From Paper Map to Geospatial Vector Layer

Peter Peller
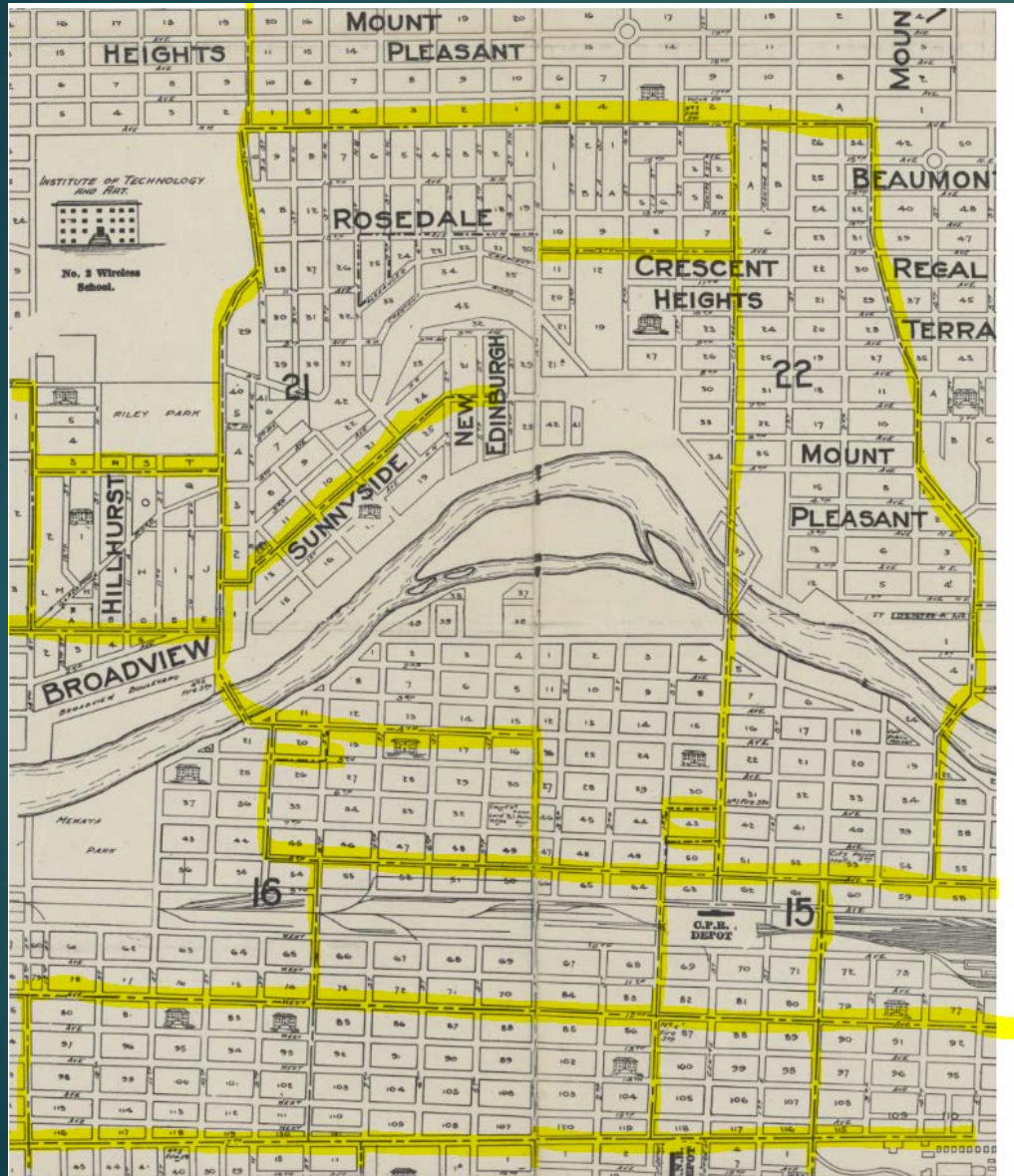
University of Calgary

May 30, 2018
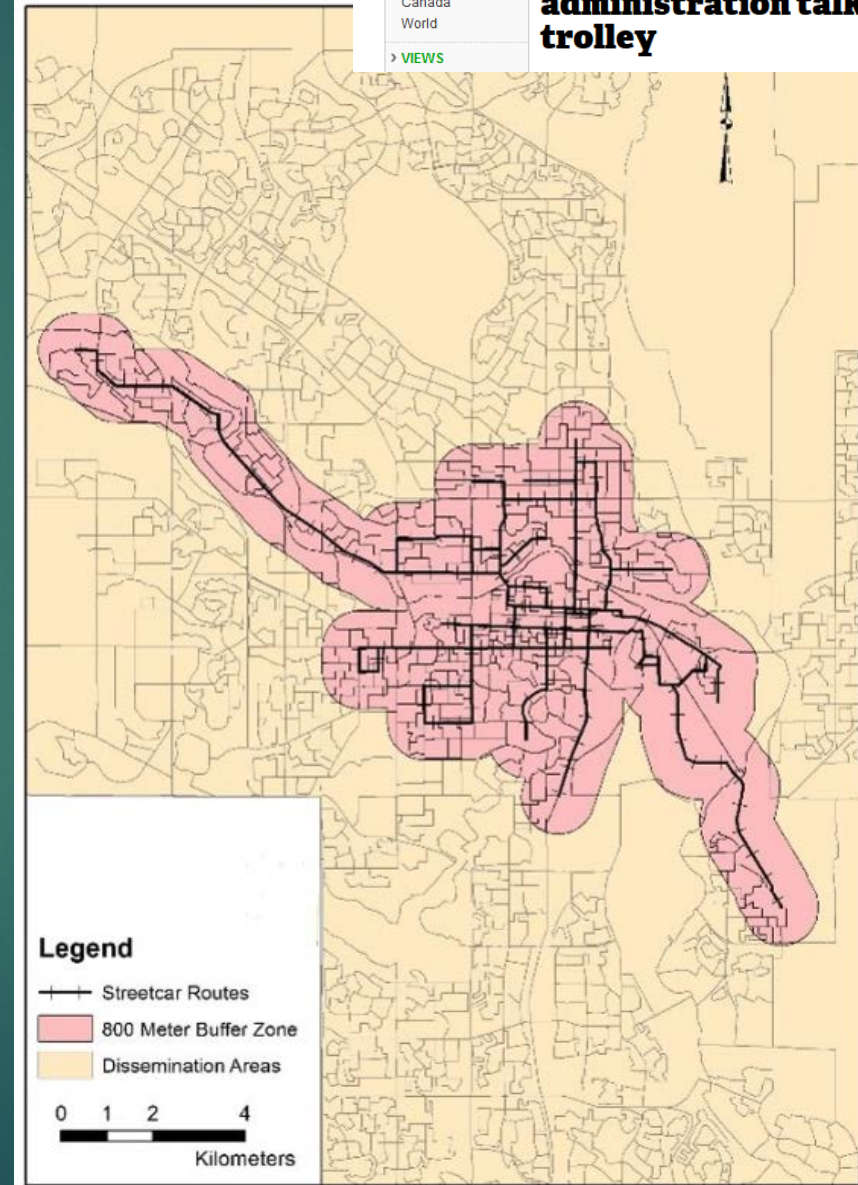
WHY?

