



# 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*.

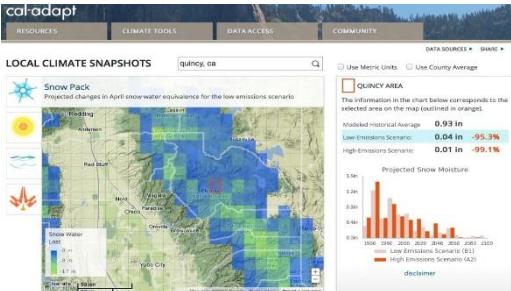
- Broader/different scales of interest – more personal and more broad
- New ways of analyzing and synthesizing data
- Multiple users and collaboration
- Agile visualization

# 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.*

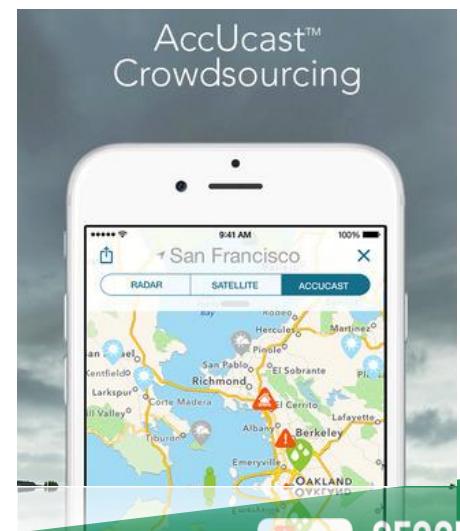
## web mapping/visualization



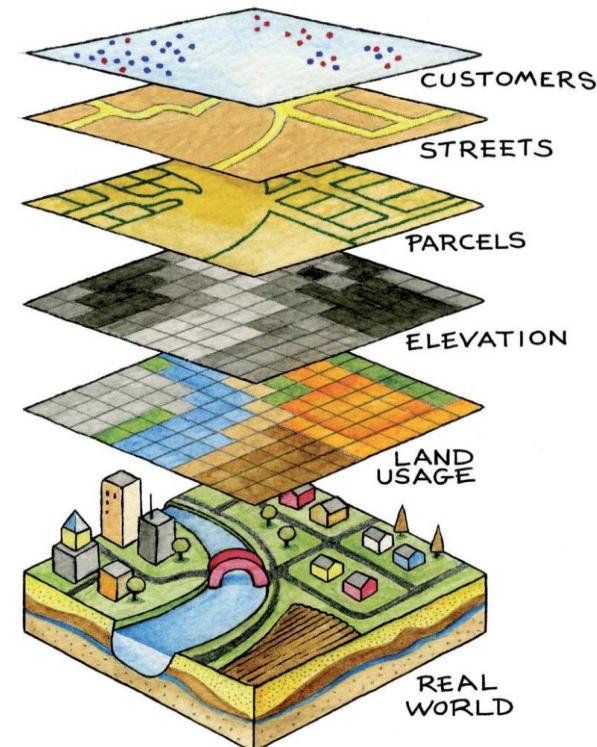
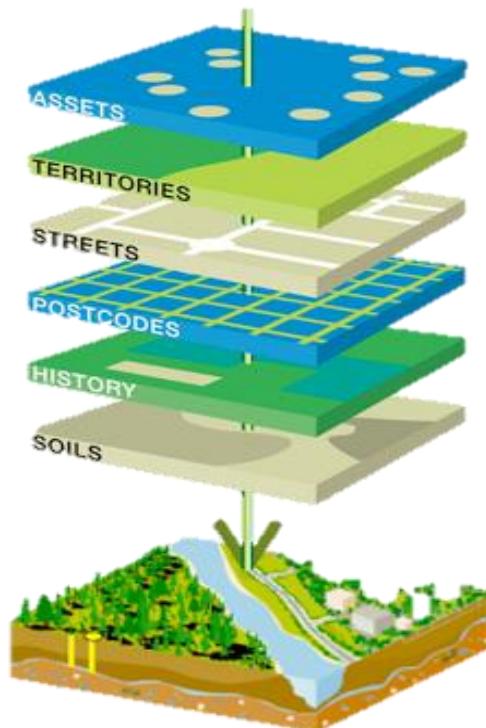
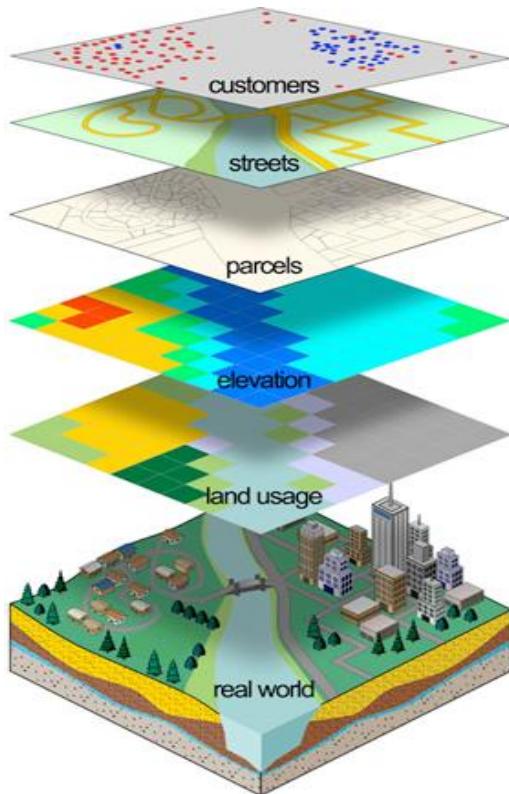
## citizen involvement



## disaster response



# The GIS Model: the 20<sup>th</sup> Century Mapping Toolkit



# We Need a 21<sup>st</sup> 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.**

# Spatial Data Science

Importance of our research being:

- Reproducible
- Repeatable
- Collaborative
- Meaningful

**Open source tools** allow us to achieve these goals more easily

# Exploring California's Climate Change Research

Cal-Adapt provides a view of how climate change might affect California. Find tools, data, and resources to conduct research, develop adaptation plans and build applications.



