Course Introduction

Michael Noonan

DATA 589: Spatial Statistics

Table of contents

1. Course Overview

2. Why Spatial Statistics?

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 Website: https://noonanm.github.io/DATA589/index.html

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 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.

How the course is structured

- For each topic, there will be a core lecture, an associated lab, and a practical assignment.
- Lectures will cover the core concepts of the course. Lecture slides will be posted 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 `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 lab will be covered in the lectures and vice versa.
- In the assignments you will have to apply the methods covered in the lectures and labs to a novel dataset.

Course Evaluation

Practicals (4)	40%	Due on a weekly basis
Participation from practicals	10%	On a weekly basis
Research project	50%	Week 5
Total	100%	

Practicals and participation

Beginning this week, you will be asked to complete practical assignments on a weekly basis. There will be a total of 4 practicals to be completed throughout the course.

The course web page on github will host the practicals, and the various datasets associated with each practical. Lectures will be given twice per week, followed by a lab. After the lab, we will have covered all of the material that is needed to complete the week's practical assignment material which is due before the **start of the following XXXX** lecture (to be submitted via canvas).

Grading: Each practical assignment is worth a total of 12.5% of your total grade. Of this, 2.5% is given for submitting the tutorial on time, irrespective of whether or not the answers are correct (participation). The remaining 10% comes from the answers provided. Late practicals will be accepted, but will only be worth a maximum of 10%.

You will be required to apply the modelling tools covered in the lectures on a dataset of your choosing and write a short report comprised of 6 sections:

- **Introduction:** Provide a brief description of the study system from which the data come and an outline of what questions you intend on exploring with the data. **(12.5%)**
- **Methods:** Describe how the data were collected, what variables are included, and what analyses were applied. **(20%)**
- **Results:** Length: Describe your statistical findings. Tables and figures should be used throughout. **(20%)**
- **Discussion:** Provide a brief summary of your findings and place them in a biological context. **(12.5%)**
- **References:** Include references to all necessary literature and statistical packages employed. **(5%)**
- **Appendix:** The appendix material should include an R markdown document that details every step of the analyses. **(30%)**

Research Project cont.

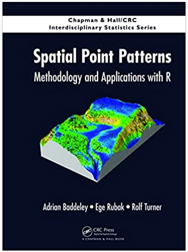
Datasets: To complete these assignments, you will have access to a number of pre-selected datasets. You can opt to use your own data to complete these assignments if you prefer, and are encouraged to do so, but you must seek instructor approval. If you intend on using your own data, it is recommended that you discuss this with me as early as possible.

Late Assignments: You are to submit your paper by the end of the day on XXXXX. Late papers will have 5% deducted per day that they are overdue, and will receive a grade of zero if more than 20 days late without a valid excuse.

Optional Material

There is no textbook for this course, but if you are interested in expanding your knowledge beyond what is covered, 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; Handling spatial data
2	Point processes 1; Point processes 2
3	Point processes 3; Spatial autocorrelation; variograms
4	Kriging; Co-Kriging; Regression-Kriging; Correlated errors

Why Spatial Statistics?

Science and statistics

THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

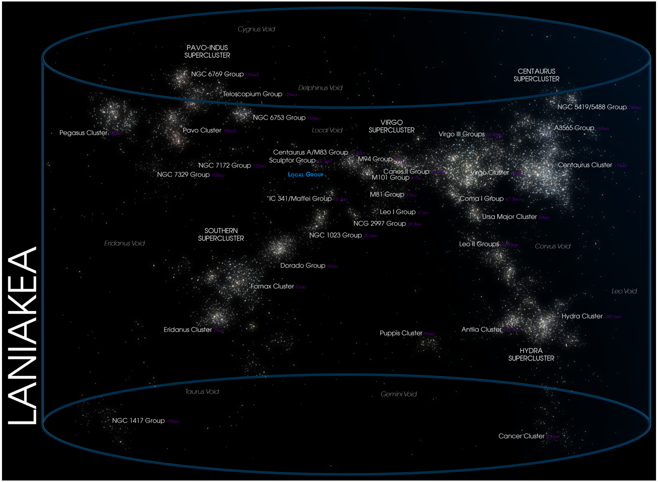
Science is a process of learning about the world around us. As scientists, we weigh competing ideas about how the world works (hypotheses) against observations (data).

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’.

