For more updated version, please go to
http://ocw.ump.edu.my/

GEOGRAPHICAL INFORMATION SYSTEMS

GIS Data Sources and Data Processing (Part 2)

by
Ngahzaifa Ab Ghani, PhD
Faculty of Computer Systems & Software Engineering

zaifa@ump.edu.my

Data Quality Assurance

- After data collection and data input, the next important process is data quality assurance.
- In GIS, data editing and data quality assurance usually consume a lot of time, sometimes longer than the process of data input
- There are several possible general data errors that can occur during data input.

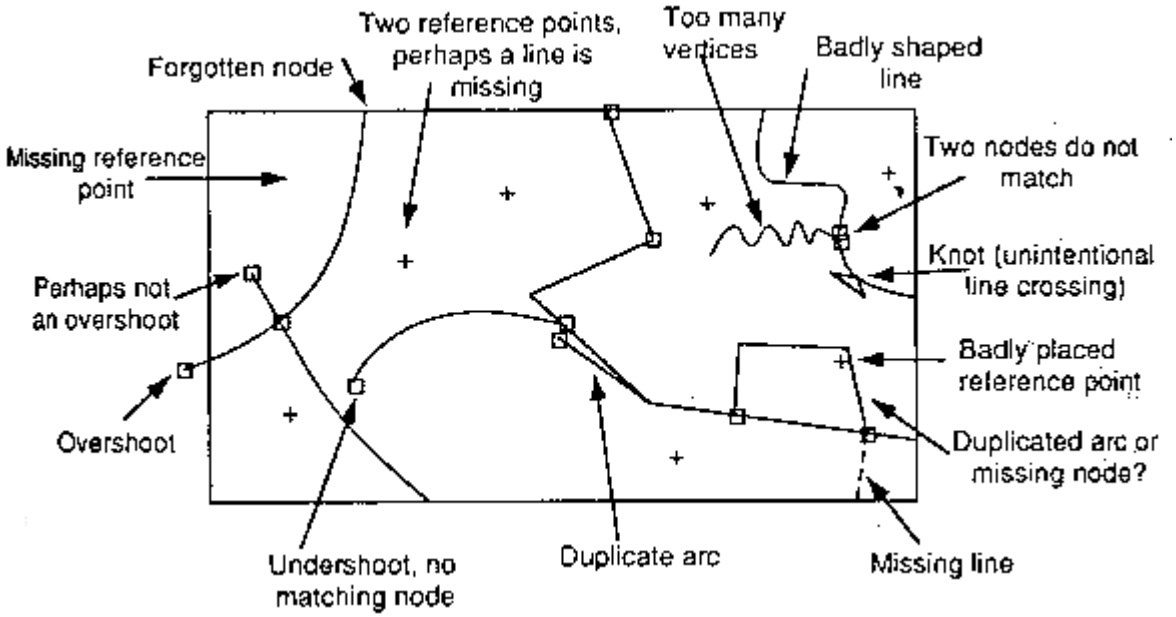
General Data Errors

- Incompleteness of the spatial data
- Locational placement errors of spatial data
- Distortion of the spatial data
- Incorrect linkages between spatial and attribute data
- Attribute data is wrong or incomplete

General Data Errors ..(cont'd)

- Incompleteness or redundancy of spatial data

  Includes missing/duplicated points, line segments, or/and polygons



Source of picture:
http://gsp.humboldt.edu/olm_2015/Lessons/GIS/04%20CreatingSpatialData/DigitizingUncertainty.html

General Data Errors ..(cont'd)

