Algorithms for GIS

Laura Toma

Bowdoin College

GIS (Geographic Information Systems)

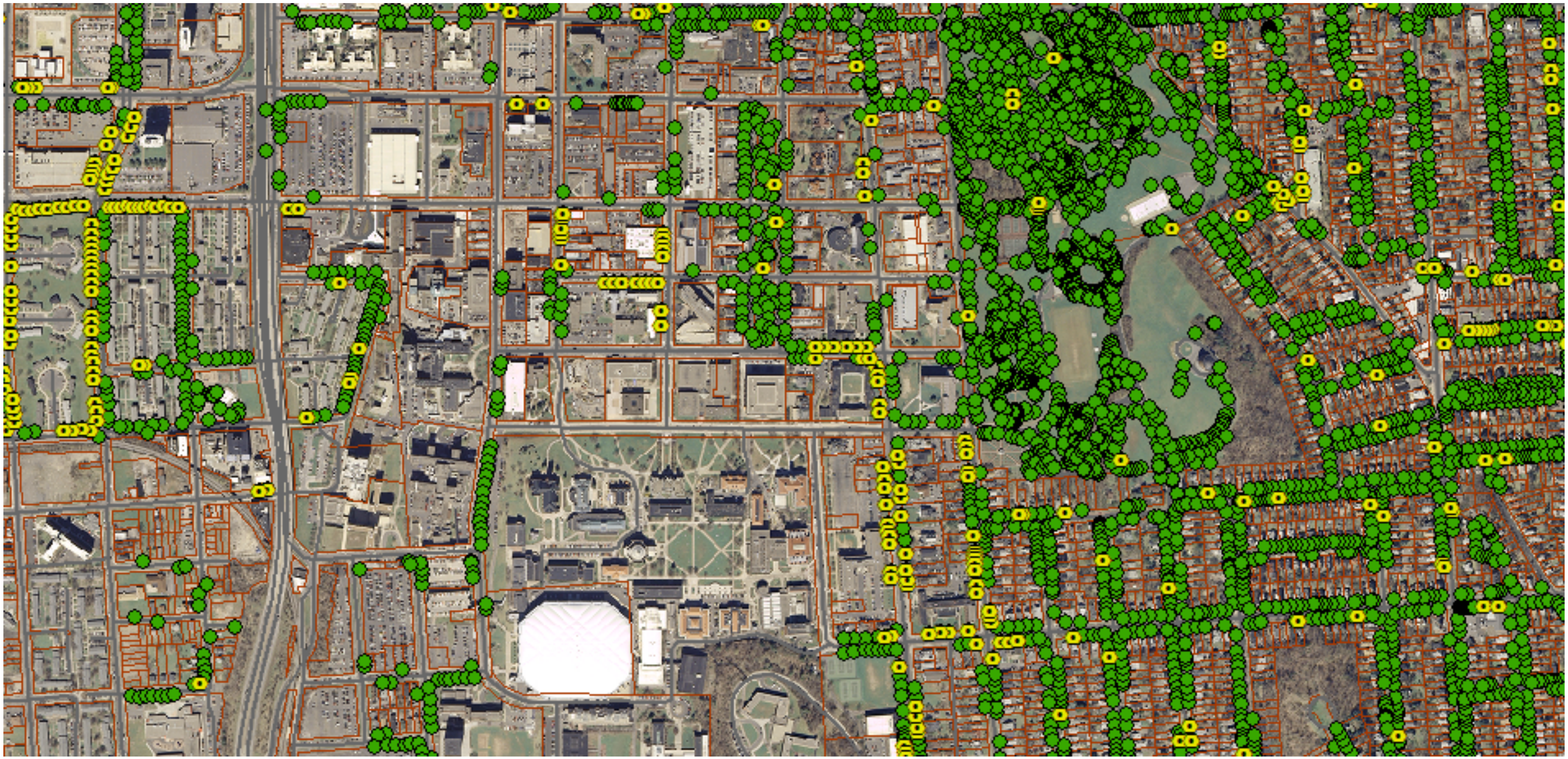
- Systems for storing, analyzing and visualizing geospatial data
- What can you do with a GIS?

What can you do with GIS?

- Mapping and visualization
    - Display different types of data, all on same location (layers); Turn layers on and off ; Zoom in/out;
    - Create beautiful, interactive maps
    - Combine data from many sources
    - E.g. ESRI: maps

all images from ESRI

http://researchguides.library.syr.edu/c.php?g=258118&p=1723814



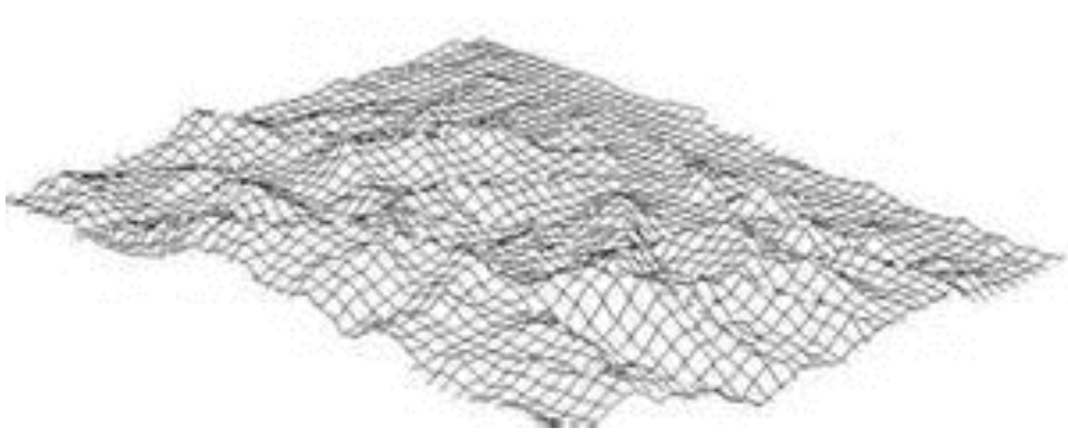
Polygons (properties), lines (streets), points (trees) and raster images (air photo) are integrated into one map.

