Review and Wrap Up

GEOG 215 - April 22, 2020

“Nothing makes the earth seem so spacious as to have friends at a distance; they make the latitudes and longitudes.” -Henry David Thoreau

“Everything we do is right, even when it's wrong. There's always a lesson to be learned.”

— Yoana Dianika , Till We Meet Again

Final Exam Format

- 150 points (plus extra credit of 10 points)
  - 3 hours
  - On sakai (email me if you have bad internet)
  - No “select multiple answers” questions”
  - True/False (provide rationale)
  - Select the correct answer
  - Short Answer questions

- Theory questions
- Application questions
- Coding questions (less focus)
  - Fill in the blanks
  - Which code is correct
  - What will be the correct output of the following code
  - I won’t ask you to write a full line of code

Final Exam Content (Lectures)

- Week 9 onwards (plus visualization - feb 12 and feb 17 lectures)
  - Intro to ESDA
  - Distance
  - Wrangling Spatial Data
  - Spatial Neighborhoods
  - Spatial Autocorrelation and clustering (general)
    - Spatial Autocorrelation
      - GLobal measures, local measures
    - Point pattern analysis
      - Density measures
      - Distance measures
      - 1st order vs 2nd order effects

Note: I am not going to ask you any direct questions from lectures before midterm 1, but i assume you are comfortable with the material pre- midterm 1 since some concepts for Final Exam build on those:

Eg: I wont ask you a question of whether something is in a projected or a geographic coordinate system, but i might give you a question where you have to decipher the reasons behind an error due to mismatch in coordinate systems.

Final Exam Content (Labs)

- Labs
  - Lab 4
    - Vector, raster operations
    - Spatial subsetting, joins, aggregation
    - Raster operations - map algebra, local,focal, zonal operations
  - Lab 5
    - I am not going to quiz you on the writing or filling code for spatial autocorrelation but more on the outputs
      - For example, I will give you a moran' I and ask you to interpret it
      - Or I will give you a sample of the code and ask you whether the neighborhood definition is relative distance, absolute distance, 1st or order neighbors or 2nd order neighbors

Final Exam Content (Class demonstrations)

- I wont ask any questions from the code I ran during class demonstrations (that code is to help you with your project)
- Point Pattern Analysis Demonstrations
    - Again, I will not ask questions about code
    - I will show you an output (quadrat test, nearest neighbor test) and ask to interpret, or show you graphs and ask you to interpret

Final Exam Content (Readings)

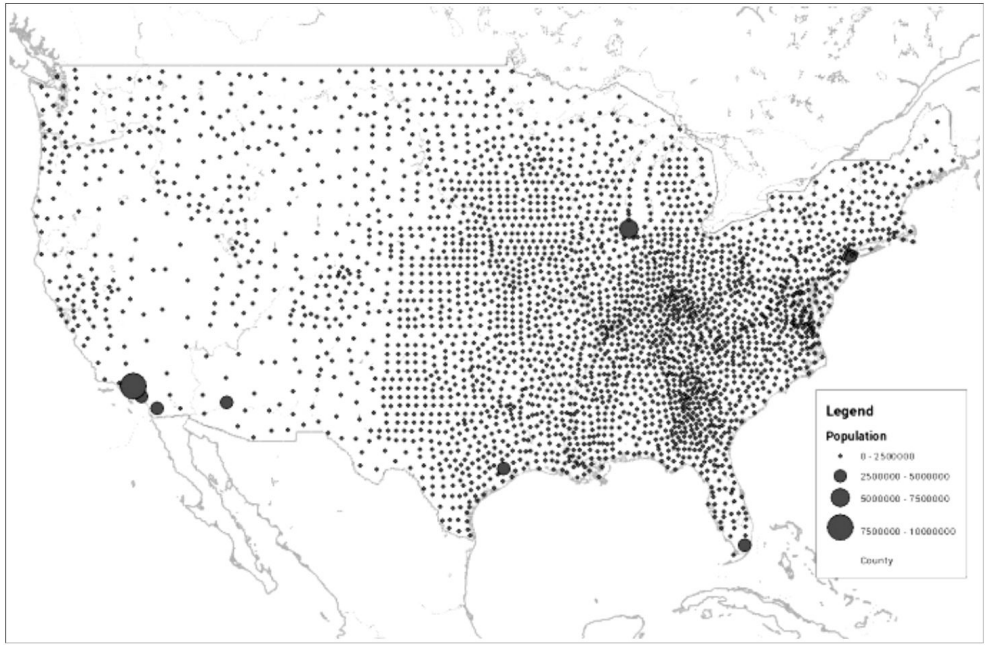
- There are few readings, the links to which I mentioned in the lecture slides. While I talked about those readings in lecture, going through the readings will help you will short answer questions, where I ask you to elucidate a bit more.

- I will also send you a compiled google doc with the specific readings i refer to in the lectures.

