Data 301 Data Analytics
GIS

Dr. Irene Vrbik

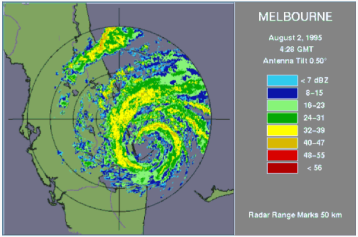
University of British Columbia Okanagan
irene.vrbik@ubc.ca

Term 1, 2018

Why learn Geographic Information Systems?

Geographic Information Systems (GIS) are used in a wide variety of areas for the analysis and display of spatial and geographical data:

- City and infrastructure planning
- Business development, planning, forecasting
- Store placement, sales trends
- Population forecasting and analysis
- Environmental and water



What is a Geographic Information System?

Geographic Information Systems are systems designed for storing, manipulating, analyzing, and displaying spatial and geographical data.

A GIS will contain components:

- for importing data from various sources in different formats
- organizing the data into layers or groups
- composing and integrating data to produce new information
- displaying the data visually as maps or 3D visualizations to help interpret the data

GIS History

- The technique of *overlaying* information on maps dates back long before computers.
- In the 1960s geographic mapping software developed the basic GIS concepts.
- The first GIS was developed by Dr. Roger Tomlinson for Canadian Department of Forestry and Rural Development.
- Tomlinson revolutionized this discipline when he introduced GIS for scanning maps into a computer and allowing not only physical landmark data, but things like population patterns, and animal migration routes in these areas to be mapped and used for statistical analysis.

GIS History

In 1969, Environmental Systems Research Institute (ESRI) founded by Laura and Jack Dangermond and developed suite of products.

- ESRI now the de facto standard for commercial GIS products (e.g. ArcGIS, ArcView).
- The software facilitates location-based analytics and provides contextual tools for mapping and spatial reasoning. Read more here.

Listen to this podcast on how the combination of data, advanced analytics, and visualization of geospatial information magnifies the power of design.

GIS Features

A GIS allows a user to add (or overlay) data on a map including:

Point a single (x,y) co-ordinate on the map

Line a connected pair of two (x,y) points

Polygon three or more (x,y) points connected to form a closed shape

These geometric shapes represent a **feature**; that is, anything you can see on the landscape.

The GIS feature consists of coordinates placing it on the map.

GIS Data types

Each *feature* can have one or more additional *attributes* (data items) describing it.

- For example, a city could be represented as a point on the map (with coordinates) and additional attributes include its name and population.

These attributes may have data types such as:

Text for names and labels

Categories for grouping similar features/classes (road, land use, etc.)

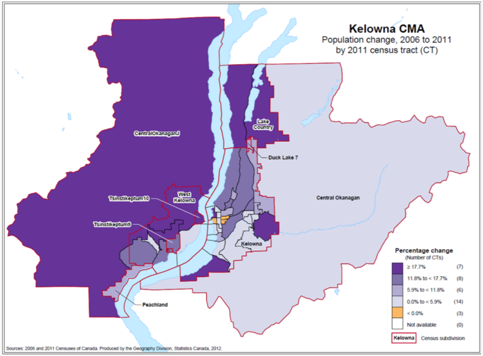
Numbers for measurements (population, rainfall, etc.)

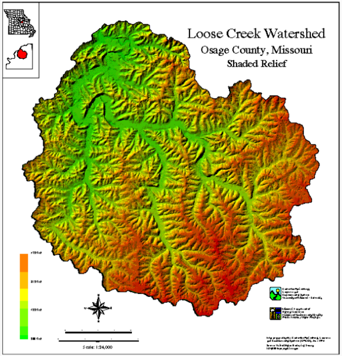
GIS Data types

- Data is often placed in categories for display.
- Each item in the category has the same symbol.
- Measurement data is also grouped into categories to ease understanding even if the data is continuous.
  - For example, elevation might be discretized into a finite number of bins in order to display this information easily on a map.