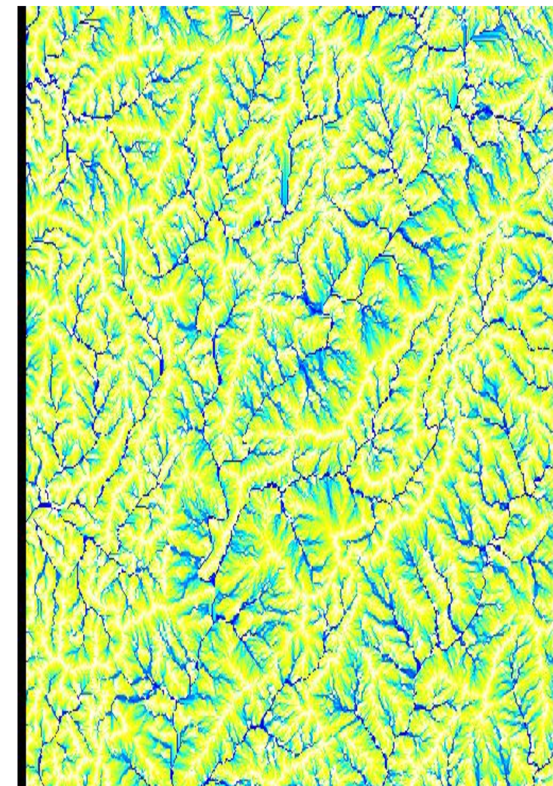
What can you do with GIS?

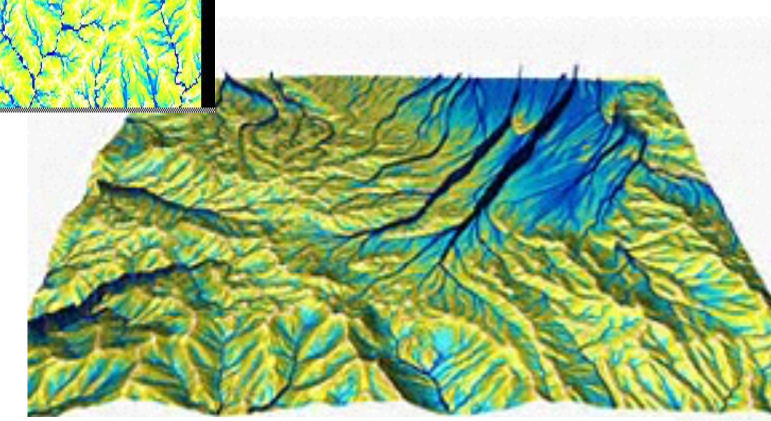
- Spatial analysis
  - Spatial queries: What lies within 5 miles of a dump site? What other crimes have occurred in this window?
  - Terrain analysis: Model phenomena like rainfall, flooding, erosion, solar radiation, sediment flow, visibility
  - Find shortest routes, connectivity
  - E.g.: ESRI: spatial analysis

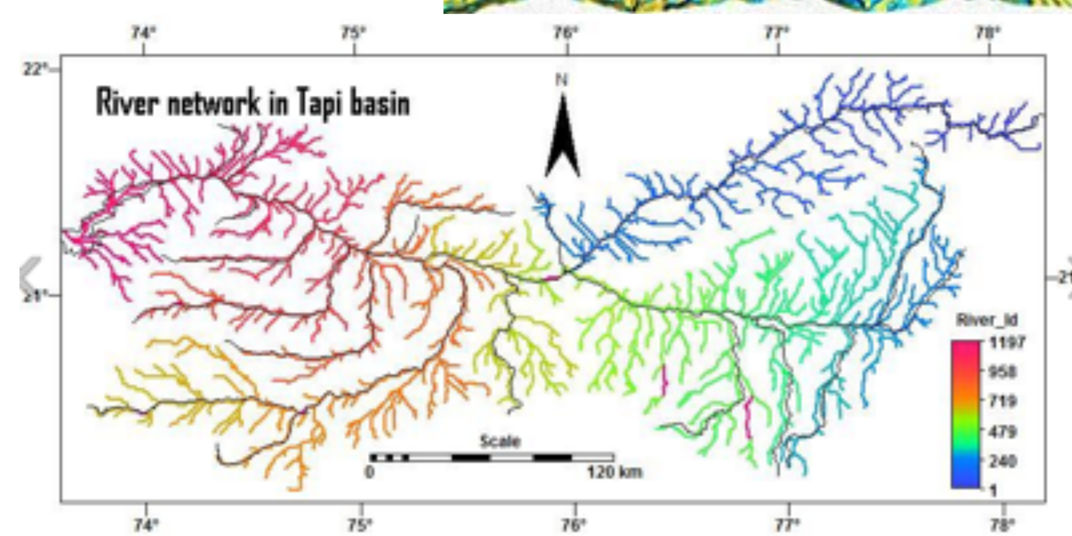
What can you do with GIS?

