

SEST-6577

Geographic Information Systems for Security Studies

Lecture 01 (GIS 101)

Yuri M. Zhukov
Associate Professor
Georgetown University

September 9, 2025

What are **Geographic Information Systems**?

1. tools for the collection, maintenance, storage, analysis, visualization and distribution of geospatial data
2. a.k.a. “geospatial data science”

Policy applications

1. GIS help us understand
 - a) where security, social, economic, public health problems occur
 - b) who is affected by them
 - c) how to monitor, manage and mitigate them

Scientific applications

1. GIS help us
 - a) acquire data
 - b) test hypotheses
 - c) make forecasts and predictions



Figure 1: GIS is super cool

History of geospatial data analysis / Public health



Figure 2: The blue wraith

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



Figure 3: The blue death

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



Figure 4: The experts

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

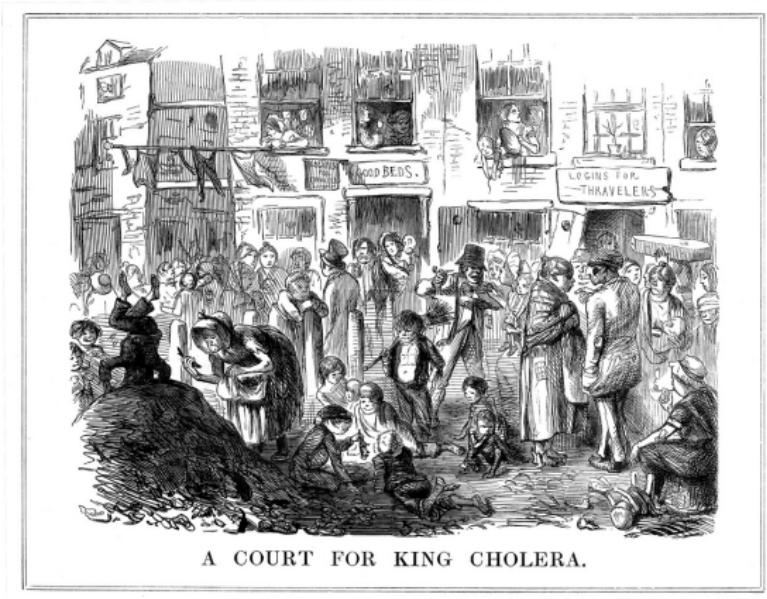


Figure 5: The scapegoats

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

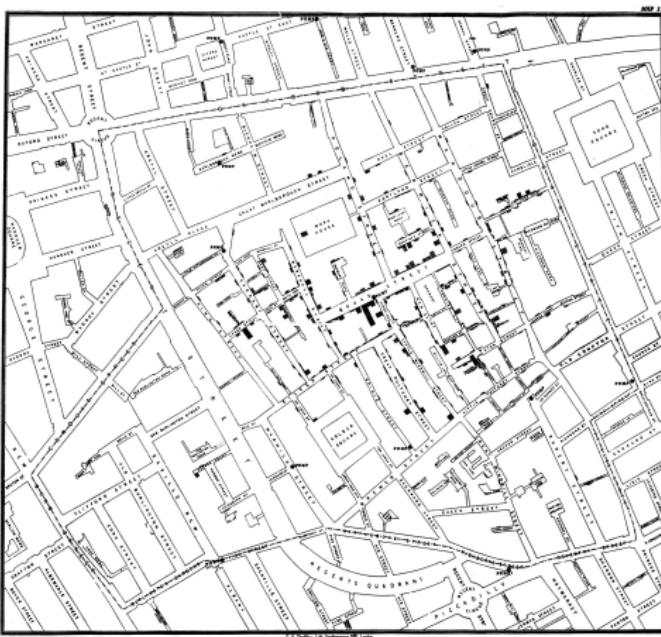


Figure 6: The geospatial data science

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 7: The pump

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



Figure 8: The public health campaign

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

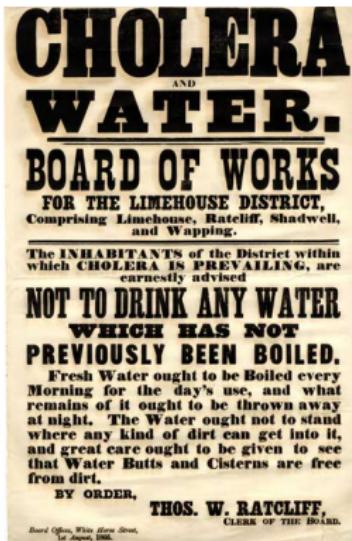


Figure 9: Data scientist has policy impact

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

History of geospatial data analysis / Military intelligence

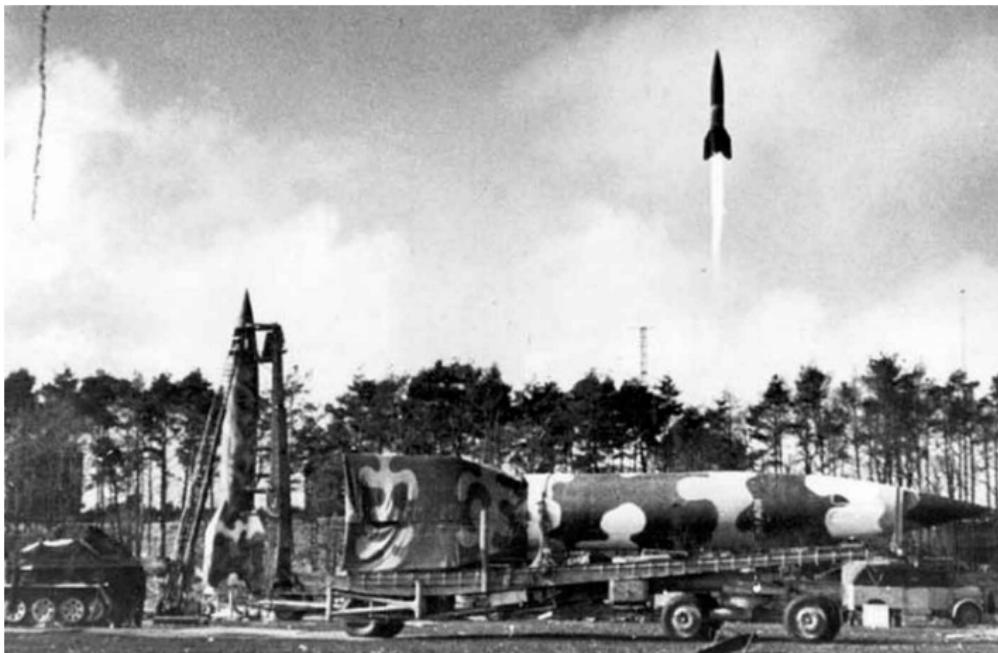


Figure 10: A new weapon

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



Figure 11: 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 12: 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 13: 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.

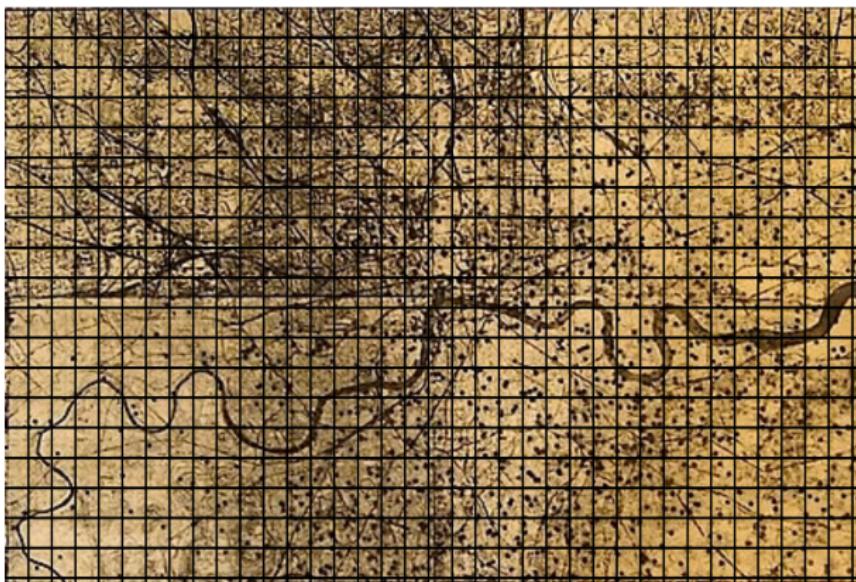


Figure 14: Wartime geostatistics

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.

