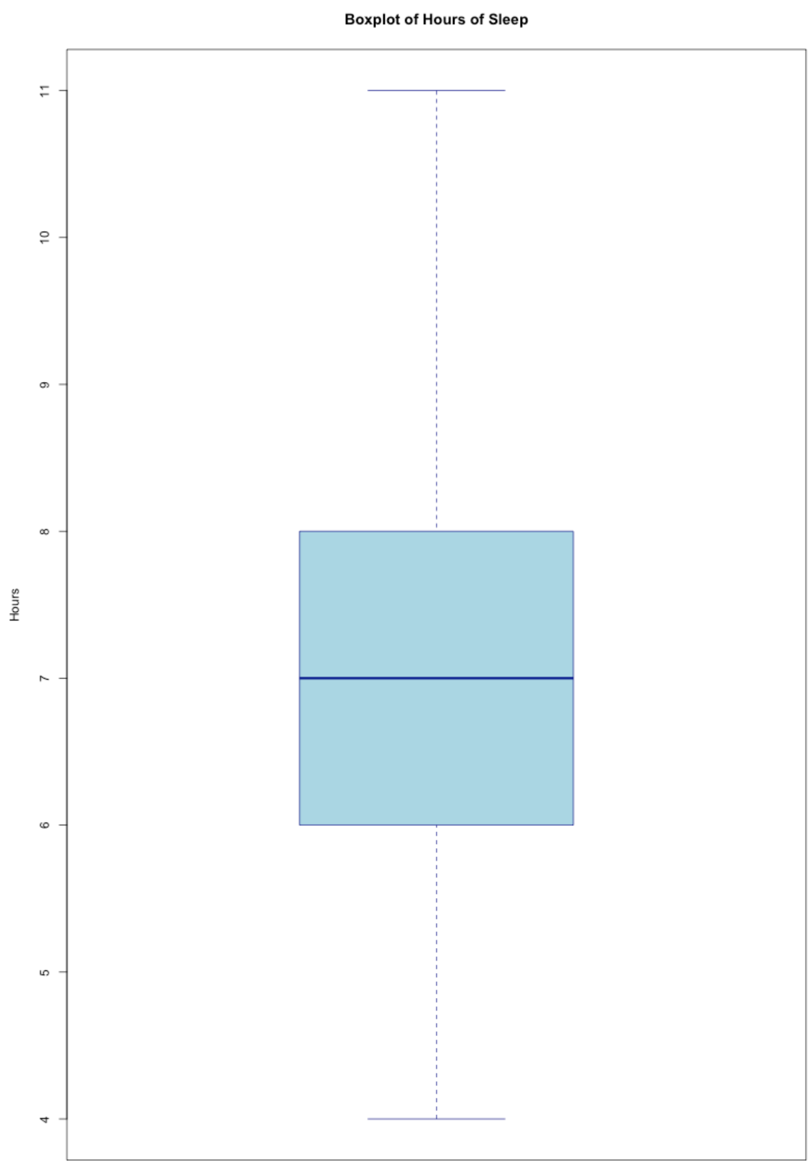
1. 15 randomly selected college students were asked to state the number of hours they slept last night. The resulting data are:

$$5,\ 6,\ 6,\ 8,\ 7,\ 7,\ 9,\ 5,\ 4,\ 8,\ 11,\ 6,\ 7,\ 8,\ 7$$

Create a boxplot in R using this dataset.

```
# Dataset: Hours of sleep for 15 college students
sleep_hours <- c(5, 6, 6, 8, 7, 7, 9, 5, 4, 8, 11, 6, 7, 8, 7)

# Create a boxplot
boxplot(sleep_hours,
        main = "Boxplot of Hours of Sleep",
        ylab = "Hours",
        col = "lightblue",
        border = "darkblue")
```