Calgary
metro
HOME
News / Calgary
NEWS
Calgary
Canada
World
VIEWS
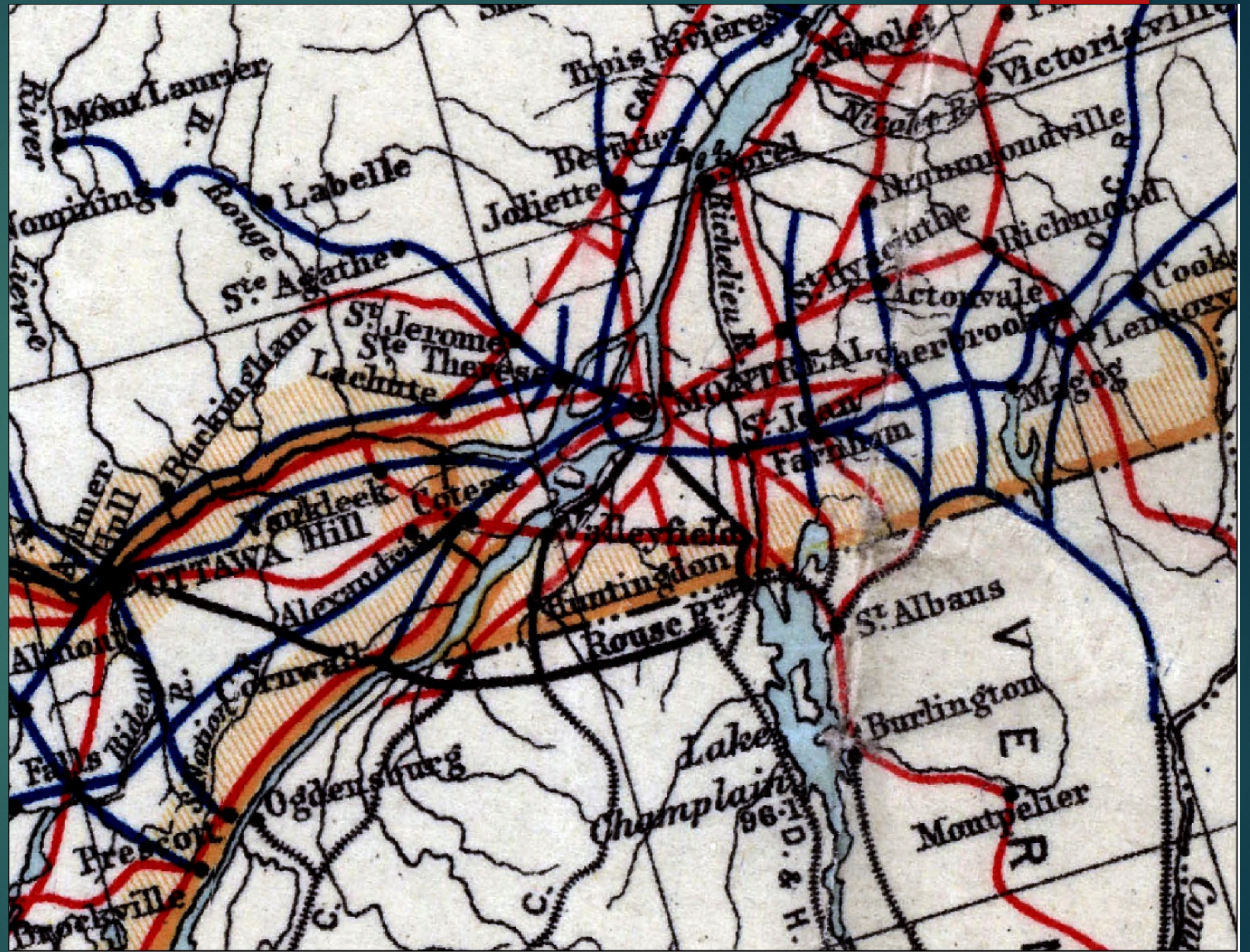
Street car revival: Calgary administration talks bringing back trolley

Legend
+—+ Streetcar Routes
800 Meter Buffer Zone
Dissemination Areas

0 1 2 4
Kilometers

Definitions

- Raster – data structure is a matrix of cells (or pixels) where each cell contains a value

- Vector – data is represented as points, lines and polygons with each feature containing information

- Scan – create a raster image (ex/ scanned map)

