

SEST-6577

Geographic Information Systems for Security Studies

Lecture 01 (GIS 101)

Yuri M. Zhukov
Associate Professor
Georgetown University

September 5, 2024



Figure 1: Nighttime luminosity



Figure 2: Ptolemy's geography of the earth, 2nd century C.E.

History of GIS

History of geospatial data analysis / Public health



Figure 3: The blue wraith

In 1854, there was an outbreak of Cholera in London's Soho district.
It killed 616 people.



Figure 4: The blue death

Cholera is an infectious disease of the small intestine.
It causes severe dehydration in its victims. It is quite deadly.



Figure 5: The experts

The Board of Health struggled to find the epidemic's cause.

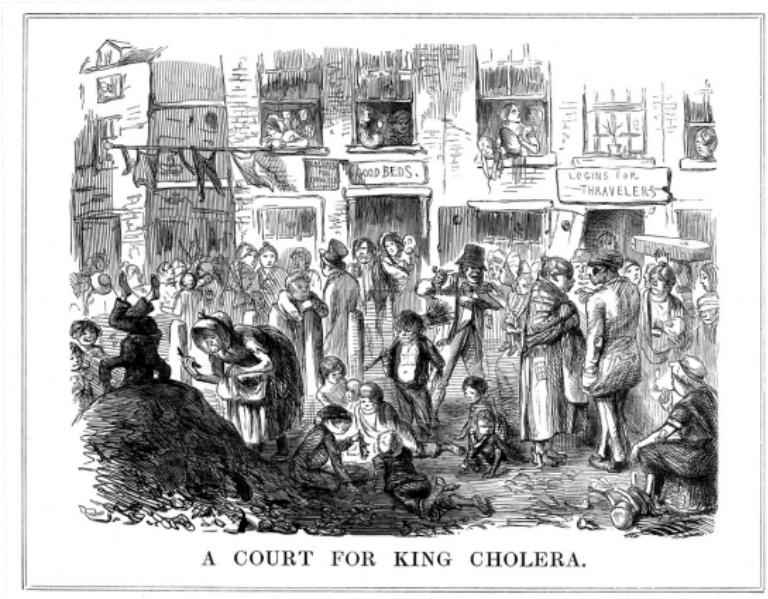


Figure 6: The scapegoats

The leading theory was that Cholera was caused by foul air and poor hygiene.

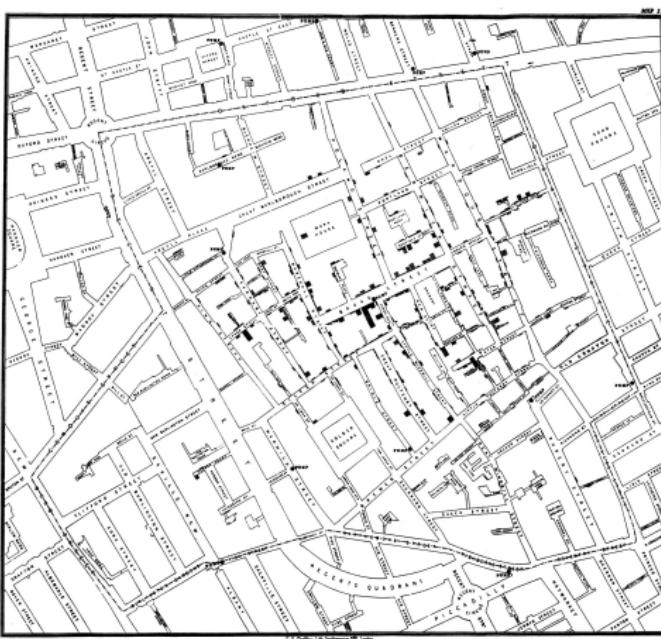


Figure 7: The geospatial data science

Dr. John Snow believed that Cholera was spread by contaminated water, not foul air. To test his theory, he created a map of Cholera cases at each address in Soho.

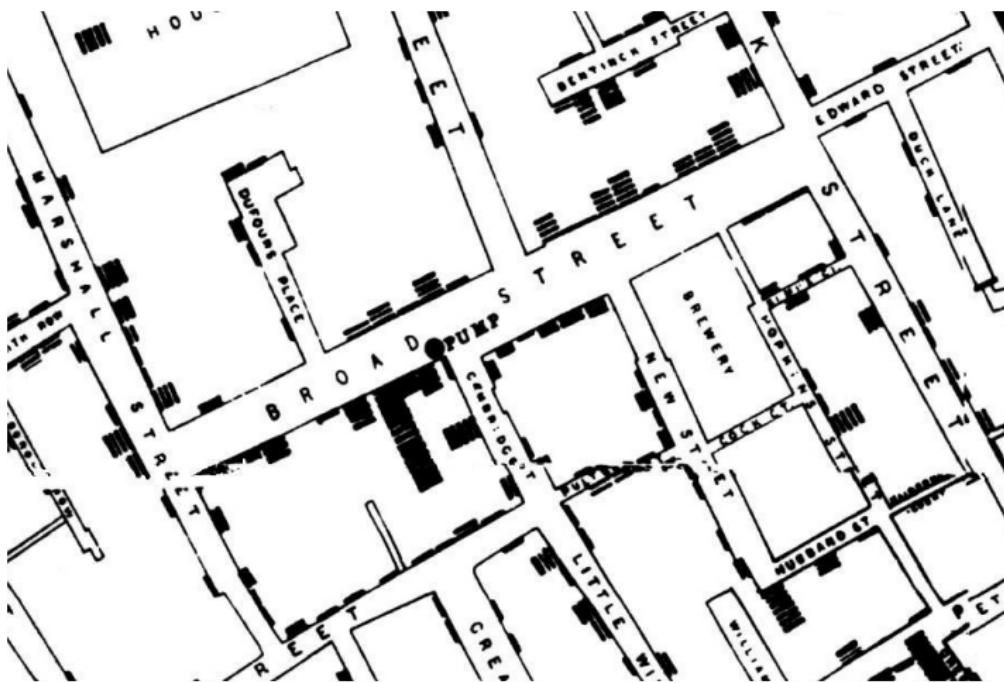


Figure 8: The pump

He found a **cluster** of cases around a water pump on Broad St.



Figure 9: The public health campaign

He told authorities to break pump's handle so people couldn't draw water from it.

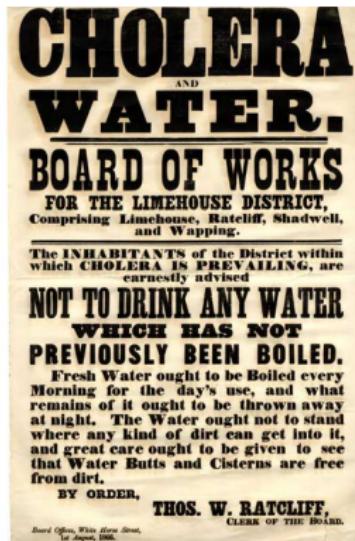


Figure 10: Data scientist has policy impact

Cholera cases drop in Soho, supporting Snow's theory about contaminated water.

History of geospatial data analysis / Military intelligence



Figure 11: A new weapon

During WWII, Germany launched 1,358 V-2 Rockets at London.



Figure 12: No way to intercept, no way to defend

V-2's speed and trajectory made it invulnerable to anti-aircraft guns and fighter jets.



Figure 13: Terror rains from the sky

V-2 strikes kill 2,724 people in UK + 6,000 civilians and military across Europe.
(+ 15,000 concentration camp prisoners died constructing the V-2)

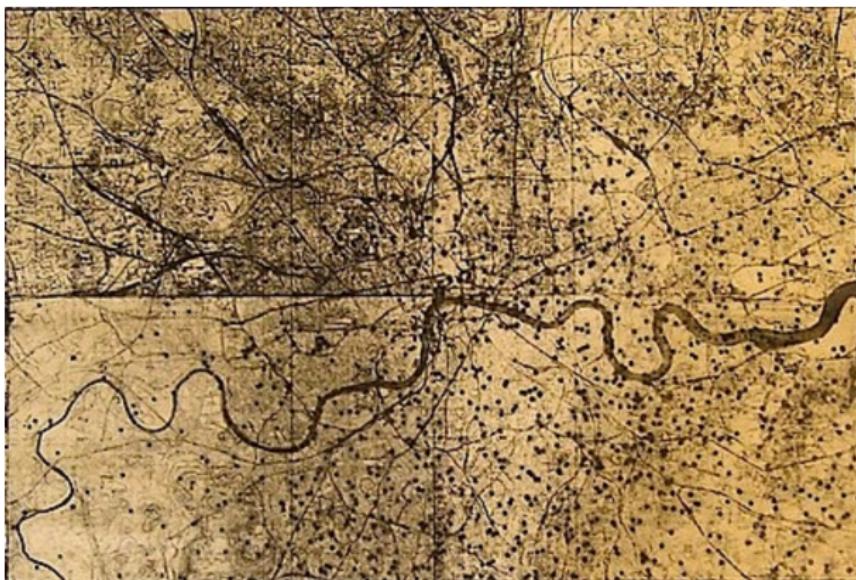


Figure 14: Locations of V-2 strikes in London

Bomb damage maps were interpreted by some analysts as showing that impact sites were **clustered**. This suggested the V-2's guidance system was more sophisticated than intel estimates thought. Allies tried to jam V-2's guidance system, to no effect.