Sample Questions

What type of map is this?

1. Reference
2. Proportional Symbol
3. Point
4. Choropleth
5. Categorical
6. None of the above



Sample Questions

Which of the following accurately describes a local clustering analysis?

1. Identifies whether events/values are clustered
2. Identifies regions with high/low values
3. Identifies geographic centroid of a region
4. None of the above

Sample Questions

If you identify the 10 nearest neighbors for each observation, you are using which of the following distance/neighborhood types?

a) **Absolute distance**

b) **Relative distance**

c) **Great Circle distance**

d) **Topology based**

e) **Spatial autocorrelation**

Sample Questions

You have county-level polygon data and you created neighbors using Rook's case neighbors. Which of the following could you use to evaluate the nature of the neighborhood connections in your data?

a) Bar chart

b) Quadrat analysis

c) Thiessen polygons

d) Histogram

e) Construct validity

f) All of the above

g) None of the above

Sample Questions

You have hospital level data. Each hospital has a hospital type, which can be either primary care, secondary care or tertiary care. You want to create a graphical display to show overall distribution of hospitals by type. You should a create a :

a) **Bar chart**

b) **Histogram**

c) **Box plot**

d) **Scatter Plot**

e) **All of the above**

Sample Questions

Choose which words fill in the blanks: Spatial autocorrelation is the degree of _____ between objects that _____.

a) Similarity | Are located near each other

b) Distance | Are in the same study area

c) Correlation | Have similar attributes

d) Connectivity | Are correlated

e) None of the above

Sample Questions

Provide 3 different measures of distance. Give 1 sentence summary describing each with an example

Sample Questions

Provide 3 different measures of distance. Give 1 sentence summary describing each with an example

Sample Questions

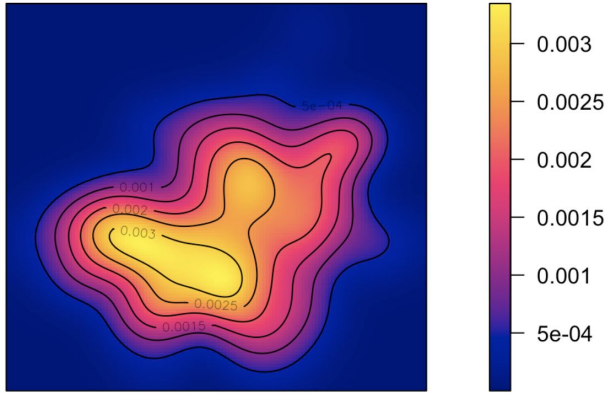
You have data of obesity prevalence [number of people who are obese per 100,000 people) by county in the US (this is continuous data). You want to create a map to show these data. Which of the following maps would you create and why? A Point Map, Graduated Symbol Map, or Choropleth map.

Sample Questions

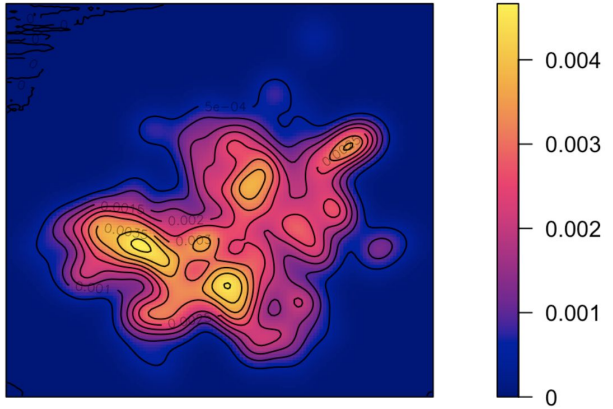
You have 2 kernel density maps describing a point process for an identical set of points. Based on visual inspection of the 2 maps, which map has a higher kernel bandwidth?

a) Map 1
b) Map 2

1



2



Sample Questions

Based on the results of a quadrat test in the image, and a VMR of 1.05, which statement is true?

a) The observed point pattern is more likely to be clustered and statistically significant

b) The observed point pattern is more likely to be dispersed and statistically significant

c) The observed point pattern is more likely to be random and statistically significant