- Locational placement errors

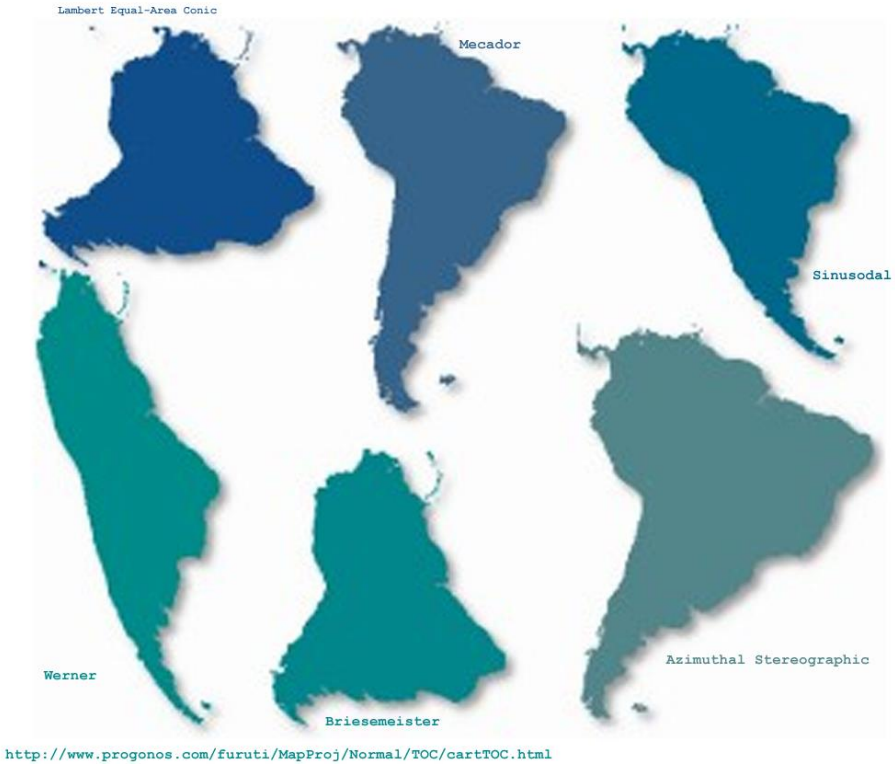
  Example:

  - A hospital is located 300m away from its location on map (mostly caused by georeferencing error)
  - An object is located on right side of the road instead of left side (the actual location) – (mostly caused by human error when inputting data)

Possible Data Errors ..(cont'd)

- Distortion of the spatial data
  - Most likely caused by wrong projection applied, or incorrect use of coordinate system
  - Sometimes caused by low quality of data (especially when data is converted from hardcopy).

- Distortion of the spatial data (cont'd)



South America with different projections

Source of picture: http://hiker.org/map_errors/diff.jpg

General Data Errors ..(cont'd)

- Incorrect linkage between spatial and attribute data

  - ✓ Mostly caused by human error (during manual key-in or digitizing).
  - ✓ Usually happens when unique identifier (unique ID) of the data is duplicated (more than one features using the same ID)

Possible Data Errors ..(cont'd)

- Wrong or incomplete attribute data
  - ✓ Most of the times caused by human error during data input
  - ✓ Sometimes caused by errors when importing/exporting data
  - ✓ Incomplete attribute data will affect the result of spatial analysis

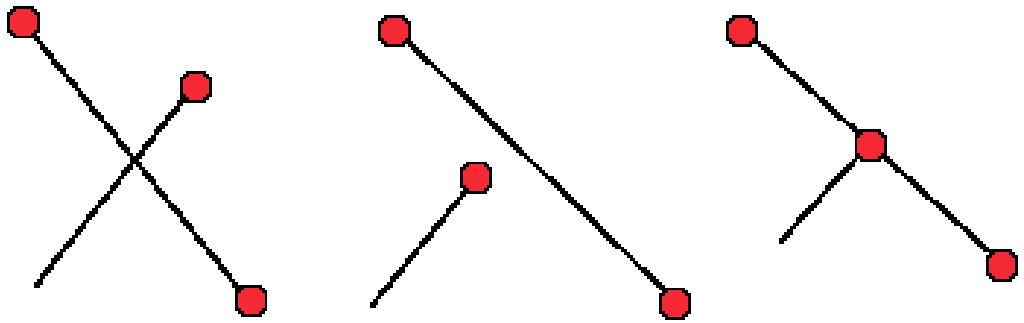
Spatial Data Errors

- Apart from the general data errors, there are some specific errors in spatial data that needs attention and correction (editing). For example,

  - ✓ Overshoot/undershoot
  - ✓ Unclosed polygon
  - ✓ Slivers and gaps in the lines
  - ✓ Extra polygons from inappropriate closing of existing polygons

Spatial Data Errors ..(cont'd)

