

Predictive Analytics

Geo Analytics – Part I

Prof. Dr. Volker Herbort

Contents

- Introduction
- Geospatial Data
- Choropleth Maps
- Points
- Geo Analytics

■ Why Geo Analytics?

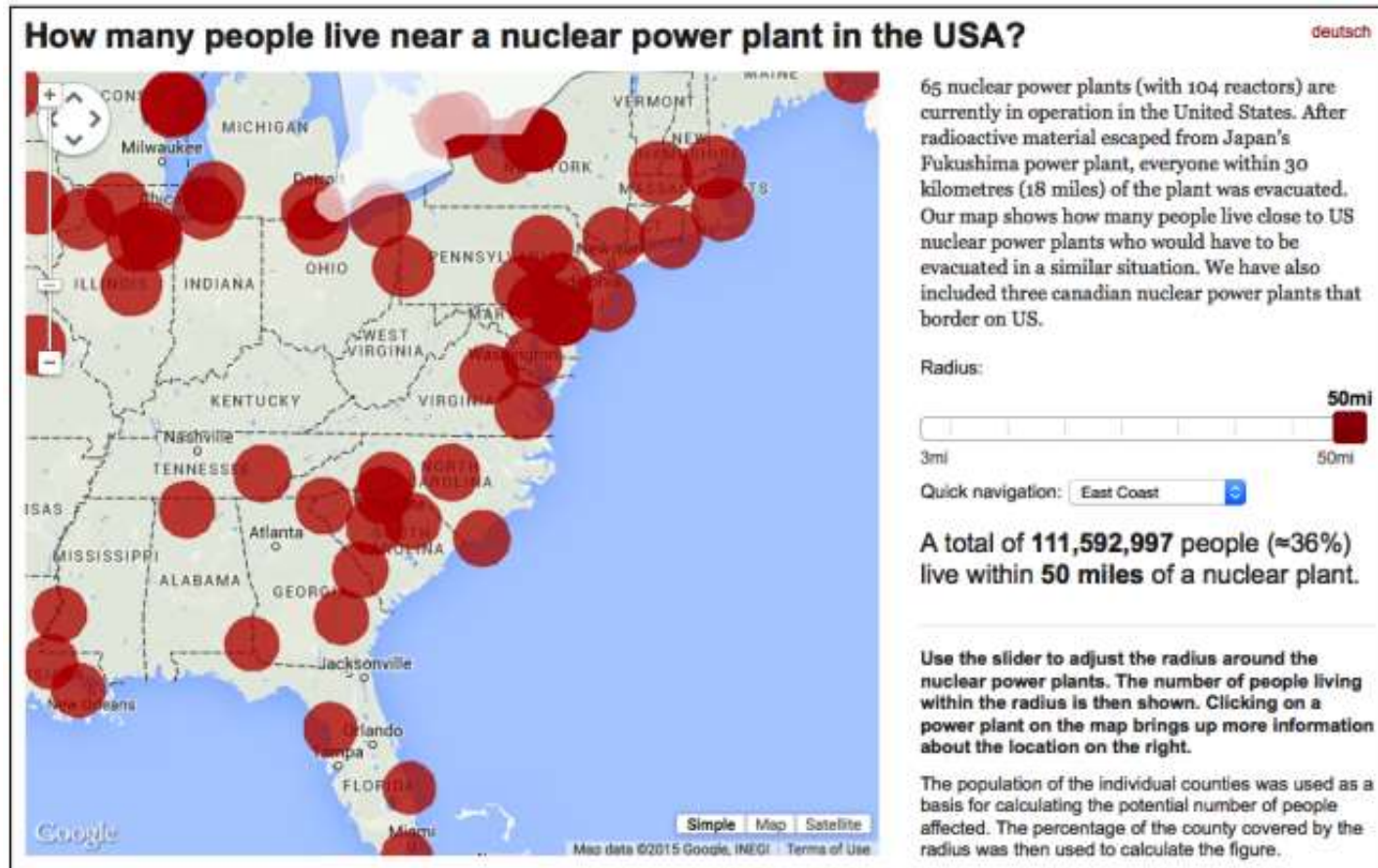
- Much of the data that is produced e.g. by business models has geospatial aspects
- ...or can at least be traced back to some location over space.
 - e.g. Transportation, Logistics, Location Based Services..

■ Geo Analytics

- Geospatial data has additional aspects.
- Analytical methods require special spatial handling.

