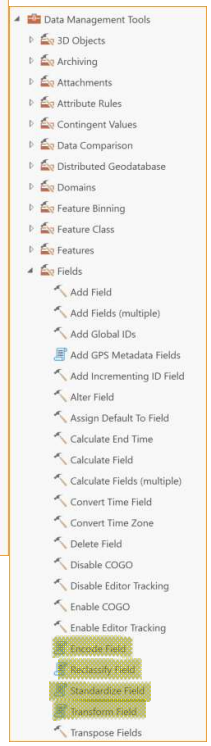
September 2: Doing data science with ArcGIS Pro

Monday, August 16, 2021 4:45 PM

ArcGIS Pro provides many tools for exploratory data analysis, visualization, machine learning, ArcPy python library, ArcGIS python API, and Notebook. Many of the tools you have learned from introductory and advanced GIS classes. We won't repeat them here, but on the right is the screen shot from all toolboxes from ArcGIS Pro for an overview.



Data Engineering: get your data ready for analysis. Data scientists spend most of their time on data engineering. Devil is in the detail. Be careful and patient.

- Missing data?
- Wrong data? Wrong data values? Wrong resolution?
- Biased?
- Need additional data?
- Transform data to a different projection, different measurement unit, different categories, different indices, different geometries, different matrix shape (rows, columns, pivot table, etc.)

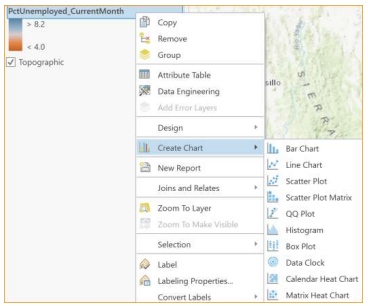
New ArcGIS Pro tools for data engineering: check out Data Management > Fields

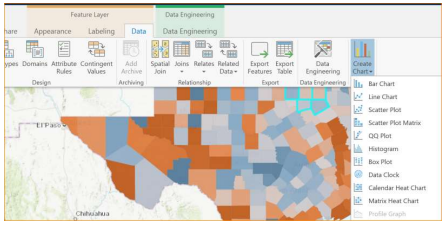
Visualization & Exploration: Explore data, Interpret analysis results, and communicate findings

- Mapping: change colors, classifications, and number of classes
- Plotting: statistical graphs and charts
- Profile plots
- 3D plots
- Animations
- Any outliers

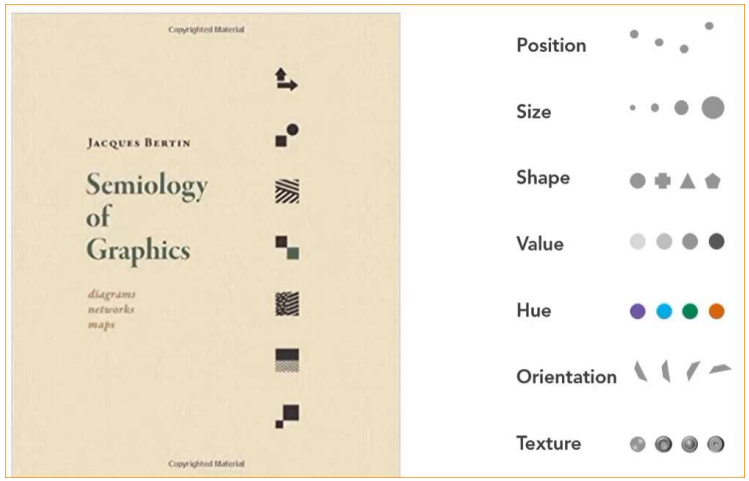
Visualizations to support spatial analysis:

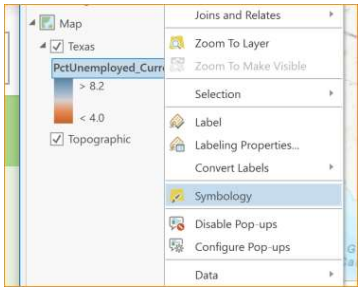
- Distribution and frequency
- Category comparison
- Relationships and correlations
- Change over distance or time