GIS Interval Data

Interval data
has values along a regular numeric scale. Source for image: Stats Can



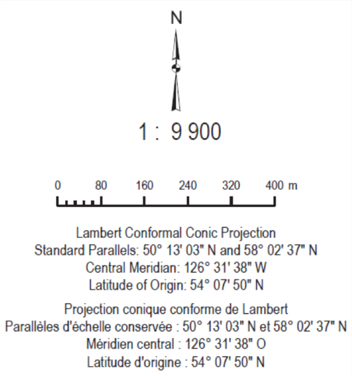


GIS Terminology: Scale and Precision

Scale is the ratio of size on the ground to size on the map.

Accuracy is a measure of how accurate the map representation is compared to the real-world.

Resolution is the smallest difference between adjacent positions that can be recorded



Example

If the scale of the map is 1:100000, and the feature on the map is 3 cm long, how long is the feature in the real-world?

A) 1 cm

B) 3 cm

C) 300000 m

D) 300 m

E) 3cm = 300000 cm = 3 km

Example

If the scale of the map is 1:100000, and the feature on the map is 3 cm long, how long is the feature in the real-world?

A) 1 cm

B) 3 cm

C) 300000 m

D) 300 m

E) *3cm = 300000 cm = 3 km*

Feature Classes and Layers

A **feature class** is a collection of objects with the same attributes.

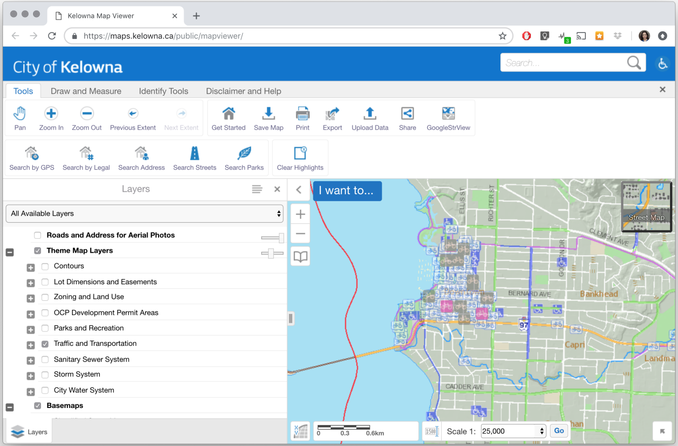
- May be stored as individual rows of a single table
- Have the same geometry (e.g. all points or all polygons).
- Example classes: states, cities, rivers.

A **layer** is a grouping of features that can be added or removed from the map (and its display visualization).

- A layer will often reference or use a feature class.

Kelowna

Feature Class and Layer Screenshot source



Representing GIS data: Raster and Vector

There are two common methods for representing GIS data.

Raster representation uses a matrix (grid of cells) of data values.

- useful for storing data that varies continuously, eg. a satellite image, a surface of chemical concentrations

Vector representation adds features (points, polygons) onto a map each with its own coordinates and attributes.

- Vector models are useful for storing data that has discrete boundaries, such as country borders, land parcels, and streets.

Vector Representation

Vector representation adds features onto a map with their own coordinates and attributes.

- Features stored as series of x-y coordinates and may be points, lines, polygons.
- Features are linked to a row in a data table which may have multiple attributes describing it.

Allows for very precise specification of features by coordinates which may have multiple attributes.

Raster Representation

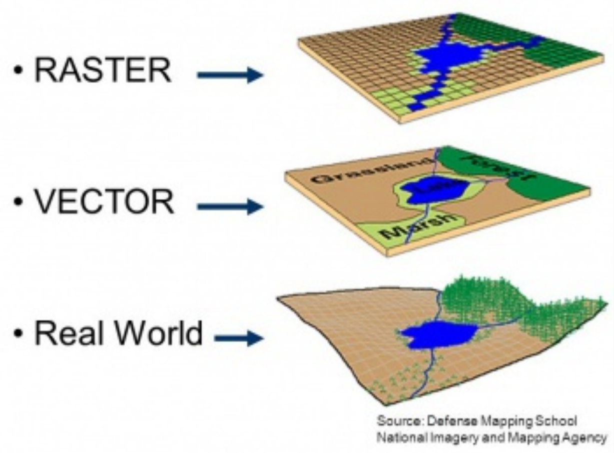
A **raster** stores data as a matrix of data points that is georeferenced to earth's surface.