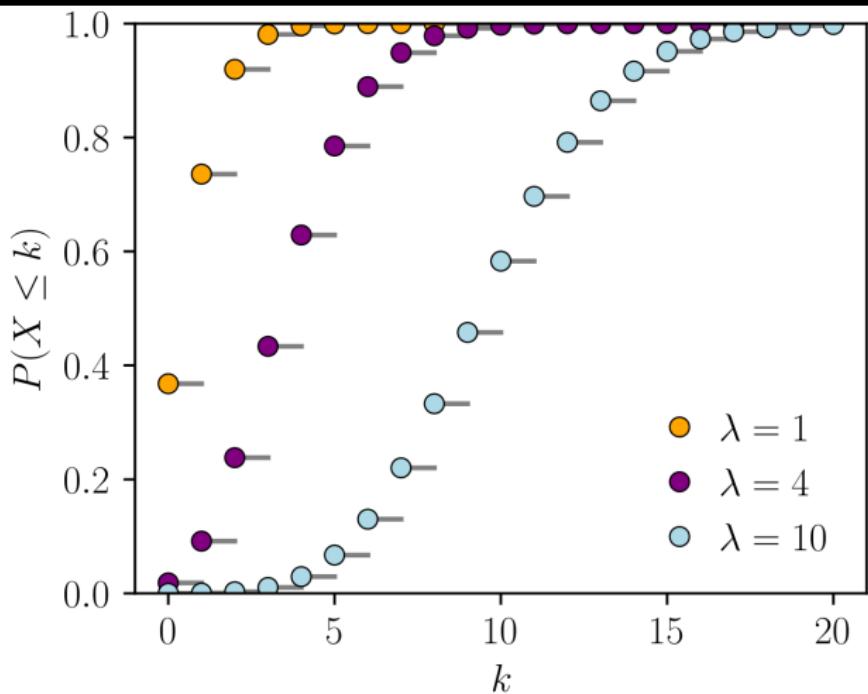


Figure 15: Poisson cumulative density function

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 16: Example: Track international shipping

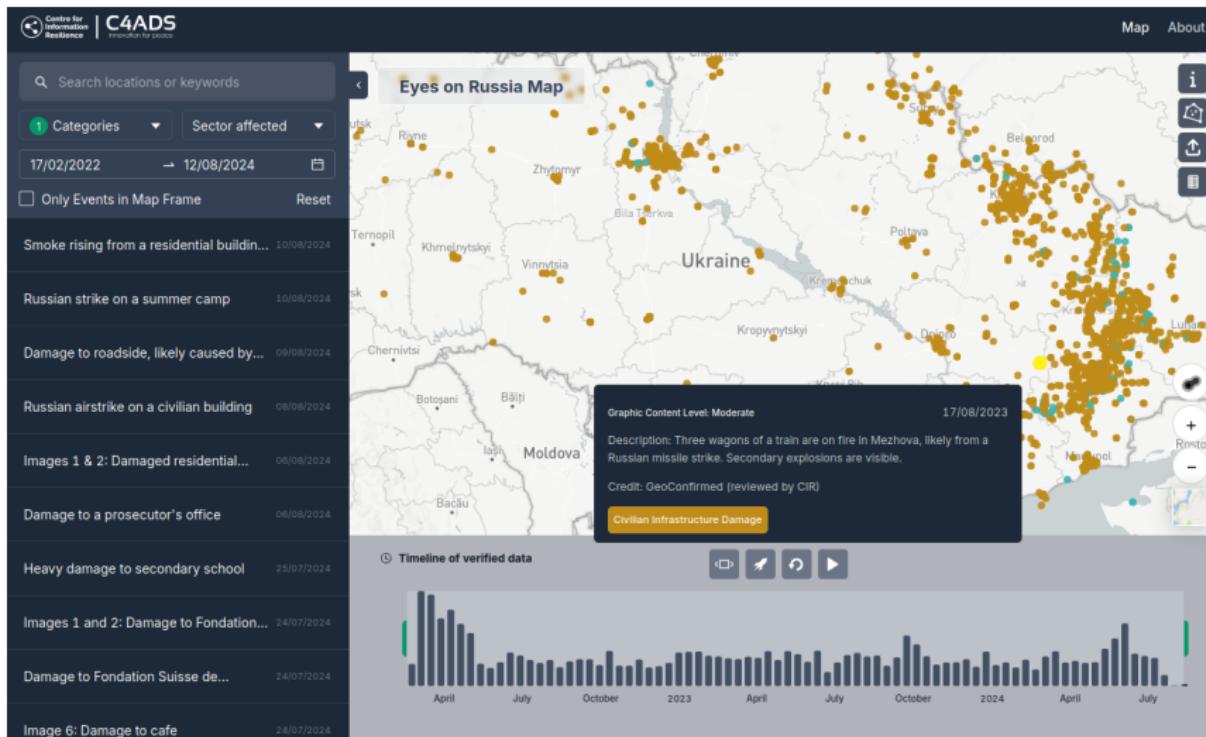


Figure 17: Example: Track civilian infrastructure damage in Ukraine

About the Class GIS Basics

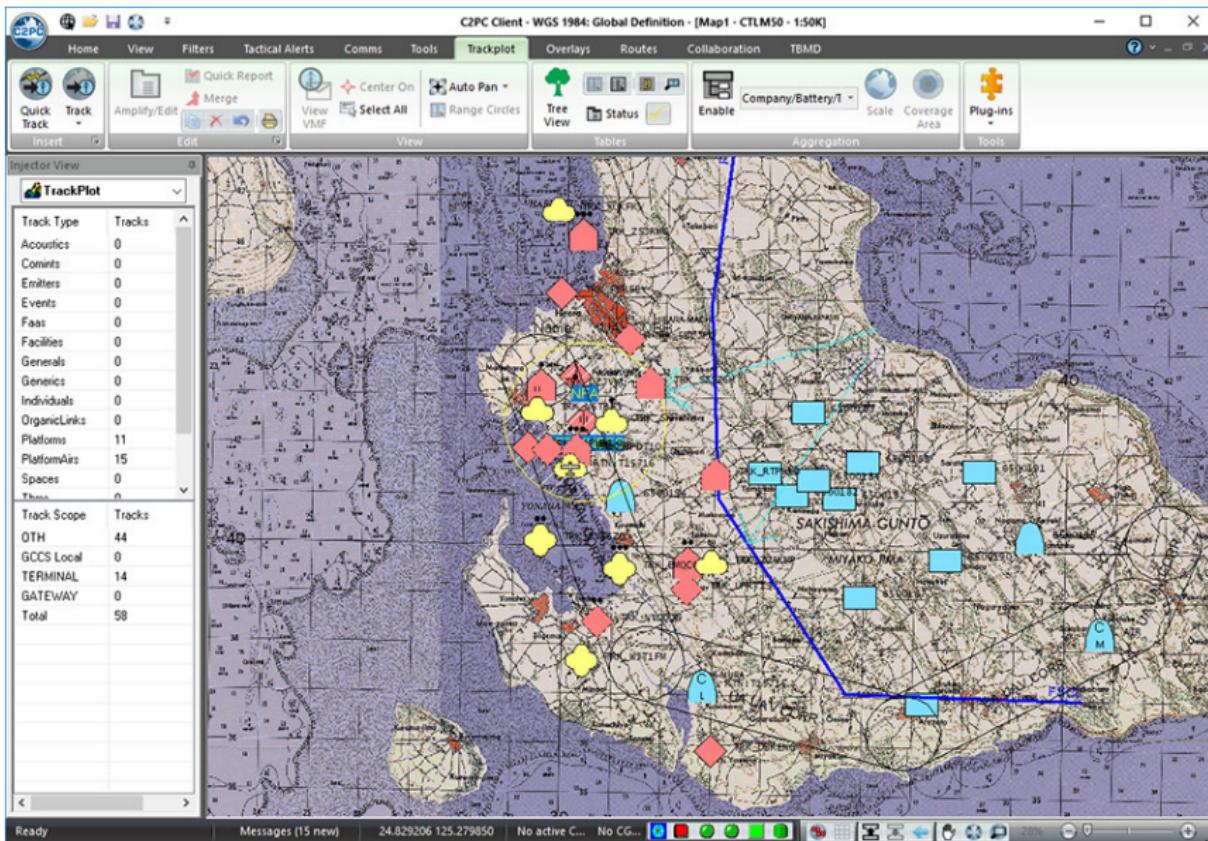


Figure 18: Example: Provide situational awareness during military operations

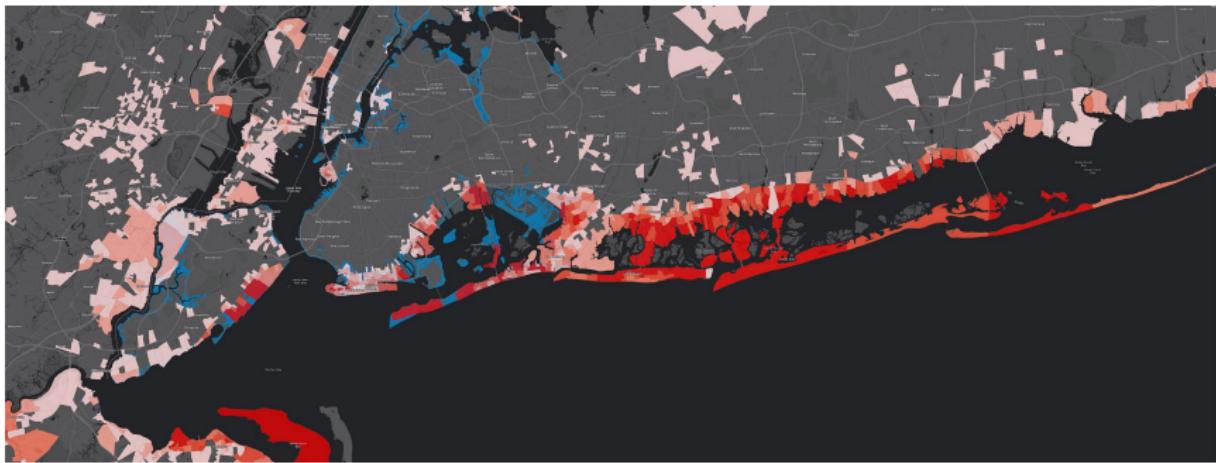


Figure 19: Example: Provide real-time information for emergency management

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.