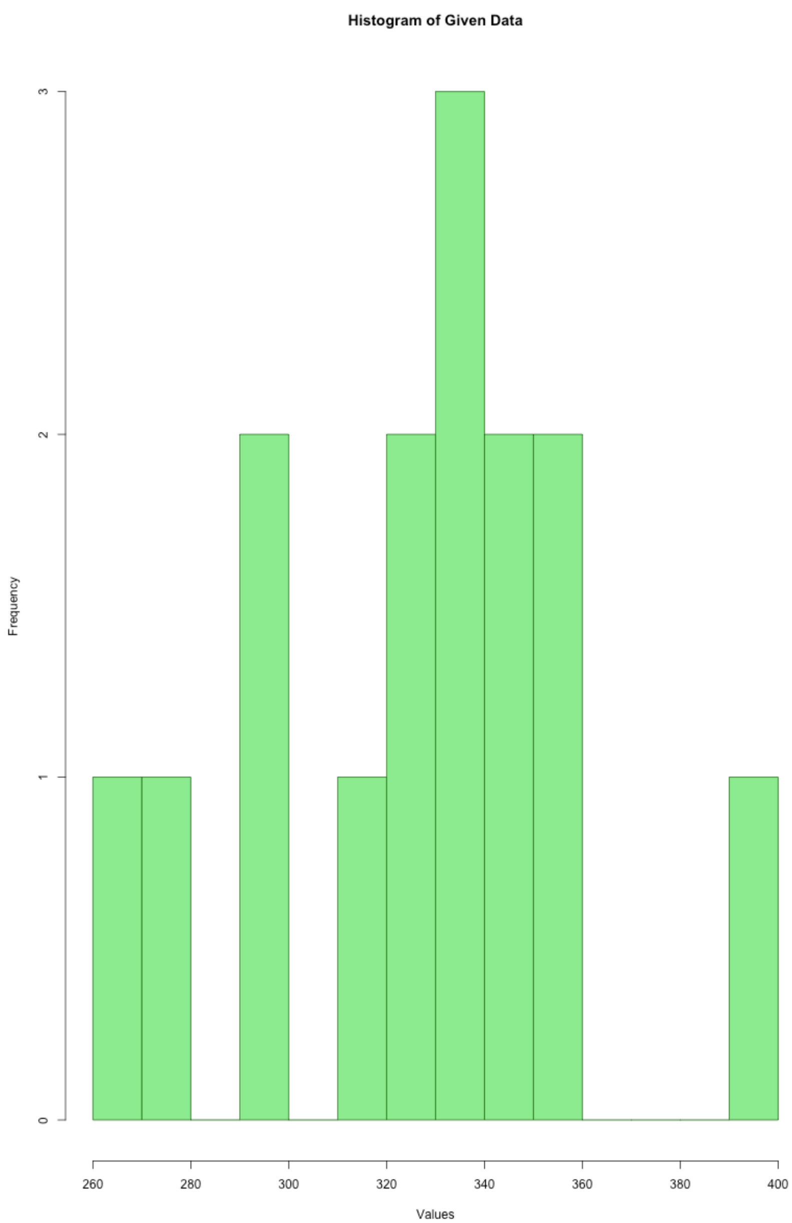
2. Find the 5 number summary, and construct a histogram in R.

$$275,\ 296,\ 299,\ 316,\ 322,\ 323,\ 332,\ 333,\ 337,\ 347,\ 350,\ 357,\ 358,\ 264,\ 393$$

```r
# Dataset: Given numbers
data <- c(275, 296, 299, 316, 322, 323, 332, 333, 337, 347, 350, 357, 358, 264, 393)

# Calculate the 5-number summary
summary_stats <- summary(data)
print(summary_stats)

# Create a histogram
hist(data,
    main = "Histogram of Given Data",
    xlab = "Values",
    col = "lightgreen",
    border = "darkgreen",
    breaks = 10)  # Adjust the number of bins as needed
```



3. How hot is it in Death Valley? The following are ground temperatures (in °F) taken from May to November. Find the mean, median, variance, and standard deviation of the data below using R.

