GEOSPATIAL INNOVATION FACILITY

Cutting-Edge Mapping Technology at UC Berkeley

Our Mission

Help people better understand the changing world through acquisition, analysis and visualization of spatial data

The GIS landscape has been changing:

- New massive data streams – from phones, from APIs, from new satellites
- Technological advances in computing- advances in database integration, cluster computing, big data, more choices in coding
- More public focus on location tech and geo

All this requires a new framework: one that is *computing-intensive, data-rich* and *collaboration-focused*.

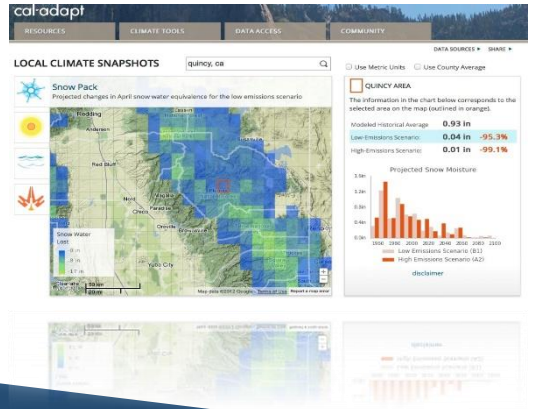
  - Broader/different scales of interest – more personal and more broad
  - New ways of analyzing and synthesizing data
  - Multiple users and collaboration
  - Agile visualization

Data in the Age of Sharing

The breadth, transparency and low comparative cost of exchanging information on the Internet provides an efficient form of communication between the public and planners, managers, and decision-makers.

Mapping makes these exchanges more powerful.

web mapping/visualization



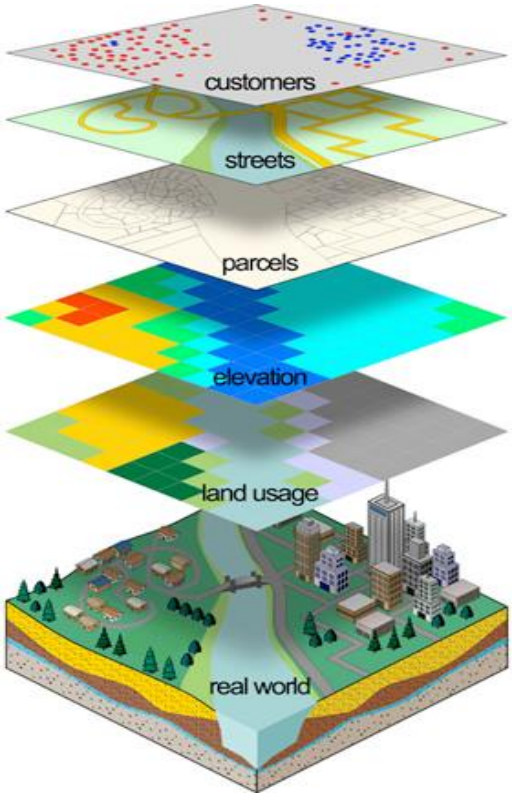
citizen involvement

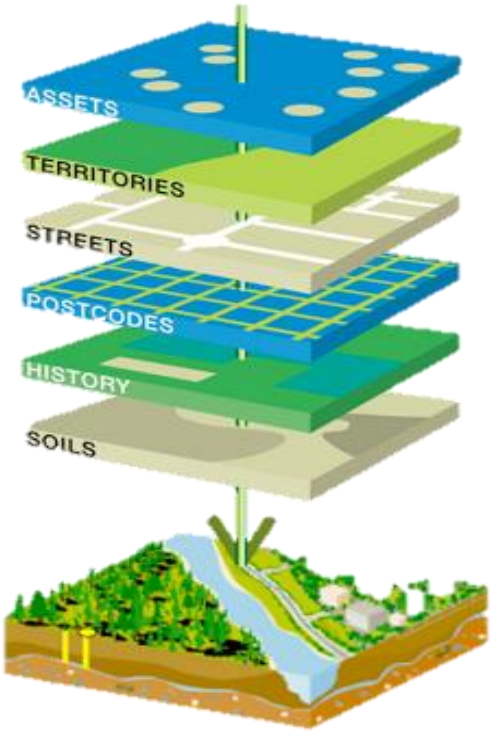
disaster response

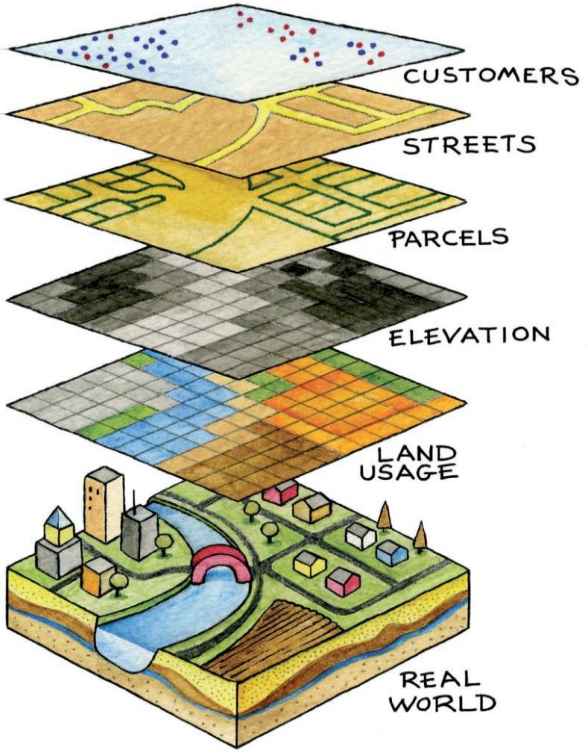
Berkeley UNIVERSITY OF CALIFORNIA College of Natural Resources

GEOSPATIAL INNOVATION FACILITY
Cutting-Edge Mapping Technology at UC Berkeley

The GIS Model: the 20th Century Mapping Toolkit

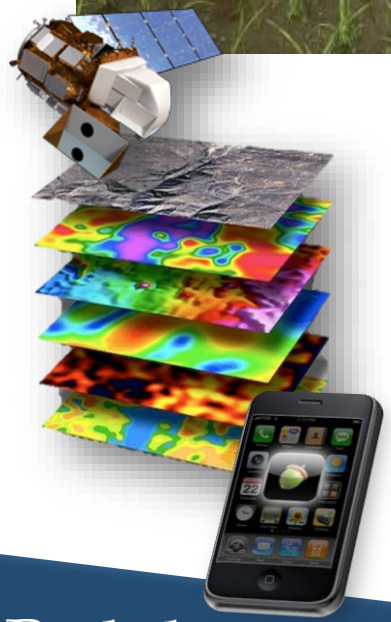






We Need a 21st Century Mapping Toolkit



Many of the global challenges that we face today – such as poverty, food and water scarcity, sustainable development, urbanization, and climate change – are large in spatial scale and impact diverse public groups.

Addressing these challenges requires a new framework: one that is ***computing intensive***, ***data-rich***, and ***collaboration-focused.***

Berkeley UNIVERSITY OF CALIFORNIA College of Natural Resources

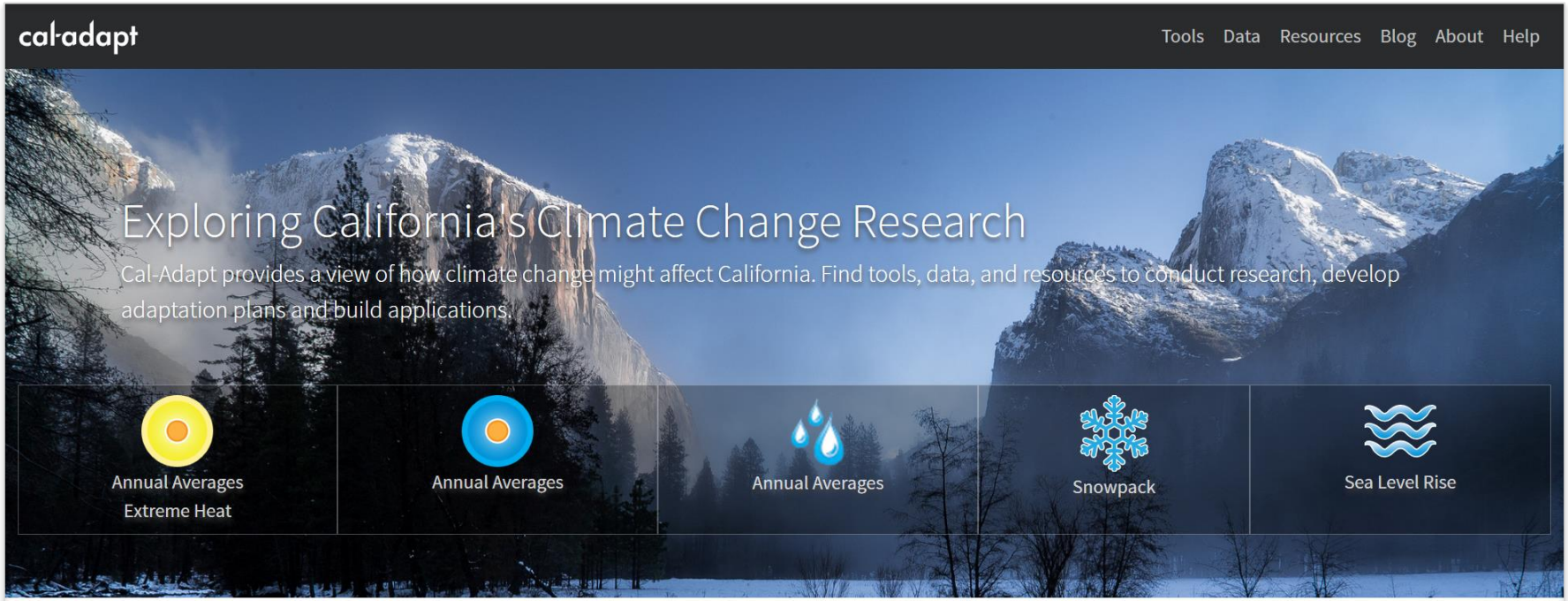
GEOSPATIAL INNOVATION FACILITY
Cutting-Edge Mapping Technology at UC Berkeley

Spatial Data Science

Importance of our research being:

- Reproducible
- Repeatable
- Collaborative
- Meaningful

Open source tools allow us to achieve these goals more easily



Climate Tools

Explore projected changes in temperature, precipitation, snowpack and sea level rise in California over this century with our interactive climate data visualizations.

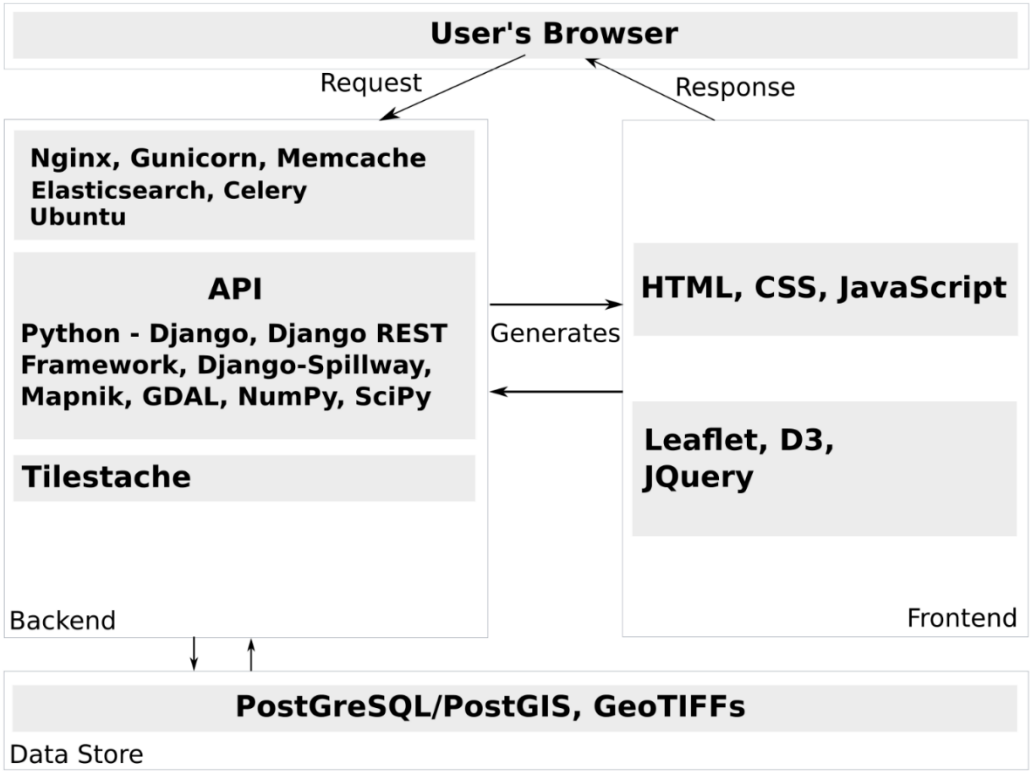
Download Data

Download high resolution downscaled daily, annual and monthly climate projections for your project area in NetCDF or GeoTiff formats.

