RenR 690 – GIS Lab

Introduction to Multivariate Statistical Tools and Techniques in ArcGIS 10.x

GIS software excels at the visualization and analyses of attributes (variables or characteristics) that are tied to location. More complex computations are best done in statistical software, and therefore efforts are made to integrate Esri's ArcGIS Desktop[1] and the open-source statistical software R[2]. However, when packages that the ArcGIS with R[3] framework relies on are no longer maintained/available or not readily accessible to the GIS, you will need to analyze your data directly in R, so knowing a reliable workflow for going back and forth between the two software applications is a useful skill. Python[4] is also increasingly used for statistics in ArcGIS.

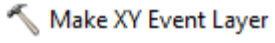
ArcGIS software contains many tools for statistical analyses. The methods and tips demonstrated here provide a preliminary ArcGIS guide on:

1. displaying and managing spatial data for/from statistical analysis
2. making a map to visualize and present
3. using the built-in multivariate tools

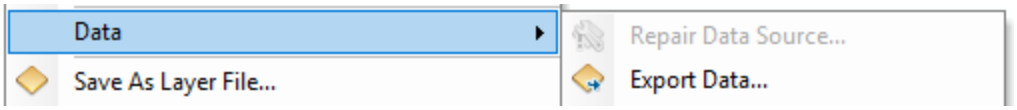
1. Visualize and Convert Spatial Data

One of the best methods to convey information from your analyses is to illustrate with graphs and maps in a GIS. If you don't already have experience with ArcGIS, you will quickly learn that getting the data in to and out of GIS is not necessarily a smooth workflow, with many issues to consider: data acquisition and file formats, map projections and coordinate systems, and types of attributes (a.k.a. categories and magnitudes of the data values), all of which are hopefully described in the associated metadata (i.e. information that describes, or documents, a geographic dataset to facilitate the intelligent use of it). Here we will work primarily with the **AB_Climate.csv** file to acquire skills in displaying, mapping, and managing spatial data.

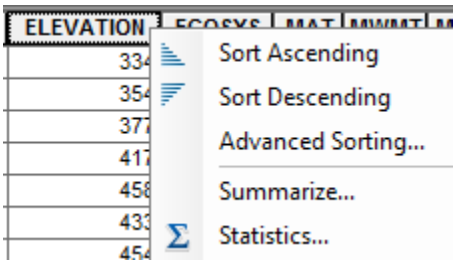
- Copy and unzip the gis.zip file to your working directory; e.g. D:\temp2
- Open ArcMap from the Start > Programs > ArcGIS menu, using a new blank map.
- Click the icon to show ArcToolbox and drag its window to dock to one side.
- Click the icon to show Catalog and drag its window to dock to one side.
- In the Catalog window, click the Connect to Folder button and navigate to your main working directly, highlight the D:\temp2\gis folder and click OK.
- Click the Toggle Contents to view the Catalog tree and Contents panel together.
- Display tabular data: In ArcToolbox, open Data Management Tools > Layers and Table Views > Make XY Event Layer (alternative is File > Add Data > Add XY Data) and specify the following:
  - XY Table: D:\temp2\gis**AB_Climate.csv**
  - X Field: X
  - Y Field: Y



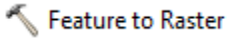
  - Spatial Reference: Click the edit button to choose Projected > National Grids > Canada > NAD_1983_10TM_AEP_Forest *(MUST be same as coordinates! Use NAD 1983 system.)*
  - Click OK
- Export to shapefile: Export the event layer so that we have a true spatial data file – this example uses the shapefile format. Right-click on the AB_Climate_Layer name and choose Data > Export Data and save the output feature class as D:\temp2\gis**abclim.shp**
  (IMPORTANT TIP: Do not accept the Default.gdb location – you can control where your output goes!)



- View attribute table: In the table of contents panel, right-click on **abclim** to open attribute table and view the available variables – these can be used for symbolizing the points in various ways (e.g. remember the Quantities – Graduated Colors method used on the event layer point representation in the PCA lab) as well as used for data conversion and analyses.
- Quick descriptive statistics: Right-click on any of the numeric field headings to access the Statistics. Close when done. Attribute values are also used for converting to raster data below. Close the table when finished examining it.



