Geographic Information Processing

GGR 321 H5S (DH 2060)

Lecture 9. Spatial Analysis II

Dr. Alex Kalynychenko
Department of Geography, Geomatics and Environment
University of Toronto Mississauga
March 11, 2025

Image courtesy: ESRI,
University of Toronto
Mississauga

Lecture 9 Outline

- Announcements (5 min): Q2, Q3, A1&2
- Textbooks and other resources
- Brief overview of Lecture 8 **Spatial Analysis I** (10-15 min): Analysis Based on Location, Analysis Based on Distance
- **Spatial Analysis I & II. Demo** (ESRI, 2023)
- Break (1.55 pm – 2 pm)
- **Spatial Analysis II (cont.)**
- Q&A (5-10 min)

Image courtesy: ESRI, University of Toronto Mississauga

Spatial Analysis I & II

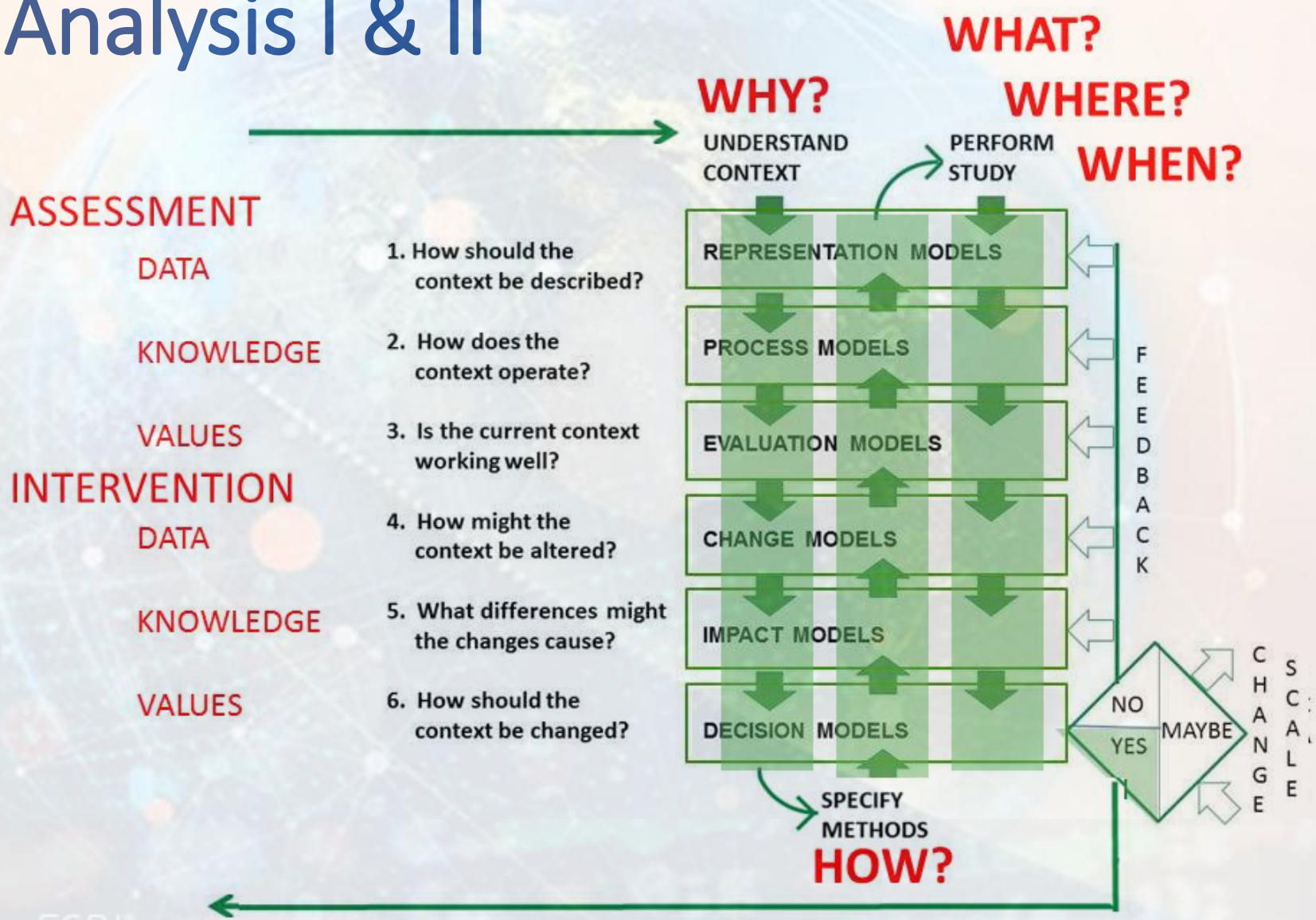
Answers scientific and practical geospatial questions:
Where? **What**? **When**? and later **Why**? **How**?

Allows to:

- **Solve complex** location-oriented **problems**
- **Explore and understand your data** from a geographic perspective
- **Determine relationships**, detect and quantify **patterns**, assess **trends**
- **Make predictions** and **decisions**
- **Address** important **questions** and **decisions** that are **beyond the scope of simple visual analysis**
- **Combine information from many sources** and derive **new information** by applying a **set of spatial operators**

Source: after ESRI 2024

Spatial Analysis I & II



A schematic model of the geodesign process, developed by Carl Steinitz to capture the various stages used in his practical applications of geodesign principles

Source: after Longley et al. (2015)

Spatial Analysis. ArcGIS Pro extensions

- The extensions are specially designed to:

