Deck 12 – Model Thinking and Problem Solving

Geostatistics and Spatial Statistics, Interpolation, Scale, Final Projects,

---

Intro to GIS – UMass Amherst – Michael F. Nelson

Overview

Geostatistics/Spatial Stats Wrap-Up

- Interpolation
- Geostatistics vs. Spatial Statistics
- Spatially-Aware Regression

Model Thinking + GIS Problem-Solving

- GIS Problem-Solving Techniques
- Computational Thinking

Final Posters

- ABT Statements
- Element Layouts

Model Thinking

It’s all about simplifying.

GIS Problem-Solving: Simplify!

It’s always easier to break down a complex problem into smaller steps.

- Don’t forget the overall goal

What are the risk criteria for lab 6?

- **High-risk criterion #1:** Within 300m of road
- **High-risk criterion #2:** Slope less than 10 degrees
- **High-risk criterion #3:** Landcover type 1
- **Mid-risk criterion:** Landcover type 2

How did you combine the landcover risk criteria?

You can solve almost any GIS problem with Model Thinking:

- Decompose the problem into the most important components
- Recognize patterns
- Reduce Complexity using Abstraction

Model Thinking with Lab 6 Problem

Decompose

- Illegal logging is complicated, but we decomposed the problem into parts:
  - Ease of Access: Infrastructure and terrain
  - Location of trees: Forest type

Recognize patterns

- Visual map reality check: should there be any risk zones far from the roads?
- How do the slope and elevation maps relate to each-other?

Reduce Complexity using Abstraction

- Composite risk can be modeled as a combination of simple criteria.

Landcover Risk Criteria
Two Approaches – Which one is better?

Reclassify: Create a single layer

- Use the Reclassify tool to create a layer with cell values:
  - 0 = no risk
  - 1 = high risk
  - 2 = moderate risk

Calculator: Create two layers

- Use Raster Calculator to make two binary layers:
- High-risk Landcover Layer
  - 0 = no risk
  - 1 = high risk
- Moderate-risk Landcover Layer
  - 0 = no risk
  - 1 = moderate risk

Landcover Risk Criteria
Two Approaches – Which one is better?

Reclassify: Create a single layer

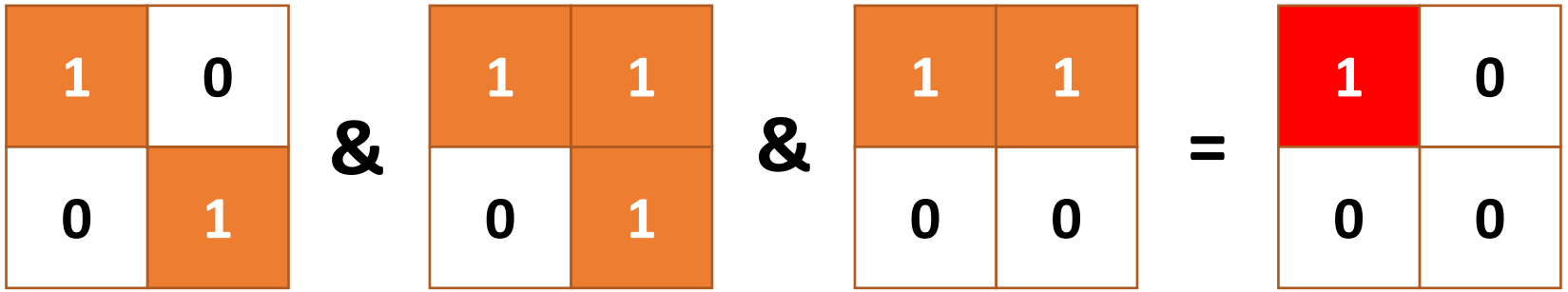
- Pro: You have only one landcover layer
- Con: More complicated to combine with road and slope layers. You can't use AND, you need a more complicated logical test.

Calculator: Create two layers

- Con: you have 2 new raster layers instead of 1.
- Pro: Since each layer is binary, you can easily use Boolean AND to combine with road/slope layers. You could even use raster multiply!

Composite Risk: Binary Layers - Boolean AND

Binary raster layers are awesome!



Risk Criterion 1 | Risk Criterion 2 | Risk Criterion 3 | Composite Risk Score

GIS Problem-Solving: Keep Your Notes

It seems obvious – but it can save you lots of time when you need to remember what you did one year, one month, or even 1 day ago.

For example: Do you remember the text of your SQL query to select towns with population higher than 10,000 and area less than 50 square km in lab 3?

I had to re-do the whole procedure. If only I had saved the query!

GIS Problem-Solving: Reality Checks

Does it make sense?

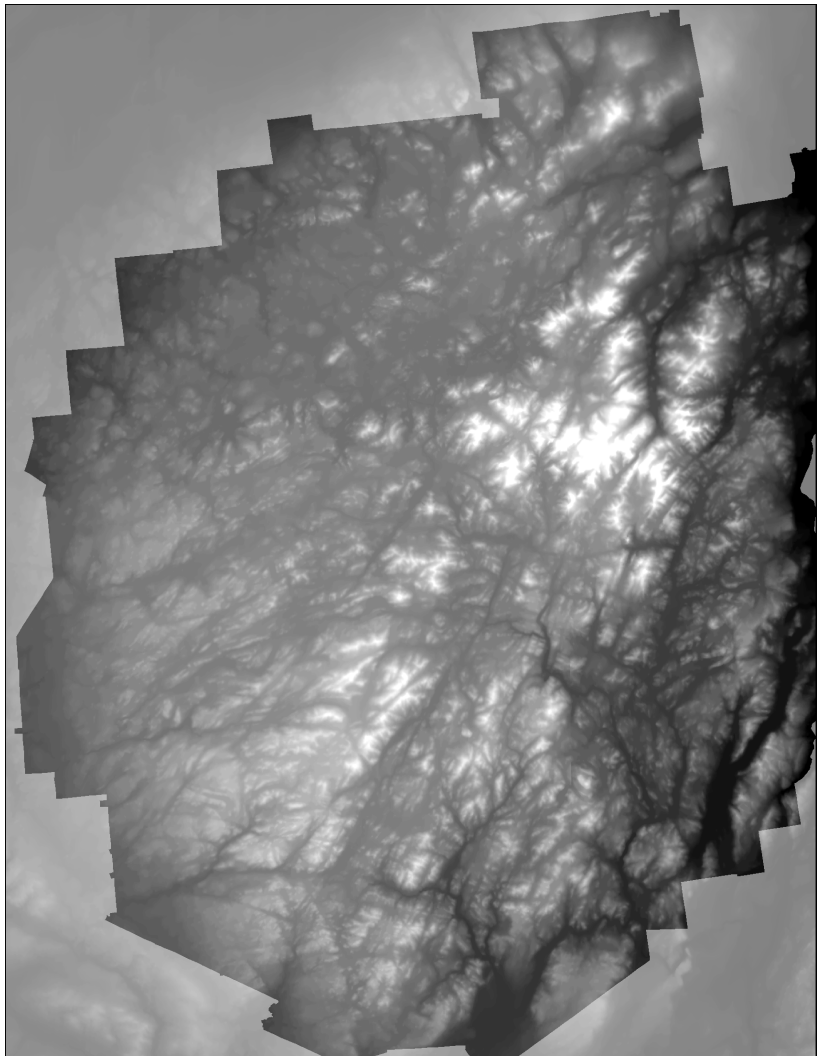
- What kind of data can I make a slope from?
- Can, or should, I do a raster multiplication on a categorical raster?

What should it look like?

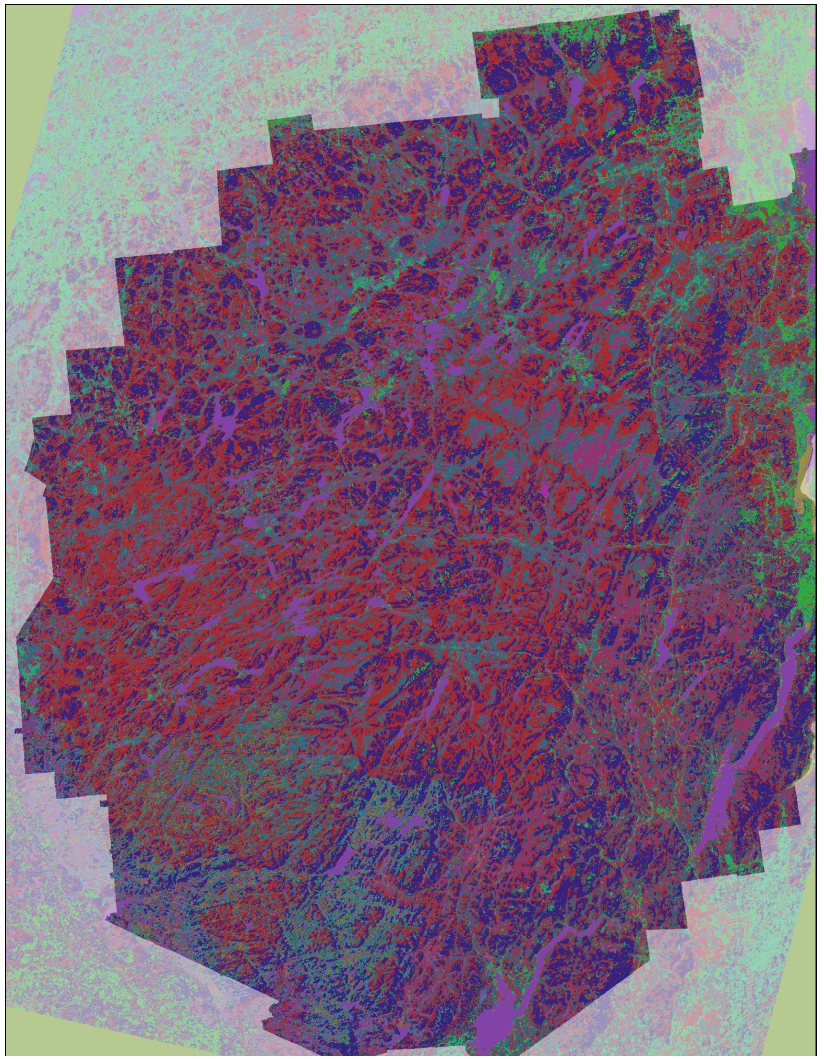
- What should a distance-to-feature layer look like?
- Should the result be a continuous surface?
- Are my selected features where I expect?