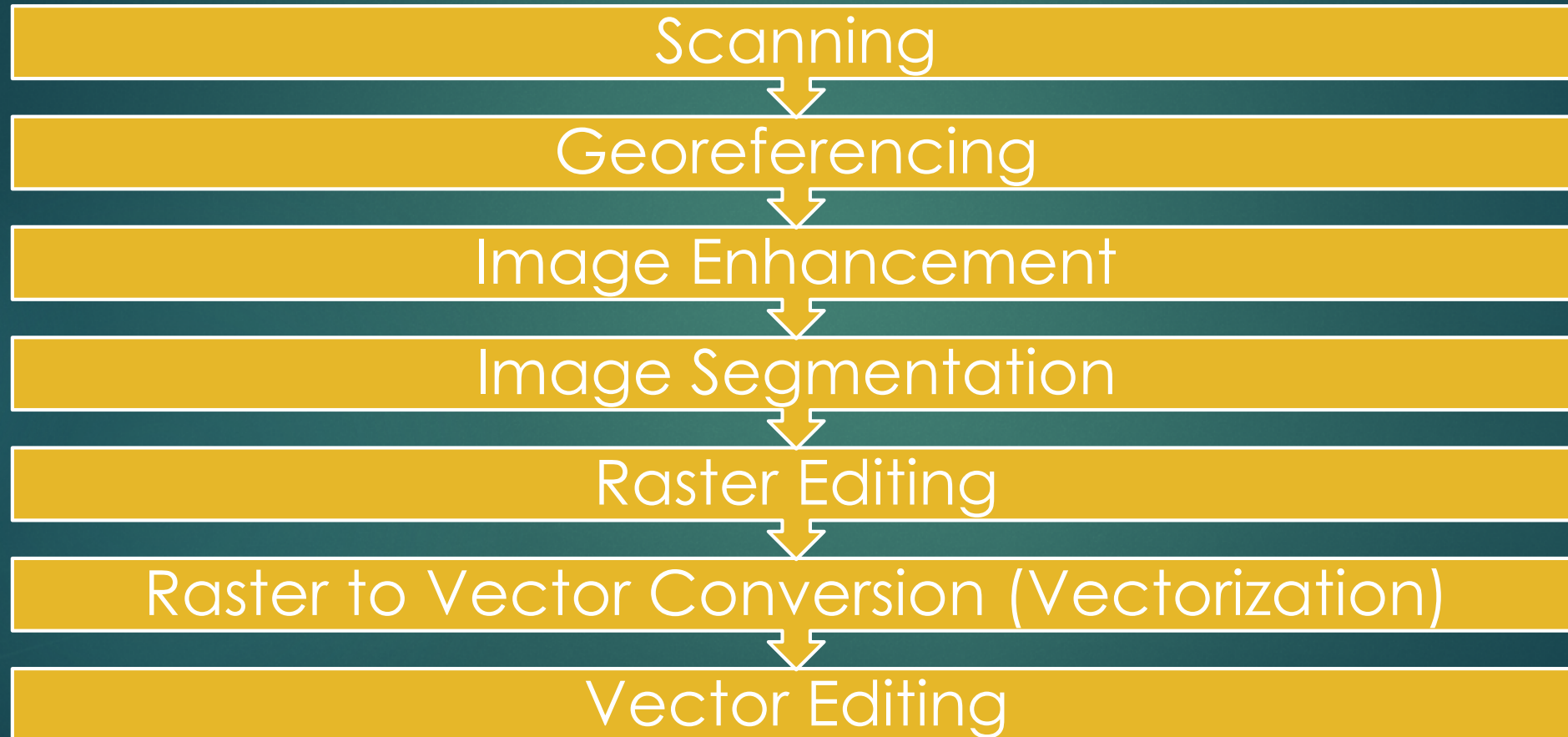- Vectorize – convert raster image pixels to a vector feature class (or layer)

Study Methodology

- Multiple databases searched

- 70 studies identified and examined

- 11 specific cases selected based on the following criteria:
  - Software available openly or commercially
  - Semi-automated process
  - Scalable to multiple maps and map types
  - Appropriate for a library and archive setting

The Process

Scanning

Georeferencing

Image Enhancement

Image Segmentation

Raster Editing

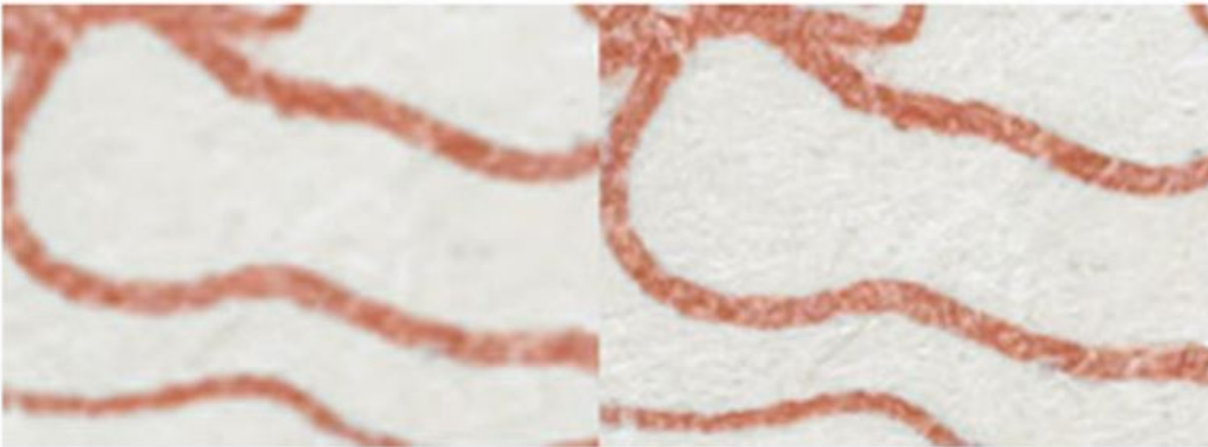Raster to Vector Conversion (Vectorization)

Vector Editing

Scanning

- Poor original maps – try to borrow it from someone else
- Higher resolutions are not always better: 300-400 dpi sufficient
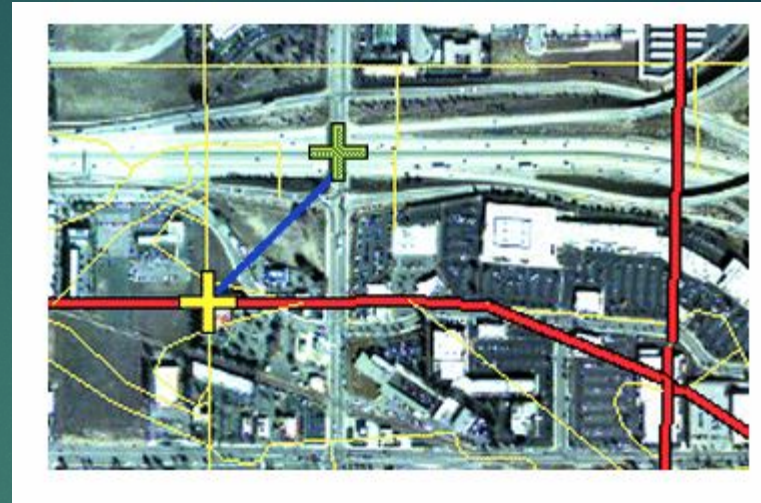- 8-bit colour for grayscale maps; 24 bit colour for colour maps
- Output file format: TIFF

FIGURE 10 Texture as noise in 300PPI (left) and 400PPI (right) variations.

Source: Pearson et al, 2013, 327.

