GEOG 204

LECTURE 4

1

Data Collection

- A GIS can contain a wide variety of geographic data types originating from many diverse sources
  - It is an important requirement for a GIS to integrate data from many forms of data from a diversity of sources
- Data collection is time consuming and expensive
  - In some cases costs are estimated to be 85% of the cost of a GIS (Longley at al)

2

2

Data Collection Classification

- Data collection can be classified by source
  - Primary Sources
    - captured by direct measurement specifically for use in GI systems
    - both raster and vector data can come from primary sources
  - Secondary Sources
    - reused from earlier studies or obtained from other systems
    - raster and vector data are created from maps, photographs, and other hardcopy documents

3

3

Data Collection Classification

	Raster	Vector
Primary	Digital satellite remote-sensing images	GPS measurements
	Digital aerial photographs	Field survey measurements
Secondary	Scanned maps or photographs	Topographic maps
	Digital elevation models from topographic map contours	Toponymy (place-name) databases

Longley, Goodchild, et al (2015) Geographic Information Science and Systems. John Wiley and Sons

4

4

Primary Data Collection

- Raster data
  - Data are collected by remote sensing
    - Remote sensing is the measurement of physical, chemical, and biological properties of objects without direct contact
  - Information is derived from measurements of the amount of electromagnetic radiation reflected, emitted, or scattered from objects.
    - Passive sensors rely on reflected solar radiation or emitted terrestrial radiation
    - active sensors (such as synthetic aperture radar) generate their own source of electromagnetic radiation
    - Sensors are mounted on earth-orbiting satellites or other airborne platforms

5

5

Primary Data Collection

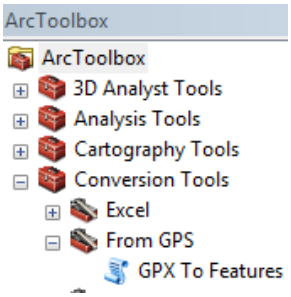
- Vector data
  - Data are captured by ground surveying, GPS and LiDAR
  - Ground surveying is based on the principle that the location of any point can be determined by measuring angles and distances from other known points.
    - It is highly accurate but time consuming and expensive
  - The GPS consists of a system of 24 satellites each orbiting the Earth every12 hours and transmitting radio pulses at precisely timed intervals
    - A receiver on the ground must make exact calculations from the signals, the known positions of the satellites, and the velocity of light in order to determine its position
    - GPS was developed by the US. Russia has GLONASS; China has BEIDU; Europe has GALILEO

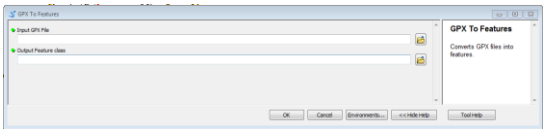
6

6

GPX File Conversion

- GPX is a standard GPS file format
- QGIS reads it directly
- ArcGIS has one intermediate step
  - Conversion tools>From GPS>GPX to Features
  - Input from GPS, output to shapefile





7

Primary Data Collection

- Vector data
  - Data are captured by ground surveying, GPS and LiDAR
  - LiDAR (light detection and ranging) employs a scanning laser range finder to collected accurate data
    - A LiDAR scanner is an active remote sensing instrument
    - It transmits electromagnetic radiation and measures the radiation that is scattered back to a receiver after interacting with the objects on the surface
    - The data collected from a LiDAR scanner is often referred to as a point cloud - a massive collection of independent points with (x, y, z)

8

8

Secondary Data Collection

- Raster Data Capture
  - Scanners
    - A scanner is a device that converts hardcopy media into digital images
    - Documents, such as building plans, CAD drawings, property deeds, and equipment photographs are scanned to reduce wear and tear, to improve access, to provide integrated database storage, and to index them geographically (e.g., building plans can be attached to building objects in geographic space).
    - Film and paper maps, aerial photographs, and images are scanned and georeferenced so that they provide geographic context for vector data layers
    - Maps, aerial photographs, and images are scanned prior to vectorization and sometimes as a prelude to spatial analysis