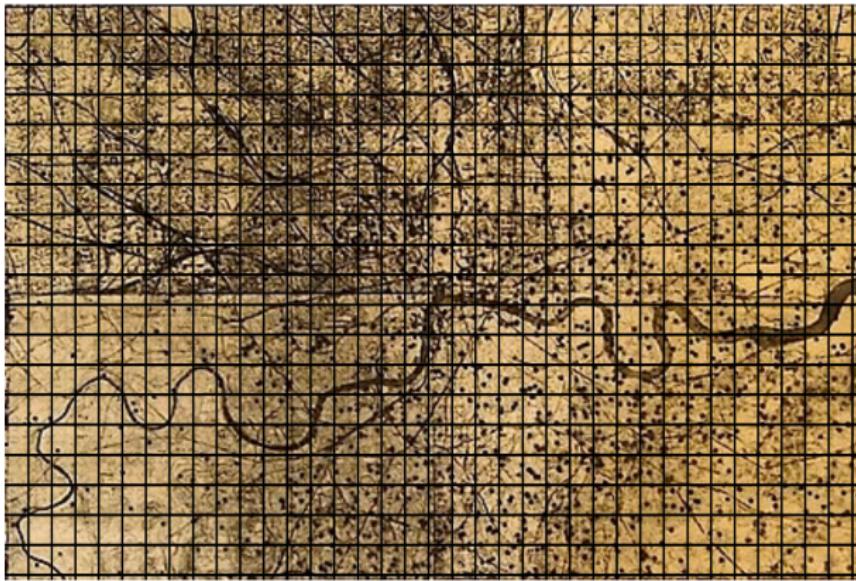


Figure 15: Wartime geostatistics

R.D. Clarke applied a statistical test to assess whether any hard evidence could be found for clustering. For each square, Clark recorded the total number of *observed* bomb hits (537 total in study area), and number of squares with $k = 1, 2, 3, \dots$ hits.

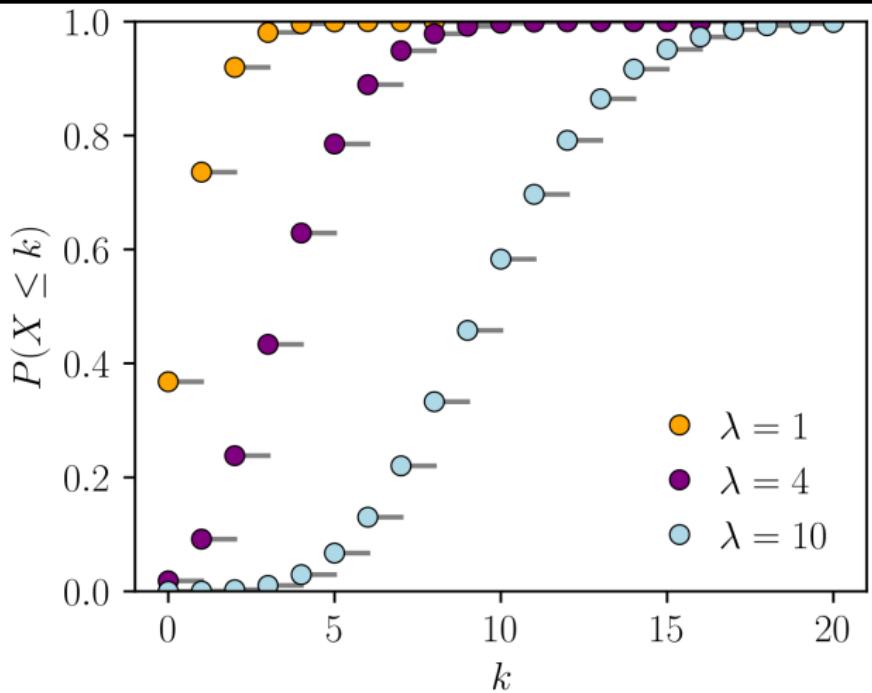


Figure 16: Poisson cumulative density function

Clarke derived the expected number of squares with k hits from the cumulative density function of the Poisson distribution $\sum_{k=1}^n \frac{e^{-\lambda} \lambda^k}{k!}$, with $\lambda = \frac{537}{576}$ and $n = 576$.

No. of bombs per square	Expected	Observed
1	226.74	229
2	211.39	211
3	98.54	93
4	7.14	7
5+	1.57	1

The distribution of observed V-2 strikes conformed quite closely to the Poisson distribution ($\chi^2 = 1.17, p = 0.88$). If strikes were clustered, we would have seen many more squares with a high number of bombs.

Conclusion: V-2 impact sites were **random**, not clustered.

Rocket strikes were indiscriminate (within city of London), not targeted.