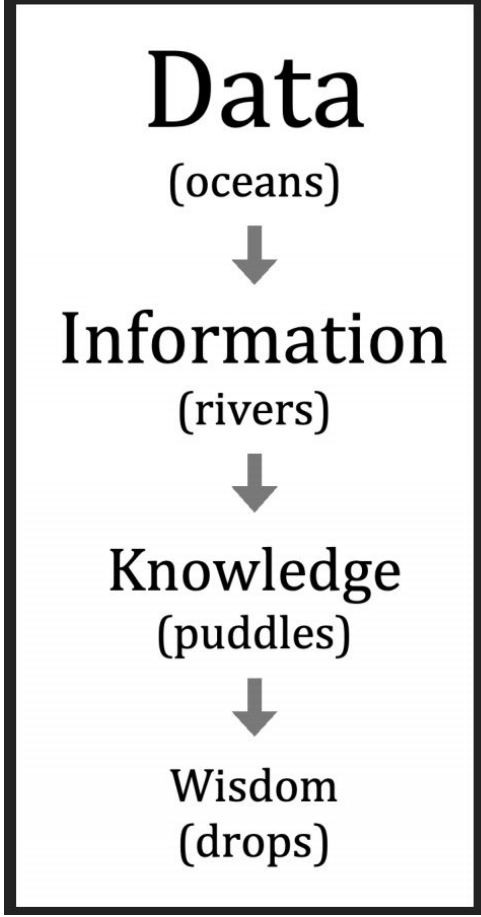
d) The observed point pattern is more likely to be clustered but not statistically significant

```
##
##  Chi-squared test of CSR using quadrat counts
##
## data:  cholera.ppp
## X2 = 940.4, df = 399, p-value < 2.2e-16
## alternative hypothesis: two.sided
##
## Quadrats: 400 tiles (irregular windows)
```

Online Zoom office hours to go through confusing materials

- Glance through your lectures
- Skim through readings/labs
- Make sure you are clear about what the code does, and why we use a certain method vs the other
- Write up your doubts
- Participate in online zoom session and ask questions
- Day/time preference?

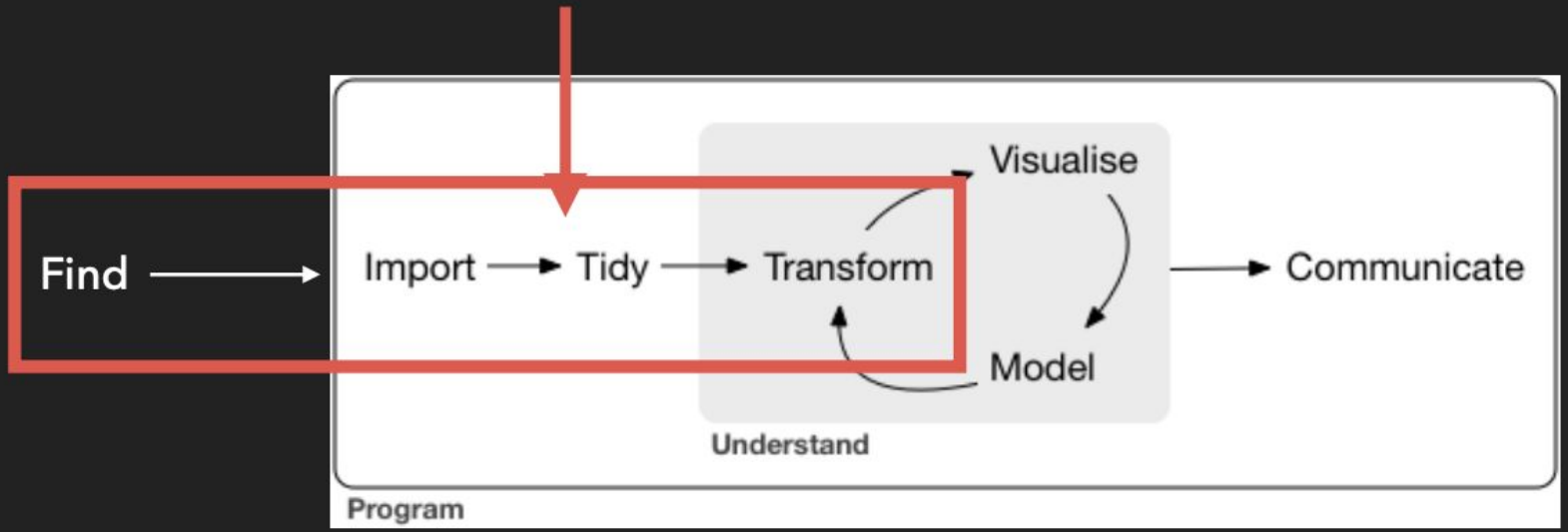
DATA SCIENCE PROCESS



- **Data**: Facts and figures, usually raw
- **Information**: data organized such that it is useful
- **Knowledge**: accumulated and integrated information on a topic over some period of time and across a broad range of situations
- **Wisdom**: application of universal principles, reason, and knowledge to discern what is true and right

