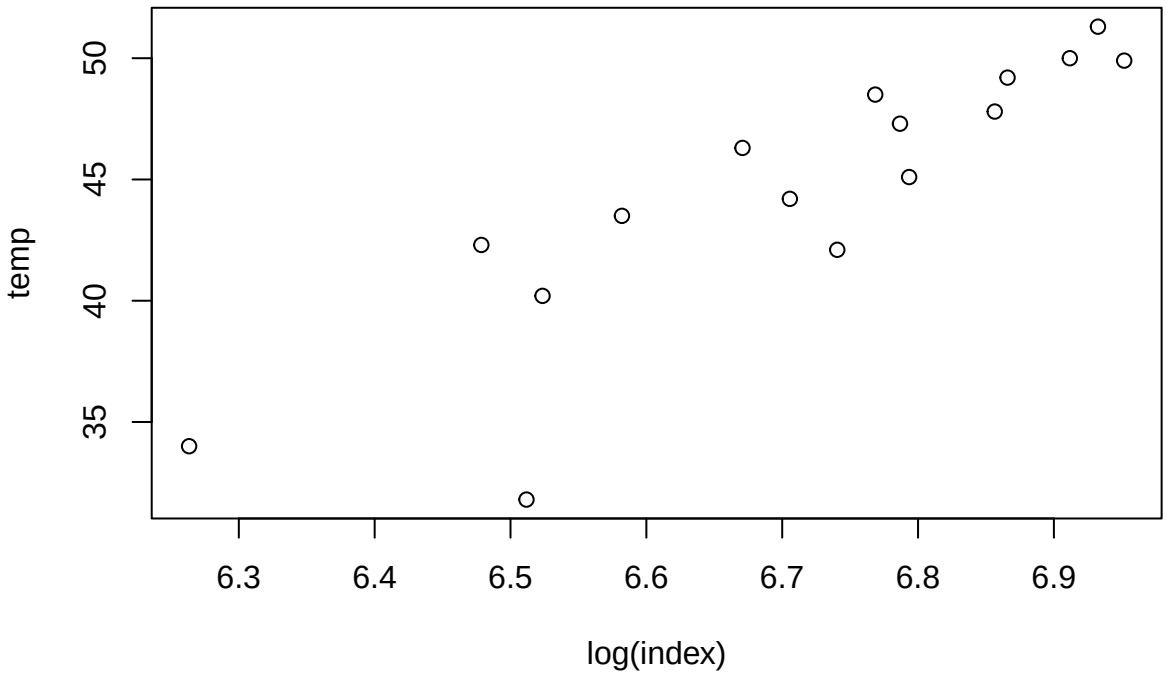
Make a plot of mortality index against mean average temperature

```
## Warning in geom_point(color = "deepskyblue4", line = 1): Ignoring unknown
## parameters: ‘line‘
```



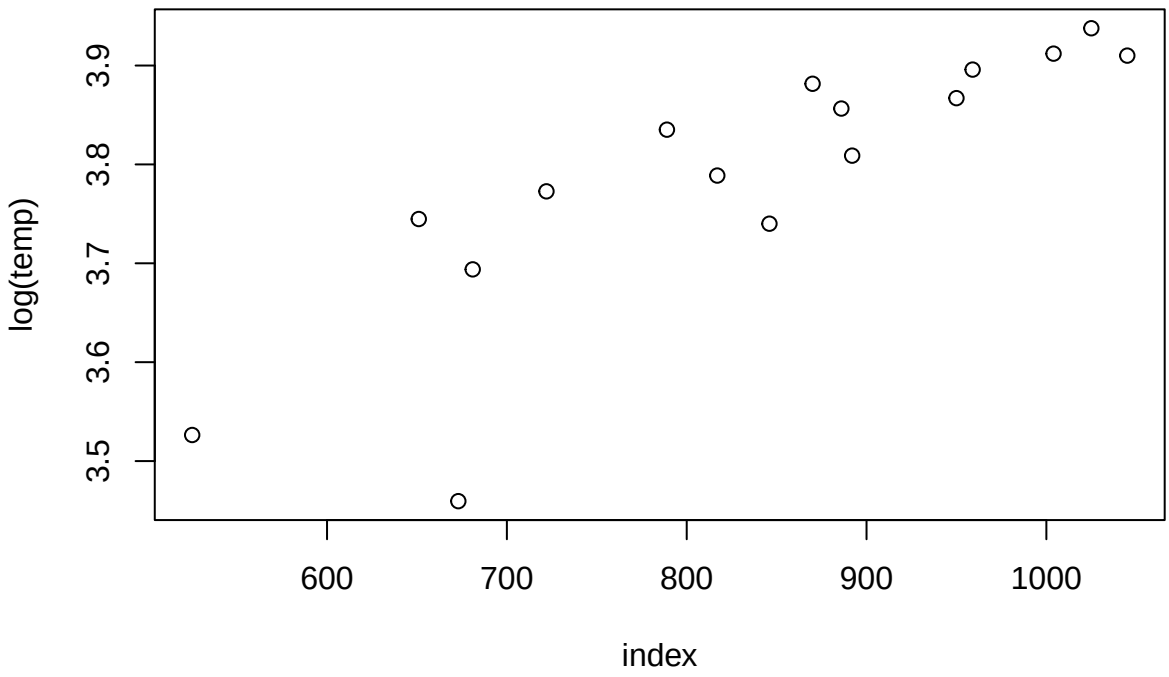
Using Log Transformation to straighten the plot:

```
plot(log(index), temp)
```



log(index)
log(temp)