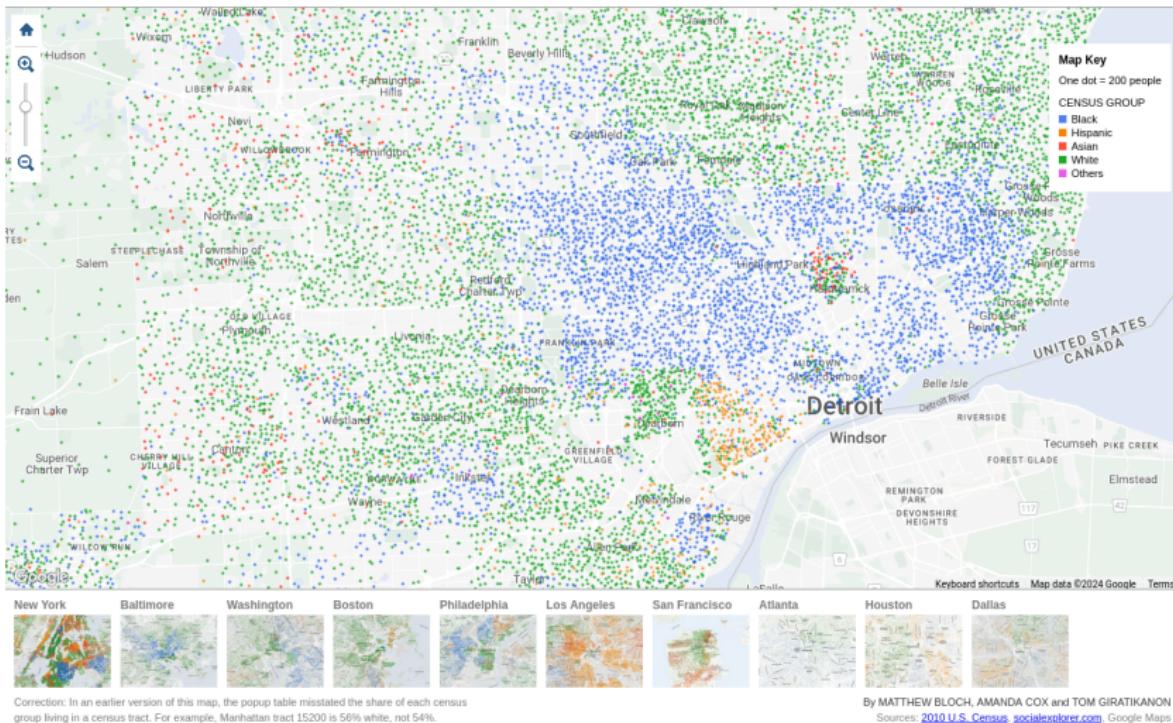


Figure 20: Example: Analyze residential segregation in American cities

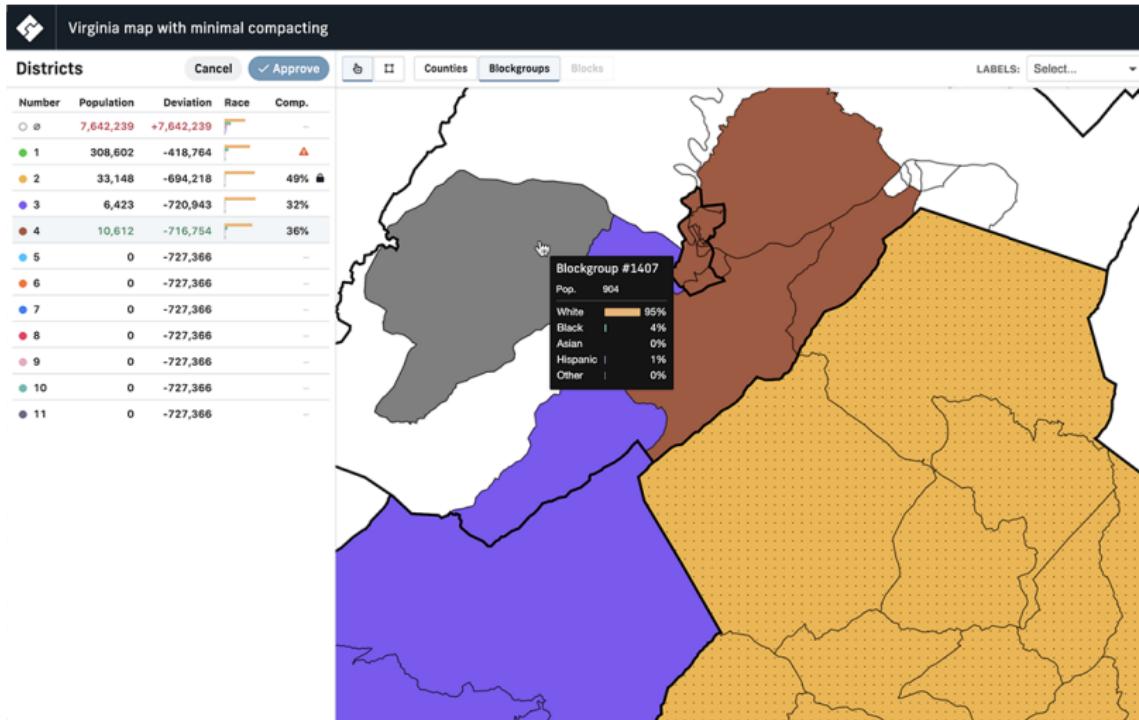


Figure 21: Example: Draw new legislative districts

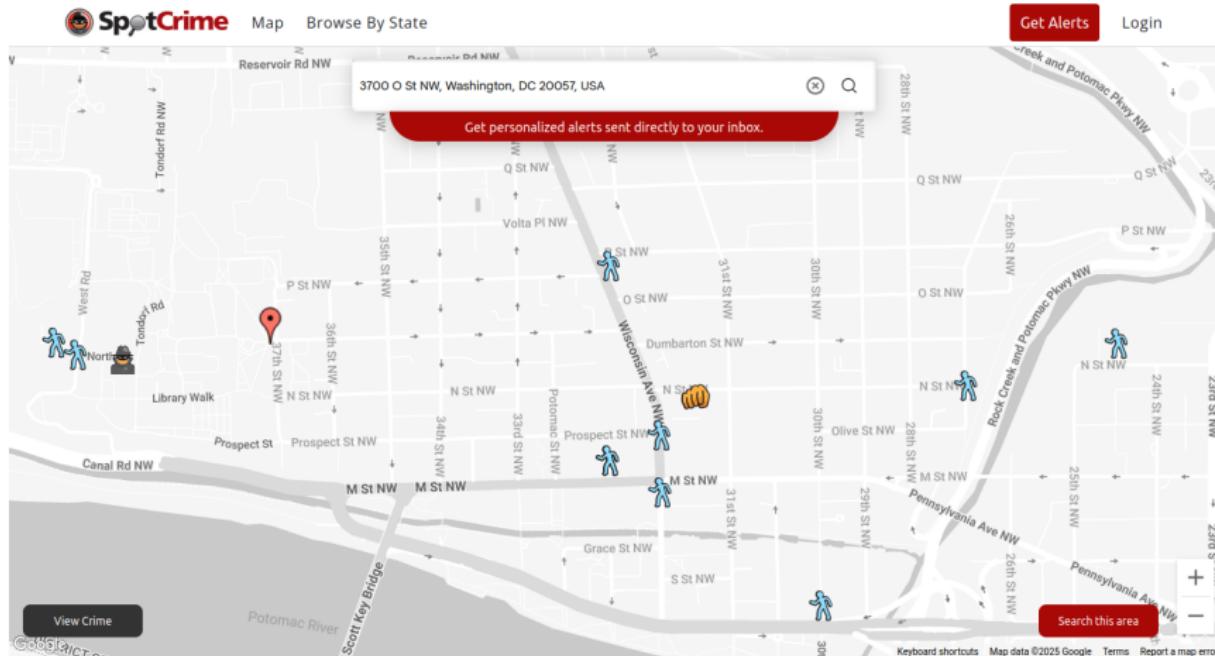


Figure 22: Example: Identify crime hotspots

About the Class GIS Basics

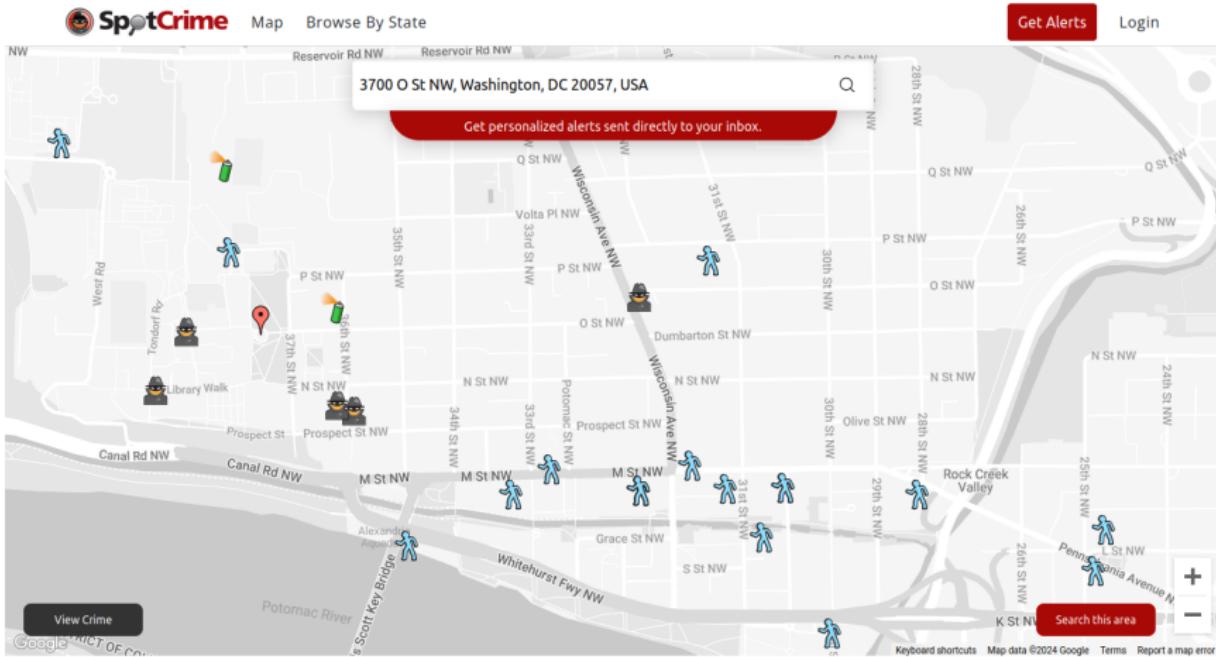


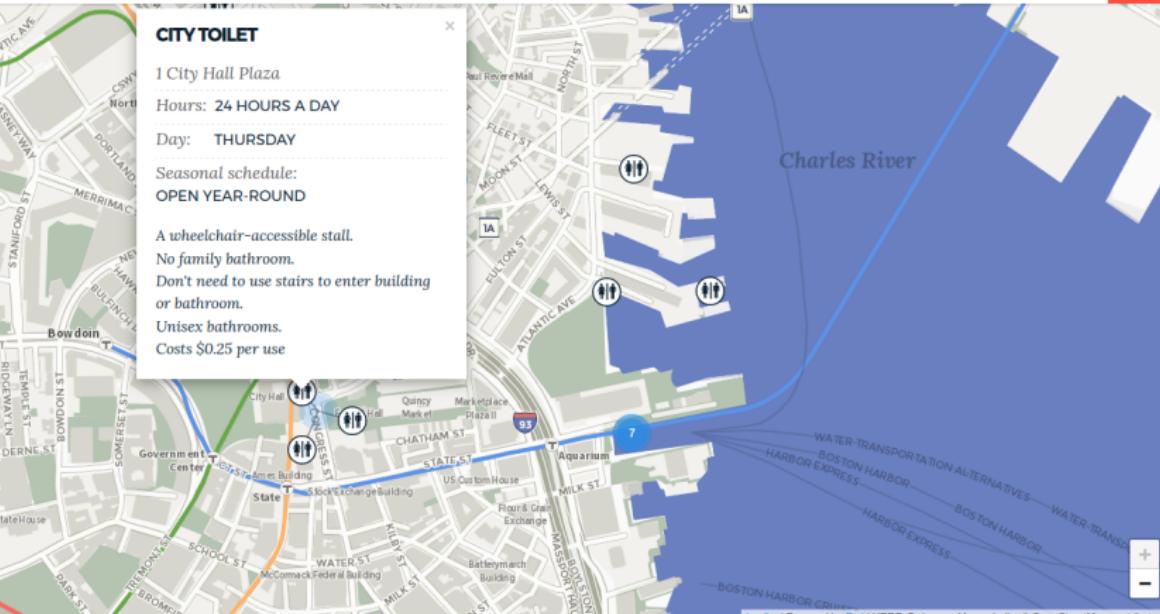
Figure 23: Example: Compare to crime from same time last year

PUBLIC RESTROOMS

CITY TOILET

1 City Hall Plaza
Hours: 24 HOURS A DAY
Day: THURSDAY
Seasonal schedule:
OPEN YEAR-ROUND

A wheelchair-accessible stall.
No family bathroom.
Don't need to use stairs to enter building or bathroom.
Unisex bathrooms.
Costs \$0.25 per use



Search for an address... X

DAY **FEATURES**

Click on a bathroom icon to find hours.

 PUBLIC RESTROOM

Figure 24: Example: Find a public restroom

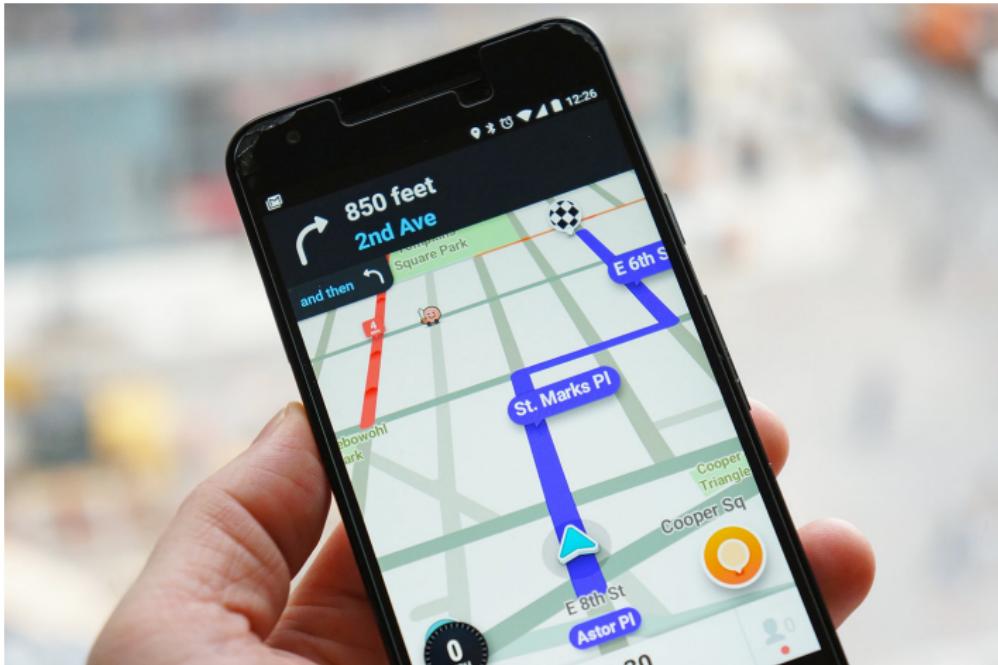


Figure 25: Example: Find your way home

About the Class

Goals and Structure

Goals of the class

1. Introduce basic GIS concepts
2. Provide hands-on experience in using open-source GIS software
3. Find, open and edit geospatial data
4. Visualize geospatial data
(make cool maps)
5. Conduct basic geospatial data analyses
6. Create new geospatial data
(georeferencing, geocoding)
7. Apply these skills to an original research project



How will we learn?

1. Methods boot camp

- a) first half of semester
- b) weekly lectures (45-75 min)
- c) weekly computational tutorials
- d) weekly problem sets

2. Research workshop

- a) second half of semester
- b) weekly “walk-throughs” of data collection & analysis on special topics
- c) no problem sets
- d) focus 100% on research project



Figure 26: Learn new methods