GIS Problem-Solving: Reality Check - Slope

Elevation

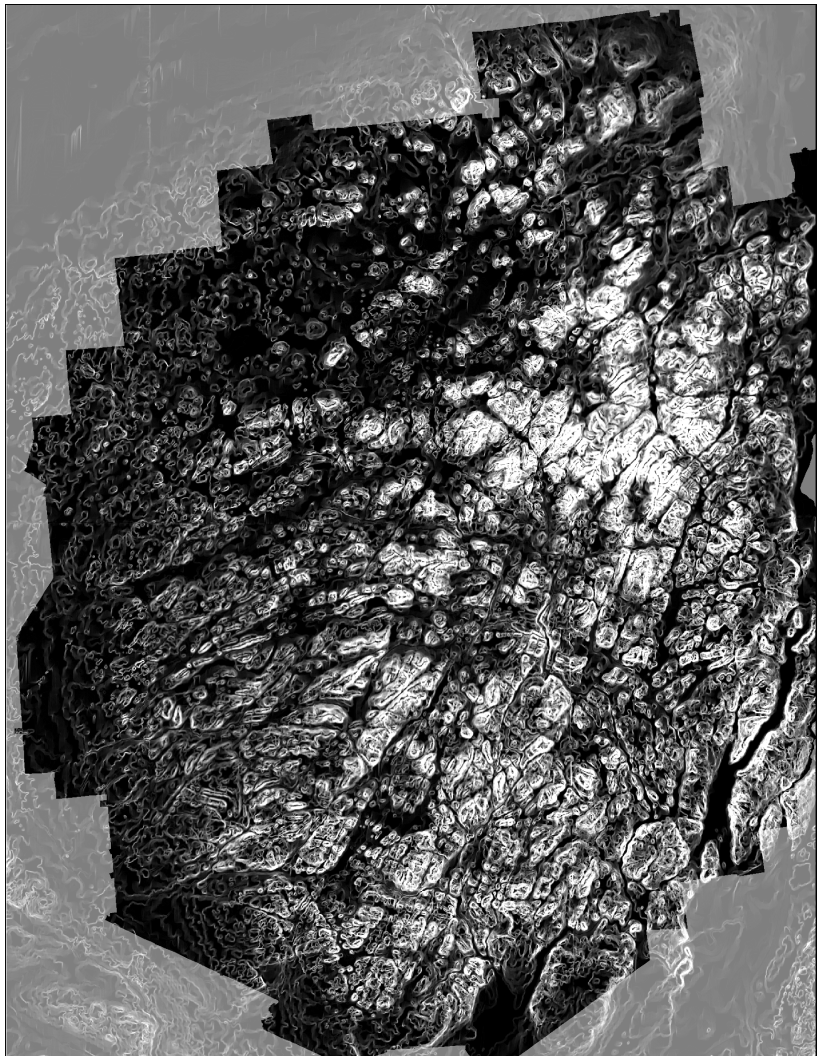


Landcover Codes

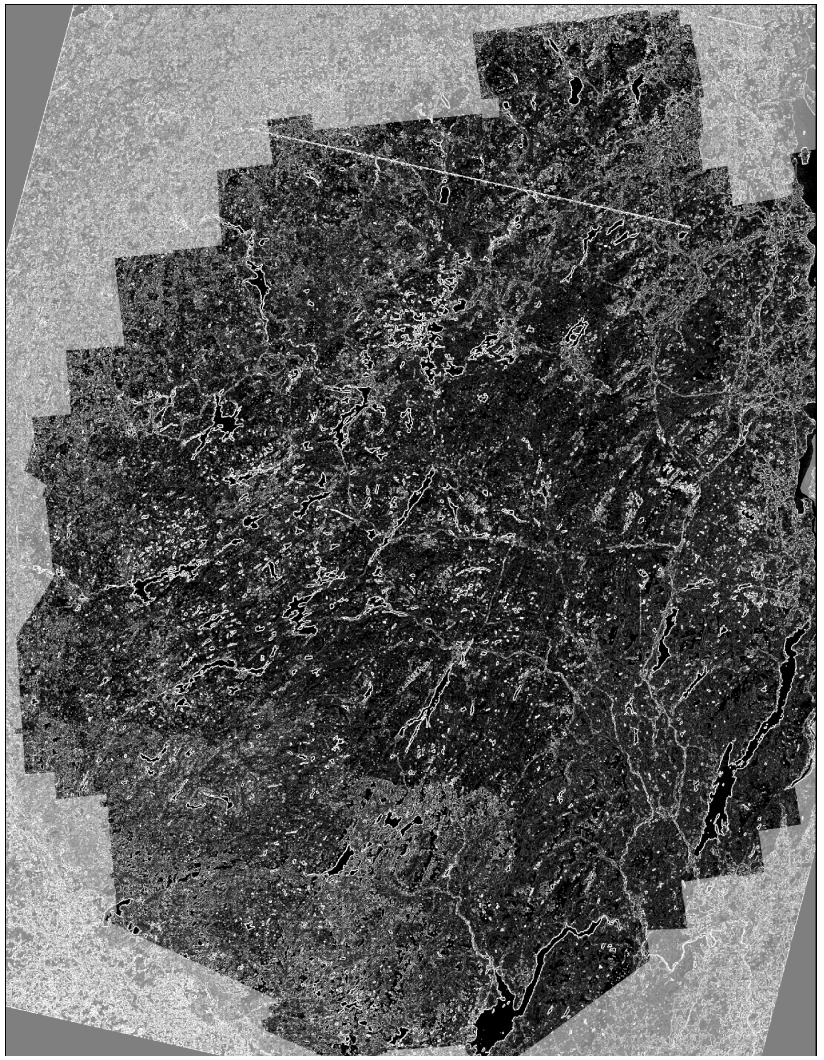


GIS Problem-Solving: Reality Check - Slope

Slope?



Slope?



Computational Thinking

It’s Model Thinking rebranded!

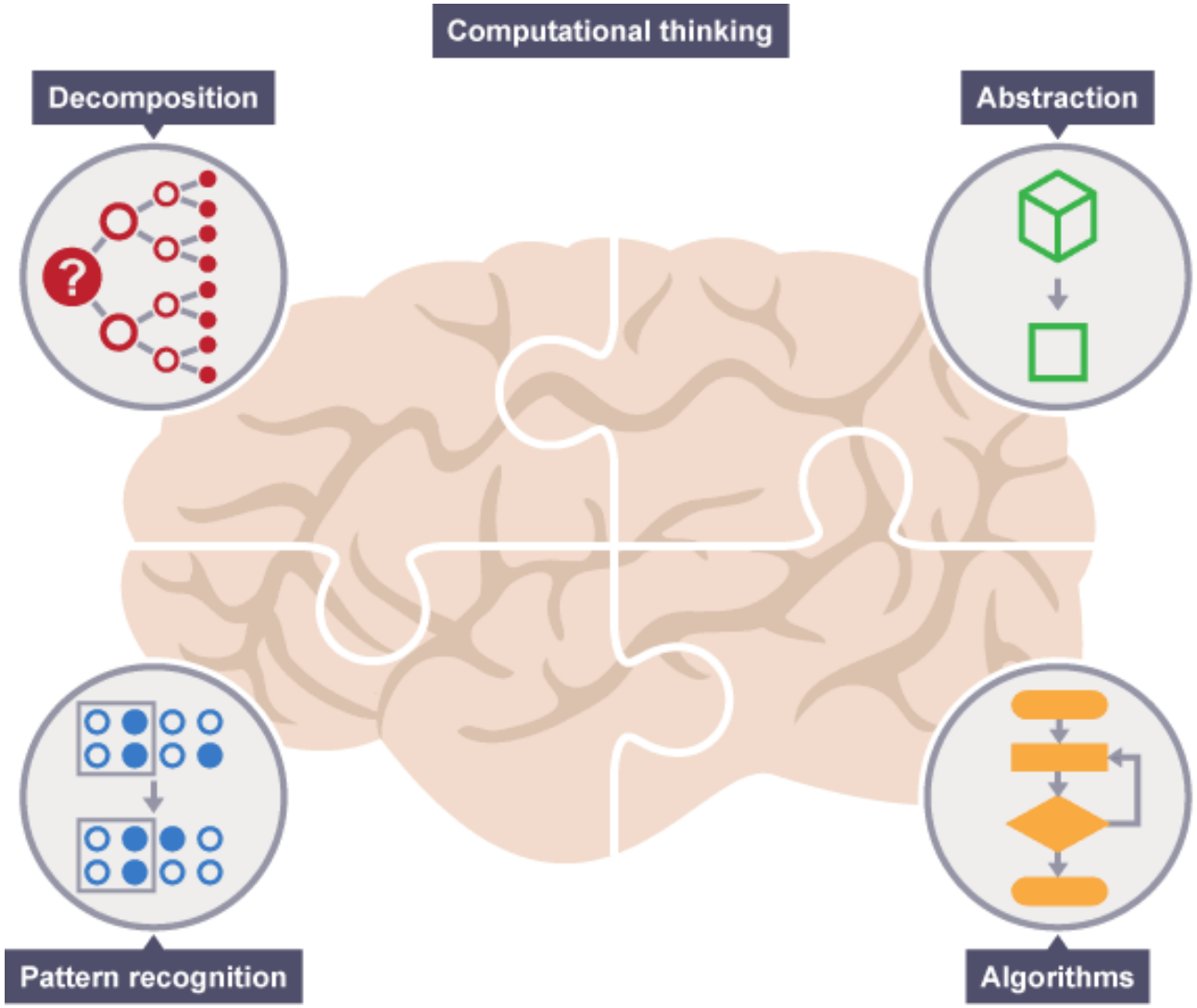
Thinking Computationally

"**Computational thinking** is a way of solving problems, designing systems, and understanding human behavior that draws on concepts fundamental to computer science. To flourish in today's world, **computational thinking** has to be a fundamental part of the way people think and understand the world."

–Carnegie Mellon University: Center for Computational Thinking

Computational Thinking and Google Earth

Recall the Foundations



http://www.bbc.co.uk/education/topics/z7tp34j

Sound Familiar?

- Gathering and analyzing data
  - Mapping, exploring symbology, plots: histograms etc.
- Using models and tools in various ways
  - Calculate slopes, distances
- Creating solutions to represent or solve real world problems.
  - Habitat modeling
  - Illegal timbering risk

Decomposition

- Involves breaking down a complex problem or system into smaller parts that are more manageable and easier to understand.
- A natural problem-solving activity.
- Model thinking perspective: ask yourself what are components of your system of problem.
  - What parts are likely to be important?
  - What parts might you be able to ignore or simplify?
  - What are the unknowns (known or unknown unknowns)?
  - What are sources of randomness

Decomposition Example: Risk Components

- You need to compute 2 different composite risk levels.
- You have 4 risk criteria. How can you break down the problem?

1. Build individual, binary, risk layers
    - Use Raster Calculator with Boolean operations to build layers where:
        - 1 = meets risk criterion, 0 = does not meet criterion
2. Combine individual criteria into composite risk layers

Decomposition Example: Suitability Analysis

- How can we represent different components of suitability?
- Boolean
  - Forest or non-forest landcover
- Categories:
  - Ordinal
    - Dry soils, medium soils, wetlands.
  - Nominal
    - Coniferous forest, prairie, deciduous forest
- Gradients?
  - Elevation

Pattern Recognition

- Involves finding the similarities or patterns among small, decomposed problems that can help us solve more complex problems more efficiently.
- Humans are (too) good at spotting certain kinds of patterns.
  - How often do you recognize faces where they aren't?

What are patterns?

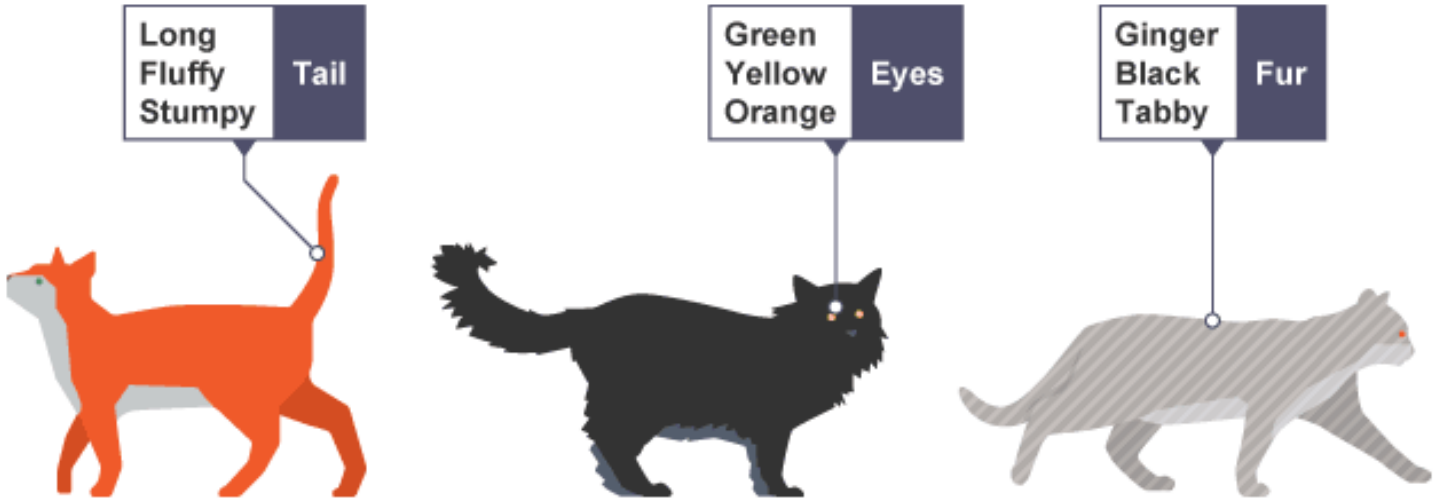
Imagine that we want to draw a series of cats.

All cats share common characteristics. Among other things **they all have eyes, tails and fur**. They also like to eat fish and make meowing sounds.

Because we know that all cats have eyes, tails and fur, we can make a good attempt at drawing a cat, simply by including these common characteristics.

