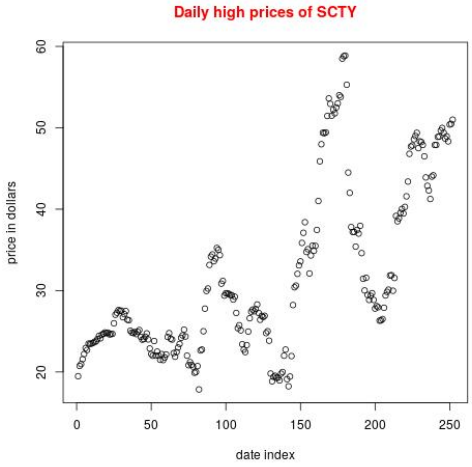
Homework 1 Solutions, Part 1

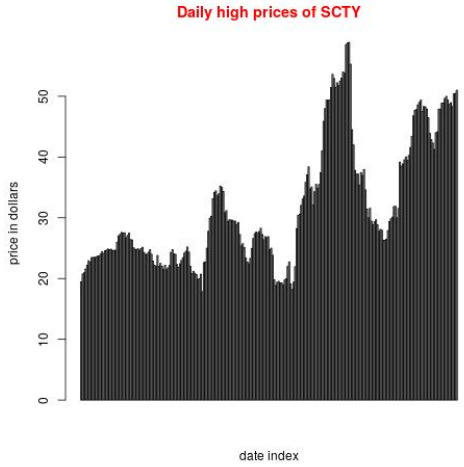
The R script, script1.R produces acceptable solutions:

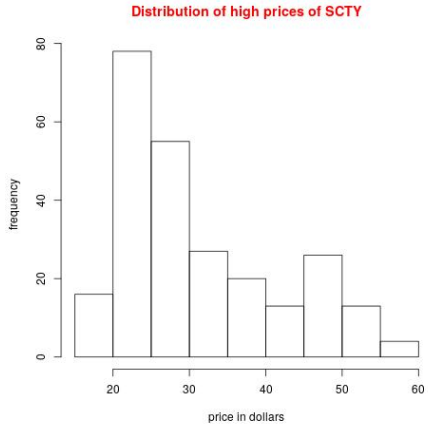
```
> source("script1.R")
[1] "Date"      "Open"      "High"      "Low"       "Close"     "Volume"
[7] "Adj.Close"
[1] "Number of data rows"
[1] 252
[1] "High price was .., on:"
[1] 58.87
[1] 2015-12-17
252 Levels: 2015-09-04 2015-09-08 2015-09-09 2015-09-10 ... 2016-09-02
[1] "Low price was .., on:"
[1] 16.31
[1] 2016-02-11
252 Levels: 2015-09-04 2015-09-08 2015-09-09 2015-09-10 ... 2016-09-02
[1] "Highest spread was .., on:"
[1] 9.240002
[1] 2015-12-16
252 Levels: 2015-09-04 2015-09-08 2015-09-09 2015-09-10 ... 2016-09-02
[1] "Lowest spread was .., on:"
[1] 0.369999
[1] 2016-08-19
252 Levels: 2015-09-04 2015-09-08 2015-09-09 2015-09-10 ... 2016-09-02
>
```

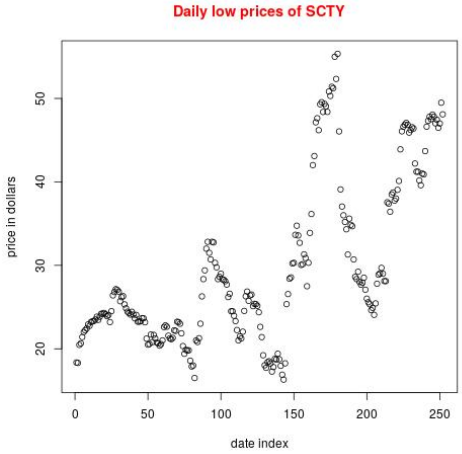
Note the first few lines of the script. In particular, look to see how the date on which the highest High price occurs is obtained using the max function and which.max.

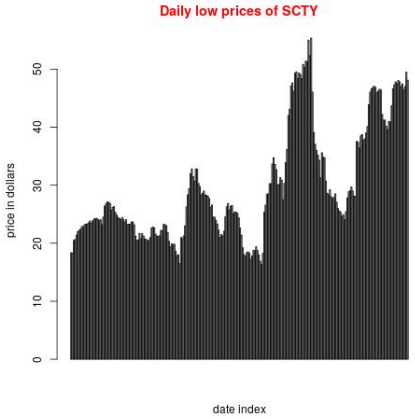
```
mydata = read.csv("table.csv");

datanames = names(mydata);
print(datanames);

print("Number of data rows");
print(length(mydata$Date))

print("High price was .., on:");
print(max(mydata$High))
ind = which.max(mydata$High)
print(mydata$Date[ind])

daily_spread = mydata$High - mydata$Low;
barplot(daily_spread,xlab="date index",ylab="spread in dollars")
title(main="Daily price spread of SCTY", col.main="red")
dev.copy(jpeg,'spread1.jpg')
dev.off()
```