Georeferencing

▶ Don't overdo the number of control points
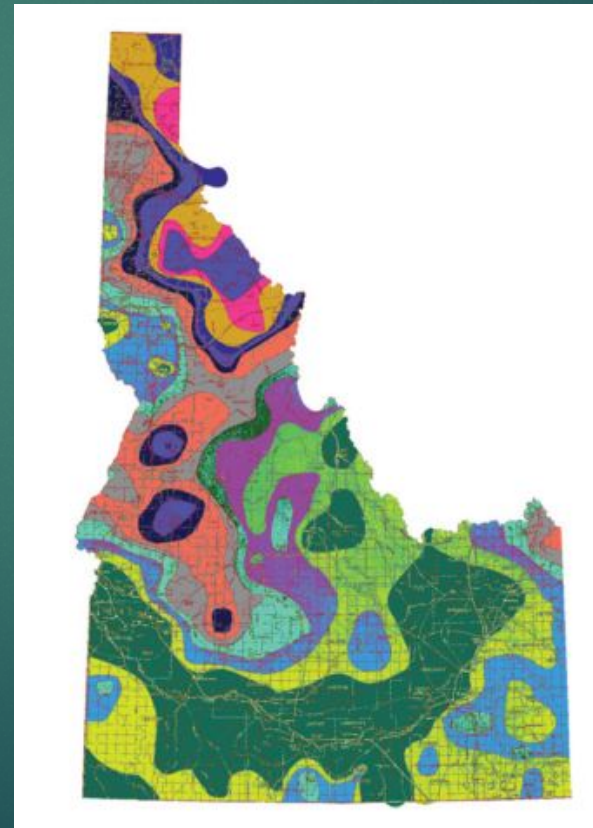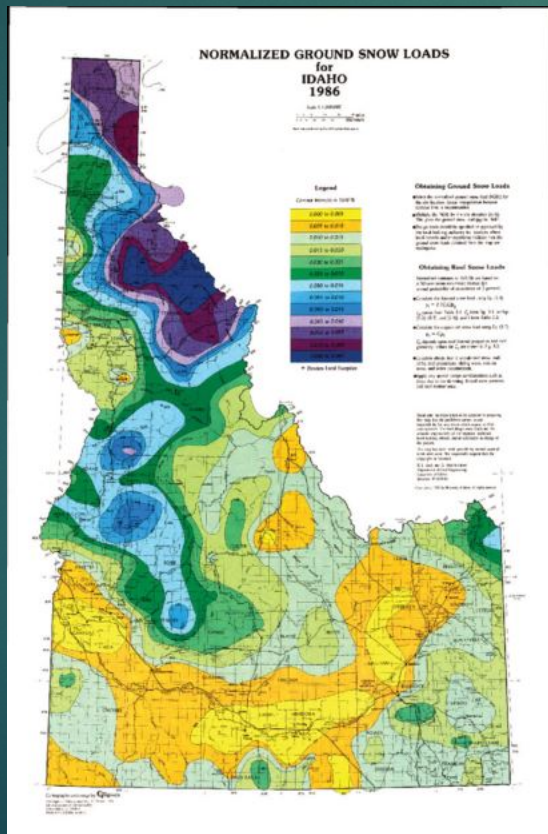
▶ Root Mean Square of all the residual errors ≤ cell size of the raster

▶ First Order Polynomial transformation usually sufficient in majority of cases for georectification

▶ Possibly employ crowdsourcing (ex/NYPL: MapWarper)

Image Enhancement

- Resample image from archival quality scanned map
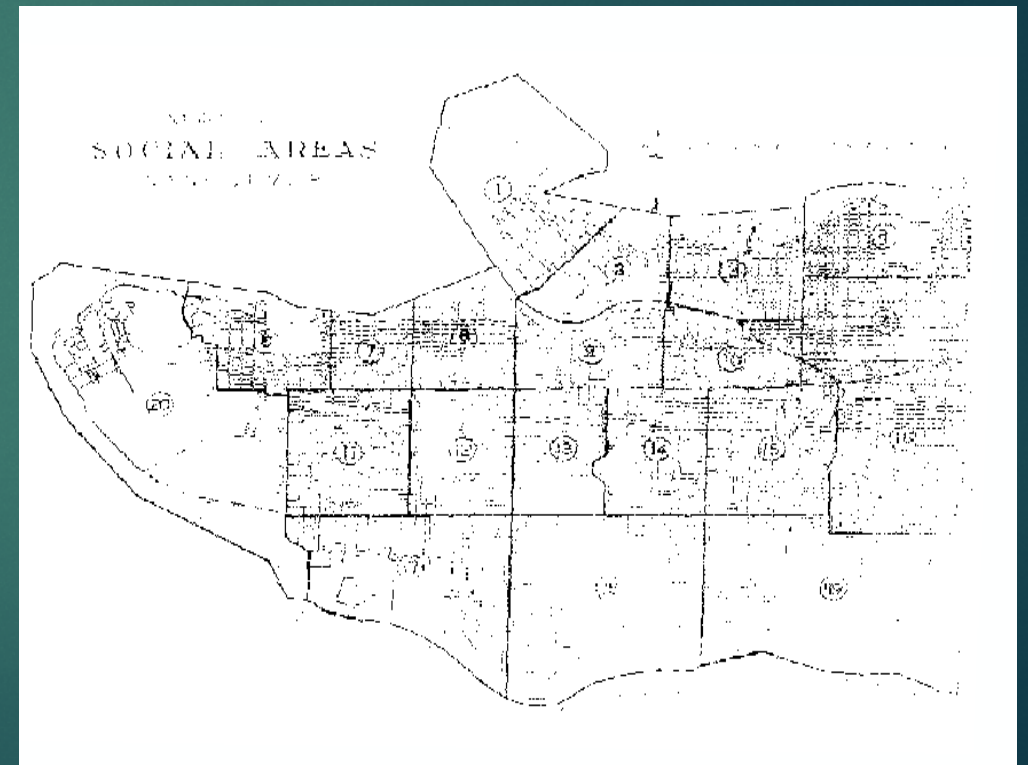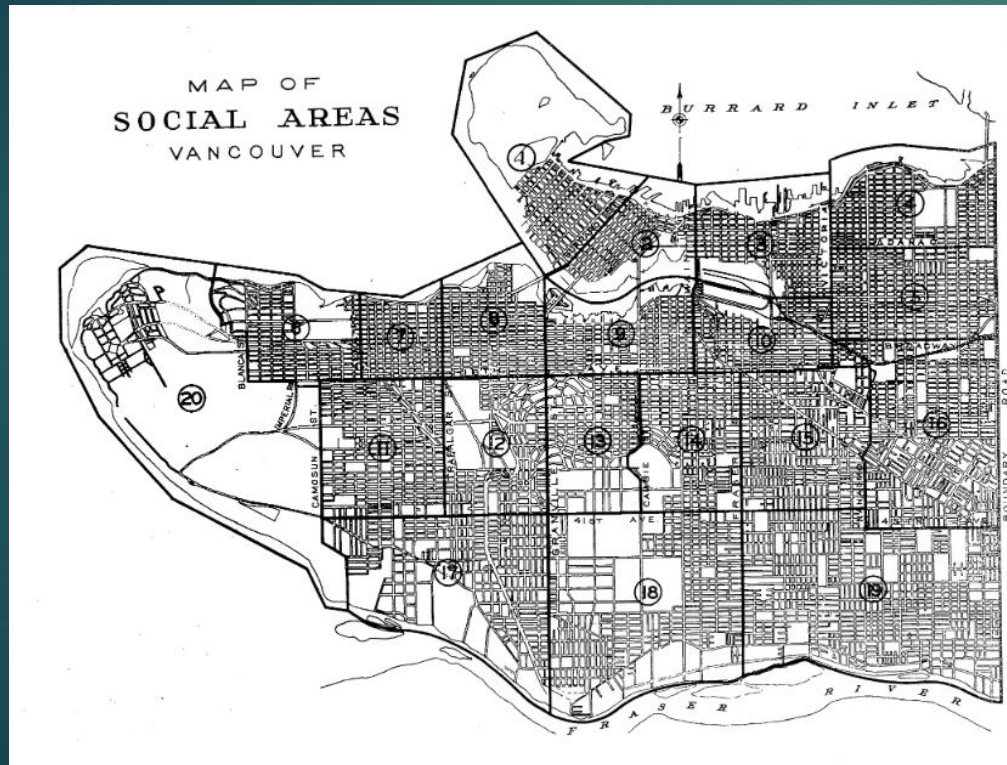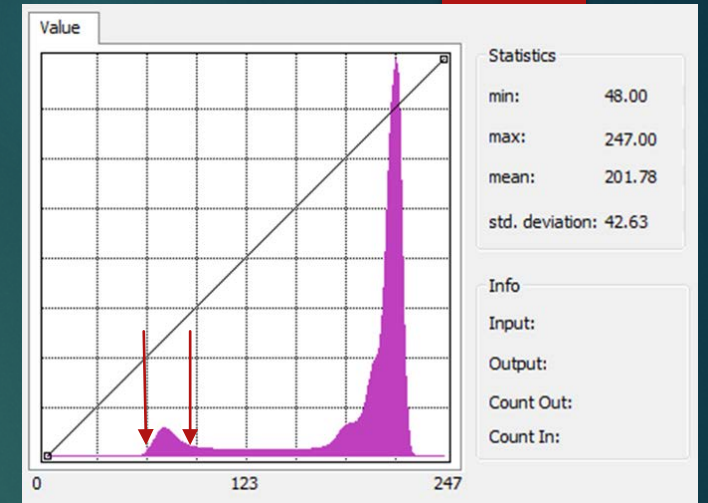- Use a "mask" to exclude extraneous parts of the map