In **computational thinking**, these characteristics are known as patterns. **Once we know how to describe one cat we can describe others, simply by following this pattern.** The only things that are different are the specifics:

- one cat may have green **eyes**, a long **tail** and black **fur**
- another cat may have yellow **eyes**, a short **tail** and striped **fur**



Public domain image

Power of Patterns

- Identifying patterns allows us to understand common features and apply common processes.
- Patterns exist among different problems and within individual problems. We need to look for both.
- Patterns can help us do a reality check
- Patterns can help us with abstraction:
  - Class: housecats have a set of characteristics
  - Instance: Buffy is male, has short hair, is friendly

Abstraction

- The process of filtering out – ignoring - the characteristics of patterns that we don't need in order to concentrate on those that we do.
- It is also the filtering out of specific details. From this we create a representation (idea) of what we are trying to solve.
- In Object-Oriented-Programming:
  - Classes vs. Instances
  - Interfaces vs. Implementation

Abstraction: patterns vs. details

General patterns	Specific details
We need to know that a cake has ingredients	We don't need to know what those ingredients are
We need to know that each ingredient has a specified quantity	We don't need to know what that quantity is
We need to know that each cake needs a specified time to bake	We don't need to know how long the time is

Algorithms

- An algorithm is a plan, a set of step-by-step instructions to solve a problem.
- Algorithms are used for many different things including calculations, data processing and automation.
- Algorithms can be computational procedures (quicksort, Newton's method), recipes (how to bake a cake), or other sets of procedures.

Resources

- https://computationalthinkingcourse.withgoogle.com/course?use_last_location=true
- https://www.edx.org/course/introduction-computational-thinking-data-mitx-6-00-2x-3
- https://www.coursera.org/learn/algorithmic-thinking-1/?source=phoenixCdp2016AbTest

Poster Construction Examples

Poster Layout

Use the principles of good map design!

Your layout should emphasize your main points.

Your maps and figures should tell the story.

Yeah, But How Do I Make It?

To… PowerPoint!

Remember the dreaded lab 0?

You all have access to PP through OneDrive!

You can save your final pptx document to a pdf for sharing.

Exporting Images

Export graphics whenever possible, don’t use screenshots!

Higher resolution images look better but take up more hard drive space.

Export vector graphics formats (pdf, svg, eps) if possible.

Be aware of default settings.

Final Poster Content

- Scientific papers/reports/presentations have a very specific organization

  - **Introduction** Tells us why the topic is interesting;
  - **Methods** Tell us what you did;
  - **Results** Tell us what you found;
  - **Discussion** Tells us why what you found matters.

- *Organize your posters this way!*

Final Poster Content - Introduction

Information to include

- Why is the topic important?
- Brief background info to provide context
- Your motivating question

Figures to include

- Your study area map
- A picture (if applicable).

Final Poster Content - Methods

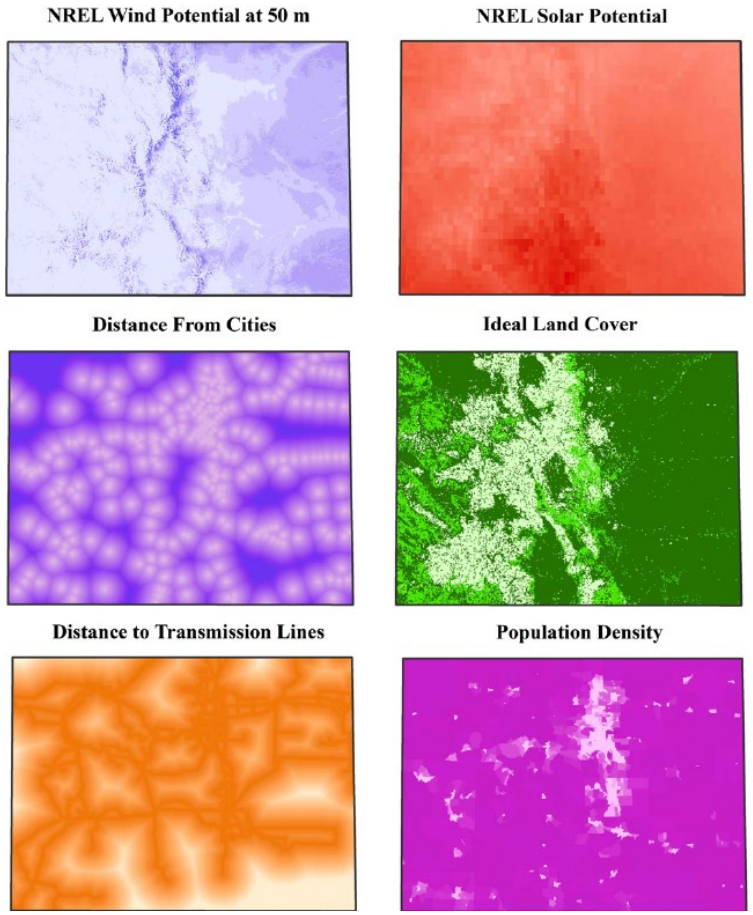
Information to include

- General steps in your methodology
- Important decisions you made in your analysis
  - Use bullets
- Table of data layers

Figures to include

- Any methods steps you want to highlight graphically, for example:
  - Flow-chart
  - Sampling Scheme

Final Poster Content - Methods



Janke, 2010

Figures to include

- Any methods steps you want to highlight graphically, for example:
  - Flow-chart
  - Sampling Scheme
- Keep them small (but not too small)
  - Your results should be the real focus

Final Poster Content - Results

Information to include

- Results of statistical analyses
- Descriptions of output maps, charts, etc.

Figures to include

- Final results map(s)
- Other results figures, for example:
  - Histograms
  - Scatterplots
  - Boxplots

Final Poster Content - Discussion

Information to include

- A summary of any important take home points from your analysis;
- What these findings might mean for the topic, or how your analysis/description/data collection (etc.) matters in this context.
- Future steps you could take;
- Notable limitations in your analysis.

Figures to include

- Figures that support your next steps

Announcements

- Poster due tomorrow. If you submit by the due date, you will have the opportunity to revise based on our comments for a better grade.
- Poster printing: Through the library
  - Unfortunately, I couldn't arrange for free printing for all Intro GIS Students.
    - If printing cost is an issue, contact me. I may be able to find some funds for a limited number of students.
  - Poster session is optional, but everybody should have equal opportunity to participate.
  - I'm going to switch one of the scheduled poster sessions to 'virtual' mode, so that everybody who wants to can particupate.

Announcements

- Last class! ☺ and ☹
- Questions/comments for me?

Do You Like GIS?

- If you’re interested in being a TA for next fall, please fill out the interest form linked on Moodle.

Spatial Autocorrelation

"everything is related to everything else, but near things are more related than distant things."

We know nearby things tend to be related, but how should we *quantify* distance?

Polygons: neighborhoods

- Primary neighbors
  - Shared edges
  - Shared vertices
- Secondary neighbors
  - Polygons that are primary neighbors of my neighbors

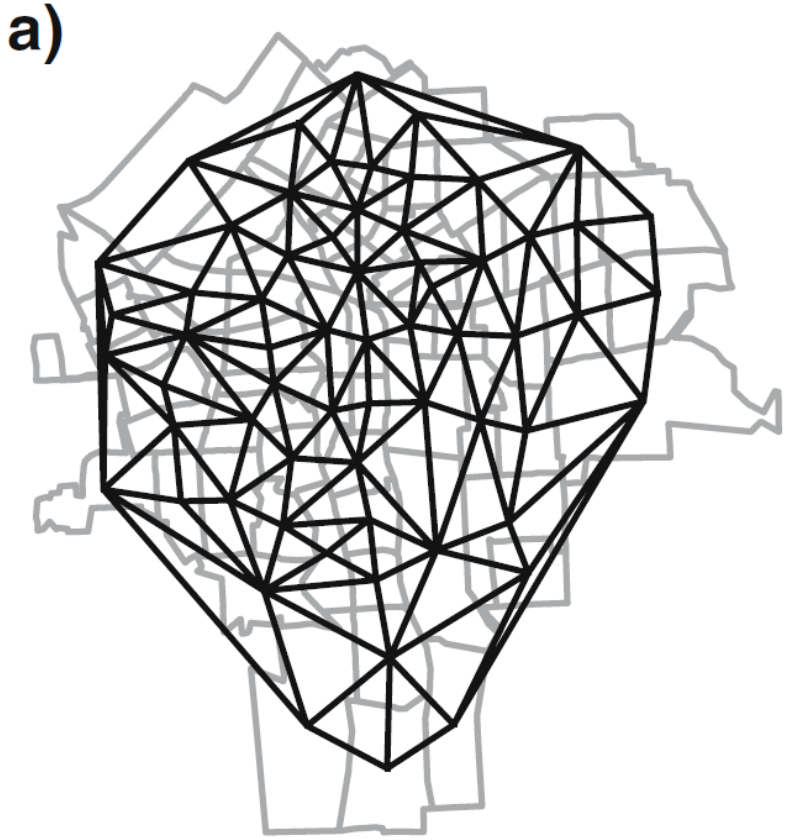
Points: neighbors and distances

- N-neighbors
  - Nearest neighbor
  - Nearest k neighbors
- Distance threshold
  - All points within x meters
- Distance decay
  - Raw distance
  - Weighted distance

Distance Matrices

We can define distances or neighborhoods using ***distance matrices***.

- Elements are the distance between row/column features.
  - Do these have to be symmetric?
- Elements can be continuous quantities: distances, weighted distances.
- Elements can be binary: 1 = in neighborhood, 0 = outside neighborhood.

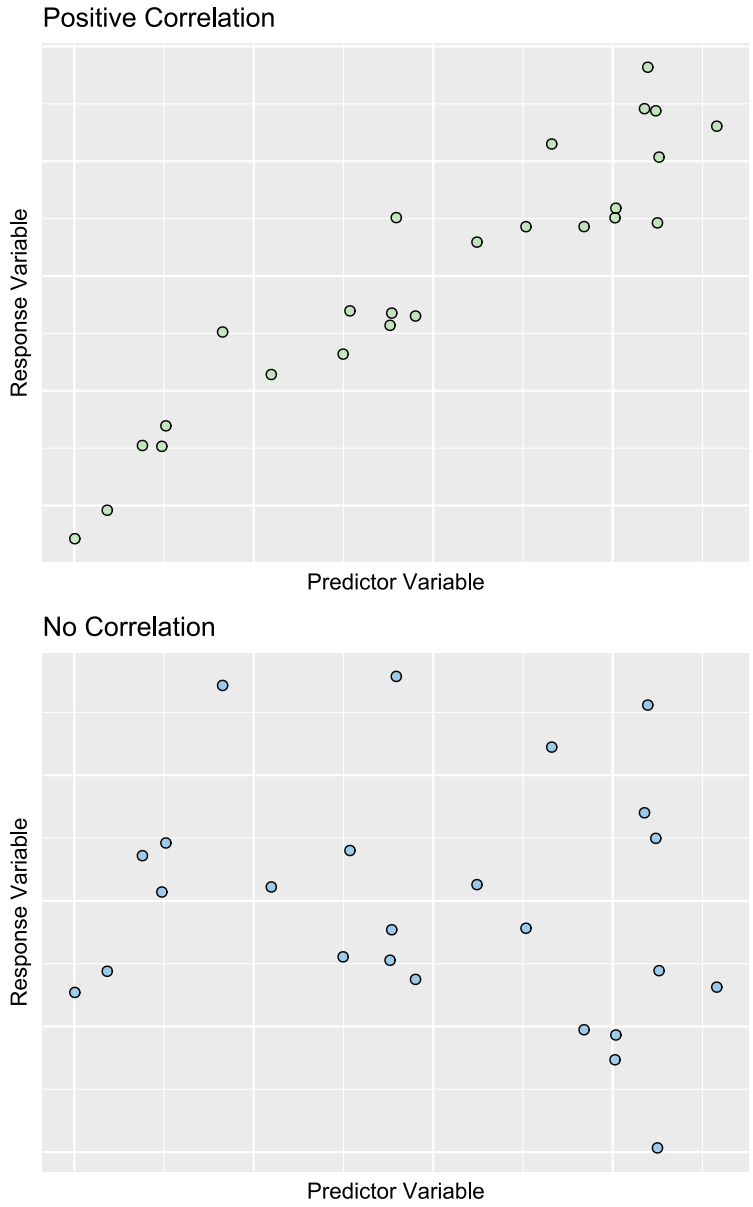


Bivand et al 2008

Correlation and Moran's I

- Correlation measures how strongly **two variables** are associated with each other.