Terrain analysis

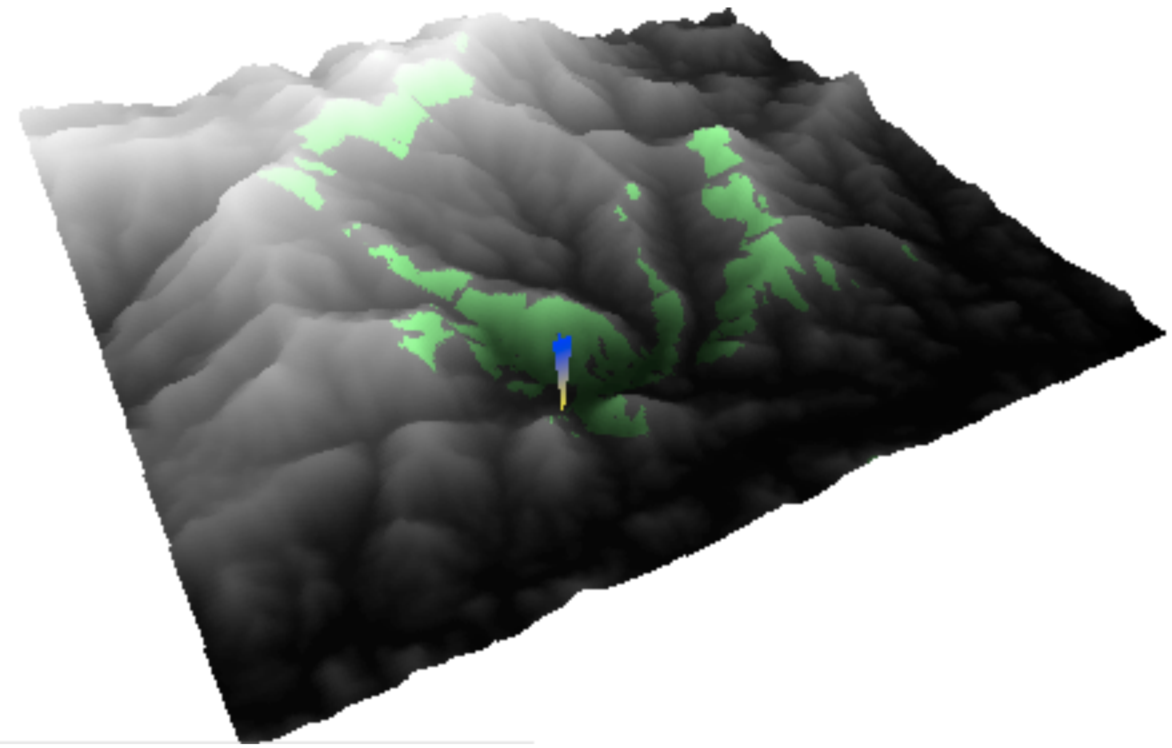




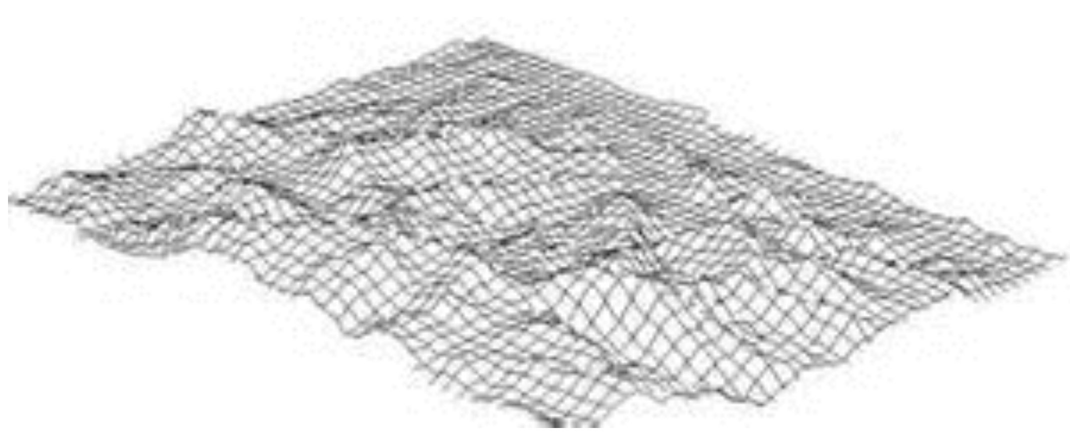


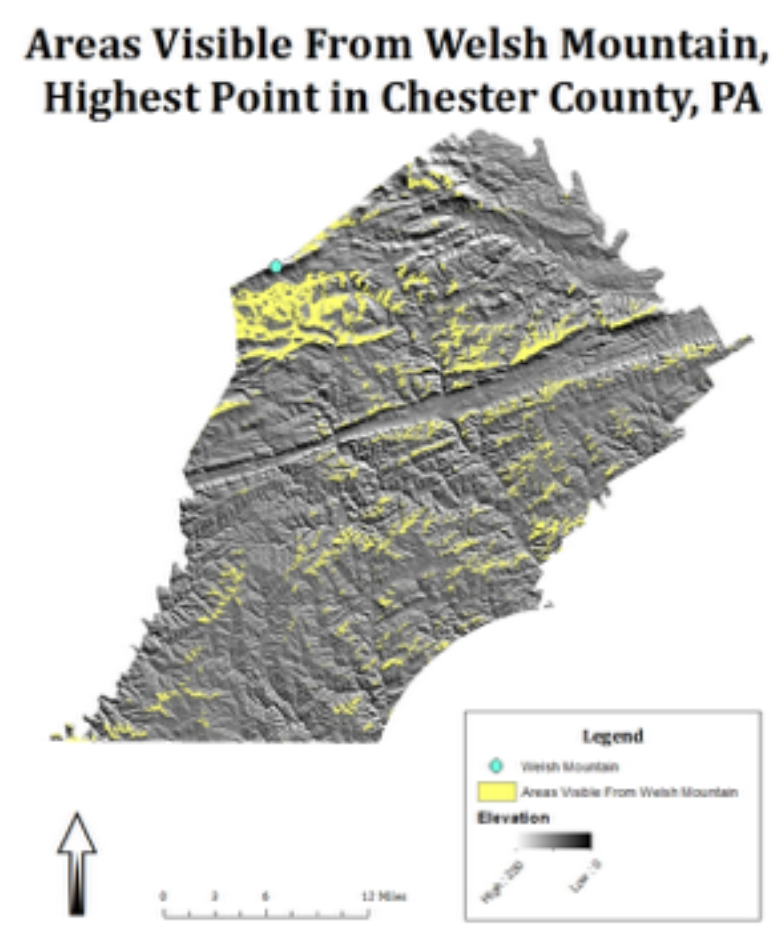