- Convert *to* raster data: Working with so many point locations can be inefficient (and cause ArcMap to crash, even when simply symbolizing), so we convert the point features to raster. A raster is data stored as a matrix of cells/pixels organized into a grid of rows and columns. Each cell value in the raster grid represents information, such as categorical (ecosystems) or numerical (elevation or climate variables). Choosing the appropriate cell size[5] is of utmost importance. In the Tools toolbar, click on the Zoom In tool to click and drag a rectangle over a very small portion of the **abclim** points and then click the Measure tool to use the ruler to learn the distance between the points in the X and Y directions. Now we are ready to use the raster conversion tools[6]. Note that the Point to Raster tool may be a better choice in some conversion instances because of the greater control over the assignment of cell values when the output cell size is large enough to include multiple input points. However, we will use the simple Feature to Raster tool because the X and Y distances are equal for our data. Click to zoom to Full Extent and then in ArcToolbox, open Conversion Tools > To Raster > Feature to Raster and specify the following:
  - Input features: **abclim**
  - Field: ELEVATION
  - Output raster: D:\temp2\gis**elevation.tif**
  - Output cell size: 4000
  - Click OK



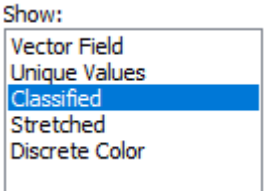
- Batch processing: Repeat the conversion for all other variables in the **abclim** attribute table. We can use batch processing[7] to reduce the repetition of opening the tool for many runs. Right-click on the Feature to Raster tool and choose Batch, double-click on the #1 left gray box to open the tool and fill in as above using a new attribute and output raster name.

Click the Add button for each new variable to be converted and complete the inputs/outputs. It is typically good practice to incorporate the name of the conversion field in the name of the output raster. When the batch grid is completely filled, click OK to run all processes.

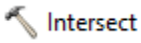
	Input features	Field	Output raster	Output cell size
1	abclim	ELEVATION	C:\Workspace\gis\elevation.tif	4000
2	abclim	MAT	C:\Workspace\gis\mat.tif	4000
3	abclim	MAP	C:\Workspace\gis\mao.tif	4000
4	abclim	DD0	C:\Workspace\gis\dd0.tif	4000
5	abclim	FFP	C:\Workspace\gis\ffp.tif	4000
6	abclim	PAS	C:\Workspace\gis\pas.tif	4000

- Add and manage data: From the Catalog window, add **geol_srf** file by dragging and dropping to the data view. Click the check box beside all layers to turn them off (TIP: Hold the Ctrl key to simultaneously turn on/off layers) and then to view each new raster that has been added to the data frame display, click the empty box to turn visibility on.

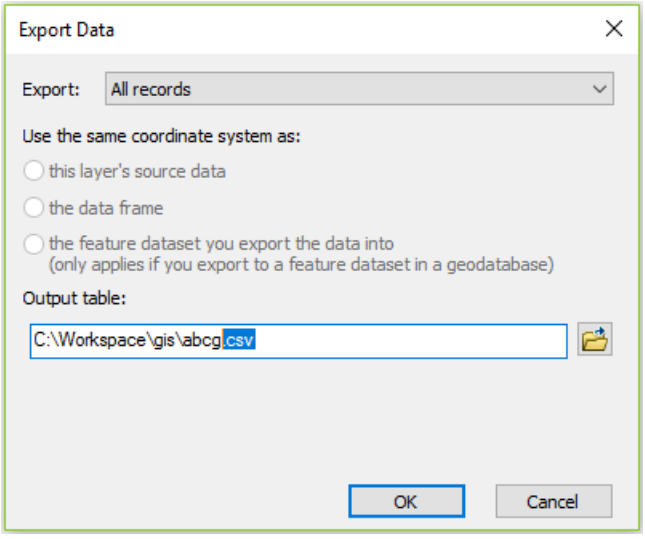
- Thematic classification of continuous rasters: Double-click any raster name to access its Layer Properties > Symbology tab. Modify to show as Classified[8], click the Classify button and to view the histogram and classification statistics; optionally, set the number of classes, method, etc. and click OK. From the drop-down choose a color ramp and click OK. With a little trial and error with the classification methods, breaks, and color ramps, each of the rasters representing the climate variables can be symbolized to best display their information. Categorical rasters, such as ecosys, should be shown as unique values by the ECOSYS field.