- Answering geospatial questions:
 - How long does it take to go from A to B?
 - What is the total area of the city of Ulm?
 - What is the best place to built a new bridge?
 - What is the shortest delivery route?
 - How much energy will be produced by a wind turbine at a certain location?
 - How many poeple live near nuclear power plants?

Answer using Mash-Up



<http://opendata.zeit.de/nuclear-reactors-usa>

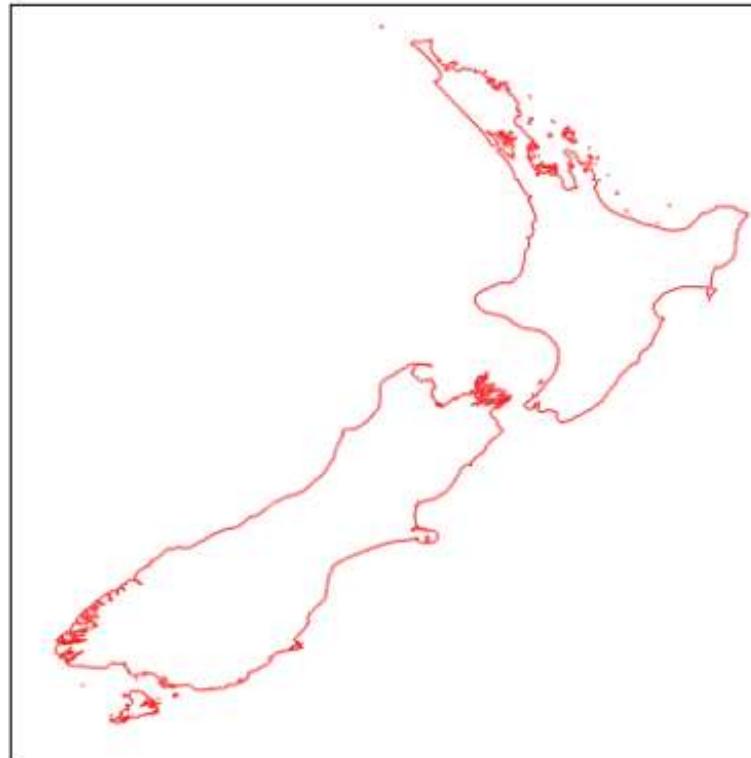
Contents

- Introduction
- **Geospatial Data**
- Mapping Data
- Points
- Geo Analytics ML

- Geospatial Data
 - Series of coordinates
(usually pairs of latitude and longitude)
 - Additional attributes such as time, height, place names or temperature
- Geographical Information Systems (GIS)
 - Dedicated systems for working with geospatial information
 - Maps
 - Networks e.g. water, gas or electricity

Example: Geospatial Data

What is displayed? How much data?



Source: [1] p. 5

Core GIS Concepts

- The earth is not flat...



...which makes GIS life a lot harder!

Location (geodetic)

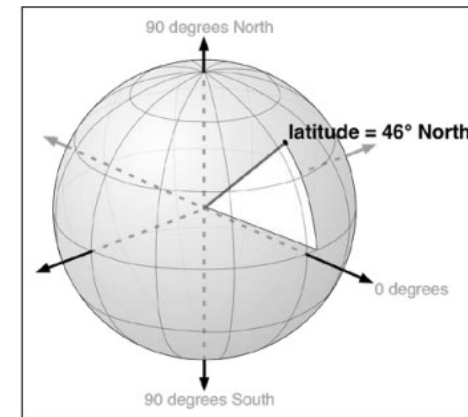
■ Points on the surface of the Earth

■ latitude (lat)

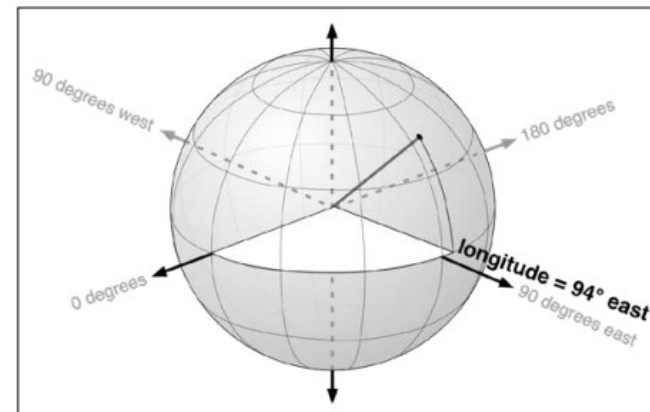
- Angle between point and equator (North-South direction)
- + North / - South