Annual Averages  
Extreme Heat



Annual Averages



Annual Averages



Snowpack



Sea Level Rise

## Climate Tools

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

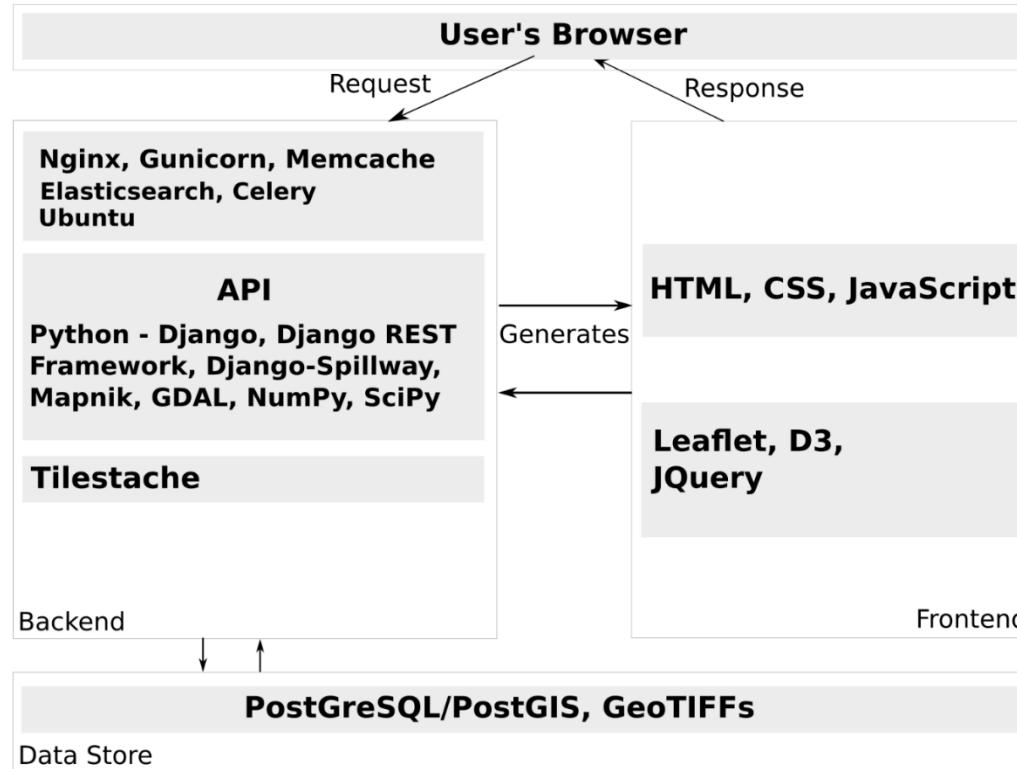
## Download Data

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

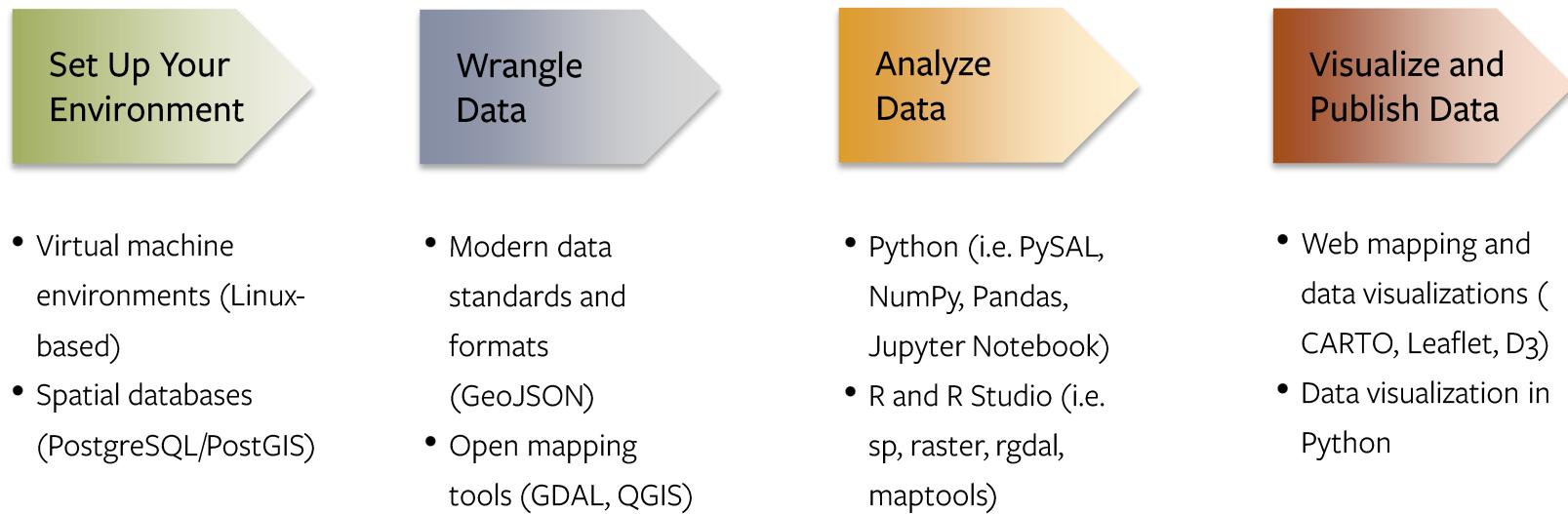
## Find Resources

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



# 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

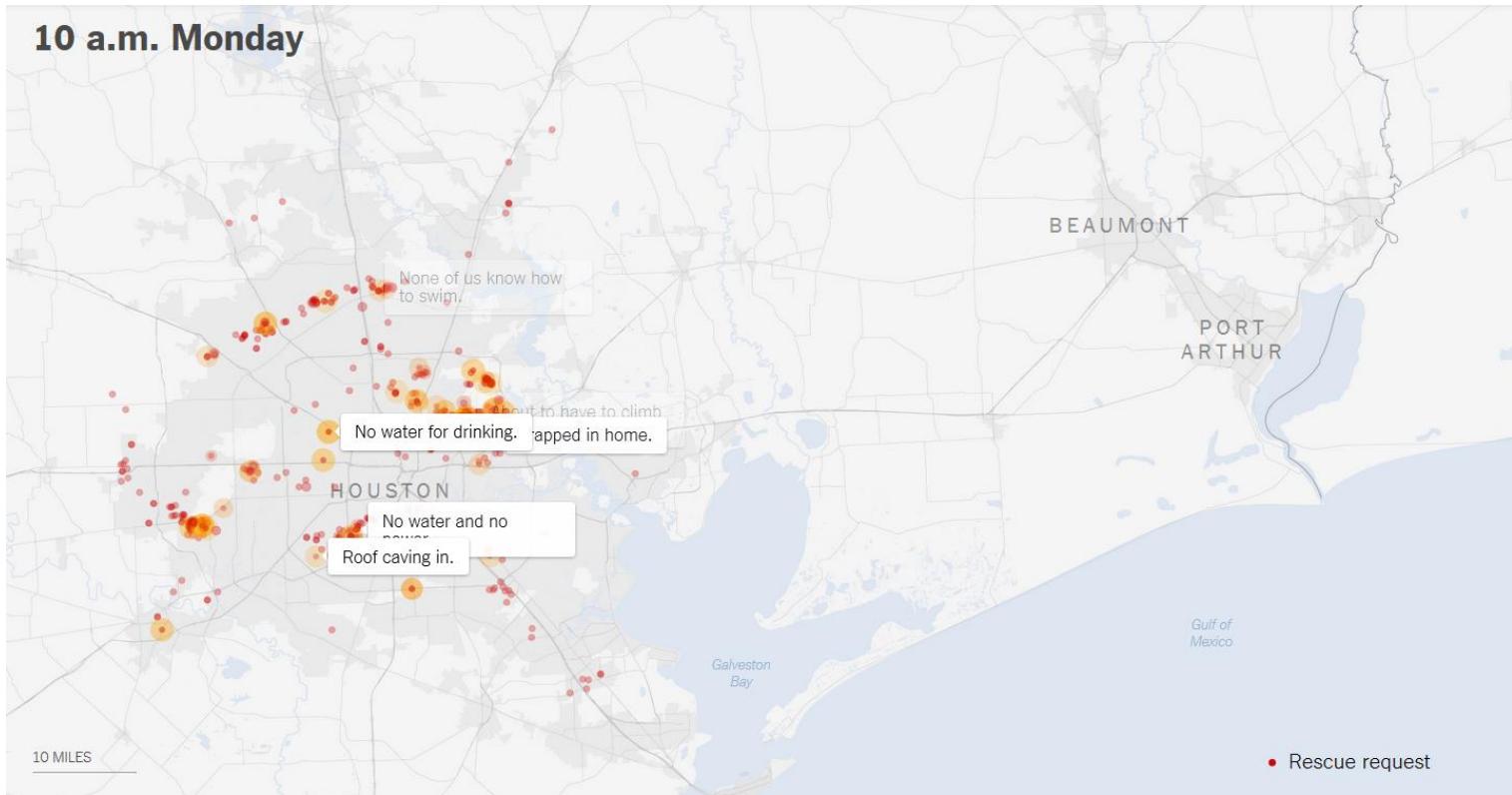