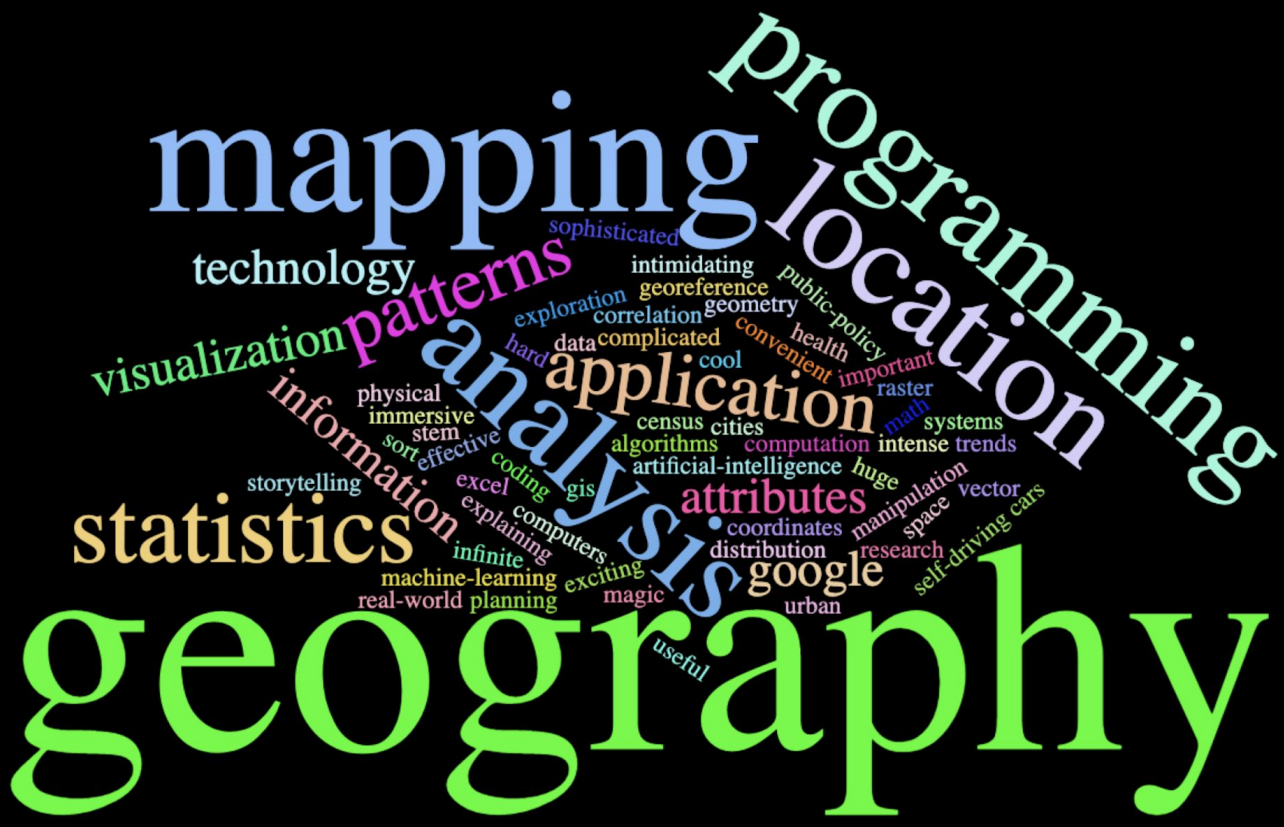
DATA SCIENCE PROCESS

"The most time consuming and least fun"

