Khan Inan
assignment 9

1. Using a poisson distribution, I would represent the data as show below

| NumberOfDeaths	Number of corps with x deaths in a given year	Poisson prediction
0	144	139
1	91	97
2	32	34
3	11	8
4	2	1

```
table2 = cbind(table1, predicted)
colnames(table2)[3]="Poisson prediction"
table2 %>%
  kable()   %>%
  column_spec(column = 1:3, width = "10em")   %>%
  kable_styling(c("striped", "bordered"))
```

The best way to plot this particular horse kick data is using a bar diagram in my opinion, since we are trying to compared the actual data to the poison data

Chi-Squared
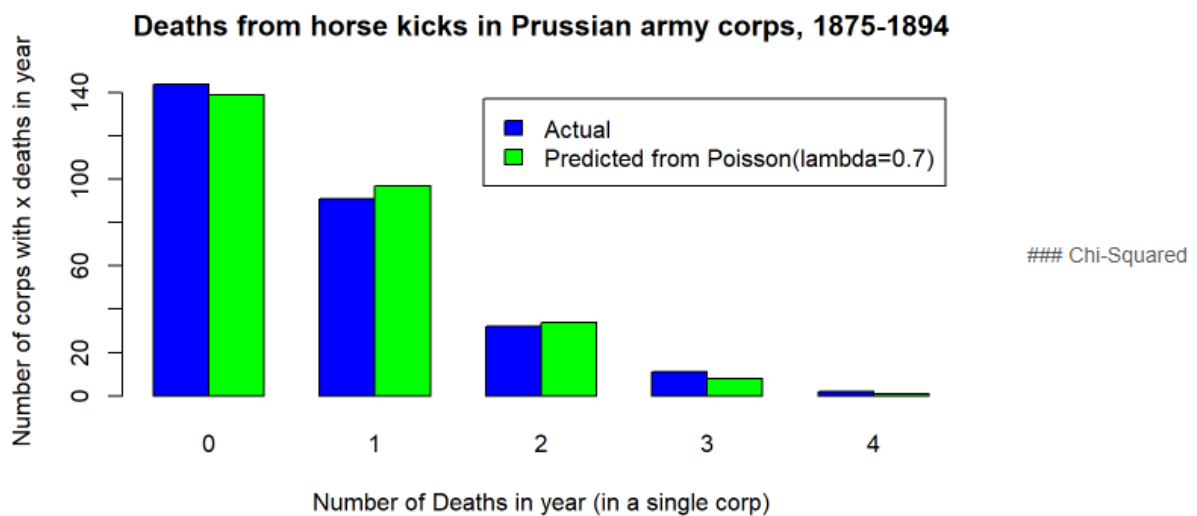
```
barplot(t(table2[,2:3]),
        beside=T, col=c("blue","green"), names.arg=NumberOfDeaths,
        xlab="Number of Deaths in year (in a single corp)",
        ylab="Number of corps with x deaths in year",
        main = "Deaths from horse kicks in Prussian army corps, 1875-1894",
        legend.text = c("Actual","Predicted from Poisson(lambda=0.7)"))
```

2.a for each column of the mrnaseq_count_data the boxcox transformation that can make them gaussian is
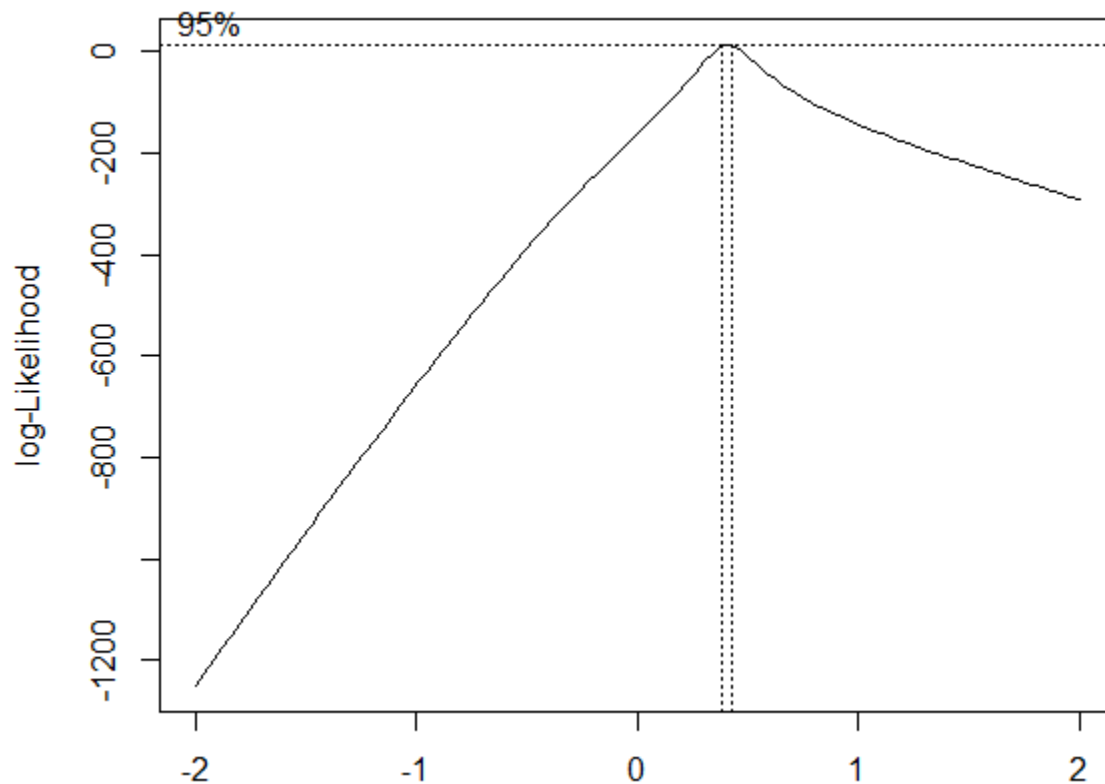
There does not seem to be a universal transform in this data due to there being an apex and a minimum in the transformation as well.

Looking at this graph I dont think a log transform would be good enough because each column of the data has a different value that would make it gassian