Course Introduction

Michael Noonan

DATA 589: Spatial Statistics

Table of contents

Course Overview

About Me

Name: Michael Noonan

Office: SCI 379

Email: michael.noonan [at] ubc.ca (use subject heading DATA589 in all email communication)

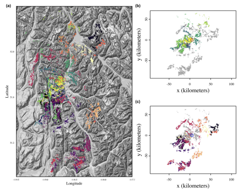
Office Hours: TBD, or by appointment arranged via email.

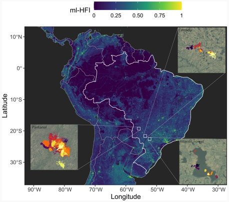
Course Resources: Canvas & GitHub

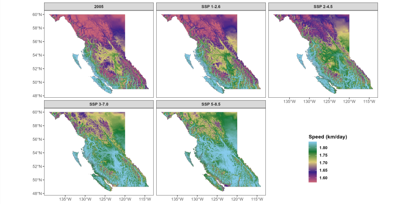
Why Me?

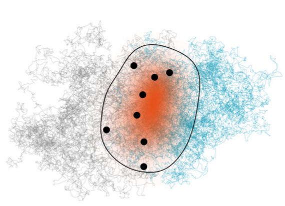
THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

Quantitative Ecology Lab
UBC Okanagan









What this course is about

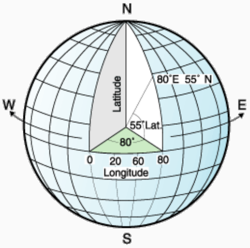
- Focus is on modelling spatial data (i.e., combining data with models to generate descriptions of spatial processes).
- You'll learn methods for handling the most common types of spatial data (Ripley's K, point processes, Kriging, regression with correlated errors, etc...).
- How to work with and visualise spatial data (not as straightforward as other data types).
- How to use open source software (R) to apply these analyses (traditional carried out in ArcGIS; $1,370/yr).

What this course is not about

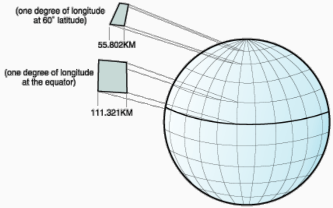
- Basic statistics (concepts like means, medians, variances, probability distributions, regression, covariance should be familiar to you).
- Remote sensing.
- Computer programming (we will be using `R`, but the course is not focused on 'how to code').
- Methods for handling *ad hoc*, corner cases.
- ArcGIS, QGIS, or cartography.

Projections

The Earth is spherical.



Flattening it to two dimensions is challenging.



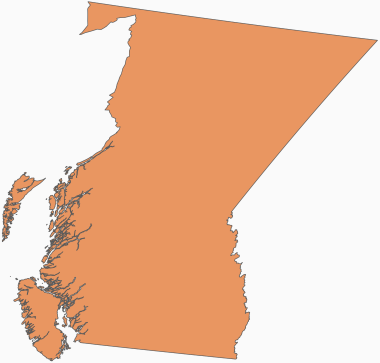
We use "projected coordinate systems" to project maps of the Earth's spherical surface onto a two-dimensional plane.

Projections cont.

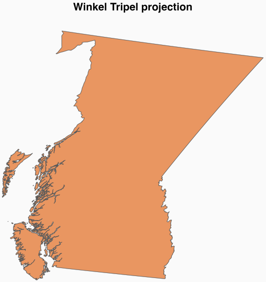
Each projected coordinate system will result in a different two-dimensional map... which can have large impacts on any downstream analyses.

Latlong Projection

Winkel Tripel projection



Projections cont.



Projection systems are not a focus of this course.

Different projections have different use cases (global, continental, city), so the choice should be guided by the study scope/aims (there are many resources to help you with this).

If you work on spatial data in the future, ensure you use a consistent projection system throughout your workflow.

How the course is structured

- For each topic, there will be a core lecture and an associated lab assignment.

