REM 429 Point pattern data lab

Background:

When learning new programs or techniques it's generally advisable to work with data that has already been prepared for use or analysis. However, in your future jobs or graduate studies you will almost certainly not be working with data that has been "quality controlled" in this manner! Instead you're more likely to be given "raw" data that still has various issues or inconsistencies that will make it either unusable or give faulty results until it has been subject to some quality control processes.

The exact methods of quality control will depend on the type of data and your general understanding of different issues that may have arisen in the acquisition of that data. For example, in remote sensing applications using Landsat data you would need to ensure that you've gotten data without cloud cover that obstructs the area you're examining and you would need to be on the lookout for data artifacts that are the result of sensor misreadings, which means that visual inspection of your images prior to analysis is crucial. With vegetation species and cover data you would need to check that all species codes have been correctly entered and that your cover percentages sum to 100 (depending on the method being used to estimate cover), so you might need to go through and use a sum() function in Excel for each point.

For this lab you will be given data from two stands (Caffey Hill [1.5 ha] and Red-Tail Ridge [1.8 ha]) of longleaf pine (*Pinus palustris*). This data is a census, meaning that all longleaf with a DBH (diameter at breast height, a commonly used measurement in forestry) greater than 2.5 cm were tagged, mapped, and measured. The stands were initially measured in 1999, and re-measured in 2005, 2008, and 2013 (although you will be dealing just with the 1999 and 2013 data). In 2013 all new in-growth of longleaf (i.e. all trees that were previously too small but were now >2.5 cm DBH) was tagged (with tags numbered >2000), mapped, and measured. Prescribed burns were conducted at Caffey Hill in 2004 and 2010, and at Red-Tail Ridge in 2006.

Quality control instructions:

With data like these we might expect that a large proportion of issues will be DBH measures that are somehow anomalous, given that in general trees should either increase in size or remain the same size through time unless they die. Therefore the best way to spot many potential issues in this data is to look at the *change* in the size of a tree between remeasurements. Another potential issue is missing data, which will result in blank cells.

Therefore, our quality control measures are concerned first with finding these "problem children" data and secondly with figuring out a way to deal with these data. There are many different potential solutions to anomalous or missing data, including:

- throwing out the datapoint completely (not preferable because it's losing you data, but could be necessary if you cannot confidently apply some "fix")
- estimating the datapoint by rounding up or down (when would this be reasonable? hint: because these measurements have the possibility of human error it's likely reasonable to assume that changes of <1 cm are human error and therefore can be rounded)
- estimating a value for the datapoint via an equation (this would likely only be reasonable if you had a large dataset and multiple years of measurements in order to calibrate your results, which you don't have in this dataset)

Certain issues are also more problematic for different types of analysis. For example, if a tree was measured but the people sampling the plot forgot to record its location (so it's missing X and Y coordinates) then those data could be used for overall stand averages but would have to be discarded for any spatial analysis.

Report instructions:

With data like this researchers and managers would likely be very interested in how stand structure has changed over time and the spatial characteristics of different size classes of trees or of mortality. Stand structure can be characterized in a variety of ways, but for a preliminary report (which is essentially what you're preparing) you should include:

- From Excel
  1) Graph showing the DBH distribution (see the Varner et al. 2003) and how that's changed over time in each stand
  2) Table showing tree density per ha, mortality density per ha, average DBH, and maximum DBH for 1999 and 2013 for each stand
- From ArcMap (See "ArcMap instructions" document on BbLearn)
  1) Stem maps with some meaningful mark (size [DBH] of the tree **or** status [live vs dead]) for 1999 and 2013 for each stand
  2) Ripley's K analysis for each stand (run on just the 2013 data). You could test the clustering or dispersion patterns of different tree size classes **or** mortality patterns. Running a Ripley's K on the entire dataset does **not** count, see "ArcMap instructions" document for further details.