- The value at each data point may be discrete (AKA thematic) or continuous.
- Often used to store continuously changing values such as elevation.

Resolution measured by cell size.

- The level of detail is often dependent on the cell (pixel) size, or spatial resolution.
- Cells must be small enough to capture adequate detail but large enough so computer storage and analysis can be performed efficiently
- Smaller cell cells implies more detail
- Raster formats: GRID, geoTIFF (both georeferenced), TIFF

Raster vs. Vector Representation



Raster vs. Vector Representation

Vector - Advantages

- precision of coordinates
- may have multiple attributes per feature
- less storage space required
- easier to scale

Vector - Disadvantages:

- implementation of overlay operations more difficult
- not easy for continuous data storage

Raster - Advantages:

- simple, robust format
- implicit georeferencing
- stores continuous data
- easy implementation of overlay operations

Raster - Disadvantages:

- storage space
- lower precision

Example

Which of the following statements are TRUE?

- In a raster if the cell size decreases by half, the resolution increases.
- A raster typically stores only one numeric data value per cell.
- A vector format may allow multiple attributes to describe each feature.
- Vector representation is better suited for continuous data than rasters.

Answer

Which of the following statements are TRUE?

- In a raster if the cell size decreases by half, the resolution increases. ✓
- A raster typically stores only one numeric data value per cell.
- A vector format may allow multiple attributes to describe each feature.
- Vector representation is better suited for continuous data than rasters.

Answer

Which of the following statements are TRUE?

- In a raster if the cell size decreases by half, the resolution increases. ✓
- A raster typically stores only one numeric data value per cell. ✓
- A vector format may allow multiple attributes to describe each feature.
- Vector representation is better suited for continuous data than rasters.

Answer

Which of the following statements are TRUE?

- In a raster if the cell size decreases by half, the resolution increases. ✓
- A raster typically stores only one numeric data value per cell. ✓
- A vector format may allow multiple attributes to describe each feature. ✓
- Vector representation is better suited for continuous data than rasters.

Answer

Which of the following statements are TRUE?

- In a raster if the cell size decreases by half, the resolution increases. ✓
- A raster typically stores only one numeric data value per cell. ✓
- A vector format may allow multiple attributes to describe each feature. ✓
- Vector representation is better suited for continuous data than rasters.✗

Representing Geographical Data

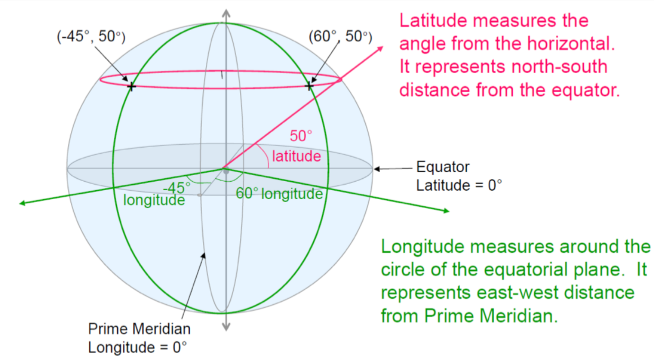
A geographic data set requires a description of its coordinate system for display and analysis, often called the **spatial reference**.

Components:

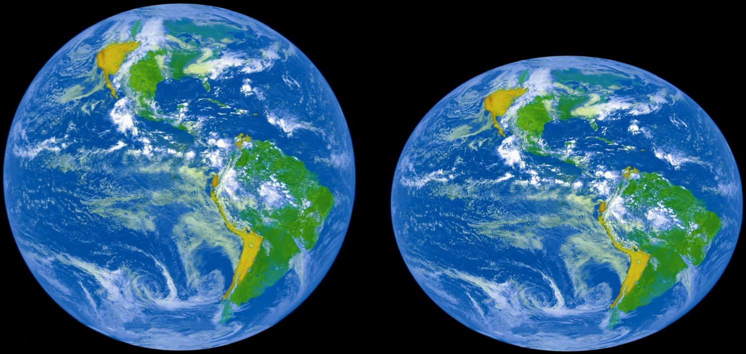
- Geographic coordinate system (GCS) / datum - for assigning coordinates to points on the earth's surface
- Projection - for mapping 3D spherical view to 2D plane