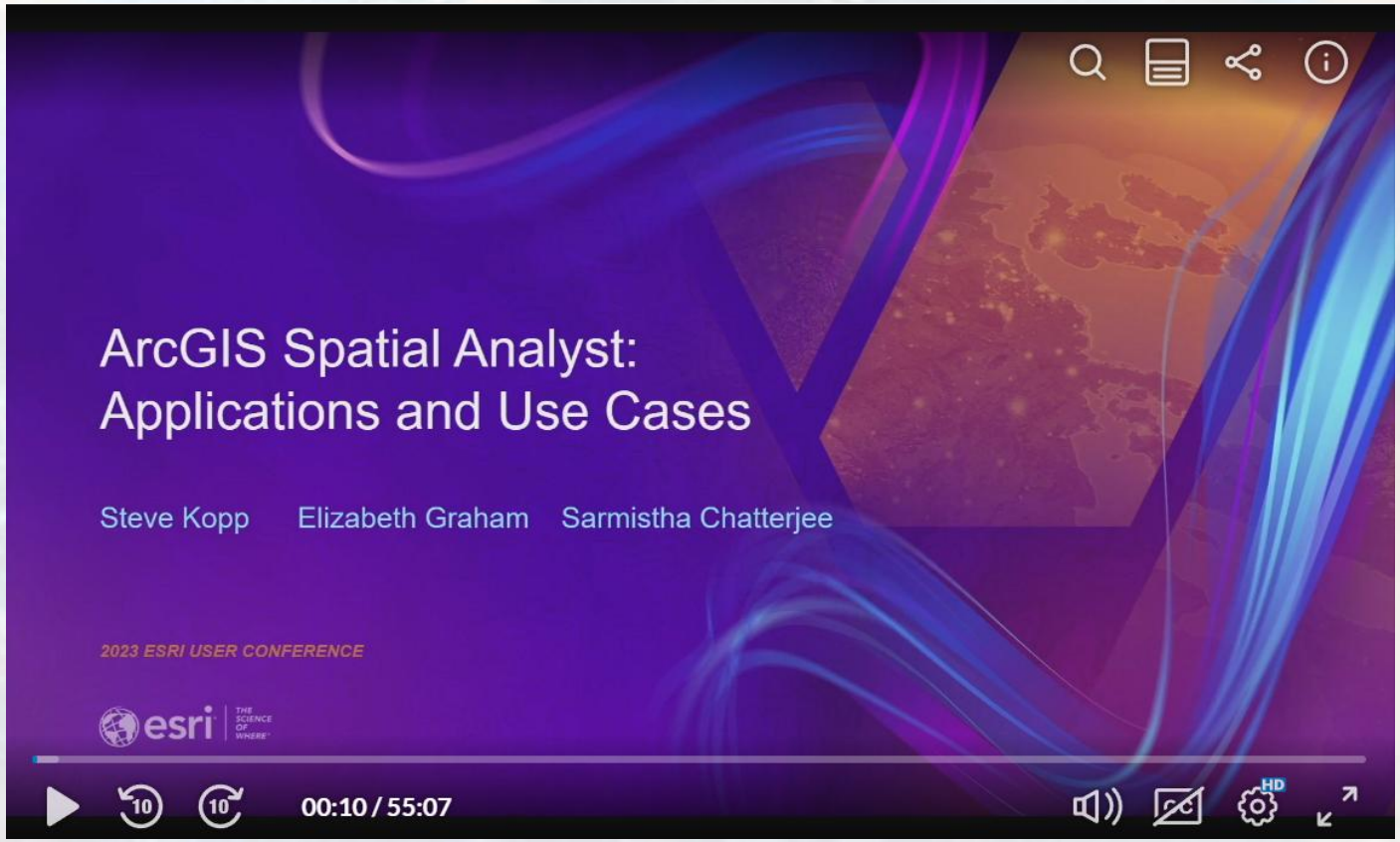
- **3D Analyst**—Analyze and create **3D GIS data** and perform 3D surface operations using rasters, TINs, terrains, and LAS datasets (lidar)
- **Business Analyst**—Analyze **market trends**, including customer and competitor analysis, site evaluation, and territory planning
- **Geostatistical Analyst**—Analyze and predict the values associated with **spatial or spatiotemporal phenomena**
- **Image Analyst**—Interpret and use **imagery**, perform feature extraction and measurement, and perform classification and object detection using machine learning
- **Network Analyst**—Measure distances and travel times **along a network** to find a route between multiple locations, create drive-time buffers or service areas, and find the best locations for facilities to serve a set of locations
- **Spatial Analyst**—Find the most suitable **locations**, calculate **distance** and determine optimal paths by incorporating the **cost of travel**, analyze and interpolate surfaces, calculate density, conduct **hydrologic analysis**, perform statistical analysis, and perform various **raster-based mathematical operations**

Image courtesy: ESRI, University of Toronto Mississauga

- Known vs unknown limitations (Aristotle)

Source: after ESRI 2024

Spatial Analysis I & II. ESRI Demo



- Spatial Analysis ESRI Webinar, April 2023 https://mediaspace.esri.com/media/t/1_g3iqfu51

Source: ESRI, 2023

Machine learning and AI

Machine learning refers to a set of data-driven algorithms and techniques that automate the prediction, classification, and clustering of data

Deep learning is an important subset of machine learning techniques that uses artificial neural networks to learn from data

Image courtesy: ESRI, University of Toronto Mississauga

Source: after ESRI 2024

Capabilities and Tools using machine learning and deep learning

- The Mapping Clusters toolset and Modeling Spatial Relationships toolset in the Spatial Statistics toolbox
- The Time Series Forecasting toolset in the Space Time Pattern Mining toolbox
- Feature, tabular, and text analysis tools in the GeoAI toolbox
- The Multidimensional Analysis toolset and Deep Learning toolset in the Image Analyst toolbox
- The Point Cloud toolset in the 3D Analyst toolbox
- The interactive deep learning-based object detection tool

Image courtesy: ESRI, University of Toronto Mississauga

Source: after ESRI 2024

Big data analytics

ArcGIS Pro includes tools that can transform **large amounts of spatial data** into manageable information

The GeoAnalytics Desktop toolbox provides a parallel processing framework for analysis on a desktop machine using Spark

The Raster Analysis toolbox contains tools for performing computationally intensive raster analysis on image services and other portal items.