- Lectures will cover the core concepts of the course. Lecture slides will be posted the evening prior to the lecture. You are encouraged to take notes, and to ask questions in the lectures. All lectures will be recorded and made available to you.

- The labs use structured tutorials to guide you on the use of the open-source software program `R` for applying the methods learned in the lectures to data. The lectures and labs are designed to be *complementary* and not all the material in the practicals will be covered in the lectures and vice versa.

Course Evaluation

Lab practicals (4)	40%	Due at the end of each lab.
Group Project	30%	End of the day Apr 28.
Final Exam	30%	In class on final day (Apr 27).
Total	100%	

Labs and assignments

There will be a total of 4 lab assignments (1 for each lab) to be completed throughout the course.

The course GitHub page and Canvas will host the labs and assignment templates, and the various datasets. The lab assignments are designed to build skills in working with and analysing spatial data.

Grading: Each lab assignment is worth a total of 10% of your total grade with some points coming from answering the questions correctly, and some from the cleanliness and documentation of your code.

Assignments are to be completed alone. You can use handwritten notes, lecture slides, and the guided lab practical to help you during the assignment. No other outside resources or communication software are permitted (e.g., slack, stack overflow, coding blogs, chatGPT etc...).

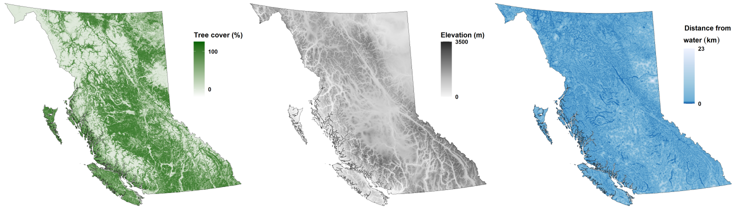
Because some lines of code can be slow to run, the time limit for submission is based on the time you start knitting your RMarkdown files.

Group Projects

Working in groups of 3, students will be required to complete a data analysis assignment and submit a report describing their work.

Each group much select a species of their choosing and download the occurrence records from GBIF data repository (https://www.gbif.org/).

Groups will be provided provided with a set of environmental variables, and must use these to describe trends in the observed point data.



Final Exam

The final exam will consist of a set of question designed to assess students' understanding of the core concepts covered in the course.

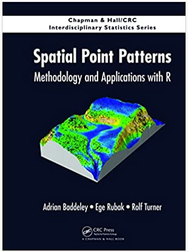
The exam will be comprehensive, and any material covered in the lectures and labs will be testable.

Grading: The final is worth a total of 30% of your total grade.

Optional Material

There is no textbook for this course, but if you are interested in expanding your knowledge beyond what is covered in lectures, the following textbook is recommended:

- Baddeley A, Rubak E, Turner R. Spatial Point Patterns: Methodology and Applications with R. 2015. CRC. $\sim$ \$120



Lecture outline

Week	Lecture Topics
1	Course introduction; Spatial Intensity
2	Spatial Correlations; Poisson Point Processes Models
3	PPP Model Validation; Spatial Autocorrelation; Variograms
4	Kriging; Co-Kriging; Regression-Kriging; Correlated Errors

Why Spatial Statistics?

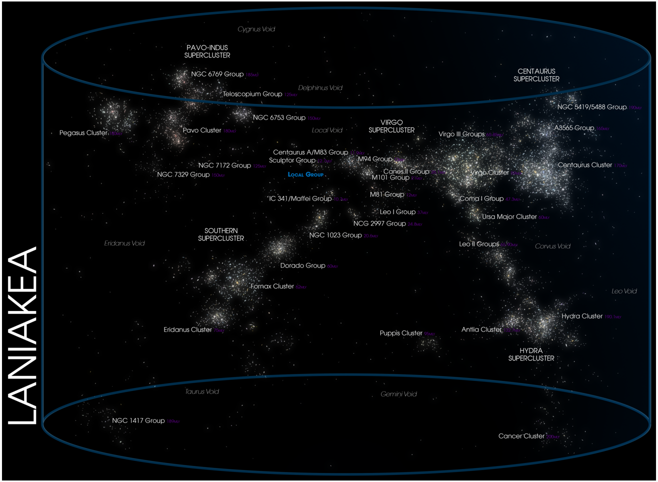
Science is a process of learning about the world around us. As scientists, we weigh competing ideas about how the world works (hypotheses) against observations (data).