Geographic Coordinate System

Latitude and Longitude



GCS - Earth is not a Perfect Sphere

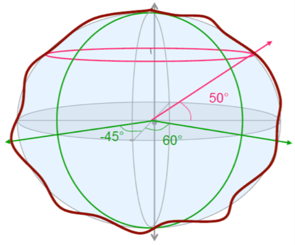


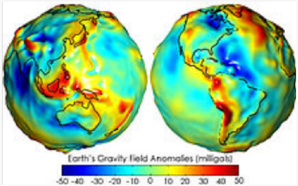
Earth is not a Perfect Ellipse

The earth has been approximated by various ellipsoids over time. Current standard was given in Maling, 1989.

Still not perfect as topography affects the height of surface features.

A geoid is an earth model that takes into account surface height from the centre of earth. (defined by gravity measurements - see here)





Datum

A **datum** is a mapping to minimize the difference between geoid and ellipsoid. Shifts ellipsoid relative to geoid for a particular location.

Datum components:

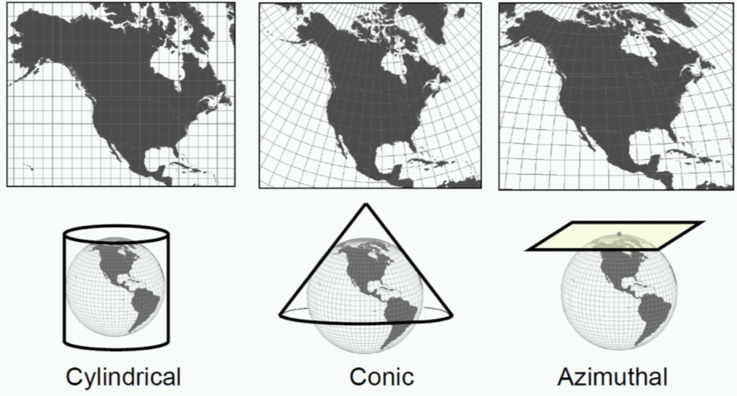
- ellipsoid used
- adjustment or fit (translation of center)

Note this means different datums are incompatible. Make sure you know your datum.

- Example: WGS 84 - reference coordinate system for GPS

Projections

A **projection** transforms a spherical coordinate system to a planar coordinate system. Each projection has different benefits and distortions.



Map Design Process

Determine objectives

- Know audience and purpose, use case

Decide on data and layers required

- What types of data: points, line, area, volume, temporal?

Plan map layout Choose colors and symbols

- Use colors consistent with understanding (red/green) and real-world
- Use bold colors sparingly. Color/size use strategically for emphasis.

Create!

Example

Which of the following statements are TRUE?

A) The earth can be modeled as a perfect spheroid.

B) Latitude measures the angle from the horizontal.

C) The zero degree for longitude is the equator.

D) A projection will cause a distortion when representing 3D as 2D.

E) A datum consists of a mapping between a geoid and an ellipsoid.

Answer

Which of the following statements are TRUE?

A) The earth can be modeled as a perfect spheroid.✗

B) Latitude measures the angle from the horizontal.

C) The zero degree for longitude is the equator. (true for latitude, 0 degree for longitude is prime meridan)

D) A projection will cause a distortion when representing 3D as 2D.

E) A datum consists of a mapping between a geoid and an ellipsoid.

Answer

Which of the following statements are TRUE?

A) The earth can be modeled as a perfect spheroid.✗

B) Latitude measures the angle from the horizontal. ✓

C) The zero degree for longitude is the equator. (true for latitude, 0 degree for longitude is prime meridan)

D) A projection will cause a distortion when representing 3D as 2D.

E) A datum consists of a mapping between a geoid and an ellipsoid.

Answer

Which of the following statements are TRUE?

A) The earth can be modeled as a perfect spheroid.✗

B) Latitude measures the angle from the horizontal. ✓

C) The zero degree for longitude is the equator. (true for latitude, 0 degree for longitude is prime meridan)✗

D) A projection will cause a distortion when representing 3D as 2D.