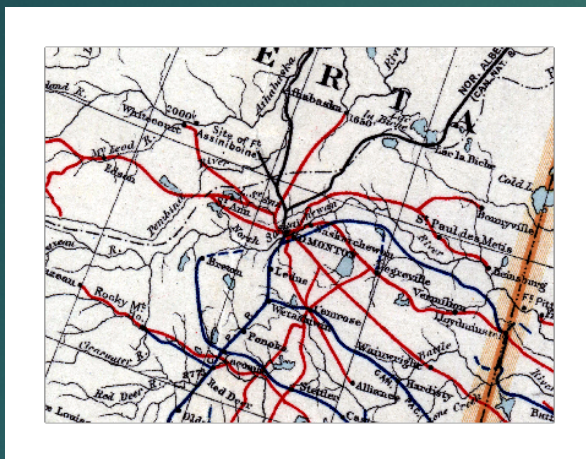

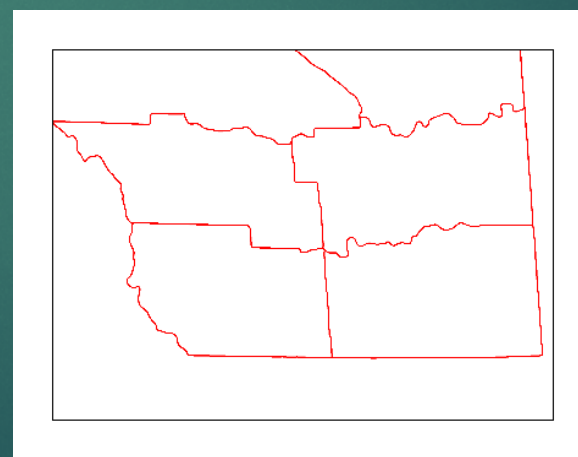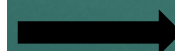Source: Godfrey and Eveleth, 2015

Image Segmentation

▶ Simple thresholding used with greyscale images
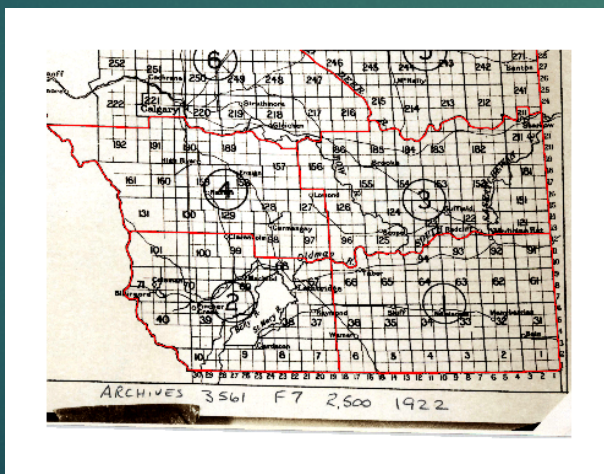
Image Segmentation – Linear Features

Image Segmentation – Areal Features

c)

d)
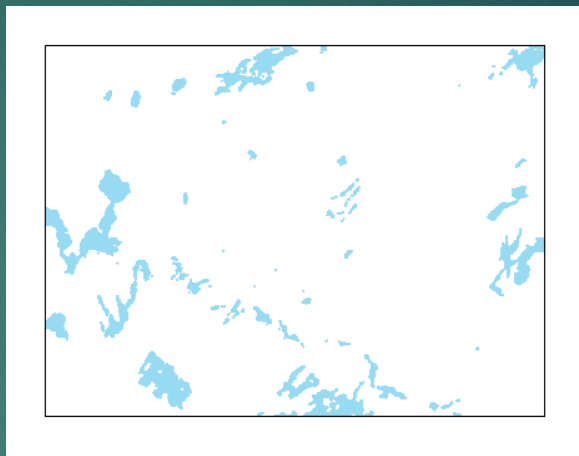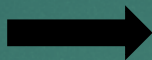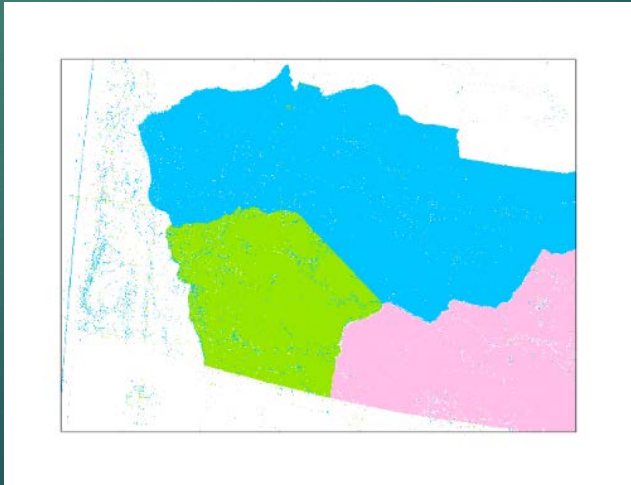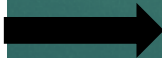
Image Segmentation: Supervised Classification