But our descriptions of the world are almost always incomplete, our observations have error, and important data are often missing... So, how do we accurately compare what we observe with what we hypothesize without bias?

Statistics

... but we exist in a physical world and everything happens 'somewhere'.

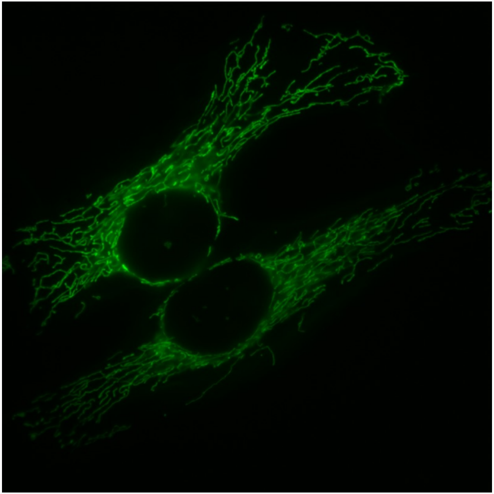
Stars are found in clusters.



Source: Wikipedia

Cellular structures

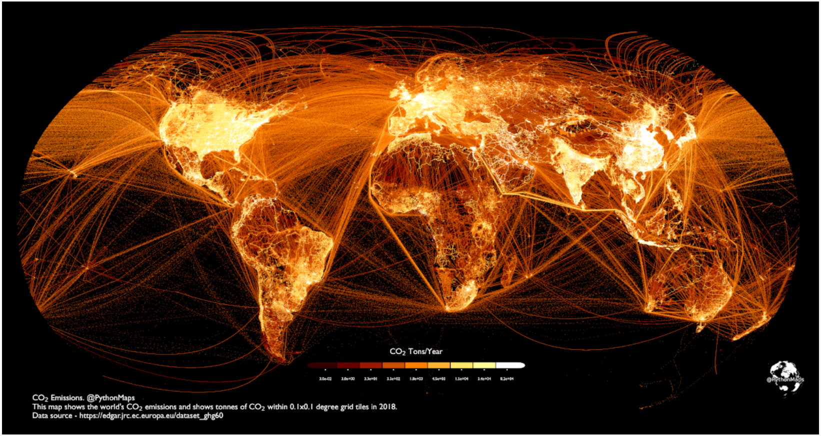
Mitochondria are located in certain areas of cells.



Source: Illinois Science Council

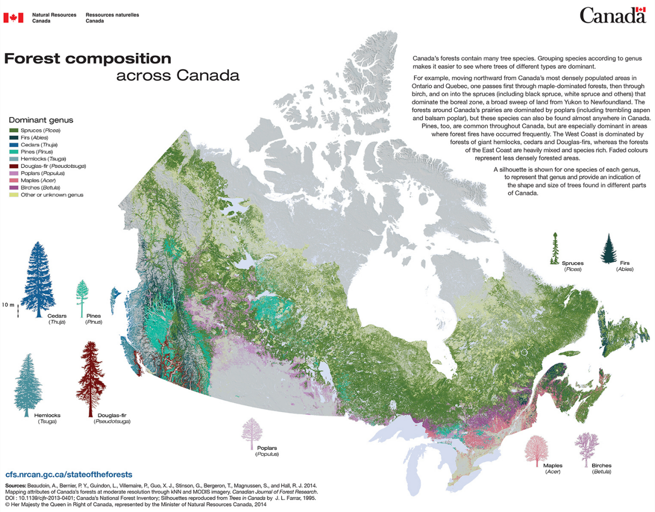
Carbon emissions

Carbon emissions are high in some places, low in others.