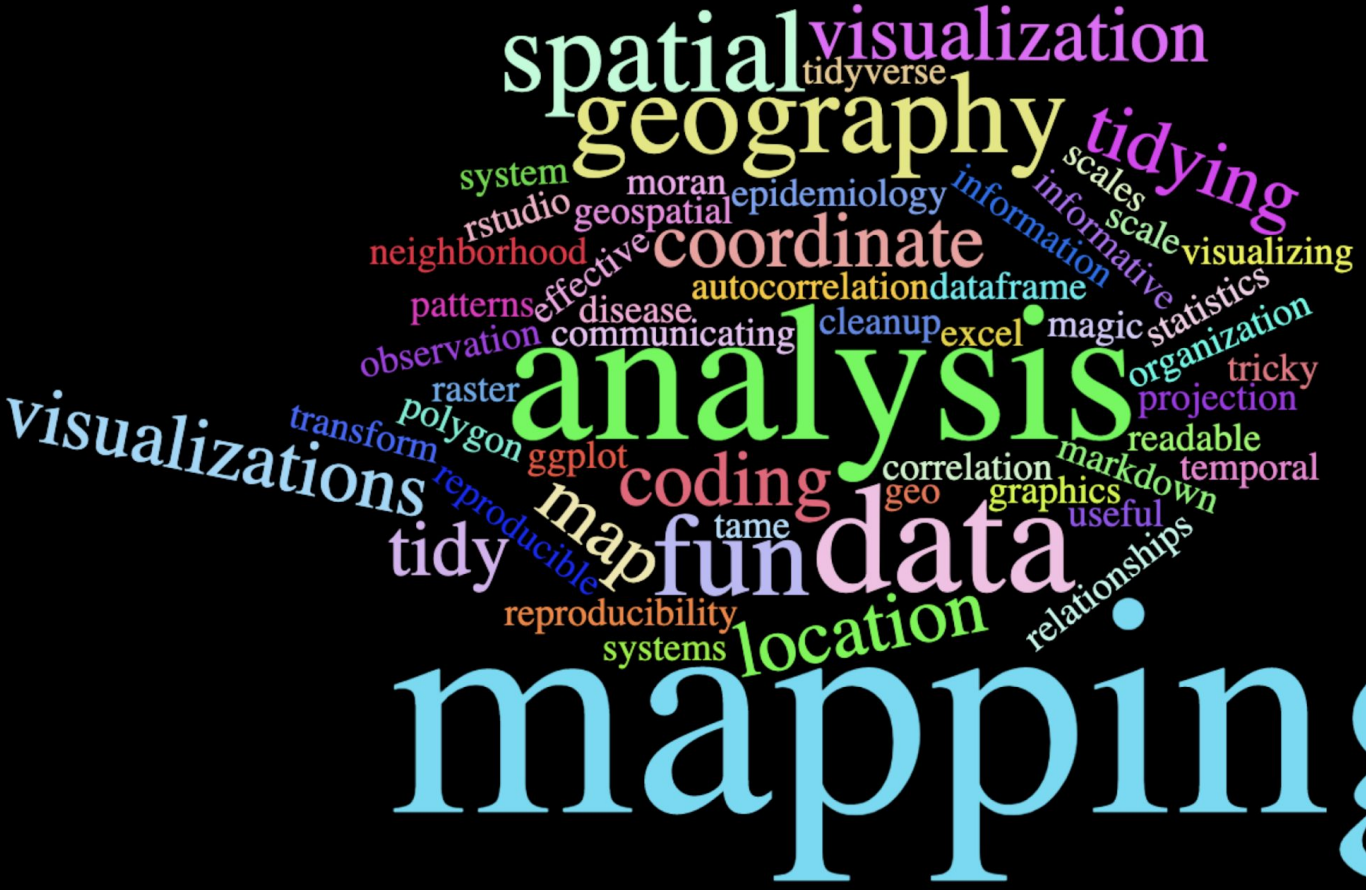


Source: R for Data Science

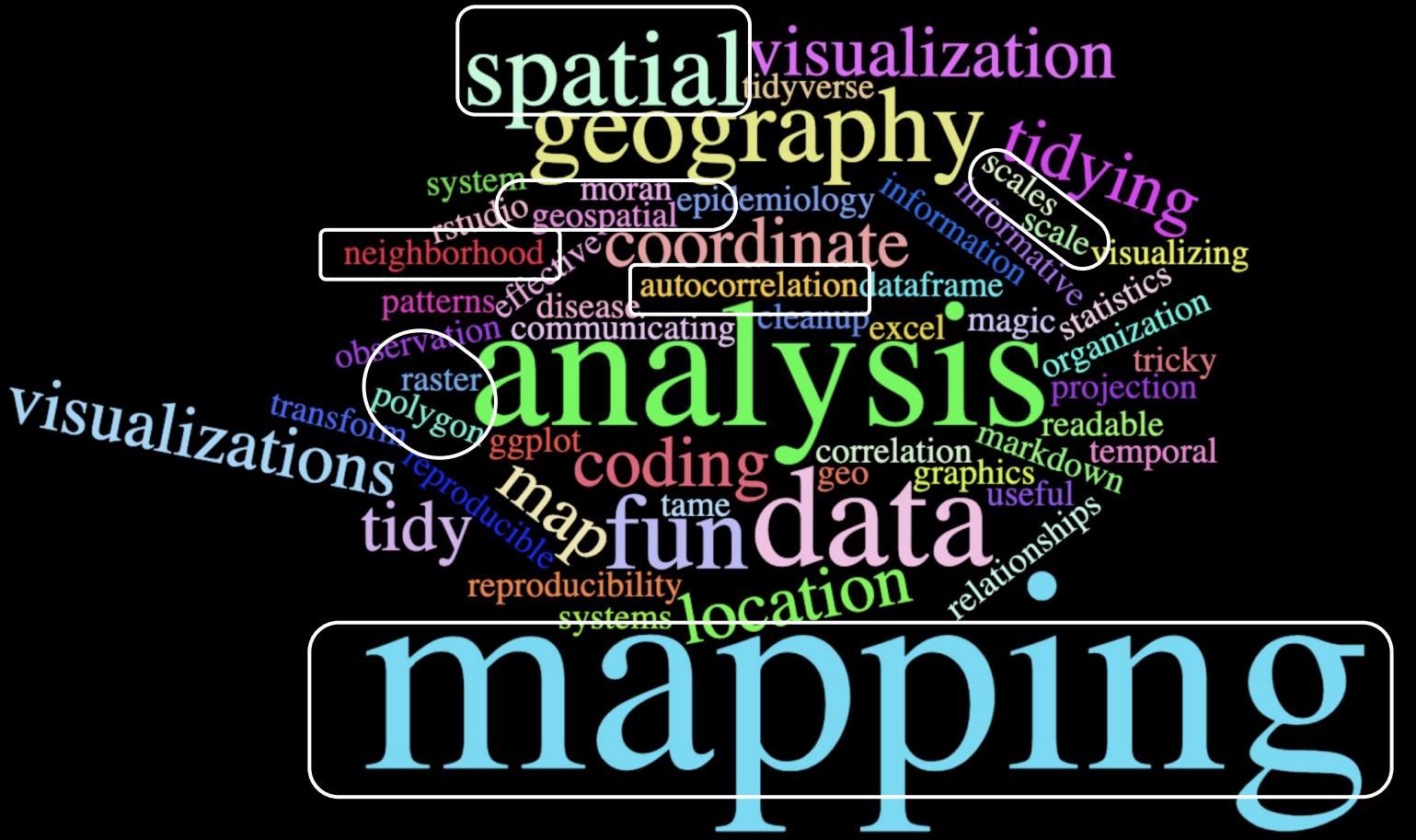
Class 1


mapping
programming
location
technology
sophisticated
intimidating
public-policy
patterns
georeference
exploration
correlation
geometry
visualization
analysis
application
hard
data
complicated
convenient
health
important
cool
raster
physical
immersive
information
census
cities
math
systems
stem
sort
effective
algorithms
computation
intense
trends
coding
artificial-intelligence
huge
storytelling
excel
gis
attributes
manipulation
vector
statistics
explaining
computers
coordinates
space
infinite
distribution
research
self-driving cars
google
machine-learning
exciting
magic
real-world
planning
urban
geography
useful