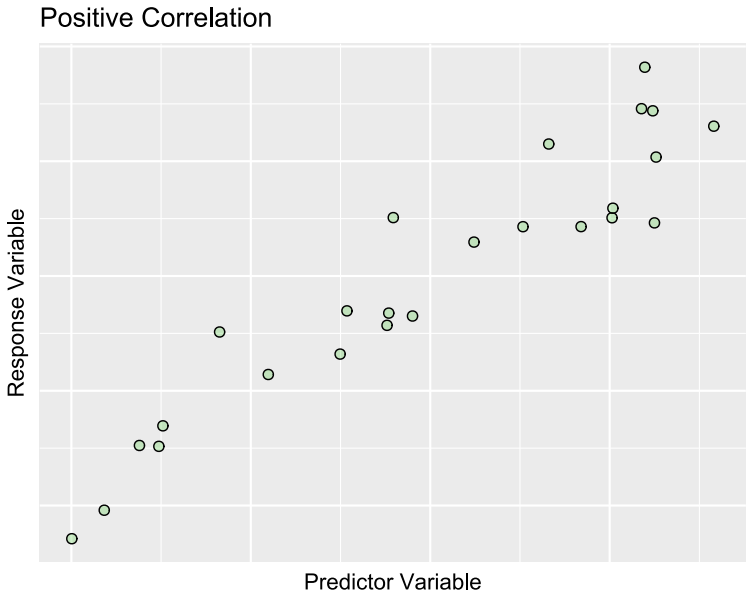
- Moran's I measures how strongly values of a single variable are associated among points **within a neighborhood *or* distance class**.

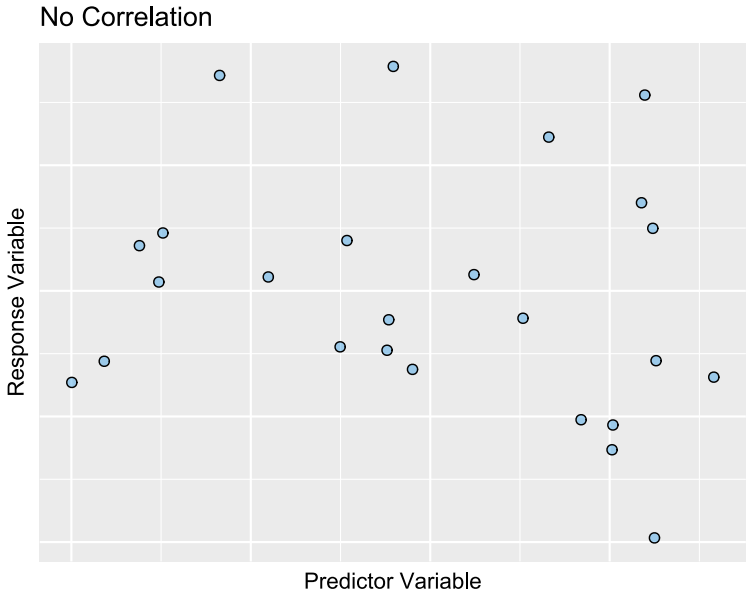


Correlation and Covariance

Covariance Intuition:

- If y tends to be high when x is high the covariance is positive.
- If y tends to be low when x is high the covariance is negative.
- If there is no relationship between x and y, the covariance zero.





Correlation and Moran's I

General idea:

$$I = \left(\frac{N_1}{\text{variance}} \right) \left(\frac{\text{covariance}}{N_2} \right) = \frac{\text{Covariance part}}{\text{Variance part}}$$

In equation form:

$$I = \left(\frac{N_{all}}{\sum_i (x_i - \bar{x})^2} \right) \left(\frac{\sum_i \sum_j (x_i - \bar{x})(x_j - \bar{x})}{N_{pairs}} \right)$$

Correlation and Moran's I

General idea:

$$I = \left(\frac{N_1}{\text{variance}}\right)\left(\frac{\text{covariance}}{N_2}\right) = \frac{\text{Covariance part}}{\text{Variance part}}$$

In equation form:

$$I = \left(\frac{N_{all}}{\sum_i (x_i - \bar{x})^2}\right)\left(\frac{\sum_i \sum_j (x_i - \bar{x})(x_j - \bar{x})}{N_{pairs}}\right)$$

Pairs of points that share a neighborhood

Correlation and Moran's I

General idea:

$$I = \left(\frac{N_1}{\text{variance}}\right)\left(\frac{\text{covariance}}{N_2}\right) = \frac{\text{Covariance part}}{\text{Variance part}}$$

In equation form:

$$I = \left(\frac{N_{all}}{\sum_i (x_i - \bar{x})^2}\right)\left(\frac{\sum_i \sum_j (x_i - \bar{x})(x_j - \bar{x})}{N_{pairs}}\right)$$

If x_i and x_j are both low or **both high**, this term will be **positive**

Correlation and Moran's I

General idea:

$$I = \left(\frac{N_1}{\text{variance}}\right)\left(\frac{\text{covariance}}{N_2}\right) = \frac{\text{Covariance part}}{\text{Variance part}}$$

In equation form:

$$I = \left(\frac{N_{all}}{\sum_i (x_i - \bar{x})^2}\right)\left(\frac{\sum_i \sum_j (x_i - \bar{x})(x_j - \bar{x})}{N_{pairs}}\right)$$

If **x_i is high** and **x_j is low** this term will be **negative**

Correlation and Moran's I

General idea:

$$I = \left(\frac{N_1}{\text{variance}}\right)\left(\frac{\text{covariance}}{N_2}\right) = \frac{\text{Covariance part}}{\text{Variance part}}$$

In equation form:

$$I = \left(\frac{N_{all}}{\sum_i (x_i - \bar{x})^2}\right)\left(\frac{\sum_i \sum_j (x_i - \bar{x})(x_j - \bar{x})}{N_{pairs}}\right)$$

If x and y are **uncorrelated**, there will be a **mix** of positive and negative values.

Correlation and Moran's I

- **Positive autocorrelation:** most pairs are similar; the sum terms are mostly positive.

- **Negative autocorrelation:** most pairs are dissimilar; the sum terms are mostly negative.

- **Zero autocorrelation:** pairs are mixed, their signs cancel out.

$$\left(\frac{\sum_i \sum_j (x_i - \bar{x})(x_j - \bar{x})}{N_{pairs}} \right)$$

Correlation and Moran's I - Intuition

General idea:

$$I = \left(\frac{N_1}{\text{variance}} \right) \left(\frac{\text{covariance}}{N_2} \right) = \frac{\text{Covariance part}}{\text{Variance part}}$$

In equation form:

$$I = \left(\frac{N_{all}}{\sum_i (x_i - \bar{x})^2} \right) \left(\frac{\sum_i \sum_j (x_i - \bar{x})(x_j - \bar{x})}{N_{pairs}} \right)$$

Examples:
Positive Spatial Autocorrelation

- Parent-offspring processes
- Land cover (vegetation types)
- Topography
- Annual rainfall
- Income level

Negative Spatial Autocorrelation?

http://gisgeography.com/spatial-autocorrelation-moran-i-gis/

Negative Spatial Autocorrelation Example

Negative spatial autocorrelation occurs when Moran's I is near -1. A checkerboard is an example where Moran's I is -1 because **dissimilar values are next to each other**. A value of 0 for Moran's I typically indicates no autocorrelation.

Checkerboard Pattern: Spatial Autocorrelation

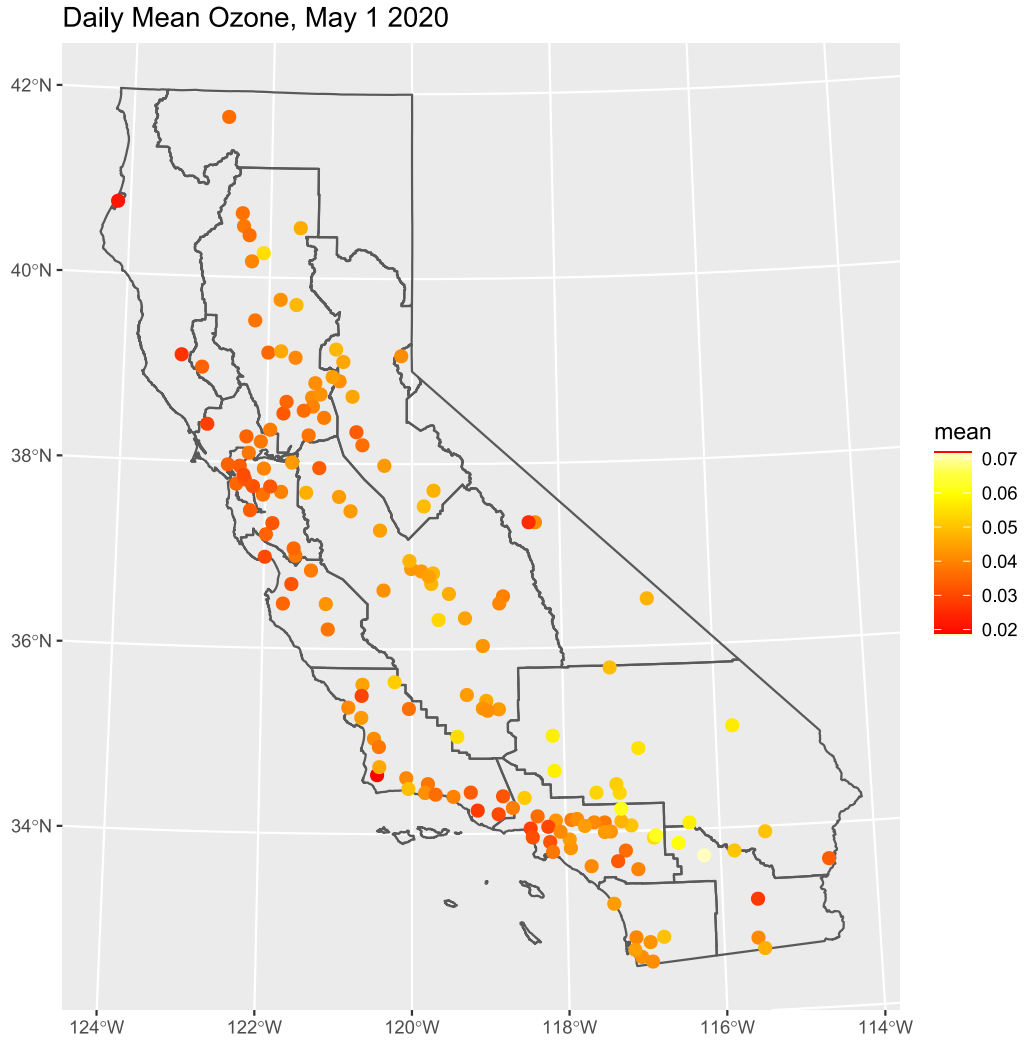
Using the spatial autocorrelation tool in ArcGIS, the checkerboard pattern generates a Moran's index of -1.00 with a z-score of -7.59. (Remember that the z-score indicates the statistical significance given the number of features in the dataset). This checkerboard pattern has a less than 1% likelihood that it is the result of random choice.