Tree species

Trees of the same species are often found in the same areas

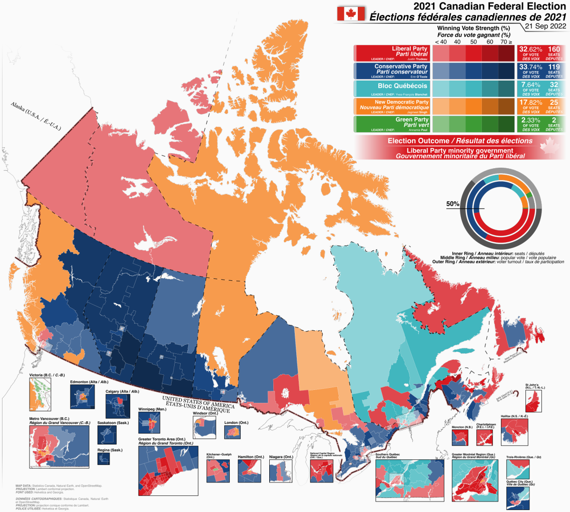


Tree species

... but have non-uniform patterns in their local distribution.

Election results

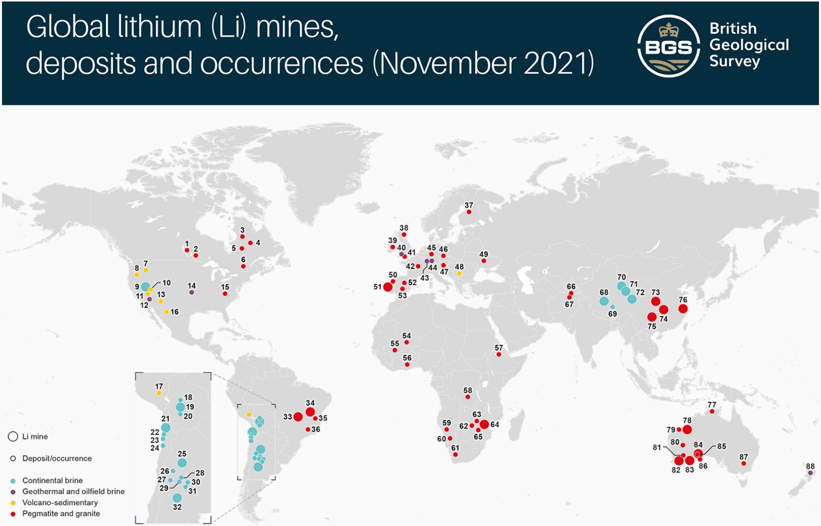
Election results are clustered in space.



Source: Wikipedia

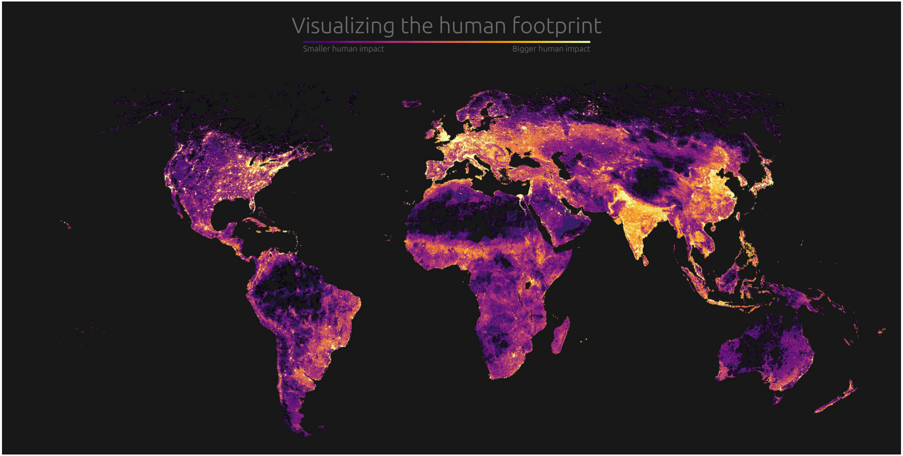
Lithium deposits

Lithium deposits only occur in certain areas.