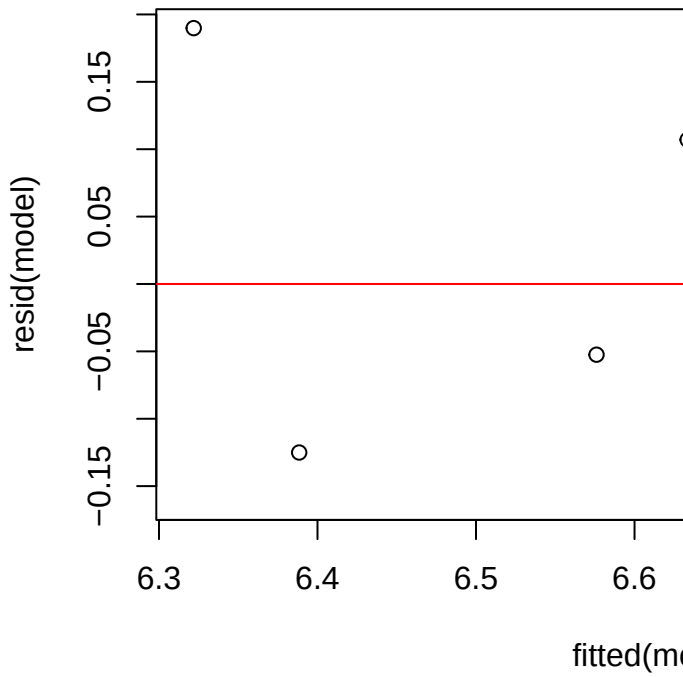
```
plot(index, log(temp) )
```



Considering log(index) as the best transformation:

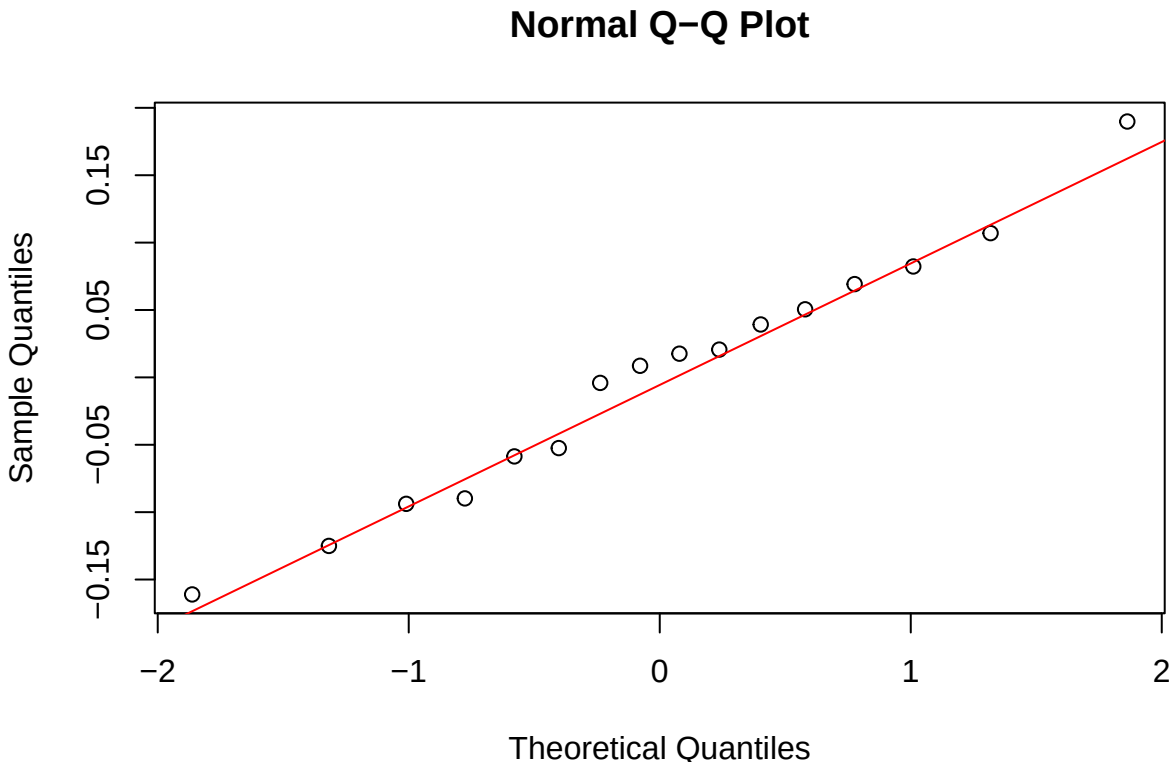
```
model <- lm(log(index) ~ temp, data = df)
```

```
plot(fitted(model), resid(model))
abline(h=0, col="red")
```