Figure 27: Apply them to research

Research “walk-throughs”

1. Step-by-step guides

- a) where to find and download data
- b) how to pre-process, integrate the data
- c) how to conduct a very rudimentary analysis of the data

2. Topics

- a) nighttime luminosity
- b) climate-conflict nexus
- c) Russian-Ukrainian War

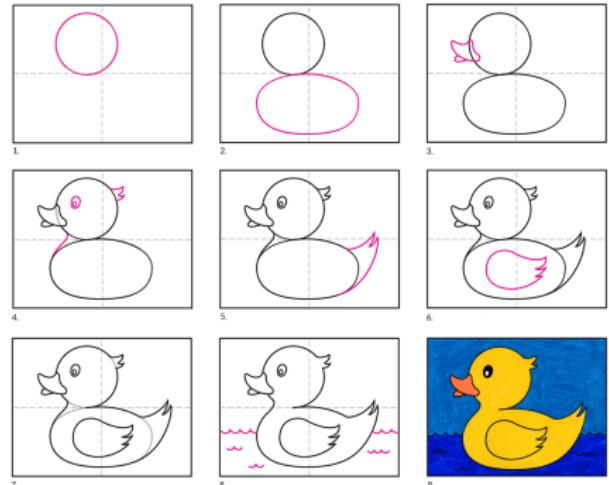


Figure 28: Like this, but for GIS

Grading

1. Problem sets (40%)
 - a) $8 \times 5\%$ each
 - b) due no later than 11:59 PM each
Monday
 - c) collaboration encouraged
2. Final project (40%)
 - a) 1-paragraph project abstract
 - due 11:59 PM, 10/24
 - b) 5-minute class presentation
 - 12/9
 - c) 5-7 page report
 - due 11:59 PM, 12/17
3. Attendance & participation (20%)
 - a) show up, ask questions, help others
 - b) utilize “troubleshooting” discussion boards on Canvas



Figure 29: Don't worry

Final Project

1. Overview

- a) goal: use GIS to answer a *political/social/economic question*
- b) *descriptive question*: answer through mapping & visualization
(e.g. "Which neighborhoods are the most violent?")
- c) *explanatory question*: answer through analysis of geospatial data
(e.g. "Why are some neighborhoods more violent than others?")
- d) collaboration/co-authorship permitted

2. Project abstract (1 paragraph)

- a) summarize research idea, needed spatial & non-spatial data

3. In-class presentation (5 min, 2 slides)

- a) slide 1: Research question
- b) slide 2: Map(s)

4. Written report (5-7 pages)

- a) section 1: Research question
- b) section 2: Data & methods
- c) section 3: Preliminary results

Who am I?

Yuri M. Zhukov

Associate Professor

School of Foreign Service

Department of Government

Georgetown University

(SSP Alum '07)

Who are you?