Data maps

Maps are also engaging, effective ways to display data and convey messages.

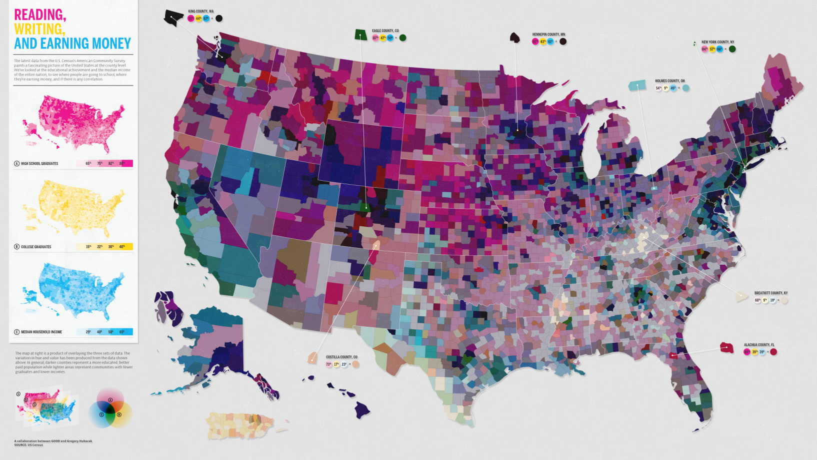


Source: Visual Capitalist

Data maps cont.

...but only if the information contained in them is meaningful

...or if meaningful information is conveyed effectively.



Source: American Community Survey

Spatial data collection

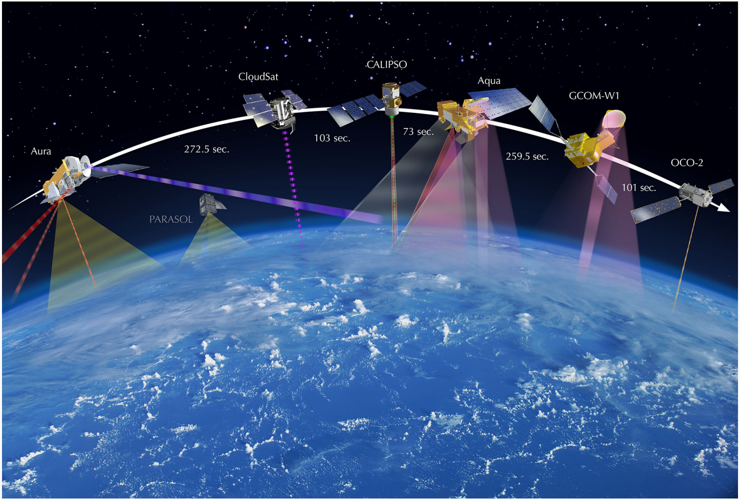
We are also in a 'Golden Age' for collecting and working with spatial data.

Spatial data collection

Historically, surveying, mapping, and spatial data collection were slow, painstaking processes.

Source: 1722 edition of William Leybourn's The Compleat Surveyor.

Satellites can remotely measure environmental characteristics.



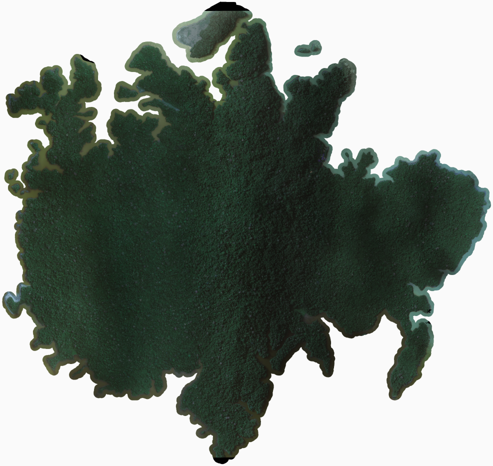
Source: Wikipedia

Camera equipped drones can help us find animals in the wild.

