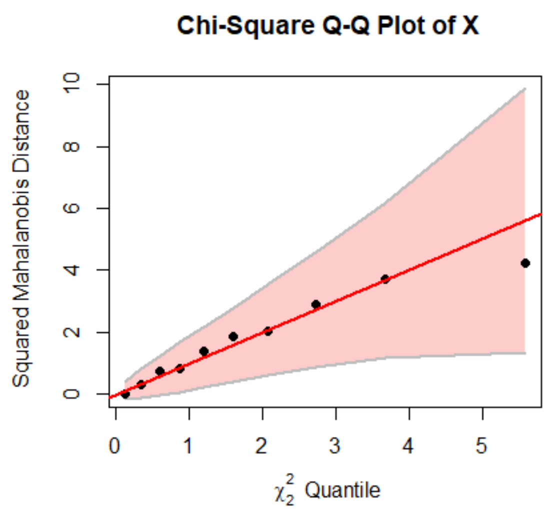
Problem #2 (4.26)

```
> x1 = c(1:3,3:6,8,9,11)
> x2 = c(18.95,19,17.95,15.54,14,12.95,8.94,7.49,6,3.99)
> X = cbind(matrix(c(x1)),matrix(c(x2)))
> mean = rbind(mean(x1),mean(x2))
> diff1 = x1-mean[1]
> diff2 = x2-mean[2]
> Diff = t(rbind(diff1,diff2))
> sum = 0
> S = matrix(c(var(x1),cov(x1,x2),cov(x1,x2),var(x2)),2,2)
> SInverse = solve(S)
> contour(X)
> chi = qchisq(0.5,2)
>
> distance = double()
> falling.contour = logical()
> for(i in c(1:10)){
+   temp = Diff[i,]
+   result = t(temp)%*%SInverse%*%temp
+   if(result>chi) {
+     text = 'fall'
+   }else{
+     text = 'not fall'
+   }
+   cat("i=", i, ':\t', format.default(result,digits = 4),'
\t',text,'\n', sep = '')
+   distance = rbind(distance,result)
+   falling.contour = rbind(falling.contour,result>chi)
+ }
i=1:        1.875       fall
i=2:        2.02        fall
i=3:        2.901       fall
i=4:        0.7353      not fall
i=5:        0.3105      not fall
i=6:        0.01762     not fall
i=7:        3.733       fall
i=8:        0.8165      not fall
i=9:        1.375       not fall
i=10:       4.215       fall
> print(sort(distance))
 [1] 0.0176162 0.3105192 0.7352659 0.8165401 1.3753379 1.87
53045 2.0203262 2.9009088 3.7329012 4.2152799
> library(heplots)
> cqplot(X)
```



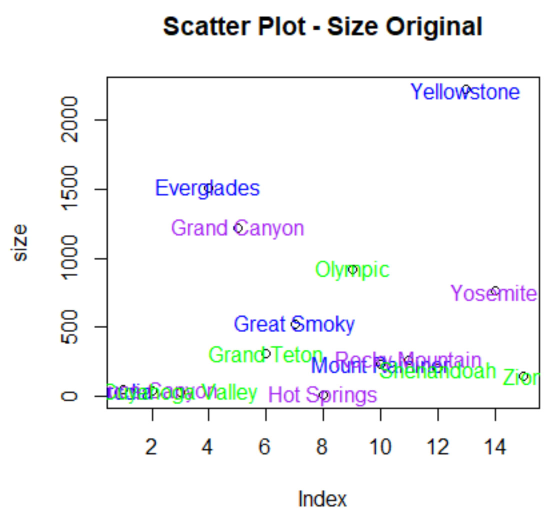
Problem #3 (4.40)

```
> park = c('Arcdia','Bruce Canyon', 'Cuyahoga Valle
y','Everglades', 'Grand Canyon', 'Grand Teton', 'Gr
eat Smoky', 'Hot Springs',
+          'Olympic', 'Mount Raininer', 'Rocky Moun
tain', 'Shenandoah', 'Yellowstone', 'Yosemite', 'Zi
on')
> size = c(47.4,35.8,32.9,1508.5,1217.4,310.0,521.
8,5.6,922.7,235.6,265.8,199.0,2219.8,761.3,146.6)
> visitors = c(2.05,1.02,2.53,1.23,4.40,2.46,9.19,
1.34,3.14,1.17,2.80,1.09,2.84,3.30,2.59)
```