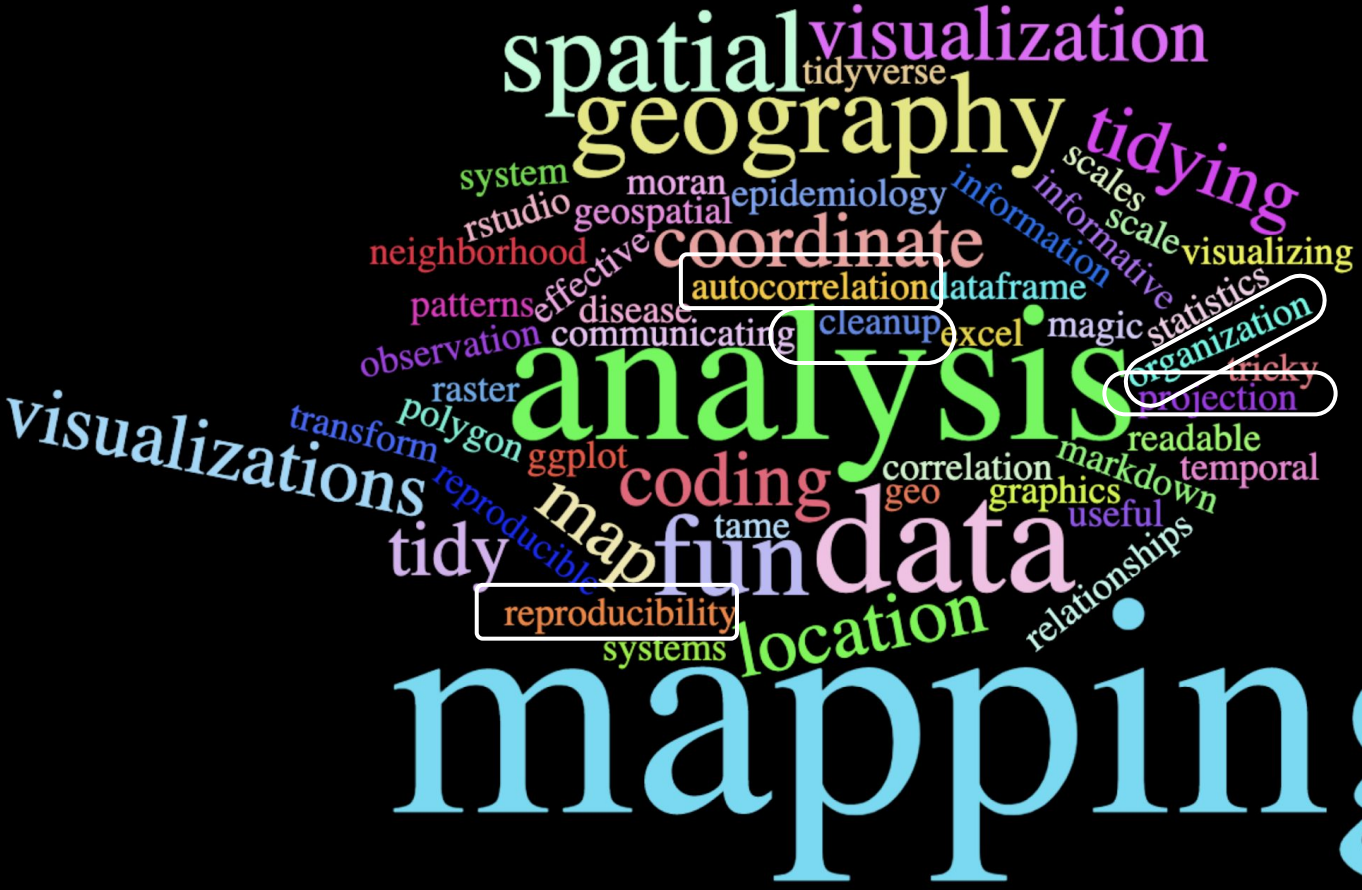
Class 24


mapping
analysis
data
spatial
visualization
geography
tidying
coordinate
coding
map
fun
location
visualizations
tidy
tidyverse
system
moran
epidemiology
rstudio
geospatial
scales
scale
visualizing
information
informative
neighborhood
effective
autocorrelation
dataframe
statistics
patterns
disease
observation
communicating
cleanup
excel
magic
organization
tricky
raster
projection
polygon
transform
reproducible
ggplot
readable
correlation
markdown
temporal
geo
graphics
useful
tame
relationships
reproducibility
systems

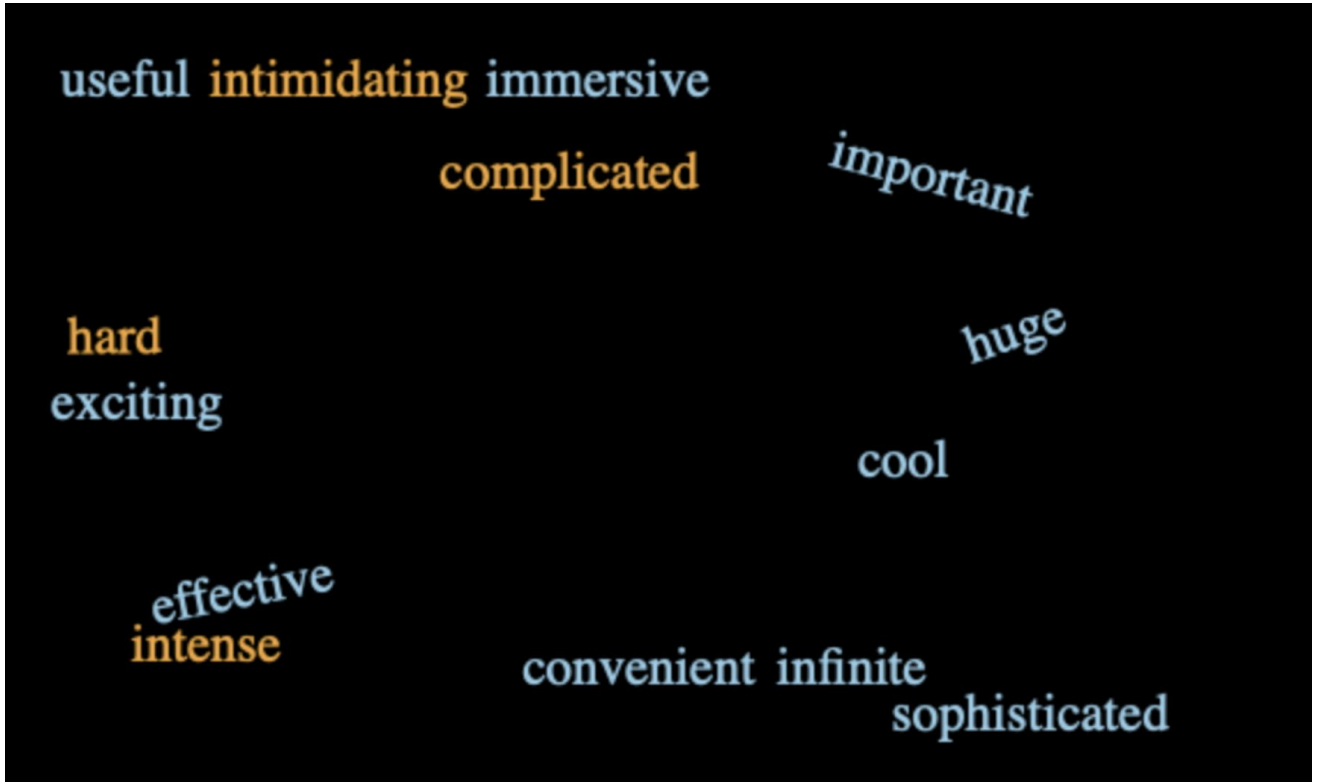
From “abstract” to “concrete”



