

BUAN 6357 (Johnston)

Homework 2b

Due: 4 Feb 2018

Points available: 100.

This assignment is only about generating synthetic data. The analysis of the data will be in your next assignment. For this assignment you may use the “tidyverse” and “broom” packages.

The first commands of your code MUST be:

```
setwd(“c:/data/BUAN6357/HW_2”); source(“prep.txt”, echo=T)
```

and the last command of your code MUST be:

```
source(“validate.txt”, echo=T)
```

Be careful with the quote characters as they must ALL be the same. Inclusion of these lines is required BEFORE your code will be tested.

Look in the UTDbox are for this class and find folder class_topics>Classification. Download the files noisy_segments.txt and longintro_RPART.pdf . Background for the data is provided in longintro_RPART Example 4.3 .

Adapt the code from the noisy_segments file to generate 5,000 samples of each of the 10 target digits. Store this result in a data frame named “rawData”. Then generate a new data frame with counts for the frequency of each unique combination of digit/segments. Call the new column “freq”. Include the digit/segments data values. Name this new data frame “countData”.

Keep the random number seed of 1.

Lastly, create a new data frame from “rawData” which has the first 250 samples for each digit. Name this new data frame “raw250”.

Submit the code to eLearning as an ASCII file which can be copied directly into R.

You may submit this assignment as many times as needed until you get full credit.