Find Resources

Search State of California's Research Catalog, explore peer-reviewed publications, understand how to use climate projections.

GIF Open Source API Architecture



Spatial Data Science Bootcamp

Set Up Your Environment

- Virtual machine environments (Linux-based)
- Spatial databases (PostgreSQL/PostGIS)

Wrangle Data

- Modern data standards and formats (GeoJSON)
- Open mapping tools (GDAL, QGIS)

Analyze Data

- Python (i.e. PySAL, NumPy, Pandas, Jupyter Notebook)
- R and R Studio (i.e. sp, raster, rgdal, maptools)

Visualize and Publish Data

- Web mapping and data visualizations (CARTO, Leaflet, D3)
- Data visualization in Python

Who you are!

Intro to Open Source Geo Tools

- Geospatial fundamentals
- Why open source?
- How do I get started using geospatial open source tools?
- Intro to QGIS and hands-on exercise

What is Geographic Data?

Data - descriptions collected through observation

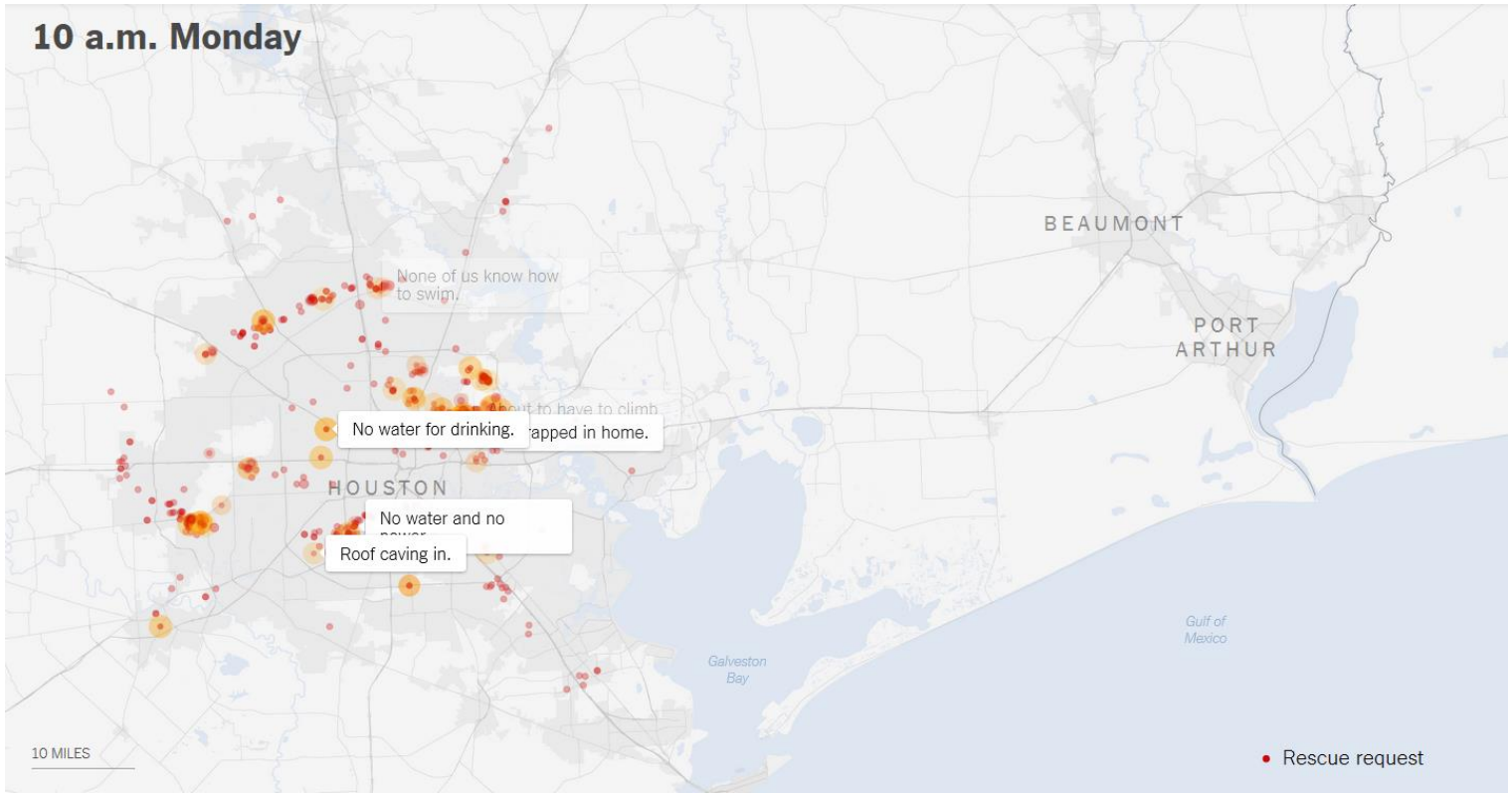
Geographic - about locations on or near the surface of the Earth

Why Geospatial?

Digital maps are everywhere & changing the way we do things

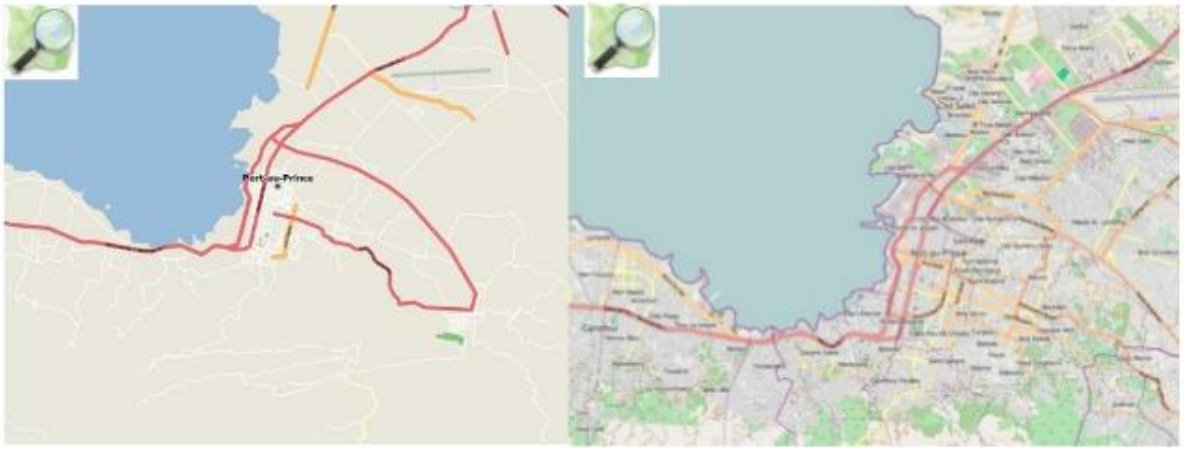
Houston flooding

Disaster Response



New York Times: August 30, 2017

Volunteered Geographic Information

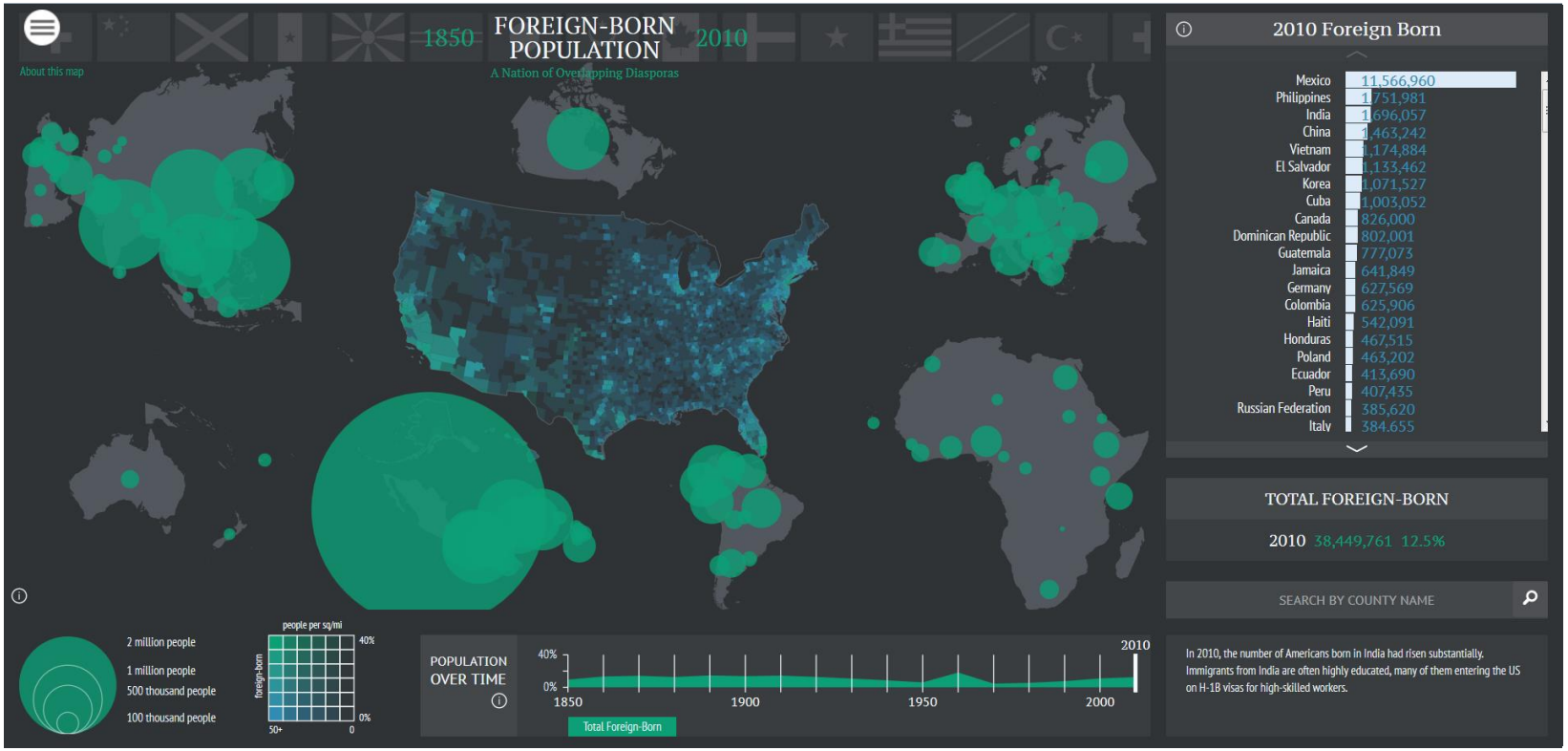


Port-au-Prince on OSM, January 12, 2010

Port-au-Prince on OSM, 28 days later

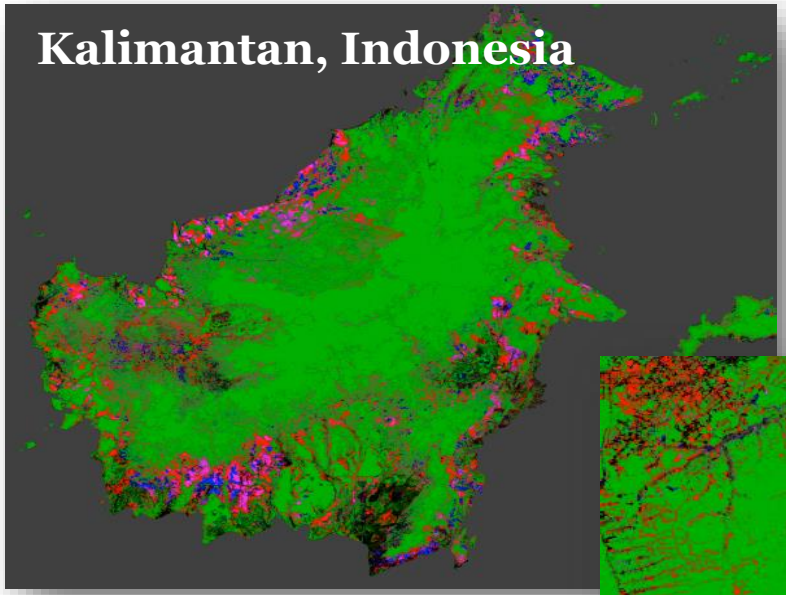
Within a few days, the response of the OSM community to map the affected areas has been intensive, as seen in this video.