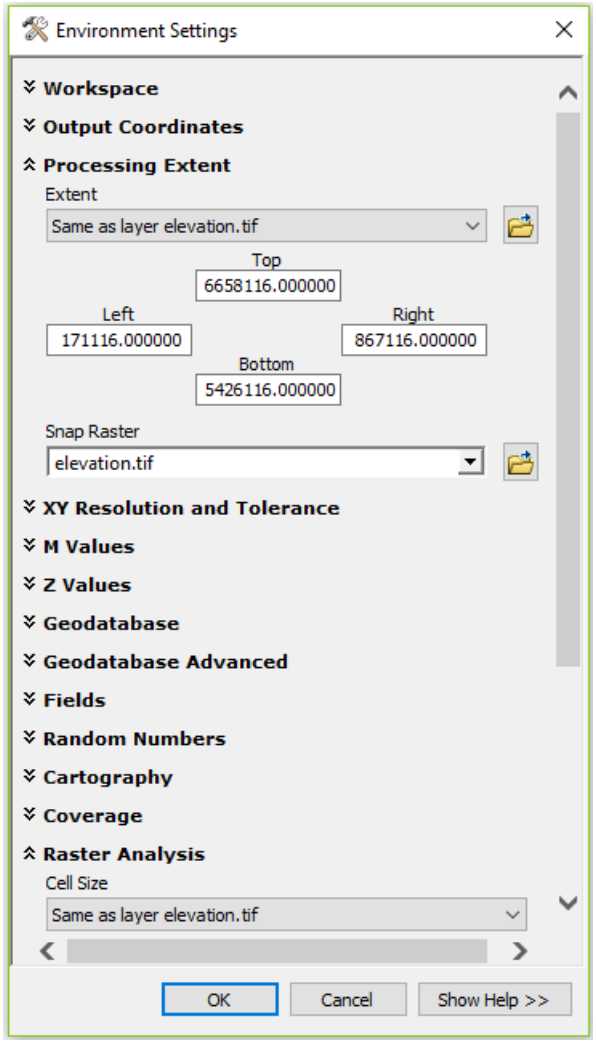
- Intersect: Geology data is provided for experimenting with the conversion tools and we have several options for organizing our data. For example, if we simply want to get the geology values at each point location, then use ArcToolbox's Analysis Tools > Overlay > Intersect **abclim** and **geol_srf** to output to a new point shapefile named **abcg.shp** – it will contain the attributes from the both inputs. If desired, use ArcToolbox's Data Management Tools > Fields > Delete Fields to permanently remove any unneeded columns of values.



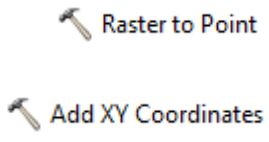
- Export to CSV for R: To save as a Comma Separated Values (CSV) text file, right-click on the **abcg** layer name to open attribute table, click the table options button, then Export, click the folder browse button to navigate to your working directory, choose to save as type: Text, type an output Name with the extension '**.csv**' and click Save. (Note: R can be set up to read *.txt and *.dbf tables, and these formats can be exported using the same attribute table dialog.)



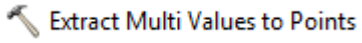
- Convert with environment settings: If we don't have another point layer to intersect with, or we need the GIS raster for statistical analysis directly in ArcGIS, then similar to the conversion of the **abclim** points above, we convert the geology polygons to raster. Note that the Polygon to Raster tool has more control over the cell assignment, but the simple Feature to Raster tool is used here. Open the attribute table for the **geol_srf** layers to decide on an attribute to convert on. In ArcToolbox, open Conversion Tools > To Raster > Feature to Raster and convert the **geol_srf** polygon shapefile to raster. It is very important that the output rasters have the same cell size, snap raster and extent as the **abclim**-based rasters, so BEFORE you click OK, click the Environments button to set these. As before, use the field name in the output raster name and save to the folder you want, before running the conversion.