Figure 30: Professor's self-image

More info

1. Check out Canvas page: georgetown.instructure.com/courses/219200
2. Email me any course-related questions: ymz2@georgetown.edu
3. Sign up for my office hours: calendly.com/ymz2-georgetown



Figure 31: Stay cool. Learn GIS

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

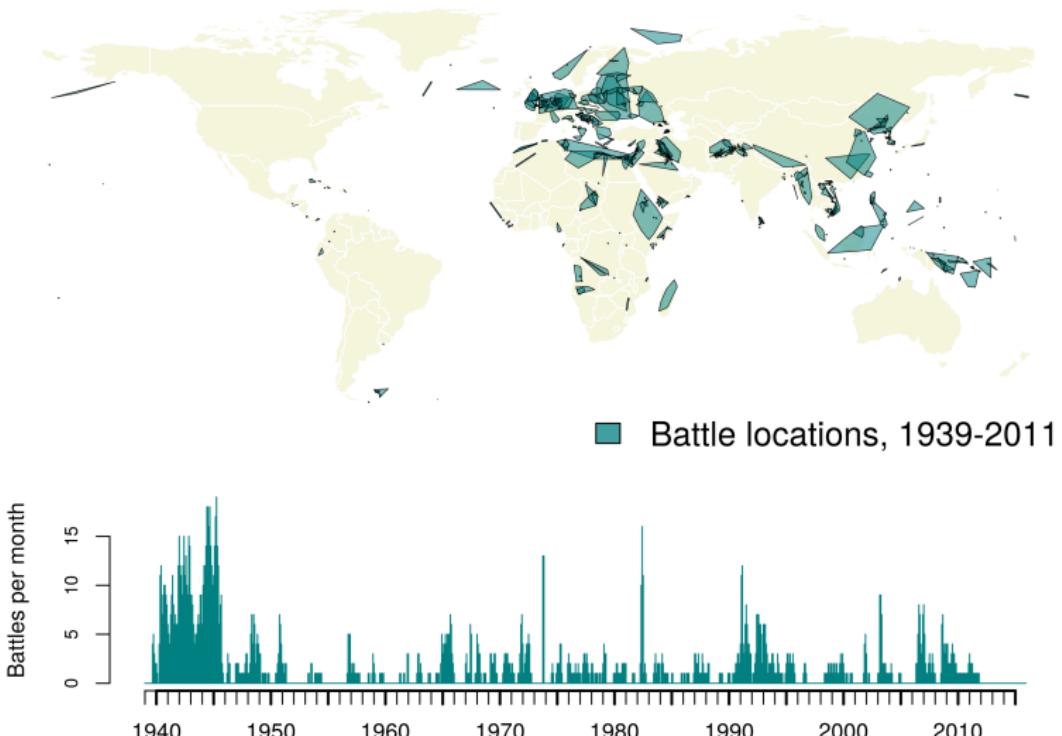




Figure 32: Let's add another dimension

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

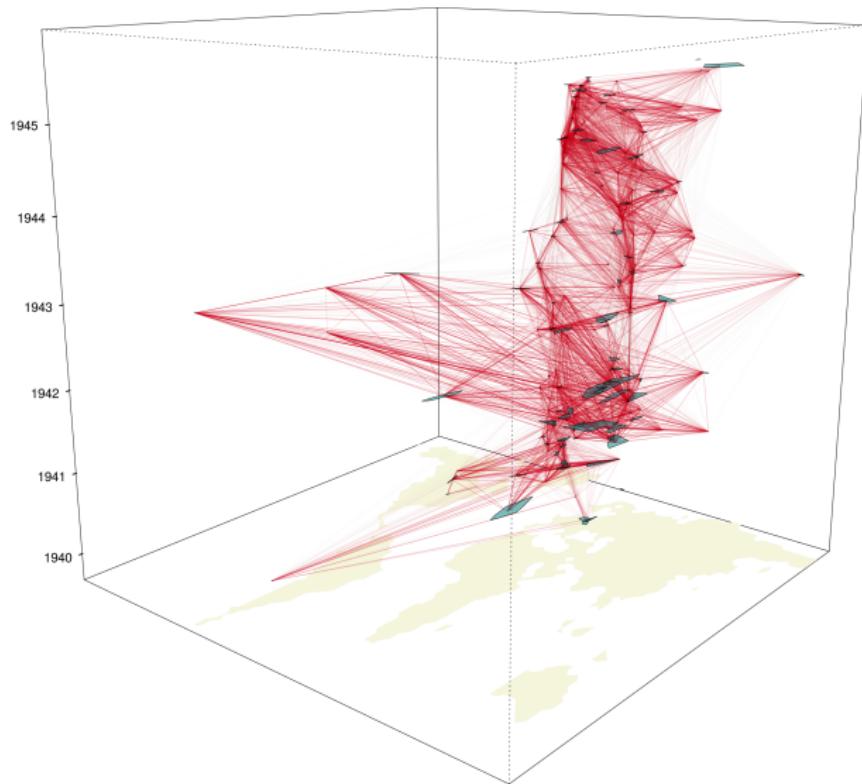


Figure 33: WWII battles in multidimensional space

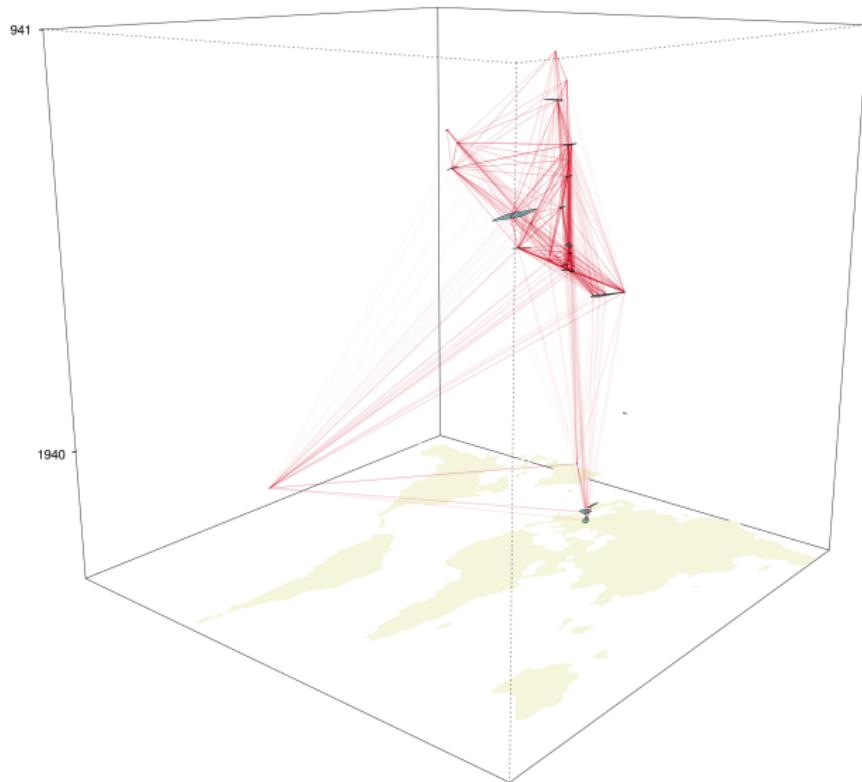


Figure 34: Battles (1939-1941), linked by combatant

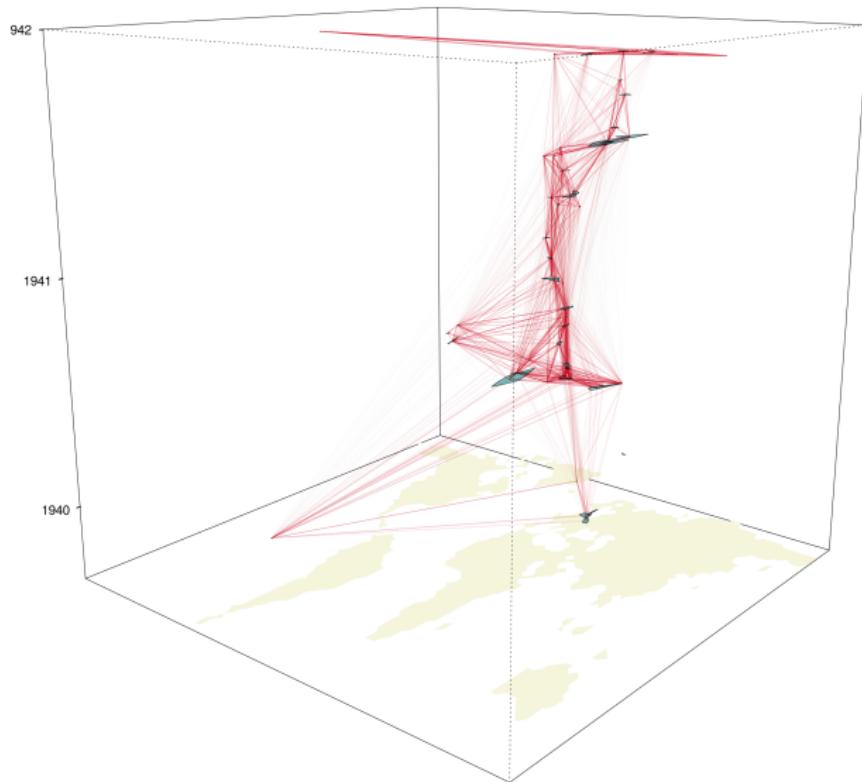