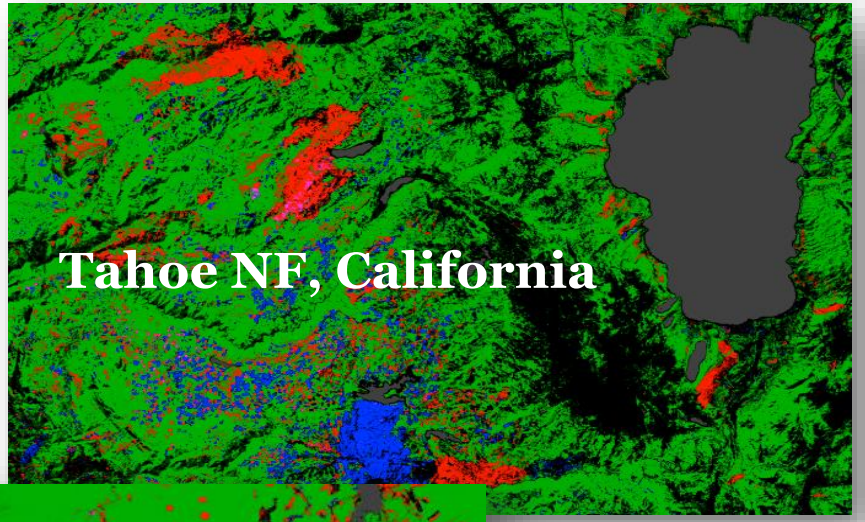
Web Mapping and Visualization



- **American Panorama:** http://dsl.richmond.edu/panorama/

Large Scale Cloud Computation: Google Earth Engine





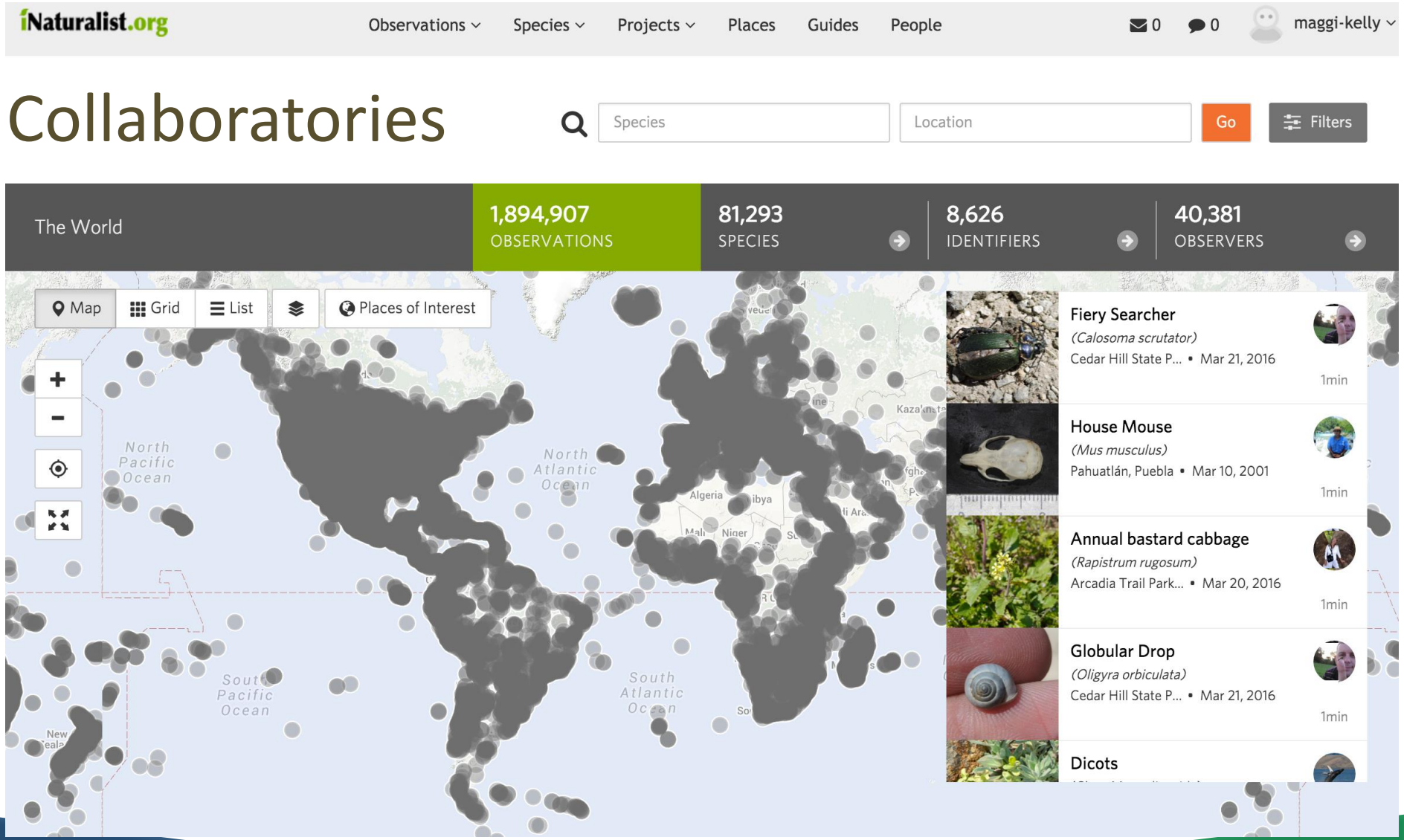
Forest Loss 2000–2012
Forest Gain 2000–2012
Both Loss and Gain
Forest Extent

10 years,
1.3M Landsat scenes,
Cloud processing power

Berkeley UNIVERSITY OF CALIFORNIA College of Natural Resources

GEOSPATIAL INNOVATION FACILITY
Cutting-Edge Mapping Technology at UC Berkeley

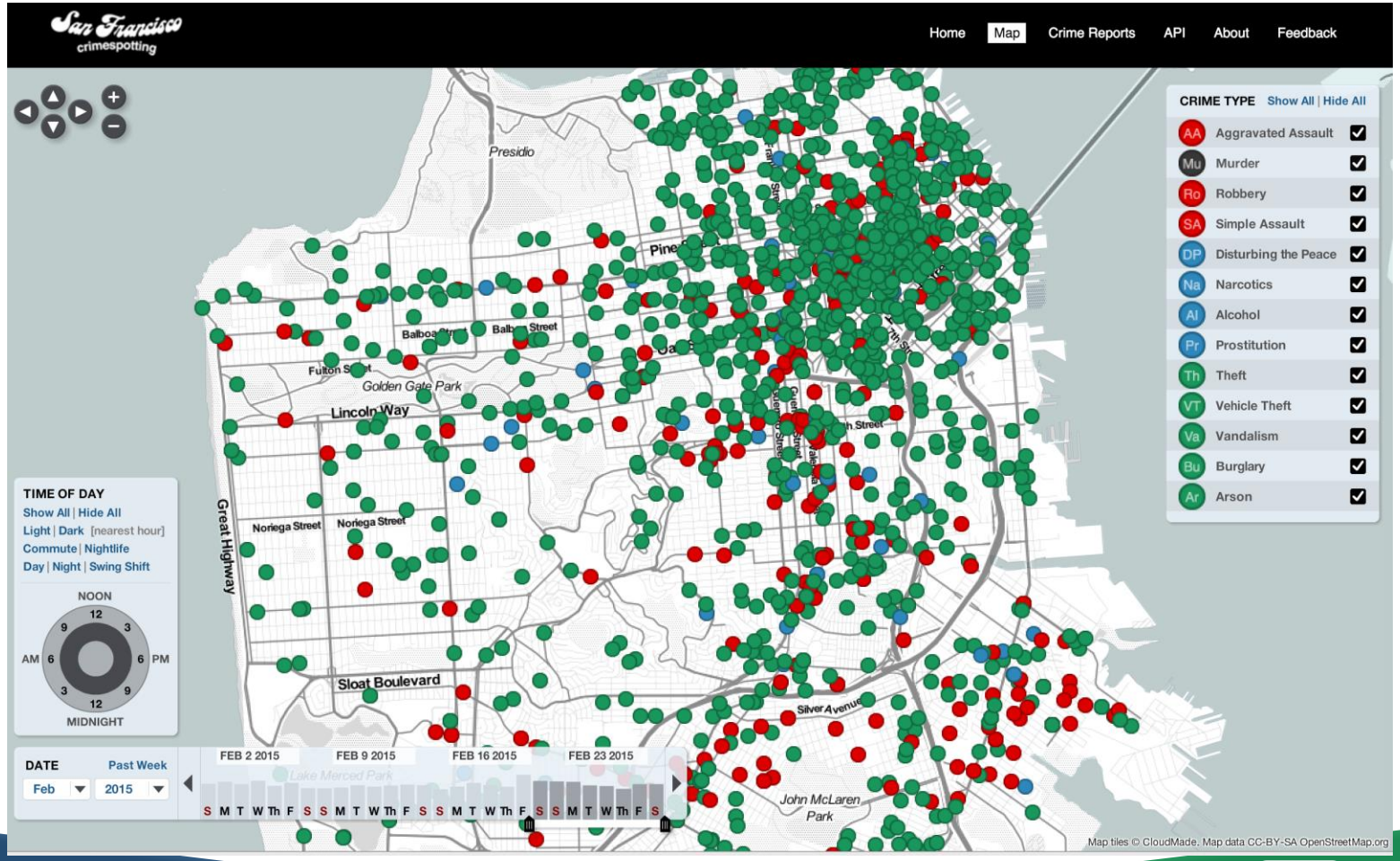
Collaboratories



Berkeley UNIVERSITY OF CALIFORNIA College of Natural Resources

GEOSPATIAL INNOVATION FACILITY
Cutting-Edge Mapping Technology at UC Berkeley

Open Government Data

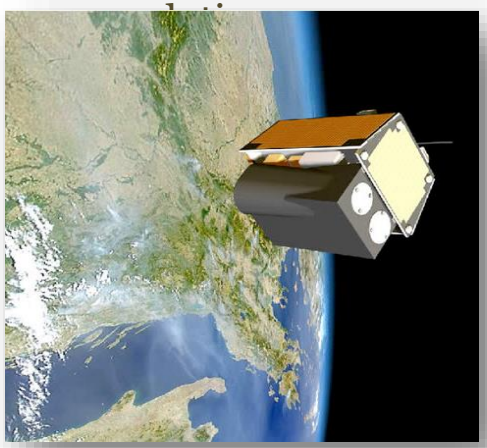


Berkeley UNIVERSITY OF CALIFORNIA College of Natural Resources

GEOSPATIAL INNOVATION FACILITY
Cutting-Edge Mapping Technology at UC Berkeley

Sensor Trends: Drones, Micro Satellites

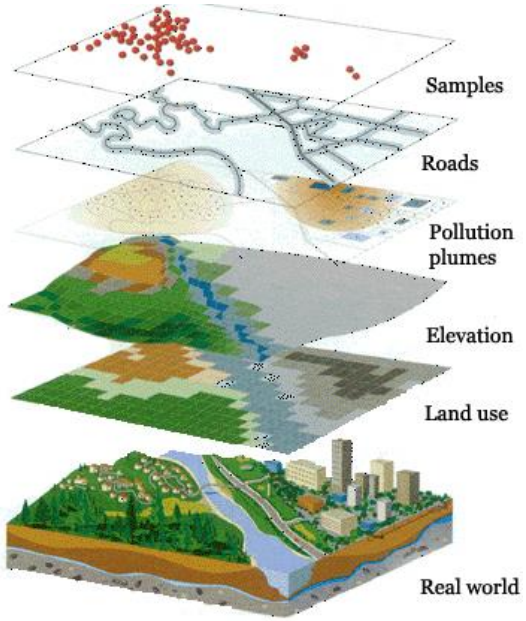
New sensors, smaller



Fundamentals of Geospatial Data & Analysis

What is GIS?