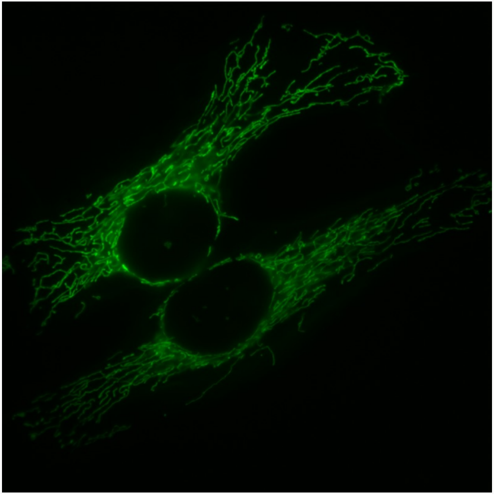
Stars are found in clusters.



Source: Wikipedia

Star clusters

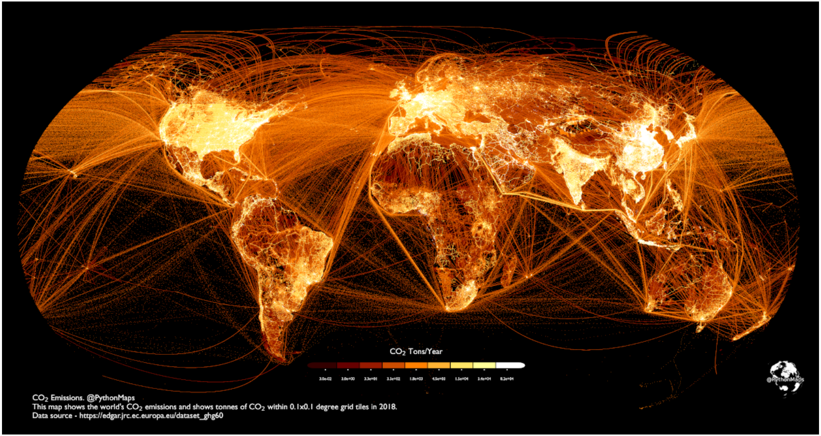
Mitochondria are located in certain areas of cells.



Source: Illinois Science Council

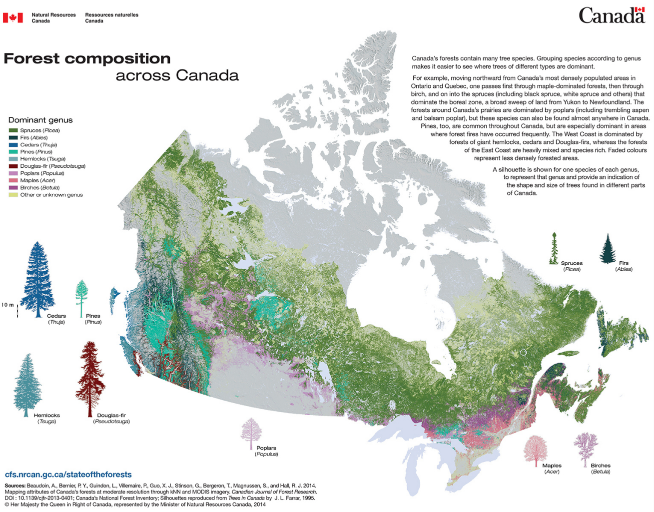
Carbon emissions

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



Tree species

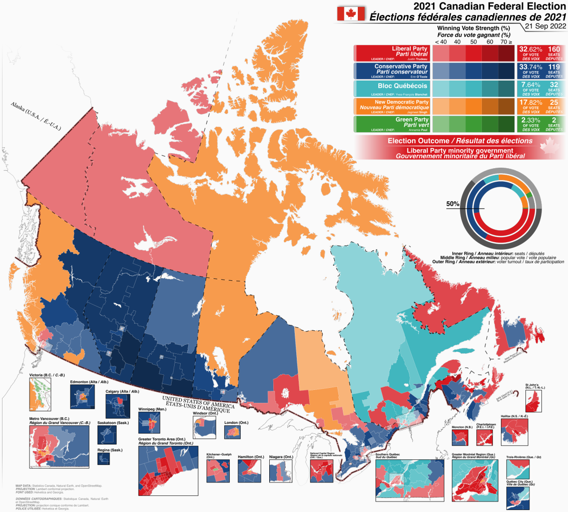
Trees of the same species are often found in the same areas