Figure 35: Battles (1939-1942), linked by combatant

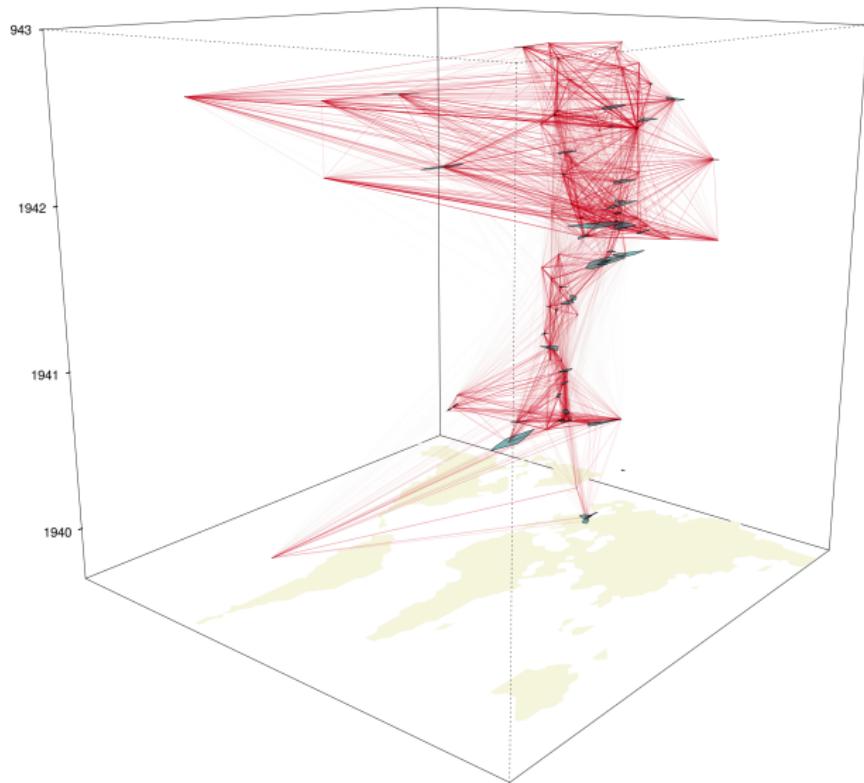


Figure 36: Battles (1939-1943), linked by combatant

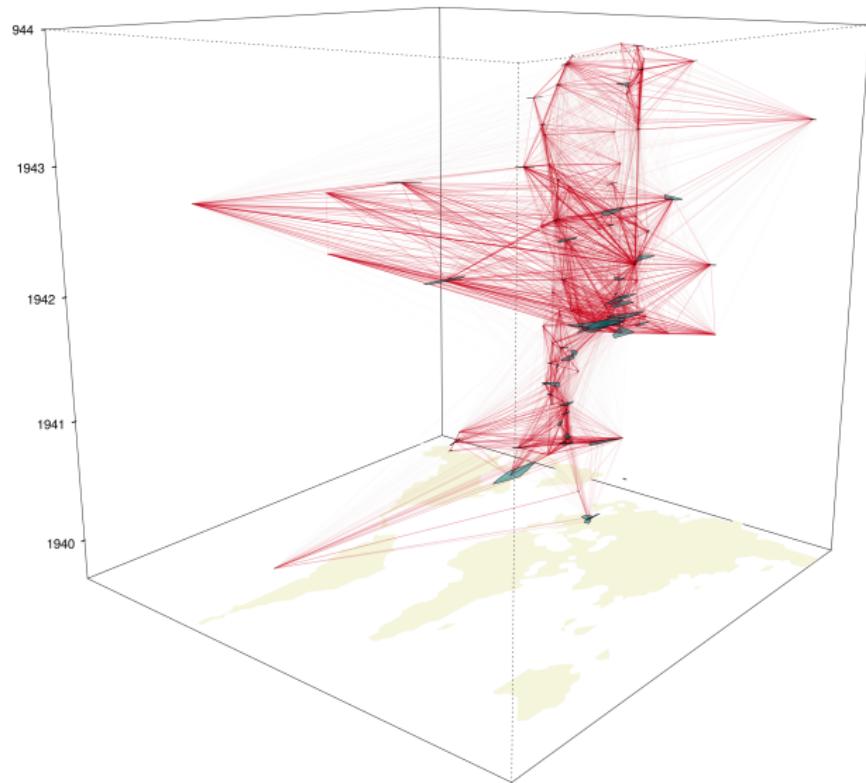


Figure 37: Battles (1939-1944), linked by combatant

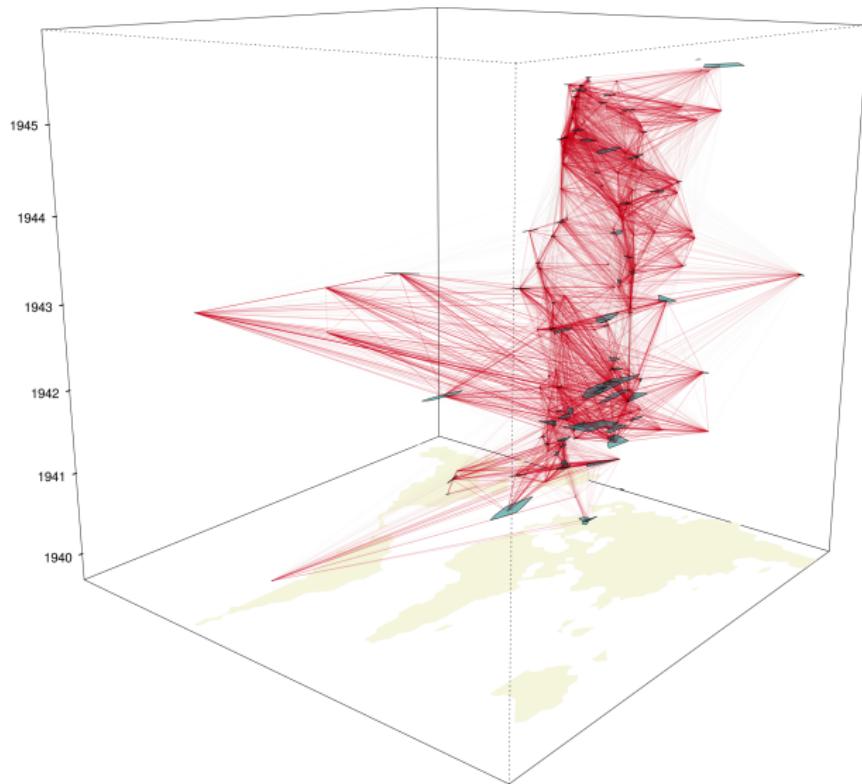


Figure 38: Battles (1939-1945), linked by combatant

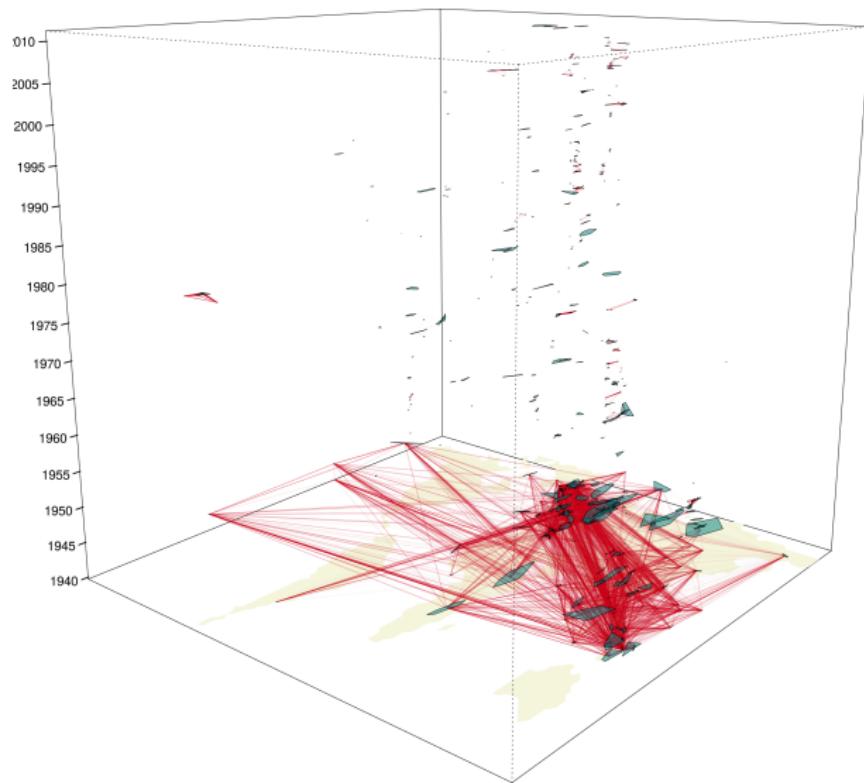


Figure 39: Battles (1939-2011), linked by combatant

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)

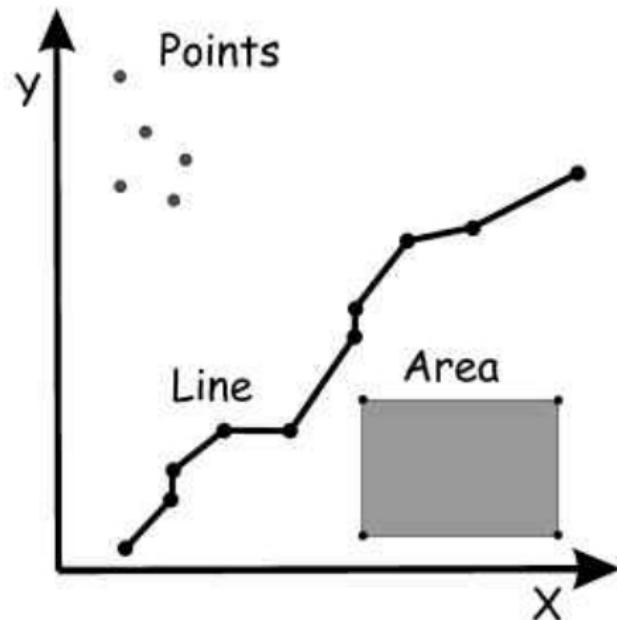


Figure 40: Vector data objects



Figure 41: Points



Figure 42: Polyline

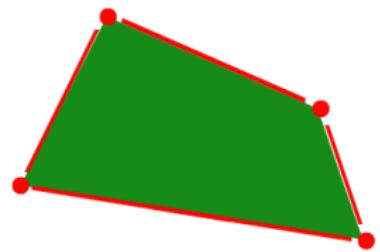


Figure 43: Polygon

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)