Contemporary uses of geospatial data analysis

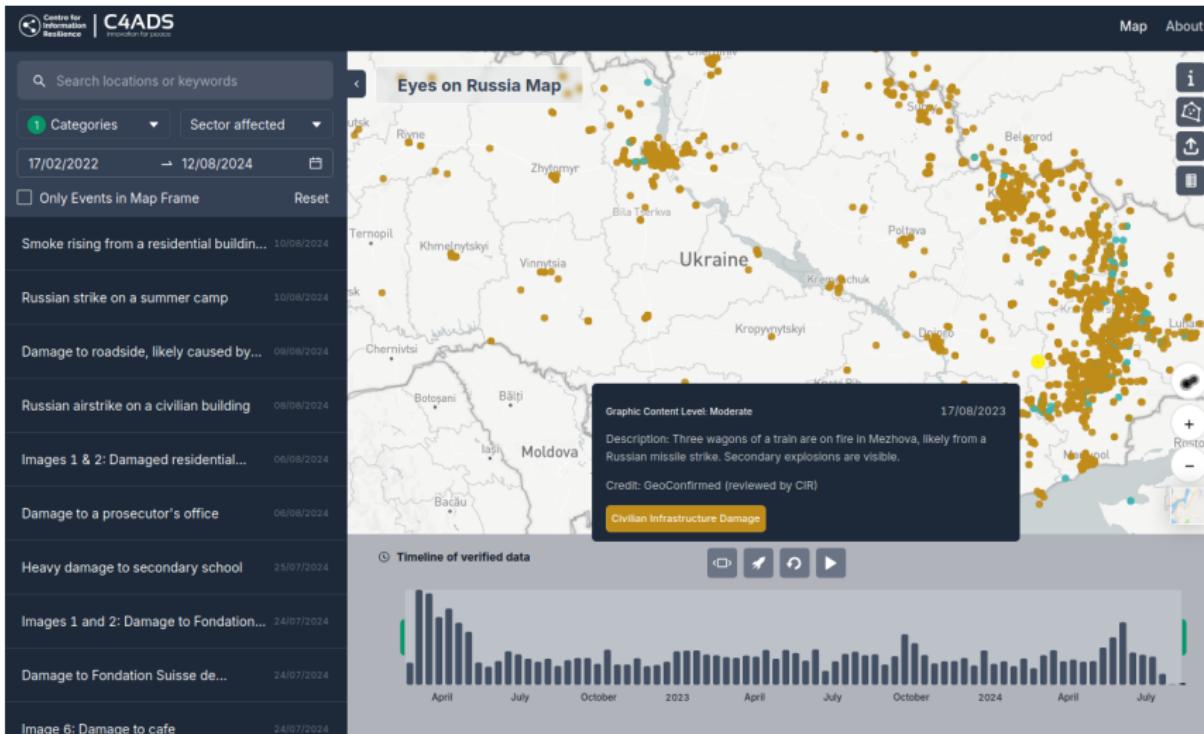


Figure 17: Example: Track civilian infrastructure damage in Ukraine

History of GIS

GIS Basics

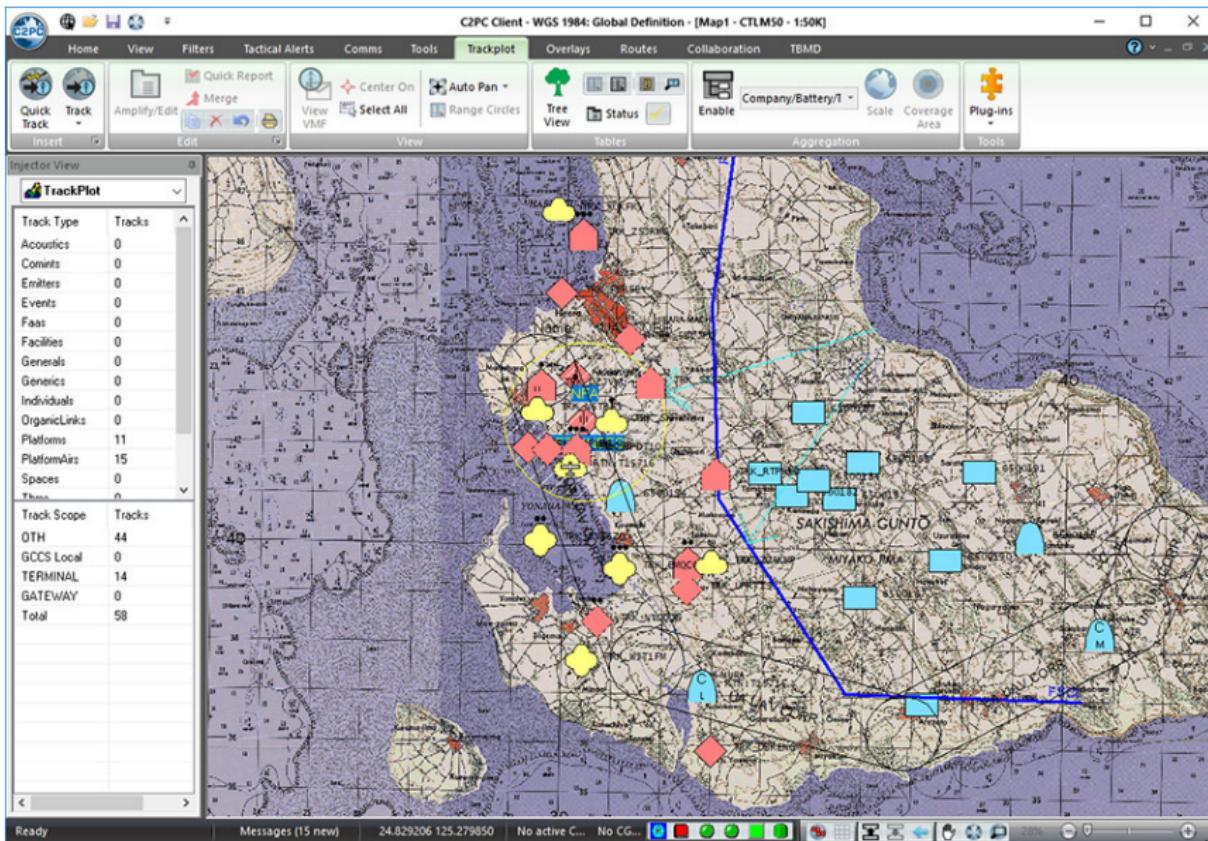


Figure 18: Example: Provide situational awareness during military operations

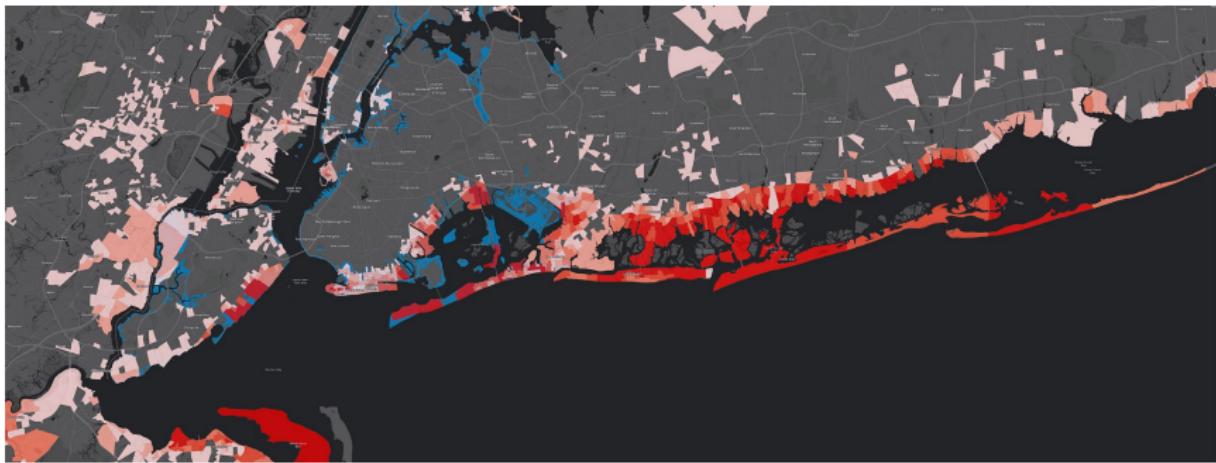


Figure 19: Example: Provide real-time information for emergency management

History of GIS

GIS Basics

The New York Times
Published: July 8, 2015

Mapping Segregation

New government rules will require all cities and towns receiving federal housing funds to assess patterns of segregation.

