

(Due: Friday, December 18th)

Lianas are woody vines that grow in tropical forests. Researchers measured liana abundance (stems/ha) in several plots in the central Amazon region in Brazil. The plots were classified into two types: plots that were near the edge of the forest (less than 100 meters from the edge) or plots far from the edge of the forest. The raw data are given in the data file *awesome_tat.csv*. Use R and code to answer the following questions. List all of your R code in the Appendix section.

- (a) Find number of rows and number of columns in the data file.
- (b) What are the column names?
- (c) Select all the stem data (first column) which are associated with category “near” and name it a variable say “x”. Hint: use subset function.
- (d) Select all the stem data (first column) which are associated with category “far” and name it a variable say “y”.
- (e) Summarize your data in a table. A style of the table given below

	sample size	mean	sd
near			
far			

- (f) Plot histograms of the data from part c and part d. Comment on the skewness of the data
- (g) Find 95% confidence interval to test whether there is a significant difference between liana abundance near and far from the edge of the forest. Use both regular text book formula $(\bar{x}_1 - \bar{x}_2) \pm ME$ and direct command `t.test()` using R. Compare your results.
- (h) Find logarithm of the variables “x i.e. $(\log(x))$ and “y i.e. $(\log(y))$ ”. Name the logarithm of x a new variable say “x1” and logarithm of y say “y1”.
- (i) Repeat the test of part (f) and (g) for the transferred data; x1 and y1.
- (j) Compare your results of part (g) and (i). What do these results indicate about the effect on a t-test(of mild skewness when the sample size is fairly large) ?