1. Download and import the dataset "finals" into R and answer the following questions.

a) What is the min, max, mean, and median of the final exams?

```
> # create a vector of the final exam scores
> finals <- c(59, 96, 61, 50, 48, 51, 80, 100, 85, 47, 75, 66, 44, 55, 62, 55, 61, 48, 80, 90, 94)
>
> # calculate the min, max, mean, and median of the final exam scores
> min_score <- min(finals)
> max_score <- max(finals)
> mean_score <- mean(finals)
> median_score <- median(finals)
>
> # print the results
> cat("Minimum score:", min_score, "\n")
Minimum score: 44
> cat("Maximum score:", max_score, "\n")
Maximum score: 100
> cat("Mean score:", mean_score, "\n")
Mean score: 67
> cat("Median score:", median_score, "\n")
Median score: 61
```

b) Attach the data distribution table. You can copy it by using the windows snipping tool.

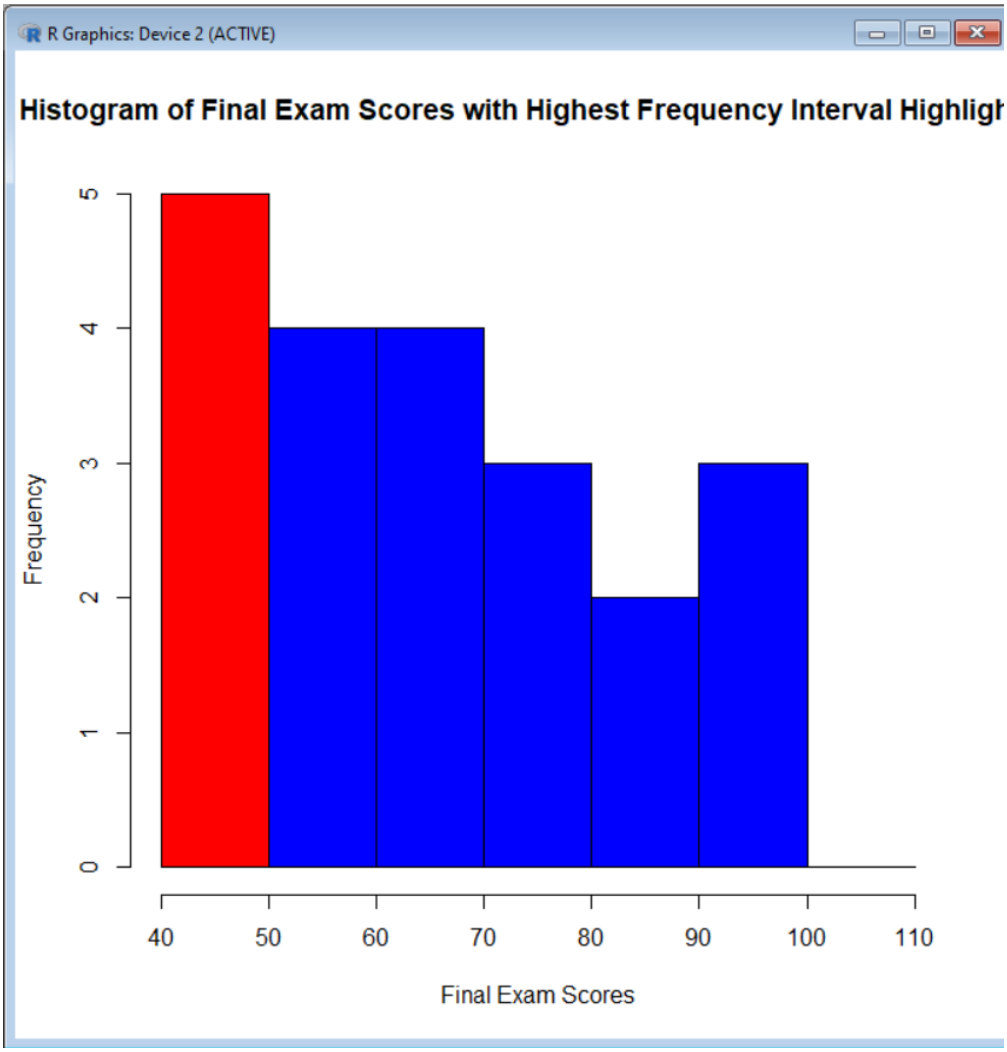
```
> # create a vector of the final exam scores
> finals <- c(59, 96, 61, 50, 48, 51, 80, 100, 85, 47, 75, 66, 44, 55, 62, 55, 61, 48, 80, 90, 94)
>
> # create a frequency table of the final exam scores
> table(finals)
finals
 44  47  48  50  51  55  59  61  62  66  75  80  85  90  94  96 100
  1   1   2   1   1   2   1   2   1   1   1   2   1   1   1   1   1
>
```