A geographic information system (**GIS**) is a computer-based tool that **links geographic information** (where things are) with **descriptive information** (what things are).



GIS Layers

Real World

Monitoring Wells

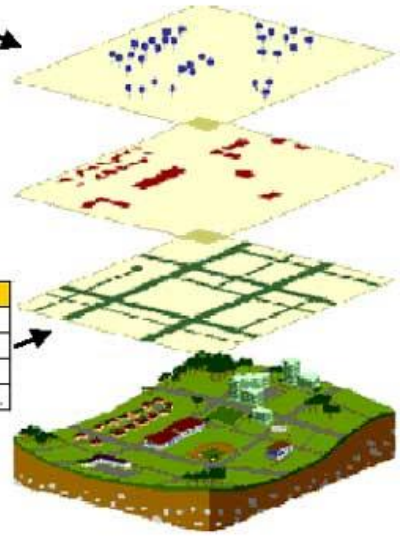
Well ID	Date Sampled	Concentration
C-6A	5/8/94	300
C-8A	5/8/94	20
C-13A	5/8/94	120
C-17A	5/8/94	560

Industries

Facility	Address
Acme	3029 Convington Dr.
Fox	742 West Lake St.
TPC	90 Aspen Dr.

Population

Family Name	Occupants	Addresss
Blake	6	79 Circuit St
Hernandez	2	148 Plain St.
Joy	4	18 Webster St.
Smith	5	4321 Tecumseh Dr.

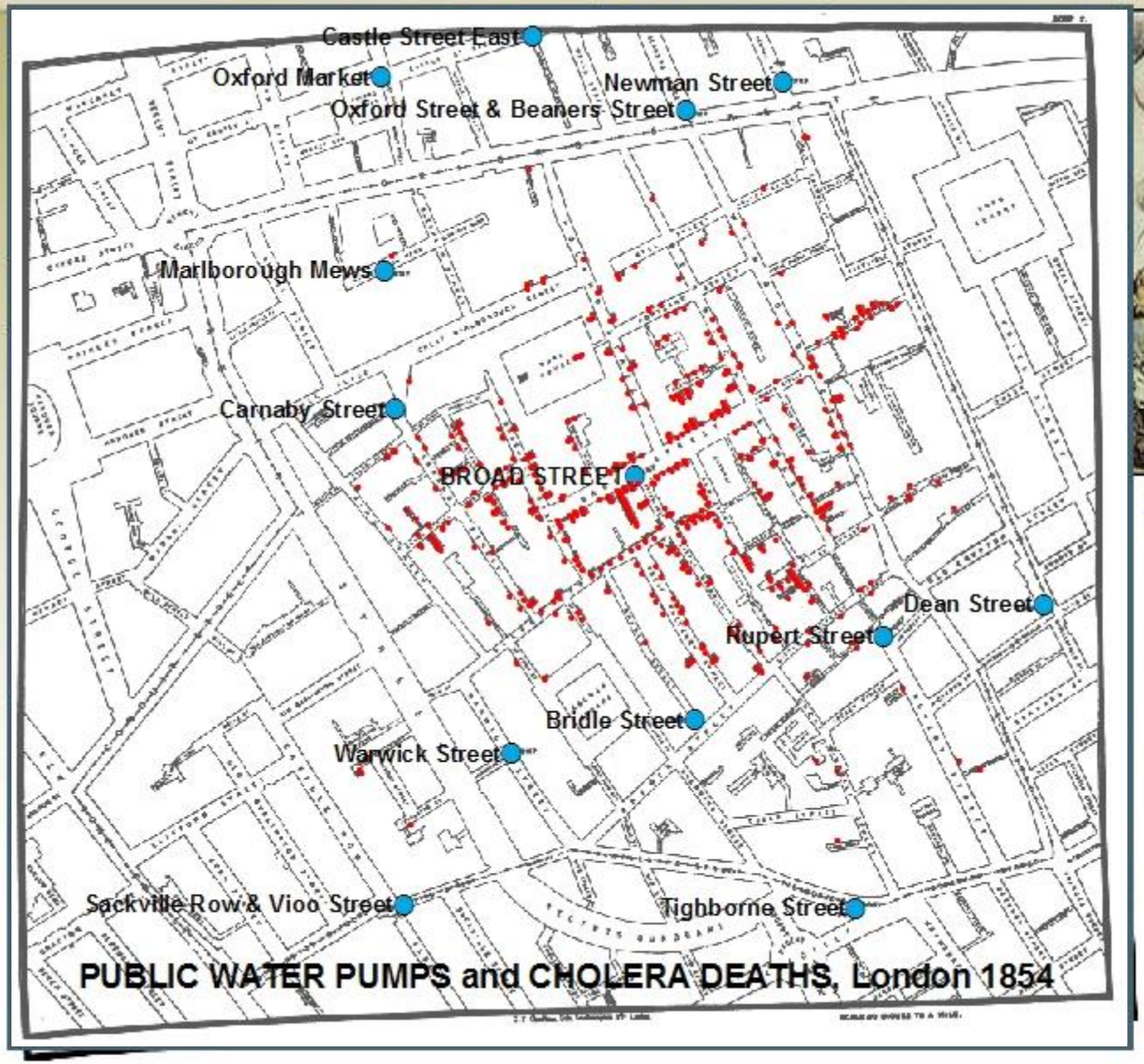


Berkeley UNIVERSITY OF CALIFORNIA College of Natural Resources

GEOSPATIAL INNOVATION FACILITY
Cutting-Edge Mapping Technology at UC Berkeley

Image credit: ESRI

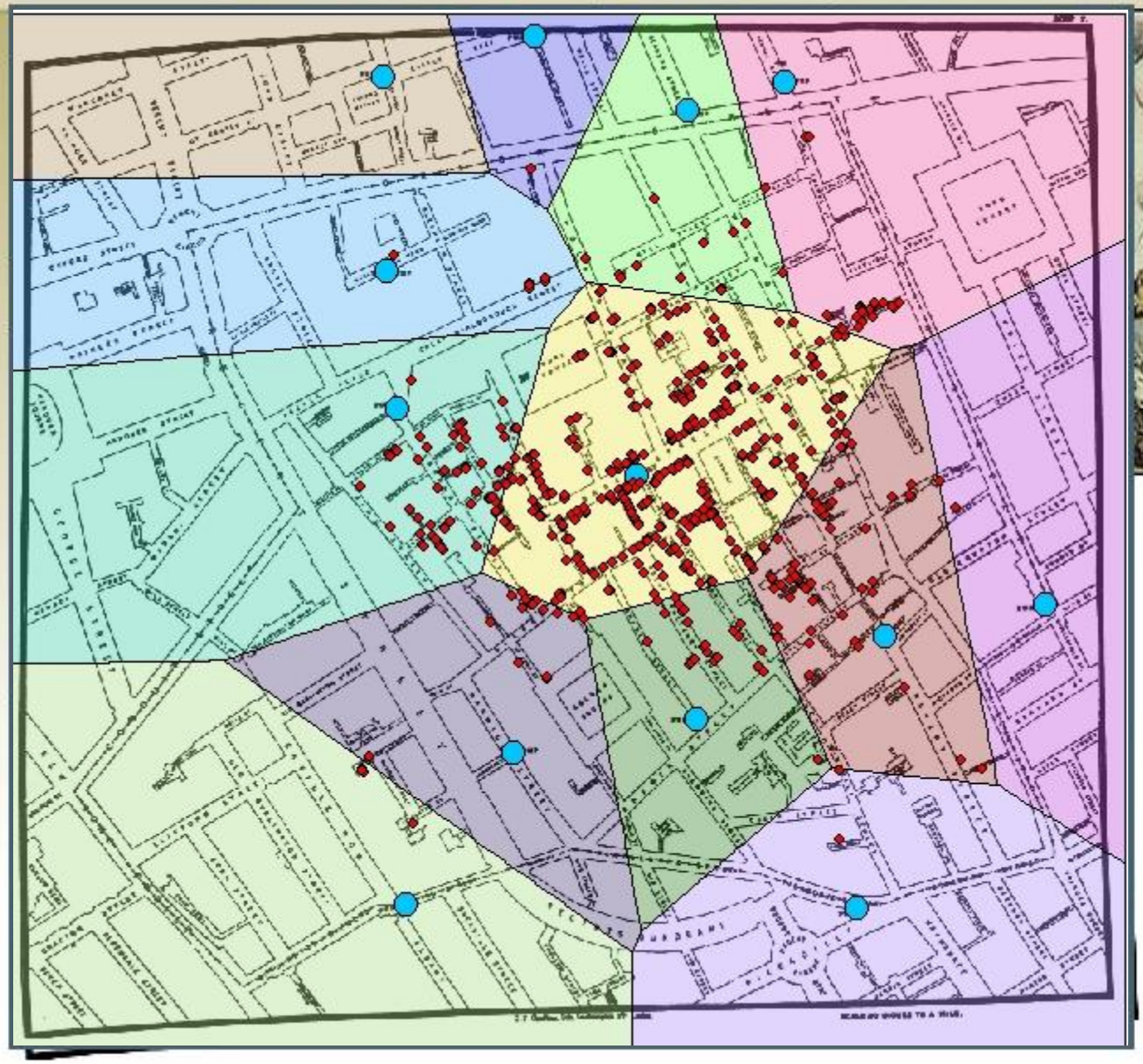
Beginnings of GIS – Snow Map



"Miasma"

http://www.udel.edu/johnmack/frec682/cholera/cholera2.html

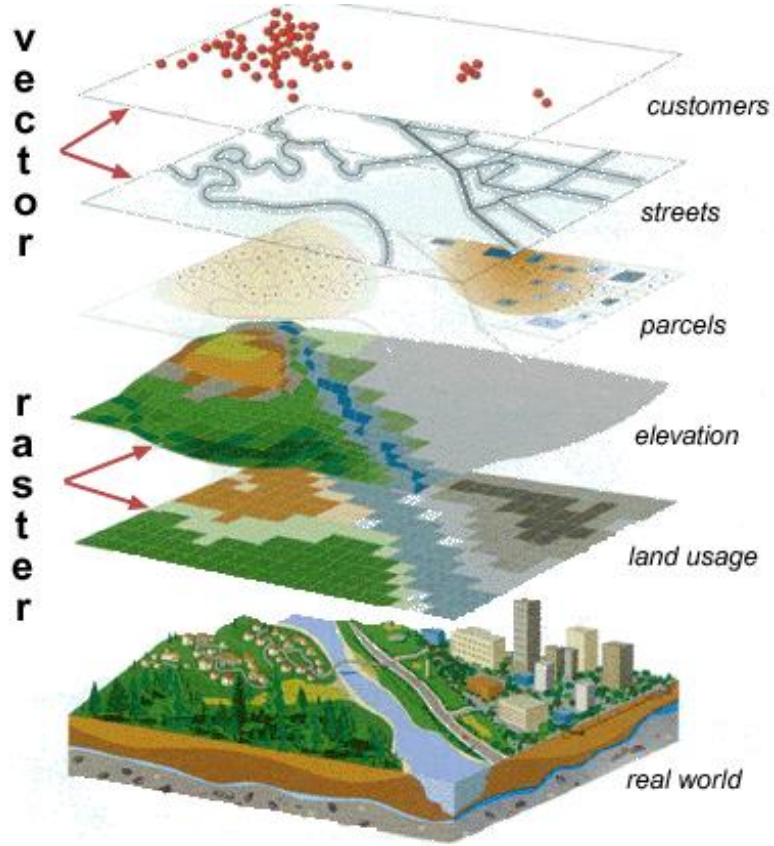
Beginnings of GIS – Snow Map



"Miasma"