E) A datum consists of a mapping between a geoid and an ellipsoid.

Answer

Which of the following statements are TRUE?

A) The earth can be modeled as a perfect spheroid. ✗

B) Latitude measures the angle from the horizontal. ✓

C) The zero degree for longitude is the equator. (true for latitude, 0 degree for longitude is prime meridan) ✗

D) A projection will cause a distortion when representing 3D as 2D. ✓

E) A datum consists of a mapping between a geoid and an ellipsoid.

Answer

Which of the following statements are TRUE?

A) The earth can be modeled as a perfect spheroid.✗

B) Latitude measures the angle from the horizontal. ✓

C) The zero degree for longitude is the equator. (true for latitude, 0 degree for longitude is prime meridan)✗

D) A projection will cause a distortion when representing 3D as 2D. ✓

E) A datum consists of a mapping between a geoid and an ellipsoid.✓

Google Maps

There are a variety of GIS tools and software to use. We will use Google Maps, specifically Google My Maps, as it is easy to use and handles many of the details for us.

Google My Maps Link: https://www.google.com/maps/d/u/1

This uses vector representation format and layers consisting of single or groups of objects (features classes) can be easily added.

Supports importing data from files, including KML, CSV, and others as well as entering data via searching or map exploring.

KML

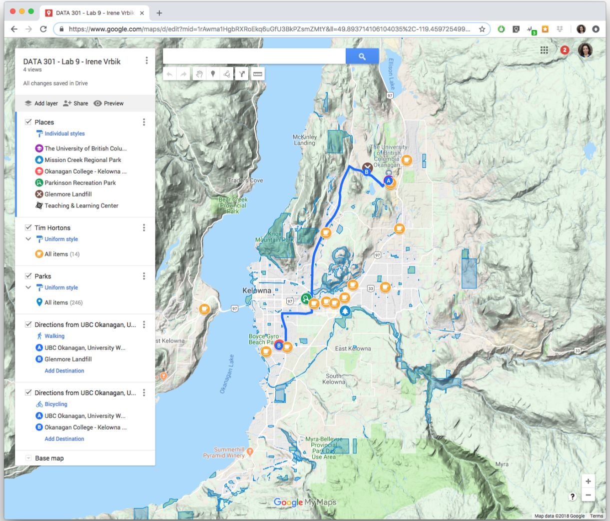
KML or Keyhole Markup Language uses XML to represent geographic information for visualization.

- Supported by Google and international standard in 2008.
- KML file represents features for display on maps with latitude/longitude coordinates.
- Data file is often in zipped form (KMZ files).

Example:

```
<?xml version="1.0" encoding="UTF-8"?>
<kml xmlns="http://www.opengis.net/kml/2.2">
<Document>
<Placemark>
  <name>New York City</name>
  <description>New York City</description>
  <Point>
    <coordinates>-74.006393,40.714172,0</coordinates>
  </Point>
</Placemark>
</Document>
</kml>
```

Layers on google maps



Google Maps API with Python

The Google Maps API can be used with Python to access and manipulate geographical data using a Python program. https://developers.google.com/maps/web-services/client-library

Services and features:

- Geocoding and reverse geocoding
- Directions (walking, driving, transit)
- Distance calculations and routes
- Elevations
- Geolocation (based on WIFI and cell towers)
- Road information and speed limits
- Times zones and places (points of interest)

Google Maps API - Getting an API Key

The first step is to get a API key that allows access to the Google services. This API key should be kept private and not shared!

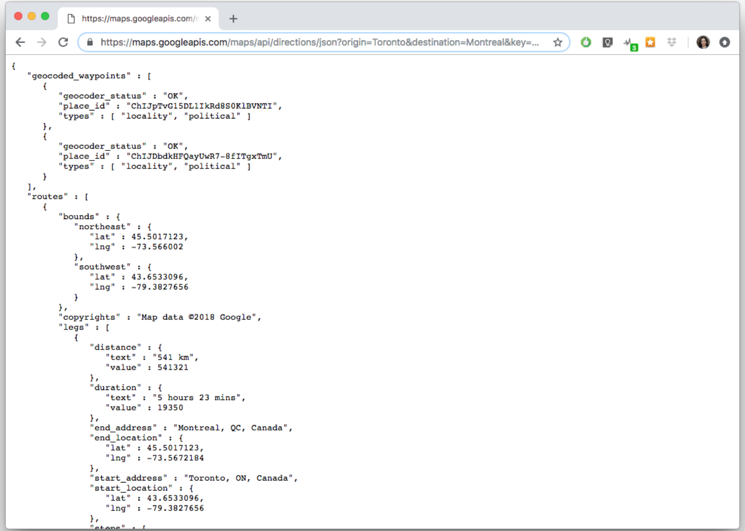
- To get a key you will need a Google account.
- For more on securing API keys see here