9

9

Scanners and Cameras

- High resolution raster

Contex

http://www.library.unt.edu/digital-projects-unit/scanners-and-scanning-systems

10

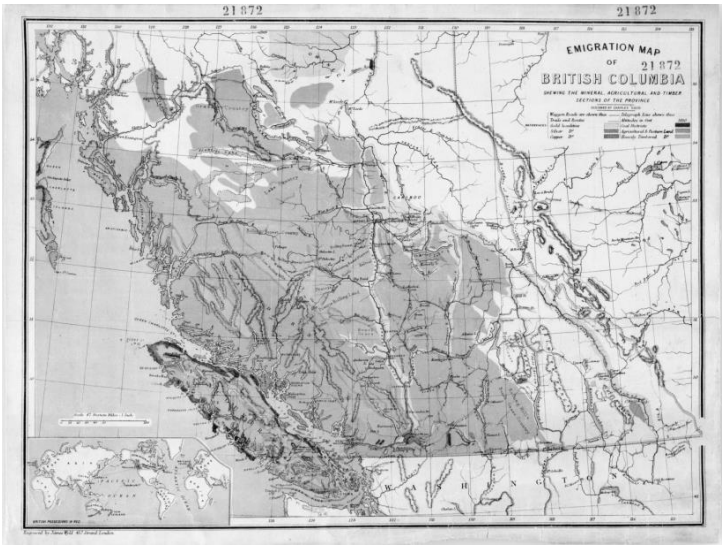
Secondary Data Collection

- Vector Data Capture
  - The digitization of vector objects from maps and other geographic data sources by heads-up digitizing and vectorization, photogrammetry, and COGO data entry
  - Heads-up digitizing and vectorization
    - creates vectors selectively from raster data
    - digitize vector objects manually straight off a computer screen using a mouse or digitizing cursor.
      - **heads-up digitizing** because the map is vertical and can be viewed without bending the head down.
    - Used to collect data for land parcels, buildings, and utility assets....

11

11

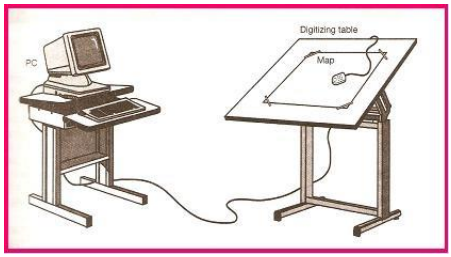
Vector Data from Historic datasets and Maps



12

Vector Data from Historic datasets and Maps

- Digitizing centuries of hand-drawn maps…
- Guess who got to do this job??
  - Prisoners
  - GIS Techs
  - Students!
- Tedious and Painstaking



13

Digitizing

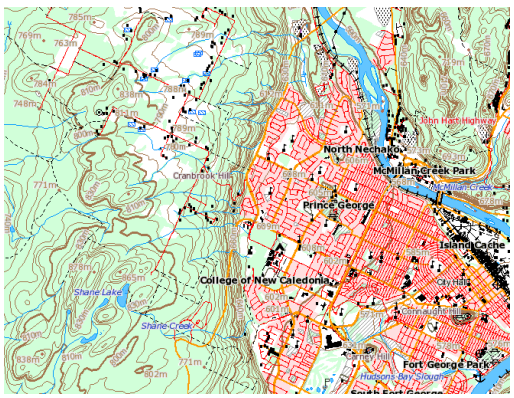
Digitizing is done in two ways:

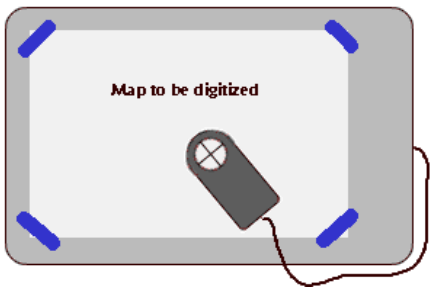
Tracing lines on maps initially using a tablet with map taped down,

or

onscreen / 'heads-up' (= copying a map) after 1995

CONNECT THE DOTS

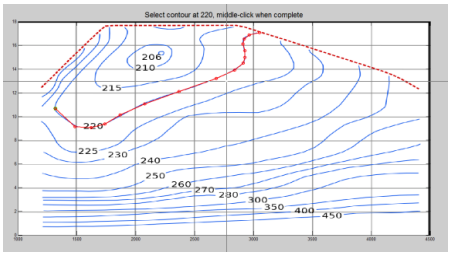




14

Digitizing Procedure

- Lines = connected points
  - Manual point selection
  - Timed point selection
  - Interval point selection