Source: RoboticsBiz.com

Spatial data collection cont.

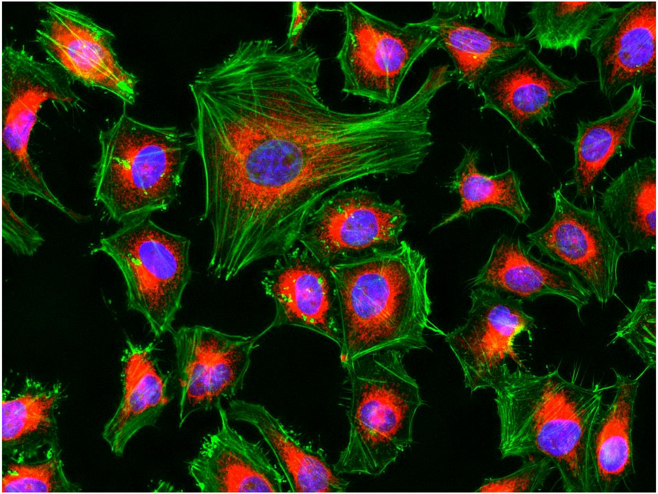
Lidar equipped drones can produce high-resolution maps.



Source: Noonan *et al.* (2021)

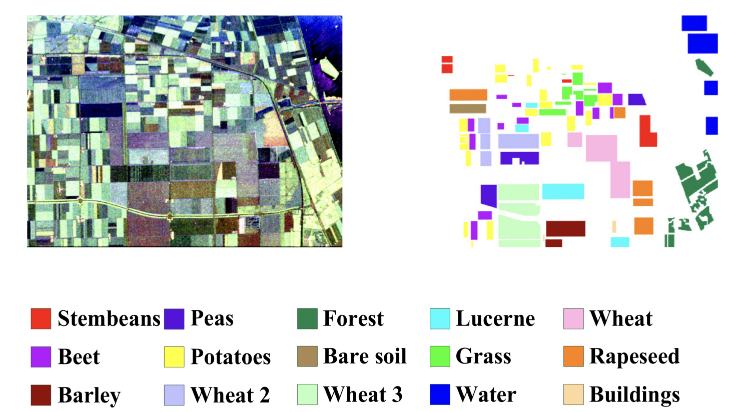
Spatial data collection cont.

Microscopes can localise microscopic cellular structures.



Spatial data processing

Machine learning algorithms can sift through these large volumes of data.



Source: Zhang *et al.* (2017)

Spatial data processing cont.

Cluster computing reduces computing time (important for spatial data).

Source: CERN

Spatial context

The spatial context of data caries important information for understanding how processes operate.

The spatial arrangement of points in a dataset is often a surrogate for unobserved variables (e.g., soil fertility and the locations of trees), or can provide us with information on unrecorded historical events (e.g., cosmological evolution).

Spatial patterns can also influence the outcomes of other processes (e.g., the distribution of schools can drive housing prices).

So how do we study spatial phenomena?