Image courtesy: ESRI, University of Toronto Mississauga

Source: after ESRI 2024

Modeling and scripting

- Understanding the geoprocessing framework is helpful in writing effective Python scripts

- Python and ModelBuilder are often used in tandem, so a good knowledge of ModelBuilder is recommended to get the most out of Python scripting

Image courtesy: ESRI, University of Toronto Mississauga

Source: after Zandbergen (2024)

The Purpose of Area-Based Analyses

Measurement of area

- One way in which humans simplify geography of the Earth's infinite complexity is by ascribing characteristics to entire areas rather than to individual points.

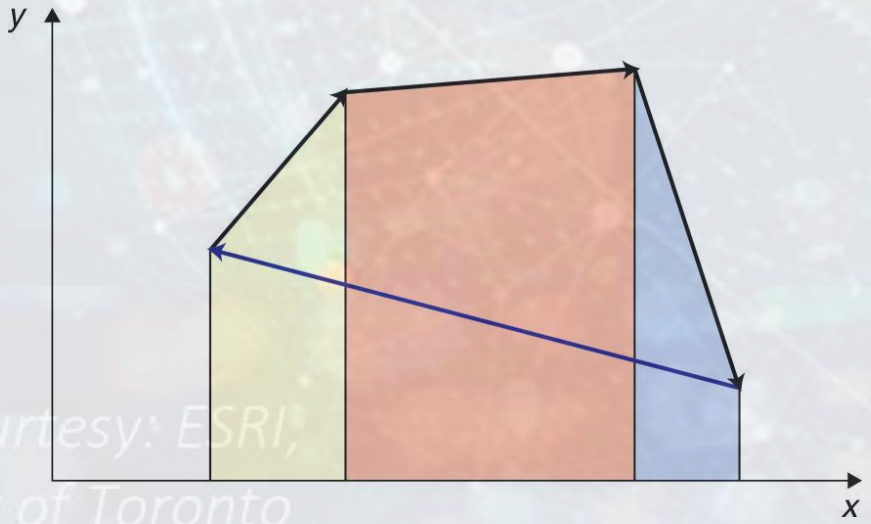


Image courtesy: ESRI, University of Toronto Mississauga

The algorithm for calculation of the area of a polygon given the coordinates of the polygon's vertices. The polygon consists of the three black arrows, plus the blue arrow forming the fourth side. Trapezia are dropped from each edge to the x-axis, and their areas are calculated as (difference in x) times (average of y).

Source: after Longley et al. (2015)

The Purpose of Area-Based Analyses

Measurement of shape

One way to bias the outcome of an election in the United States is to draw (or *gerrymander*) district boundaries to include a majority of people likely to vote for a particular party

Redistricting following the 2010 U.S. census, coupled with a new opposition to *gerrymandering* in California, transformed the old 23rd Congressional District (yellow) into a new, more compact 24th District (outlined in red)

Source: after Longley et al. (2015)

Centrality

Centers

- There are several numerical measures of *central tendency* including mean, median, and mode
- *Centers* are the two-dimensional equivalent of the mean
- The *centroid* or *mean center* is the most convenient way of summarizing the locations of a set of points
  - Found by taking the weighted average of the *x* and *y* coordinates

Image courtesy: ESRI, University of Toronto Mississauga

Source: after Longley et al. (2015)

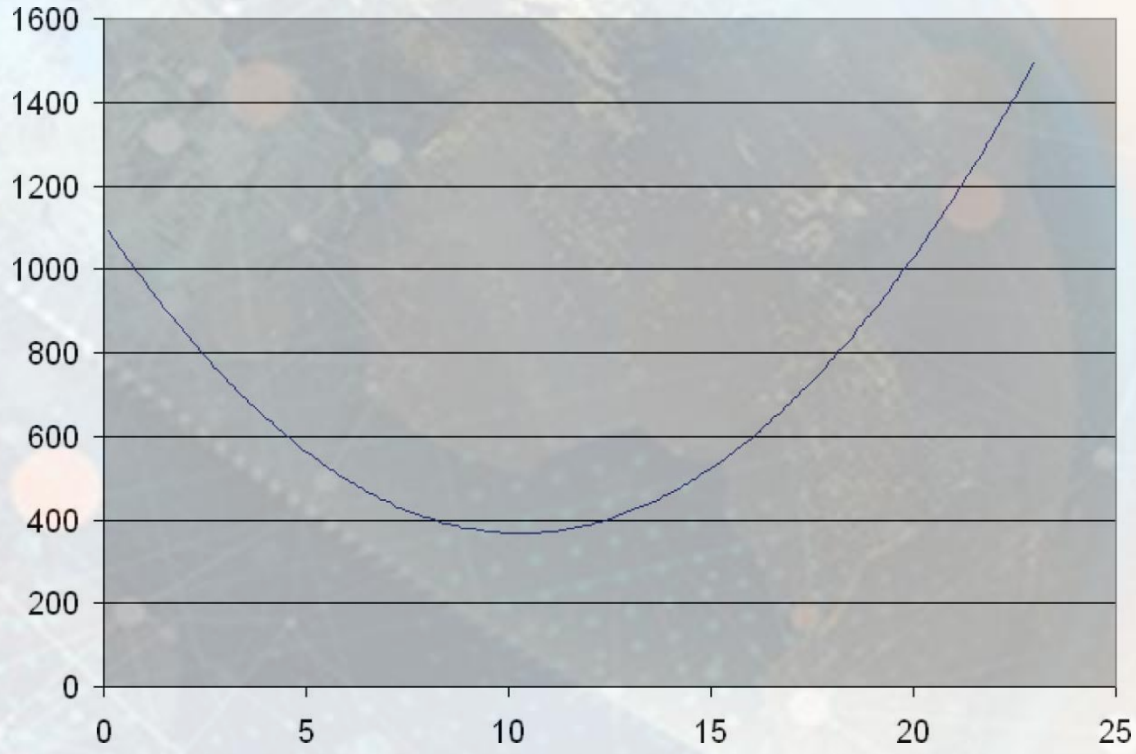
Centrality

The mean centre of the U.S. population, determined from the results of the 2010 census. Also shown is the median centre, such that half of the population is to the north and half to the south, half to the east and half to the west