# Why Geospatial?

Digital maps are everywhere & changing the way we do things

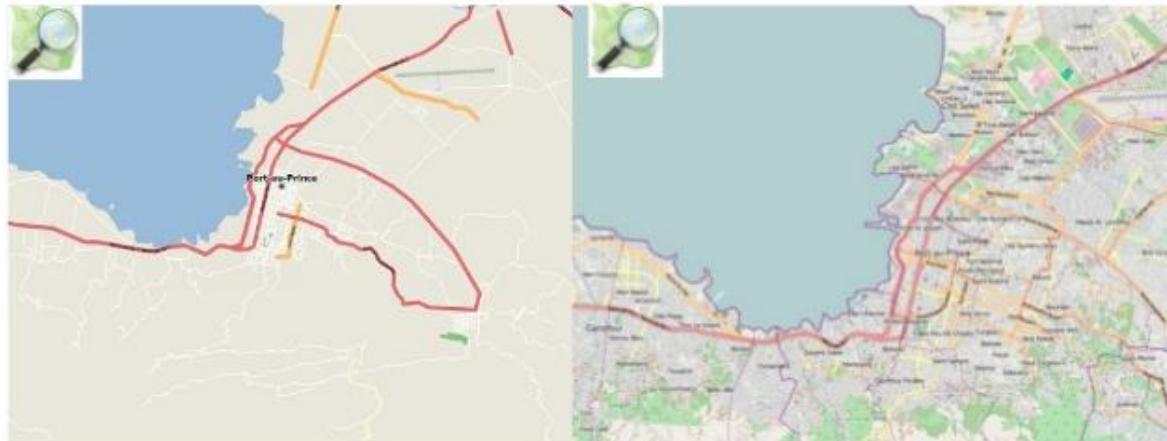
Houston flooding

# Disaster Response



New York Times: August 30, 2017

# Volunteered Geographic Information

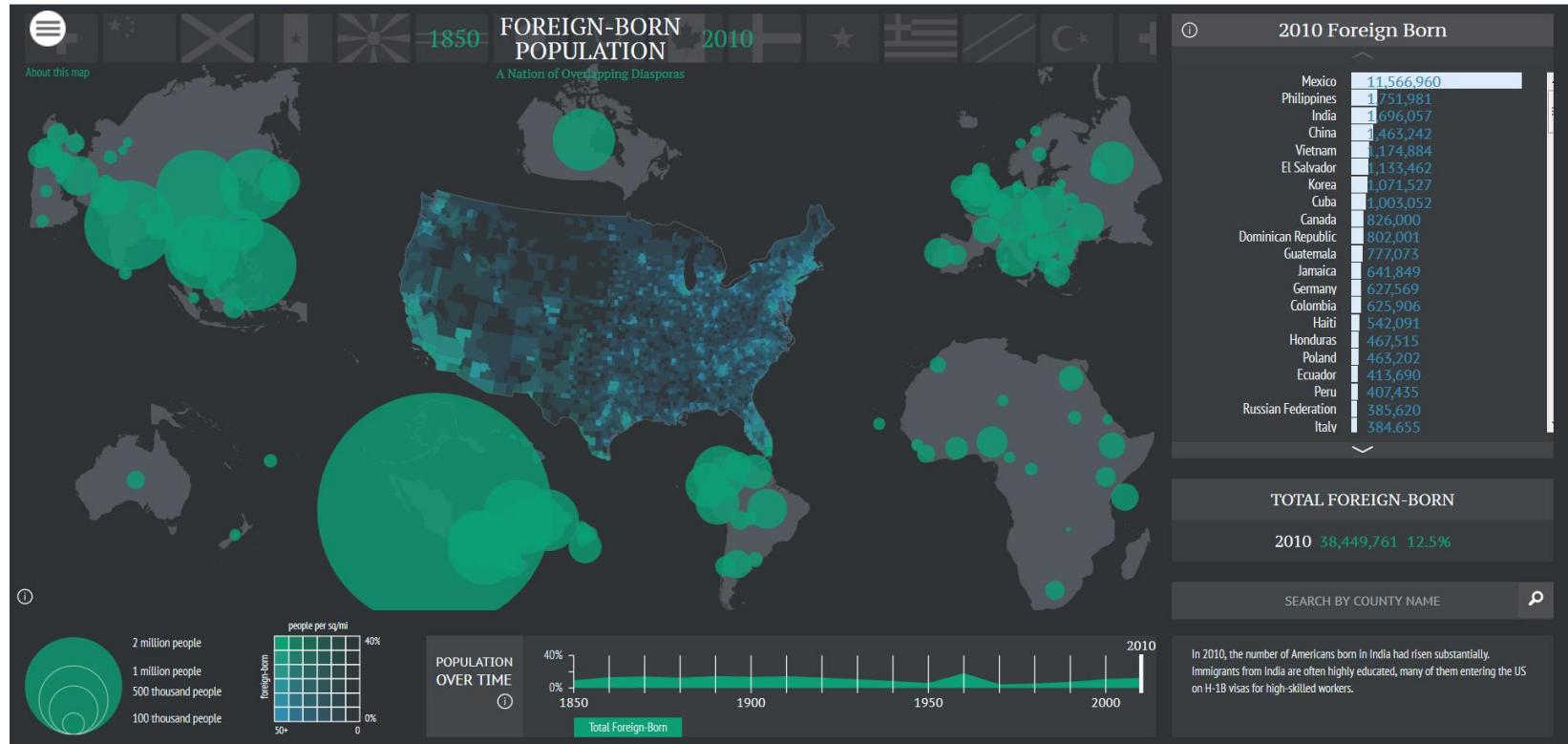


Port-au-Prince on OSM,  
January 12, 2010

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.

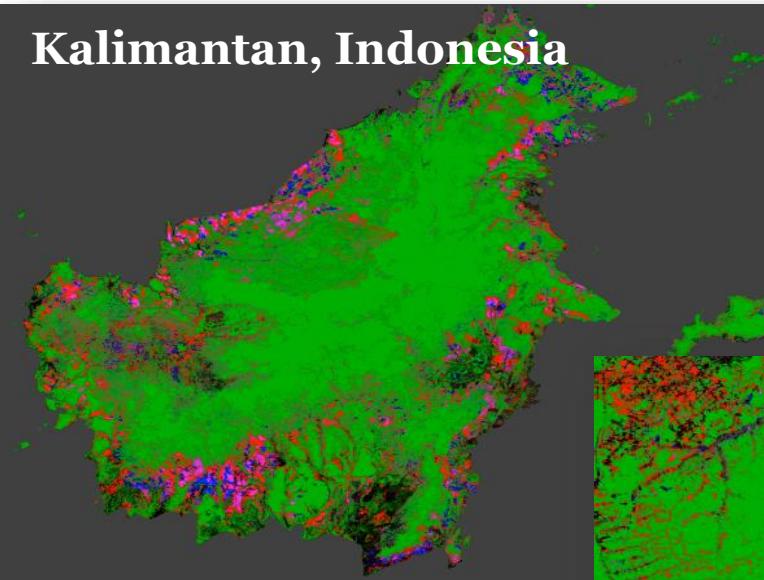
# Web Mapping and Visualization



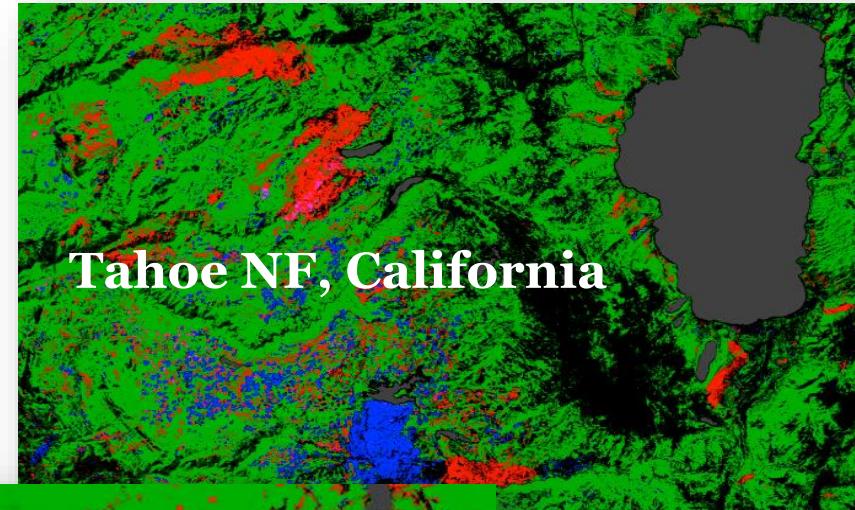
- American Panorama: <http://dsl.richmond.edu/panorama/>

