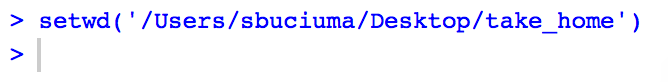
STAT 8030: Midterm Exam, Take-home portion

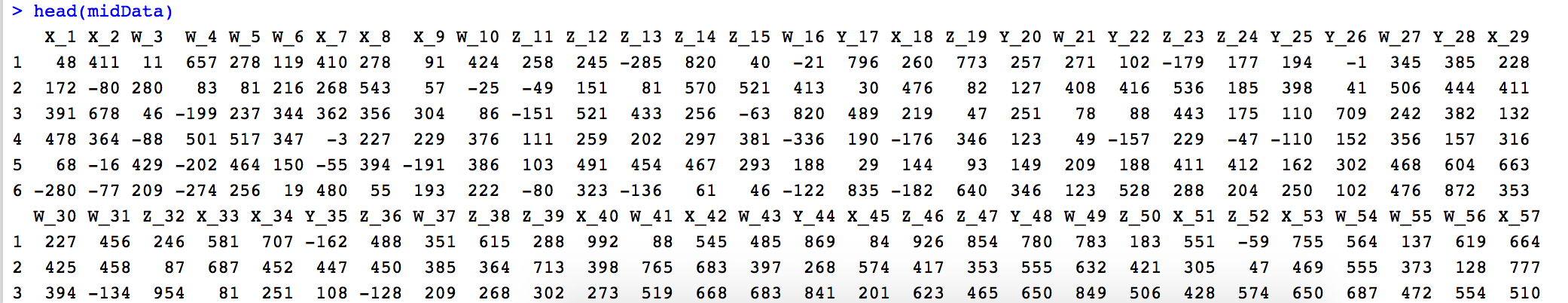
Upload this completed word document along with your R script into D2L by Monday, October 17th at 6:00 PM. You may use notes, the textbook, etc. You can also search online. But you can’t solicit or receive help from a classmate or anyone else—you must work alone and cannot discuss the exam contents with anyone.

The csv file *Midterm.csv* has the data you will use to get started on questions 1-6. Download this data into a folder on your computer. Then complete the following in R:

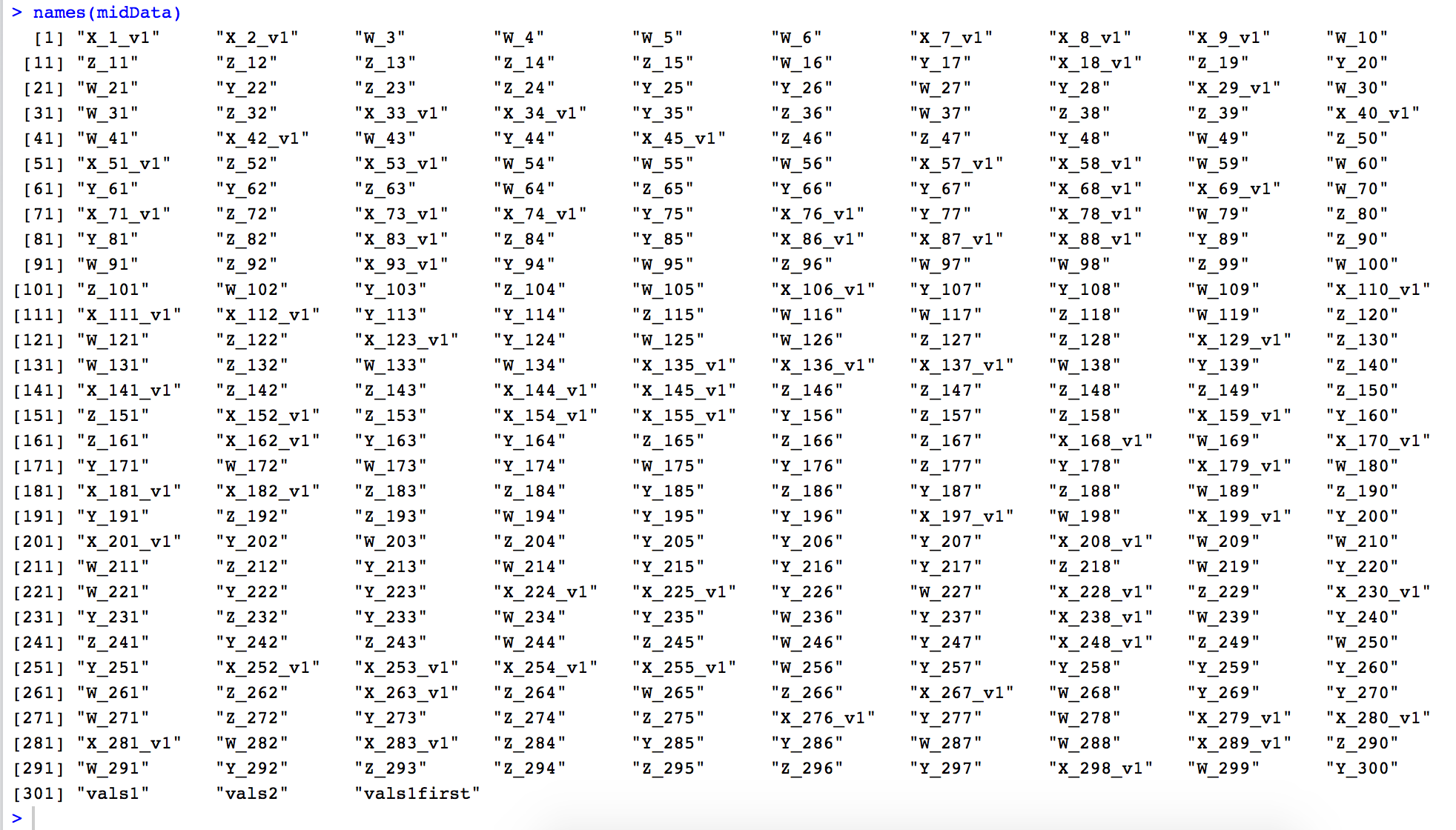
1. Change the working directory to be the same as the folder that contains the downloaded csv file.



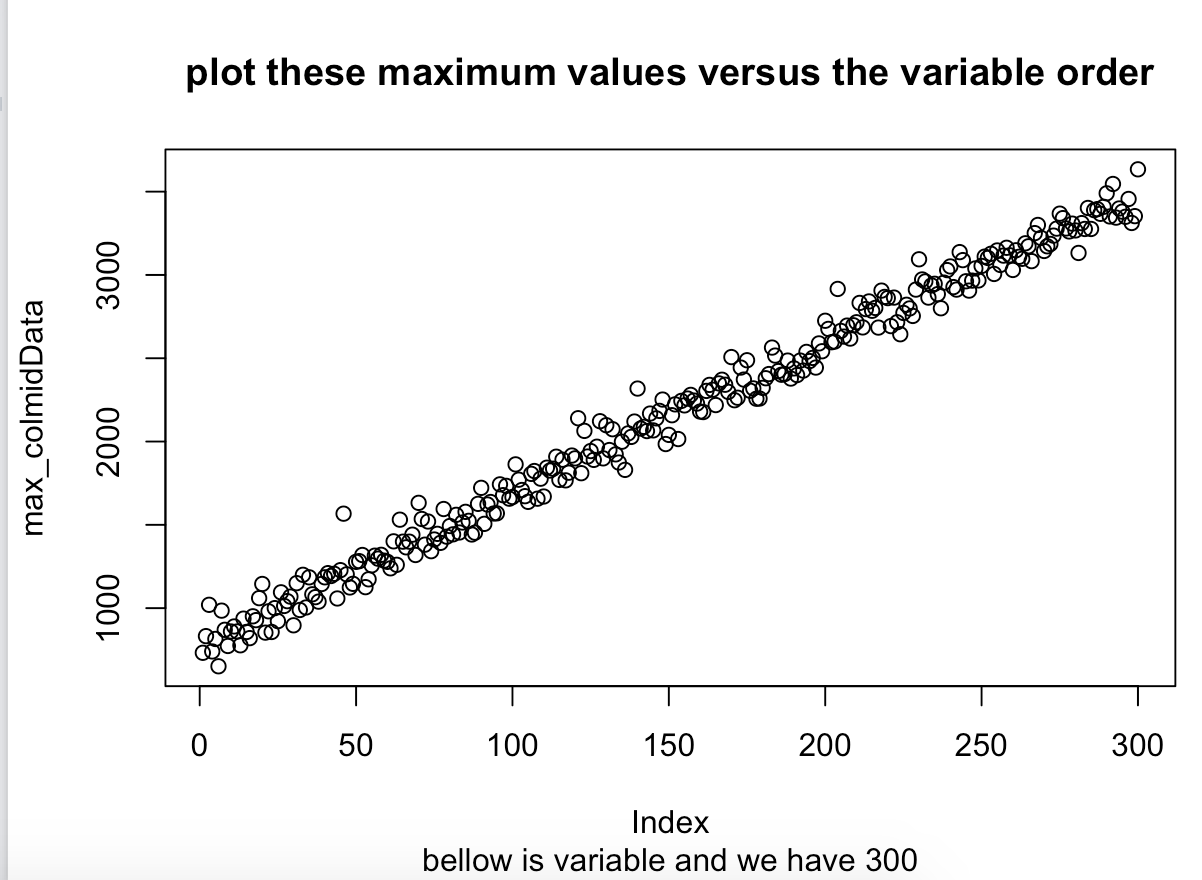
1. Read in the data set as an R data frame with the following name: midData