(Source: U.S. Bureau of the Census, public domain)

Source: after Longley et al. (2015)

Centrality



Seven points are distributed along a line, at coordinates 1, 3, 5, 11, 12, 18, and 22. The curve shows the sum of distances squared from these points, and how it is minimized at the mean [(1+3+5+11+12+18+22)/7 = 10.3]

Source: after Longley et al. (2015)

Centrality

The mean is also the balance point, the point about which the distribution would balance if it were modeled as a set of equal weights on a weightless, rigid rod

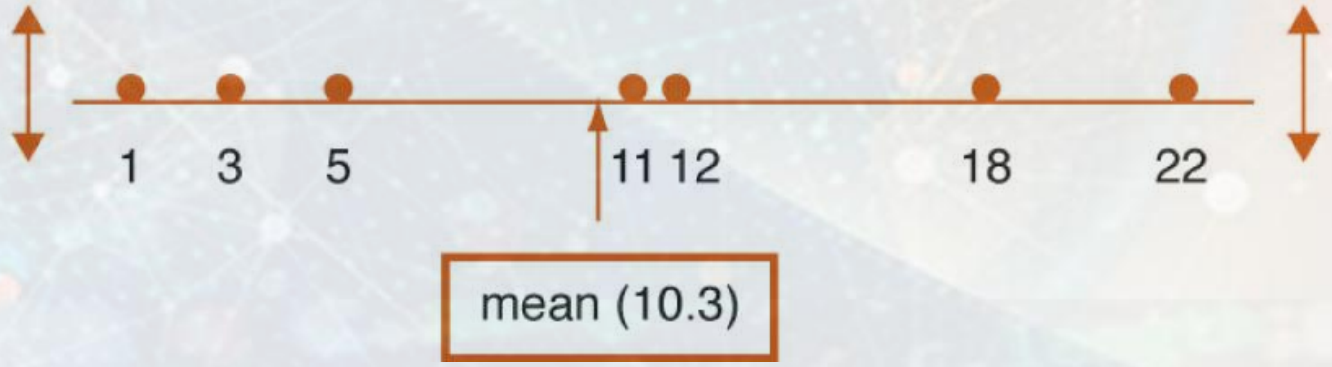


Image courtesy: ESRI, University of Toronto Mississauga

Source: after Longley et al. (2015)

Centrality

The **centroid** or **mean center** replicates the balance-point property in two dimensions—the point about which the two-dimensional pattern would balance if it were transferred to a weightless, rigid plane and suspended

Image courtesy: ESRI, University of Toronto Mississauga

Source: after Longley et al. (2015)

Centrality

Dispersion

- The measure of choice for numbers with interval or ratio properties is the standard deviation, or the square root of the mean squared difference from the mean
  - Standard deviation and variance are considered more appropriate measures of dispersion than the range (difference between the highest and lowest numbers) because as averages they are less sensitive to the specific values of the extremes.
  - RMSE is a similar measure of dispersion
- A simple measure of dispersion in two dimensions is the mean distance from the centroid

Source: after Longley et al. (2015)

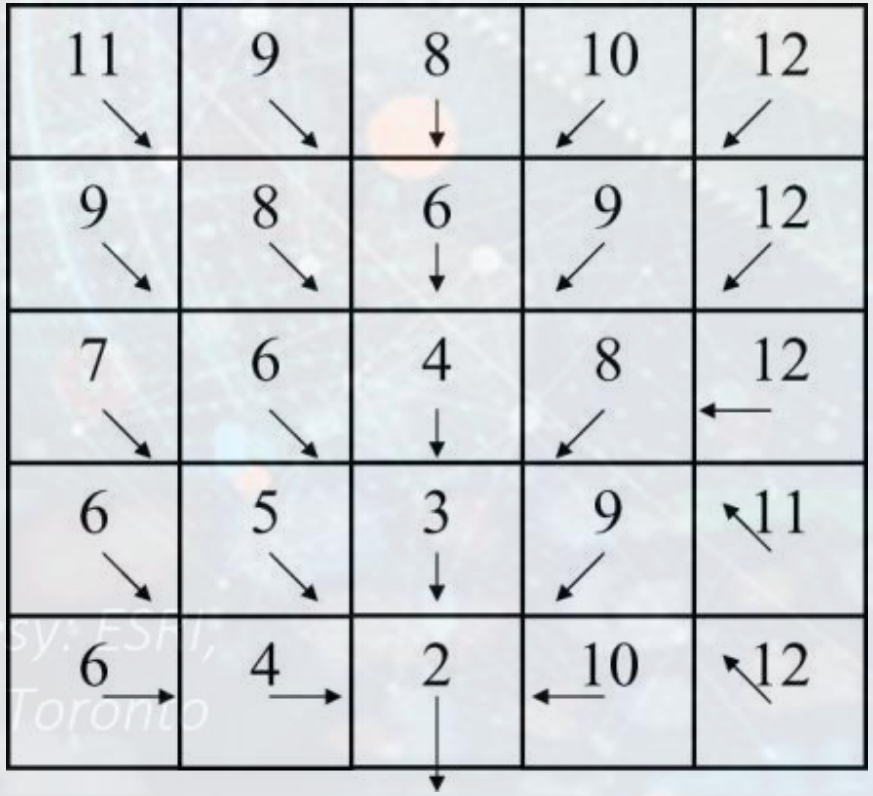
Analysis of Surfaces

- Slope and aspect ☑
- Modeling travel on a surface ☑
- **Computing watersheds and channels** ☐
- Computing visibility ☑