Training Sample Selection

- Select from cluttered areas
- Select a number of samples
- Can be reused with maps in same series
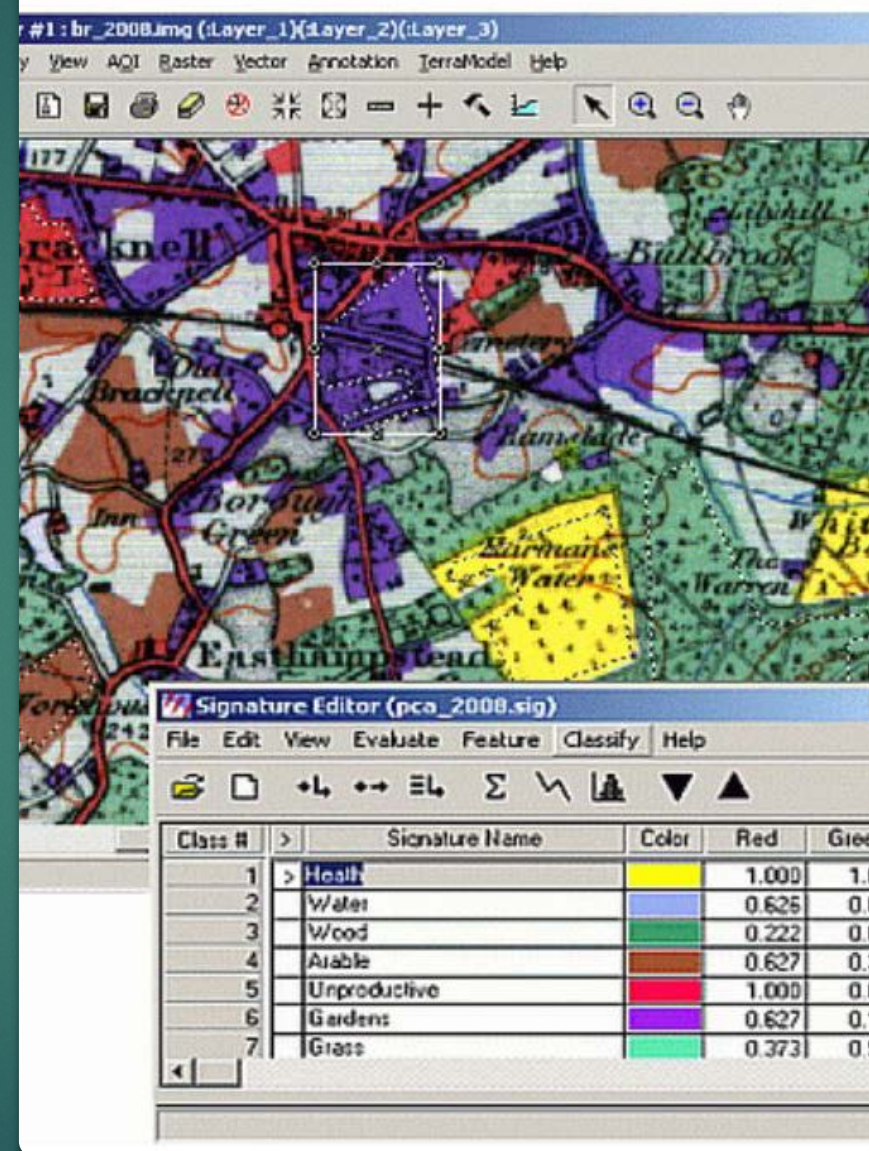
Source: Southall et al., 2003, 24