# Large Scale Cloud Computation: Google Earth Engine

Kalimantan, Indonesia



Tahoe NF, California



State of Para, Brazil



10 years,  
1.3M Landsat  
scenes,  
Cloud processing  
power

# Collaboratories



Species

Location

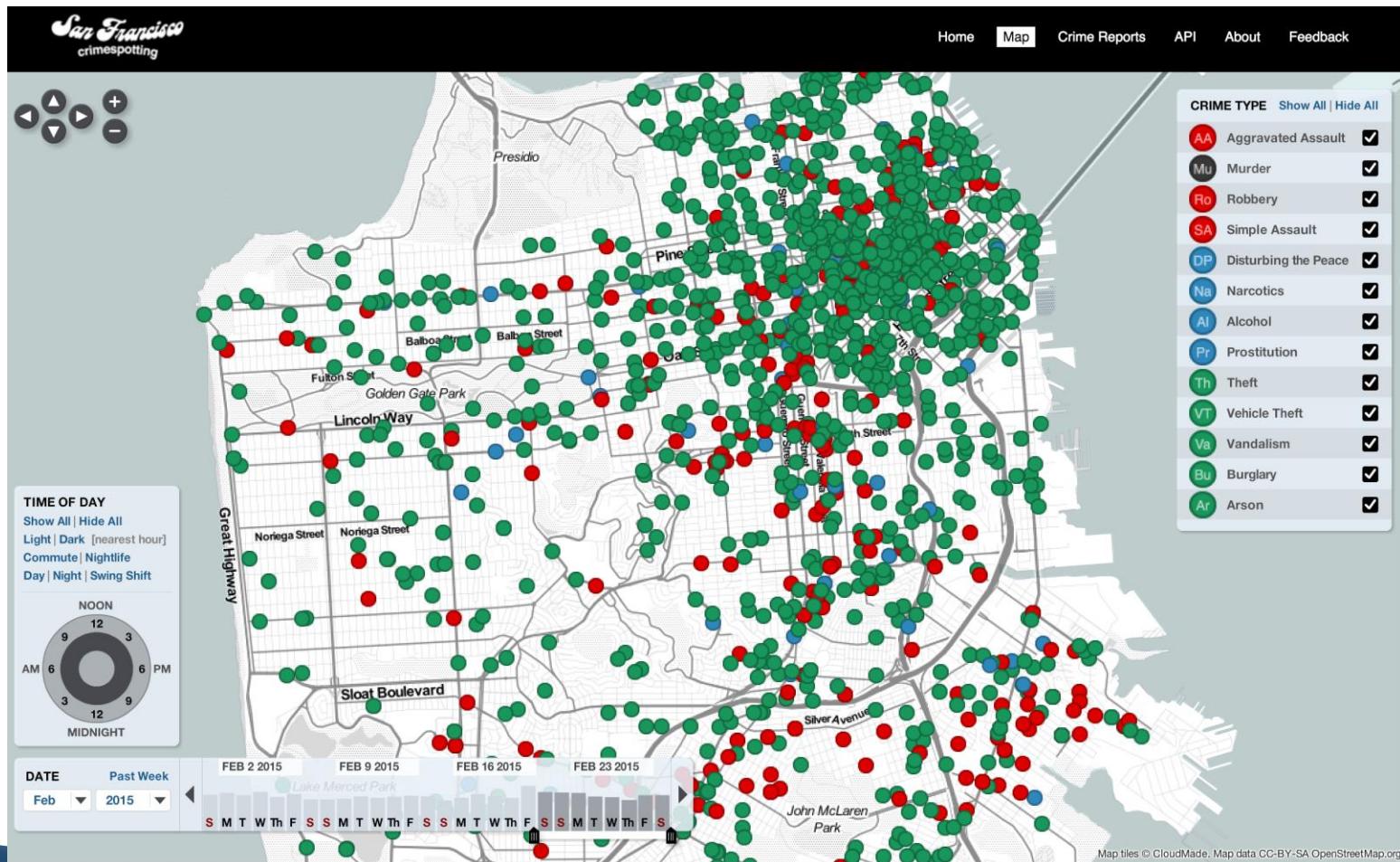
Go

Filters

The World

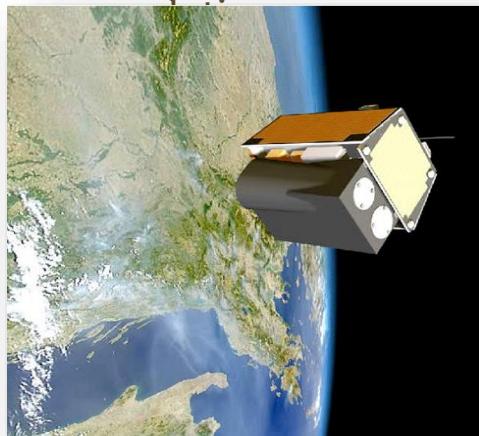
1,894,907  
OBSERVATIONS81,293  
SPECIES8,626  
IDENTIFIERS40,381  
OBSERVERS

# Open Government Data



# Sensor Trends: Drones, Micro Satellites

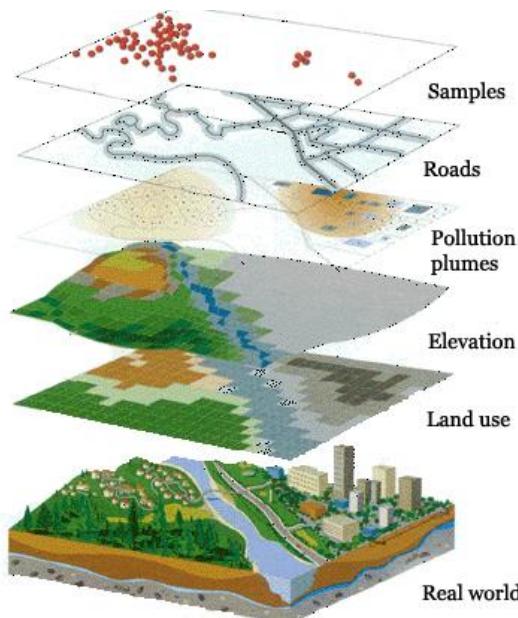
New sensors, smaller



# 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

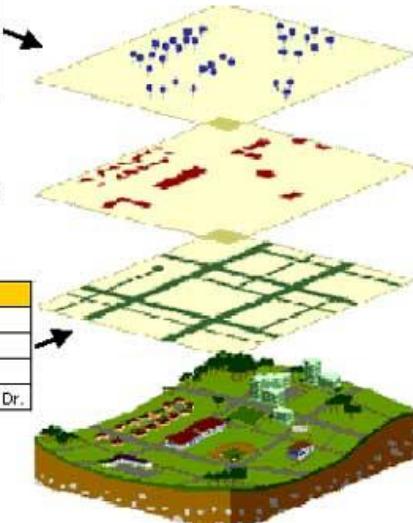


Real  
World

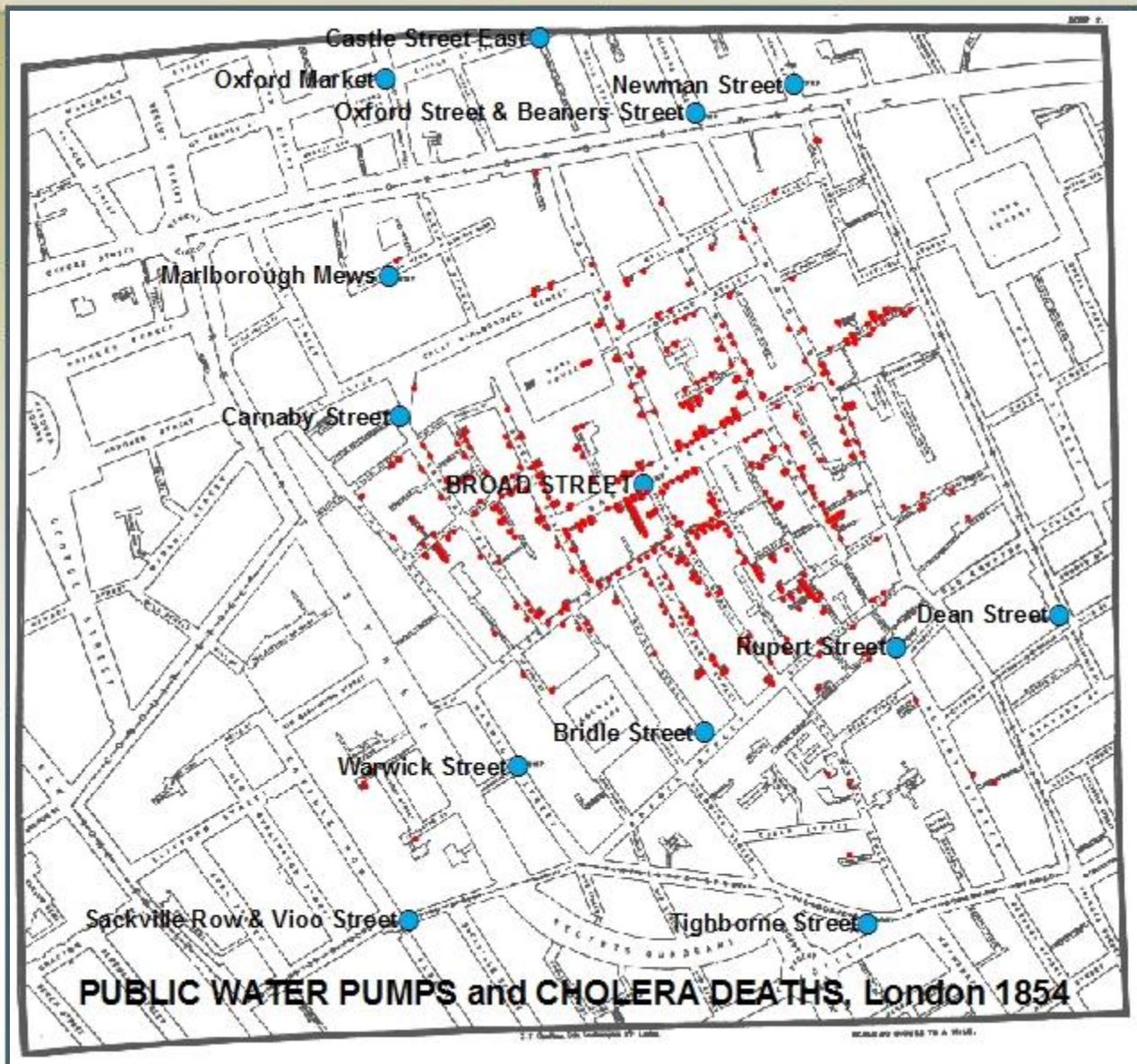
| Monitoring Wells |              |               |
|------------------|--------------|---------------|
| Well ID          | Date Sampled | Concentration |
| C-6A             | 5/8/94       | 300           |
| C-8A             | 5/8/94       | 20            |
| C-13A            | 5/8/94       | 120           |
| 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 | Address           |
| Blake       | 6         | 79 Circuit St     |
| Hernandez   | 2         | 148 Plain St.     |
| Joy         | 4         | 18 Webster St.    |
| Smith       | 5         | 4321 Tecumseh Dr. |