c) What value is the most frequent? What are the frequency and relative frequency of that value? Note that relative frequency is the ratio between frequency and the total counts.

```
> # create a vector of the final exam scores
> finals <- c(59, 96, 61, 50, 48, 51, 80, 100, 85, 47, 75, 66, 44, 55, 62, 55, 61, 48, 80, 90, 94)
>
> # calculate the mode of the final exam scores
> mode_score <- names(sort(-table(finals)))[1]
>
> # calculate the frequency of the mode score
> freq_mode <- table(finals)[mode_score]
>
> # calculate the relative frequency of the mode score
> rel_freq_mode <- freq_mode / length(finals)
>
> # print the results
> cat("The most frequent value is:", mode_score, "\n")
The most frequent value is: 48
> cat("Frequency of the most frequent value:", freq_mode, "\n")
Frequency of the most frequent value: 2
> cat("Relative frequency of the most frequent value:", rel_freq_mode, "\n")
Relative frequency of the most frequent value: 0.0952381
>
```

d) Attach the frequency histogram.

```
> # create a histogram of the final exam scores, with the highest frequency interval highlighted
> hist(finals, breaks = seq(40, 110, by = 10), col = ifelse(levels(intervals) == levels(intervals)[which.max(table(intervals))], "red", "blue"), xlab = "Final Exam Scores", ylab = "Frequency", main = "Histogram of Final Exam Scores with Highest Frequency Interval Highlighted")
>
```



e) What interval is associated with the highest frequency? What is that frequency?