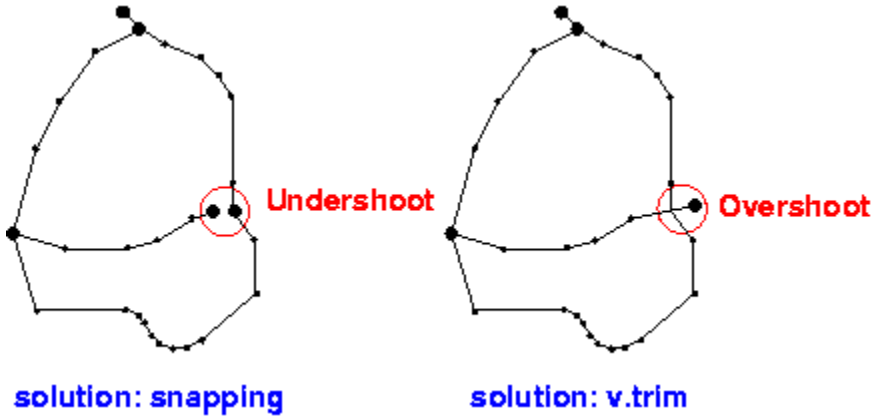
- Overshoot/undershoot
  - ✓ Overshoot can be solved by 'trimming'
  - ✓ Undershoot can be solved by 'snapping'



Source of picture: http://www.onlinegeographer.com/303_f_10/303_lecture7.html

Spatial Data Errors ..(cont'd)

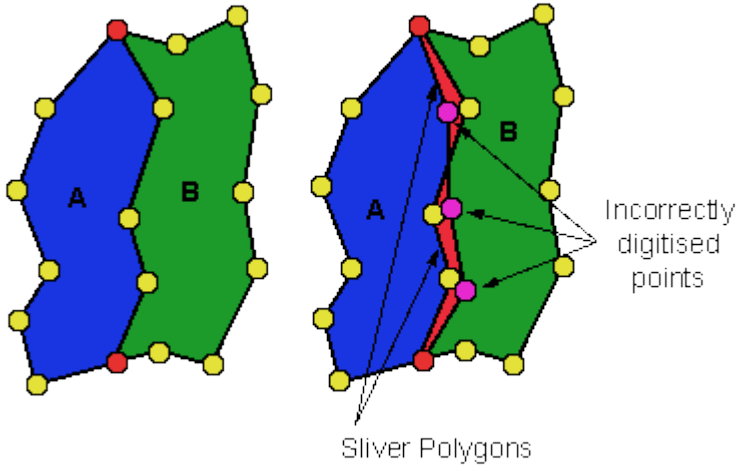
- Unclosed polygons
  - ✓ Will cause the polygon not recognized as polygon.
  - ✓ Often caused by undershoot and can be solved by 'snapping'



Source of picture: https://grass.osgeo.org/gdp/grass5tutor/HTML_en/c922.html

Spatial Data Errors ..(cont'd)

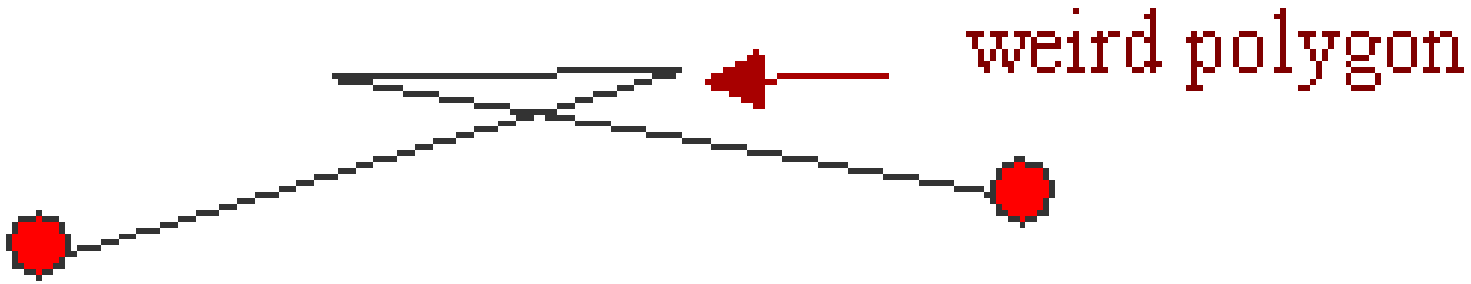
- Slivers and gaps
  - Improper joining of polygon will likely cause sliver polygons



Source of picture: http://www.onlinegeographer.com/303_f_10/303_lecture7.html

Spatial Data Errors ..(cont'd)

- Extra polygons (from inappropriate closing of existing polygons; or inappropriate intersection of lines)



Source of picture: http://www.onlinegeographer.com/303_f_10/303_lecture7.html

Think GIS way…

You have a vector file for land parcels that should have 4000 parcels. When you check, the attribute indicates only 3980 parcels. Why does this happen?

Source of picture: pixabay.com