Kriging and IDW

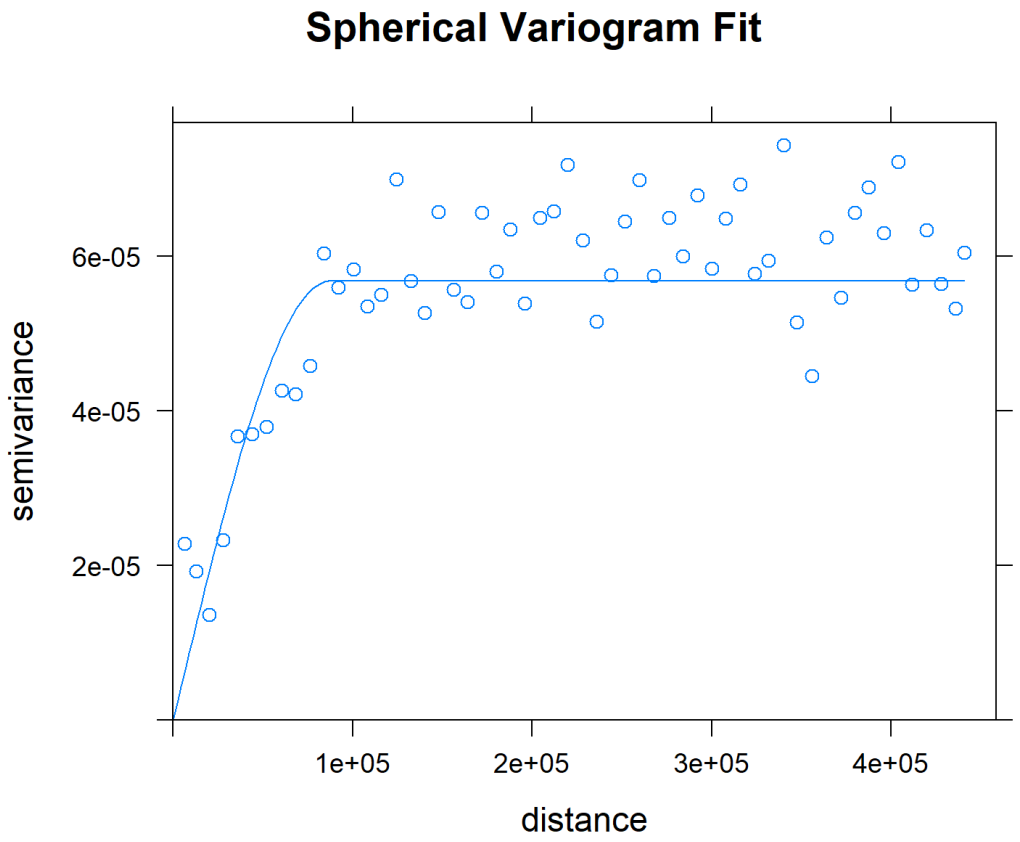
Ozone Pollution

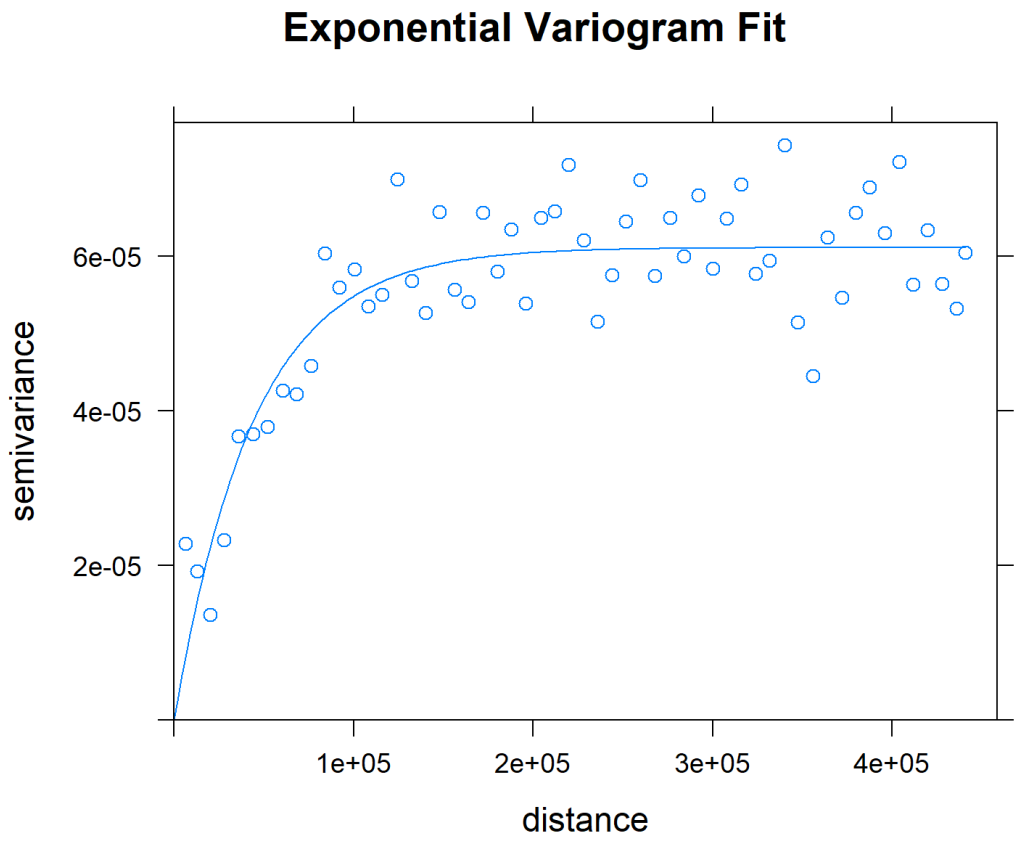
Daily Mean Ozone from 162 pollution monitoring stations