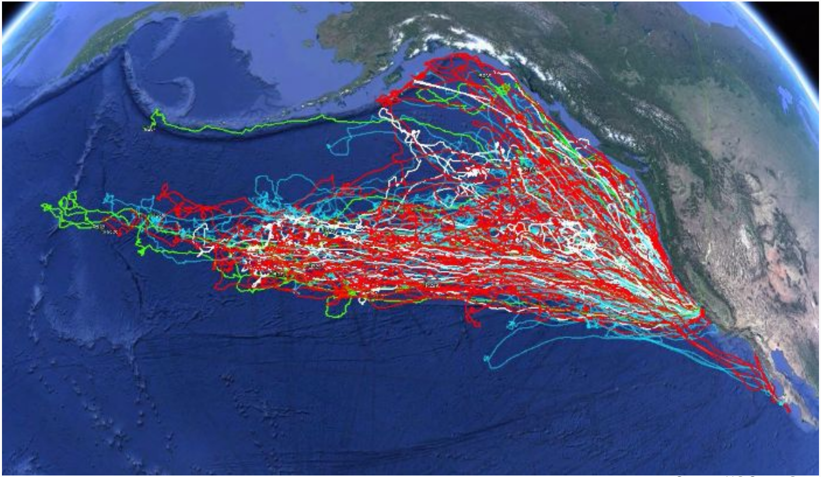
John Snow, cholera deaths

Sometimes plotting the data is enough to learn something new.

Source: Wikipedia

Elephant seal movement

Sometimes plotting the data is enough to learn something new

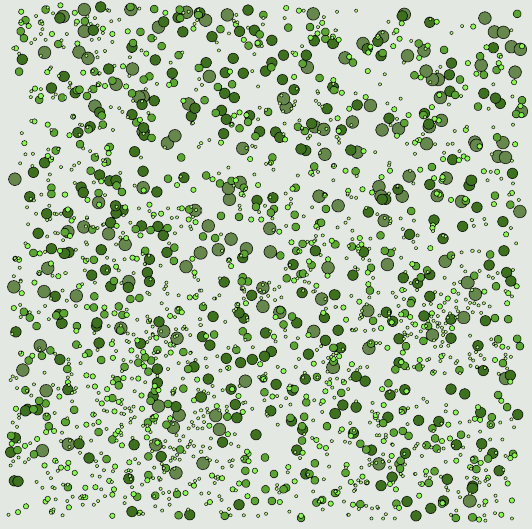


Source: UC Santa Cruz

Trees in southern Washington

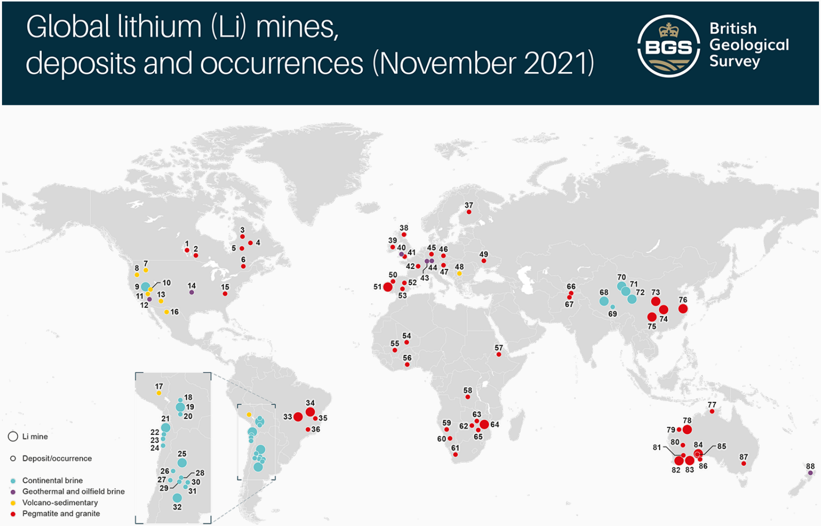
THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

...but not always



Source: Conservation Biology Institute

...and we can't make predictions from simple figures.



Sometimes visualising the data is enough to learn something new... but in most cases the patterns are difficult to see and simple visualisations can be uninformative, or even misleading.

So instead, we rely on

Spatial Statistics

A set of formal techniques for studying processses using their spatial properties.

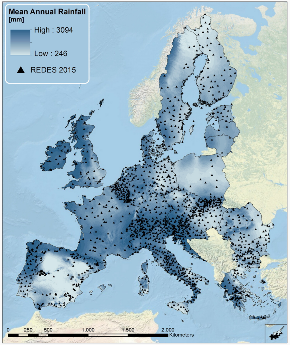
Spatial data

THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

In general, there are two types of spatial data:

Point data where the location conveys information about the process.