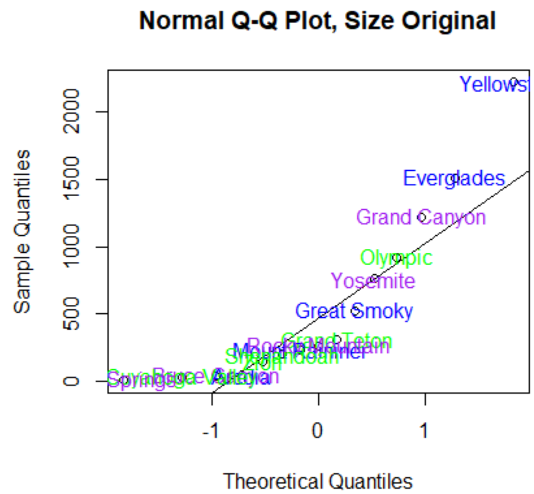
A)

Let's first observe the scatterplot of *size* (x_1): We can see that Yellowstone, Everglades, and Grand Canyon maybe potential outliers. This can be proved by the Normal Q-Q Plot.

```
> plot(size, main='Scatter Plot - Size Original')
> color = c('blue', 'purple', 'green')
> text(size, park, col = color)
```

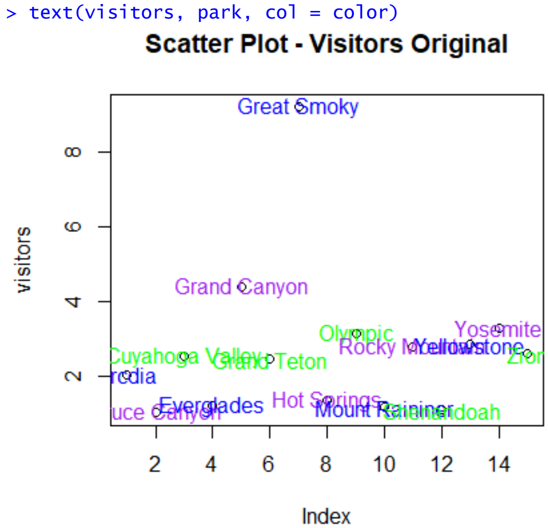



```
 > plot1 = qqnorm(size, main = 'Normal Q-Q Plot, Si
ze Original')
> qqline(size)
> text(plot1$x,plot1$y,park,col=color)
```

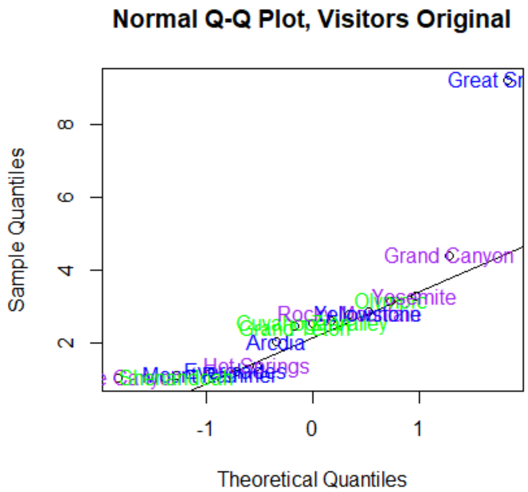


Then, observe the scatterplot of *visitors* (x_2): We can see that Grand Smoky maybe a potential outlier. This can also be seen in the normal Q-Q Plot

```
> plot(visitors, main='Scatter Plot - Visitors Orig
inal')
> text(visitors, park, col = color)
```

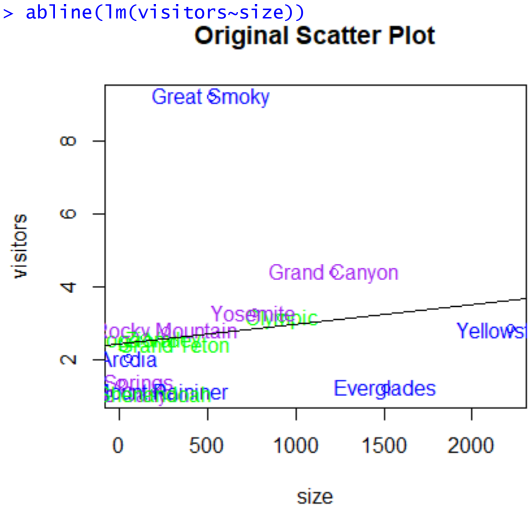



```
> plot1=qqnorm(visitors, main = 'Normal Q-Q Plot, V
isitors Original')
> qqline(visitors)
> text(plot1$x,plot1$y,park,col=color)
```