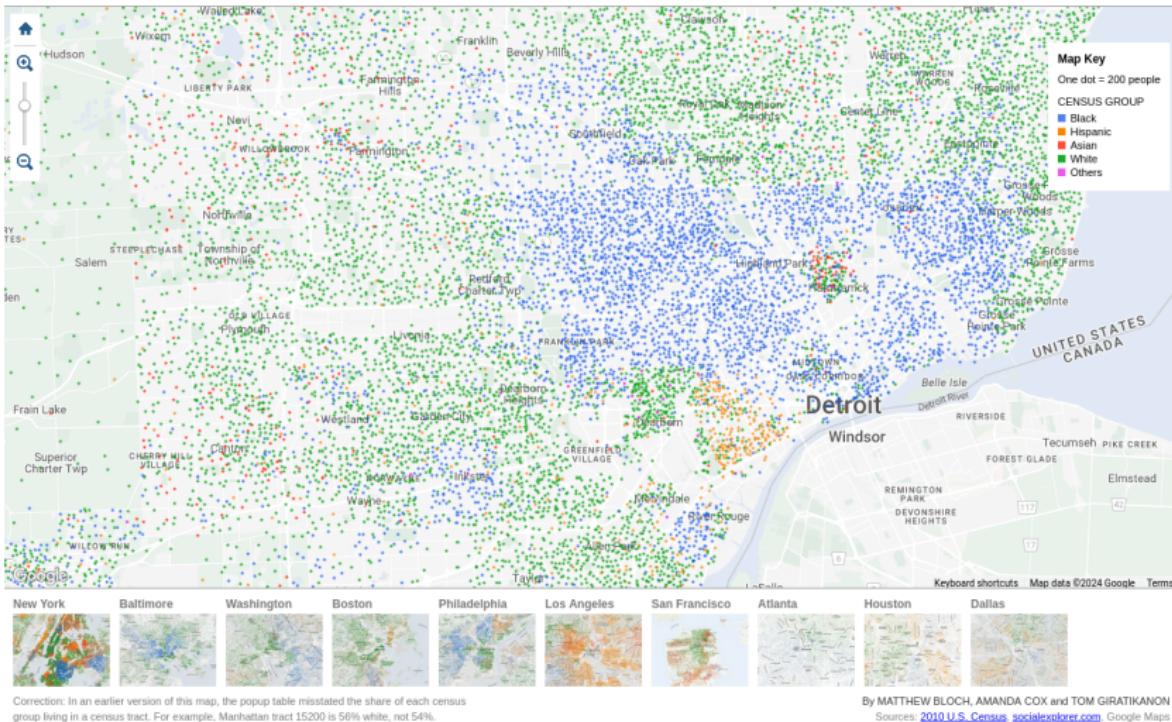


Figure 20: Example: Analyze residential segregation in American cities

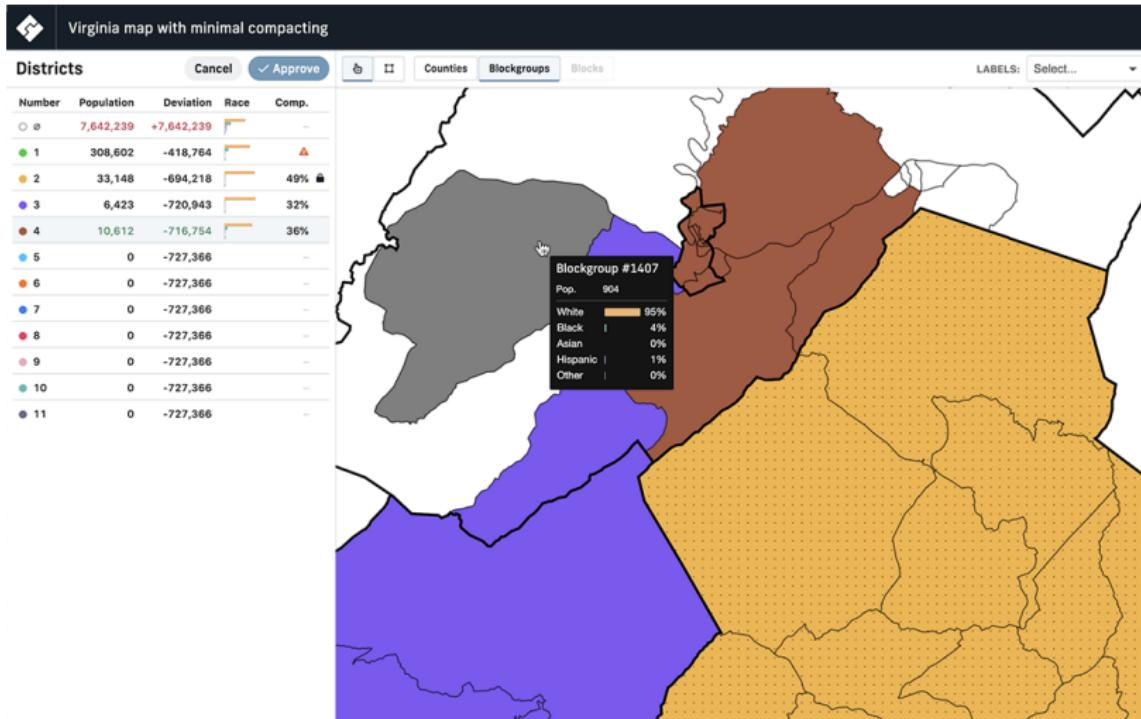


Figure 21: Example: Draw new legislative districts

History of GIS

GIS Basics

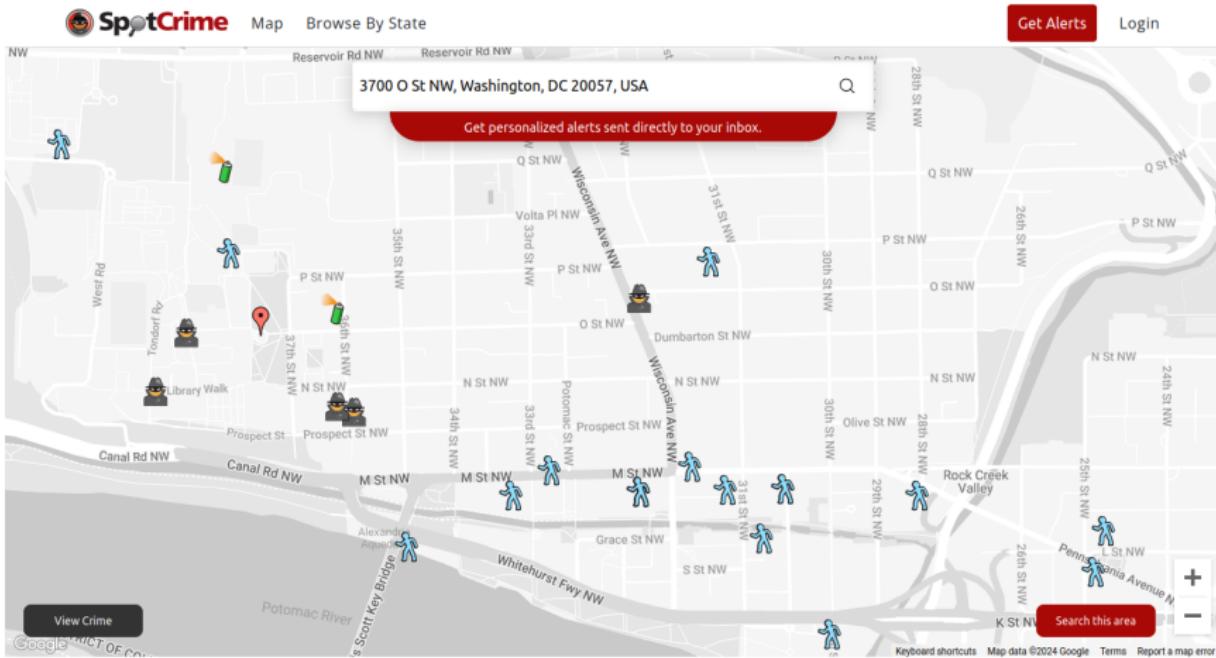


Figure 22: Example: Identify crime hotspots

GIS Basics

Types of spatial data

geospatial = situated in *geographic space*

space is about more than geography

- “space” refers to any dimension for which a notion of distance between objects can be defined (e.g. social networks, trade, culture, ideology)
- “geographic space” refers to Earth’s surface and near-surface

geospatial data = information on “where” + “what”

where: absolute and relative locations of features

(e.g. coordinates, distance, clustering, dispersion)

- dimension 1: x , horizontal position, longitude, easting
- dimension 2: y , vertical position, latitude, northing
- dimension 3: z , elevation, altitude, depth

what: properties and attributes of those features

(e.g. vote share, number of fatalities, temperature)

spatio-temporal data = info on “where” + “when” + “what”

when: absolute and relative timing of observation

(e.g. year, day, electoral cycle, round)