146, 152, 168, 174, 180, 178, 179, 180, 178, 178, 168, 165, 152, 144

```
# Dataset: Ground temperatures in Death Valley (°F)
temperatures <- c(146, 152, 168, 174, 180, 178, 179, 180, 178, 178, 168, 165, 152, 144)

# Calculate the mean
mean_temp <- mean(temperatures)
print(paste("Mean:", mean_temp))

# Calculate the median
median_temp <- median(temperatures)
print(paste("Median:", median_temp))

# Calculate the variance
variance_temp <- var(temperatures)
print(paste("Variance:", variance_temp))

# Calculate the standard deviation
sd_temp <- sd(temperatures)
print(paste("Standard Deviation:", sd_temp))

> source("/Users/ran/cs/github/DSA/R/3.r", encoding = "UTF-8")
[1] "Mean: 167.285714285714"
[1] "Median: 171"
[1] "Variance: 178.373626373626"
[1] "Standard Deviation: 13.35565896441"
```

4. The students in a biology class kept a record of the height (in centimeters) of plants for a class experiment. Here are the heights:

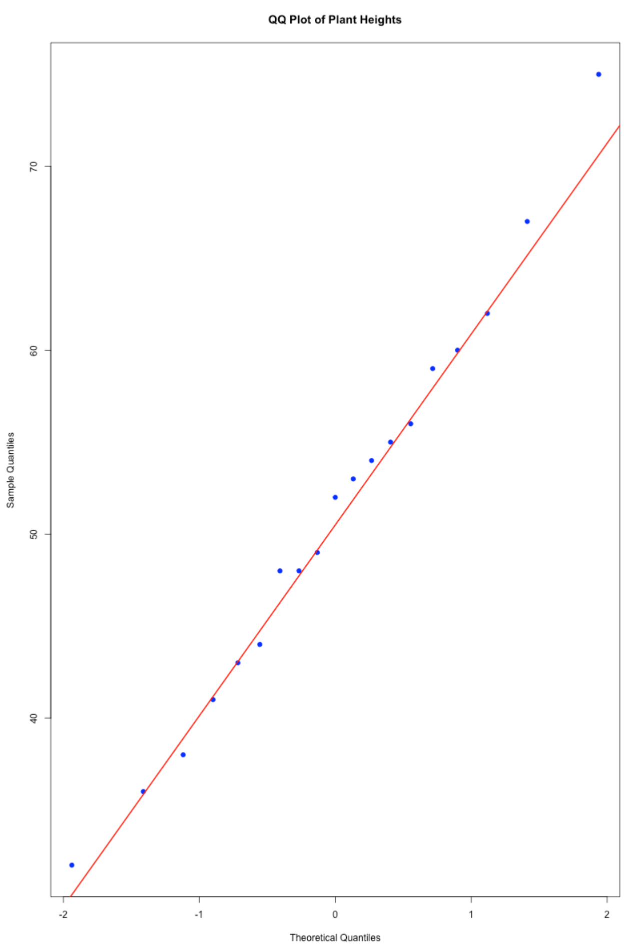
32, 36, 38, 41, 43, 44, 48, 48, 49, 52, 53, 54, 55, 56, 59, 60, 62, 67, 75

Based on this data, use R to construct a qqplot of the data. Do the data appear approximately normal?

```
# Dataset: Heights of plants (in cm)
plant_heights <- c(32, 36, 38, 41, 43, 44, 48, 48, 49, 52, 53, 54, 55, 56, 59, 60, 62, 67, 75)

# Create a QQ plot
qqnorm(plant_heights,
       main = "QQ Plot of Plant Heights",
       xlab = "Theoretical Quantiles",
       ylab = "Sample Quantiles",
       col = "blue",
       pch = 19)

# Add a reference line
qqline(plant_heights, col = "red", lwd = 2)
```



it is normal

5. A machine that cuts corks for wine bottles operates in such a way that the distribution of the diameter of the corks produced is well approximated by a normal distribution with mean 3 cm and standard deviation of 0.1 cm. The specifications call for corks with diameters between 2.9 and 3.1 cm. A cork not meeting the specifications is considered defective. What proportion of corks produced by this machine are defective?

```r
# Parameters of the normal distribution
mean_diameter <- 3      # Mean diameter in cm
sd_diameter <- 0.1      # Standard deviation in cm

# Specifications for acceptable corks
lower_limit <- 2.9      # Lower bound of acceptable diameter
upper_limit <- 3.1      # Upper bound of acceptable diameter

# Calculate the proportion of corks within the acceptable range
within_spec <- pnorm(upper_limit, mean = mean_diameter, sd = sd_diameter) -
               pnorm(lower_limit, mean = mean_diameter, sd = sd_diameter)

# Proportion of defective corks
proportion_defective <- 1 - within_spec

# Print the result
print(paste("Proportion of defective corks:", round(proportion_defective, 4)))
```

```
> source("/Users/ran/cs/github/DSA/R/5.r", encoding = "UTF-8")
[1] "Proportion of defective corks: 0.3173"
```

