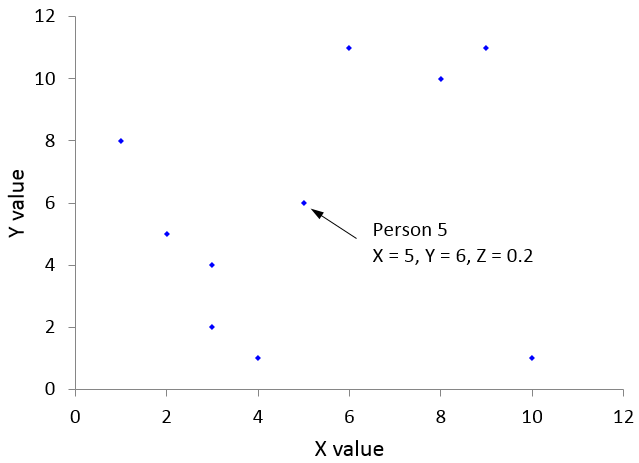
**Golf analytics research position: sample tasks**

**Task 1. Interactive data visualization on the web**

Create a graph that looks something like this on a web page:

  
  
where the user can hover the cursor over a data point and have the information (person number, X, Y, and Z values) about that data point displayed. Here is some sample data:  
  
Person X Y Z  
Person 1 9 11 0.1  
Person 2 1 8 -0.3  
Person 3 2 5 -0.1  
Person 4 4 1 0.4  
Person 5 5 6 0.2  
Person 6 6 11 0.1  
Person 7 3 4 -0.5  
Person 8 8 10 0.4  
Person 9 3 2 -0.4  
Person 10 10 1 0.4  
  
Design your code so that it is easy to create similar interactive graphs with similar, but different, data sets.

**Task 2. Statistical analysis**

Read in data from a text file, compute simple stats and use Monte Carlo simulation to compute other stats. The preferred language is MATLAB. It is not enough to get correct answers. Your code must illustrate great programming technique (modular, commented, efficient, etc.).  
  
**Task description**

1. Read the data from the file round-2014-small.txt. Note that the first row contains column headings. The columns of interest are:

Score, column 16  
 GIR, column 80  
 Putt, column 123

1. Run the regression: Score = a + b\*GIR + c\*Putt  
   and report the coefficients a, b, c, standard errors, R-squared and number of data points.
2. Estimate the probability that the sum of four random scores is less than or equal to 270. For this part, assume independence (i.e., pick four rows at random using a uniform distribution), and estimate the probability using a Monte Carlo simulation with 10,000 trials (i.e., replications). Report the probability estimate and the standard error of the estimate. The simulation code should have a random number seed so the output is exactly reproducible.

The code should be well-commented with appropriate variable names and modular. (See the sample function\_name below for a recommended format.) The output should be written to a text file. The code should be able to run on similar input files with different numbers of rows. Don’t write your own function to do the regression – use a built-in function (you might want to look at more than one to find which one is best for this particular task).   
  
**More details**  
  
1. Modular code

Instead of having one routine, please split into functions for separate tasks (e.g., one for reading in data, one for the regression, one for the simulation and one for writing the results). **Design the routines so they can be easily re-used in other programs.**

2. Efficient code

For this task in MATLAB, there does not need to be any loops, not in reading in the data nor in the simulation. 

1. No numerical values in the body of the code – all constants (and filenames) should be at the top of the code

Example 1:    
 prob = mean(result<=270);

should instead be:

% the next line should be near the top of the code  
score\_cutoff = 270;

 % the body of the code should have the next line

 prob = mean(result<= score\_cutoff);

  Example 2:

prob = sum(res\_vec) / 10000;

should instead be:

% the next line should be near the top of the code  
   num\_trials = 10000;

  % the body of the code should have the next line

  prob = sum(res\_vec) / num\_trials;

Example 3.  The name of the output file should be given at the top of the code.

   output\_filename = ‘golf\_output.txt’

**Recommended MATLAB function format**

function [output\_1 output\_2 ] = function\_name(input\_1, input\_2)  
% function\_name does this:  
% description of the function here  
%  
% input variables  
% input\_1 description of input\_1 here  
% input\_2 description of input\_2 here  
%  
% output variables  
% output\_1 description of output\_1 here  
% output\_2 description of output\_2 here  
%

% body of the function here

% end of function\_name