- dimension 4: t , time

Example of multi-dimensional data / Battles in space and time

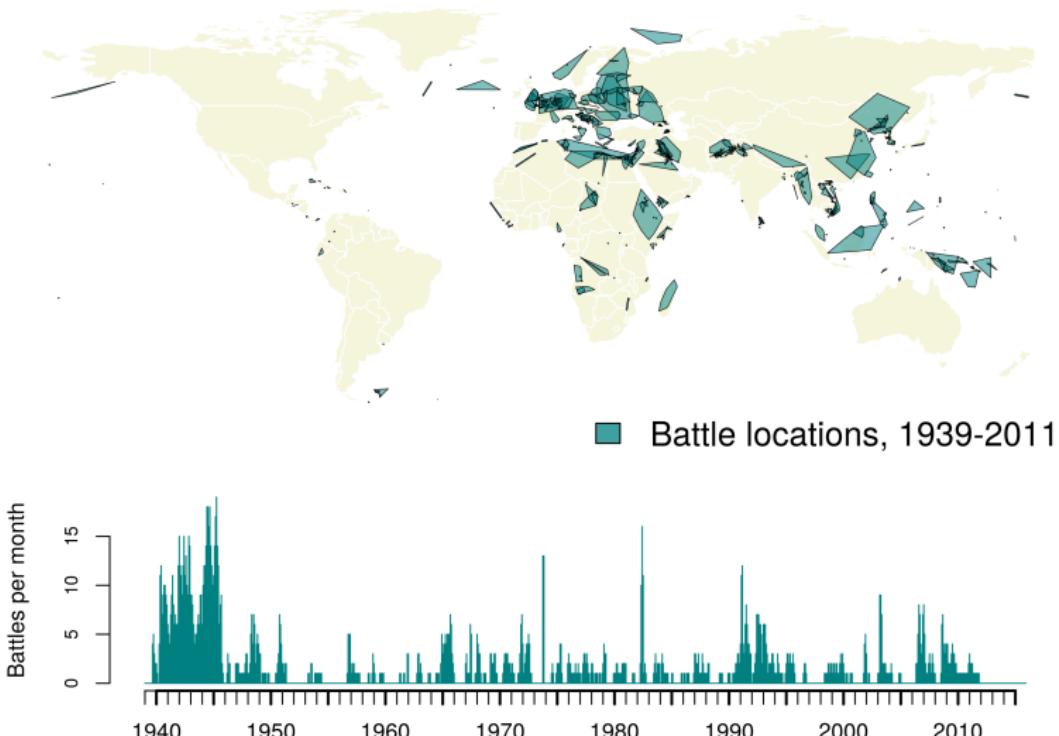




Figure 23: Let's add another dimension

Key: **red lines** denote pairs of battles with common participant

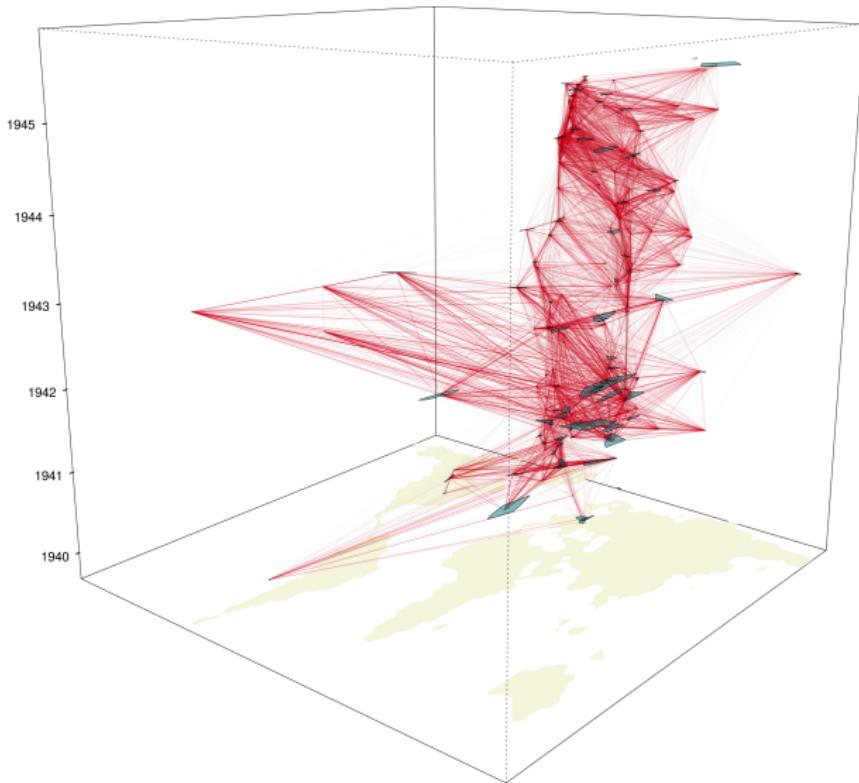


Figure 24: WWII battles in multidimensional space

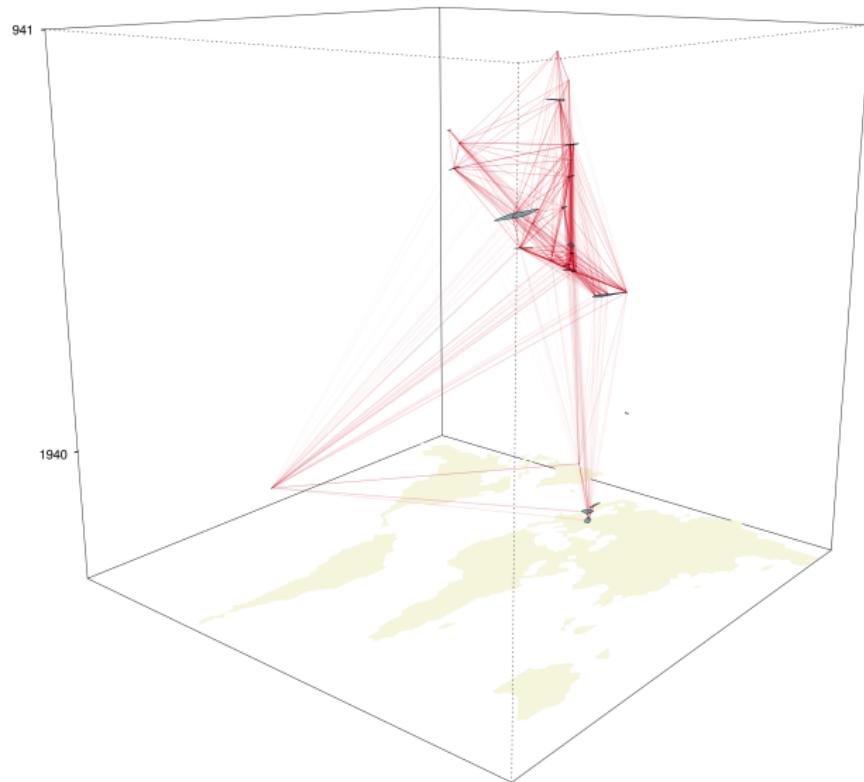


Figure 25: Battles (1939-1941), linked by combatant

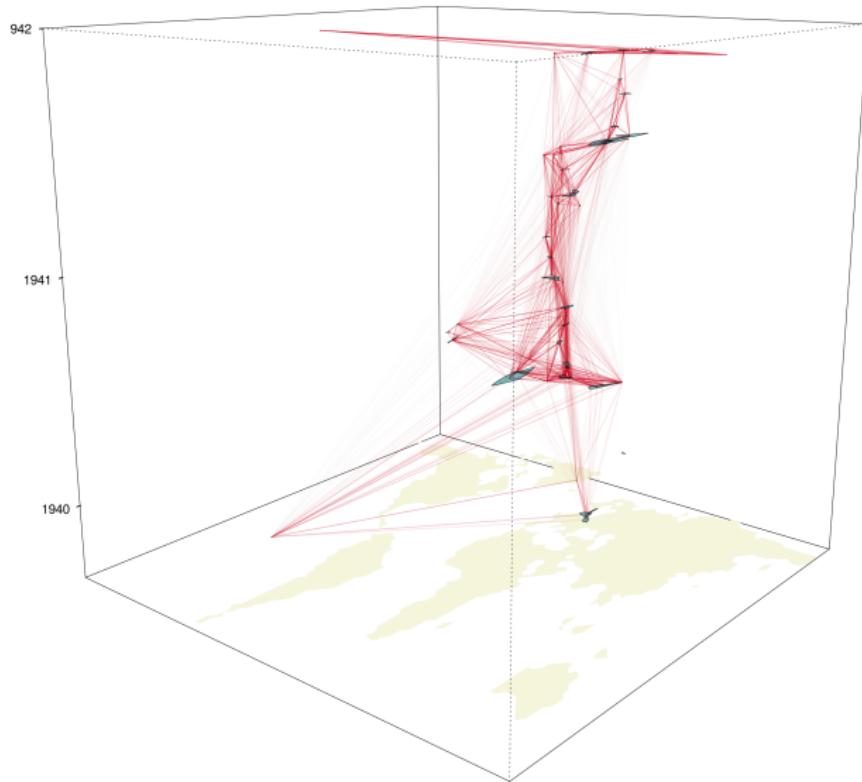


Figure 26: Battles (1939-1942), linked by combatant

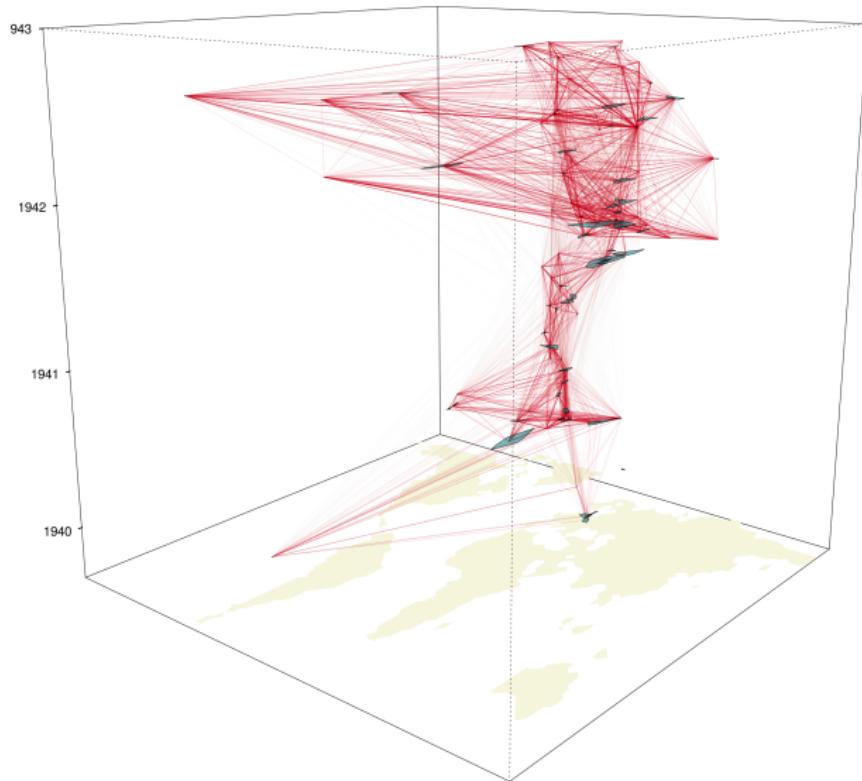


Figure 27: Battles (1939-1943), linked by combatant

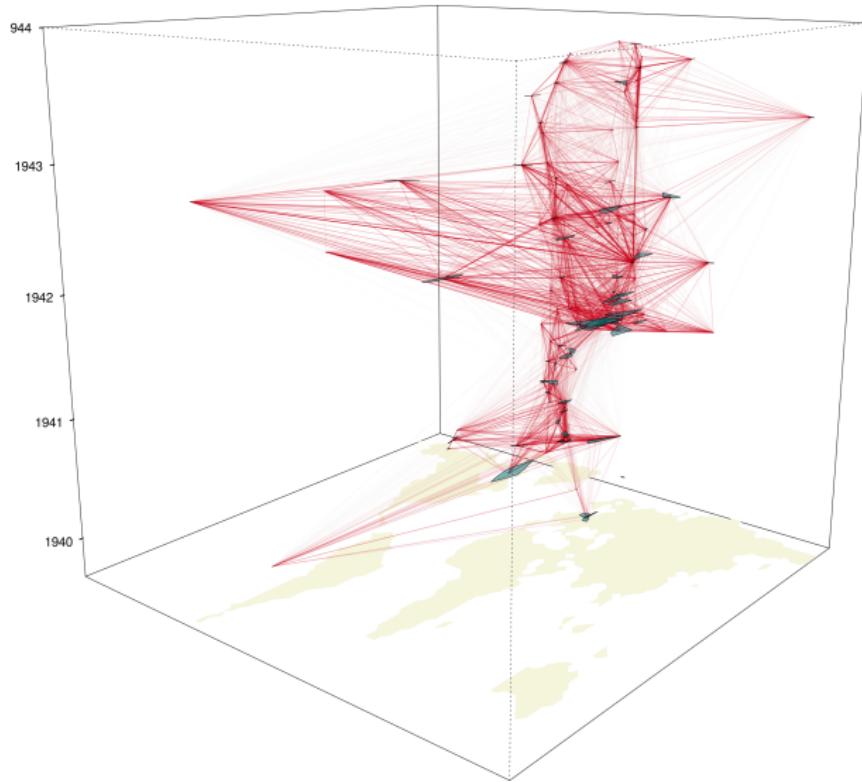


Figure 28: Battles (1939-1944), linked by combatant

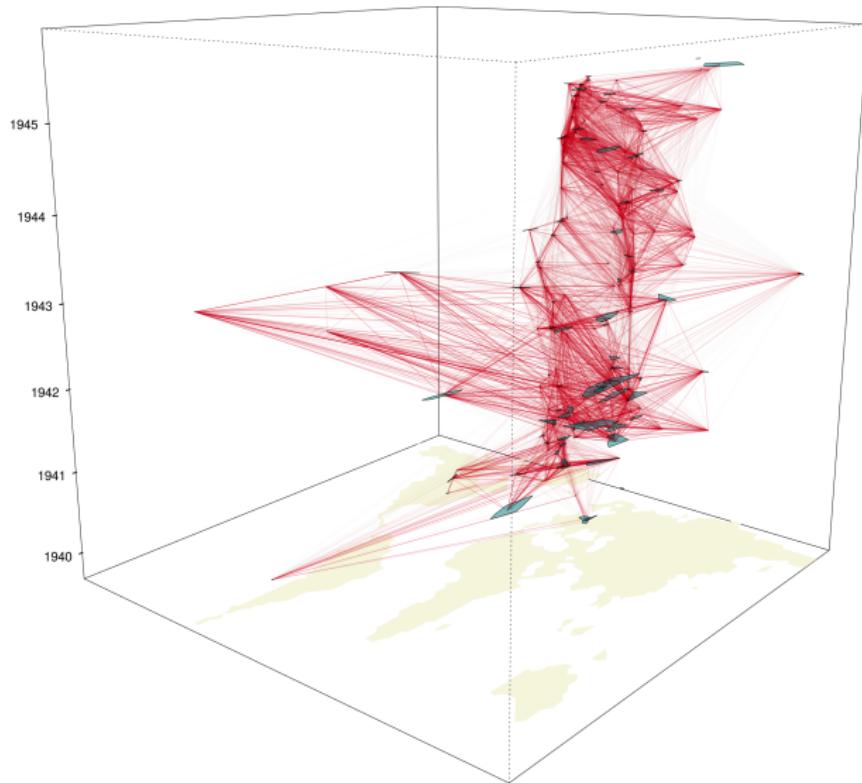


Figure 29: Battles (1939-1945), linked by combatant

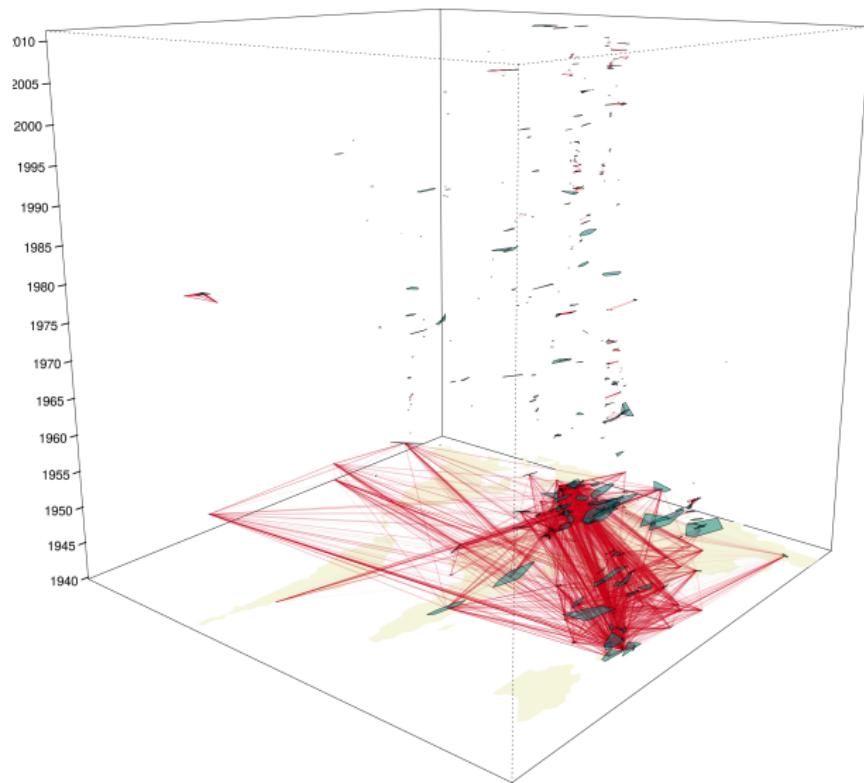