N.B. Google Maps web APIs have 25,000 free requests per day.

Get an API key using Google Developer Console (get API key here, see 1 minute tutorial here)

With directions API, test with:
`https://maps.googleapis.com/maps/api/directions/json?origin=Toronto&destination=Montreal&key=yourkey`
replace `yourkey` above!

The directions API will return your result in **json** format



Google Maps and Python

Google My Maps is an easy-to-use tool for displaying geographical information. The Google Maps API can be used with Python.

We can install google maps for python using the following command:

Install google maps via command line

```
pip install -U googlemaps
```

- One common problem when working with spatial data is going from human-understood locations to specific locations on a map.
- **Geocoding** is the process of converting addresses (like a street address) into geographic coordinates (like latitude and longitude), which you can use to place markers on a map, or position the map.
- This module provides an easy way to do this which uses the the Geocoding API.

Python Google Maps API Example

Geocoding with Python

```
import googlemaps
from datetime import datetime

# TODO: Replace <yourkey> with a valid API key.
gmaps = googlemaps.Client(key='yourkey')

# Use Geocoding API to look up latitude, longitude
address = '3333 University Way, Kelowna, BC, Canada'
geocode_result = gmaps.geocode(address)

print("Geocoding address...")
print("Address:",address, "Coordinates:",
        geocode_result[0]["geometry"]["location"])
```

Python Google Maps API Example

```
>>> import googlemaps
>>> from datetime import datetime
>>>
>>> # TODO: Replace the API key below with a valid API key.
... gmaps = googlemaps.Client(key='[redacted]')
>>>
>>> # Use Geocoding API to look up latitude, longitude
... address = '3333 University Way, Kelowna, BC, Canada'
>>> geocode_result = gmaps.geocode(address)
>>>
>>> print("\n Geocoding address... \n Address:",address, "\n Coordinates:",g
eocode_result[0]["geometry"]["location"])

 Geocoding address...
 Address: 3333 University Way, Kelowna, BC, Canada
 Coordinates: {'lat': 49.9399807, 'lng': -119.395521}
>>>
>>>
>>>
```

Python Google Maps API Example 2

reverse geocode with python

```
# Look up an address with reverse geocoding (UBC Van)
lat = 49.2683043
lon = -123.2489377
reverse_geocode_result=gmaps.reverse_geocode((lat, lon))

print("Reverse geocoding...")
print("Coordinates: ",lat,lon,"Address:",
        reverse_geocode_result[0]["formatted_address"])
```

Python Google Maps API Example 2

Reverse geocoding is the process of converting geographic coordinates into a human-readable address.

```
irene — python — 76×19
>>> # Look up an address with reverse geocoding (UBC Van)
... lat = 49.2683043
>>> lon = -123.2489377
>>> reverse_geocode_result=gmaps.reverse_geocode((lat, lon))
>>>
>>> print("\nReverse geocoding... \n Coordinates: ",lat,lon,"\n Address:", r
everse_geocode_result[0]["formatted_address"])

Reverse geocoding...
 Coordinates:  49.2683043 -123.2489377
 Address: UBC Student Recreation Centre, 6000 Student Union Blvd, Vancouver,
 BC V6T 1Z1, Canada
>>>
>>> 
```

Python Google Maps API Example 3

- The **Directions API** is a service that calculates directions between locations.
- You can search for directions for several modes of transportation, including transit, driving, walking, or cycling.
- Remember: we 'only' have 25,000 free requests per day.