What can you do with GIS?



Terrain analysis



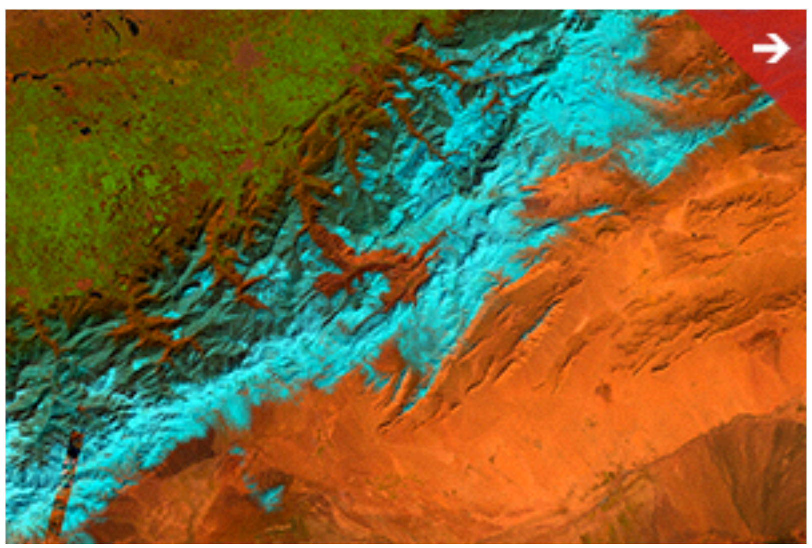


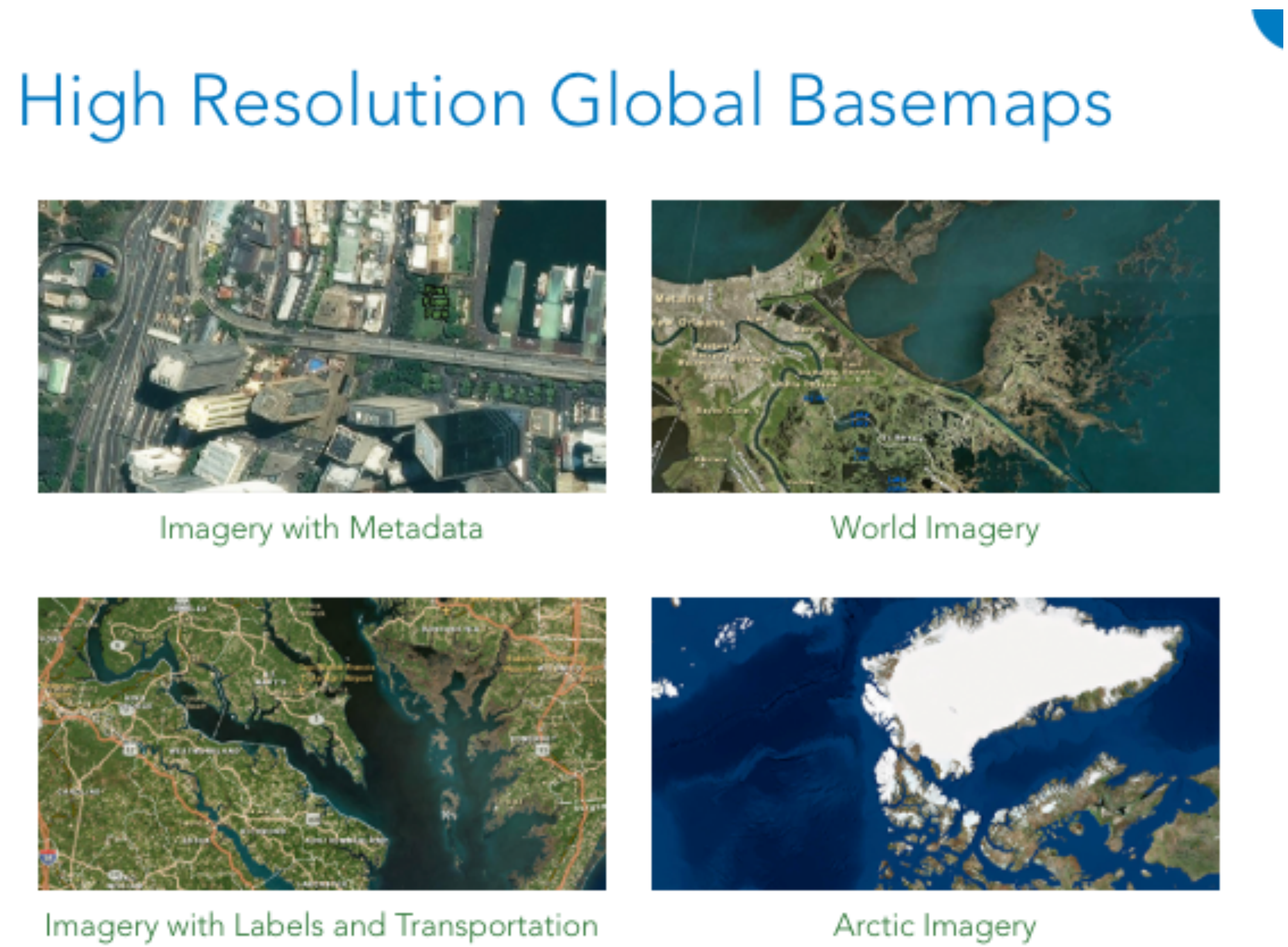
What can you do with GIS?