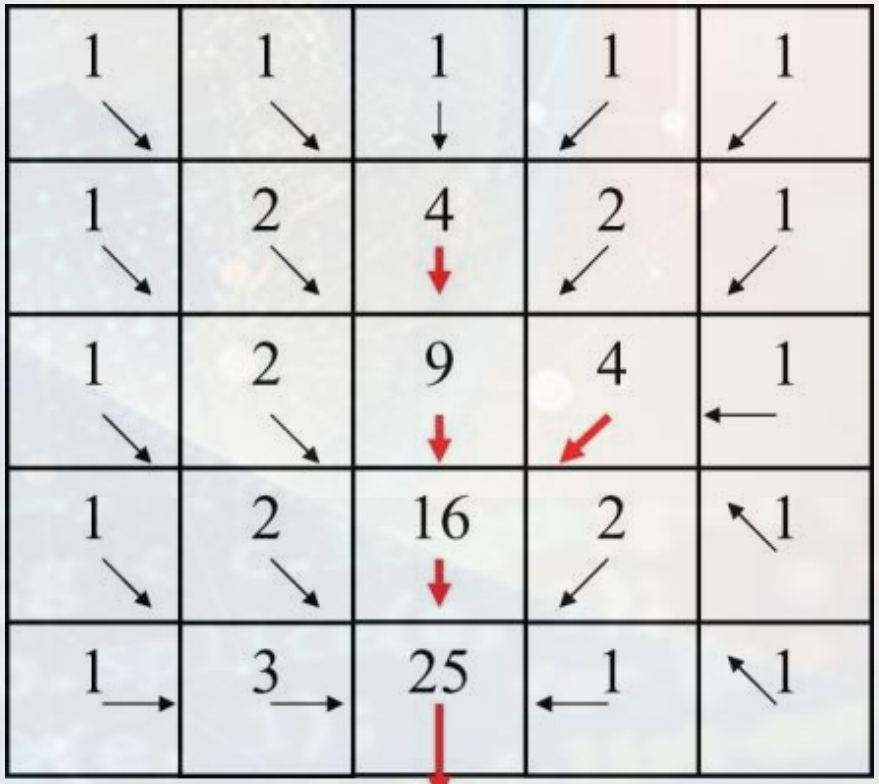
Image courtesy: ESRI, University of Toronto Mississauga

Computing watersheds and channels: hydrologic analysis of a sample DEM

The DEM and inferred flow directions using the queen's case move set



Accumulated flows in each cell and eroded channels based on a threshold flow of 4 units



Source: after Longley et al. (2015)

Computing watersheds and channels: hydrologic analysis of a sample DEM

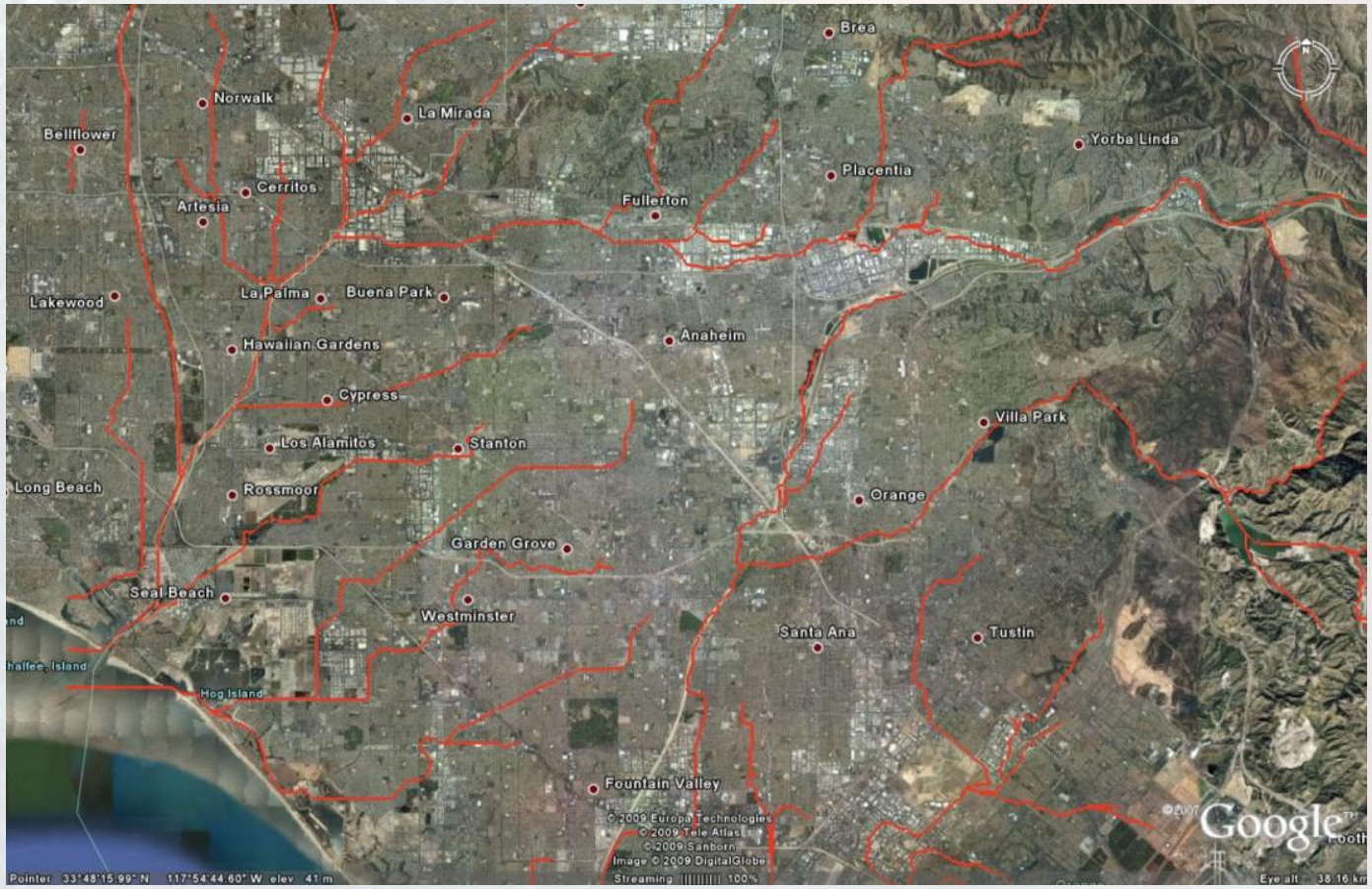
The DEM and inferred flow directions using the queen's case move set