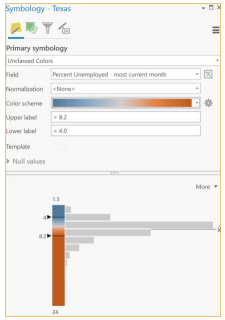




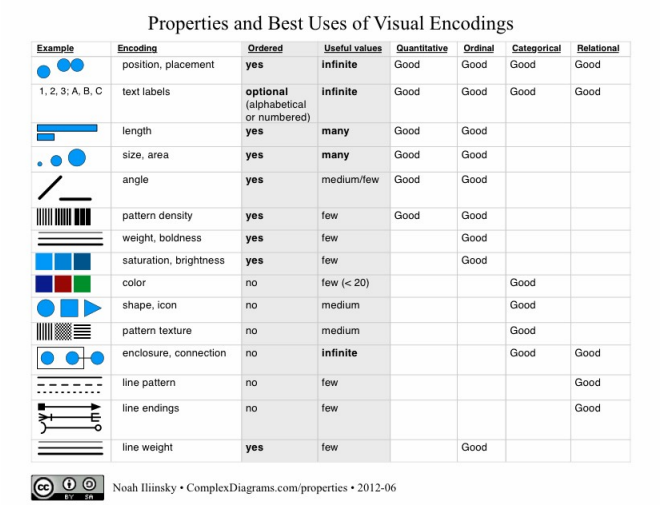
Visualization is in its simplest term to create graphical representation of data, commonly using visual variables. Bertin (1983) *Semiology of Graphics: Diagrams, Networks, Maps* is the classic for graphic representation, including geovisualization.







Noah Iliinsky has an excellent podcast on the topic: https://ellessmedia.com/csi/noah-iliinsky/ and he also summarizes the proper use of visual variables:

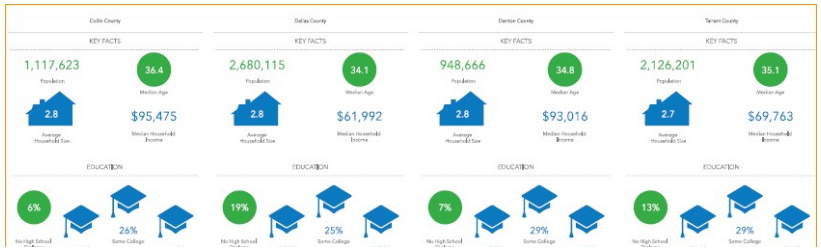


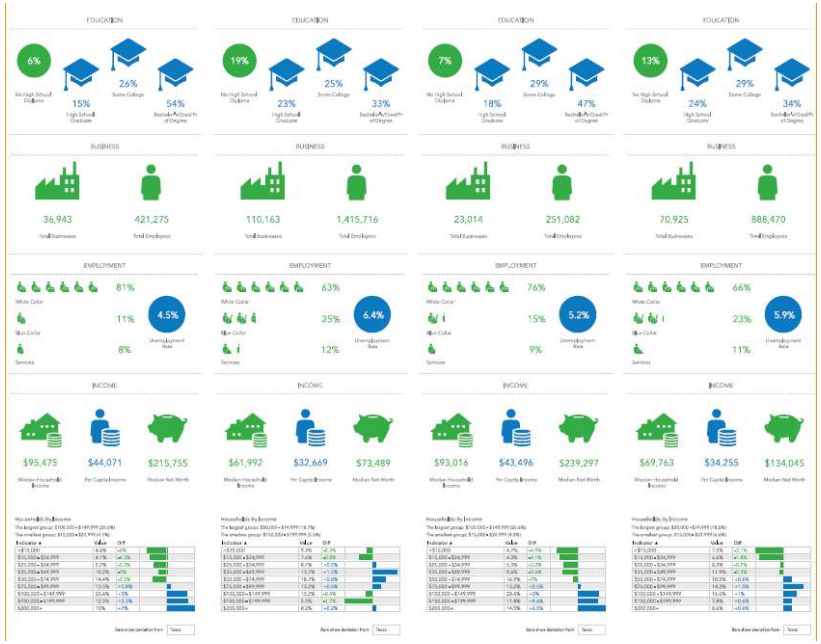
Properties and Best Uses of Visual Encodings

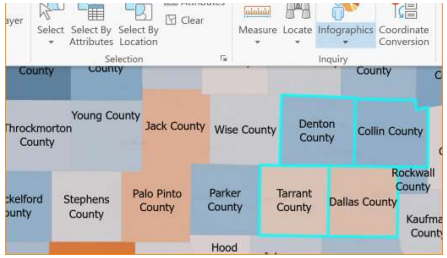
Example	Encoding	Ordered	Useful values	Quantitative	Ordinal	Categorical	Relational
	position, placement	**yes**	**infinite**	Good	Good	Good	Good
1, 2, 3; A, B, C	text labels	**optional** (alphabetical or numbered)	**infinite**	Good	Good	Good	Good
	length	**yes**	**many**	Good	Good		
	size, area	**yes**	**many**	Good	Good		
	angle	**yes**	medium/few	Good	Good		
	pattern density	**yes**	few	Good	Good		
	weight, boldness	**yes**	few		Good		
	saturation, brightness	**yes**	few		Good		
	color	no	few (< 20)			Good	
	shape, icon	no	medium			Good	
	pattern texture	no	medium			Good	
	enclosure, connection	no	**infinite**			Good	Good
	line pattern	no	few				Good
	line endings	no	few				Good
	line weight	**yes**	few		Good		

Noah Iliinsky • ComplexDiagrams.com/properties • 2012-06

InfoGraphics combines statistical charts and icons in a table to summarize data for comparison.