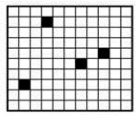
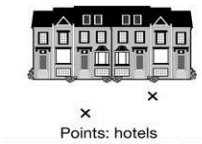
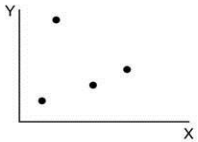
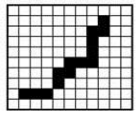
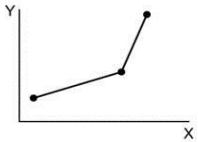
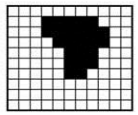
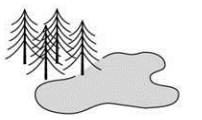
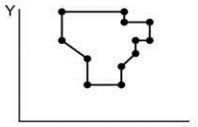
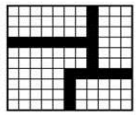
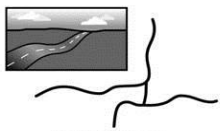
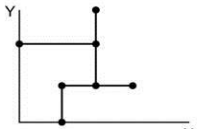
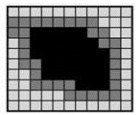
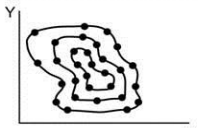
How GIS Works

A GIS stores information about the world as a collection of thematic layers that can be linked together by geography



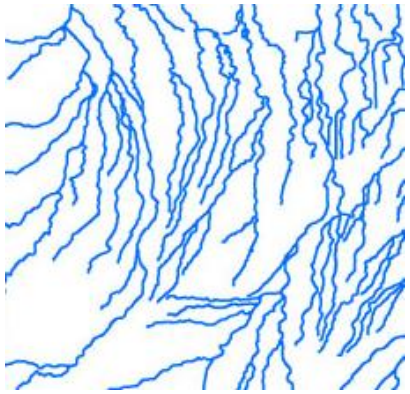
There are 2 basic spatial data types representing the real world:

Raster Vector

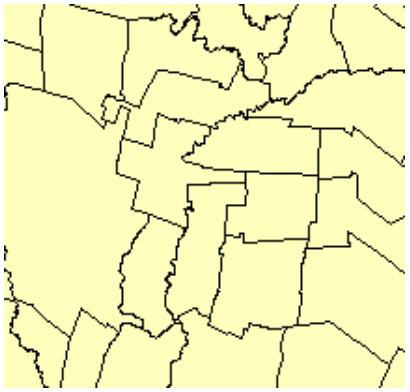
The raster view of the world	Happy Valley spatial entities	The vector view of the world
	 Points: hotels	
	 Lines: ski lifts	
	 Areas: forest	
	 Network: roads	
	 Surface: elevation	

Vector data examples
(points, lines, and polygons)

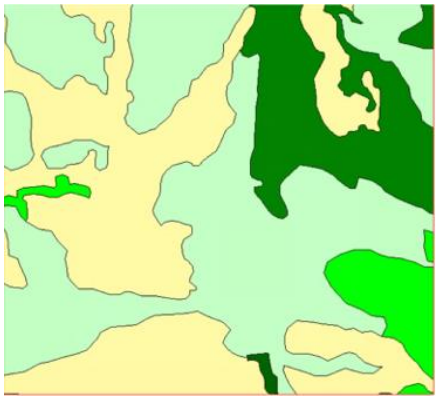
GPS data



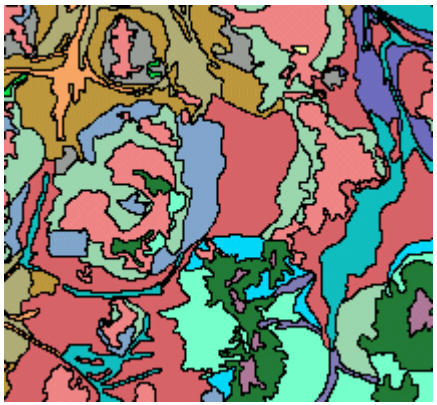
Rivers



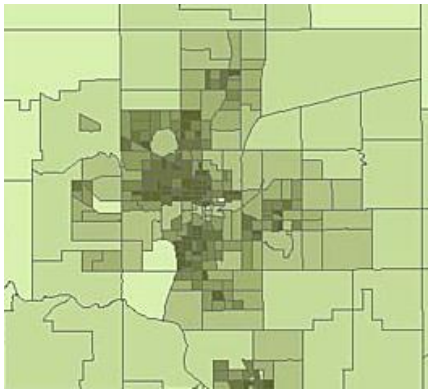
Counties



Habitat boundaries



Soil type

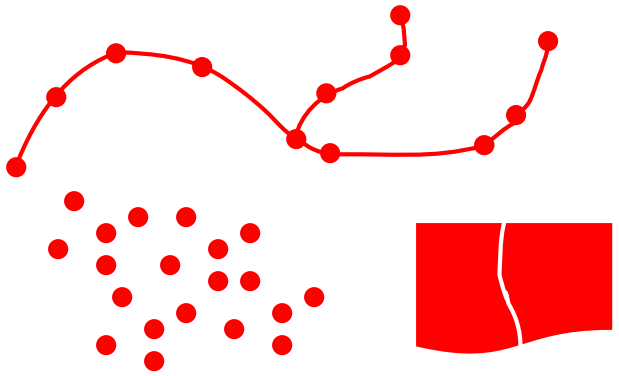


Census data

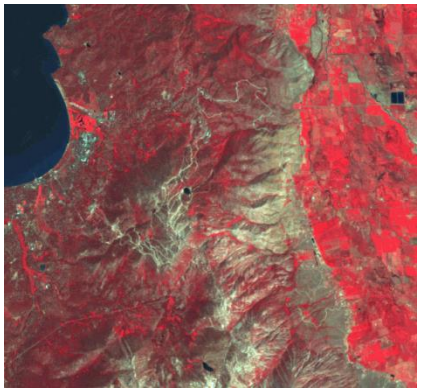
Shapefiles

File type extension	*What it is*	*Description*	*Example*
.shp	Main file	Stores each shape with a list of its vertices	counties.shp
.shx	Index file	Each record contains the offset of the corresponding main file record from the beginning of the main file	counties.shx
.dbf	dBASE table	Contains feature attributes with one record per feature	counties.dbf
.prj	Projection file	Stores projection information; doesn't define the data projection, only describes it. Recommended, but not mandatory.	counties.prj
.xml	Metadata file	Stores metadata information created by data creator/editor/distributor. Recommended, but not mandatory	counties.xml
.sbn	Unnecessary file, created automatically, doesn't need to be moved/copied/renamed		counties.sbn
.sbx	Unnecessary file, created automatically, doesn't need to be moved/copied/renamed		counties.sbx

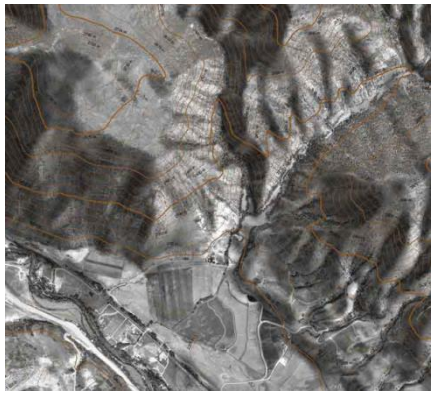
- Vector Data type
- Features: points, lines & polygons
- Attributes: size, type, length, etc.