Data from the California Air Resource Board: https://ww2.arb.ca.gov/

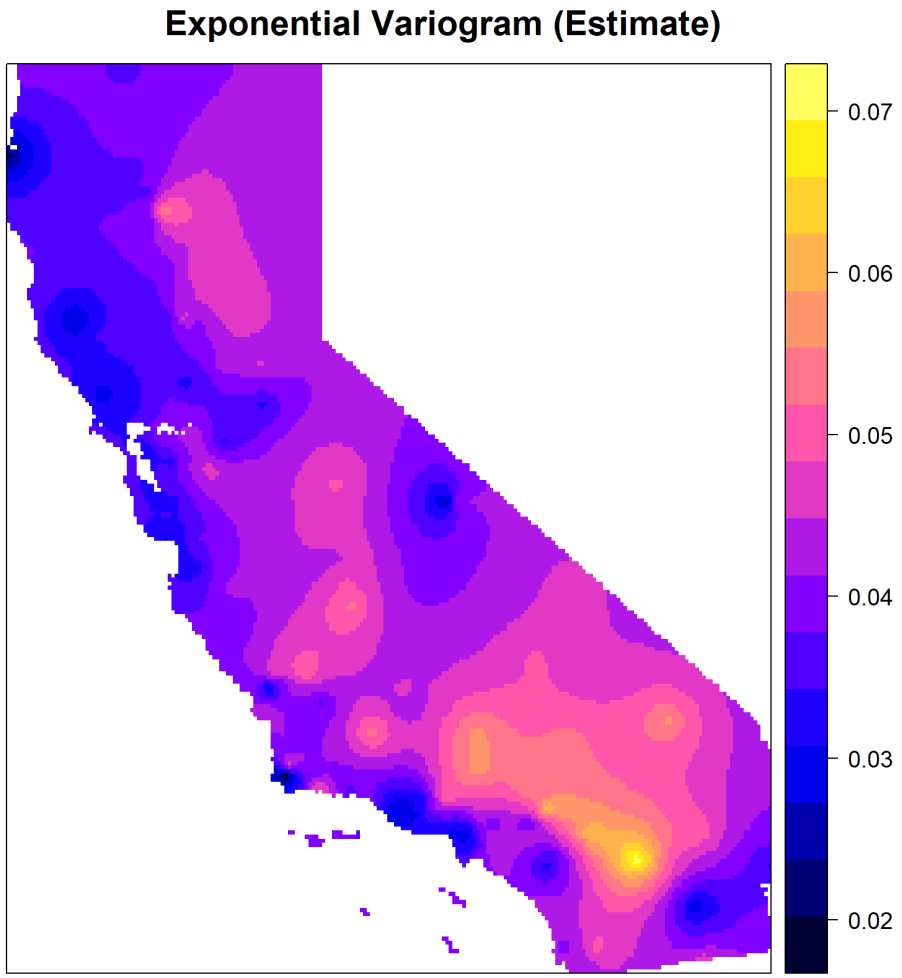


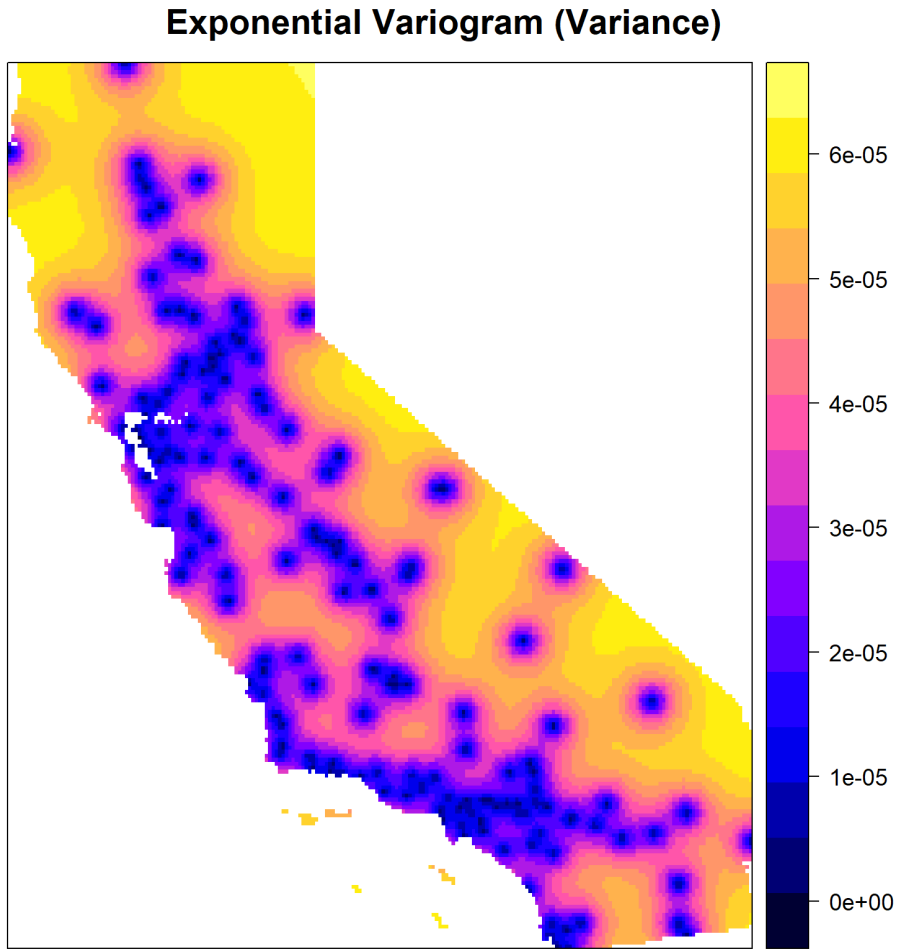
Kriging: The Variogram



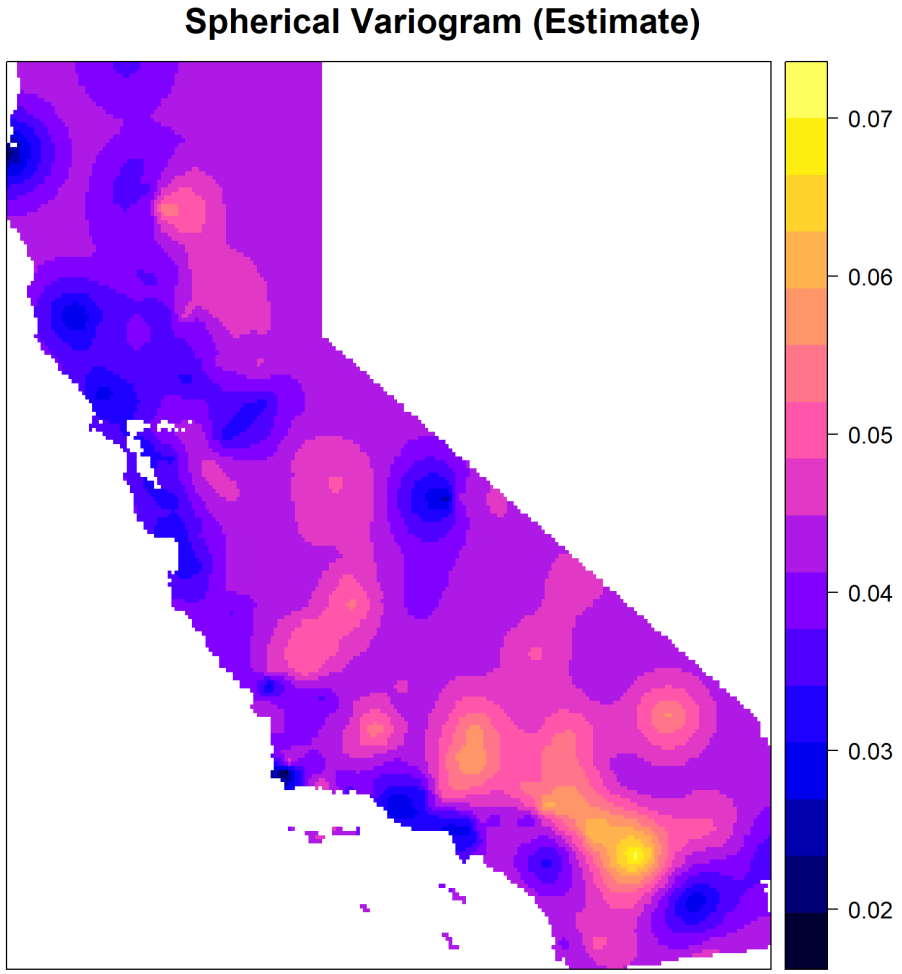


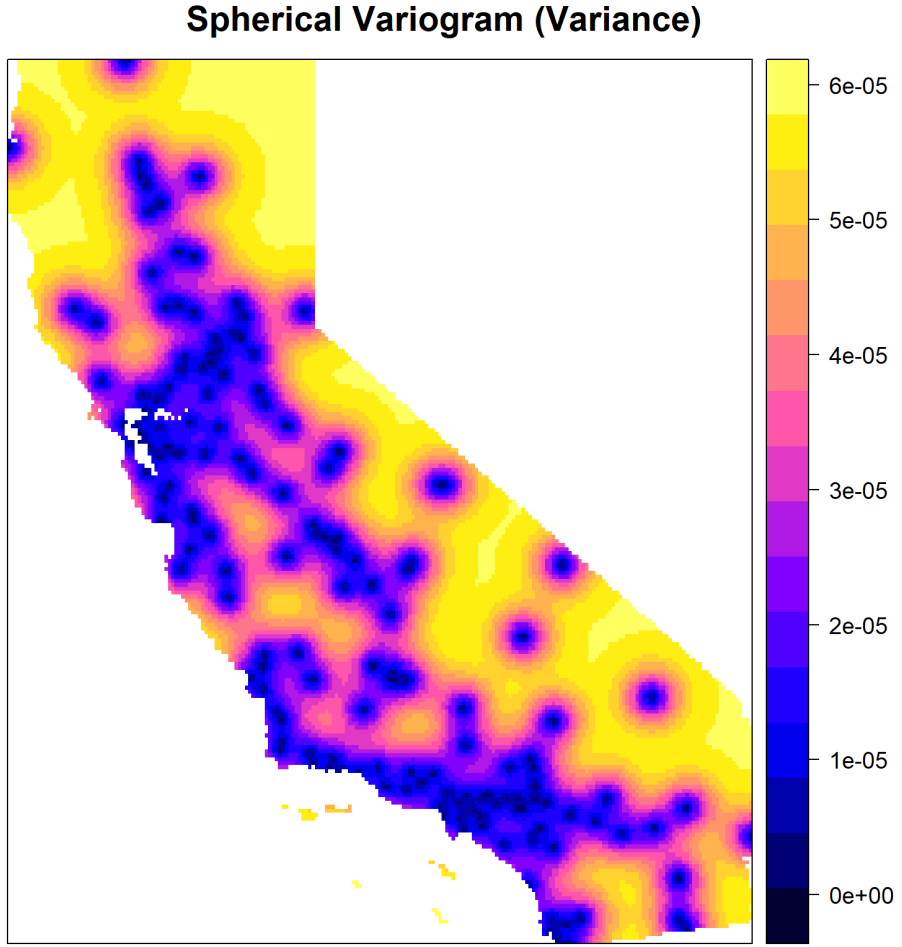
Kriging: Estimate and Variance



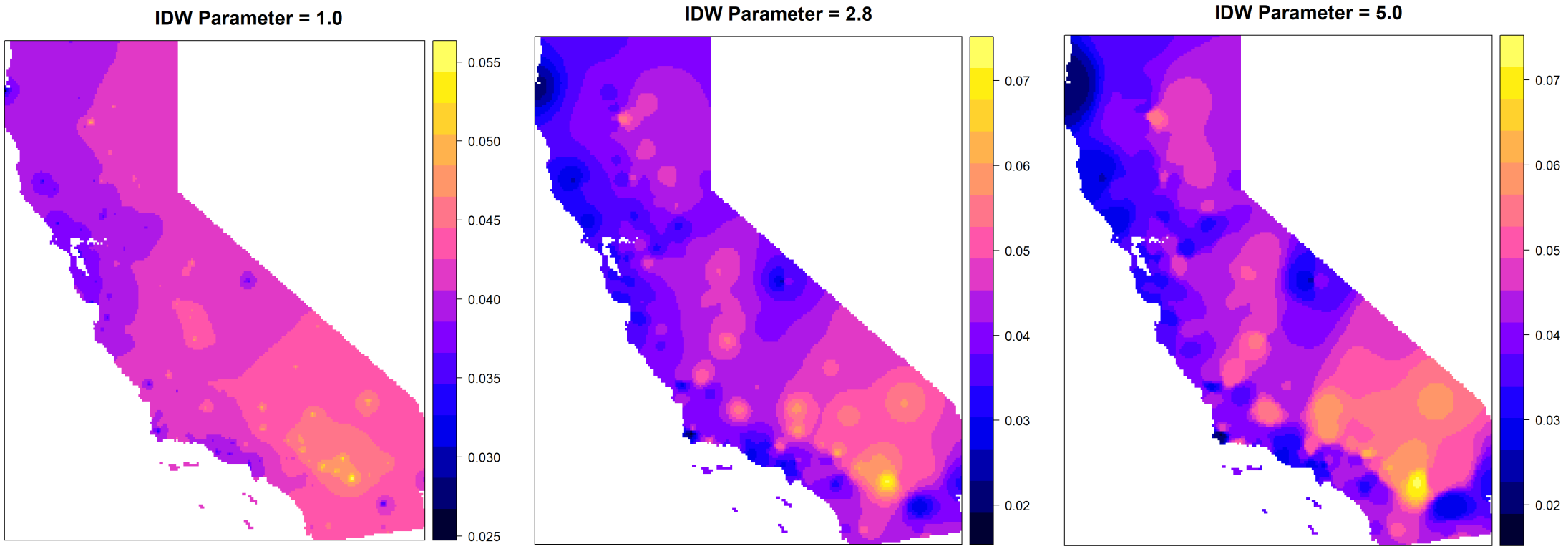


Kriging: Estimate and Variance





Inverse Distance Weighting



Interpolation: Take home messages

- There are multiple ways to interpolate surfaces between data points – each with its own strengths and weaknesses
- Your choice of interpolation method will depend on your data, your modeling purpose, your judgment.
- The best interpolation in the world can’t make up for lack of data!

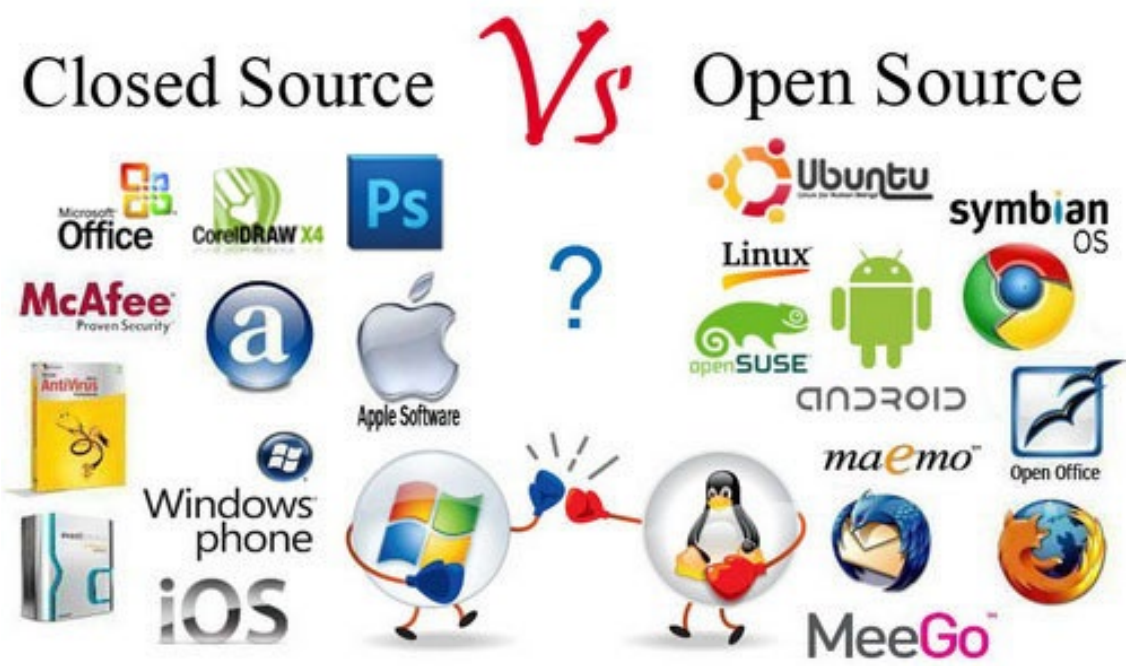
This course is just the beginning!

So, what else should you know?

- We did not cover the entirety of the GIS landscape in this course
  - Shocking
- In fact, there is other software out there you might be interested in…

Open Source GIS

- Because it only *feels* like ESRI controls the world.



Open Source vs. Closed Source

- Open Source:
  - Source code is available;
  - Allows study, change, distribution of software;
  - Encourages collaborative development
- Closed Source:
  - i.e. Proprietary
  - Restrictions on use, analysis, modification, or distribution.
  - Can also refer to file formats, workflows, etc.

The Monolith

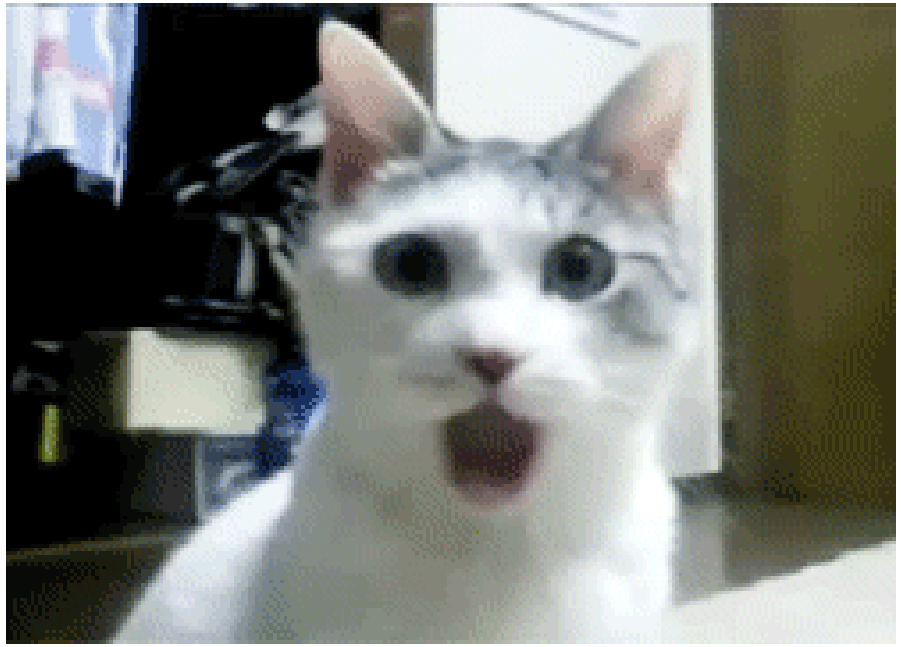
- ESRI products (ArcMap et al.) are often criticized for their proprietary nature.

- The closed source means that:
  - Licenses may restrict what you can do with your program;
  - You might not know what your processes are actually doing.

Sure, ESRI is a Big Deal…

- With ~45% of the GIS market share, ESRI products are everywhere, but:
  - You might not always be able to use ESRI products;
  - You might not always have the budget for ESRI products;
  - You might not WANT to use them!

This Course is not [supposed to be] Principles of ArcGIS!



Open Source Knowledge for Your GIS Future

- You are more marketable and employable if you can at least demonstrate knowledge of other software packages.

- Your GIS skills will grow if you can abstract your problem solving!
  - Problem-solving skills are transferrable to other software.

What's Different?