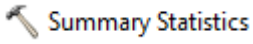
- Future reference – convert *from* raster data: If you want to export the data from any raster to a CSV table (i.e. to take data from ArcGIS for analysis in R), there are two common workflows:
  1. To create a new point file use ArcToolbox's Conversion Tools > From Raster > Raster to Point, followed by Data Management Tools > Features > Add XY Coordinates (to calculate fields of the coordinate values) and then open the attribute table to export as text type with output filename**.csv**. This method creates a table in which each row has values for each raster cell center, and the coordinates.



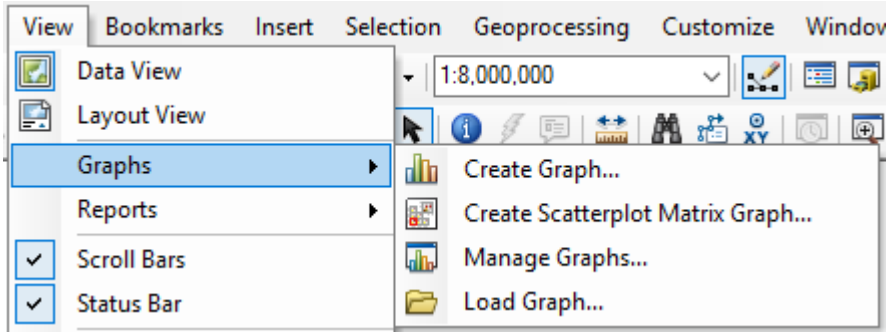
  2. Use an existing point file or the point file output from #1 above as the input point features in Spatial Analyst Tools > Extract > Extract Multi Values to Points – this tool allows for many rasters at a time! Then export the attribute table as text with filename.csv. This method is the raster equivalent of the overlay intersect done above for vector data.



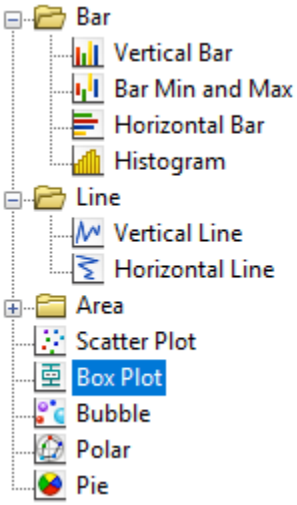
- Summary statistics: In ArcToolbox, click Analysis Tools > Statistics > Summary Statistics and specify **abcg** as the input. Select **ECOSYS** (or SurfGeol) as the case field, and select the following fields and statistics type:
  - MAP: MEAN
  - MAP: STD
  - DD0: MIN
  - DD0: MAX

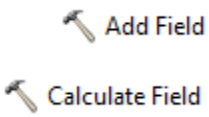
  Specify the output table as **ecosys_stats.dbf** (or surfgeol_stats.dbf) and click OK. Right-click the table name in the table of contents to open and view. Close when finished.
- Create scatterplots: In the main menu, click View > Graphs > Create Scatterplot Matrix Graph[9] and specify the following:
  - Layer/Table: **abclim**
  - Fields: MAT, MAP, DD0, FFP, PAS

  (NOTE: Clicking the Add All Fields buttons can slow the computer considerably.)
  - Show Histograms: check on and set the Number of bins: 5
  - Click Apply, then Next and then Finish



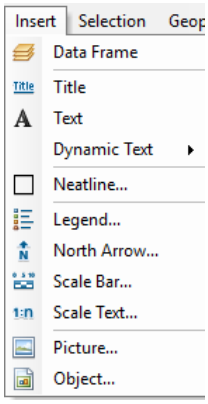
- Manage graphs in ArcGIS: Resize the panel by dragging on the corner. The Scatter Plot Matrix can be closed and to easily find it again click View > Graphs > Scatter Plot Matrix (or whatever title we give it). Right-click on the open graph's title bar for a context menu that will allow you to copy, export, modify properties, etc. Click on the scatterplot of interest to have it enlarged in the preview pane at top-right. See the ArcGIS Help link[9] below for more details on creating and interacting with scatterplots. Repeat the Create Scatterplot Matrix Graph for other combinations of variables. If you have time and interest, read the help links on graphs in ArcGIS[10] and box plots[11] and try some of them via Click View > Graphs > Create Graph.
- Generally, the data for the multivariate analyses below should be normally distributed for optimal results – the graphing tools available in ArcGIS have become quite helpful for determining whether the data are normal, bimodal, multi-modal, or severely skewed. If you decide to apply a transformation, this is easily done in the attribute table via the Add Field and Calculate Field tools in ArcToolbox's Data Management Tools > Fields toolset[12].