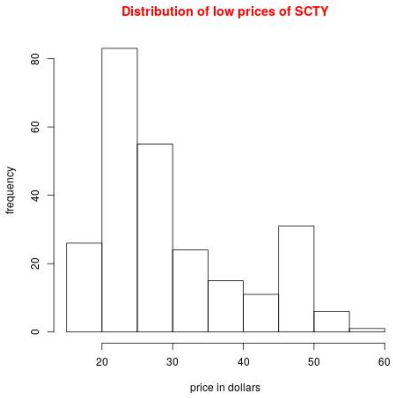

Daily high prices of SCTY
price in dollars
date index

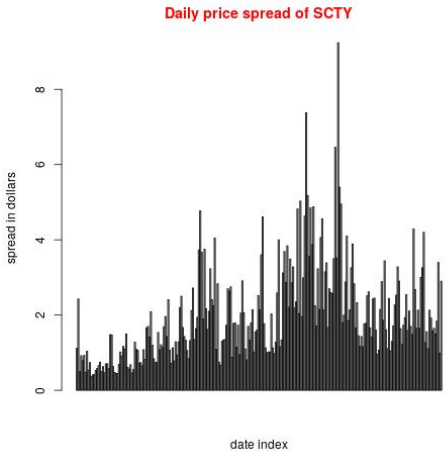

Daily high prices of SCTY
price in dollars
date index


Distribution of high prices of SCTY
frequency
price in dollars


Daily low prices of SCTY
price in dollars
date index


Daily low prices of SCTY
price in dollars
date index


Distribution of low prices of SCTY
frequency
price in dollars


Daily price spread of SCTY
spread in dollars
date index

Section 1-2

1. Statistical significance is indicated when methods of statistics are used to reach a conclusion that some treatment or finding is effective, but common sense might suggest that the treatment or finding does not make enough of a difference to justify its use or to be practical. Yes, it is possible for a study to have statistical significance but not a practical significance.
2. If the source of the data can benefit from the results of the study, it is possible that an element of bias is introduced so that the results are favorable to the source.
3. A voluntary response sample is a sample in which the subjects themselves decide whether to be included in the study. A voluntary response sample is generally not suitable for a statistical study because the sample may have a bias resulting from participation by those with a special interest in the topic being studied.
5. There does appear to be a potential to create a bias.
8. There does appear a potential to create a bias.
9. The sample is a voluntary response sample and is therefore flawed.
10. The sample is a voluntary response sample and is therefore flawed.
17. The male and female pulse rates in the same column are not matched in any meaningful way. It does not make sense to use the difference between any of the pulse rates that are in the same column.
18. Yes, the source of the data is likely to be unbiased.
19. The data can be used to address the issue of whether males and females have pulse rates with the same average (mean) value.

Section 1-3

5. Statistic
6. Parameter
8. Statistic
10. Parameter
15. Discrete
16. Continuous
17. Discrete
24. Ordinal

Section 1-4

1. No. Not every sample of the same size has the same chance of being selected. For example, the sample with the first two names has no chance of being selected. A simple random sample of (n) items is selected in such a way that every sample of same size has the same chance of being selected.
3. The population consists of the adult friends on the list. The simple random sample is selected from the population of adult friends on the list , so the results are not likely to be representative of the much larger general population of adults in the United States.

Section 2-2

10. Class width: 2.
    Class midpoints: 3.95, 5.95, 7.95, 9.95, 11.95.
    Class boundaries: 2.95, 4.95, 6.95, 8.95, 10.95, 12.95, 14.95.

22. Yes. The pulse rates of males appear to be generally lower than the pulse rates of females.

Pulse Rate (Females)	Frequency
50 – 59	1
60 – 69	8
70 – 79	18
80 – 89	5
90 – 99	6
100 – 109	2

24. No, the distribution does not appear to be a normal distribution.

Depth (km)	Frequency
1.00 – 4.99	7
5.00 – 8.99	21
9.00 – 12.99	4
13.00 – 16.99	12
17.00 – 20.99	6

29.

Category	Relative Frequency
Male Survivors	16.2%
Males Who Died	62.8%
Female Survivors	15.5%
Females Who Died	5.5%

31. Pilot error is the most serious threat to aviation safety. Better training and stricter pilot requirements can improve aviation safety.

Cause	Relative Frequency
Pilot Error	50.5%
Other Human Error	6.1%
Weather	12.1%
Mechanical	22.2%
Sabotage	9.1%