■ longitude (lon)

- Angle between point and Greenwich (England)
- + East / - West

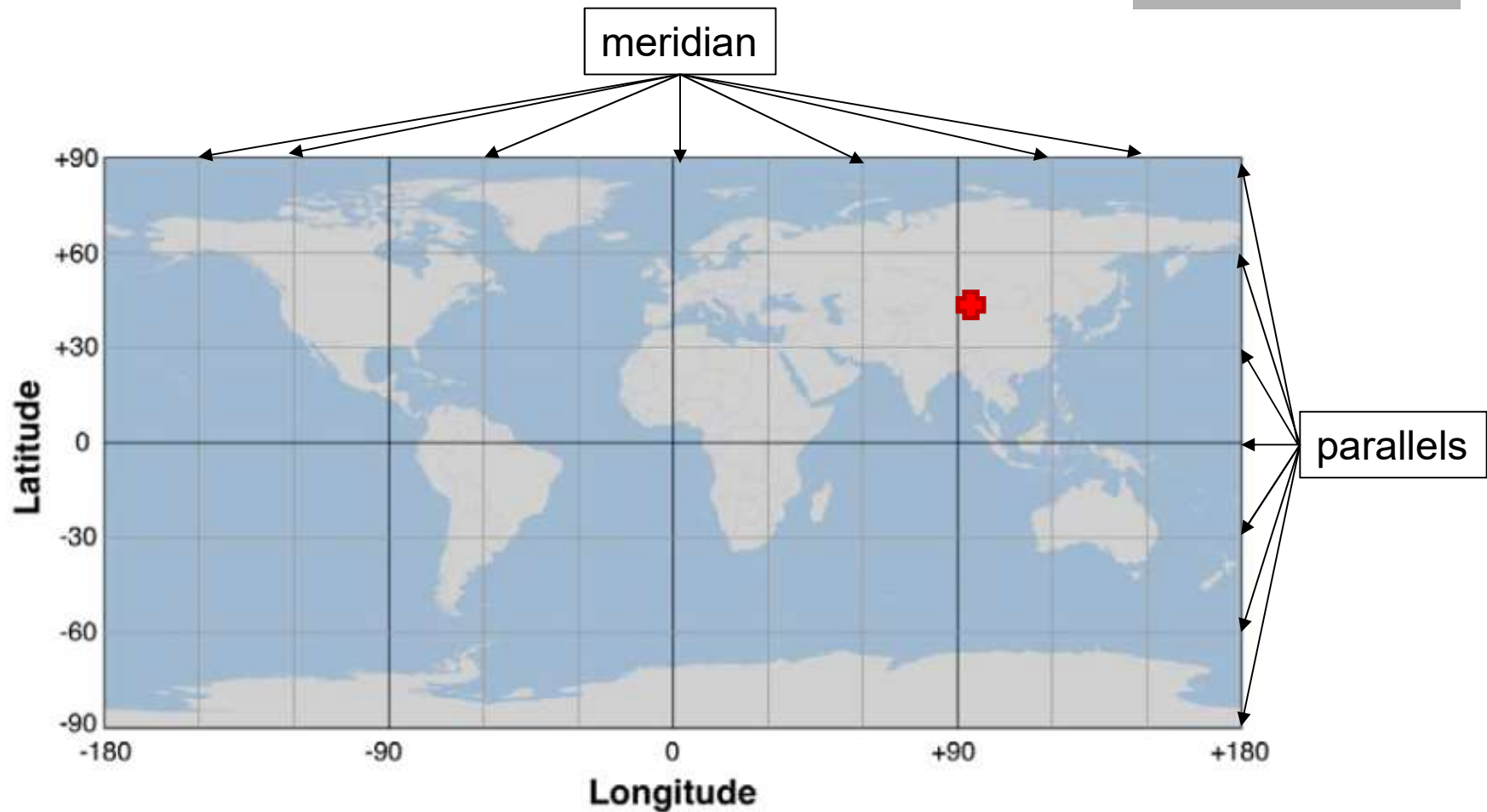


Source: [1] p. 17



Source: [1] p. 18

Location Ctd.