11 ↘	9 ↘	8 ↓	10 ↙	12 ↙
9 ↘	8 ↘	6 ↓	9 ↙	12 ↙
7 ↘	6 ↘	4 ↓	8 ↙	12 ←
6 ↘	5 ↘	3 ↓	9 ↙	11 ↖
6 →	4 →	2 ↓	10 ←	12 ↖

Source: after Longley et al. (2015), https://www.travelassociates.com/

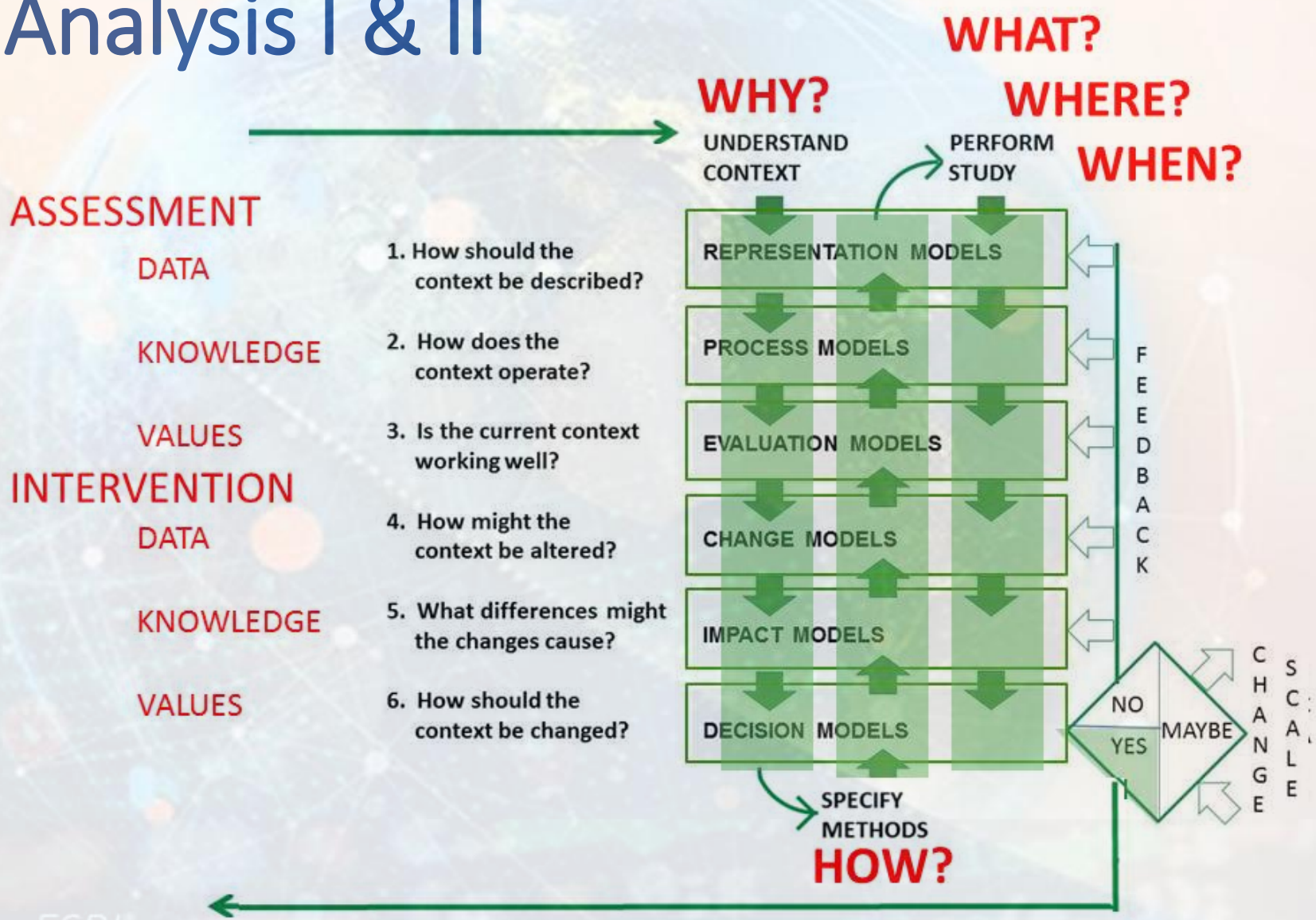
Computing watersheds and channels:
hydrologic analysis of a sample DEM

- **Analysis of the Orange County DEM predicts the river channels shown in red**
- **The Santa Ana River appears to flow out of the gorge shown in the upper right, and then far to the west before emptying into the Pacific near Seal Beach**
- **In reality, it turns south and empties near Newport Beach in the bottom center**



Source: after Longley et al. (2015)

Spatial Analysis I & II



A schematic model of the geodesign process, developed by Carl Steinitz to capture the various stages used in his practical applications of geodesign principles

Source: after Longley et al. (2015)

Point locations

Location-allocation problems which involve where to locate and how to allocate demand to central facilities include

- The *p*-median problem which seeks optimum locations for any number *p* of central facilities such that the sum of the distances between each weight and the nearest facility is minimized.
- The coverage problem which seeks to minimize the furthest distance traveled