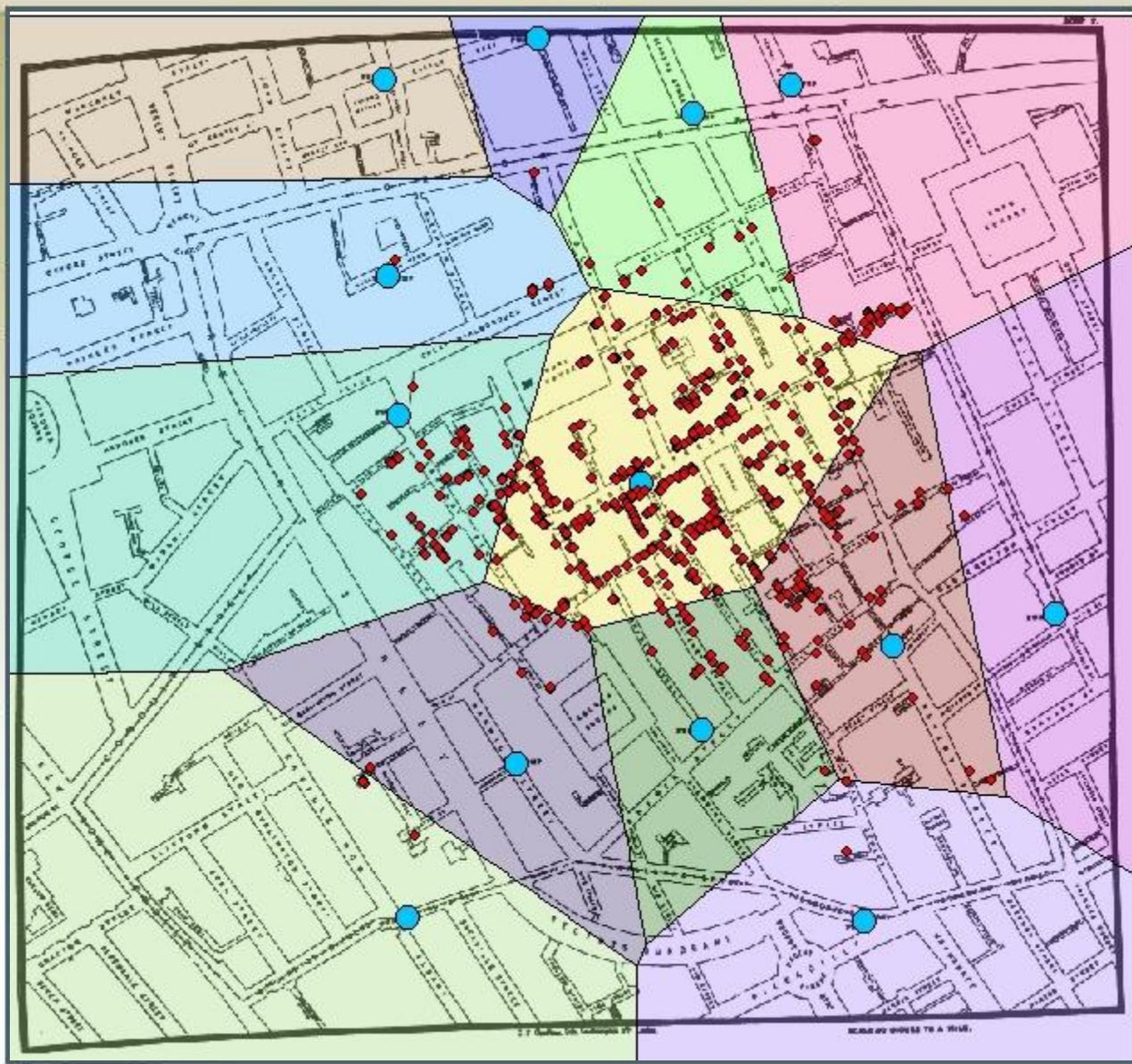


# Beginnings of GIS – Snow Map



**"Miasma"**

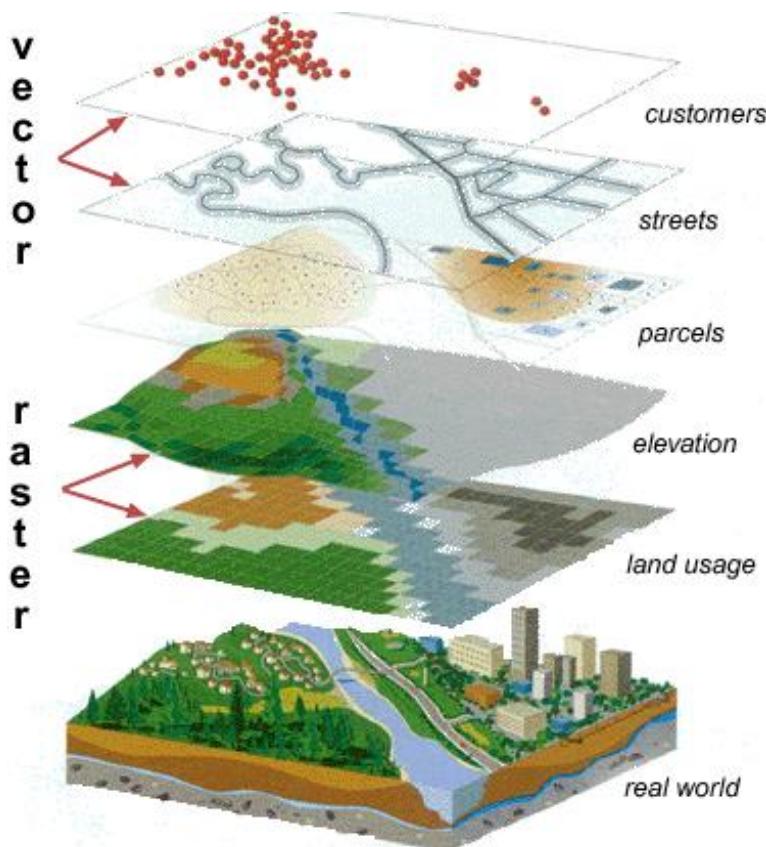
# Beginnings of GIS – Snow Map



“Miasma”

# How GIS Works

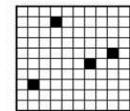
**A GIS stores information about the world as a collection of thematic layers that can be linked together by geography**



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

Raster

The raster view of the world



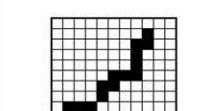
Happy Valley spatial entities



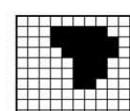
Points: hotels

Vector

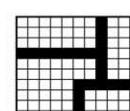
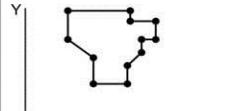
The vector view of the world



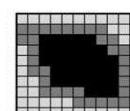
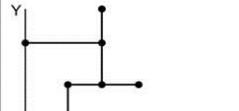
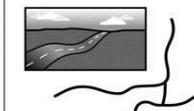
Lines: ski lifts



Areas: forest



Network: roads

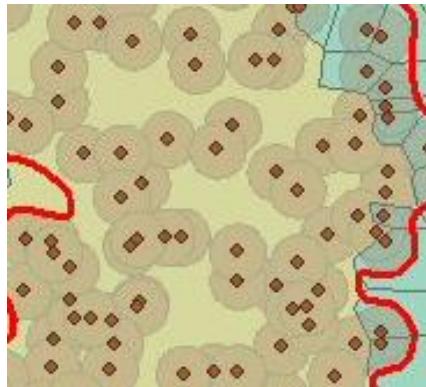


Surface: elevation

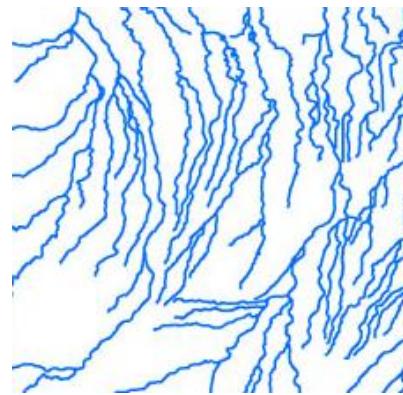


# Vector data examples

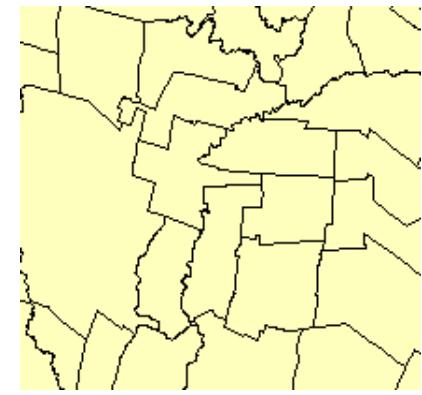
(points, lines, and polygons)



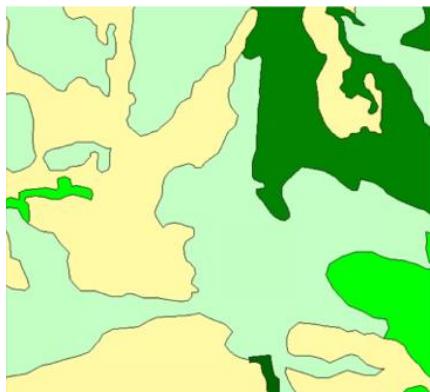
GPS data



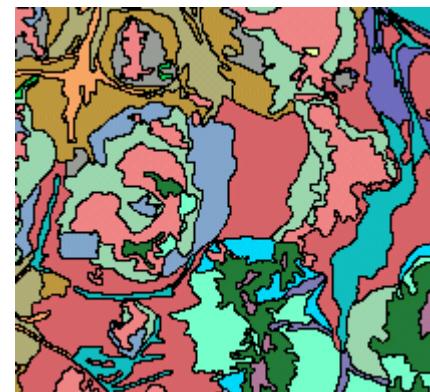
Rivers



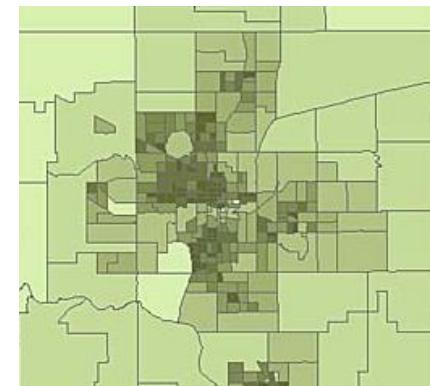
Counties



Habitat boundaries



Soil type

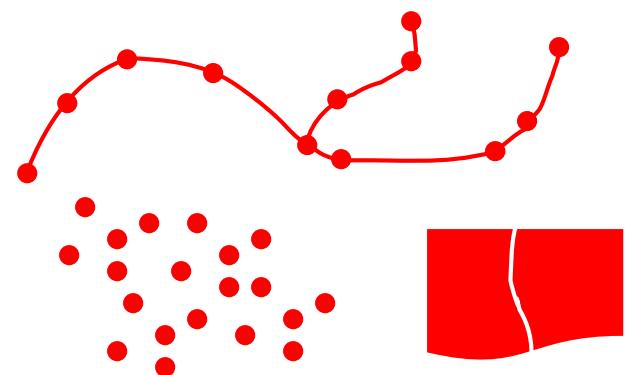


Census data

# Shapefiles

| <i>File type extension</i> | <i>What it is</i>  | <i>Description</i>  | <i>Example</i> |
|----------------------------|--|---|----------------|
| .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   |

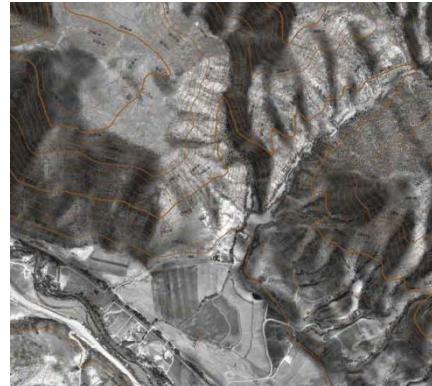
- Vector Data type
- Features: points, lines & polygons
- Attributes: size, type, length, etc.