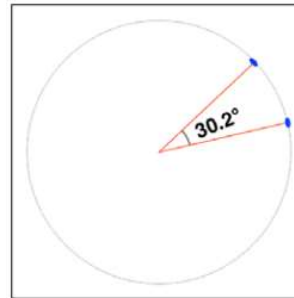
Source: based on [1] p. 18

Other Locations

- Civic Location
 - Street Address
- Jurisdictional Location
 - governmental boundary
 - Electoral ward, county or city
- Translate different types of location
 - Geocoding: Address -> lat, lon
 - Reverse Geocoding: lat, lon -> (nearest) Address

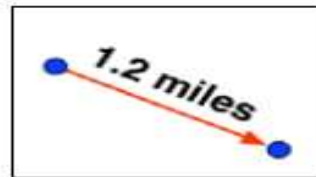
■ Different interpretation

■ Angular distance



Source: [1] p. 19

■ Linear distance ("as the crow flies")



Source: [1] p. 20

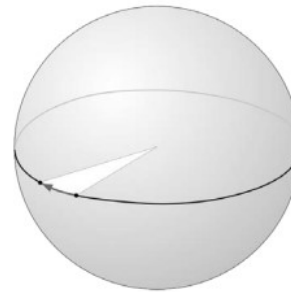
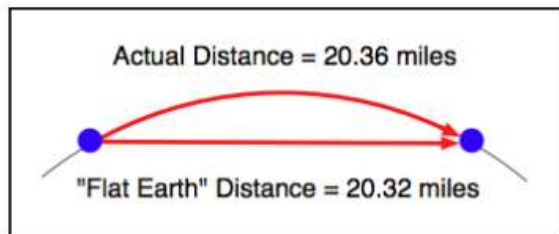
■ Traveling distance



Source: [1] p. 20

Distance ctd.

- Okay, but the Earth is not flat



Source: [1] p. 21

- Actually, **great-circle Distance**
 - Can be calculated from lat lon
 - Haversine Formula

Haversine Formula[4]

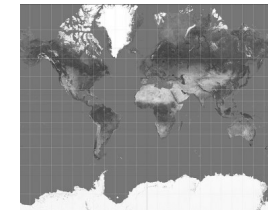
- Calculate distance d between two points $A(\text{lon}_1, \text{lat}_1)$ and $B(\text{lon}_2, \text{lat}_2)$ and Earth with radius R
- Algorithm
 - $d_{\text{lon}} = \text{lon}_2 - \text{lon}_1$
 - $d_{\text{lat}} = \text{lat}_2 - \text{lat}_1$
 - $a = (\sin(d_{\text{lat}}/2))^2 + \cos(\text{lat}_1) * \cos(\text{lat}_2) * (\sin(d_{\text{lon}}/2))^2$
 - $c = 2 * \text{atan2}(\text{sqrt}(a), \text{sqrt}(1-a))$
 - $d = R * c$
- Explanation
 - d has same unit as R
 - R is 6.371 km

Projection[1]

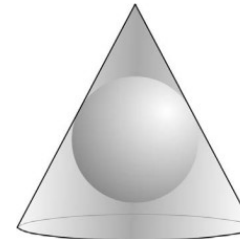
- Maps are two-dimensional projections of the Earth
- Different types of projection
 - cylindrical
 - Mercator projection
 - equal-area cylindrical projection
 - universal transverse Mercator projection
 - conical
 - Albers equal-area projection
 - Lambert conformal conic projection
 - equidistant projection.
 - azimuthal
 - gnomonic projection
 - Lambert equal-area azimuthal projection
 - orthographic projection



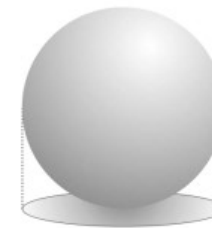
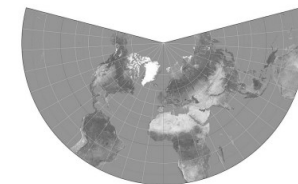
Source: [1] p. 24