6. Using the same scenario as above, 75% of the corks cut by the machine in the earlier example, $N(3, 0.1)$, have a diameter smaller than what value?

```r
# Parameters of the normal distribution
mean_diameter <- 3      # Mean diameter in cm
sd_diameter <- 0.1      # Standard deviation in cm

# Find the 75th percentile
diameter_75th <- qnorm(0.75, mean = mean_diameter, sd = sd_diameter)

# Print the result
print(paste("Diameter below which 75% of corks fall:", round(diameter_75th, 4)))
```

```
> source("/Users/ran/cs/github/DSA/R/6.r", encoding = "UTF-8")
[1] "Diameter below which 75% of corks fall: 3.0674"
```

7. Below is a distribution for number of visits to dentists in one year:

X	0	1	2	3	4
P	0.1	0.3	0.4	0.15	0.05

Determine the expected value, the variance, and the standard deviation of the distribution.

```
# Distribution for number of visits to dentists in one year
X <- c(0, 1, 2, 3, 4)          # Number of visits
P <- c(0.1, 0.3, 0.4, 0.15, 0.05)  # Corresponding probabilities

# Calculate the expected value (mean)
expected_value <- sum(X * P)

# Calculate the variance
variance <- sum((X - expected_value)^2 * P)

# Calculate the standard deviation
std_deviation <- sqrt(variance)

# Print the results
print(paste("Expected Value:", round(expected_value, 4)))
print(paste("Variance:", round(variance, 4)))
print(paste("Standard Deviation:", round(std_deviation, 4)))
```

```
> source("/Users/ran/cs/github/DSA/R/7.r", encoding = "UTF-8")
[1] "Expected Value: 1.75"
[1] "Variance: 0.9875"
[1] "Standard Deviation: 0.9937"
```

8. A genetic trait of one family manifests itself in 25% of the offspring. If eight offspring are randomly selected, find the probability that the trait will appear in exactly three of them.

```
# Parameters for the binomial distribution
n <- 8       # Number of trials (offspring)
p <- 0.25    # Probability of success (trait appearing)
x <- 3       # Number of successes (trait appearing in exactly 3 offspring)

# Calculate the probability
probability <- dbinom(x, size = n, prob = p)

> source("/Users/ran/cs/github/DSA/R/8.r", encoding = "UTF-8")
[1] "Probability of the trait appearing in exactly 3 offspring: 0.2076"
```

9. The number of inaccurate gauges in a group of four is a binomial random variable. If the probability of the defect is 0.1, what is the probability that only 1 is defective? How about more than 1?

```
# Parameters for the binomial distribution
n <- 4       # Number of trials (gauges)
p <- 0.1     # Probability of defect
x <- 1       # Number of defective gauges

# Probability that only 1 is defective
probability_one_defective <- dbinom(x, size = n, prob = p)

# Probability that more than 1 is defective
probability_more_than_one <- 1 - (dbinom(0, size = n, prob = p) + dbinom(1, size = n, prob = p))

# Print the results
print(paste("Probability that only 1 is defective:", round(probability_one_defective, 4)))
print(paste("Probability that more than 1 is defective:", round(probability_more_than_one, 4)))
```

```
> source("/Users/ran/cs/github/DSA/R/9.r", encoding = "UTF-8")
[1] "Probability that only 1 is defective: 0.2916"
[1] "Probability that more than 1 is defective: 0.0523"
>
```

10. Bob is a recent law school graduate who intends to take the state bar exam. About 57% of people who take the state bar exam pass. Determine the probability distribution.

   (a) What is the probability that Bob passes the bar exam on his second try?

   (b) What is the probability that Bob needs 3 attempts to pass the bar exam?