Raster Editing – Problems

Bridges
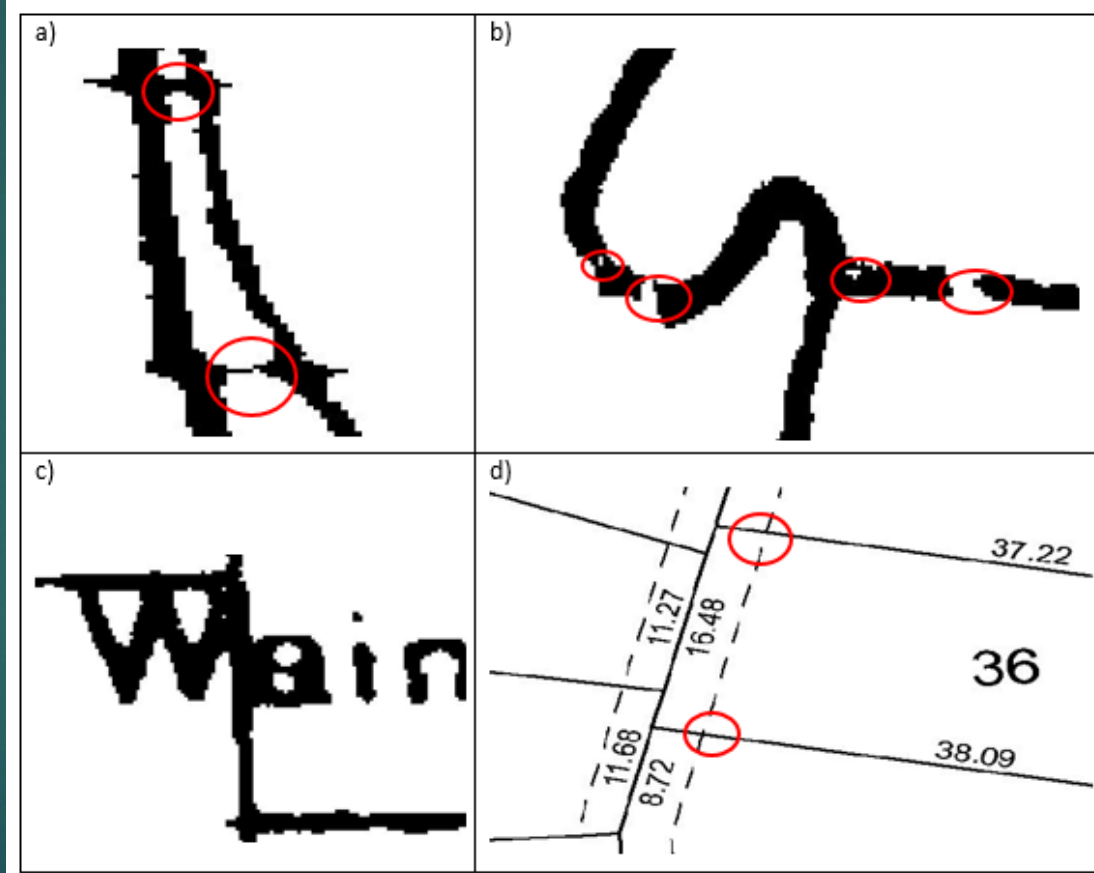
Gaps and holes

Text overlapping feature

Intersecting features

Raster Editing – Morphological Operators

Dilate

Erode

Raster Editing – GIS Functions

Original | After Segmentation | Applying Majority Filter, Expand & Shrink

Vectorization

- ArcScan and R2V – lines and outlines
  - ❖ Automated vectorization and interactive raster tracing
- ArcGIS' "RasterToPolygon" – polygons
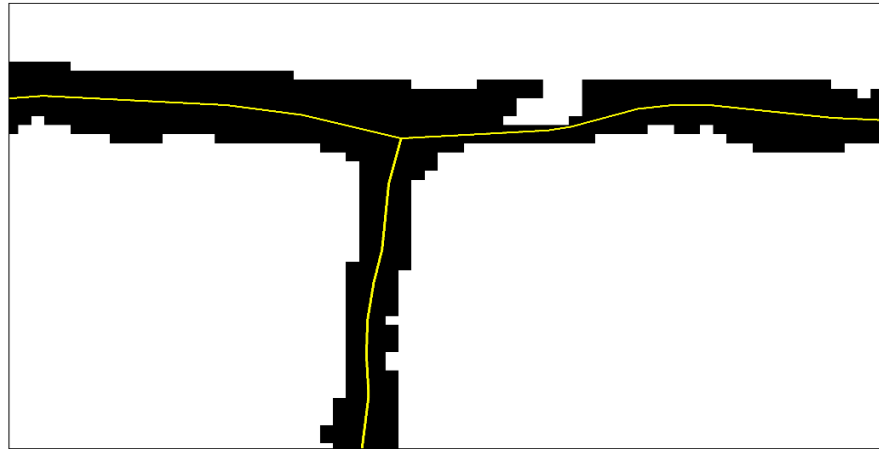
Vector Editing

- ▶ Use ArcGIS Eliminate function to dissolve areas of unwanted map elements (below a minimum threshold) and No Data gaps

- ▶ Use Smooth Line and Smooth Polygon tools to make these features look better
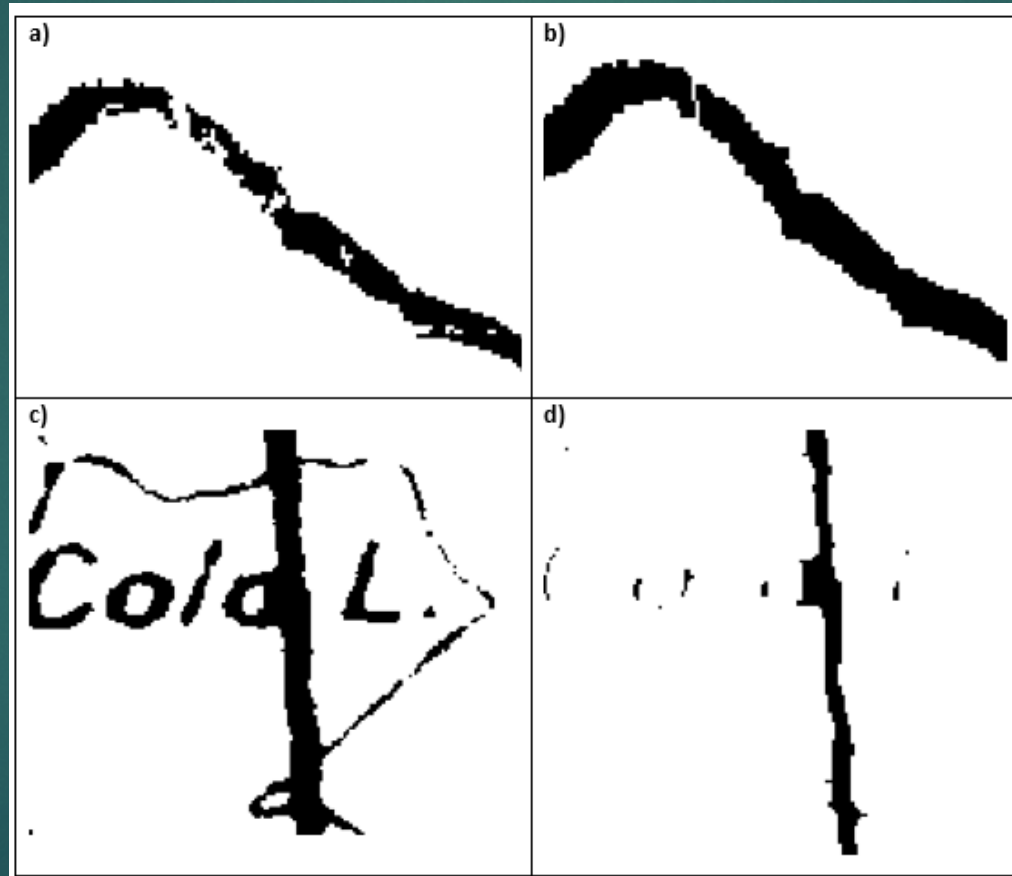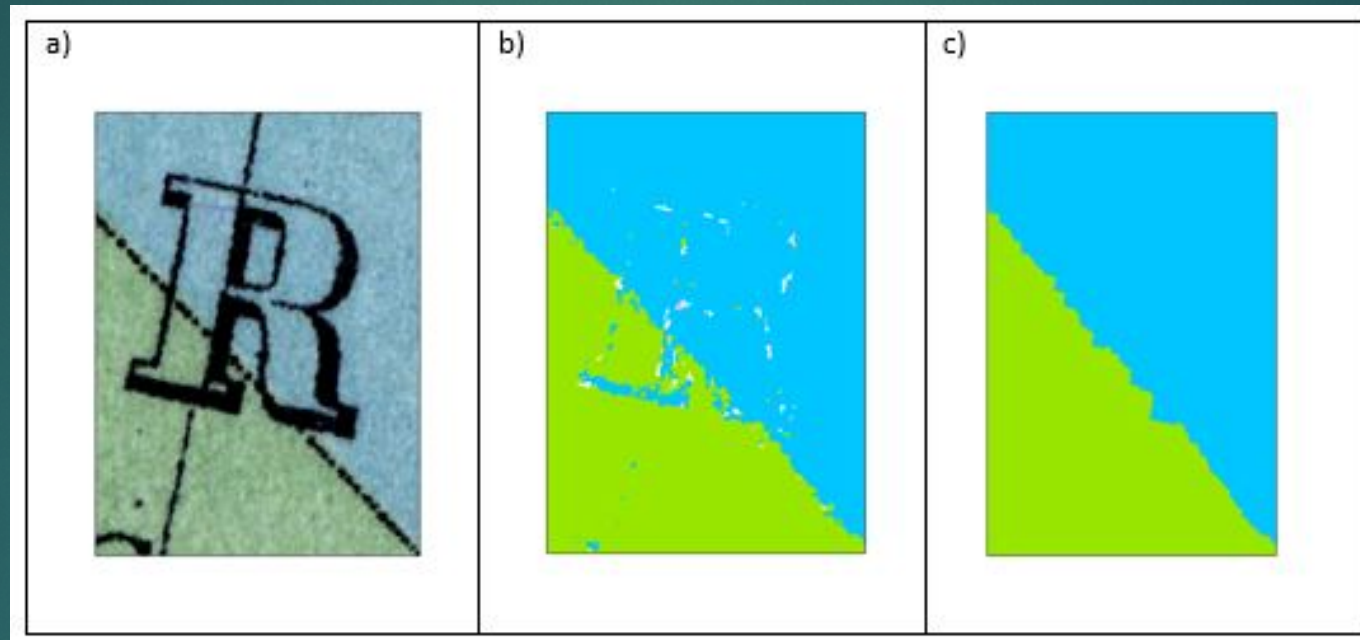
- ▶ Add attribute data to vector feature classes through the use of lookup tables to transfer multi-variable data to features