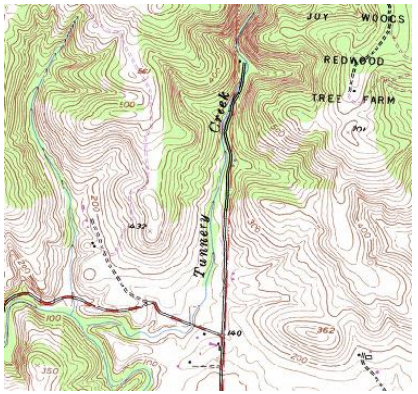
Raster data examples



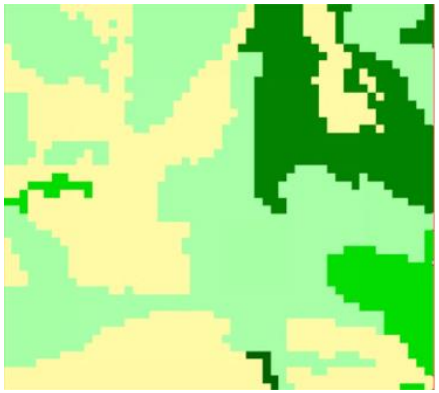
Satellite imagery



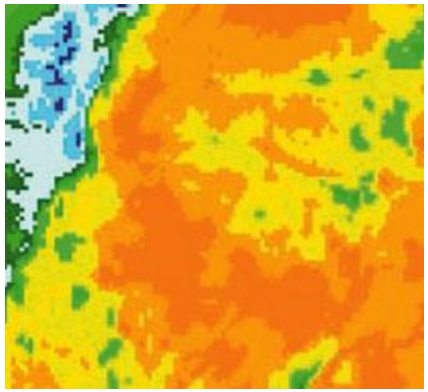
Elevation



Digital USGS topo map



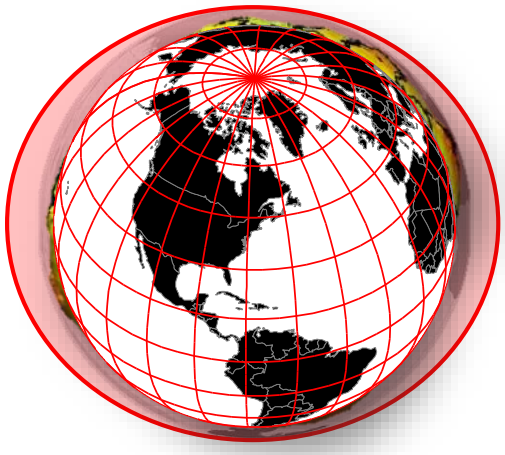
Landcover/landuse



Precipitation

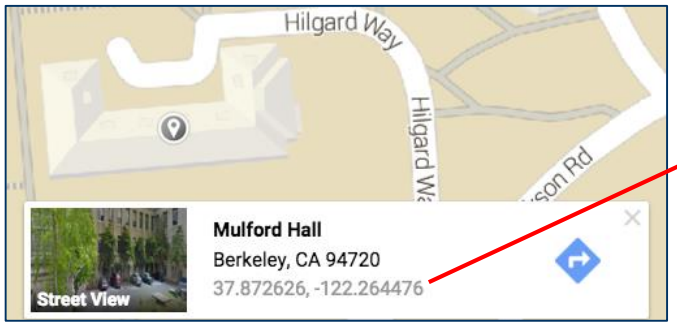
Aerial photography

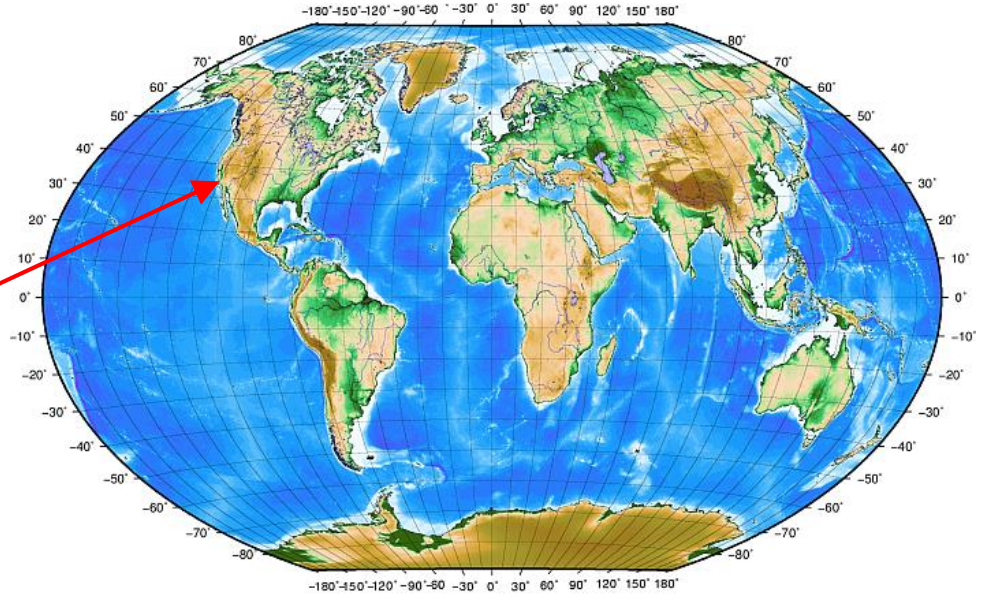
Coordinate Systems & Projections



Geographic Coordinate Reference System (CRS)

Specifies precise locations as **longitude** and **latitude** values.



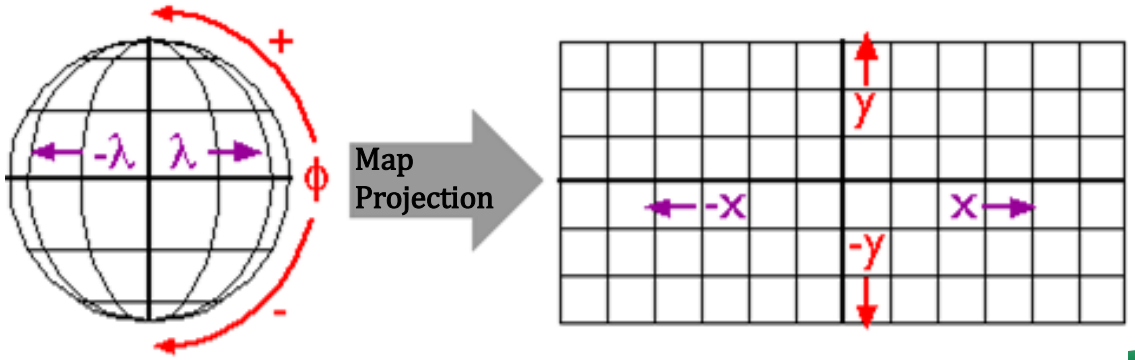


Map Projection

- **Map** = flat representation of the non-flat Earth
- **Map projection** = mathematical transformation from 3D surface to 2D plane.

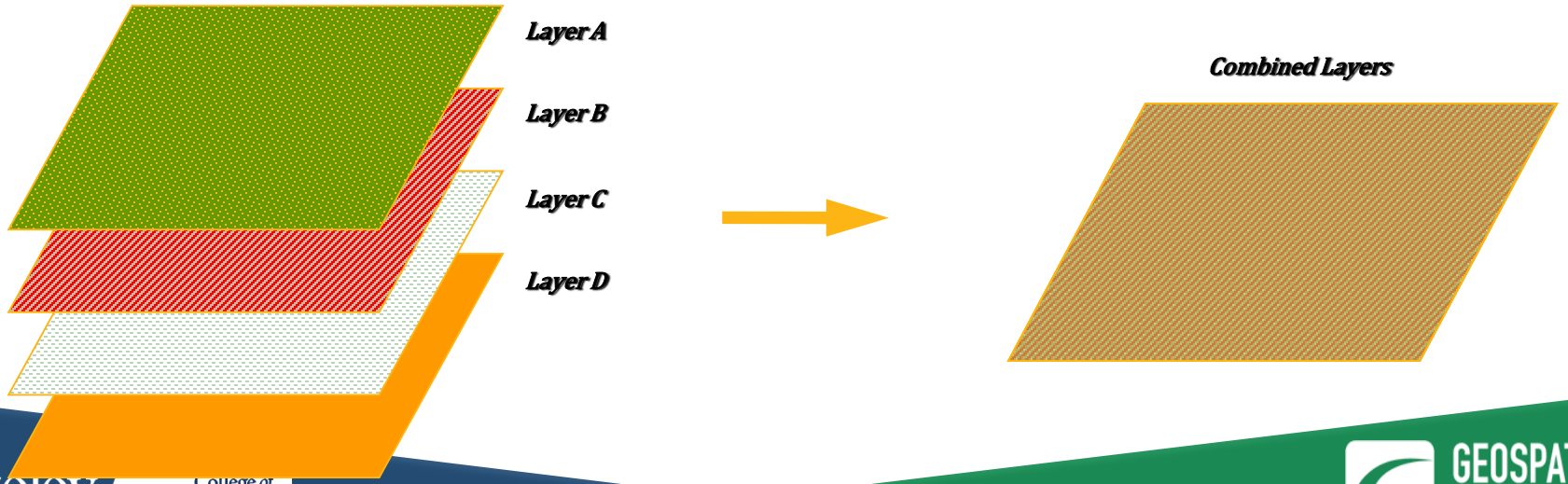
Longitude λ > X

Latitude ϕ > Y



Spatial Analysis

- A critical function in GIS analysis is the integration of datasets
  - determining where different geographic phenomena are coincident
- Overlay tools are powerful analysis tools for this
  - vertical stacking and merging of spatial data
  - combine, erase, modify features from multiple datasets



Spatial Relationship Queries

How we reason & communicate about space

Key Concepts:

Proximity, Nearness, Adjacency, Connectivity, Containment, Arrangement

The spatial aspects of an environment…

(e.g. *location, amount, distance, adjacency, isolation, fragmentation, pattern*)

…impact ecological/human/environmental function.

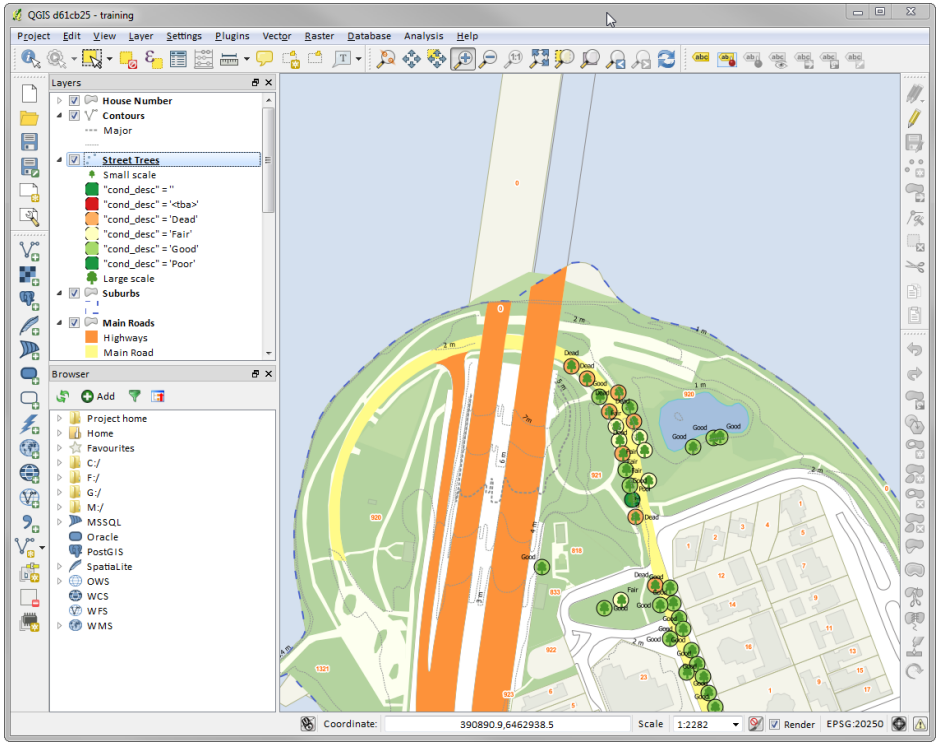
Open Source Tools for Geo

Open Source Software

- Source code freely available, and modifiable
- Open source geospatial includes
  - Spatial databases
  - Web map servers
  - Desktop applications
- Strength comes from a strong community of users
- Participate by
  - Providing programming help
  - Writing documentation
  - Simply by providing feedback

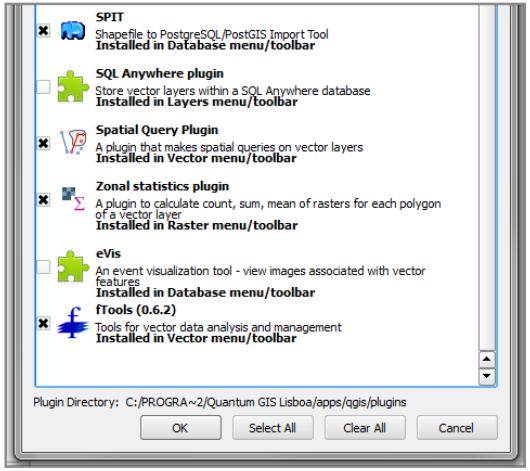
QGIS

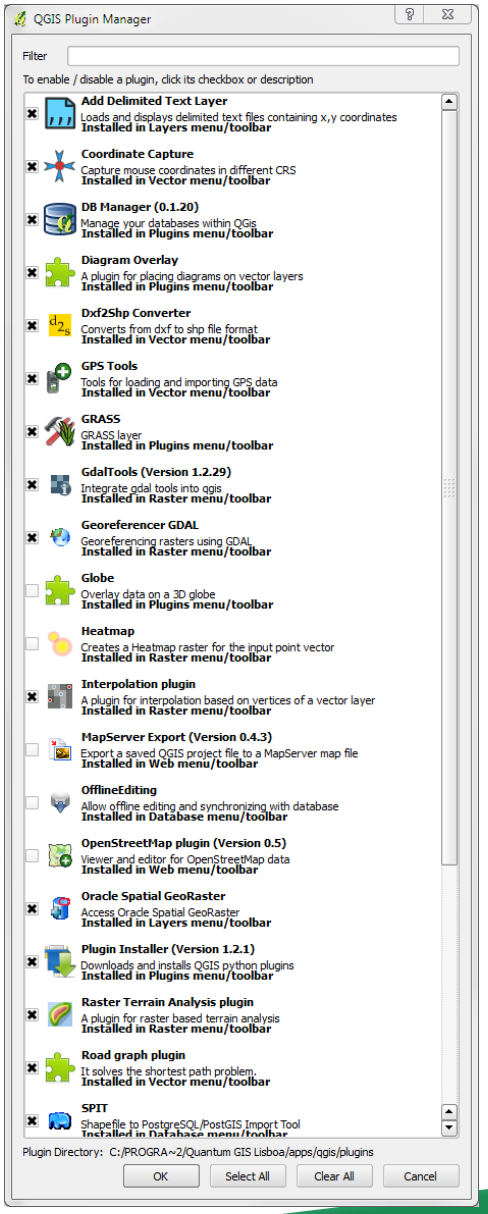
- Free and Open Source
- Easily Translatable
- Plugin Environment
- Easy to use/learn
- Powerful
- Multiplatform
  - Windows
  - Mac
  - Linux



QGIS Features

- Growing Core Functionality
  - Vector Support
  - Raster Support
  - Projection Support
  - Vector editing
  - Map Composition
- Plugins!