   (c) What is the probability that Bob needs more than 3 attempts to pass the bar exam?

```r
# Probability of passing the bar exam
p <- 0.57

# (a) Probability that Bob passes on his second try
prob_second_try <- (1 - p)^(2 - 1) * p

# (b) Probability that Bob needs 3 attempts to pass
prob_third_try <- (1 - p)^(3 - 1) * p

# (c) Probability that Bob needs more than 3 attempts to pass
prob_more_than_three <- 1 - ( (1 - p)^(1 - 1) * p + (1 - p)^(2 - 1) * p + (1 - p)^(3 - 1) * p )

# Print the results
print(paste("Probability that Bob passes on his second try:", round(prob_second_try, 4)))
print(paste("Probability that Bob needs 3 attempts to pass:", round(prob_third_try, 4)))
print(paste("Probability that Bob needs more than 3 attempts to pass:",
round(prob_more_than_three, 4)))
```

```
> source("/Users/ran/cs/github/DSA/R/10.r", encoding = "UTF-8")
[1] "Probability that Bob passes on his second try: 0.2451"
[1] "Probability that Bob needs 3 attempts to pass: 0.1054"
[1] "Probability that Bob needs more than 3 attempts to pass: 0.0795"
```

11. Each person in a random sample of 1026 adults in the United States was asked the following question. "Based on what you know about the Social Security system today, what would you like Congress and the President to do during the next year?" The response choices and the percentages selecting them are shown below.

Completely overhaul the system:	19%
Make some major changes:	39%
Make some minor adjustments:	30%
Leave the system the way it is now:	11%
No opinion:	1%

Find a 95% confidence interval for the proportion of all United States adults who would respond "Make some major changes" to the question. Give an interpretation of the confidence interval and give an interpretation of the confidence level. Make sure to use correct "=F" in the prop.test function.

```
# Sample size and number of people who responded "Make some major changes"
n <- 1026  # Total sample size
p_major_changes <- 0.39  # Proportion selecting "Make some major changes"
x <- n * p_major_changes  # Number of people selecting "Make some major changes"

# Perform the proportion test
result <- prop.test(x = x, n = n, conf.level = 0.95, correct = FALSE)

# Print the confidence interval
print(result$conf.int)

# Print the interpretation
print(paste("The 95% confidence interval for the proportion of adults who would respond
'Make some major changes' is:",
        round(result$conf.int[1], 4), "to", round(result$conf.int[2], 4)))
```

```
> source("/Users/ran/cs/github/DSA/R/11.r", encoding = "UTF-8")
[1] 0.3606182 0.4202024
attr(,"conf.level")
[1] 0.95
[1] "The 95% confidence interval for the proportion of adults who would respond 'Make some
major changes' is: 0.3606 to 0.4202"
```

12. A researcher wants to determine the mean incubation time for eggs of the broad-tailed hummingbird. An SRS of 30 eggs resulted in a mean incubation time of 17.83 days with standard deviation 2.20 days. Determine a 99% confidence interval for the mean incubation time for eggs for the broad-tailed hummingbird.

```r
# Sample statistics
n <- 30                # Sample size
mean_incubation <- 17.83  # Sample mean
sd_incubation <- 2.20     # Sample standard deviation
confidence_level <- 0.99  # Confidence level

# Calculate the critical t-value
t_critical <- qt((1 + confidence_level) / 2, df = n - 1)

# Calculate the margin of error
margin_of_error <- t_critical * (sd_incubation / sqrt(n))

# Calculate the confidence interval
lower_bound <- mean_incubation - margin_of_error
upper_bound <- mean_incubation + margin_of_error

# Print the results
print(paste("The 99% confidence interval for the mean incubation time is:",
          round(lower_bound, 4), "to", round(upper_bound, 4)))
```

```
> source("/Users/ran/cs/github/DSA/R/12.r", encoding = "UTF-8")
[1] "The 99% confidence interval for the mean incubation time is: 16.7229 to 18.9371"
```