Plot *visitors* - *size*, we can see that Grand Canyon, Great Smoky, Yellowstone, and Everglades can also be outliers.

```
> plot(size, visitors,type='p', main = 'Original Sc
atter Plot', col= c('blue', 'purple', 'green'))
> text(size, visitors, park, col = color)
> abline(lm(visitors~size))
```

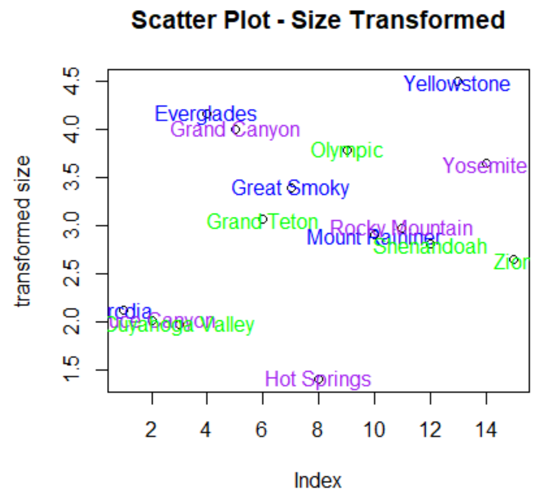


B)

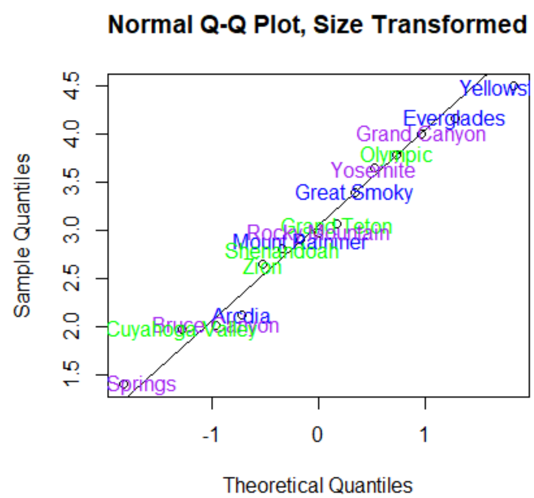
Using Box-cox to transform *size* (x_1). Lambda is 0.194947. Examine the results using a scatter plot and a normal Q-Q Plot, we can see remarkable improvements. The data is almost normal now

```
> library(car)
> trans=powerTransform(size)
> summary(powerTransform(size))
bcPower Transformation to Normality

     Est Power Rounded Pwr Wald Lwr bnd Wald Upr Bnd
size    0.1949           0       -0.129       0.5189

Likelihood ratio tests about transformation parameters
                          LRT df         pval
LR test, lambda = (0)  1.485416  1 2.229290e-01
LR test, lambda = (1) 17.221093  1 3.327216e-05
```

```
> plot(size**trans$lambda, ylab = 'transformed size
')
> text(size**trans$lambda, park, col = color)
```




```
> plot1=qqnorm(size**trans$lambda, main = 'Normal Q
-Q Plot, Size Transformed')
> qqline(size**trans$lambda)
> text(plot1$x,plot1$y,park,col=color)
```

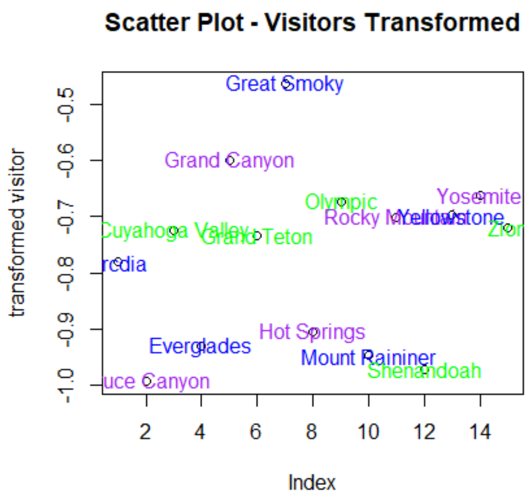


C)