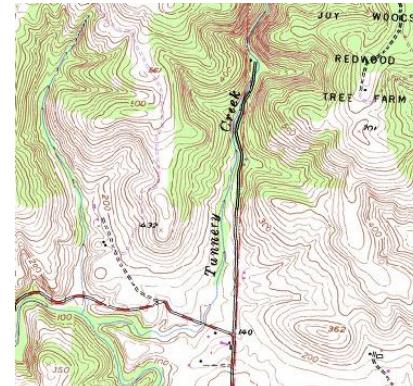
# Raster data examples



Satellite imagery



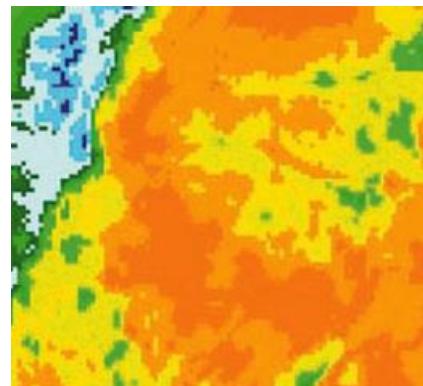
Elevation



Digital USGS  
topo map



Landcover/landuse



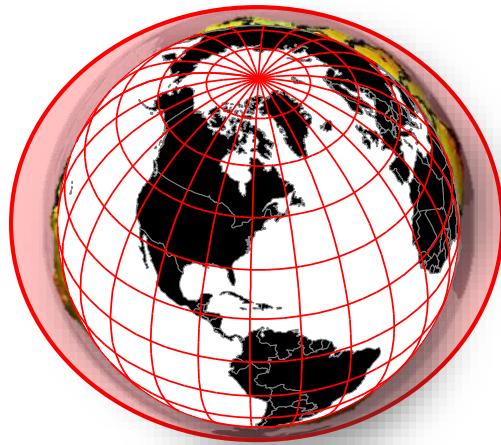
Precipitation



Aerial photography

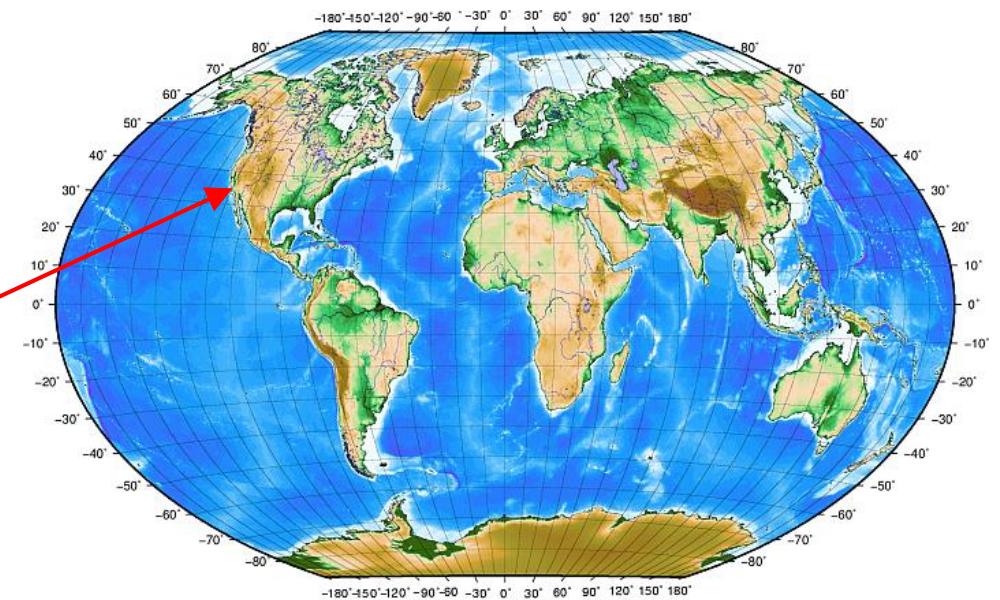
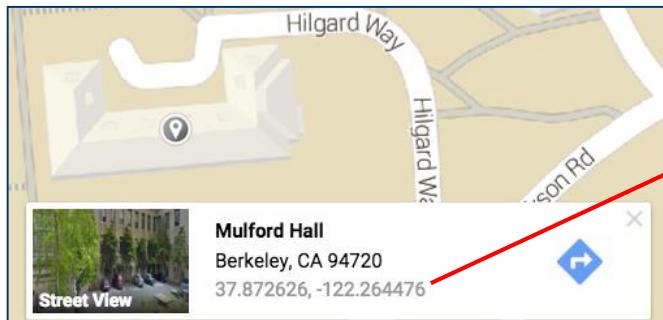