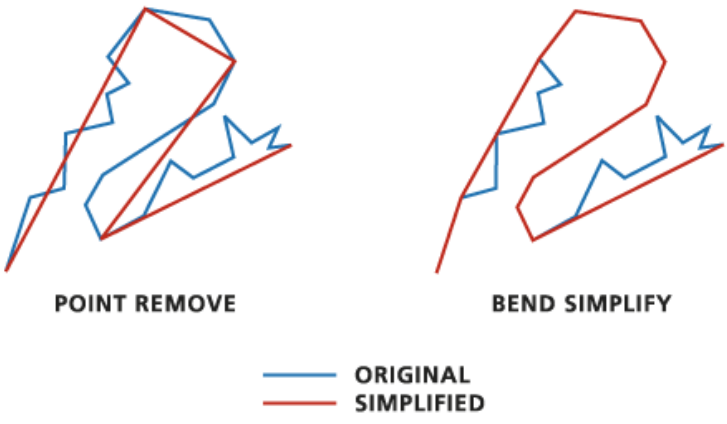


http://forum.imagej.net/t/digitizing-contour-map/118

15

Simplifying Lines

- Each vertex has a storage cost
- How much is enough? Too many?
- If too many, simplify in post process
  - Point remove: maintain essential shape
  - Bend simplify: maintain "important" bends

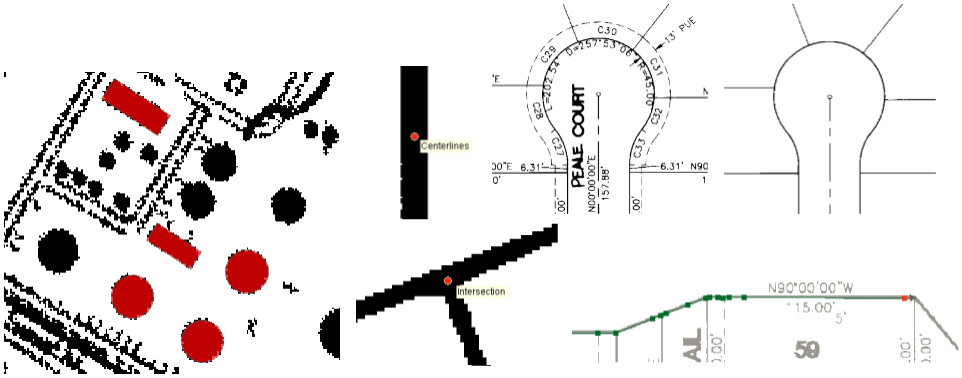


http://pro.arcgis.com/en/pro-app/tool-reference/cartography/how-simplify-line-works.htm

16

Automatic Feature Recognition

- ArcScan toolbar in ArcMap
  - Automated, semi-automated, or manual modes
  - Scanned image must be georeferenced
  - Toolbar: 



17

Digitising – editing is still needed: updates and errors

Coordinate locations are based on underlying georeferencing e.g. NAD 1983

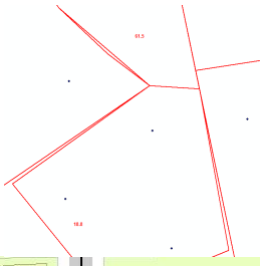
Edits: e.g. adding new features, modifying existing features, creating a new layer

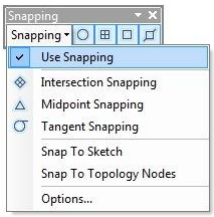
ArcEdit: https://www.youtube.com/watch?v=6dY3x-5qX6U

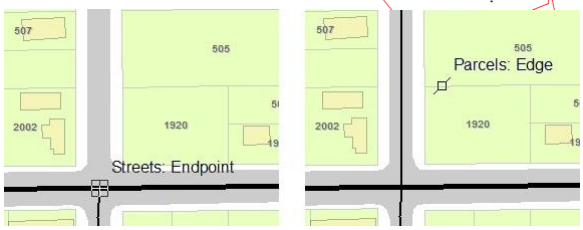
18

Snapping

- Automatic connection to other features
  - Any features, selected features, feature class
  - Same feature class (roads)
- Prevents slivers and disconnects
- User-defined tolerance, radius…
- https://blogs.esri.com/esri/arcgis/2010/09/20/using-snapping-effectively-in-arcgis-10/



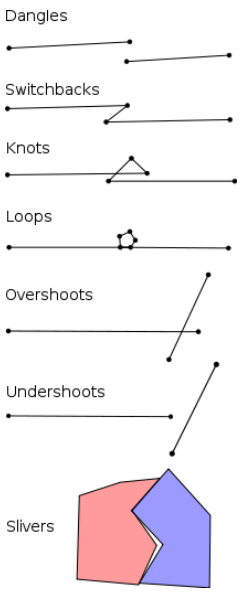




19

Digitizing errors

- Common errors
  - Dangles
  - Switchbacks
  - Knots
  - Loops
  - Overshoots
  - Undershoots
  - Slivers



