```r
# STAT 3093 Assignment 1
# Question 3
# Albert Lockett (3254354, k44if@unb.ca)

q3_data <- c(
  0.736, 0.863, 0.865, 0.913, 0.915, 0.937, 0.983, 1.007,
  1.011, 1.064, 1.109, 1.132, 1.140, 1.153, 1.253, 1.394
)

n <- length(q3_data)

# ~~~ PART A ~~~~
# (output is below)

# calculate the sample mean:
```

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

```r
sample_mean <- sum(q3_data) / n
print(sprintf('The sample mean is: %.3f',sample_mean))

# calculate the sample median:
# since n is even, it is the average of the two middle values
sample_median <- (q3_data[n/2] + q3_data[n/2 + 1]) / 2
print(sprintf('The sample median is: %.3f',sample_median))

# calculate sample standard deviation
```

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

```r
squared_deviations <- unlist(
  purrr::map(
    q3_data,
    function(x) (x - sample_mean)^2
  )
)
sample_variance <- sum(squared_deviations) / (n - 1)
```

$$S = \sqrt{S^2}$$

```r
sample_stdev <- sqrt(sample_variance)
print(sprintf('The sample standard deviation is: %.3f', sample_stdev))

# calculate the sample range:
# the difference of the largest and smallest numbers
sample_range <- max(q3_data) - min(q3_data)
print(sprintf('The sample range is: %.3f', sample_range))

# calculate the five number summary:
smallest_xi <- min(q3_data)
lower_fourth <- median(q3_data[1:(n/2)])
upper_fourth <- median(q3_data[(n/2 + 1):n])
largest_xi <- max(q3_data)
```

```r


print('The five number summary is ... ')
print(sprintf(' - smallest xi   %.3f', smallest_xi))
print(sprintf(' - lower fourth  %.3f', lower_fourth))
print(sprintf(' - sample median %.3f', sample_median))
print(sprintf(' - upper fourth  %.3f', upper_fourth))
print(sprintf(' - largest xi    %.3f', largest_xi))

# calculate the inter quartile range:
inter_quartile_range <- upper_fourth - lower_fourth
print(sprintf('The inter quartile range is: %.3f', inter_quartile_range))


# OUTPUT:
# > source('~/School/STAT_3093/r/assignment_1/question3.R')
# [1] "The sample mean is: 1.030"
# [1] "The sample median is: 1.009"
# [1] "The sample standard deviation is: 0.165"
# [1] "The sample range is: 0.658"
# [1] "The five number summary is ... "
# [1] " - smallest xi   0.736"
# [1] " - lower fourth  0.914"
# [1] " - sample median 1.009"
# [1] " - upper fourth  1.132"
# [1] " - largest xi    1.394"
# [1] "The inter quartile range is: 0.218"


# ~~~ PART B ~~~~

# An observation is an outlier if it is more than 1.5 f_s (interquartile
# ranges) from the closest fourth.

upper_outlier_value <- upper_fourth + 1.5 * inter_quartile_range
lower_outlier_value <- lower_fourth - 1.5 * inter_quartile_range

upper_outliers <- q3_data[q3_data > upper_outlier_value]
lower_outliers <- q3_data[q3_data < lower_outlier_value]
num_outliers <- length(upper_outliers) + length(lower_outliers)

print(sprintf('upper outlier value: %.3f', upper_outlier_value))
print(sprintf('lower outlier value: %.3f', lower_outlier_value))
print(sprintf('there are %d outliers.', num_outliers))

# OUTPUT:
# [1] "upper outlier value: 1.469"
# [1] "lower outlier value: 0.581"
# [1] "there are 0 outliers."

# There are no outliers because no values in the sample are greater
# than 1.469 or less than 0.581
```