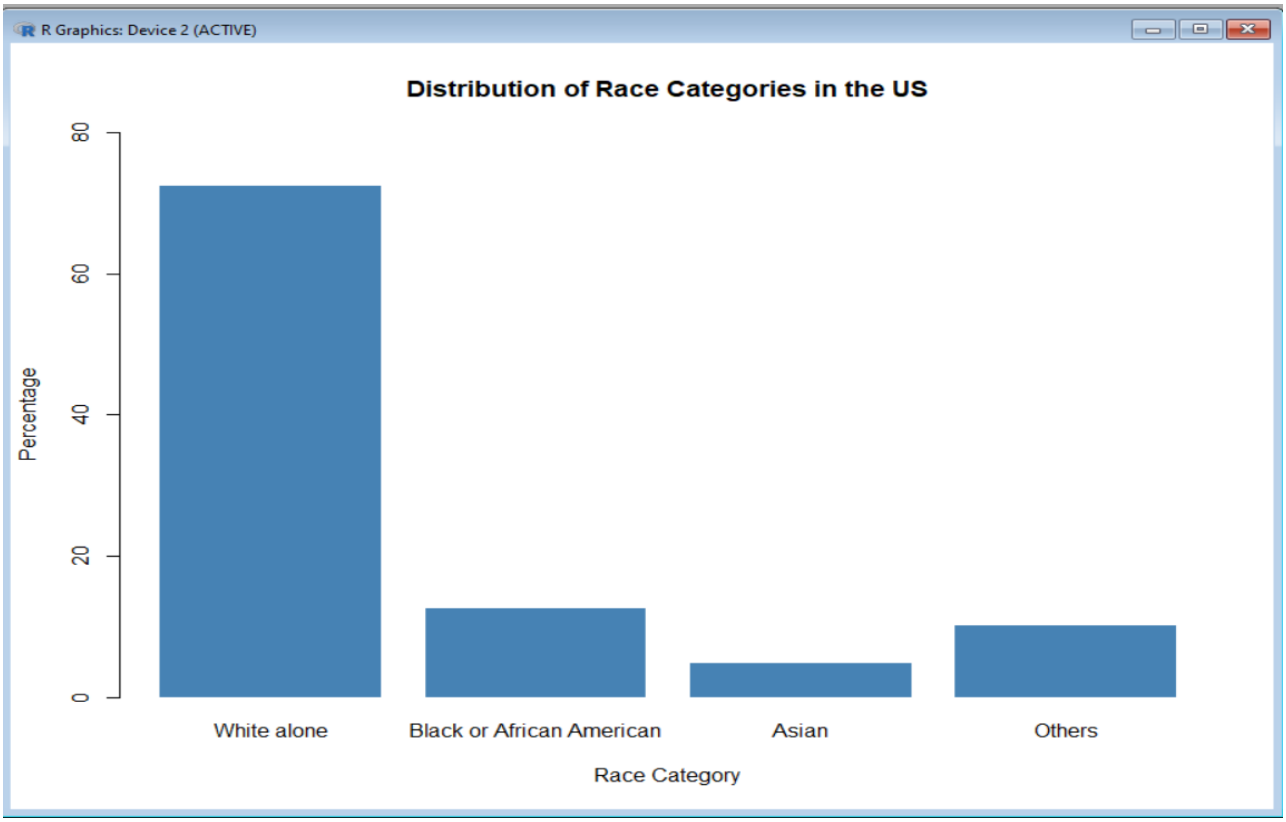
```
> # create a vector of the final exam scores
> finals <- c(59, 96, 61, 50, 48, 51, 80, 100, 85, 47, 75, 66, 44, 55, 62, 55, 61, 48, 80, 90, 94)
>
> # divide the data into intervals of width 10
```

```
> intervals <- cut(finals, breaks = seq(40, 110, by = 10), include.lowest = TRUE)
>
> # count the frequency of each interval
> freq_table <- table(intervals)
>
> # find the interval with the highest frequency
> max_freq_interval <- names(which.max(freq_table))
>
> # find the frequency of the interval with the highest frequency
> max_freq <- freq_table[max_freq_interval]
>
> # print the results
> cat("The interval with the highest frequency is:", max_freq_interval, "\n")
The interval with the highest frequency is: [40,50]
> cat("The frequency of the interval with the highest frequency is:", max_freq, "\n")
The frequency of the interval with the highest frequency is: 5
```

2. According to 2010 U.S Census, the percentages of self-identified races are: "White alone": 7 2.4%; "Black or African American": 12.6%; "Asian": 4.8%; and others. Use the R function barplot to draw a bar chart to visualize the distribution of the percentages of the four categories.

```
> # Create a data frame with the percentages for each category
> percentages <- data.frame(Category = c("White alone", "Black or African American",
"Asian", "Others"),
+                           Percentage = c(72.4, 12.6, 4.8, 10.2))
>
> # Create a bar plot using the barplot function
> barplot(percentages$Percentage, names.arg = percentages$Category,
+         main = "Distribution of Race Categories in the US", xlab = "Race Category",
ylab = "Percentage",
+         col = "steelblue", ylim = c(0, 80), border = NA)
>
> # Add percentage labels to the top of each bar
> text(x = 1:4, y = percentages$Percentage + 2, paste(percentages$Percentage, "%"),
cex = 0.8, col = "white")
```