Spatial Data Analysis Options

Proprietary: ESRI ArcDesktop, MapInfo, IDRISI (low cost), Manifold (low cost)

Open Source Desktop: QGIS, Grass, uDig, SAGA, gvSIG, GMT, R Spatial Packages, Python spatial packages (GeoPandas, PySAL)

Web-based: MapGuide (open source), CartoDB (limited free edition)

Mobile: Amigo Cloud, Open Data Kit, QGIS for Andriod

Spatial Data Analysis Options for Remote Sensing

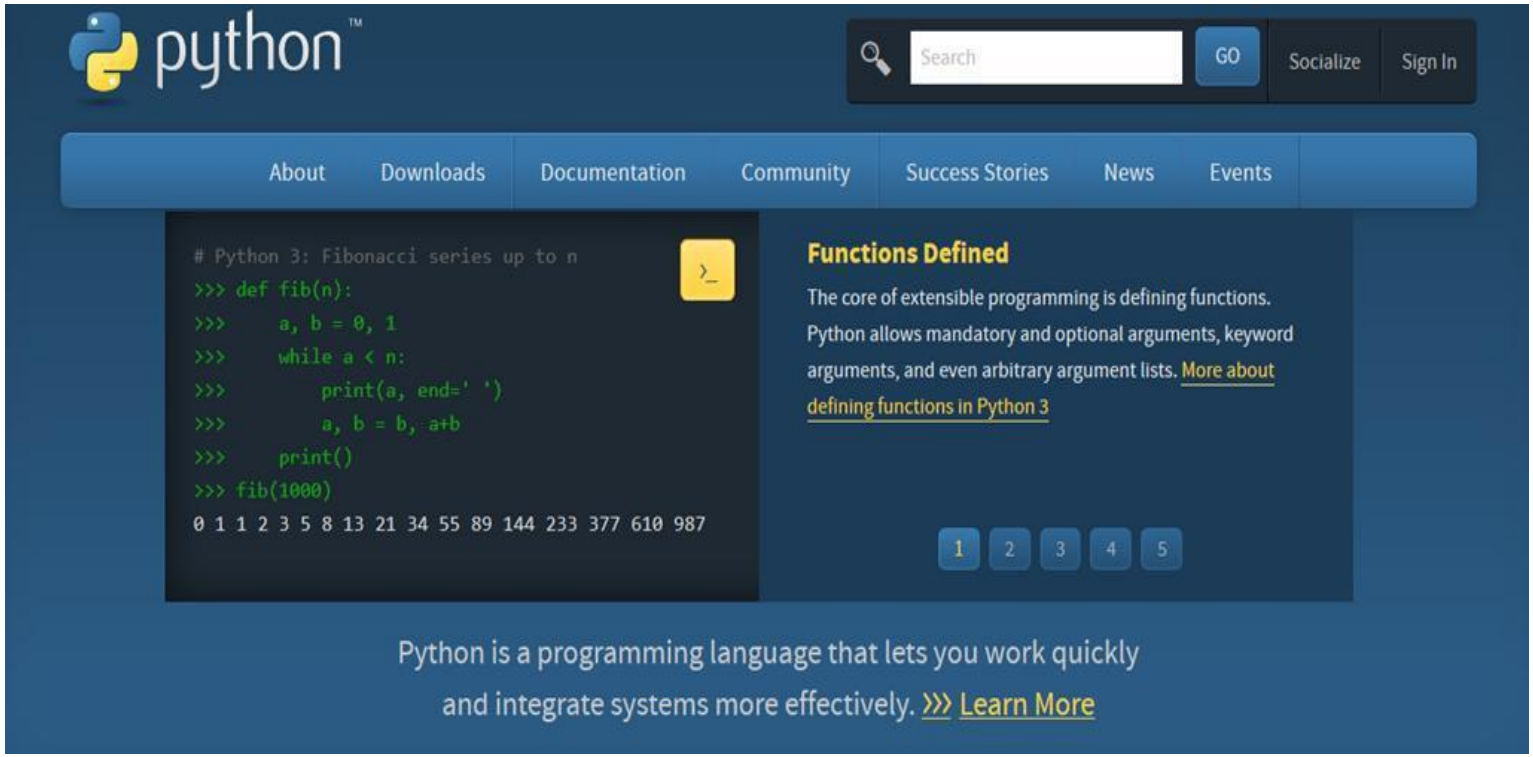
Proprietary: ENVI, ERDAS Imagine, PCI Geomatica, IDRISI, ESRI ArcDesktop

Open Source Remote Sensing Tools: Orfeo Toolbox, Grass, R, SAGA, QGIS, Opticks, OSSIM, ILWIS, Python packages (Rasterio, scikit-learn, scikit-image)

Free (but not open source) Remote Sensing Tools:

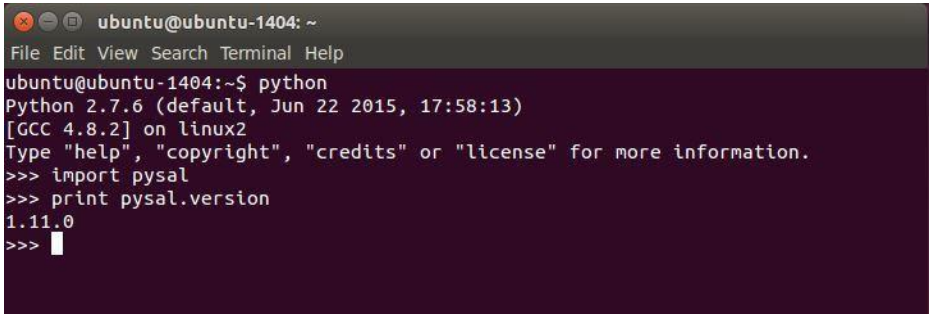
- Object-based image analysis: Spring
- Web-based: Google Earth Engine

What is Python?

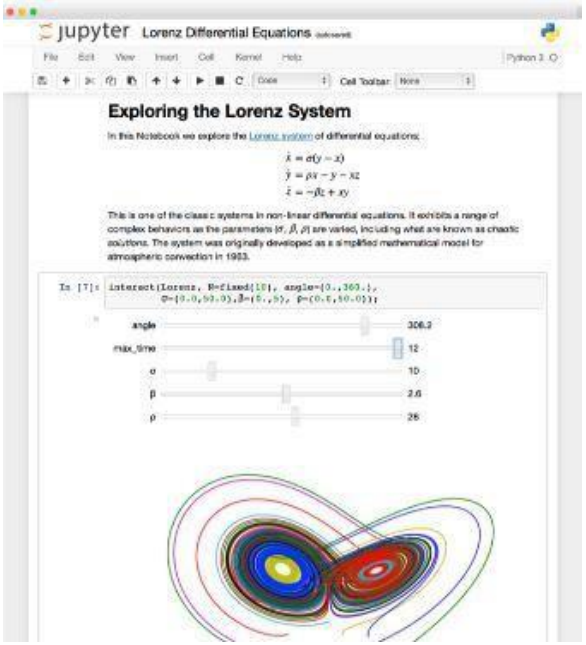


https://www.python.org/

Using Python

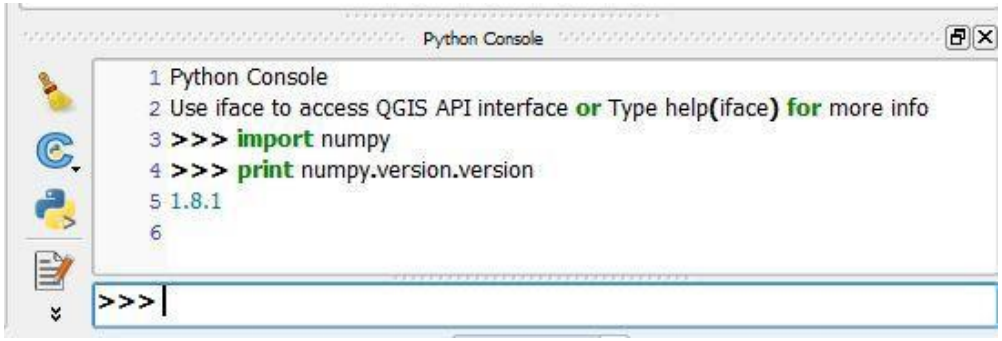


Via terminal

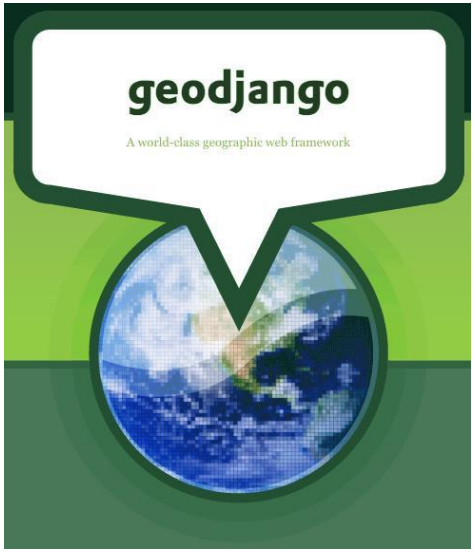


Jupyter Notebook
(browser-based shell installed locally or run in the cloud)

Using Python: 3rd Party Tools



QGIS Python Console



GeoDjango (web mapping)

http://geodjango.org/

What is R?

[Home]

Download

CRAN

R Project

About R
Logo
Contributors
What's New?
Mailing Lists
Bug Tracking
Development Site
Conferences
Search

R Foundation

Foundation
Board
Members
Donors
Donate

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.