Python Google Maps API Example 3

directions with python

```
# Request driving directions between UBCO and UBCV
directions_result = gmaps.directions(address,
          reverse_geocode_result[0]["formatted_address"],
          mode="driving", departure_time=datetime.now())
leg = directions_result[0]['legs'][0]

print("Driving directions...")
print("Start address:",leg['start_address'],
      "Destination address:",leg['end_address'])
print("Distance:",leg['distance']['text'],
      "Time:",leg['duration']['text'])

for step in leg['steps']:
    print("Step:",step['duration']['text'],
          step['html_instructions'])
```

Python Google Maps API Example 3

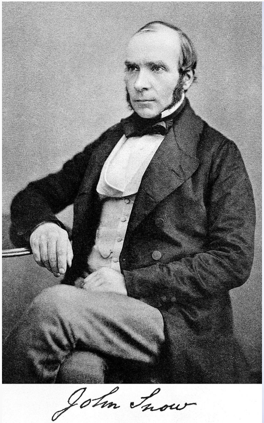
```
irene — python — 77×27
>>> # Request driving directions between UBCO and UBCV
... directions_result = gmaps.directions(address,

...          reverse_geocode_result[0]["formatted_address"],
...          mode="driving", departure_time=datetime.now())
>>> leg = directions_result[0]['legs'][0]
>>>
>>>
>>> print("\nDriving directions... \n Start address:",leg['start_address'],
...       "\nDestination address:",leg['end_address'],
...       "\nDistance:",leg['distance']['text'],
...       "Time:",leg['duration']['text'])

Driving directions...
 Start address: 3333 University Way, Kelowna, BC V1V 1V7, Canada
Destination address: UBC Student Recreation Centre, 6000 Student Union Blvd,
Vancouver, BC V6T 1Z1, Canada
Distance: 412 km Time: 4 hours 34 mins
>>>
>>> for step in leg['steps']:
...     print("Step:",step['duration']['text'],
...            step['html_instructions'])
...
```

Python Google Maps API Example 3

```
irene — python — 77×27
N</b>
Step: 1 min At the roundabout, take the <b>2nd</b> exit onto the <b>Hwy 97 N<
/b>/<b>Okanagan Hwy</b>/<b>BC-97 S</b> ramp to <b>UBC Okanagan</b>
Step: 33 mins Merge onto <b>Hwy 97 N</b>/<b>Okanagan Hwy</b>/<b>BC-97 S</b><d
iv style="font-size:0.9em">Continue to follow Hwy 97 N/Okanagan Hwy</div>
Step: 1 min Take the <b>BC-97C W</b> ramp to <b>Merritt</b>/<b>Kamloops</b>/<
b>Hope</b>
Step: 56 mins Continue onto <b>BC-97C</b>
Step: 1 hour 0 mins Turn <b>left</b> to merge onto <b>BC-5 S</b> toward <b>Va
ncouver</b>/<b>Yellowhead Highway S</b>
Step: 4 mins Continue onto <b>BC-3 W</b>
Step: 1 hour 22 mins Continue onto <b>Trans-Canada Hwy</b>/<b>BC-1 W</b>
Step: 1 min Take exit <b>27</b> for <b>First Ave</b>
Step: 1 min Keep <b>left</b> at the fork to continue toward <b>E 1st Ave</b>
Step: 8 mins Turn <b>left</b> onto <b>E 1st Ave</b>
Step: 3 mins Continue onto <b>Terminal Ave</b>
Step: 1 min Turn <b>left</b> onto <b>Quebec St</b>
Step: 1 min Continue straight to stay on <b>Quebec St</b>
Step: 3 mins Turn <b>right</b> onto <b>E 2nd Ave</b>
Step: 2 mins Continue onto <b>W 6th Ave</b>
Step: 13 mins Slight <b>right</b> onto <b>W 4th Ave</b>
Step: 2 mins Continue onto <b>Chancellor Blvd</b>
Step: 1 min Turn <b>left</b> onto <b>Wesbrook Mall</b>
Step: 1 min Turn <b>right</b> onto <b>Student Union Blvd</b>
Step: 1 min Turn <b>left</b> to stay on <b>Student Union Blvd</b><div style="
font-size:0.9em">Destination will be on the right</div>
>>>
```