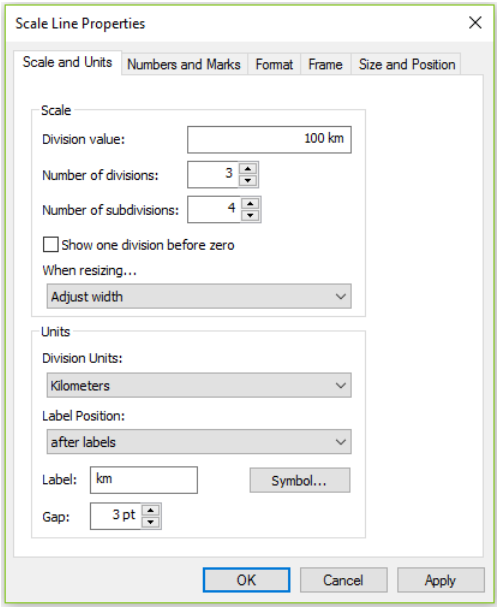
  (TIP: To transform data values in raster cells, use tools available in Spatial Analyst's Math toolset[13])

2. Cartographic Output

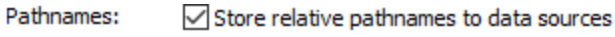
Inserting elements of cartography into a layout will turn the *picture of data* into a true map. R has some incredible graphing abilities, but ArcGIS is made for mapping.



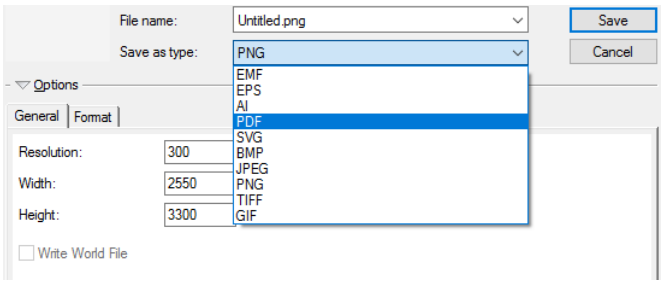
- Map layout: Click View > Layout View and if needed, click Customize > Toolbars to show the Layout toolbar. Use the main menu Insert pull-down list to add the cartographic elements: title, legend, north arrow, and scale bar.
  Wizards walk you through how to set each up. Right-click each element to access and modify the Properties – the Scale Line figure suggests properties for your Alberta scale bar.
  In Layout view you can move and resize the graphic elements similar to a power point slide.
  For more tips, please see the map elements help and related online documentation for creating maps[14].



- Save the map document: Choose File > Map Document Properties and check “Store relative pathnames to data sources” and click OK.



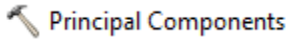
  Then choose File > Save As, navigate to the D:\temp2\gis directory and type a name; e.g. RenR690_todaysdate.mxd and click SAVE.
  (IMPORTANT TIP: Relative path names specify the location of the spatial data relative to the current location on disk of the map document (.mxd file) itself. Since relative paths don't contain drive names (e.g. D:\temp2), they make it easier for you to move the map and its associated data to any disk drive without the map having to be repaired. As long as the same directory structure is used at the new location (e.g. \gis), the map will still be able to find its data by traversing the relative paths.)
- Export map to graphic: While in the Layout View, choose File > Export Map. Specify the folder to save in, the type, resolution, any other properties, and the filename.
  (TIP: 300 dpi is generally good for Power Point or web, 300 dpi TIFF is usually typical for publications, 300 dpi PDF is handy for sharing by email.)



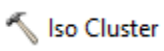
3. Spatial Analyst's Multivariate Statistics Toolset

The Spatial Analyst extension provides additional specialized tools for raster processing that are organized by categories or groups of related functionality, called toolsets. The statistical analysis[15] groups include cell (local functions), neighborhood (focal functions), zonal overlay (group by zones), and multivariate statistical analysis techniques to the modeling of your spatial data. Explore relationships between many different data layers or types of attributes using the Multivariate Statistics[16] tools, which include classification (supervised and unsupervised) and principal components analysis (PCA), to transform a multispectral image into a categorized landcover map or terrain stratification or habitat analysis. We will work with the rasters that we converted above.