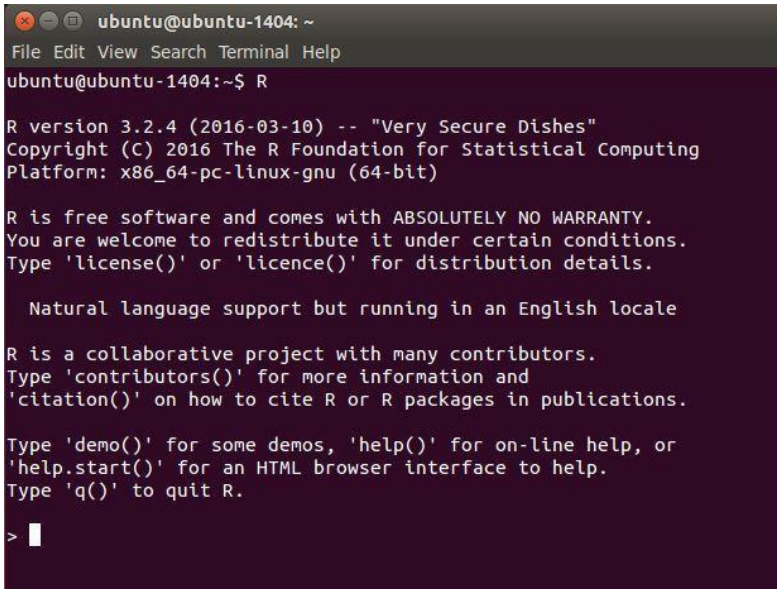
If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

News

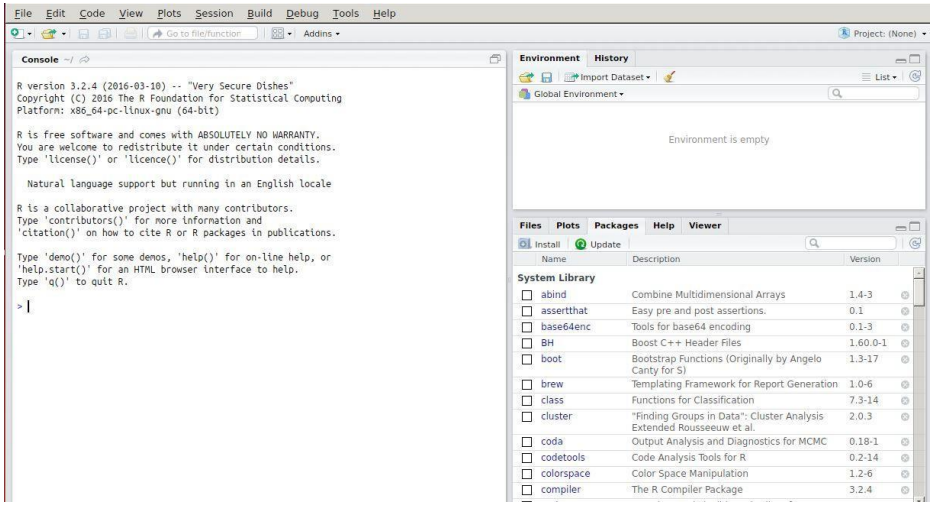
- **Notice XQuartz users (Mac OS X)** A security issue has been detected with the Sparkle update mechanism used by XQuartz. Avoid updating over insecure channels.
- **R version 3.2.4 (Very Secure Dishes)** has been released on Thursday 2016-03-10.
- **R version 3.3.0 (Supposedly Educational) prerelease versions** will appear starting Monday 2016-03-14. Final release is scheduled for Thursday 2016-04-14.
- The **R Logo** is available for download in high-resolution PNG or SVG formats.
- **useR! 2016**, will take place at Stanford University, CA, USA, June 27 - June 30, 2016.
- **The R Journal Volume 7/2** is available.
- **R version 3.2.3 (Wooden Christmas-Tree)** has been released on 2015-12-10.
- **R version 3.1.3 (Smooth Sidewalk)** has been released on 2015-03-09.

https://www.r-project.org/

Using R: Traditional Install

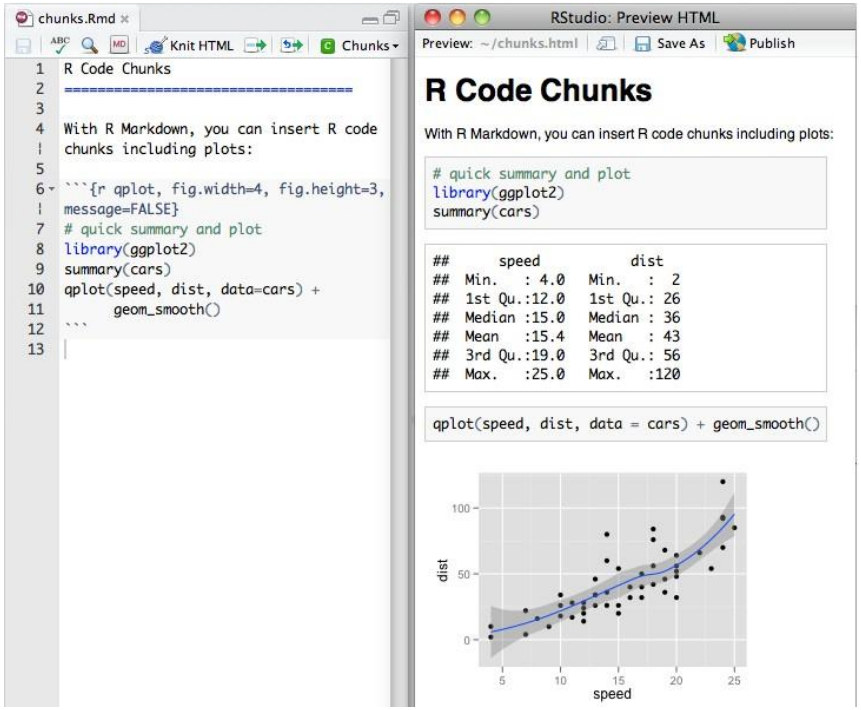


Via terminal

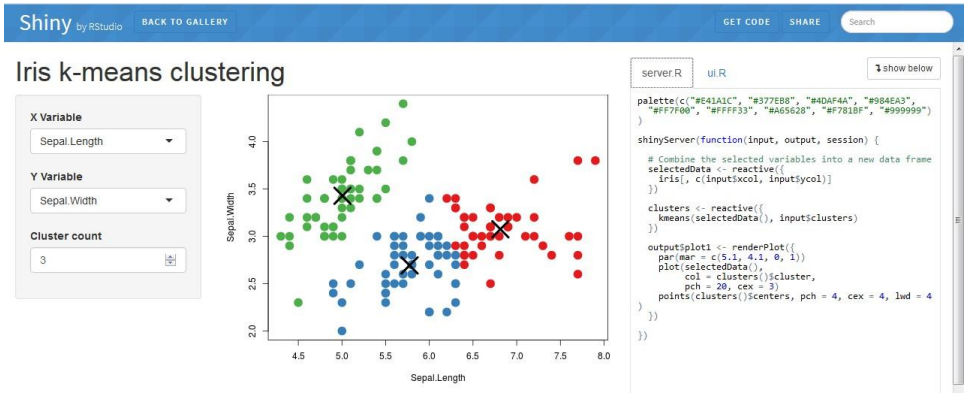


RStudio

Using R: Modern Options



R Markdown



R Shiny

	ArcDesktop Suite	*Open Source GIS*	*Python*	*R Spatial*
Setting Up Working Environment	Easy Install; Very clear GUI including Model Builder; restricted to Windows	**Easy Install; Relatively clear GUIs; Lots of plug-ins; OS independent**	Often many dependencies for installation of packages; (too?) many GUI and text editor options; OS independent	Requires installation of packages, but relatively easy; R Studio is nice GUI; OS independent
Spatial Databases	Needs ArcSDE for DBMS	work with major open source DBMS	**Works VERY well with major open source DBMS**	Can work with major open source DBMS
Editing Data	**Easy-to-use editing toolbar for all edits**	Multi-step editing process not as clear	Easier to create new spatial data than to edit existing	Easier than Python to create new spatial data
Analyzing Data	Good for beginners ; ArcPy good for advanced users; not much statistical power; solid data handling	Good for beginners; Grass allows more advanced analysis, as do plug-ins	**Statistical platform of choice for most physical sciences; handles "Big Data" the best**	**Statistical platform of choice for env and ecological sciences**
Making Maps	**Easiest option for Beginners**	Good option for Beginners	Requires willingness to learn programming fundamentals	Good for scientists already using R
Visualizing Data	Some functionality for easy charts and graphs; 3D tools are improving	Plugins available (i.e., Diagram Overlay)	Great options available but requires willingness to learn programming fundamentals	**Best option, esp. for those already using R**
Publishing Data (web)	ArcGIS Online (free but limited); ArcGIS Server (robust but expensive)	Doesn't have built-in option	**Web full of options (i.e., GeoDjango, Kartograph, Mapnik, MapServer, Flask)**	R Shiny is getting a lot of attention

GDAL/OGR

Geospatial Data Abstraction Library (GDAL/OGR) is a library for data handling of raster and vector formats (the current count is 142 and 84 formats, respectively. It is installed on the operating system and accessed via command line or scripting languages such as Python.

Primarily developed by Frank Warmerdam (now at Planet), it is now maintained by the GDAL/OGR Project Management Committee under the Open Source Geospatial Foundation.

For projections, it can use PROJ4 definition, EPSG code, and Well-known Text Coordinate Reference System (WKT CRS).

http://www.osgeo.org/gdal_ogr

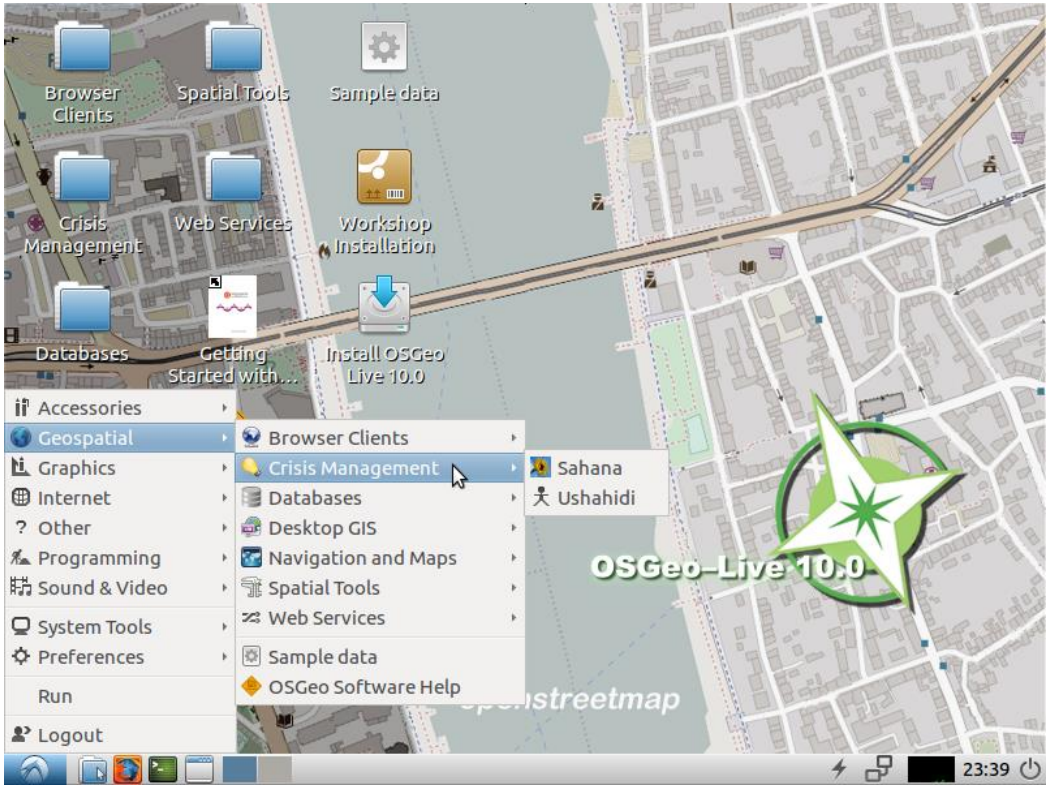
Raster Processing with GDAL

- Get started with this great GDAL/OGR cheat sheet
- Gdalinfo
  - Get raster info on file
  - Calculate stats (mean, standard deviation, histograms) which is stored in external XML file
    - Note options to reduce amount of info you get
- Gdal_translate
  - Converts data formats, rescales and subsets bands and spatial extents, assigns no-data values and the spatial reference system!
  - Compresses raster files (loseless compression, either LZW or Deflate)
  - Subsets
  - Writes out to a new file

Vector Processing with OGR

- Ogrinfo
  - Lists information about data file
- Ogr2ogr
  - Converts simple features data between file formats

OSGeo Live 10.0 -
Virtual Machine for Open Source and Web



GitHub

Code management and storage web-based platform with a focus on:

1. Versioning (tracking changes)
2. Collaboration (ability to clone and add to existing repositories)

All repositories are publicly available, unless you create it with a paid (upgraded account) that can create private repositories.

"GitHub is the largest code host on the planet with over **22.8 million** repositories."

If you haven't already, create your free account: https://github.com/join

Hands-on exercise

https://github.com/nethomas-geo/QGIS-Training