Vancouver Social Areas 1951
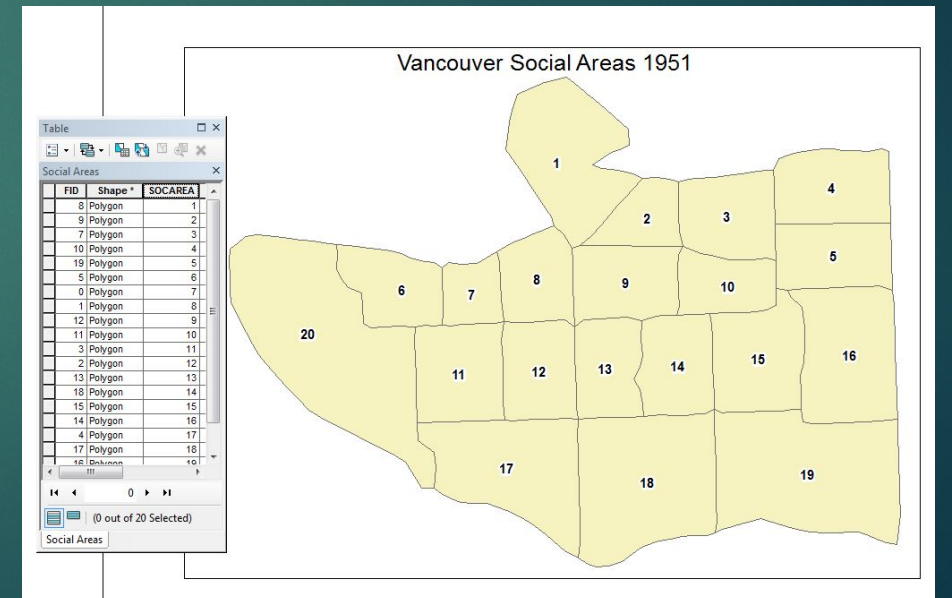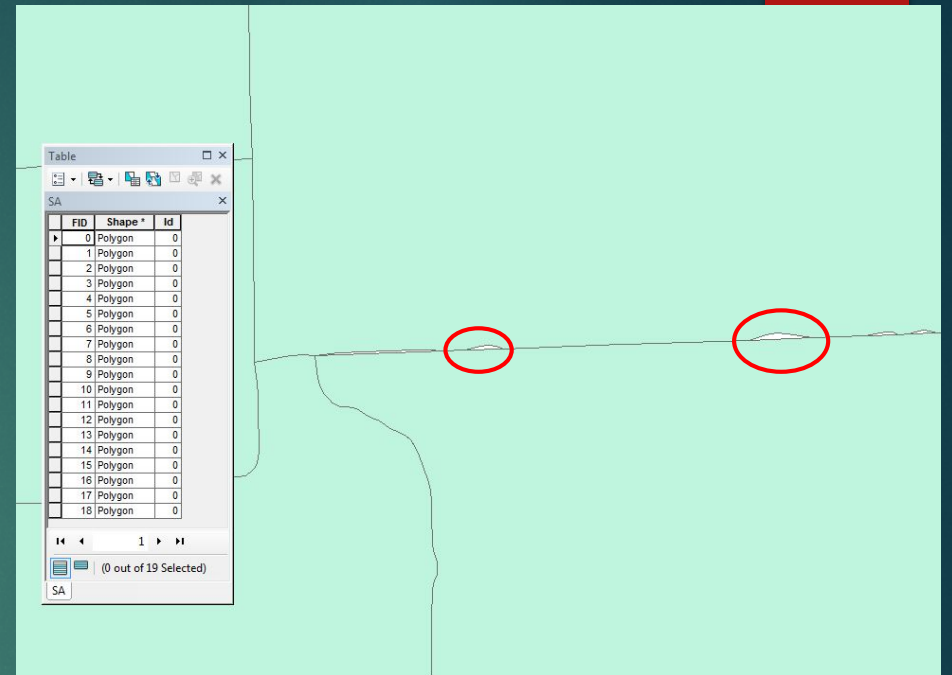
Conclusions

- Bad News:
  - There is no easy-to-use one-size-fits-all automated solution that works for all map types and features.

- Good News:
  - There is a semi-automated process that can be adopted by libraries and archives to vectorize scanned maps with a few caveats:
    - The process does require some GIS skills.
    - Some experimentation required to determine optimal process.
    - Although significantly quicker than heads-up digitizing, it still takes time and effort.
  - Vectorization enables all kinds of new possibilities for map features.