Software Installation Instructions for the Introduction to R Workshop (Instructor: Shannon E. Albeke)

**Installing R for the first time:**

- Typically, having the most up-to-date version of R is preferable, as algorithms are updated to be more efficient or bugs are fixed. For this workshop we will be using R 3.1.0. To download and install R, follow these steps:
    - Go here: http://cran.us.r-project.org/
    - Click on the operating system that matches your CPU (i.e. Windows)
    - Click on the link 'install R for the first time'
    - Download and Install the newest version (should be the top most hyperlink)
        - Simply accept all of the defaults while installing R
        - It is OK to have more than one version of R installed on a single CPU

**Installing R Studio**

- Since R is command line, having an Integrated Development Environment (IDE) can be extremely helpful. I have come to really appreciate R Studio as an IDE. To download and install R Studio, follow these steps:
    - Go here: http://www.rstudio.com/ide/download/
    - Select the 'Desktop' version
    - Download and Install the 'Recommended Version'
        - Simply accept all of the defaults while installing R Studio

**Setting up R Studio**

- R can be customized to perform additional calculations if you install 'Packages' that others have developed. For this workshop we are focusing on using spatial data within R. Thus we need to install some of the available spatial packages. To install the packages:
    - Open R Studio
    - In the bottom right pane, click on the 'Packages' tab
    - Next click the Install Packages button and install, one at a time, each of the following packages (you can try to install multiple packages concurrently, but I have had some issues at times, so singular installation is 'safer')

        - '***ROCR***': ROC graphs, sensitivity/specificity curves, lift charts, and precision/recall plots are popular examples of trade-off visualizations for specific pairs of performance measures. ROCR is a flexible tool for creating cutoff-parameterized 2D performance curves by freely combining two from over 25 performance measures new performance measures can be added using a standard interface). Curves from different cross-validation or bootstrapping runs can be averaged by different methods, and standard deviations, standard errors or box plots can be used to visualize the variability across the runs. The parameterization can be visualized by printing cutoff values at the corresponding curve positions, or by coloring the curve according to cutoff. All components of a performance plot can be quickly adjusted using a flexible parameter dispatching mechanism. Despite its flexibility, ROCR is easy to use, with only three commands and reasonable default values for all optional parameters.

        - '***SDMTools***': This package provides a set of tools for post processing the outcomes of species distribution modeling exercises. It includes novel methods for comparing

models and tracking changes in distributions through time. It further includes methods for visualizing outcomes, selecting thresholds, calculating measures of accuracy and landscape fragmentation statistics, etc.. This package was made possible in part by financial support from the Australian Research Council & ARC Research Network for Earth System Science.

- '***plotrix***': Lots of plots, confidence intervals, various labeling, axis and color scaling functions.

- '***sp***' : A package that provides classes and methods for spatial data. The classes document where the spatial location information resides, for 2D or 3D data. Utility functions are provided, e.g. for plotting data as maps, spatial selection, as well as methods for retrieving coordinates, for subsetting, print, summary, etc.

- '***rgdal***' : Provides bindings to Frank Warmerdam's Geospatial Data Abstraction Library (GDAL) (>= 1.6.3) and access to projection/transformation operations from the PROJ.4 library. The GDAL and PROJ.4 libraries are external to the package, and, when installing the package from source, must be correctly installed first. Both GDAL raster and OGR vector map data can be imported into R, and GDAL raster data and OGR vector data exported. Use is made of classes defined in the sp package. Windows and Mac Intel OS X binaries (including GDAL, PROJ.4 and Expat) are provided on CRAN.

- '***rgeos***' : Interface to Geometry Engine - Open Source (GEOS) using the C API for topology operations on geometries. The GEOS library is external to the package, and, when installing the package from source, must be correctly installed first. Windows and Mac Intel OS X binaries are provided on CRAN.

- '***raster***' : Reading, writing, manipulating, analyzing and modeling of gridded spatial data. The package implements basic and high-level functions and processing of very large files is supported.

- '***spatstat***' : A package for analyzing spatial data, mainly Spatial Point Patterns, including multitype/marked points and spatial covariates, in any two-dimensional spatial region. Also supports three-dimensional point patterns, and space-time point patterns in any number of dimensions. Contains over 1000 functions for plotting spatial data, exploratory data analysis, model-fitting, simulation, spatial sampling, model diagnostics, and formal inference. Data types include point patterns, line segment patterns, spatial windows, pixel images and tessellations. Exploratory methods include K-functions, nearest neighbour distance and empty space statistics, Fry plots, pair correlation function, kernel smoothed intensity, relative risk estimation with cross-validated bandwidth selection, mark correlation functions, segregation indices, mark dependence diagnostics etc. Point process models can be fitted to point pattern data using functions ppm, kppm, slrm similar to glm. Models may include dependence on covariates, interpoint interaction, cluster formation and dependence on marks. Fitted models can be simulated automatically. Also provides facilities for formal inference (such as chi-squared tests) and model

diagnostics (including simulation envelopes, residuals, residual plots and Q-Q plots).

- '***spdep***' : A collection of functions to create spatial weights matrix objects from polygon contiguities, from point patterns by distance and tessellations, for summarizing these objects, and for permitting their use in spatial data analysis, including regional aggregation by minimum spanning tree; a collection of tests for spatial autocorrelation, including global Moran's I, APLE, Geary's C, Hubert/Mantel general cross product statistic, Empirical Bayes estimates and Assunção/Reis Index, Getis/Ord G and multicoloured join count statistics, local Moran's I and Getis/Ord G, saddlepoint approximations and exact tests for global and local Moran's I; and functions for estimating spatial simultaneous autoregressive (SAR) lag and error models, impact measures for lag models, weighted and unweighted SAR and CAR spatial regression models, semi-parametric and Moran eigenvector spatial filtering, GM SAR error models, and generalized spatial two stage least squares models.

- '***RANN***' : Finds the k nearest neighbours for every point in a given dataset in O(N log N) time using Arya and Mount's ANN library (v1.1.3). There is support for approximate as well as exact searches, fixed radius searches and bd as well as kd trees.

- '***gstat***' : Variogram modelling; simple, ordinary and universal point or block (co)kriging, sequential Gaussian or indicator (co)simulation; variogram and variogram map plotting utility functions.

- '***adehabitatHR***' : A collection of tools for the estimation of animals home range