Lecture 1: R Review & RMarkdown

Problem Set for Lecture 1

These questions will assist in bolstering your understanding of the material in Lecture 1. Emphasis should be placed on having the correct code/output as well as communication. The deliverable should be a knitted RMarkdown document to pdf without any code running off the page. You will be able to present these in office hours as a practice oral assessment interview during the second week of class.

***** First, make sure you have read the Lecture Material's writeup *****

1. Answer the following questions:

   (a) Create a vector named `exam_grades` which has the values:

   $$85,\ 92,\ 35,\ 94,\ 88,\ 75,\ 99,\ 86,\ 83,\ 91$$

   (b) Calculate the mean, median, and standard deviation value for the vector.

   (c) Display all values of `exam_grades` which are less than or equal to 85 or equal to 88, and determine the proportion of values which meet this criteria.

   (d) Calculate the mean for the values of `exam_grades` which are between 80 and 95 (inclusive).

   (e) Create a new vector named `exam_grades_root` which is the square-root of each element of `exam_grades` rounded to 1 decimal place.

   (f) Display the last 3 observations of `exam_grades_root` in a manner that could be scaled up for larger datasets.

2. Answer the following questions:

   (a) Create a vector named `sun` by looking at the `sunspot.year` dataset (it is currently time series data, coerce it into a vector when creating `sun`).

   (b) Display the values of `sun` which are greater than or equal to 2 times the median of `sun`, and determine the proportion which meet this criteria.

   (c) Display the values of `sun` which are within 1 standard deviation of the mean of `sun`, and calculate the proportion which meet this criteria.

   (d) Create a new vector named `sun_log` which contains the natural log of all elements of `sun` that are greater than 10.

   (e) Calculate the product of the first 10 elements of `sun_log` and round the result to the nearest whole number.

   (f) Calculate the sum of the differences `sun - sun_log`

3. Answer the following questions:

   (a) Create a vector named `internet` by looking at the `WWWusage` dataset (it is currently time series data, coerce it into a vector when creating `internet`).

   (b) Calculate the mean, median, and standard deviation value for the vector.

   (c) Display all of the values of `internet` which have indices that are a multiple of 3.

   (d) Calculate the median for the values of `internet` whose index is either below 20 or above 90 (note: this is asking about indices, not element values).

   (e) Create a new vector named `internet_mult` which consists of the `internet` vector multiplied by 0.85.

   (f) Display all of the values from `internet_mult` whose leading decimal place is greater than 0.5 (you might need to create a temporary vector to help you do this).