13. Some boxes of a certain brand of breakfast cereal include a voucher for a free video rental inside the box. The company that makes the cereal claims that a voucher can be found in 20% of the boxes. However, based on their experiences eating this cereal at home, a group of students believes that the proportion of boxes with vouchers is less than 0.2. This group of students purchased 65 boxes of the cereal to investigate the company's claim! The students found a total of 11 vouchers for free video rentals in the 65 boxes. Suppose it is reasonable to assume that the 65 boxes purchased by the students are a random sample of all boxes of this cereal. Based on this sample, is there support for the students' belief that the proportion of boxes with vouchers is less than 0.2? In your R code using prop.test, make sure to specify correct=F.

```
# Sample data
x <- 11      # Number of boxes with vouchers
n <- 65      # Total number of boxes
p_null <- 0.2 # Null hypothesis proportion

# Perform the one-sided proportion test
result <- prop.test(x = x, n = n, p = p_null, alternative = "less", correct = FALSE)

# Print the results
print(result)
```

```
> source("/Users/ran/cs/github/DSA/R/13.r", encoding = "UTF-8")

	1-sample proportions test without continuity correction

data:  x out of n, null probability p_null
X-squared = 0.38462, df = 1, p-value = 0.2676
alternative hypothesis: true p is less than 0.2
95 percent confidence interval:
 0.000000 0.258559
sample estimates:
        p
0.1692308
```

14. Here are the measurements (mm) of a critical dimension on a sample of automobile engine crankshafts:

224.120, 224.001, 224.017, 223.982, 223.960, 224.089, 223.987, 223.976,

224.098, 224.057, 223.913, 223.999, 223.989, 223.902, 223.961, 223.980

The manufacturing process is known to vary normally. The process mean is supposed to be 224mm. Do these data give evidence that the process mean is greater than the target value of 224mm? Use a 10% level of significance.

```
# Data: Measurements of crankshaft dimensions (mm)
measurements <- c(224.120, 224.001, 224.017, 223.982, 223.960, 224.089,
                  223.987, 223.976, 224.098, 224.057, 223.913, 223.999,
                  223.989, 223.902, 223.961, 223.980)

# Perform a one-sided t-test
result <- t.test(measurements, mu = 224, alternative = "greater", conf.level = 0.90)

# Print the results
print(result)

> source("/Users/ran/cs/github/DSA/R/14.r", encoding = "UTF-8")

        One Sample t-test

data:  measurements
t = 0.1254, df = 15, p-value = 0.4509
alternative hypothesis: true mean is greater than 224
90 percent confidence interval:
 223.9812      Inf
sample estimates:
mean of x
 224.0019
```

15. A large electric power plant uses ocean water for its cooling system and returns the water to the ocean. An SRS of 10 temperature readings showed the changes in water temperature to be (in $^\circ$F)

$$6, \quad 8, \quad 4, \quad 5, \quad 10, \quad 3, \quad 9, \quad 11, \quad 7, \quad 9$$

A new generator was added to the plant, and environmentalists fear that the average change in water temperature has changed. An SRS of 12 temperature readings after the generator was added showed (in $^\circ$F)

$$9, \quad 11, \quad 15, \quad 12, \quad 7, \quad 12, \quad 10, \quad 13, \quad 8, \quad 11, \quad 14, \quad 8$$

(a) Test the environmentalists claim at the 2% level.

(b) Give a 98% confidence interval. Do the results of your interval and test agree?

```
# Temperature changes before the generator was added
before <- c(6, 8, 4, 5, 10, 3, 9, 11, 7, 9)

# Temperature changes after the generator was added
after <- c(9, 11, 15, 12, 7, 12, 10, 13, 8, 11, 14, 8)

# (a) Perform a two-sample t-test (two-sided) at the 2% significance level
test_result <- t.test(before, after, alternative = "two.sided", conf.level = 0.98, var.equal = FALSE)

# Print the test results
print(test_result)

# (b) Extract and print the 98% confidence interval
conf_interval <- test_result$conf.int
print(paste("98% Confidence Interval for the difference in means:",
        round(conf_interval[1], 4), "to", round(conf_interval[2], 4)))
```

```
> source("/Users/ran/cs/github/DSA/R/15.r", encoding = "UTF-8")

	Welch Two Sample t-test

data:  before and after
t = -3.2702, df = 18.857, p-value = 0.004057
alternative hypothesis: true difference in means is not equal to 0
98 percent confidence interval:
 -6.4567823 -0.8098844
sample estimates:
mean of x mean of y
  7.20000  10.83333

[1] "98% Confidence Interval for the difference in means: -6.4568 to -0.8099"
```