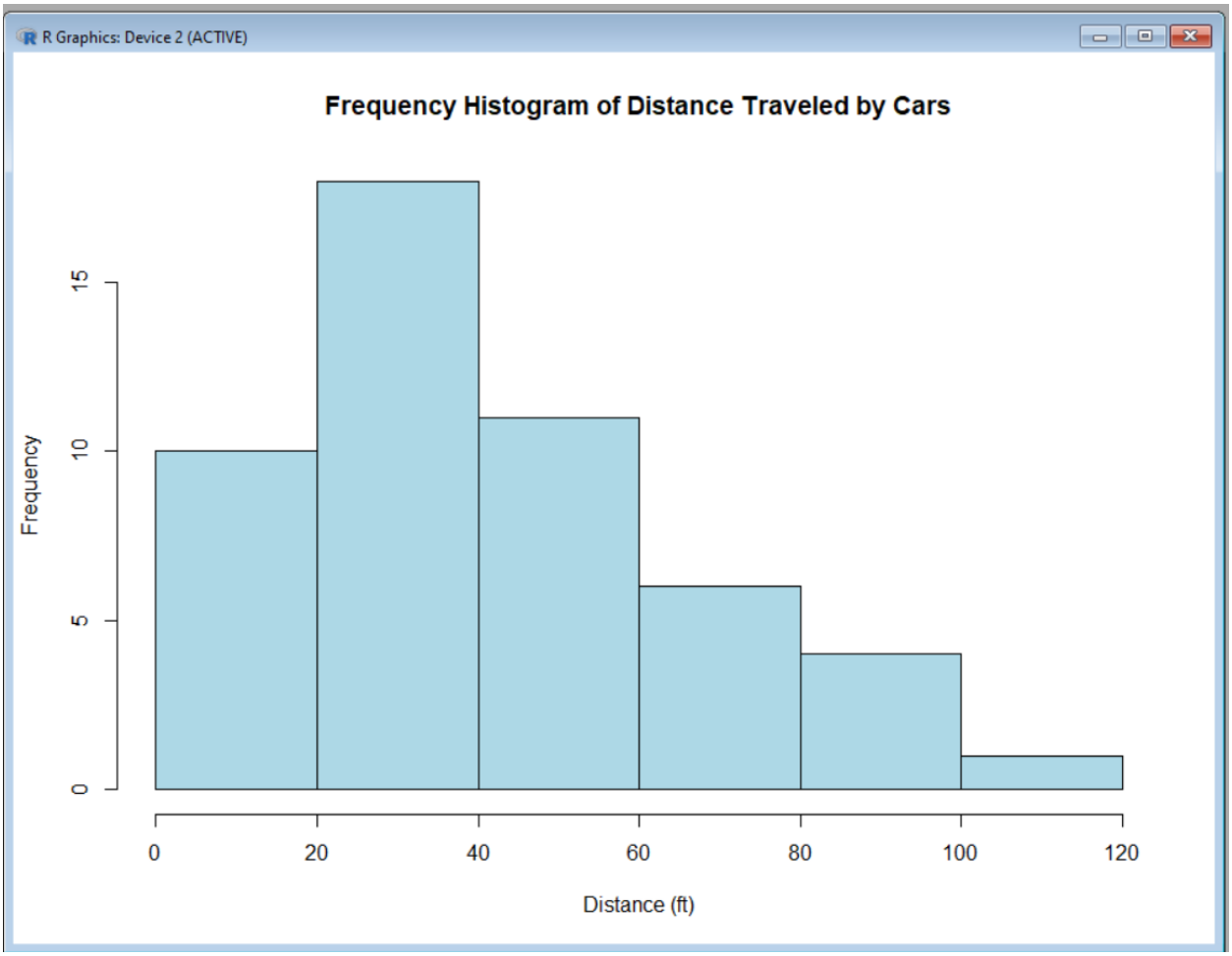
>



3. Consider the variable dist in data "cars" (in built-in R). Draw a frequency histogram and sum marize its distribution pattern in terms of modality, symmetry/skewness, centre, spread, etc. What do you learn from the histogram?

```
# Load the "cars" dataset
data("cars")

# Create a frequency histogram for the "dist" variable
hist(cars$dist, main = "Frequency Histogram of Distance Traveled by Cars",
     xlab = "Distance (ft)", ylab = "Frequency", col = "lightblue")
```


R Graphics: Device 2 (ACTIVE)
Frequency Histogram of Distance Traveled by Cars
Frequency
0
5
10
15
0
20
40
60
80
100
120
Distance (ft)