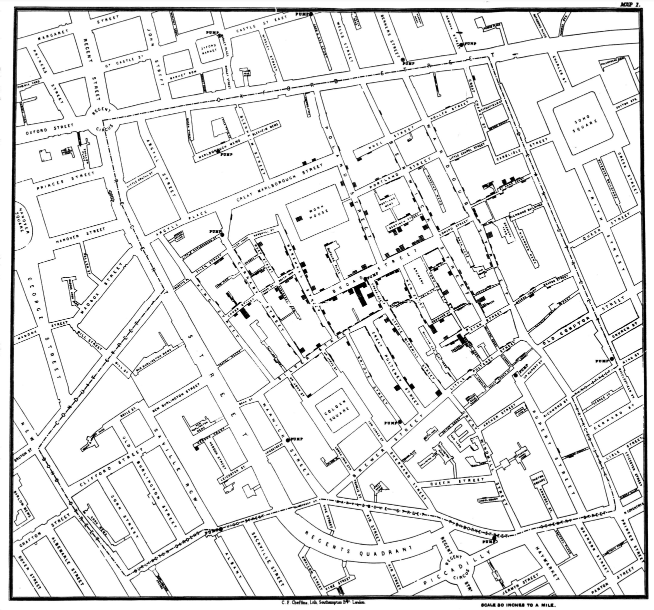
John Snow Example

- When could we use this?
- Lets consider the data analysis preformed by John Snow:



- (not to be confused with Jon Snow from GoT)

- John Snow (March 1813 — June 1858) was was an English physician considered to be one of the fathers of modern epidemiology.
- He is cited as one of the earliest known examples of using GIS by famously tracing the source of a cholera outbreak in Soho, London, in 1854.
- This outbreak took the lives of 127 people in only three days, killing over 500 by by September 10, 1854.
- The map on the following page is the original showing the clusters of cholera cases in the London epidemic. Image source: Wikipedia.


MAP 1.
OXFORD STREET
REGENT CIRCUS
GREAT MARLBOROUGH STREET
BROAD STREET
WORK HOUSE
SOHO SQUARE
GOLDEN SQUARE
HANOVER SQUARE
CONDUIT STREET
SAVILLE ROW
REGENT STREET
REGENTS QUADRANT
QUEEN STREET
PICCADILLY
PUMP
C.F. Cheffins, Lith, Southampton Bldgs London.
SCALE 30 INCHES TO A MILE.

- Perhaps made more clear in an updated version of this map, on the following page, is that these outbreaks were clustered around the water pump on Broad Street.
- By observing these these outbreaks spatially on his original map, Snow was able to form his hypothesis that a contaminated water wells was the lead contributor to the cholera death count.
- Subsequently, Snow convince the local council to remove the handle to prevent its use.

Image created by Robin Wilson of Southampton University: source

Conclusion

Geographic Information Systems are systems designed for storing, manipulating, analyzing, and displaying spatial and geographical data. A GIS supports:

- importing data from various sources in different formats
- organizing the data into layers or groups and integrating data from sources
- displaying the data visually as maps or 3D visualizations to help interpret the data

Understanding how GIS data is encoded using a geographical coordinate system and datums is important when interpreting and combining data from sources.

Google My Maps is an easy-to-use tool for displaying geographical information. The Google Maps API can be used with Python.

Obejectives

- Provide examples where a GIS is used
- Define GIS and list some of its features/components
- Appreciate history of GIS including Canadian connection
- List and use GIS features: text, point, line, polygon
- Explain the relationship between features, coordinates, and attributes
- Provide an example on how interval and categorical data is displayed
- Define: scale, precision, resolution and perform simple calculations
- Define: feature class, layer
- Compare and contrast raster versus vector representations

Obejectives

- Define and use latitude and longitude
- Explain the challenge in modeling a point on the earth's surface given that it is not a perfect sphere and has topography
- Explain role and connection between a geoid, spheroid, datum
- Explain the purpose of a projection and understand different projections have different benefits and distortions
- Apply a map design process to produce visually appealing maps
- Define and use KML
- Create a map with Google My Maps with various features
- Write a program to access the Google Maps API using Python