Image courtesy: ESRI, University of Toronto Mississauga

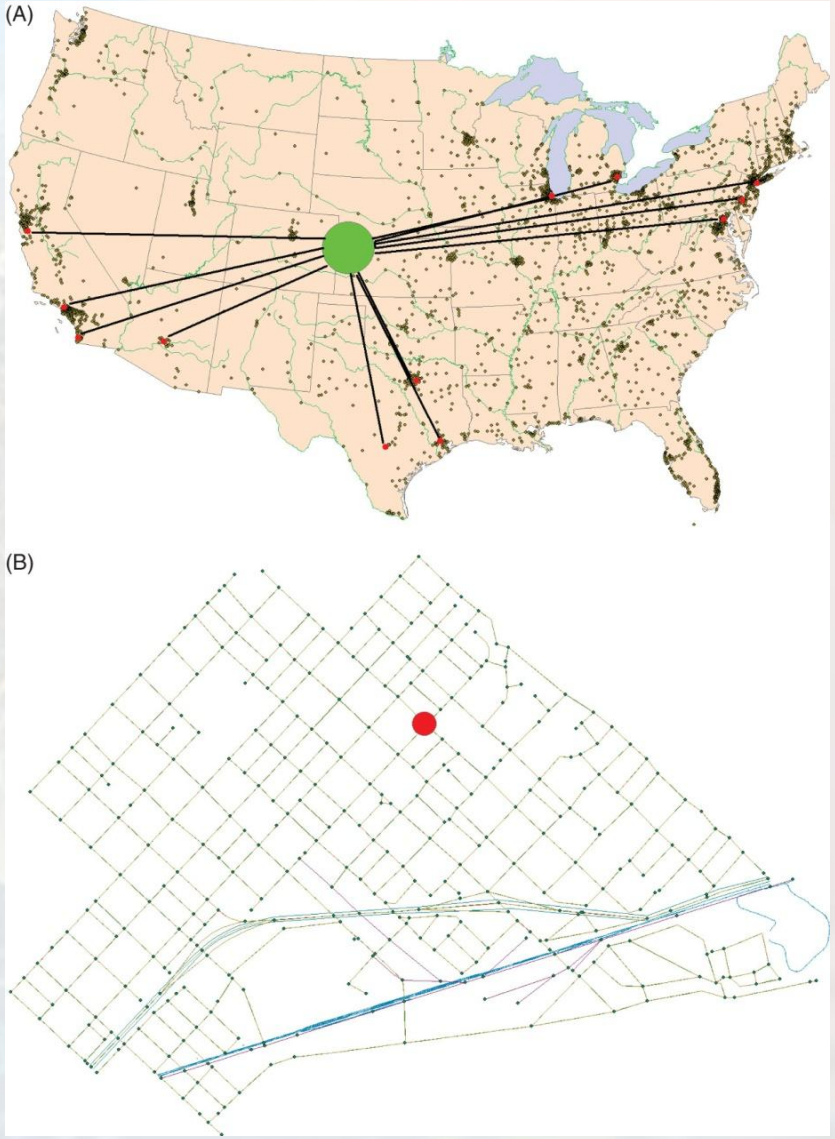
Source: after Longley et al. (2015)

Point locations (cont.)

Search for the best locations for a central facility to serve dispersed customers

(A) the problem is solved in continuous space, with straight-line travel, for a warehouse to serve the 12 largest U.S. cities. In continuous space there is an infinite number of possible locations for the site

(B) a similar problem is solved at the scale of a city neighborhood on a network, where Hakimi's theorem states that only junctions (nodes) in the network and places where there is weight need to be considered, making the problem much simpler, but where travel must follow the street network



Source: after Longley et al. (2015)

Routing problems

Routing and scheduling involves decisions about the optimum tracks followed by vehicles

- The shortest-path problem finds the path through the network between a defined origin and destination that minimizes distance or some other measure based on distance
- The traveling-salesman problem (TSP) is to select the best tour out of all possible orderings of places to visit, in order to minimize the distance (or other measure) traveled
- The orienteering problem is similar to TSP but the objective is to maximize the rewards associated with visiting a selection of the stops while minimizing the total distance traveled

Image courtesy: ESRI, University of Toronto Mississauga

Source: after Longley et al. (2015)

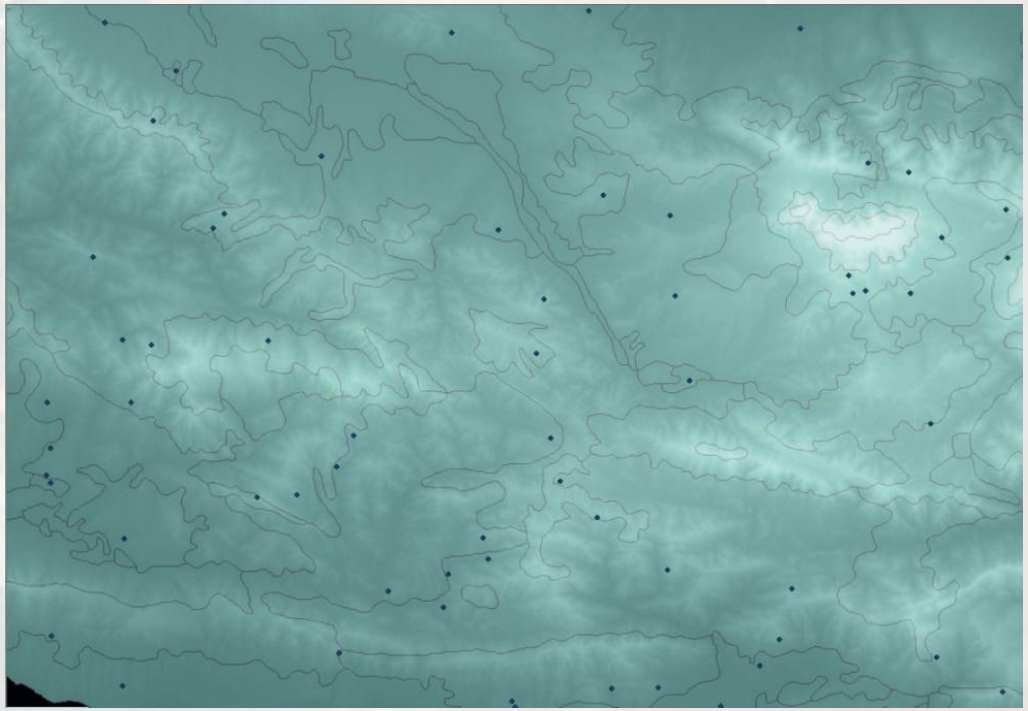
Routing problems

- Screenshot of the system used by drivers for Sears to schedule and navigate a day's workload