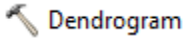
- PCA: Open ArcToolbox's Spatial Analyst Tools > Multivariate Tools > Principal Component Analysis[17] and specify the following:



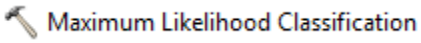
  - Input raster bands: select all you are interested in *(TIP: Same or more than components)*
  - Output multiband raster: D:\temp2\gis\rasters**pca_clim**
  - Number of principal components: 3
  - Output data file: D:\temp2\gis**pca_clim.TXT**
  - Click OK
- View the PCA outputs: The PCA output is a multiband raster, or stack of raster layers stored in a single file. The default symbolization shows the first three principal components (PC) as an RGB composite where the color mixing depends on which PC has higher values for each color. Click Add Data and add the individual rasters: pc_clim**c1**, pc_clim**c2**, and pc_clim**c3**. In the layer properties choose to symbolize the single bands as Stretched using any of the color ramps, and stretch methods that best convey the information. Or perhaps the classified renderer might help in understanding the PC. In the table of contents List By Source, right-click on the **pca_all** table and click Remove *(it's not really a table)*. In Windows Explorer, find the **pca_clim.txt** file and open it in Notepad or similar text editor, with word wrap turned off and a fairly small font size. Reformat as desired (e.g. place the raster file names above the numbers in the matrices), and save. Use what you learned in the PCA lab using R to help interpret the report. Optionally, repeat the PCA with a different set of variables.
- Clustering: In the previous lab using R we learned how to classify using cluster commands. We can also create a classification[18] of our data with Spatial Analyst by grouping raster cells into classes (i.e. a known category, such as forest or water) or clusters (i.e. a grouping of cells based on the statistics of their attributes). ArcGIS analyzes a sampling of cells (from the input rasters) that are representative of a class or cluster, and stores the statistics from it in a signature file to be used in classifying all cells among the input rasters. Here we will classify with clusters using the unsupervised multivariate classification available in ArcGIS, specifically the Iterative Self Organizing (ISO) method of performing clustering. Open ArcToolbox's Spatial Analyst > Multivariate Tools > Iso Cluster[19] and specify the following:
  - Input raster bands: **pas**, **ffp**, **dd5**, **map**, **mat** (or any combination you wish to investigate)



  - Output signature file: D:\temp2\gis**iso_5v.gsg**
  - Number of classes: 21
  - Number of iterations: 30
  - Click OK
- Evaluate with dendrogram: The signature file from Iso Cluster is the input for the Dendrogram tool, which then creates an ASCII text file that illustrates how the classes or clusters relate to one another using the multidimensional distance separating the classes in attribute space. Open ArcToolbox's Spatial Analyst > Multivariate Tools > Dendrogram[20] and specify the following:



  - Input signature file: D:\temp2\gis**iso_5v.gsg**
  - Output dendrogram file: D:\temp2\gis**dendrog_iso_5b.txt**
  - Line width of dendrogram: 100
  - Click OK
- Classify: If the dendrogram shows acceptable results then it is time to assign each cell in the study area to a cluster, resulting in a map that groups the study area into naturally occurring classes, corresponding to clusters (a.k.a. stratification). (NOTE: You may need to try different inputs rasters and numbers of classes, etc. if the dendrogram shows that the signature file statistics are not separated adequately, or alternatively merge the groupings by editing the signature file.) To go ahead with the signature file, as is, open ArcToolbox's > Spatial Analyst Tools > Multivariate Tools > Maximum Likelihood Classification[21] and specify the following:



  - Input rasters bands: **pas**, **ffp**, **dd0**, **map**, **mat** (must be same rasters as used in Iso Cluster)
  - Input signature file: D:\temp2\gis**iso_5v.gsg**
  - Output classified raster: D:\temp2\gis\rasters**class_5v**
  - Output confidence raster: D:\temp2\gis\rasters**conf_5v**
  - Click OK

A good read through of the ArcGIS Desktop online help files on the multivariate toolset will provide you with more details on what each tool needs and does, and occasionally there are further references. See the Online References section.