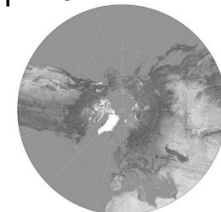
Source: [1] p. 25



Source: [1] p. 26

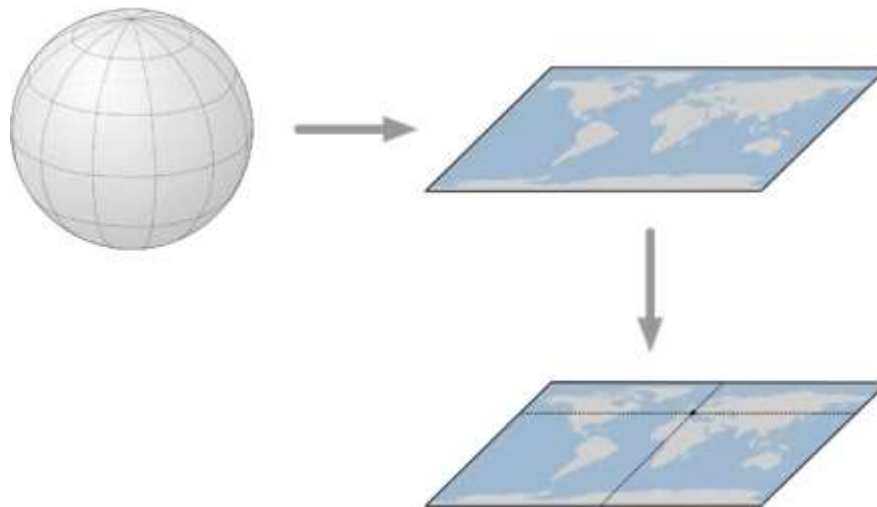


Source: [1] p. 27



Projected Coordinate Systems

- Points on the map but not the earths surface



Source: [1] p. 29

- Mathematical Model of the Earth
- Important Datums
 - NAD 27: North American Datum of 1927.
 - only USA
 - NAD 83: North American Datum of 1983.
 - USA, Canada, Mexico and Central America
 - WGS 84: World Geodetic System of 1984.
 - Global datum
 - Used by Global Positioning System (GPS) satellites
 - All data captured by GPS units use this datum.

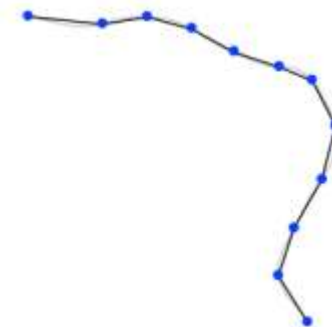
■ Geospatial data is often represented as shapes

■ Point •

- A coordinate, described by two or more numbers
- Projected or unprojected coordinate system

■ Path

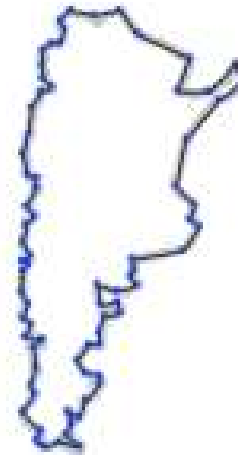
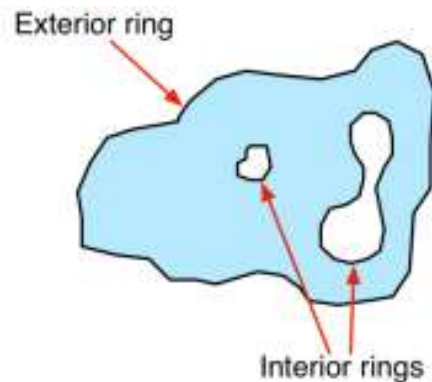
- generally described using what is called a *LineString*
- connected series of line segments
- a.k.a. PolyLines
- e.g. roads, rivers



Source: [1] p. 32

■ Outline

- usually a polygon
- e.g. country, lakes, cities



Source: [1] p. 33

- Geospatial Data contains
 - Geographical features
 - Meta-data (projection, datum, coordinate system, ...)
 - Attributes (name, population, average temp., ...)
 - Display information (color or line style)
- Two different main types of data
 - Raster format data
 - Vector format data

Geospatial Data Formats ctd.

- Raster format data: bitmapped images
 - **Digital Raster Graphic (DRG)**
 - **Digital Elevation Model (DEM)**
 - **Band Interleave by**
 - Line (BIL)
 - Pixel (BIP)
 - **Band Sequential**

Geospatial Data Formats ctd.

- Vector format data: Points, Lines and Polygons
 - **Shapefile** from Environmental Systems Research Institute (ESRI)
 - **Simple Features** as defined by Open Geospatial Consortium (<http://www.opengeospatial.org/>)
- Additional data exchange formats:
 - Well-known Text (WKT)
 - Well-known Binary (WKB)
 - GeoJSON
 - Geography Markup Language (GML)

OGC Simple Features