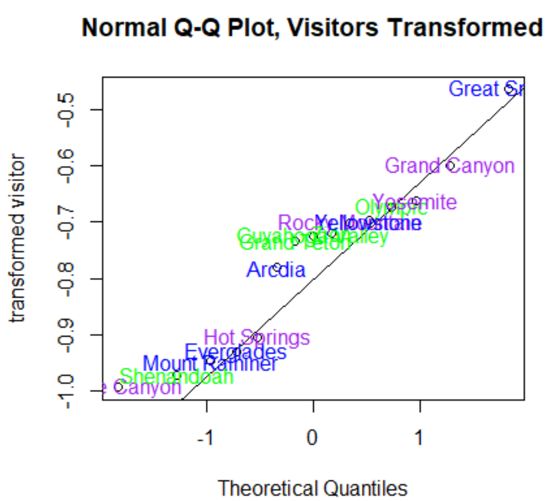
Using Box-cox to transform *visitors* (x_2). Lambda is -0.3456574. Examine the results using a scatter plot and a normal Q-Q Plot, we can see remarkable improvements. The data is almost normal now.

```
> trans = powerTransform(visitors)
> summary(powerTransform(visitors))
bcPower Transformation to Normality

         Est Power Rounded Pwr Wald Lwr bnd Wald Upr Bnd
visitors   -0.3457           0       -1.165       0.4737

Likelihood ratio tests about transformation parameters
                          LRT df         pval
```

```
LR test, lambda = (0)  0.7365957  1 0.3907535765
LR test, lambda = (1) 12.5728818  1 0.0003913846
> plot(-visitors**trans$lambda, ylab = 'transformed
 visitor', main = 'Scatter Plot - Visitors Transfor
med')
> text(-visitors**trans$lambda, park, col = color)
```




```
> plot1=qqnorm(-visitors**trans$lambda, ylab = 'tra
nsformed visitor',main = 'Normal Q-Q Plot, Visitors
 Transformed')
> qqline(-visitors**trans$lambda)
> text(plot1$x,plot1$y,park,col=color)
```

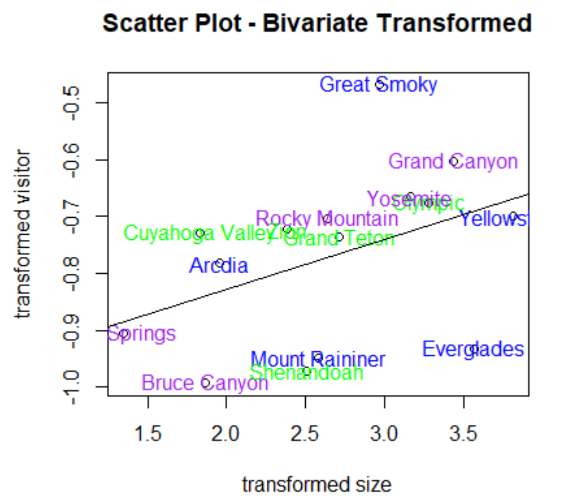


D)

To achieve bivariate normality, we have the followings. The scatter plot shows some improvement, but it seems Great Smoky may still be an outlier. But the Bivariate Normal Q-Q Plot shows the data looks fine.

```
> summary(powerTransform(cbind(size, visitors)))
bcPower Transformations to Multinormality
         Est Power Rounded Pwr Wald Lwr bnd Wald Upr Bnd
size        0.1738           0      -0.1355       0.4830
visitors   -0.3410           0      -1.1198       0.4378
```

```
Likelihood ratio tests about transformation parameters
                              LRT df         pval
LR test, lambda = (0 0)  2.068243  2 3.555386e-01
LR test, lambda = (1 1) 32.167033  2 1.035184e-07
> trans.visitor=-visitors**trans$lambda[2]
> trans.size = (size**trans$lambda[1])
> plot(trans.size,trans.visitor,xlab = 'transformed
 size', ylab = 'transformed visitor', main = 'Scatt
er Plot - Bivariate Transformed')
> text(trans.size,trans.visitor, park, col = color)
> abline(lm(trans.visitor~trans.size))
```




```
> plot1 = qqplot(trans.size,trans.visitor)
> text(plot1$x,plot1$y,park,col=color)
```