- 3D modeling
  - visualize data in 3D
  - E.g.: ESRI: 3d-gis

What can you do with GIS?

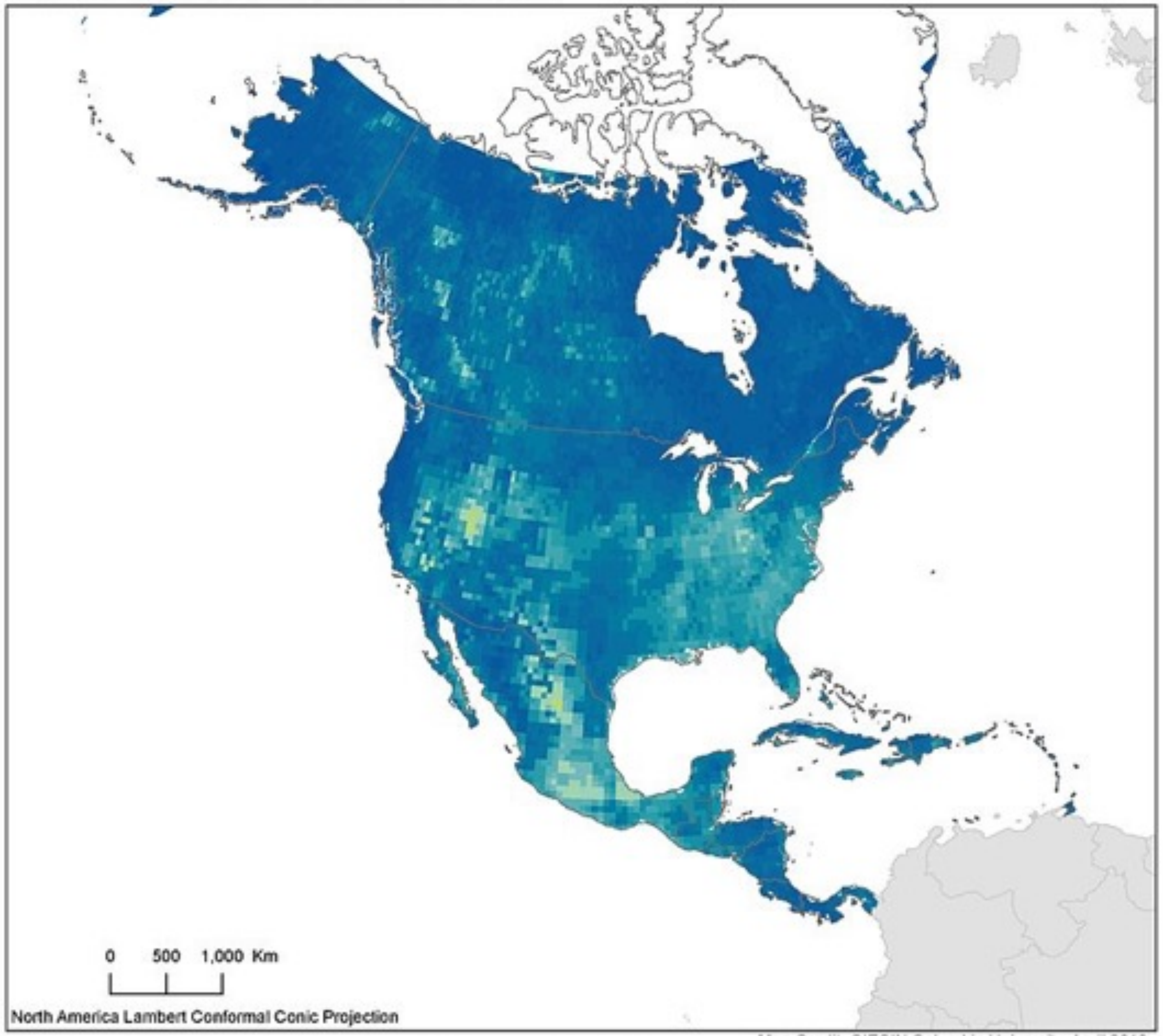
- Work with satellite imagery
  - access existing data collection (e.g. Landsat, Modis)
  - visualization and analysis



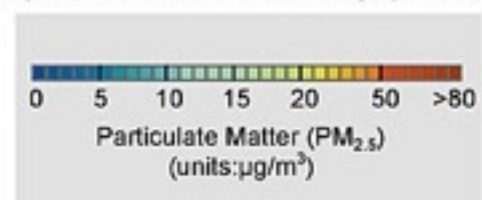


Global Annual Average $PM_{2.5}$ Grids from MODIS and MISR Aerosol Optical Depth (AOD), 2010: North America

Satellite-Derived Environmental Indicators



Map Credit: CIESIN Columbia University, April 2013.