1. For all variable names in midData that begin with “X”, add “\_v1” to the end of the variable name. If the variable name does not begin with “X”, don’t change the variable’s name



1. For each of the first 300 variables in midData, find the maximum value for the variable. Then plot these maximum values versus the variable order. (That is, the first point would be at (1, maximum of V1), the second point would be at (2, maximum of V2), etc.)

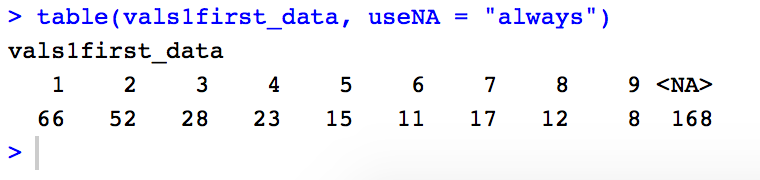


1. Notice that the variable vals1 contains numeric values. The values happen to either be 0, missing, or greater than 0.000001 in absolute value. Here is what I want you to do in creating the new variable, vals1first:
   * If the value of vals1 is known and nonzero, then vals1first should equal the first nonzero digit in vals1 (for example, for -0.0003 it would be 3, for 45.728 it would be 4, and for -999.998 it would be 9)
   * If the value of vals1 is equal to 0 or if it is missing, then there is no available nonzero digit, so the first nonzero digit would be missing. In this instance, vals1first should have a missing value.

(Hint: I did not idly tell you that the nonmissing, nonzero values are at least .000001 in magnitude—this fact can be exploited to your advantage.)

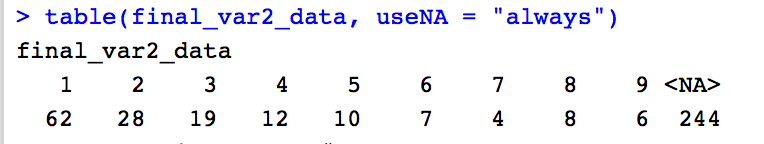
After creating the variable vals1first, create a table of the values in vals1first that shows how many times the first nonzero digit was 1, 2, 3, …,9 and also shows how many times vals1first was missing.

(Hint: the number of 2’s should be 52 and the number of NA’s should be 168; you can use these to help check if you were correct.)



1. Write a function called Nonzero1 that takes as input a numeric vector satisfying the same properties as vals1, and returns a vector with values created according to the same methodology as described in question 5 (that is, the first nonzero digit or a missing value, as the case may be). Use this function with the input vals2 to create the variable vals2first. Create a table of the values in vals2first.

(Hint: To help check that your function works, you might also apply it to vals1 and make sure you get the same vector of output as you did for question 5.)



Source code in .R file.