Source: Caitlin Dempsey, GIS Lounge

20

20

Sources of Error

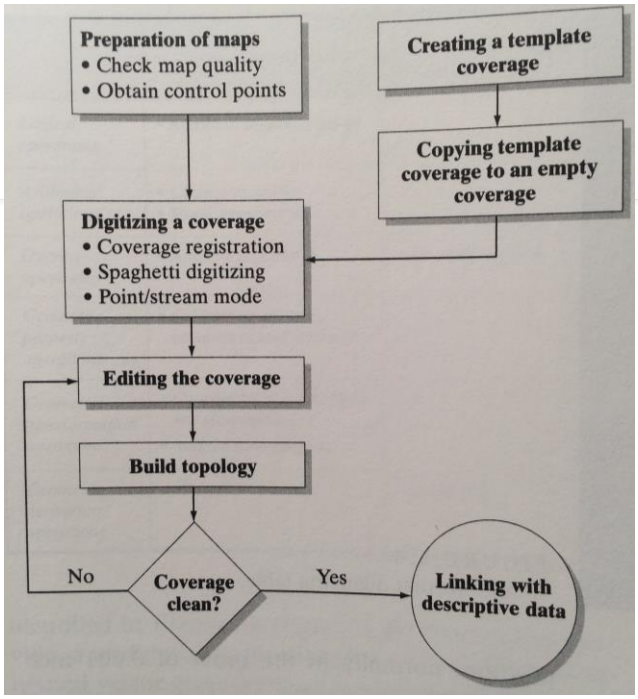
- Precision:
  - If points +/- 25m on creation
    - Similarly +/- 25m error introduced on digitization
    - Conceivably 50m total error
- Accuracy:
  - Paper may have shrunk, stretched or torn
  - Symbols rearranged to prevent overlap
  - Map sheet boundaries
  - Human boredom, fatigue, humor or malice

21

Input Error

- Very susceptible to errors
  - Does not cause error messages in digitization process
  - Outlier analysis sometimes catches mistakes
  - Easily goes unnoticed until publication

22



Coverage: A layer that can contain points, lines and polygons

23

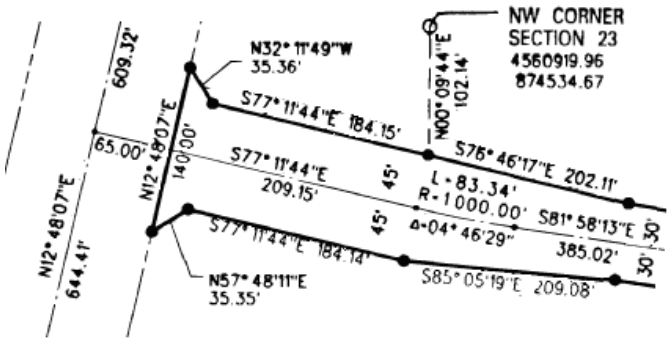
Secondary Data Collection

- Vector Data Capture
  - Photogrammetry
    - Measurements are taken from pictures, aerial photographs, and images
    - Measurements are captured from overlapping pairs of images using stereoplotters.
  - COGO and Other Data Entry
    - COGO is short for coordinate geometry and a method for data entry
    - Uses bearings and distances to define each part of an object
    - The COGO system is widely used in North America to represent land records and property parcels

24

Secondary Data Collection

- COGO descriptions for a road centerline and parcel boundaries adjoining the road



Source: ESRI

25

File Conversion

- FME Universal Translator
  - GIS Lab has a license
- ArcMap File formats
  - Read-only
  - Read + Write
- Raster: http://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/supported-raster-dataset-file-formats.htm
- Vector: http://desktop.arcgis.com/en/arcmap/10.3/manage-data/datatypes/about-geographic-data-formats.htm
- QGIS: https://docs.qgis.org/2.2/en/docs/user_manual/working_with_vector/supported_data.html