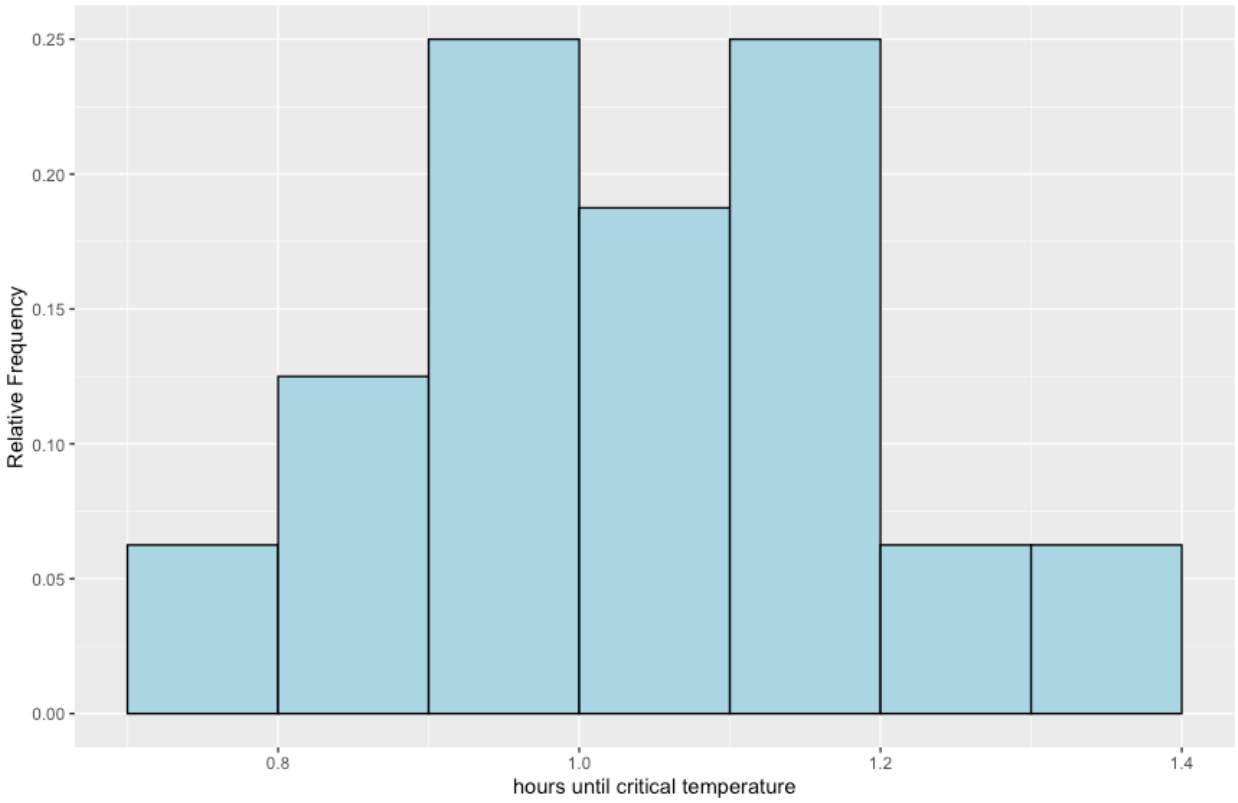
```r
# ~~~ PART C ~~~

histogram <- ggplot(data.frame(q3_data), aes(x=q3_data)) +
  geom_histogram(
    color='black',
    fill='lightblue',
    binwidth=0.1,
    center=1.05,
    aes(y = (..count..)/sum(..count..))
  ) +
  xlab('hours until critical temperature') +
  ylab('Relative Frequency')


boxplot(q3_data)
```

Question 3, part C output:

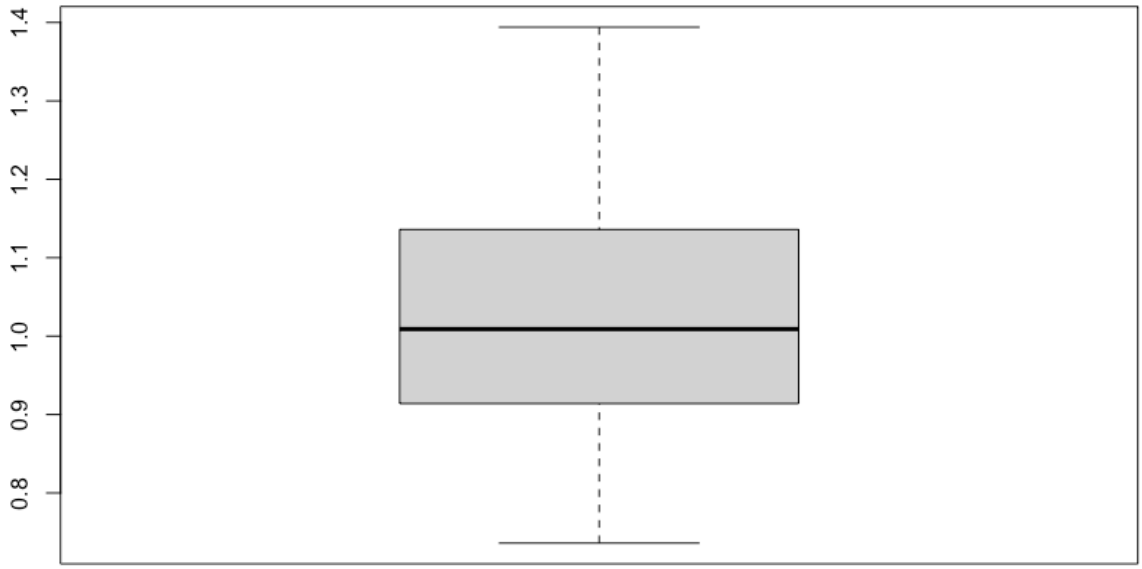
Histogram:



Stem and Leaf Diagram:

```
0.7 | 36
0.8 | 63 65
0.9 | 13 15 37 83
1.0 | 07 11 64
1.1 | 09 32 40 53
1.2 | 53
1.3 | 94
```

Box Plot



Question 3 - PART D

Based on the plots, I think the shape of the distribution looks symmetrical.