Fitting log(index) and plotting the residuals respectively

```
qqnorm(resid(model))
qqline(resid(model), col="red")
```



DIAMOND DATASET

```
head(df2,8)
```

```
## # A tibble: 8 x 10
##   carat cut       color clarity depth table price     x     y     z
##   <dbl> <ord>     <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl>
## 1  0.23 Ideal     E     SI2      61.5    55   326  3.95  3.98  2.43
## 2  0.21 Premium   E     SI1      59.8    61   326  3.89  3.84  2.31
## 3  0.23 Good      E     VS1      56.9    65   327  4.05  4.07  2.31
## 4  0.29 Premium   I     VS2      62.4    58   334  4.2   4.23  2.63
## 5  0.31 Good      J     SI2      63.3    58   335  4.34  4.35  2.75
## 6  0.24 Very Good J     VVS2     62.8    57   336  3.94  3.96  2.48
## 7  0.24 Very Good I     VVS1     62.3    57   336  3.95  3.98  2.47
## 8  0.26 Very Good H     SI1      61.9    55   337  4.07  4.11  2.53
```

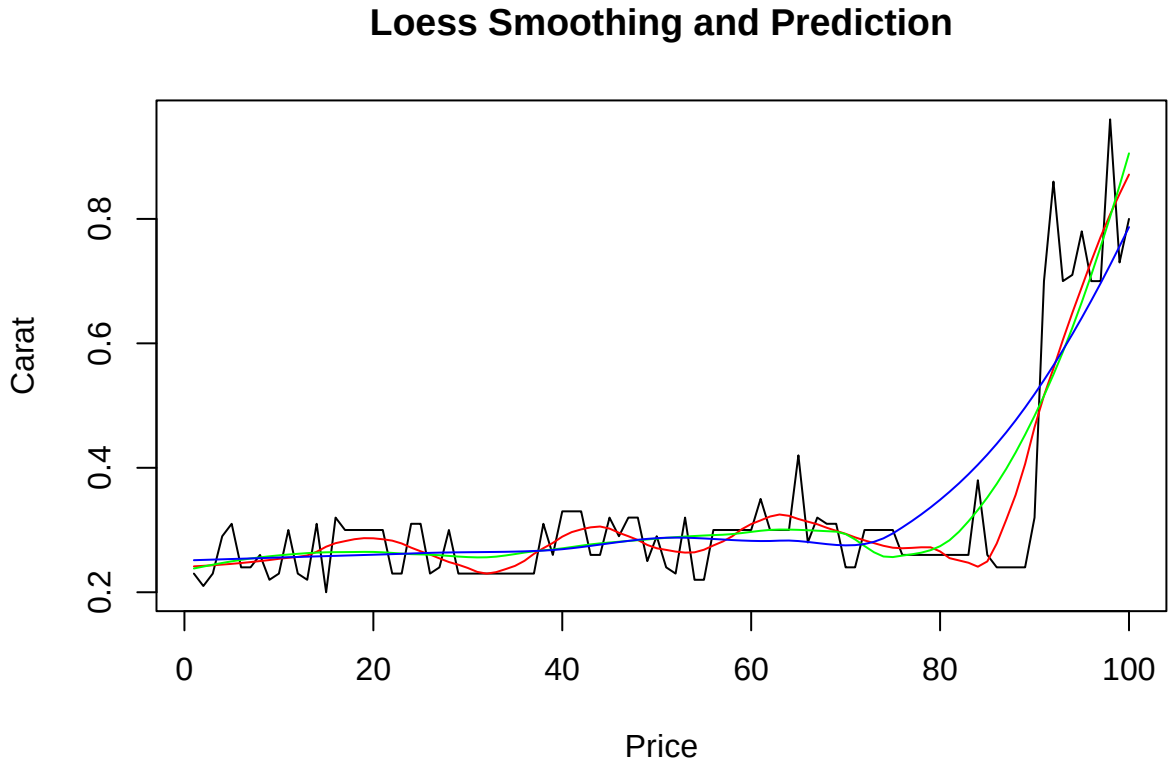
LOESS SMOOTHER

```
df2$price <- 1:nrow(df2)
df2 <- df2[1:100, ]
loessMod25 <- loess(carat ~ price, data=df2, span=0.25)
loessMod50 <- loess(carat ~ price, data=df2, span=0.50)
loessMod75 <- loess(carat ~ price, data=df2, span=0.75)
```

span values: 0.25, 0.50, and 0.75

```
smoothed25 <- predict(loessMod25)
smoothed50 <- predict(loessMod50)
smoothed75 <- predict(loessMod75)

plot(df2$carat, x=df2$price, type="l", main="Loess Smoothing and Prediction", xlab="Price", ylab="Carat
lines(smoothed25, x=df2$price, col="red")
lines(smoothed50, x=df2$price, col="green")
lines(smoothed75, x=df2$price, col="blue")
```


Loess Smoothing and Prediction
Carat
0.2
0.4
0.6
0.8
0
20
40
60
80
100
Price

Plotting