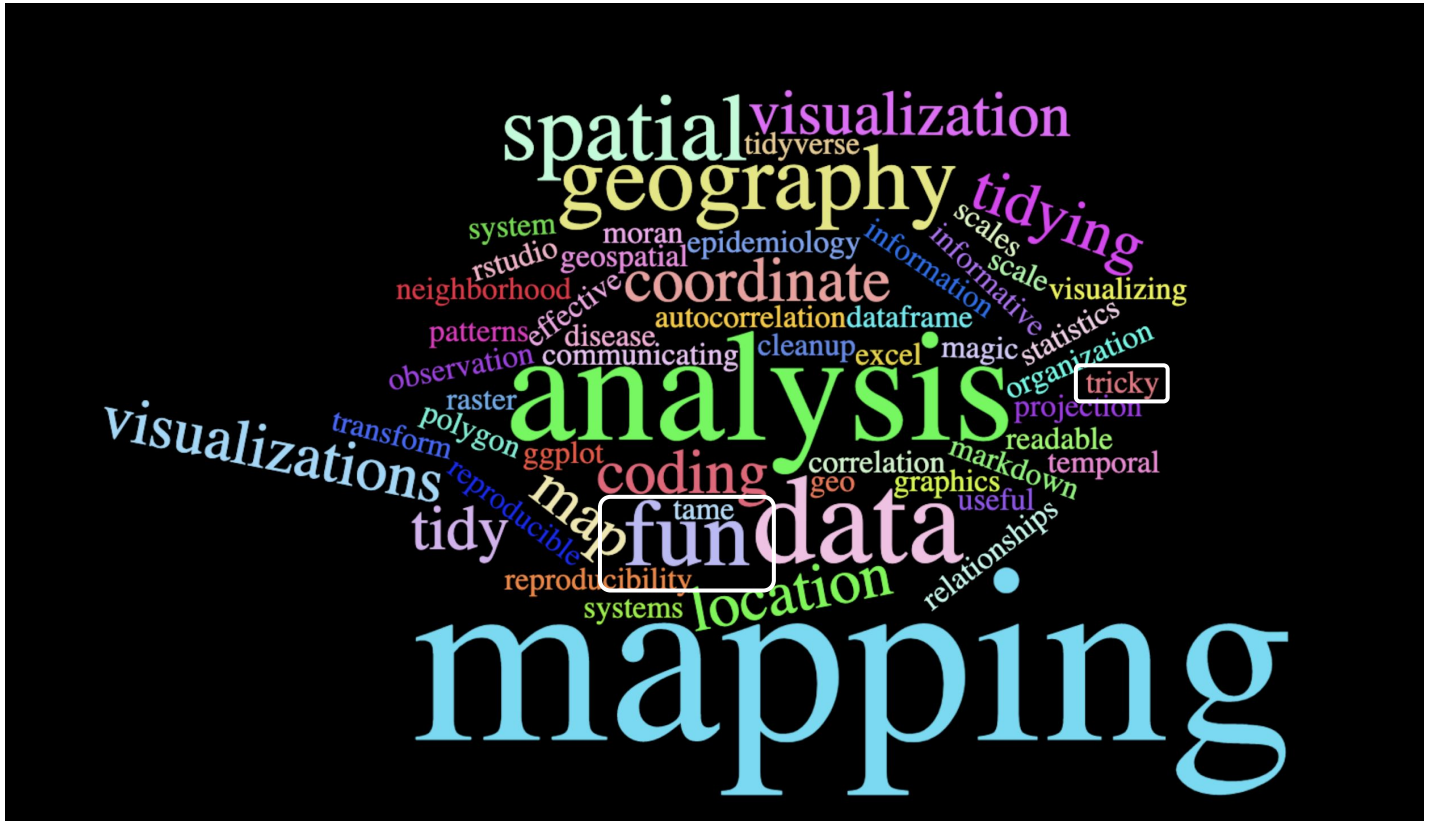
Much more than just mapping



From potentially not so “fun”



To mostly Fun but still tricky??



You all accomplished a lot

- You wrote 3600 lines of code on average!
- You learnt how to use 24 R Packages
- Conducted a complete analysis on your own (one more on its way)
- Contributed (and read through) 362 piazza posts
- Tolerated me :)
- Most importantly , you all taught me a lot!

Where can you go with you have learned

- Advanced classes
  - GIS, and advanced GIS classes in the department to learn more about advanced spatial operations
  - Take advanced spatial stats and programming class (in the future)
  - Take R classes in other departments (stats, ecology, environmental science, public health)

- Showcase your projects/homeworks as part of portfolios for jobs and internships

- Push yourself - dabble with new problems, new data, think of a problem in multiple ways

- Take more online courses
  - Data camp (free for you till end of July)

- Show off how cool you are to your friends

QUESTIONS/CONCERNS ?

HOW CAN I HELP ?