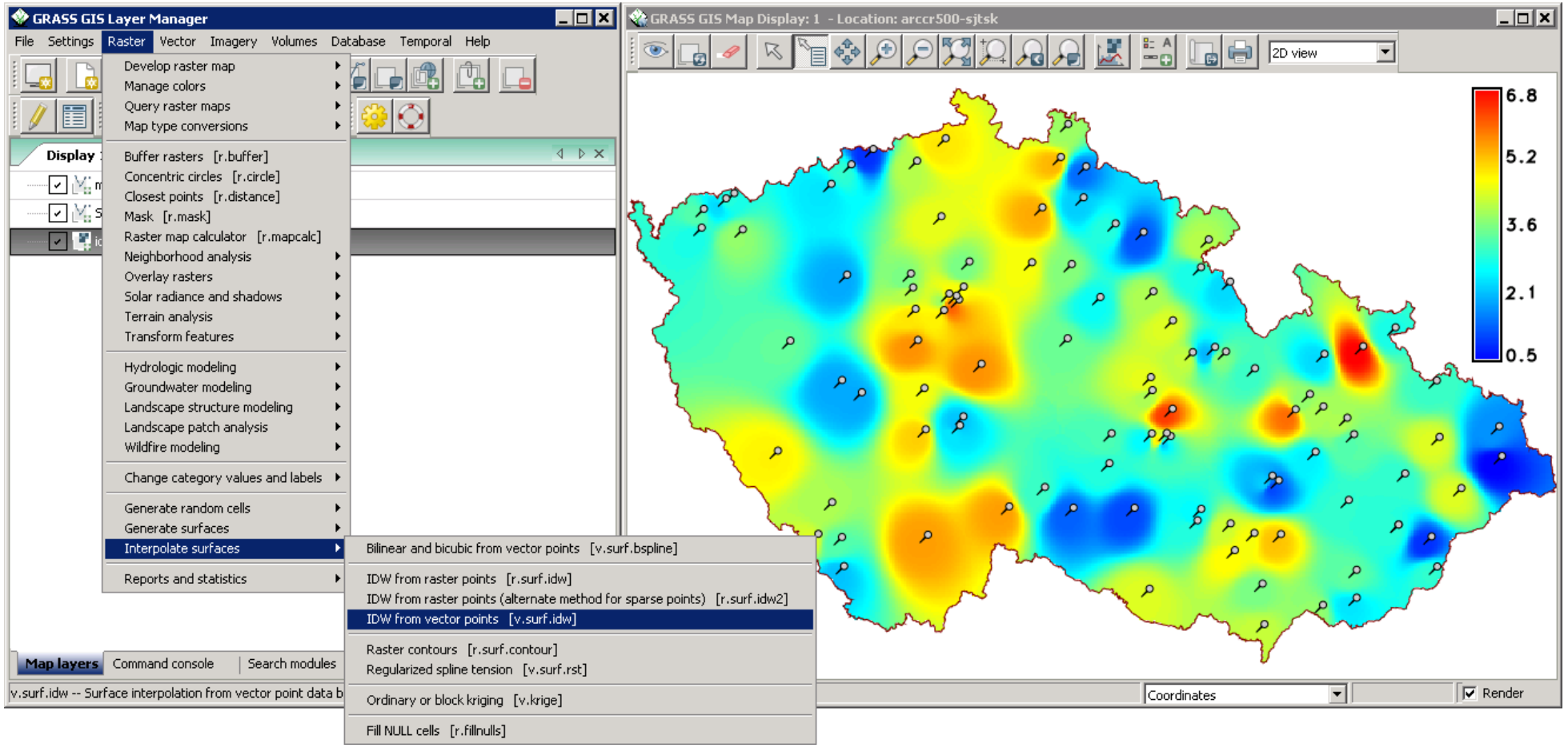
- Specifics:
  - Interface
  - File formats (sometimes)
  - Capabilities
  - Operating System Support
  - How things are Calculated (sometimes)

What's the Same?

- Concepts:
  - Fundamentals of GIS
  - Geographic Questioning
  - Cartographic Technique (but not capability)
  - Spatial-ness
  - Problem-solving skills

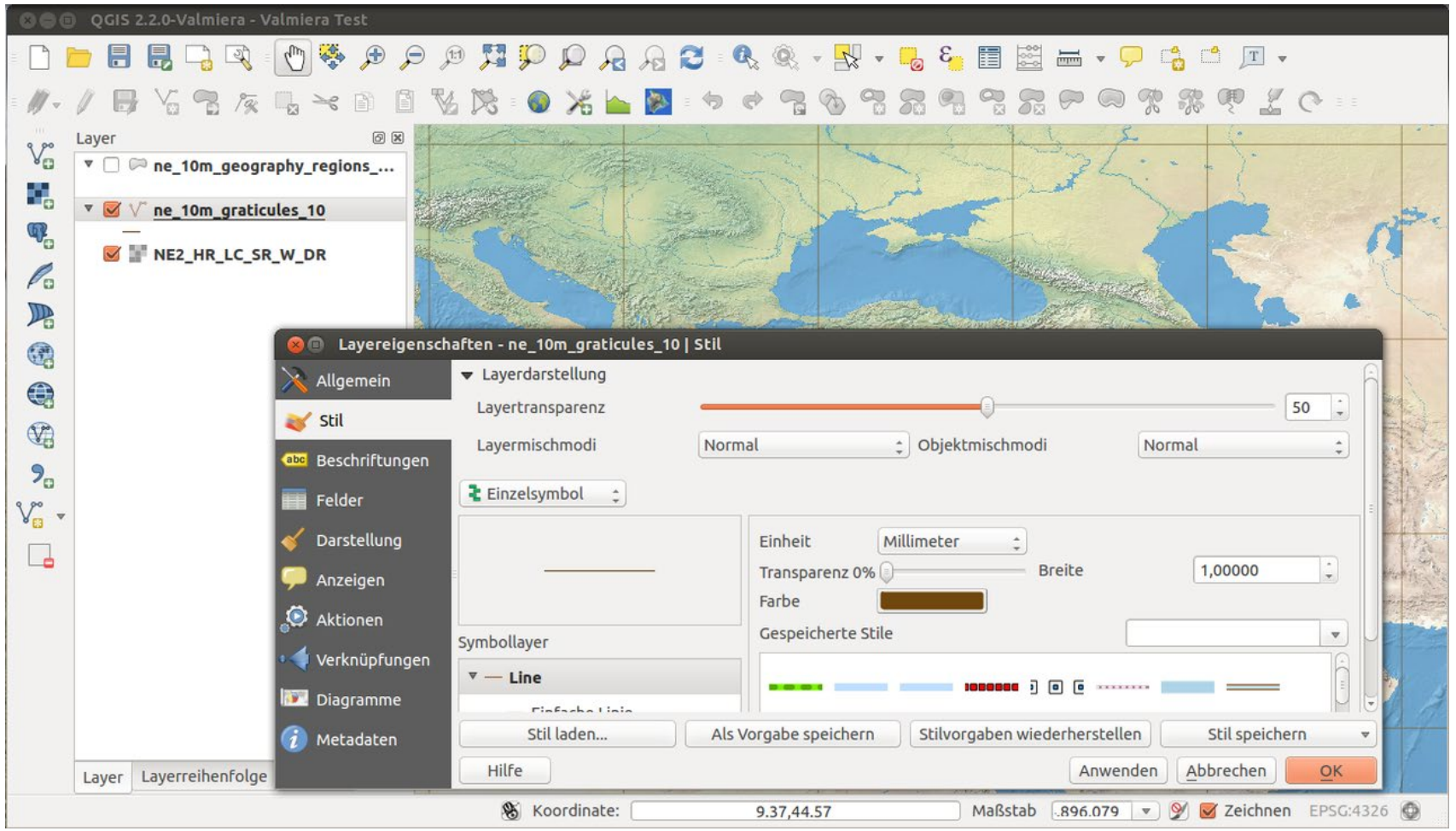
Example: GRASS GIS

Developed by the [Army Corps of Engineers]



Example: QGIS

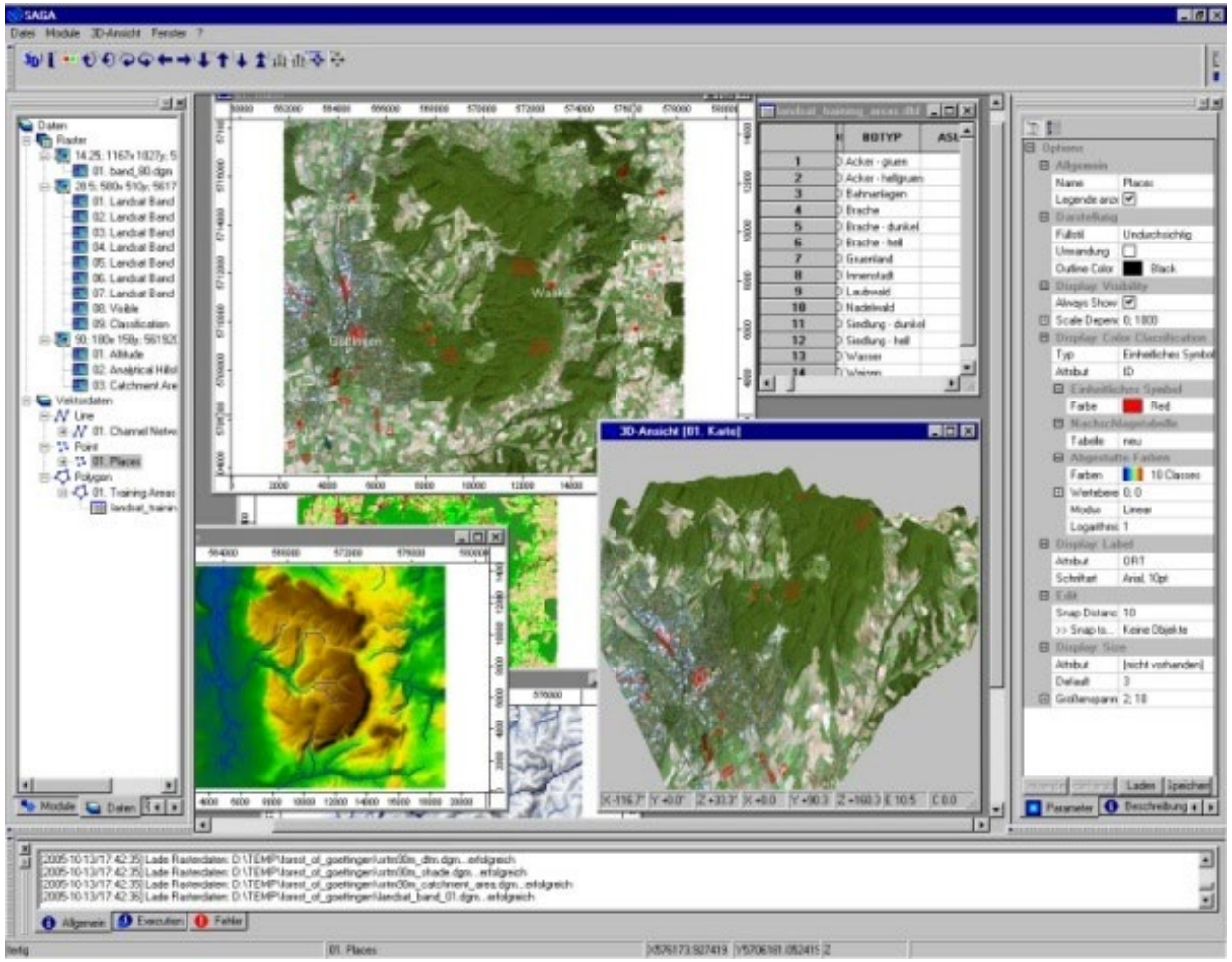
Previously Quantum GIS, a product of the Open Source Geospatial Foundation



Example: SAGA

Developed at the University of Göttengin.

Only needs 10mb!

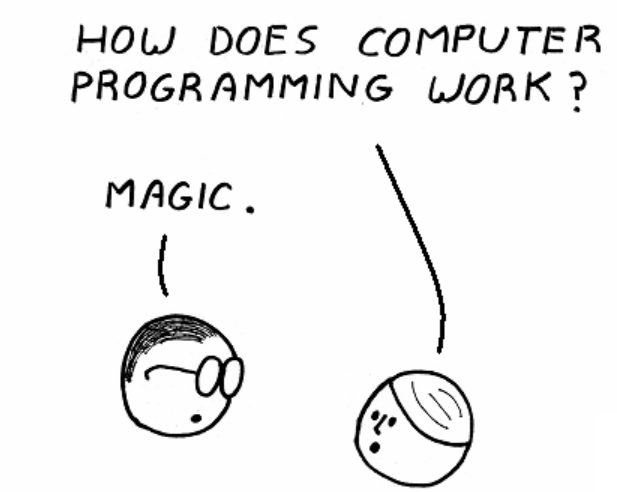


What other skills might you need?