26

Keyboard Data Entry

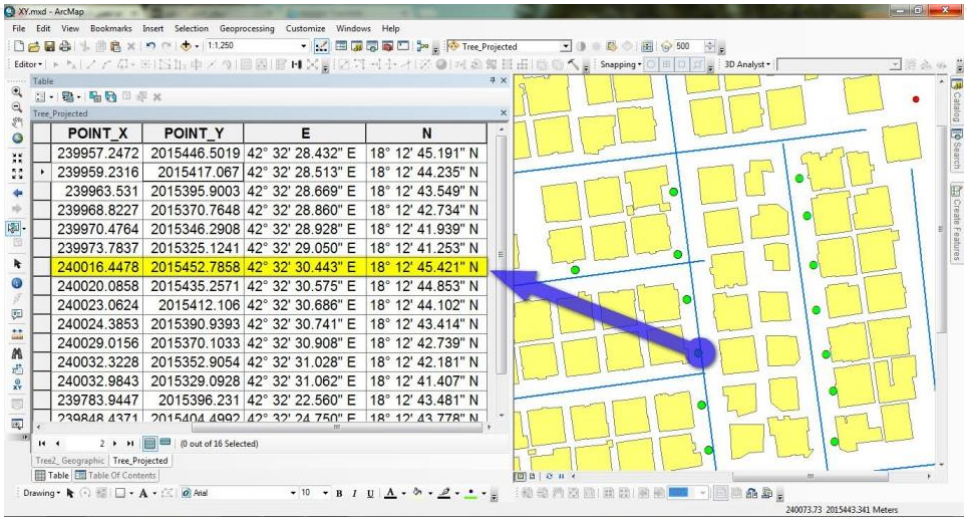
POLYGON	ESA_1	SPC1	PCT1	SPC2	PCT2	AGE_CL	HT_CL_IN	SITE_IDX	CRNCL_CL	SitePrep	Dist	YearDist	Regen	STTEND
67		HW	40	S	40	2	1	16.6	8	B	R	1985	1999	F
133			0		0	0	0	0	0			0	0	
199		HM	40	HW	30	9	3	7.2	5		L	1980	0	
353		HW	90	BA	10	9	4	11.6	1	B	L	1980	1999	F
229		HW	70	HM	20	9	3	9.5	5	B	L	1980	1999	F
264		HM	50	HW	30	9	3	7.5	5	H	L	1980	1999	F
162			0		0	0	0	0	0			0	0	
393		HW	60	HM	20	9	3	8.5	5	H	L	1980	1999	R
165		HM	80	BL	20	9	3	7	4	H	L	1980	1999	R

- **Sp**
  - Spatial reference: GPS
    - GPS point # = row ID # (“Join” function)
  - Row ID links spatial and tabular data
    - Field data usually entered in Excel or similar
    - GPS data straight to Arc

27

Keyboard Data Entry 2

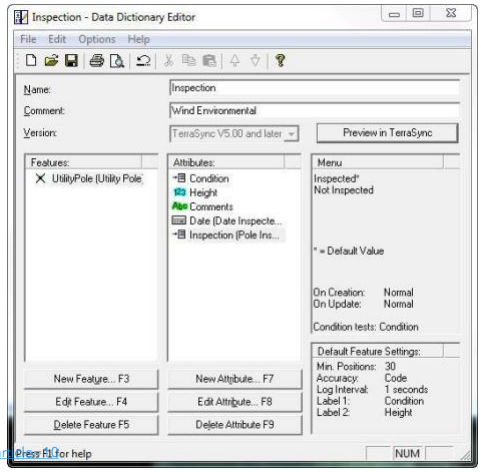
- Coordinates + all data in spreadsheet
- Geographic data, no projection (unless…)



28

Data Dictionaries (1)

- Trimble and other survey/map grade GPS
- Establish database design first
  - Features may be culverts, bridges, signs, poles…
  - In this case, poles:

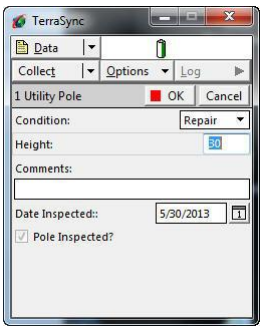


http://support.windenvironmental.com/knowledgebase.php?article=10

29

Data Dictionaries (2)

- Populate attributes while collecting points
  - Takes ~90 seconds to average enough points
  - Data entry taken care of at no additional time cost
- Attribute table ready to use
- Requires proper prior planning



http://support.windenvironmental.com/knowledgebase.php?article=10

30

Data Acquisition

- Data BC
- ESRI Open Data
- GeoDiscover Alberta
- BC MoFLNRO
- Google
- Openstreet Map Data
- Municipal Open Data Portals
- Spatially referenced 99% of the time
- Most data publicly available (no $$)

