Lab Exercise #2

Now that you have gone through "The R Guide.pdf" last week, you are familiar with basic R functions now. Meanwhile, you can refer to a list of R references on Canvas for help.

Download Lab2data.txt from Canvas and copy it into your working directory (e.g.\Geog580\Lab\). The file contains 6 columns: (1)Date; (2)Month; (3)Temp2 (daily average temperature in Celsius degrees); (4) preci (precipitation in mm); (5)pollen (daily pollen count); and (6) EVI (The enhanced vegetation index which is an 'optimized' vegetation index designed to enhance the vegetation signal with improved sensitivity in high biomass regions and improved vegetation monitoring through a de-coupling of the canopy background signal and a reduction in atmosphere influences).

Open R or RStudio. Setup your work directory using the setwd() function. For example, setwd("G:/ Geog580/Lab"). For references on any functions, use help().

Load the Lab2data.txt into RStudio using read.table() (Hint: you can assign this table to a new variable. For example: lab2<- read.table("Lab2data.txt ", header = TRUE).

1. How many variables, and how many observations are there in the data set? (Hint: dim()) **(5 points)**
2. What are the pollen count values for the first 10 days? (Hint: lab1$pollen[1:10]) **(5 points)**
3. Calculate the following statistics for pollen count: (i) mean, (ii) variance, (iii) std deviation, (iv) quantile. **(5 points)**
4. Calculate the following statistics for pollen count in April: (i) mean, (ii) variance, (iii) std deviation, (iv) quantile. (Hint: subset(lab1,Month=="April")) **(10 points)**
5. Use hist() to plot 3 count histograms (in frequencies) of pollen count in April, May, and June (Assign a main title for each histogram). Arrange the histograms in 1 single graph (Hint: par(mfrow=c(1,3)). Take a screenshot of the graph and describe the results. (Hint: use subset() function to extract data in May and June first) **(15 points)**
6. Use qqnorm() and qqline() to construct plots for temperature and pollen count for the whole dataset. Take a screenshot of the graphs. Based on the graphs, are they normally distributed? You can further perform Shapiro-Wilk Normality Test by using shapiro.test(). **(10 points)**
7. Use plot() to construct two graphs for pollen count and EVI for the whole dataset. Use date as the x axis and assign a title for the x axis, a title for the y axis, and an overall title. Take a screenshot of the graphs and describe the patterns. **(15 points)**
8. Conduct a two-sample t-test to test whether pollen count in April is significantly different from that in May. (Hint: t.test()) **(10 points)**
9. Fit a linear regression between the pollen count (dependent variable) and EVI (independent variable). Please provide the formula of the regression with the coefficient. Is EVI a significant predictor? What is the adjusted R square? Plot the data (pollen count Vs EVI) and add the fitted regression line to the plot. Please also plot the residual in a separate graph. **(15 points)**
10. Fit a linear regression between the pollen count (dependent variable) and EVI, precipitation, temperature (independent variables). Please provide the formula of the regression with the coefficient. Which independent variables are significant? What is the adjusted R square? Plot the residual. **(10 points)**

Your answers should be in a Word document with each answer numbered. The format for naming your labs is:

lastname_ lab#.doc ie Smith_ Lab2