Figure 30: Battles (1939-2011), linked by combatant

Vector data

discrete objects in space

- *point*: pair of coordinates
(e.g. small objects, events)
- *polyline*: open, connected set of points
(e.g. roads, rivers)
- *polygon*: closed, connected set of points
(e.g. countries, administrative units)

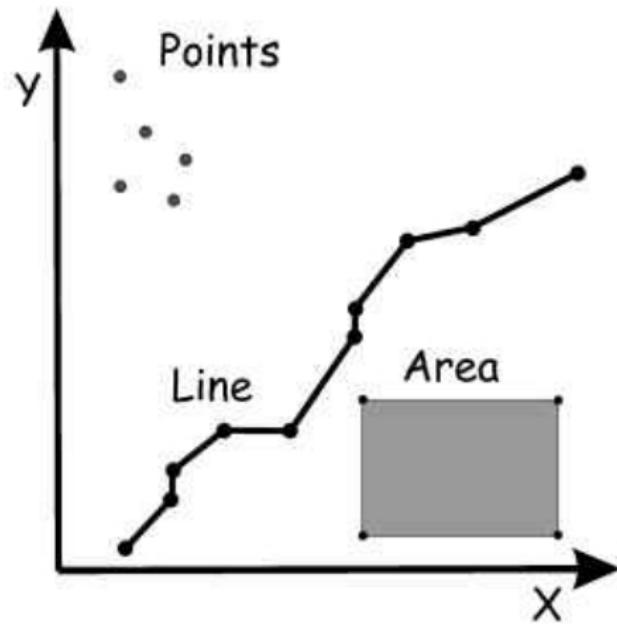


Figure 31: Vector data objects



Figure 32: Points



Figure 33: Polyline

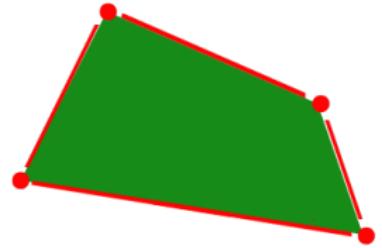


Figure 34: Polygon

Raster data

space as continuous field

- *image*: regular, equally-spaced grid
- *pixel*: individual grid cell
- each pixel represents value or presence/absence of some quantity of interest (e.g. temperature, rainfall, elevation, land cover)

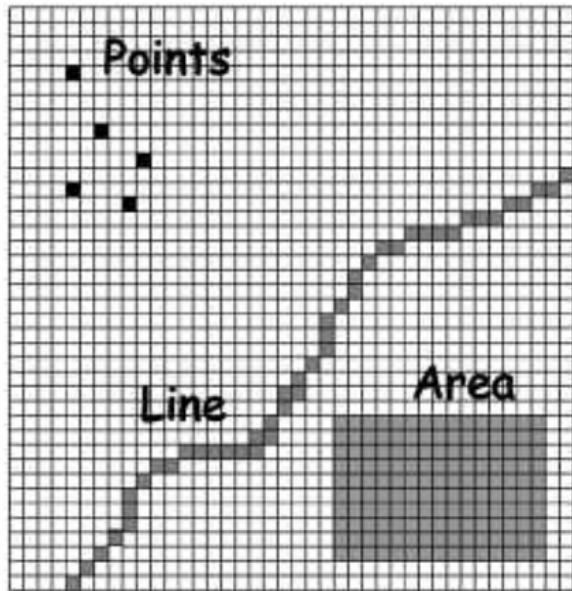


Figure 35: Raster data

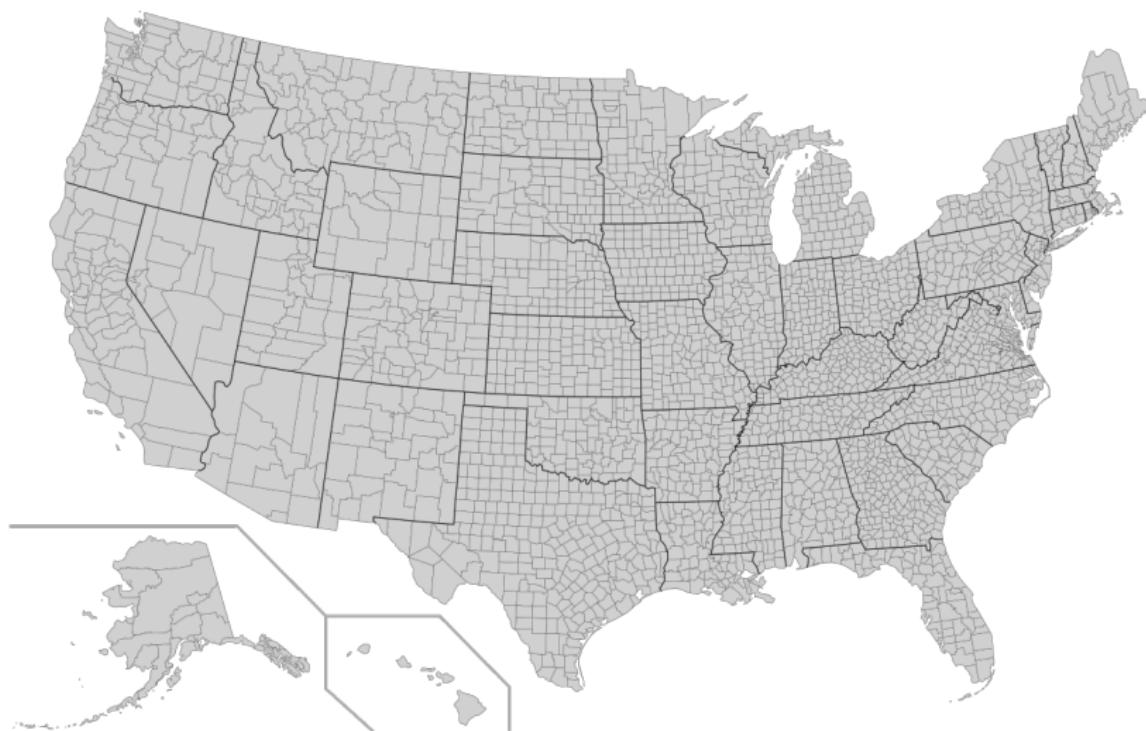


Figure 36: Vector or raster?



Figure 37: Vector or raster?