Global Annual $PM_{2.5}$ Grids from MODIS and MISR Aerosol Optical Depth (AOD) data sets provide annual "snap shots" of particulate matter 2.5 micrometers or smaller in diameter from 2001–2010. Exposure to fine particles is associated with premature death as well as increased morbidity from respiratory and cardiovascular disease, especially in the elderly, young children, and those already suffering from these illnesses. The grids were derived from Moderate Resolution Imaging Spectroradiometer (MODIS) and Multi-angle Imaging SpectroRadiometer (MISR) Aerosol Optical Depth (AOD) data. The raster grid cell size is approximately 50 sq. km at the equator, and the extent is from 70°N to 60°S latitude.

Center for International Earth Science Information Network
EARTH INSTITUTE | COLUMBIA UNIVERSITY

Data Source: Battelle Memorial Institute, and Center for International Earth Science Information Network (CIESIN)/Columbia University. 2013. Global Annual Average PM2.5 Grids from MODIS and MISR Aerosol Optical Depth (AOD), 2001–2010. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://sedac.ciesin.columbia.edu/data/set/sdei-global-annual-avg-pm2-5-2001-2010.

What can you do with GIS?

- Real-time GIS
  - Analyze real-time sensor data and put it on interactive maps for real-time decision making
  - Track dynamic assets such as vehicles, aircrafts and vessels

Who Uses Real-Time GIS?

From http://www.esri.com/products/arcgis-capabilities/real-time-gis

Real-time GIS is used by many organizations. Here's a few examples:

Dong Energy, a leading renewable energy group, will be operating more than 1,800 offshore wind turbines by 2020.

DHL Express is one of the world's largest shippers, handles more than 30 million items with a fleet of about 1,500 vehicles.

Port of Rotterdam, one of the busiest ports in the world, handles more than 130,000 vessels and processes 440 million tons of cargo.

OSIsoft, a global leader in operational intelligence, enables businesses to leverage sensor-based data across their enterprises.

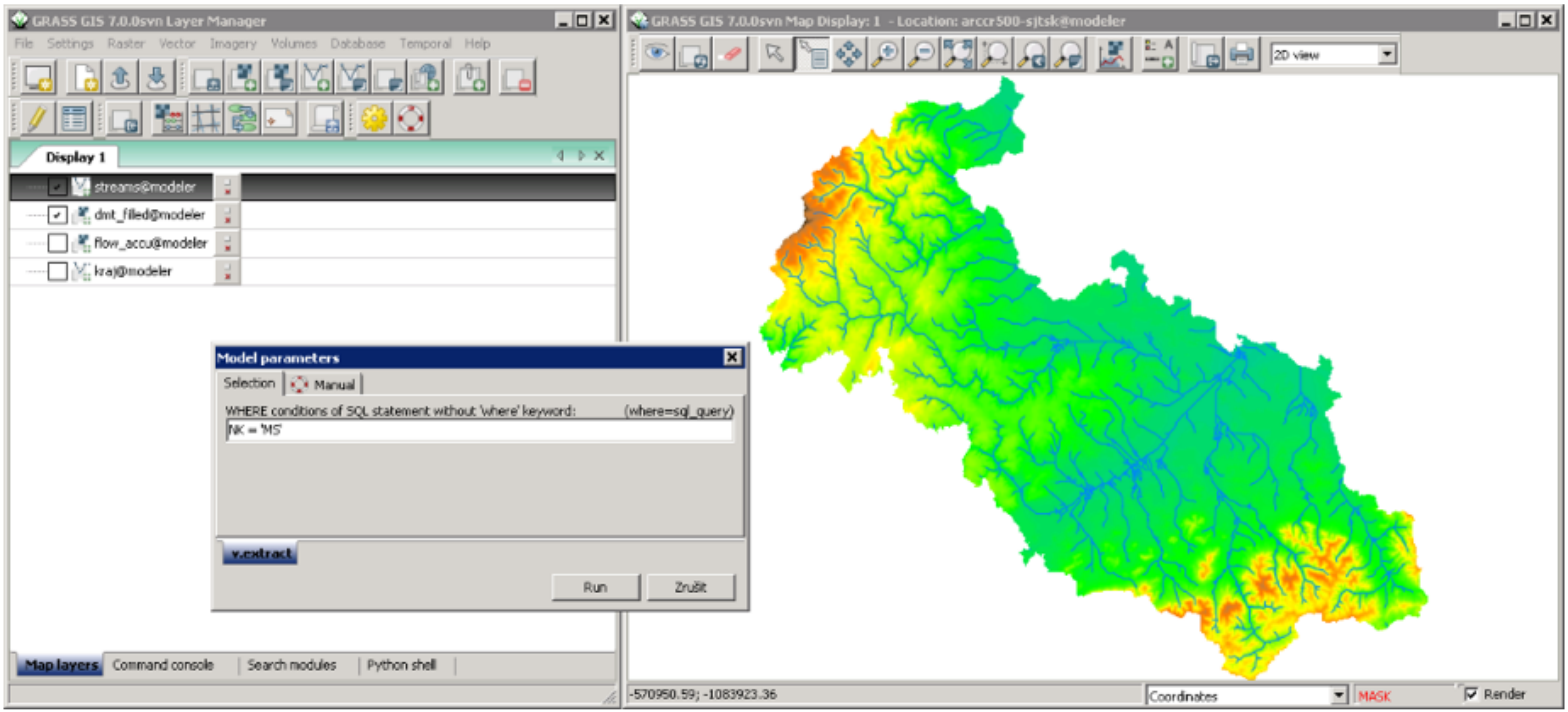
AccuWeather, the leading global weather info provider, uses tens of thousands of new data elements per hour to predict weather patterns.