# Coordinate Systems & Projections



# Geographic Coordinate Reference System (CRS)

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

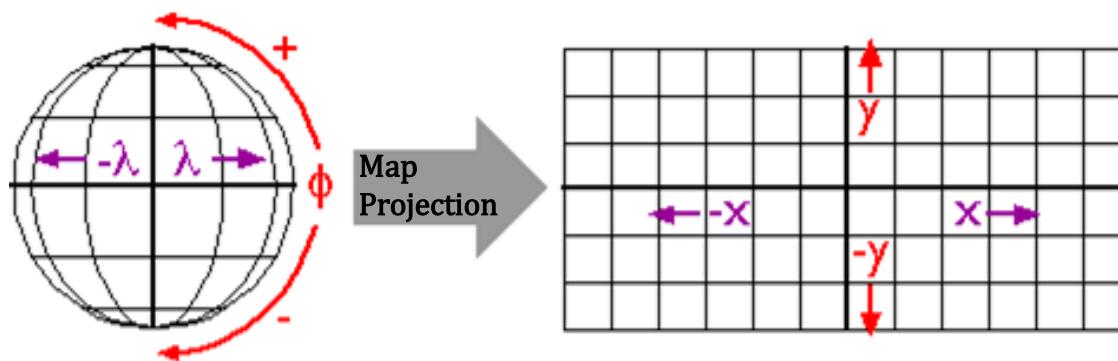


# Map Projection

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

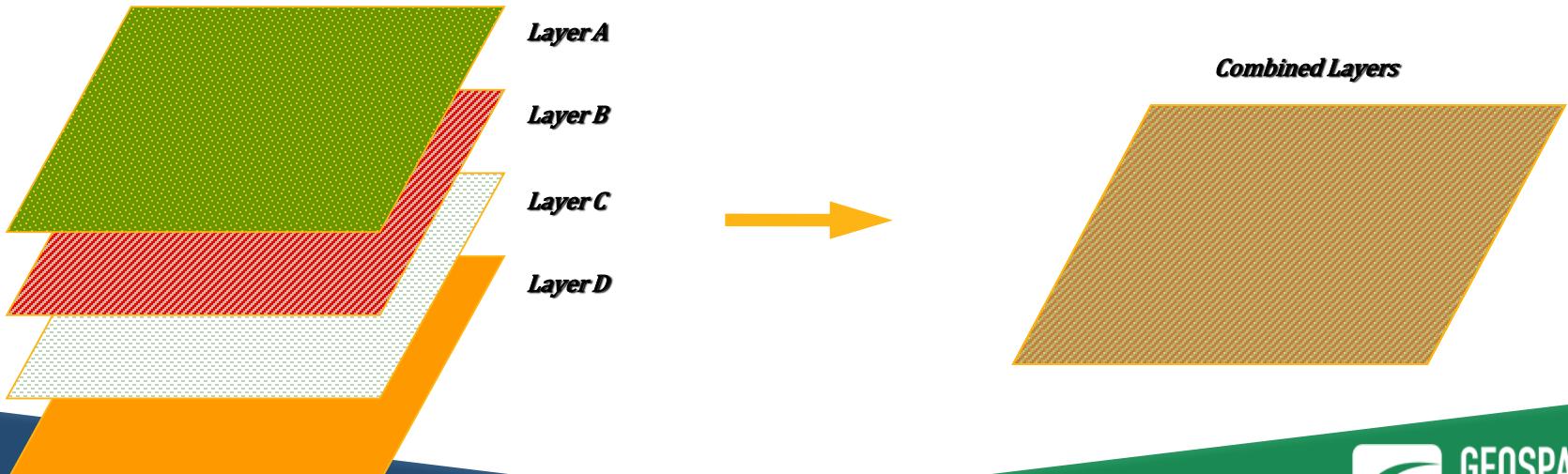
Longitude  $\lambda > X$

Latitude  $\phi > Y$



# 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.

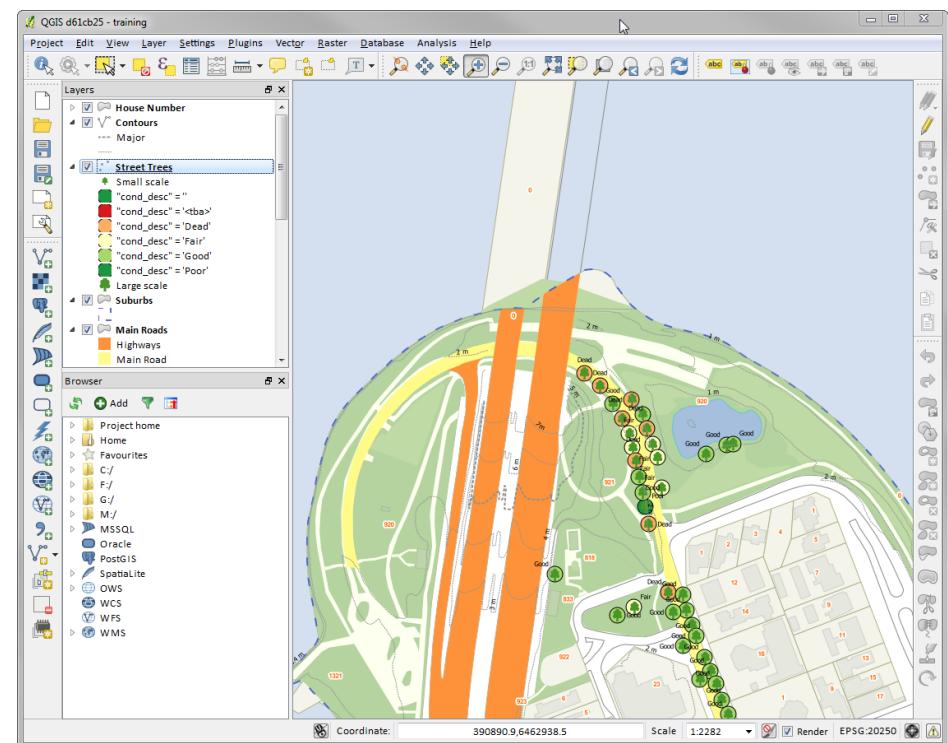
# Open Source Tools for Geo

# 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

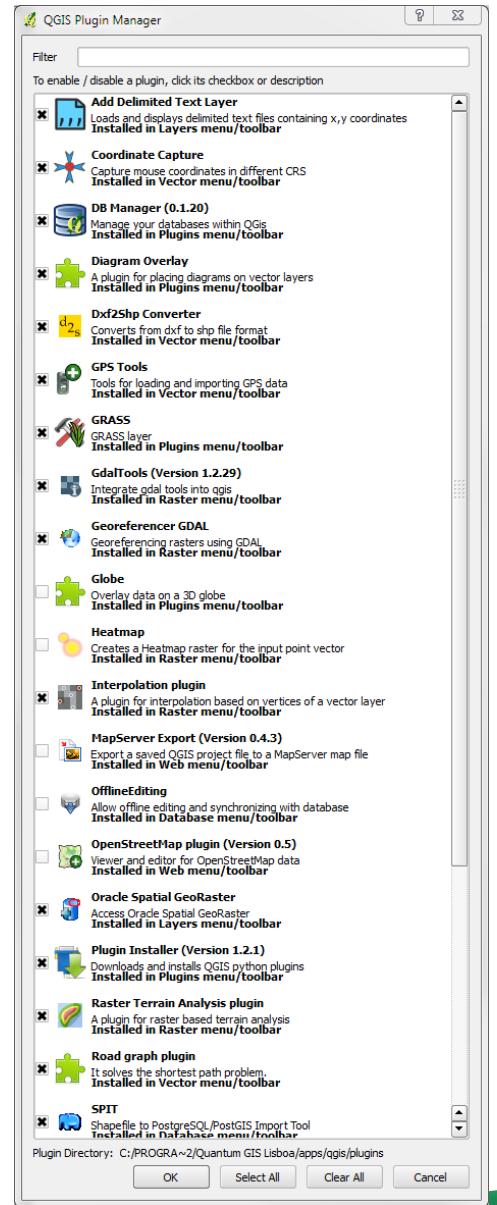
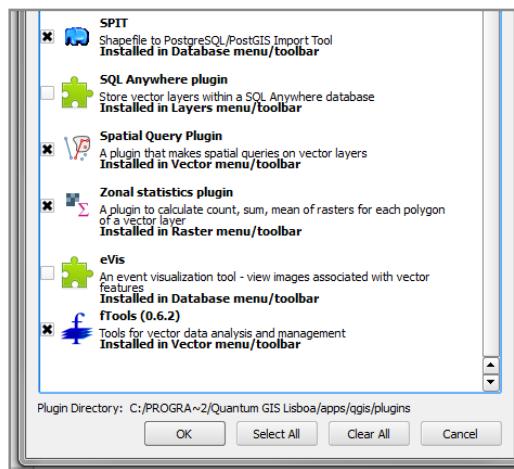
# QGIS

- 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?

The screenshot shows the Python.org homepage. At the top, there's a navigation bar with links for About, Downloads, Documentation, Community, Success Stories, News, and Events. Below the navigation is a search bar with a magnifying glass icon and a 'GO' button. To the right of the search bar are 'Socialize' and 'Sign In' buttons. The main content area features a dark blue background with white text. On the left, there's a code snippet in a terminal window:

```
# Python 3: Fibonacci series up to n
>>> def fib(n):
    >>>     a, b = 0, 1
    >>>     while a < n:
    >>>         print(a, end=' ')
    >>>         a, b = b, a+b
    >>>     print()
    >>> fib(1000)
0 1 1 2 3 5 8 13 21 34 55 89 144 233 377 610 987
```

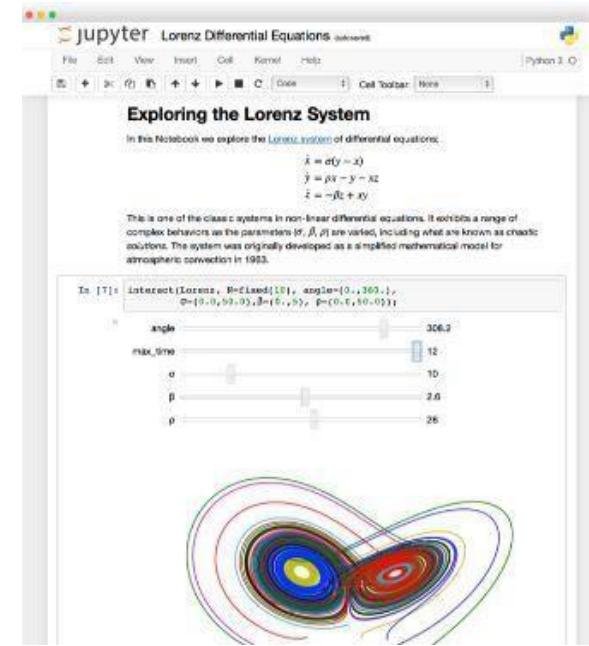
To the right of the code, there's a yellow button with a right-pointing arrow. Below the code, the text "Functions Defined" is highlighted in orange. The text explains the core of extensible programming and provides a link to "More about defining functions in Python 3". At the bottom of the main content area, there are five numbered buttons (1, 2, 3, 4, 5) and a call-to-action text: "Python is a programming language that lets you work quickly and integrate systems more effectively. [» Learn More](#)".

<https://www.python.org/>

# Using Python

```
ubuntu@ubuntu-1404:~  
File Edit View Search Terminal Help  
ubuntu@ubuntu-1404:~$ python  
Python 2.7.6 (default, Jun 22 2015, 17:58:13)  
[GCC 4.8.2] on linux2  
Type "help", "copyright", "credits" or "license" for more information.  
>>> import pysal  
>>> print pysal.version  
1.11.0  
>>> █
```

Via  
terminal



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

# Using Python: 3<sup>rd</sup> Party Tools

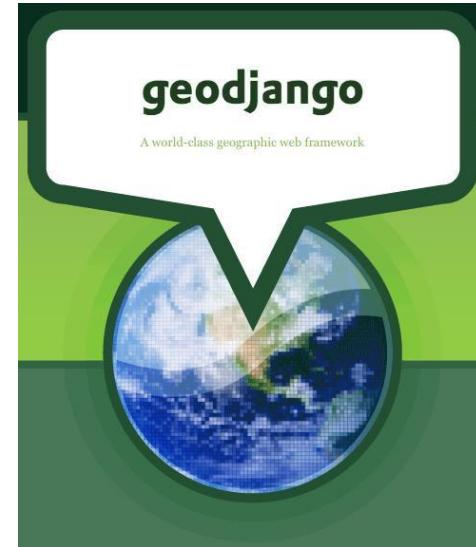


A screenshot of the QGIS Python Console window. The title bar says "Python Console". The main area contains the following text:

```
1 Python Console
2 Use iface to access QGIS API interface or Type help(iface) for more info
3 >>> import numpy
4 >>> print numpy.version.version
5 1.8.1
6
```

The bottom input field shows ">>> |". On the left side of the window, there are icons for file operations: a lightbulb, a document with a 'C' and a 'P', and a clipboard.