http://www.pixelstech.net/article/images/Python.png

Programming is not ‘Beyond’ You

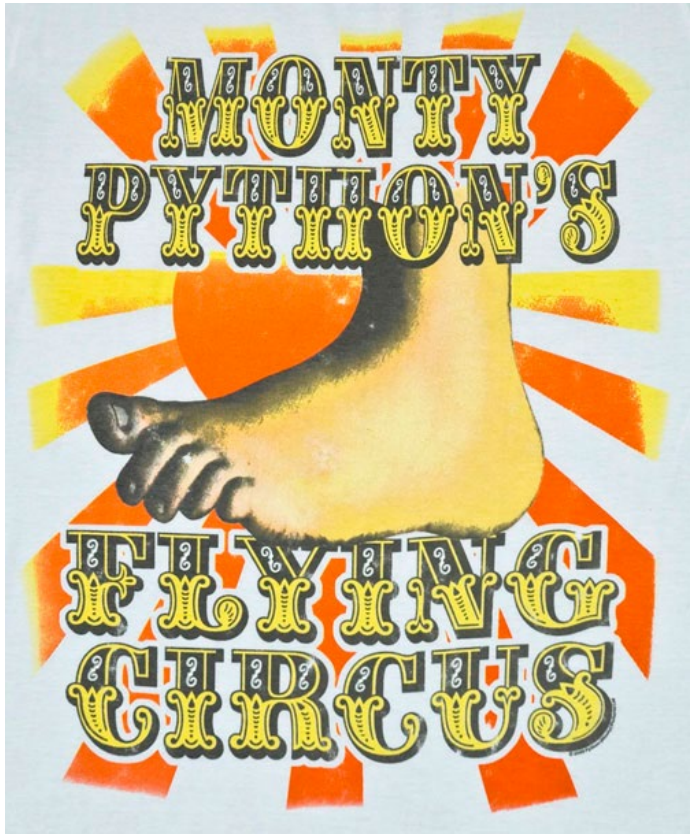
Not Like This



https://raw.github.com/tcyrus/python/master/en/assets/intro.png

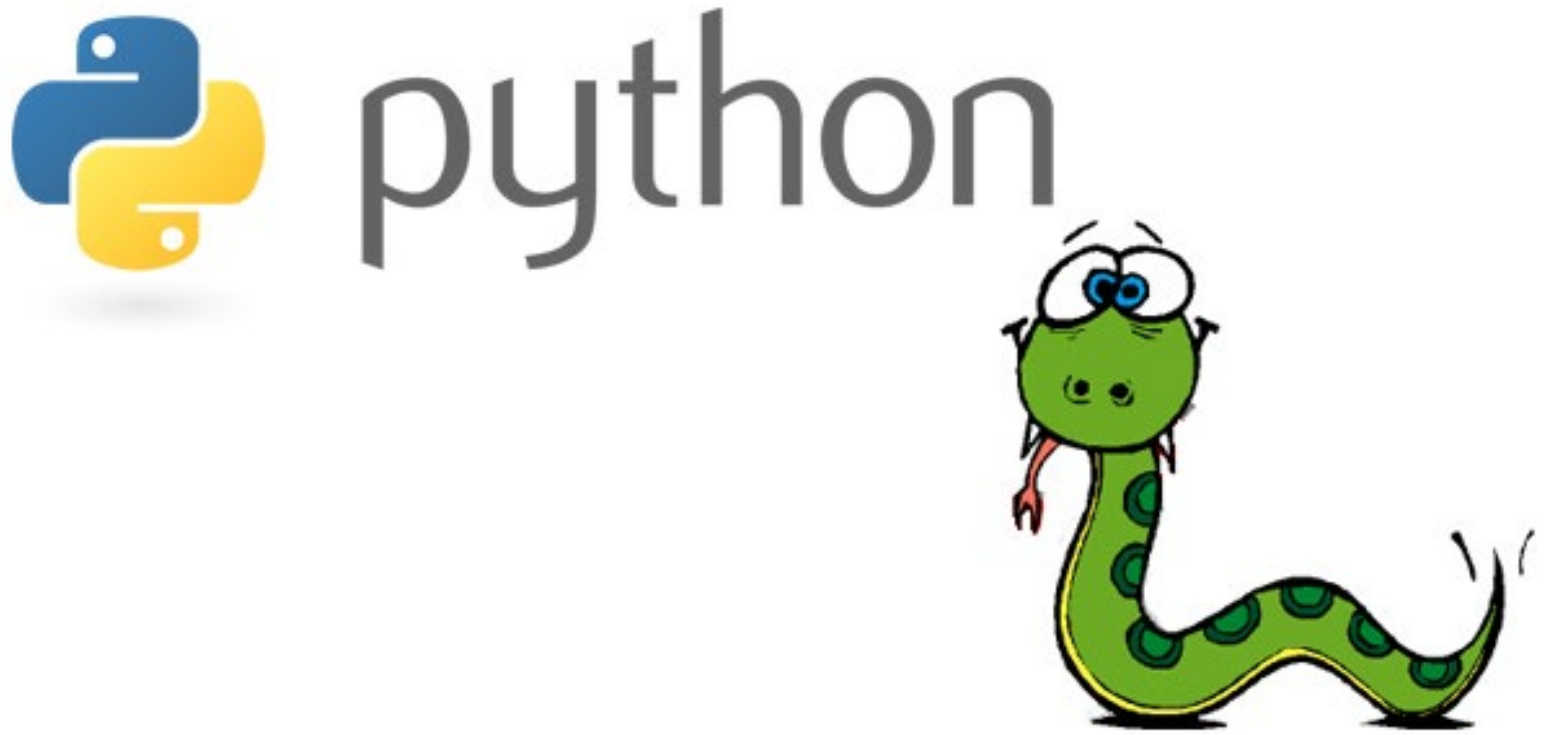
So About This Python, Anyway

- Python is a:
  - High-level
  - General-purpose
  - Interpreted
  - Dynamic
- Programming Language

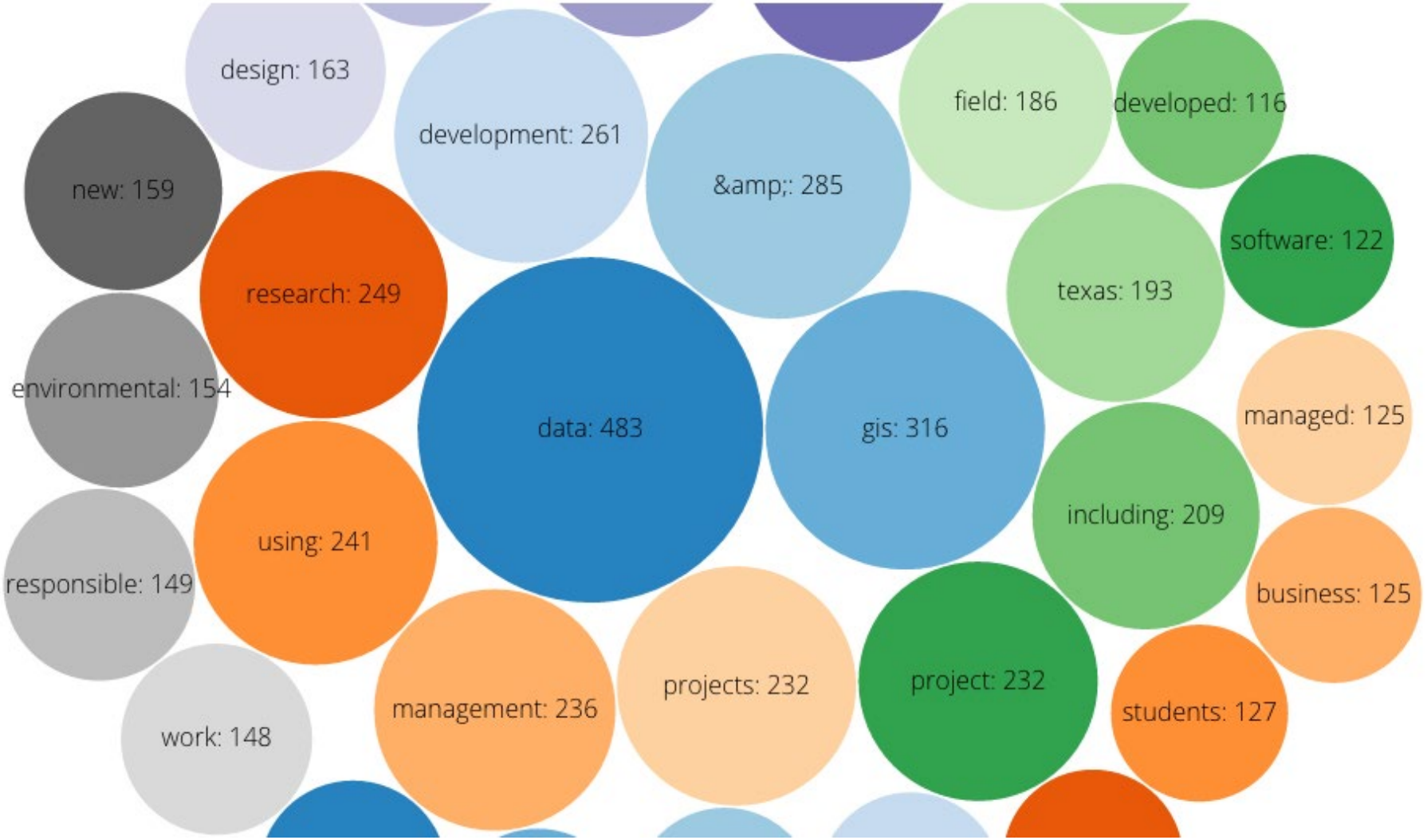


There are lots of great resources for learning Python, just Google it!

- http://www.learnpython.org/en/Welcome



What might you do?



Soft Skills vs. Hard Skills

- Hard working
- Good communication
- Team player
- “Ability to comprehend, analyze, and interpret documents.”

- Ability to analyze data in R
- Cartographic capabilities in ArcGIS
- SQL database management experience

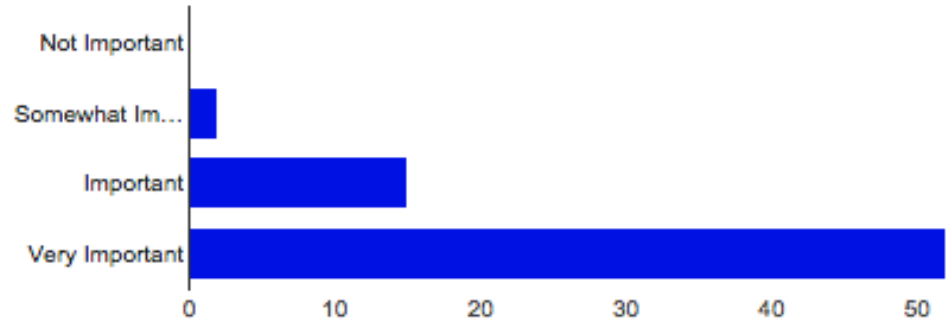
Degree Requirements

“Bachelor's degree from an accredited college or university with major work in physical or natural science, geography, engineering, planning, computer science or a related field.”

AKA ‘Have skills will hire’

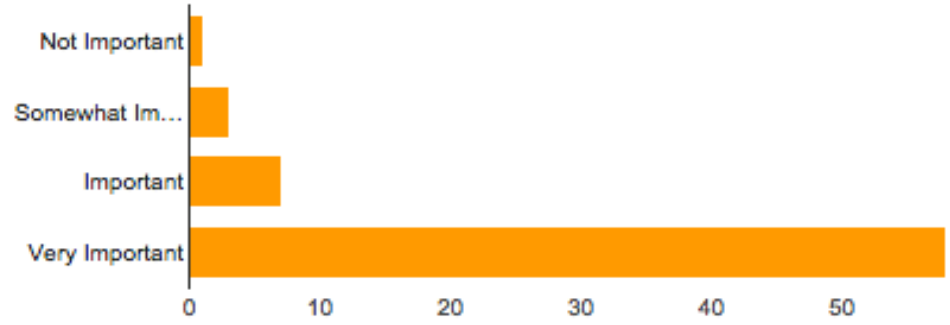
Thoughts from the Energy Domain

GIS Data Creation & Editing [Rank the importance of the following GIS skills for a new GIS professional in your group/division]



Not Important	**0**	0%
Somewhat Important	**2**	2.9%
Important	**15**	21.7%
Very Important	**52**	75.4%

GIS Data Management [Rank the importance of the following GIS skills for a new GIS professional in your group/division]

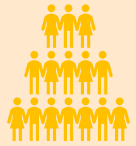


Not Important	**1**	1.4%
Somewhat Important	**3**	4.3%
Important	**7**	10.1%
Very Important	**58**	84.1%

Thoughts from Natural Resources

“Most, if not all, wildlife projects benefit from the use of GIS-based tools. Many depend on GIS-based analysis. If you want a wildlife job you need to understand GIS.”

Thoughts from the Energy Domain



Q: What would you consider the biggest deficiency when students join your group/division as new hires?

“Lack of critical thinking skills. Asking "How do I do this" rather than finding the answers themselves from help files or the web.”

Thoughts from Natural Resources

- “Learn how to code and learn how databases are structured + how to work with them. Knowledge of Python, Javascript, C++ is now a base-level job requirement for many GIS and related jobs, and, to be honest, makes many GIS tasks much easier.”

Fin

- Thank you!