Use the length() function to find out how many observations there are.

length(dates)

length(dayOfMonth)

length(month)

length(temp)

length(year)

There are 5534 observations in all.

1. (1) What was the coldest temperature recorded in this time period?

```
> min(temp, na.rm = TRUE)
[1] 38.3
```

The coldest temperature is 38.3.

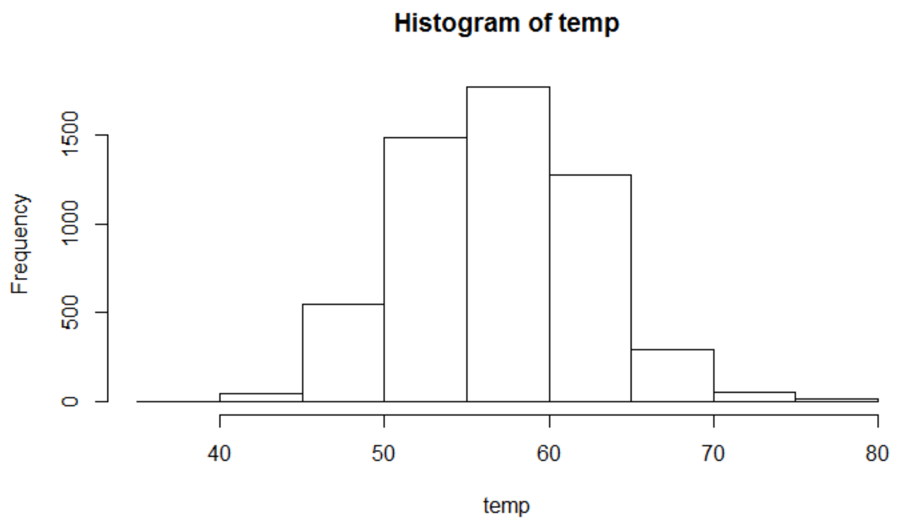
2. (1) What was the average temperature recorded in this time period?

```
> summary(temp)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
  38.30   53.00   57.00   56.96   60.80   79.60      36 
```

The average temperature recorded in this time period is 56.96.

3. (2) What does the distribution of temperatures look like, i.e. are there roughly as many warm as cold days, are the temps clustered around one value or spread evenly across the range of observed temperatures, etc.?

```
> hist(temp)
```



The distribution of temperatures looks like normal. Since the average temperature is 56.96, there are roughly as many warm as cold days. The temperatures are clustered mostly between 50 and 65 degrees, instead of evenly spreading across the range. As temperature becomes lower than 55 degrees or greater than 60 degrees, the number of corresponding days becomes less. There are quite few days which have temperatures lower than 45 degrees or higher than 75 degrees.

4. (1) Examine the first few values of dates. These are a special type of data. Confirm this with class().

```
> head(dates)
[1] "1995-01-01" "1995-01-02" "1995-01-03" "1995-01-04" "1995-01-05" "1995
-01-06"
> class(dates)
[1] "Date"
```

The special type of data is called "Date".

5. (1) We would like to convert the temperature from Fahrenheit to Celsius.

Below are several attempts to do so that each fail.

Try running each expression in R.

Record the error message in a comment

Explain what it means.

Be sure to directly relate the wording of the error message with the problem you find in the expression.

(temp -32)

Error message here: No error message in R.

Explanation here: But the results are temperatures that are 32 smaller than the original ones. NA does not change. They are not the correct Celsius degrees.

(temp - 32)5/9

Error message here: Error: unexpected numeric constant in "(temp - 32)5"

Explanation here: It means numeric constant "5" right after the right parenthesis is not correct. There should be "*" after the right parenthesis.

5/9(temp - 32)

Error message here: Error: attempt to apply non-function

Explanation here: "5/9(temp-32)" is regarded as a function with arguments inside the parentheses, which is not what we want.

[temp - 32]5/9

Error message here: Error: unexpected '[' in "["

Explanation here: "[" is usually used for subsetting, and there should be a name before "[" when applying. This expression is not correct.