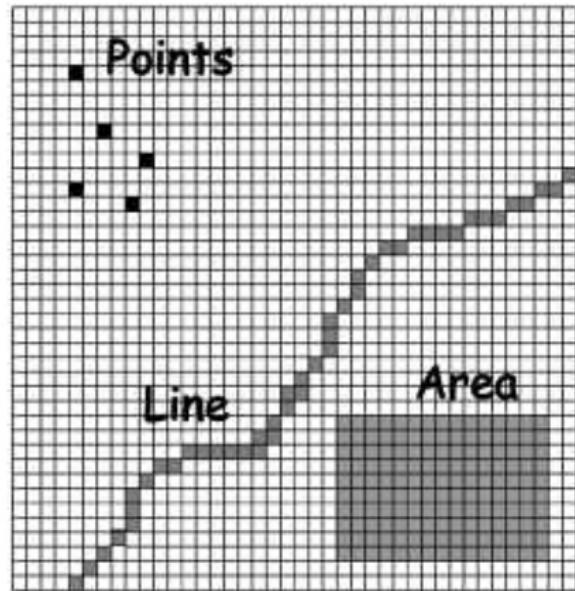


Figure 44: Raster data

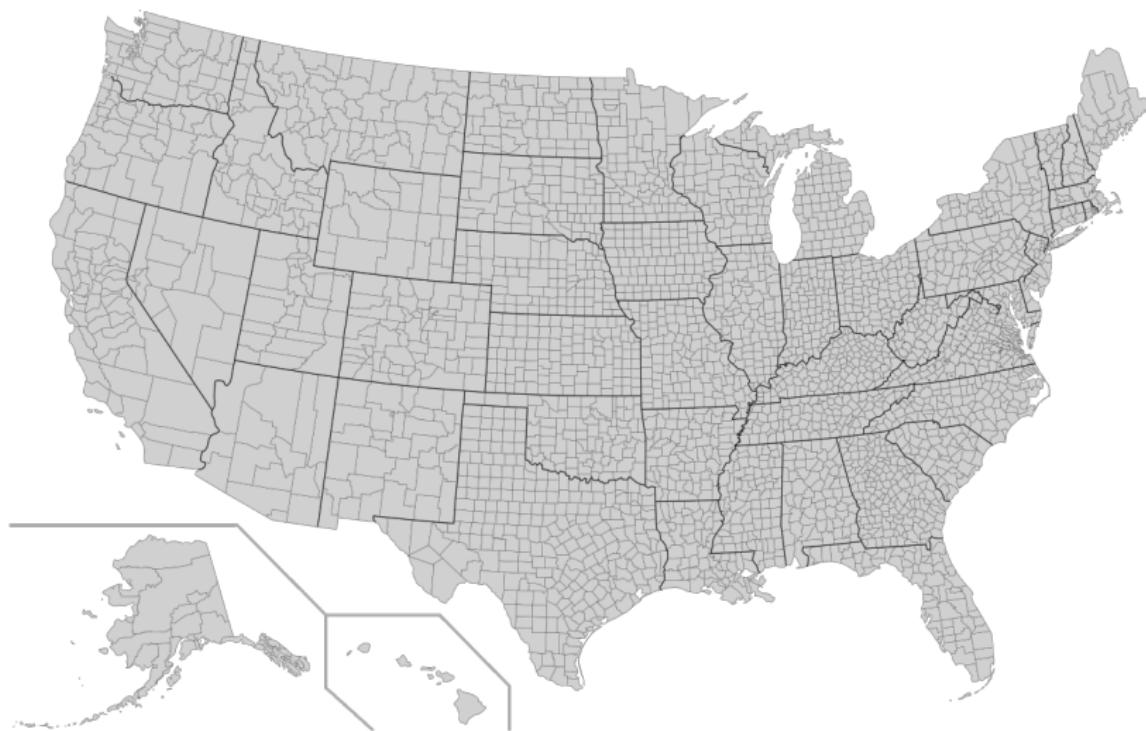


Figure 45: Vector or raster?



Figure 46: Vector or raster?

Red Army soldiers in WWII, by birth location

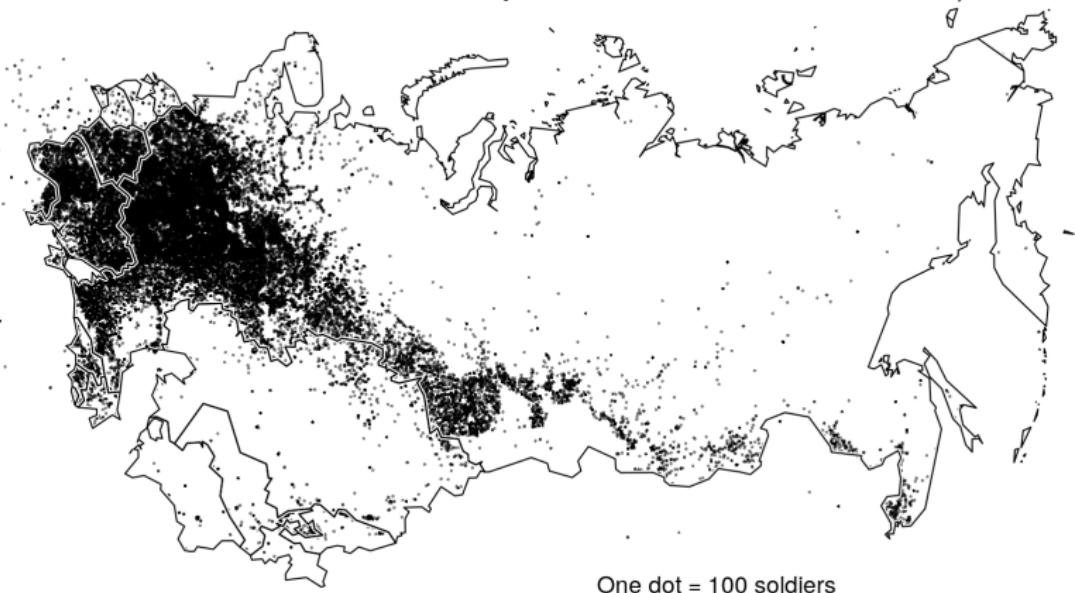


Figure 47: Vector or raster?

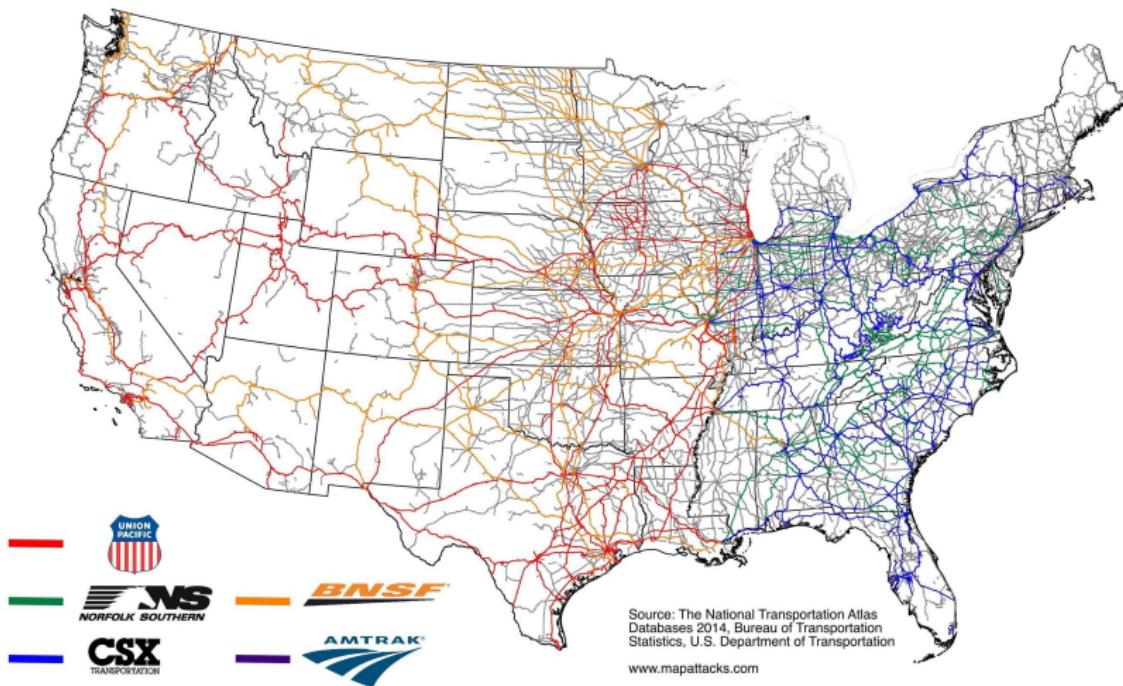


Figure 48: Vector or raster?