Source: after Longley et al. (2015)

Hypothesis testing

- Much work in statistics is ***inferential*** which uses information obtained from samples to make general conclusions about a larger population, on the assumption that the sample came from that population
- ***Randomization*** tests which simulate a large number of random arrangements of the data offer a good alternative
- A randomly placed sample of points used to examine the relationship between vegetation cover class (delimited by the boundaries shown) and elevation (whiter areas are higher), in an area north of Santa Barbara



Source: after Longley et al. (2015)

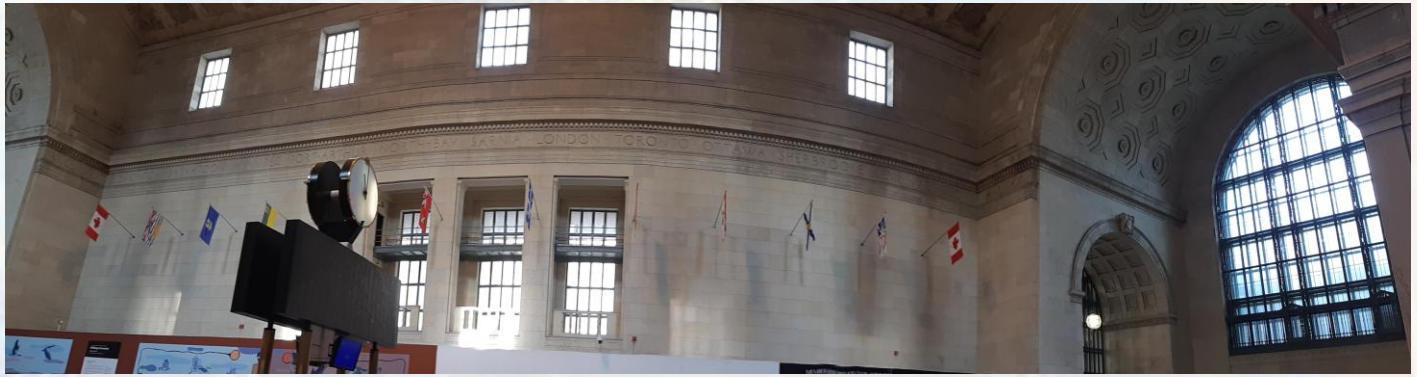
Conclusions

- Spatial Analysis is a powerful set of methods that can be applied to spatial data **to add value to them**, to **reveal patterns and anomalies** that are not immediately obvious, to find **solutions for complex** spatial **problems** and support **decision-making**

- Spatial Analysis allows **to combine spatial information from various sources** and derive **new information** by applying a **set of spatial operators**

- Selection of appropriate geospatial tools and extensions for Spatial Analysis in ArcGIS Pro environment depends on the scope of the problem/project, available data and resources, level of expertise and follows the key geodesign principles

Spatial Analysis I & II. Which route is yours?



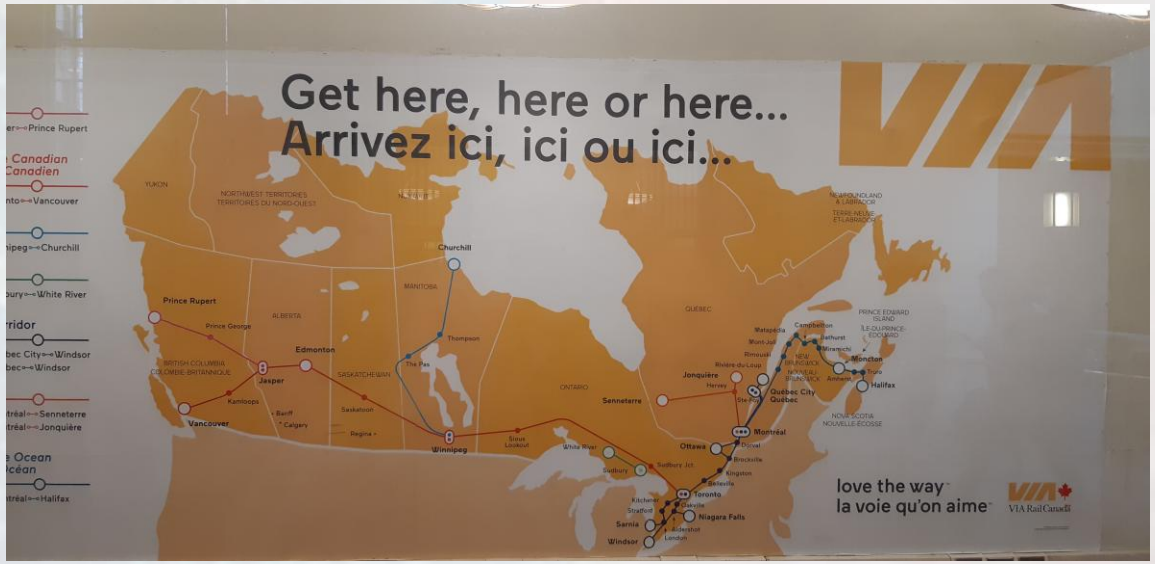


Image courtesy: ESRI, University of Toronto Mississauga

Thank you!

Reminder: ***Quiz 4 will be administered via Quercus*** *next week, covering* ***materials of Week 9&10*** *Spatial Analysis I&II and* ***recommended readings****, so please plan your time accordingly*

Next lab 3 is on Thursday, March 13, 2025

Image courtesy: ESRI, University of Toronto Mississauga