**QGIS Python  
Console**



**GeoDjango  
(web mapping)**

<http://geodjango.org>  
L

# 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

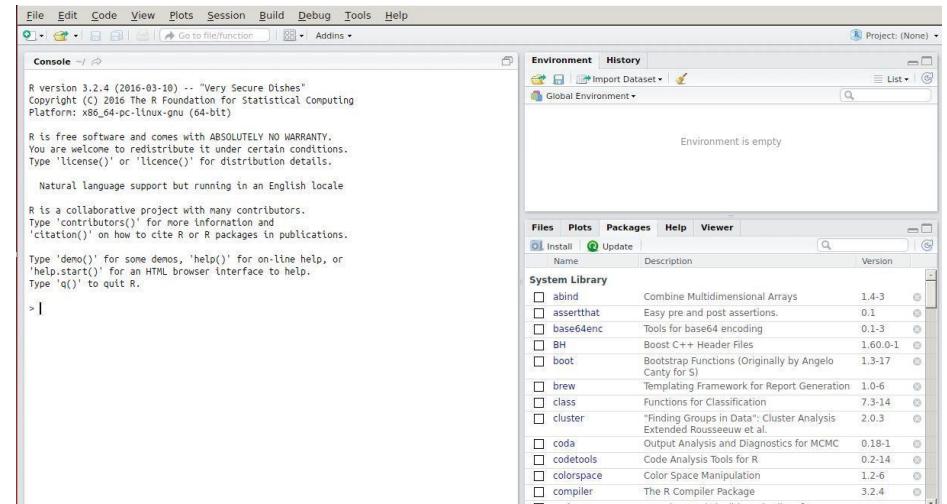
- [Notice XQuartz users \(Mac OS X\)](#) A security issue has been detected with the Sparkle update mechanism used by XQuartz. Avoid updating over insecure channels.
- [R version 3.2.4 \(Very Secure Dishes\)](#) has been released on Thursday 2016-03-10.
- [R version 3.3.0 \(Supposedly Educational\) prerelease 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

```
ubuntu@ubuntu-1404:~  
File Edit View Search Terminal Help  
ubuntu@ubuntu-1404:~$ R  
  
R version 3.2.4 (2016-03-10) -- "Very Secure Dishes"  
Copyright (C) 2016 The R Foundation for Statistical Computing  
Platform: x86_64-pc-linux-gnu (64-bit)  
  
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.  
  
Natural language support but running in an English locale  
  
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.  
  
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.  
  
> █
```

Via  
terminal



RStudio

# Using R: Modern Options

The screenshot shows the RStudio interface. On the left, the code editor displays an R Markdown file named 'chunks.Rmd' with the following content:

```
1 R Code Chunks
2 =====
3
4 With R Markdown, you can insert R code
5 chunks including plots:
6
7 ```{r qplot, fig.width=4, fig.height=3,
8 message=FALSE}
9 # quick summary and plot
10 library(ggplot2)
11 summary(cars)
12 qplot(speed, dist, data=cars) +
13   geom_smooth()
14 ````
```

Below the code, a preview window titled 'R Studio: Preview HTML' shows the rendered content:

## R Code Chunks

With R Markdown, you can insert R code chunks including plots:

```
# quick summary and plot
library(ggplot2)
summary(cars)

##      speed          dist
## Min.   :4.0   Min.   : 2
## 1st Qu.:12.0  1st Qu.: 26
## Median :15.0  Median : 36
## Mean   :15.4  Mean   : 43
## 3rd Qu.:19.0  3rd Qu.: 56
## Max.   :25.0  Max.   :120

qplot(speed, dist, data = cars) + geom_smooth()
```

A scatter plot is displayed below the code, showing 'dist' on the y-axis and 'speed' on the x-axis. The plot includes a blue smoothing line.

The screenshot shows a Shiny application titled 'Iris k-means clustering'. The interface includes dropdown menus for 'X Variable' (Sepal.Length), 'Y Variable' (Sepal.Width), and 'Cluster count' (set to 3). The main area displays a scatter plot of Sepal.Length (x-axis, 4.5 to 8.0) vs Sepal.Width (y-axis, 2.0 to 4.0). Data points are colored by cluster (green, blue, red) and marked with 'X' symbols. A legend at the bottom right identifies the clusters. The right side of the screen shows the 'server.R' and 'ui.R' files for the application.

```
shinyServer(function(input, output, session) {
  selectedData <- reactive({
    iris[, c(input$xcol, input$ycol)]
  })
  clusters <- reactive({
    kmeans(selectedData(), input$clusters)
  })
  output$plot1 <- renderPlot({
    par(mar = c(5, 1, 4, 1, 0, 1))
    plot(selectedData(),
         col = clusters()$cluster,
         pch = 20, cex = 2)
    points(clusters()$centers, pch = 4, cex = 4, lwd = 4)
  })
})
```

## R Shiny

R  
Markdown

|                                       | <i>ArcDesktop Suite</i>   | <i>Open Source GIS</i>   | <i>Python</i>   | <i>R Spatial</i>   |
|---------------------------------------|---|--|---|--|
| <i>Setting Up Working Environment</i> | Easy Install; Very clear GUI including Model Builder; restricted to Windows                         | <b>Easy Install; Relatively clear GUIs; Lots of plug-ins; OS independent</b> | 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 |
| <i>Spatial Databases</i>              | Needs ArcSDE for DBMS   | work with major open source DBMS   | <b>Works VERY well with major open source DBMS</b>  | Can work with major open source DBMS   |
| <i>Editing Data</i>                   | <b>Easy-to-use editing toolbar for all edits</b>  | 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  |
| <i>Analyzing Data</i>                 | 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      | <b>Statistical platform of choice for most physical sciences; handles “Big Data” the best</b>                 | <b>Statistical platform of choice for env and ecological sciences</b>                        |
| <i>Making Maps</i>                    | <b>Easiest option for Beginners</b>   | Good option for Beginners  | Requires willingness to learn programming fundamentals  | Good for scientists already using R  |
| <i>Visualizing Data</i>               | Some functionality for easy charts and graphs; 3D tools are improving                               | Plugins available (i.e., <a href="#">Diagram Overlay</a> )                   | <a href="#">Great options available</a> but requires willingness to learn programming fundamentals            | <b>Best option, esp. for those already using R</b>   |
| <i>Publishing Data (web)</i>          | ArcGIS Online (free but limited); ArcGIS Server (robust but expensive)                              | Doesn't have built-in option   | <b>Web full of options ( i.e., GeoDjango, Kartograph, Mapnik, MapServer, Flask)</b>                           | <a href="#">R Shiny</a> 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](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

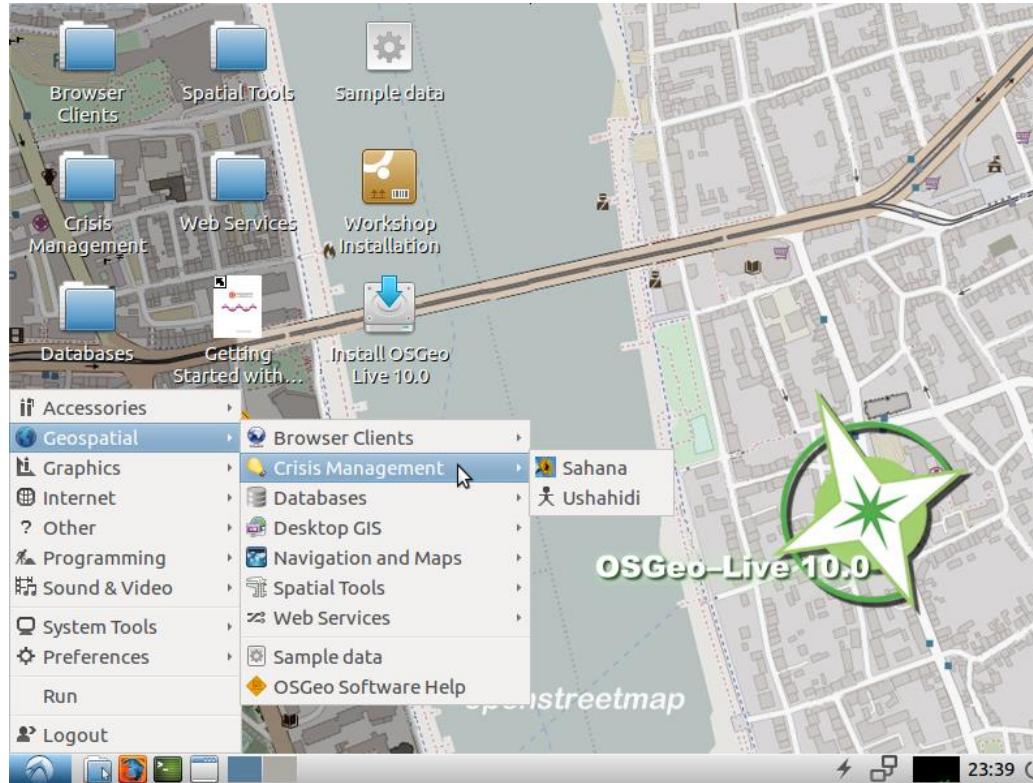
# Vector Processing with OGR



- 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>