Red Army soldiers in WWII, by birth location

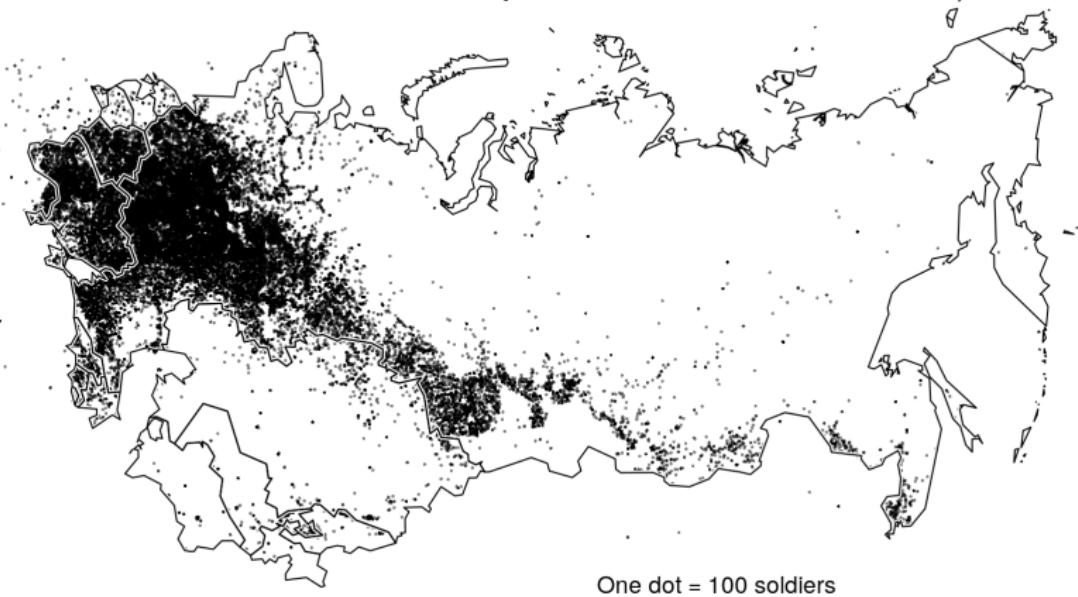


Figure 38: Vector or raster?

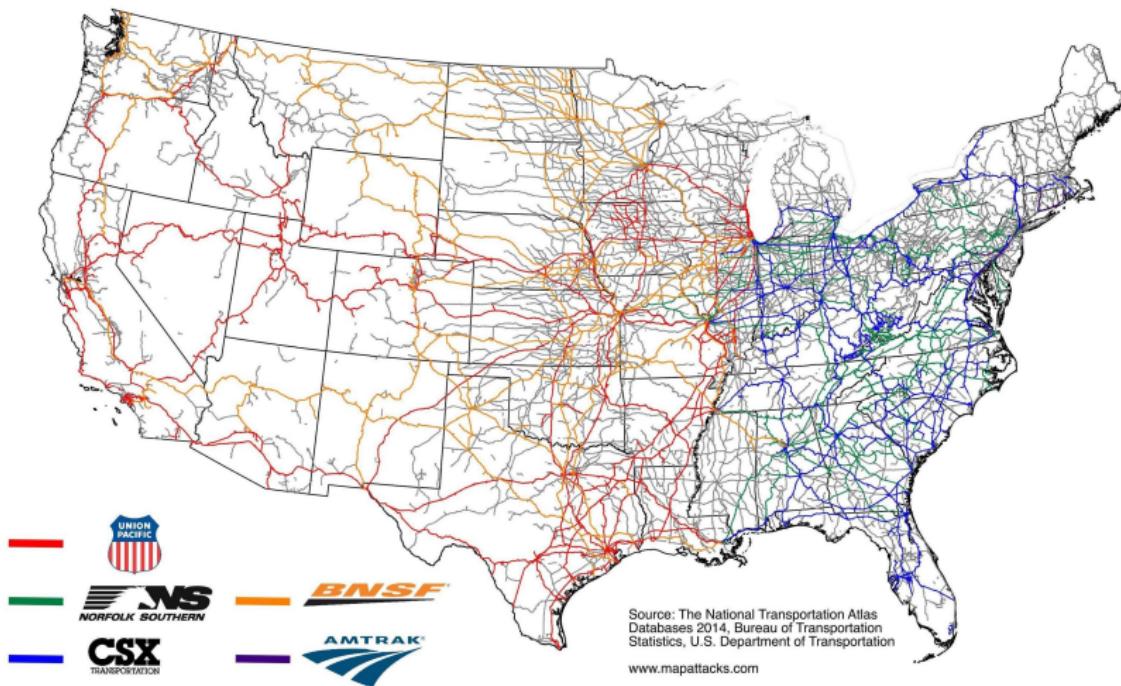


Figure 39: Vector or raster?

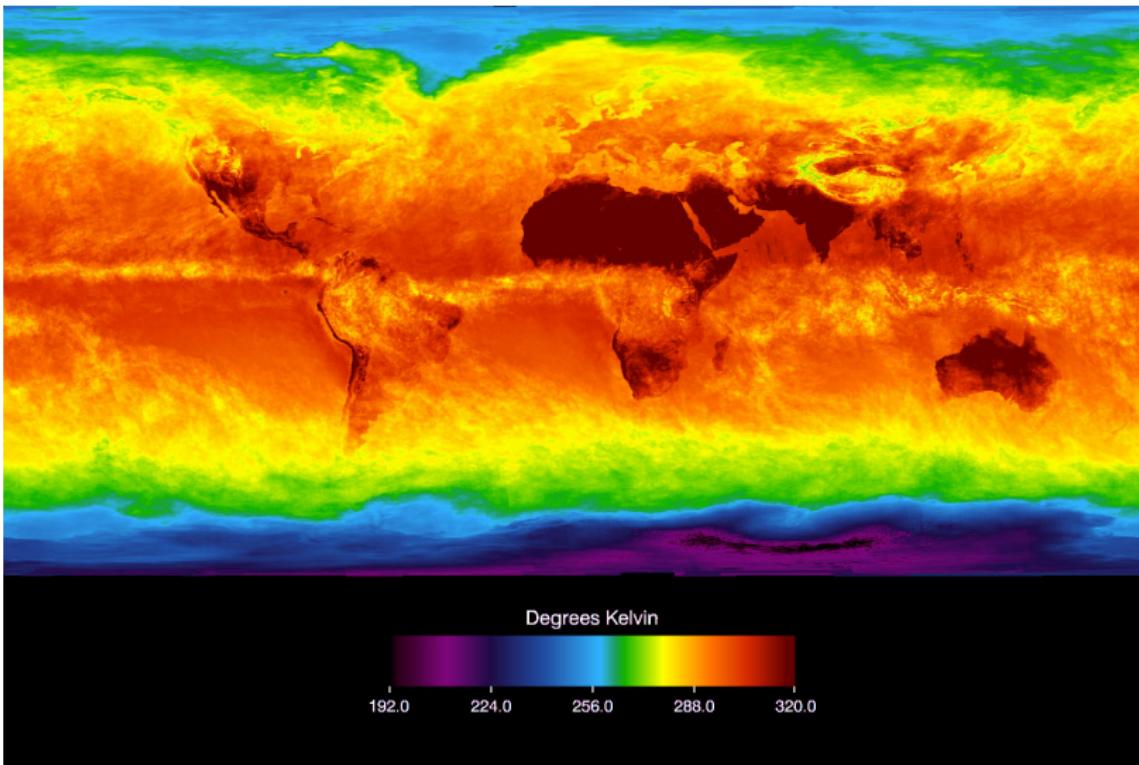


Figure 40: Vector or raster?

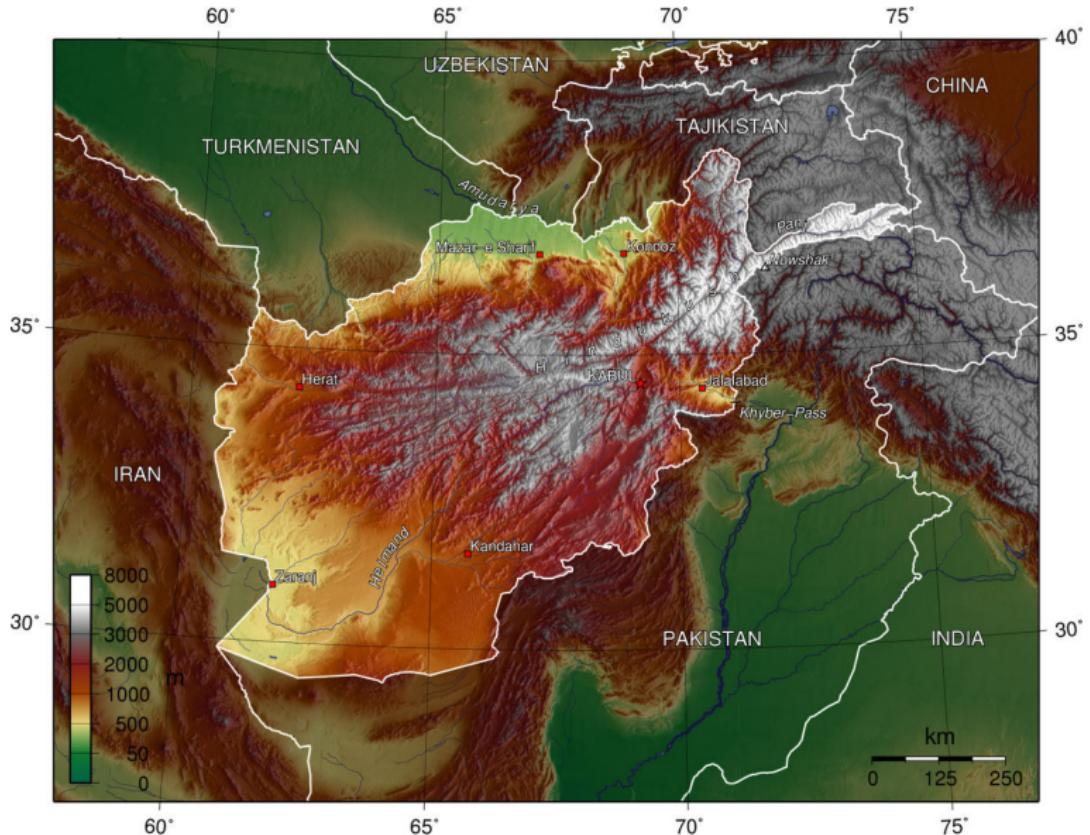


Figure 41: Vector or raster?