Valarm, a startup, creates apps and sensor-enabled devices that do everything from optimizing truck routes to tracking humidity in vineyards.

What can you do with GIS?

GRASS: module of the day

v.overlay: overlays two vector maps, offering clip, intersection, difference and union operators

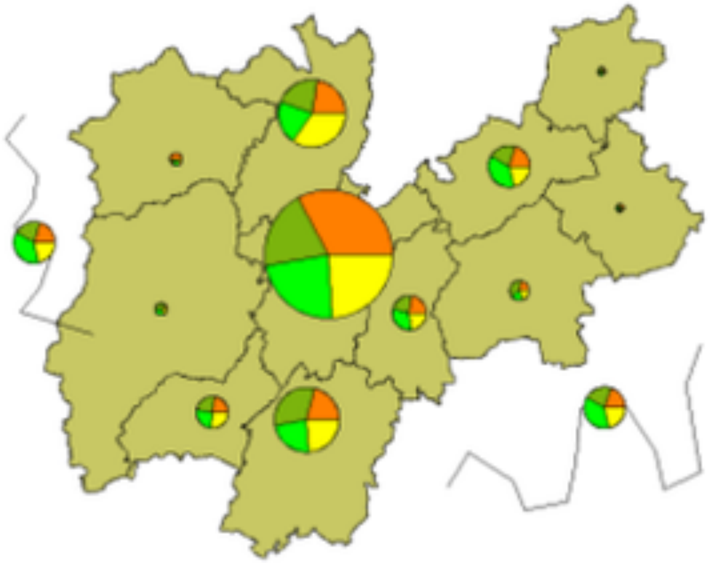


What can you do with GIS?

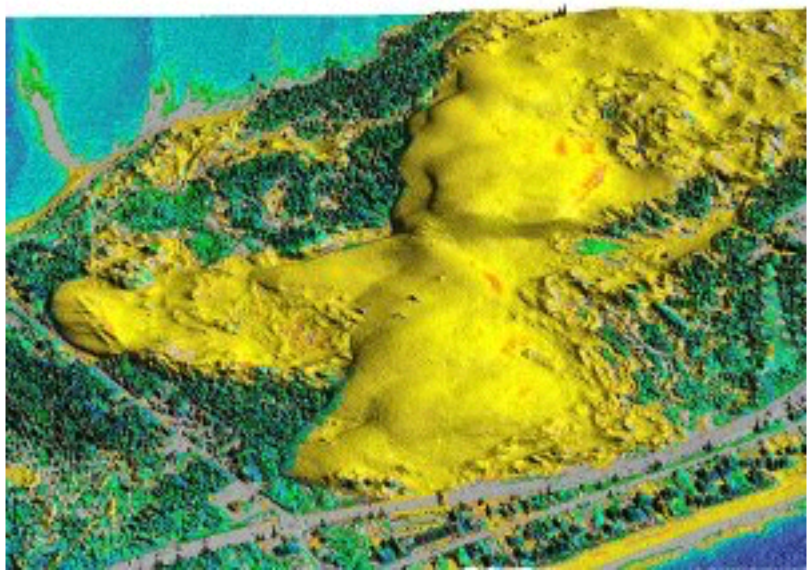
GRASS screenshots

Raster map operations

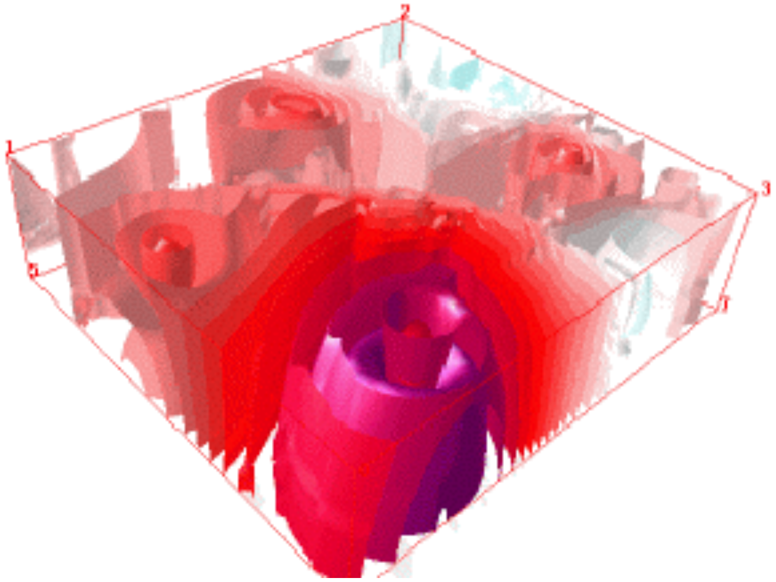
Vector map operations



LiDAR data Processing



3D Visualization



Cartography

GIS

- Used by a growing number of disciplines
  - earth, atmospheric and oceanographic sciences
  - environmental studies
  - digital humanities, …
- Also used by city planners, government, …

With the explosion of digital data ==> GIS has seen tremendous growth

Existing GIS software

- **ArcGIS**
  - largest system, developed by ESRI
  - complex interface
  - very comprehensive
  - available in Bowdoin labs ; IT/ES offer tutorials