6. (1) Provide a well-formed expression that correctly performs the calculation that we want. Assign the converted values to tempC.

tempC = (temp - 32) * 5/9

7. Run the following code to make a plot.

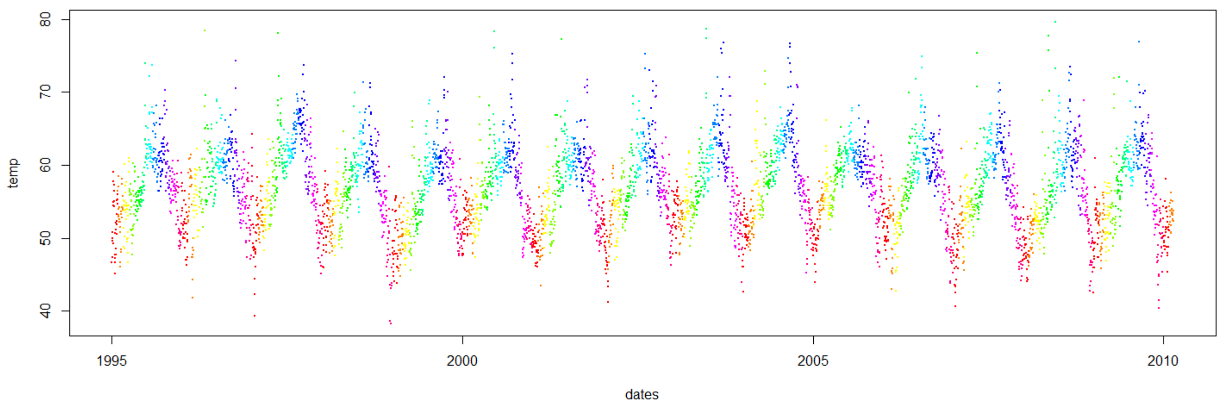
(don't worry right now about what this code is doing)

plot(temp~dates, col = rainbow(12)[month], type="p", pch=19, cex = 0.3)

(1) Use the Zoom button in the Plots window to enlarge the plot.

Resize the plot so that it is long and short, so it is easier to read.

Include this plot in the homework your turn in.



(1) Make an interesting observation about temp in the Bay Area

based on this plot (something that you couldn't see with

the calculations so far.)

Your answer goes here

The temperatures from 1995 to 2010 have a periodic cycle appearance. For each year, the temperature

goes down in winter (around Jan) and goes up in summer (around Jul) with some fluctuations. The ranges of temperatures in different

years are basically equal, and ranges of some certain years may be a bit larger.

(1) What interesting question about the weather in the SF Bay Area

would you like to answer with these data, but don't yet know

how to do it?

Your answer goes here

My question is that how the curve coincides with the perfect periodic distribution.

What are the values of basic parameters in statistics such as period of this curve?

8. (5). Use the following information about you to generate some random values:

#a. Use the day of the month you were born for the mean of the normal.

mean = 2

#b. Use your year of birth for the standard deviation (sd) of the normal curve.

sd = 1993

#c. Generate 5 random values using the parameters from a and b.

#d. Assign the values to a variable named with your first name.

#e. Provide the values generated.

```
> hui = rnorm(5, mean = 2, sd = 1993)
> hui
[1]  -274.33807  2138.20555  -570.00985 -1092.01491    54.14029
```

9. (1). Generate a vector called "normsamps" containing

100 random samples from a normal distribution with

mean 2 and SD 1.

```
> normsamps = rnorm(100, mean = 2, sd = 1)
```

10. (1). Calculate the mean and sd of the 100 values.

```
> mean(normsamps)
[1] 1.995412
> sd(normsamps)
[1] 0.9554894
```

The values may change every time execute the expressions because the numbers are random.

11. (1). Use implicit coercion of logical to numeric to calculate

the fraction of the values in normsamps that are more than 3.

```
> install.packages("Rlab")
> library("Rlab", lib.loc="C:/Program Files for operation/R-3.3.1/library
")
> count(normsamps>3)
[1] 17
> count(normsamps>3)/100
[1] 0.17
```

So the fraction of values more than 3 in normsamps is 0.17. The values may change every time execute the expressions because the numbers are random.

12. (1). Look up the help for rnorm.

You will see a few other functions listed.

Use one of them to figure out about what answer you

should expect for the previous problem.

That is, find the area under the normal(2, 1) curve

to the right of 3. This should be the chance of getting

a random value more than 3. What value do you expect?

What value did you get? Why might they be different?

```
> 1-pnorm(3, mean = 2, sd = 1)
[1] 0.1586553
```

Number 3 exactly equals the mean adds one sd. Because of 3 sigma rule, the probability of

getting a random more than 3 should be roughly 1 - (0.6826/2 + 0.5) = 1 - 0.8413 = 0.1587.

The value I actually get is 0.1586553, just as expected. The value of answer in Question 11

is 0.17, which is different from 0.1586553. That's because normsamps only includes 100 sample

values of the normal distributed curve.