Online References

[1] www.esri.com/software/arcgis/arcgis-for-desktop

[2] www.r-project.org

[3] blogs.esri.com/esri/esri-insider/2015/07/20/building-a-bridge-to-the-r-community/

r-arcgis.github.io/

[4] www.python.org , www.scipy.org , http://spatial.uchicago.edu/software

http://proceedings.esri.com/library/userconf/proc15/tech-workshops/tw_295-280.pdf

http://proceedings.esri.com/library/userconf/proc16/tech-workshops/tw_1868-100.pdf

[5] http://desktop.arcgis.com/en/arcmap/latest/tools/conversion-toolbox/an-overview-of-the-to-raster-toolset.htm

http://desktop.arcgis.com/en/arcmap/latest/tools/conversion-toolbox/converting-features-to-raster-data.htm

[6] http://desktop.arcgis.com/en/arcmap/latest/manage-data/raster-and-images/cell-size-of-raster-data.htm

[7] http://desktop.arcgis.com/en/arcmap/latest/analyze/executing-tools/a-quick-tour-of-batch-processing.htm

[8] http://desktop.arcgis.com/en/arcmap/latest/manage-data/raster-and-images/drawing-thematic-raster-datasets-representing-continuous-data.htm

http://desktop.arcgis.com/en/arcmap/latest/manage-data/raster-and-images/renderers-used-to-display-raster-data.htm

[9] http://desktop.arcgis.com/en/arcmap/latest/map/graphs/scatter-plot-matrix-graphs.htm

[10] http://desktop.arcgis.com/en/arcmap/latest/map/graphs/exploring-and-visualizing-data-with-graphs.htm

[11] http://desktop.arcgis.com/en/arcmap/latest/map/graphs/creating-box-plot-graphs.htm

[12] http://desktop.arcgis.com/en/arcmap/latest/tools/data-management-toolbox/an-overview-of-the-fields-toolset.htm

[13] http://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/an-overview-of-the-math-tools.htm

[14] http://desktop.arcgis.com/en/arcmap/latest/map/page-layouts/what-is-a-page-layout.htm

http://desktop.arcgis.com/en/arcmap/latest/map/page-layouts/map-elements.htm

http://desktop.arcgis.com/en/arcmap/latest/map/map-export-and-print/exporting-your-map.htm

[15] http://www.esri.com/software/arcgis/extensions/spatialanalyst/key-features

[16] http://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/an-overview-of-the-multivariate-tools.htm

[17] http://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/how-principal-components-works.htm

[18] http://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/understanding-multivariate-classification.htm

[19] http://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/how-iso-cluster-works.htm

[20] http://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/how-dendrogram-works.htm

[21] http://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/how-maximum-likelihood-classification-works.htm

Extra References

A really useful book is 'Designing Better Maps: A Guide for GIS Users' by Esri Press: http://esripress.esri.com/display/index.cfm?fuseaction=display&websiteID=293&moduleID=0

The UofA has free, unlimited access to self-paced e-Learning on the Training website. Check out the catalog here: https://www.esri.com/training/catalog/search/ – search for 'map' or 'statistics'.

If you don't already have one, you will need to create a free account on https://my.esri.com/. This is different than the ArcGIS Online subscription for organizations account (that you may or may not have).

You will need to request permissions to join the University of Alberta organization (if you don't have it already) via https://my.esri.com/#/request-access. Once you have access, you will be able to complete any and all e-learning courses and verify your completion by printing a certificate.

Esri Canada Integrating R with ArcGIS webinar series:

Part 1 slide deck - http://slides.com/camplouffe/integrating-r-with-arcgis/

Part 1 video recording - https://www.youtube.com/watch?v=9lpPuCmVmxc

Part 2 slide deck - https://slides.com/camplouffe/integrating-r-with-arcgis-part-2/

Part 2 video recording - https://www.youtube.com/watch?v=TNnCgQ-jG2I

Scripts, project files, documentation, and data used during the webinars - https://github.com/cplouffe/r-arcgis-webinar

R sample tools repository - https://github.com/R-ArcGIS/r-sample-tools

Join the University of Alberta GIS list: http://www.mailman.srv.ualberta.ca/mailman/listinfo/gis