- Open source systems
  - e.g. **GRASS**, **QGIS**

- Other proprietary modules, with specialized functions
  - e.g. **LAStools**

Studying GIS

- Users
  - Navigate the functionality provided by {ArcGIS, GRASS, QGIS,…}
  - Use existing functions to model and solve problem
- Research
  - GIS = rich source of problems in CS, spanning from theory, to databases, to interfaces, visualization, cloud computing and systems
  - Add new functions
  - Improve performance

Class overview

- This class explores fundamental GIS applications and the algorithms and data structures involved.
- As the size of data in GIS keeps growing, the design of efficient algorithms becomes more important. Focus will be on the interplay between theory and practice and scalability to large data.
- Topics:
    - Data models (raster, vector, TIN) and representations
    - Flow modeling on terrains and applications (river network, watersheds, basins, flooding, sea-level rise).
    - Visibility on terrains and applications: viewsheds, multiple viewsheds, guarding, and approximation.
    - Simplification: 2D (line simplification) and 3D (terrain simplification).
    - LIDAR data.
    - Spatial data structures: B-trees and quadtrees, and applications.
    - Space-filling curves.
    - Dealing with large data: IO-efficient algorithms, COB-algorithms, parallel algorithms

Class info

- No textbook. Papers, slides, and other online materials.
- Pre-requisites: 1101, 2101 and 2200 (algorithms)
- CS curriculum:
  - Projects class
  - Satisfies the “theory” requirement
- Collaboration policy:
  - For the programming assignments I encourage you to find a partner (working alone is fine too).

Work for the class and grading policy

- The work for the class consists of
  - programming assignments, culminating with a final project
  - reading research papers, writing reviews, class discussions
  - class presentations (such as project proposal, updates, and final presentation)
  - a project report
- There will be no exams
- The final project will occupy the last 3 weeks of the class and there will be presentations and reports associated with it.
- All programming is in C/C++
- The final project demo is on December 18th (check exact time in polaris)

Expectations

- Attendance in class is strongly encouraged and will be weighed in the final grade. Generally speaking, I hope that not only you will show up in class, but also contribute to the learning experience.

- Be engaged in class and participate in group work.
- This is a 3000-level class and everyone's programming background varies. I am aiming the assignments somewhere in the middle. I expect that you'll be pro-active and you'll come talk to me about tailoring the assignments to your experience.
  - If the assignments seem overwhelming and you don't know where to start, you need to come talk to me.
  - If you have more experience and you finish the assignments fast, you need to come talk to me about ways to extend
- Give me feedback!

My research

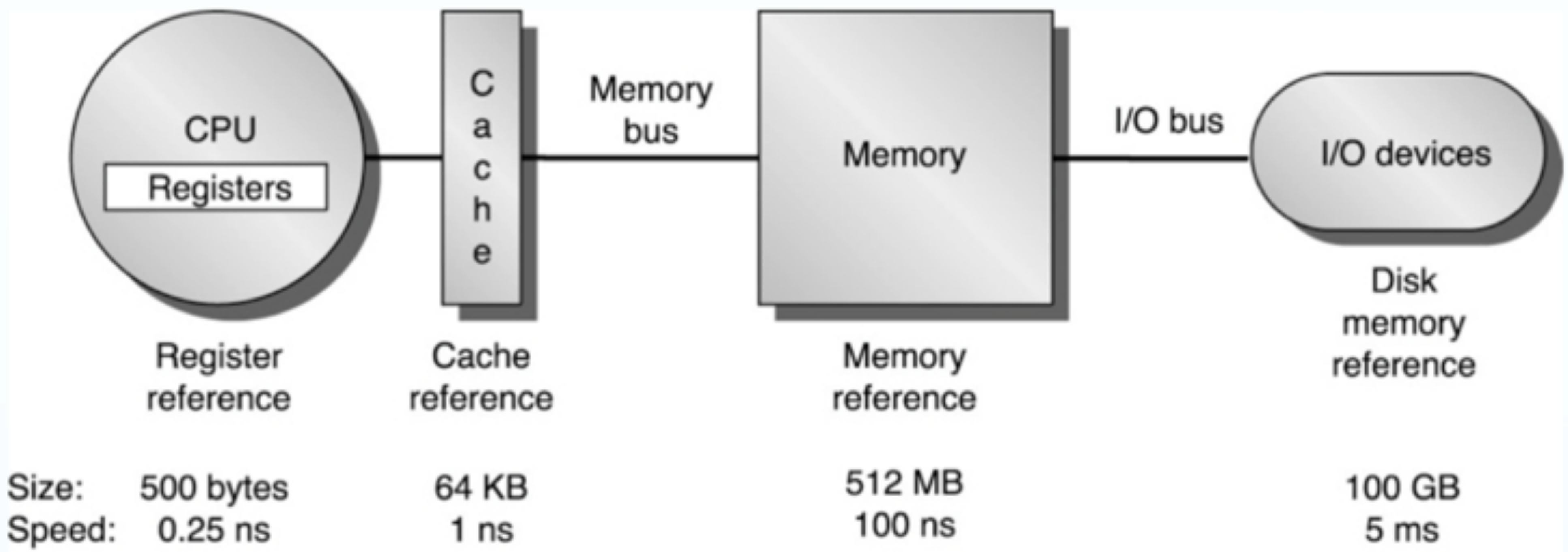
Terrain modeling in GIS

Algorithms for large data

Algorithm engineering

High performance computing

Large data


CPU
Registers
C a c h e
Memory bus
Memory
I/O bus
I/O devices
Register reference
Cache reference
Memory reference
Disk memory reference
Size: 500 bytes
Speed: 0.25 ns
64 KB
1 ns
512 MB
100 ns
100 GB
5 ms

High performance computing