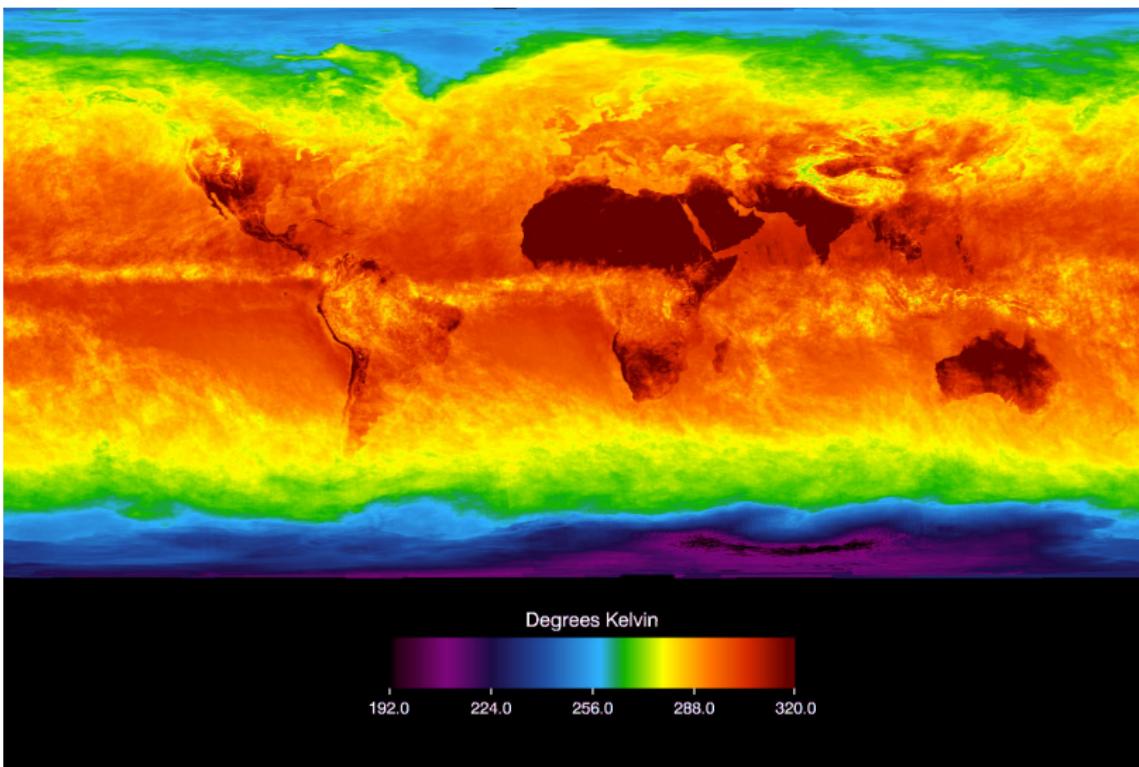


Figure 49: Vector or raster?

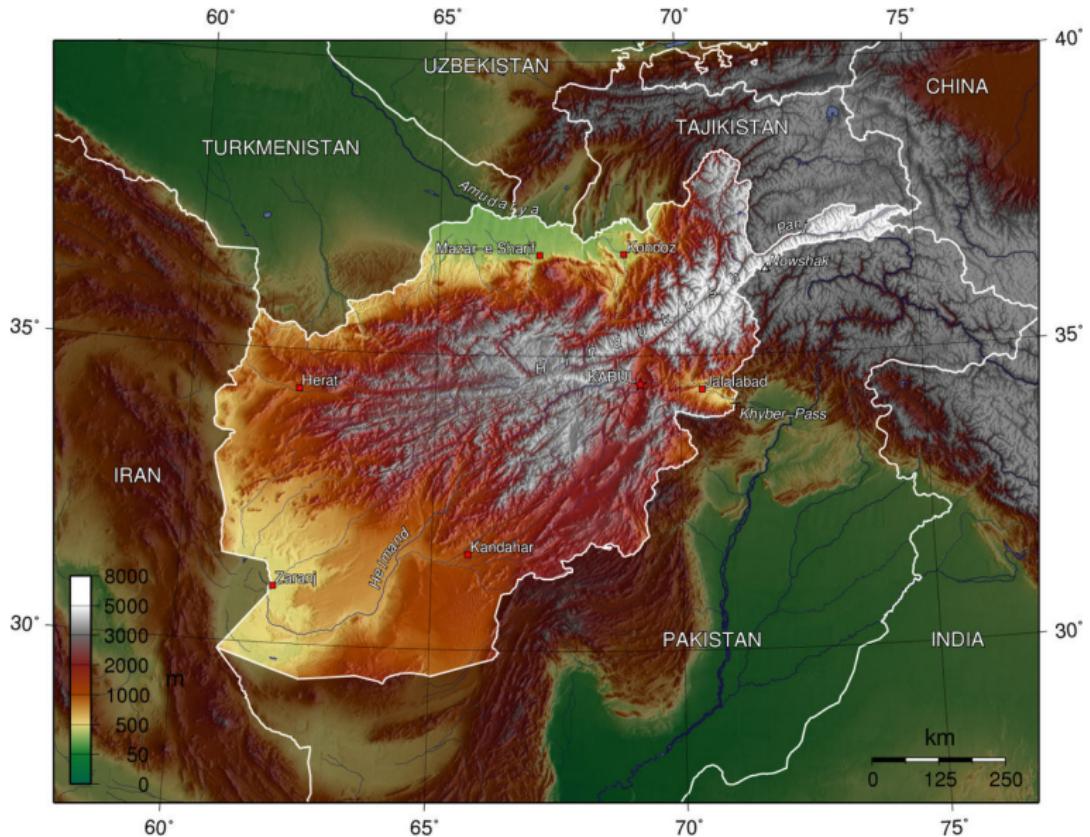


Figure 50: Vector or raster?

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
R	Free	High	Geoprocessing, visualization, geostatistics, spatial econometrics, point processes
Python	Free	High	Geoprocessing, visualization, geostatistics, spatial econometrics, point processes

You will be using **QGIS**, **R** and/or **Python** (choose your own adventure)

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

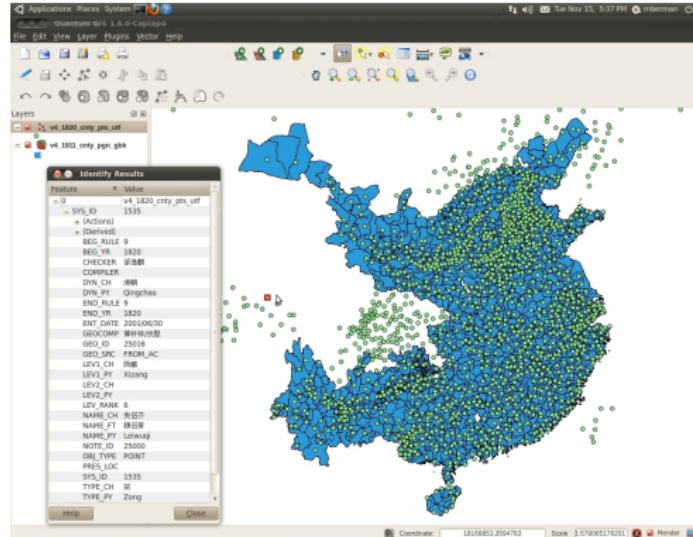
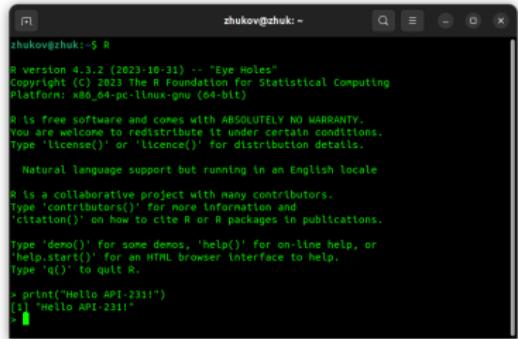


Figure 51: QGIS

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



A screenshot of a terminal window titled 'zhukov@zhuk: ~'. It displays the R command-line interface. The output shows the R version (4.3.2), copyright information, and a welcome message. It then shows a demonstration of the 'print()' function, which outputs the string 'Hello API-23!!'.

```
zhukov@zhuk: ~
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-23!!")
[1] "Hello API-23!!"
>
```

Figure 52: R

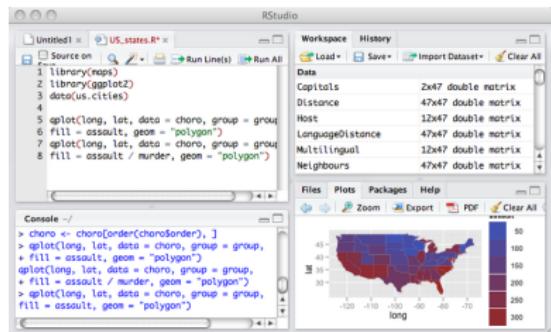


Figure 53: RStudio

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

Software & programming

3. Python (option 3)

- a) open-source programming language
- b) powerful ecosystem for geospatial analysis (GeoPandas, Rasterio)
- c) excellent for automating workflows and large datasets
- d) can run Python from command line
 - ... or using code editors
(e.g. VS Code, Sublime Text)
 - ... or using Jupyter notebooks
(interactive analysis)
- e) multiplatform
(runs on Linux, Mac, Windows)
- f) download Python here:
python.org
 - ... or Anaconda here:
anaconda.com

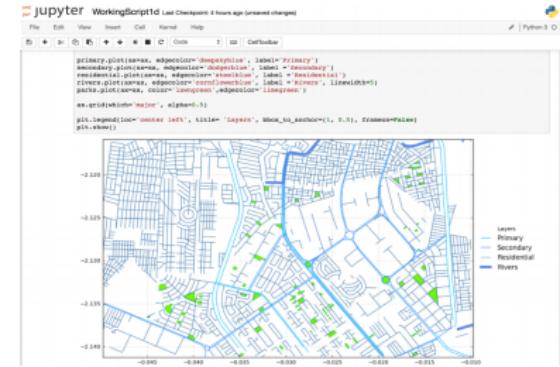


Figure 54: Jupyter notebook

Geospatial analysis in Python

Task	Python Libraries
Data management	geopandas, shapely, fiona, pyproj, mapclassify
Integration with other GIS	arcpy, ogr, rasterio, geopandas, geoalchemy2
Access spatial data	osmnx, geonamescache, contextily, rasterio, pyproj
Point pattern analysis	pointpats, scikit-mobility, PySAL
Geostatistics	PyKrig, gstatools, PySAL
Disease mapping	PySAL, GeoDa, statsmodels
Spatial regression	PySAL, spreg, statsmodels, mgcvpy

→ LAB EXERCISE 1