Cherry Blossom Ten Mile Run & Walk

Examining the Impact of Age on Physical Performance

The Raw Data

data is pre-processed from the web in \02_data.r

Further Analysis

1.)

Write a function that uses *read.fwf()* to read the 28 text tables in MenTxt/ and WomenTxt/ into R. These are called 1999.text, 2000txt, etc. and are described in greater detail in 2.2. Examine the tables in a plain text editor to determine the start and end position of each column of interest (name, hometown, age and gun and net time).

Use statistics to explore the results and confirm that you have extracted the information from the correct positions in the text.

2.)

Revise the extractVariables function (see section 2.2) to remove the rows in *menTables* that are blank. In addition, eliminatee the rows that begin with a '*' or a '#'. You may find the following regular expression helpful for locating blank rows in a table.

grep("[1]*$", body)

The pattern uses several meta characters. The ^ is an anchor for the start of the string, the $ anchors to the end of the string, the [[:blank:]] denotes the equivalence class of any space or tab character, and * indicates that the blank character can appear 0 or more times. All together the pattern [2]*$ matches a string that contains any number of blanks from start to end.

[1] [:blank:]
[2] [:blank:]

3.)

Find the record where the time is only 1.5. What happened? Determine how to handle the problem and which function needs to be modified: *extractResTable()*, *extractVariables()*, or *cleanUp()*. In your modification, include code to provide a warning message about the rows that are being dropped for having a time that is too small.

4.)

Examine the head and tail of the 2006 men's file. Look at both the character matrix in the list called *menResMat* and the character vector in the list called *menFiles* (see Sec 2.2). (Recall that the desired character matrix in *menResMat* and the character vector in *menFiels* both correspond to the element named "2006"). What is wrong with the hometown? Examine the header closely to figure out how this error came about. Modify the *extractVariables()* function to fix the problem.