31

Canadian Data Sources

- Canada: https://www.mcgill.ca/library/find/maps/geospatial-online
- BC: https://catalogue.data.gov.bc.ca/dataset
- Alberta: https://geodiscover.alberta.ca/geoportal/catalog/main/home.page
- Saskatchewan: https://www.isc.ca/Pages/Content%20Gallery/GeoSask.aspx
- Manitoba: http://www.manitoba.ca/iem/geo/gis/index.html
- Ontario: https://www.ontario.ca/page/land-information-ontario
- New Brunswick: http://www.snb.ca/geonb1/e/DC/catalogue-E.asp
- Nova Scotia: https://geonova.novascotia.ca/
- Quebec: https://www.mcgill.ca/library/find/maps (Gov't data in a non-ESRI format)
- Yukon: http://www.geomaticsyukon.ca/
- Northwest Territories: http://www.geomatics.gov.nt.ca/dldsoptions.aspx (must register)
- Nunavut: http://ntilands.tunngavik.com/maps/

32

U.S. and Other Data Sources

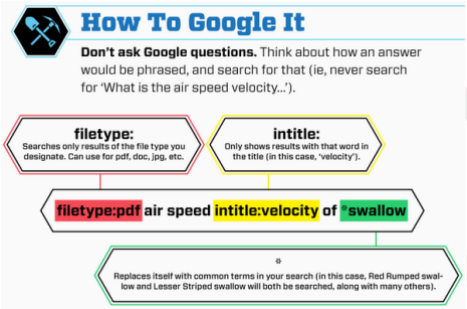
- NASA: http://data.giss.nasa.gov/
- Census Data
- US Geological Society

- Wikipedia's List of GIS Data Sources

33

Searching the Web: Keywords

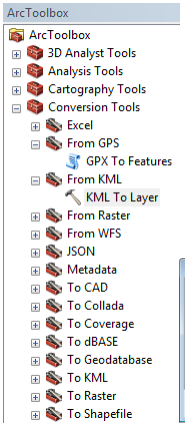
- Google tips and tricks
- Shapefile British Columbia Wildfire



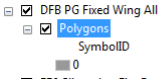
34

Publicly Available KMLs

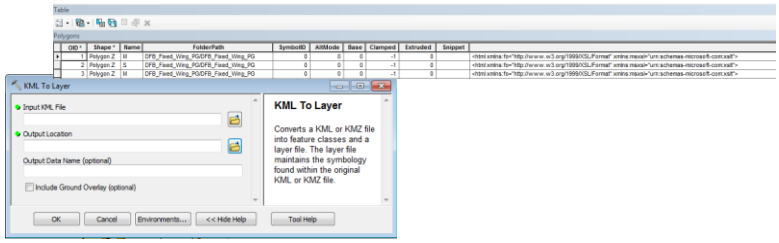
- Sometimes Google Earth is all you get
- It's enough…



- It comes in like this:



- Right click Polygons>File>Export data>Export as shapefile
- Result: usable shapefile/attribute table (turn off extra items)



35

Shapefiles

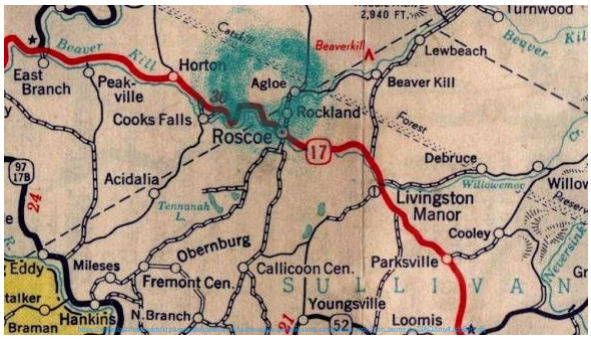
- One "shapefile" = 3 or more files
- **.shp**: shape format/geometry
- **.shx**: shape index format (file navigation)
- **.dbf**: attribute data (the 'spreadsheet')
- **.prj**: projection data
- .sbn, .sbx, .fbn, .fbx, .ain, .aih, .ixs, .mxs, .atx, .shp.xml, .cpg, .qix: other formatting files

- **ALL HAVE TO MOVE TOGETHER**

36

Paper Towns: False Input

- Agloe, New York
  - Copyright “trap”
    - Agloe General Store later built at location
- Beatosu (Beat OSU) and Goblu (Go Blue)
  - Also copyright trap



37