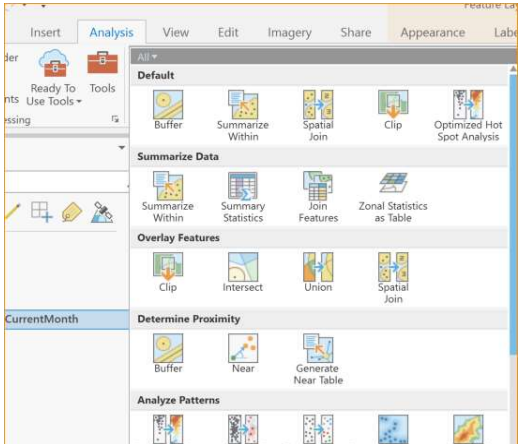


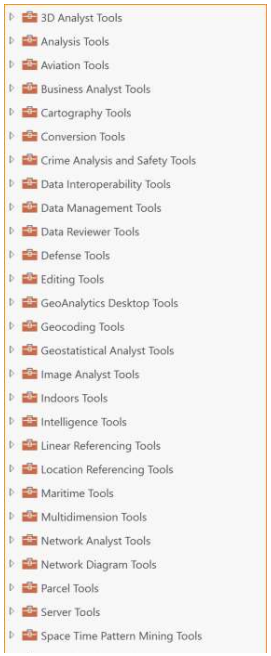


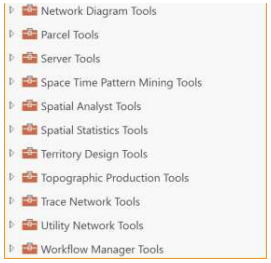


Spatial Analysis and Machine Learning & AI

- Spatial statistics-based methods
- Learning-based methods
- Find spatial patterns, spatial correlation, or spatial relationships
- Spatial operations: spatial join, buffer, overlays or spatial binning
- Classification and regression
- Modeling and Prediction





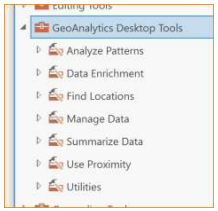


For deep learning tools that are used to identify objects in images, we need to install deep learning frameworks for ArcGIS. Instructions and the libraries installer are available at https://pro.arcgis.com/en/pro-app/latest/help/analysis/deep-learning/install-deep-learning-frameworks.htm

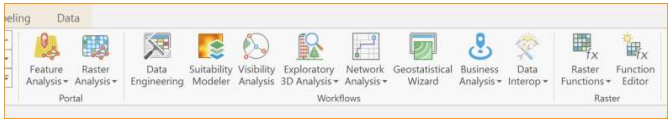
Big Data Analytics

Since the tools deal with big data, which by definition does not fit the memories of regular computers. Therefore, the tools are installed with ArcGIS GeoAnalytics Server and the computation is done over the server with parallel processing on more than one machine.

A reduced-scale reversion is GeoAnalytics Desktop Toolbox that uses Apache Spark to handle parallel processing on a desktop computer with multiple CPU cores. GeoAnalytics Desktop Toolbox has similar tools in Analysis, Spatial Analyst, and Spatial Statistics toolboxes. The key difference is that GeoAnalytics tools use parallel algorithms so the tools can take the full advantage of computing power on a machine.

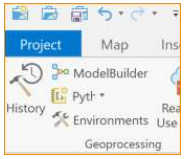


Modeling and prediction



Modeling and Scripting

The key tools are Model Builder, ArcPy python Library, ArcGIS R-bridge, ArcGIS Notebook, and ArcGIS API.



- Model builder provides a graphic interface to connect tools, input data, output data, and parameters to construct a workflow.
- ArcPy python library and ArcGIS R-bridge allow users to write python scripts or R scripts to use ArcGIS tools.
- ArcGIS Notebook is included in the desktop and at ArcGIS online to facilitate scripting.

- ArcGIS API is for writing python scripts with ArcGIS online.

Sharing and Collaboration
A suite of tools for sharing and collaborating data, maps, and tools on ArcGIS online or ArcGIS web portals.