Measurements where the sampling locations are artefactual.



Point data

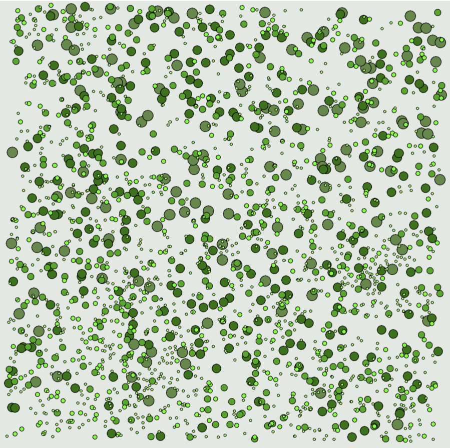
The analysis of point data requires a special set of statistical tools that fall under a general umbrella termed 'point pattern processes'.

This family of methods treats the locations as informative and attempts to identify patterns in the spatial arrangement of the locations, as well as underlying causes for these patterns (usually by relating them to covariates).

The first half of the course will be focused on this family of methods.

Point data cont.

Point data typically look like this, and our goal is usually to describe and understand how/why this pattern emerged.



Source: Conservation Biology Institute

Measurements in space

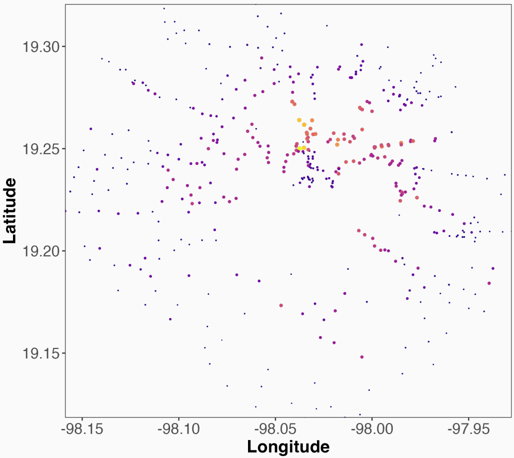
The second type of spatial data are measurements of 'things' taken across space.

This family of methods treats the measurement as informative and attempts to identify patterns in the spatial arrangement of the values by modelling the spatial autocorrelation structure (i.e., points measured close together in space will usually have more similar values than points measured further apart in space).

The second half of the course will be focused on this family of methods.

Spatial measurements cont.

Spatial measurements typically look like this, and our goal is usually to interpolate the values or relate them to covariates.



Source: Fusaro *et al.* (2019)

Take home messages

We use statistics to interpret data and make inference about the world around us.

Spatial statistics is a branch of statistics that is focused on studying processses based on their spatial properties.

Spatial data are very diverse and there is no one-size-fits-all approach for working with them (need to match the tools to the study questions and the properties of the data).

Next lecture we will begin to learn how to work with ‘point’ data.

References

Fusaro, C., Sarria-Guzmán, Y., Chávez-Romero, Y.A., Luna-Guido, M., Muñoz-Arenas, L.C., Dendooven, L., Estrada-Torres, A. & Navarro-Noya, Y.E. (2019). Land use is the main driver of soil organic carbon spatial distribution in a high mountain ecosystem. *PeerJ*, 7, e7897.

Noonan, M.J., Martinez-Garcia, R., Davis, G.H., Crofoot, M.C., Kays, R., Hirsch, B.T., Caillaud, D., Payne, E., Sih, A., Sinn, D.L. *et al.* (2021). Estimating encounter location distributions from animal tracking data. *Methods in Ecology and Evolution*, 12, 1158–1173.

Zhang, Z., Wang, H., Xu, F. & Jin, Y.Q. (2017). Complex-valued convolutional neural network and its application in polarimetric sar image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55, 7177–7188.