Where to find (free/open-source) spatial data?

Coordinates and basemaps:

- Geographical place names: geonames.org
- Global administrative units: geoboundaries.org, gadm.org
- Land cover and elevation: www.usgs.gov/centers/eros

Geo-referenced data:

- Geocoded U.S. Census data: nhgis.org
- City of Cambridge GIS data: cambridgema.gov/GIS
- Environmental data: sedac.ciesin.columbia.edu
- Armed conflict data: x-sub.org, ucdp.uu.se
- Nighttime lights and fires: payneinstitute.mines.edu/eog/
- Electoral districts: electiondataarchive.org, cdmaps.polisci.ucla.edu

A large number of links is also available at

- freegisdata.rtwilson.com
- hgl.harvard.edu
- guides.library.upenn.edu/globalgis

This is *not* a comprehensive list

Data file formats

Vector data:

- GeoJSON (JavaScript Object Notation) is the new standard for vector data
- but points, polylines, polygons are often stored in older Shapefile format
 - each Shapefile includes: shapes/geometries (.shp), positional index (.shx), attribute table (.dbf)
 - sometimes also includes: projection (.prj), spatial index (.sbx), metadata (.shp.xml), other elements
- other common formats include
 - GDB/MBD (File/Personal Geodatabase)
 - KML/KMZ (Keyhole Markup Language, used for Google Earth)
 - OSM (OpenStreetMap's XML-based file format)

Raster data:

- common formats include
 - ASC (ASCII delimited text file)
 - GeoTIFF (georeferenced TIFF image file)
 - IMG (ERDAS Imagine file)
 - DEM (Digital Elevation Model)
 - DTED (Digital Terrain Elevation Data)

Software options

Popular software for the analysis of spatial data

Application	Availability	Learning Curve	Key Functionality
ArcGIS	License	Medium	Geoprocessing, visualization, georeferencing
QGIS	Free	Medium	Geoprocessing, visualization, georeferencing
GRASS	Free	High	Image processing, spatial modeling
Matlab	License	High	Spatial econometrics, basic visualization
Stata	License	Medium	Spatial econometrics, basic visualization
Python	Free	High	Geoprocessing, visualization, geostatistics, spatial econometrics, point processes
R	Free	High	Geoprocessing, visualization, geostatistics, spatial econometrics, point processes

We will be using **QGIS** and **R**

Software & programming

1. QGIS (option 1)

- a) free, open-source alternative to ESRI ArcGIS
- b) visualize, manage, edit, analyze spatial data, create maps
- c) intuitive graphical user interface (GUI)
- d) multiplatform (runs on Linux, Mac, Windows, Android)
- e) download it here: qgis.org

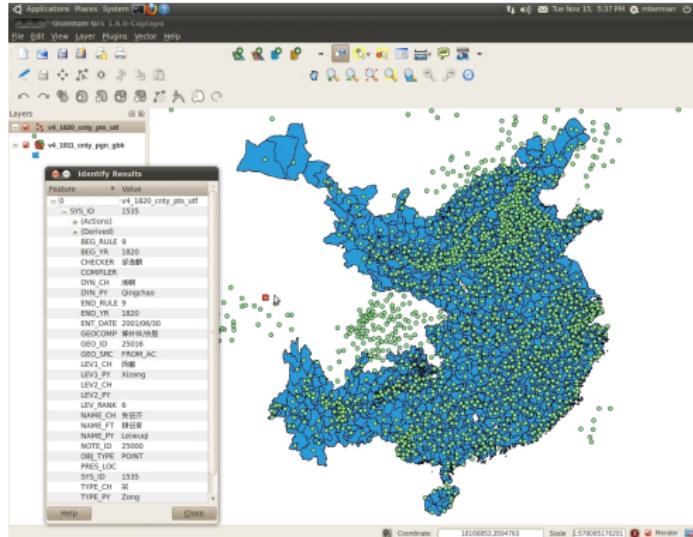
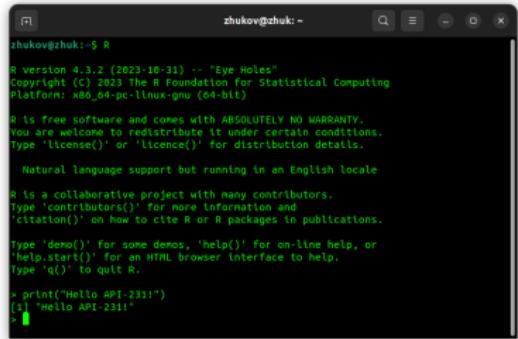


Figure 42: QGIS

Software & programming

2. R (option 2)

- a) open-source statistical programming language
- b) can do (most) of what you can do in QGIS, and lots more
- c) can run R from the command line
 - ... or using source code editor (e.g. Sublime Text, XEmacs)
 - ... or using integrated development environment (e.g. RStudio Cloud)
- d) also multiplatform (runs on Linux, Mac, Windows, Android)
- e) download R here: r-project.org
 - ... or RStudio here: posit.co



```
zhukov@zhuk:~$ R
R version 4.3.2 (2023-10-31) -- "Eye Holes"
Copyright (C) 2023 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.

> print("Hello API-231!")
[1] "Hello API-231!"
>
```

Figure 43: R

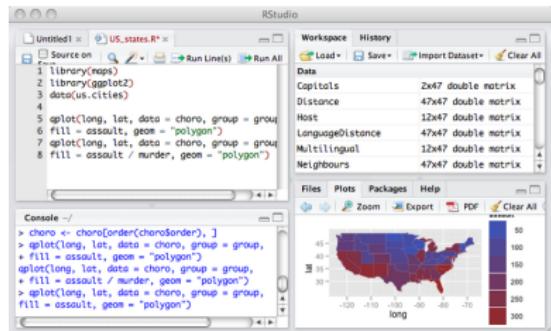


Figure 44: RStudio

3. RStudio Cloud (option 2.5)

- a) same as RStudio, but accessible through web browser
- b) advantages:
 - packages/dependencies already installed
 - no software to download
- c) all R lab exercises will be made available through RStudio Cloud
- d) you can access it through link posted on Canvas
- e) set up RStudio Cloud account w/ your georgetown.edu credentials
- f) link to sign-up page: posit.cloud

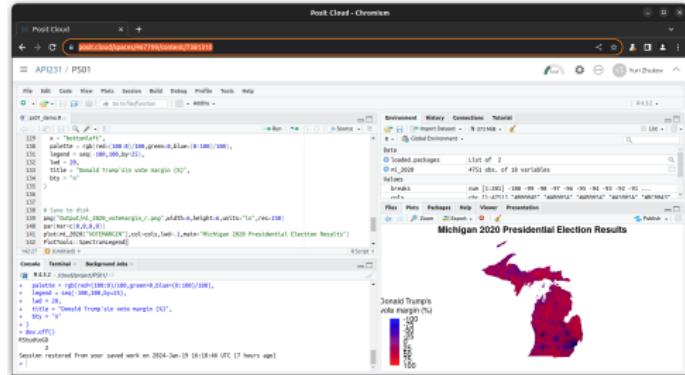


Figure 45: RStudio Cloud

Geospatial analysis in R

Task	R Packages
Data management	<code>sf</code> , <code>terra</code> , <code>rgdal</code> , <code>rgeos</code> , <code>rmapshaper</code>
Integration with other GIS	<code>rgdal</code> , <code>RArcInfo</code> , <code>SQLiteMap</code> , <code>spgrass6</code> , <code>rpostgis</code> , <code>RPyGeo</code> , <code>RQGIS</code> , <code>R2WinBUGS</code>
Access spatial data	<code>RgoogleMaps</code> , <code>rnaturalearth</code> , <code>geonames</code> , <code>OpenStreetMap</code>
Point pattern analysis	<code>spatstat</code> , <code>splancs</code> , <code>spatialkernel</code>
Geostatistics	<code>gstat</code> , <code>geoR</code> , <code>geoRglm</code> , <code>spBayes</code>
Disease mapping	<code>DCluster</code> , <code>spgwr</code> , <code>glmmBUGS</code> , <code>diseasemapping</code>
Spatial regression	<code>spdep</code> , <code>spatcounts</code> , <code>McSpatial</code> , <code>splm</code> , <code>spatialprobit</code> , <code>mgcv</code> , <code>spatialreg</code>

Full(-ish) list: cran.r-project.org/web/views/Spatial.html

→ LAB EXERCISE 1