Tree species

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

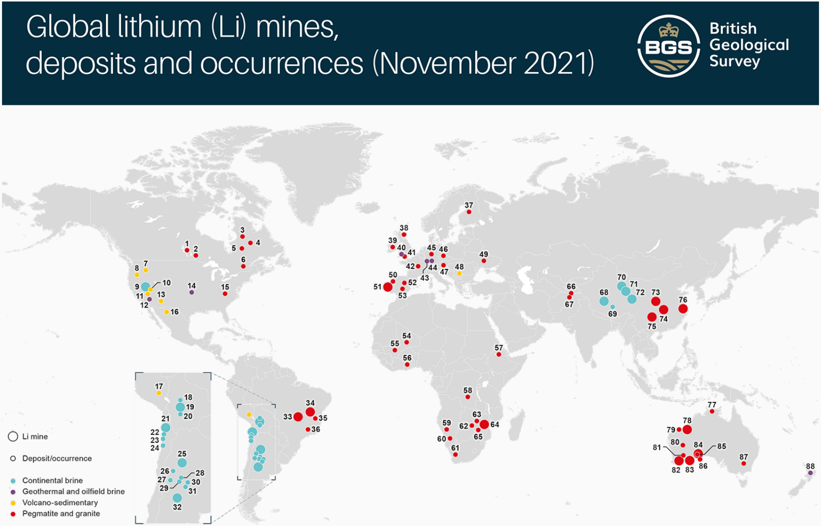
Election results are clustered in space.



Source: Wikipedia

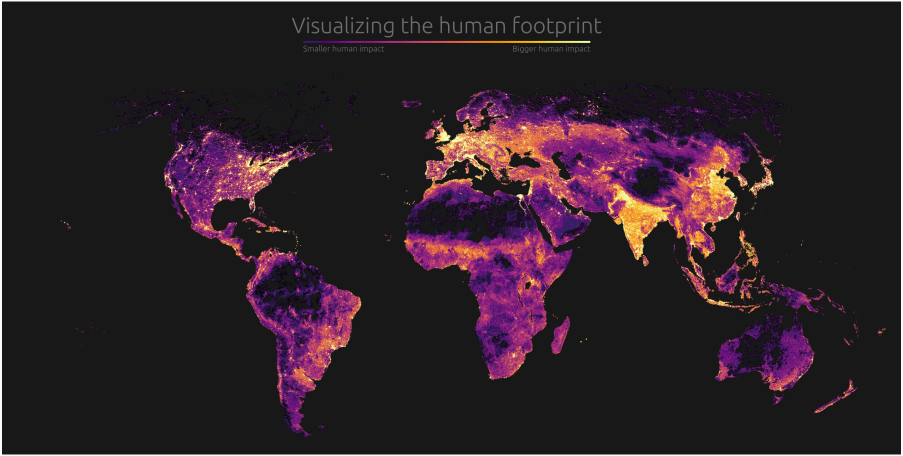
Lithium deposits

Lithium deposits only occur in certain areas.



Data maps

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



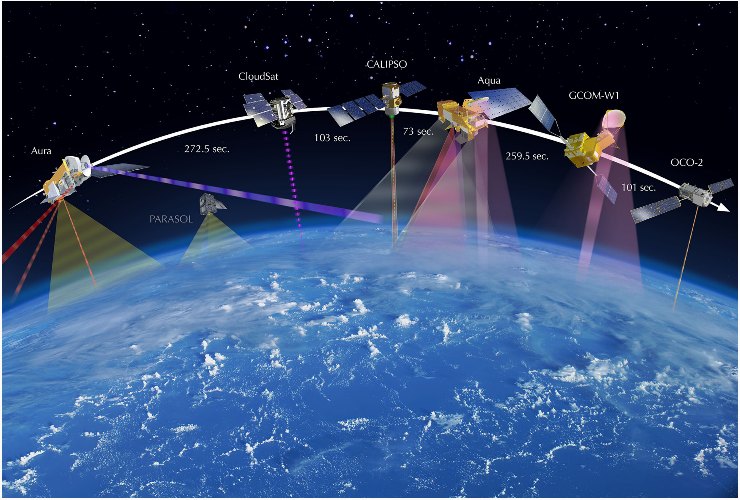
Source: Visual Capitalist

Data maps cont.

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

...and we are in a ‘Golden Age’ for collecting and working with spatial data.

Satellites can remotely measure environmental characteristics.



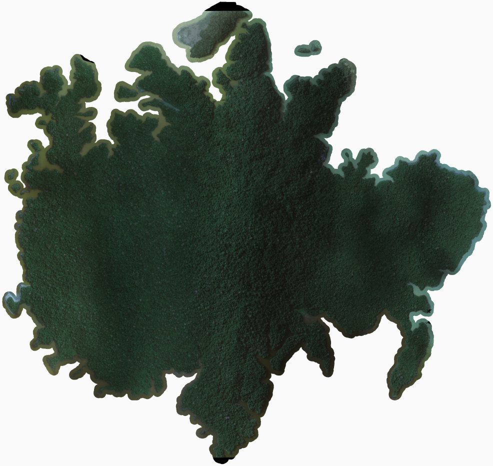
Source: Wikipedia

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

Source: RoboticsBiz.com

Spatial data collection cont.

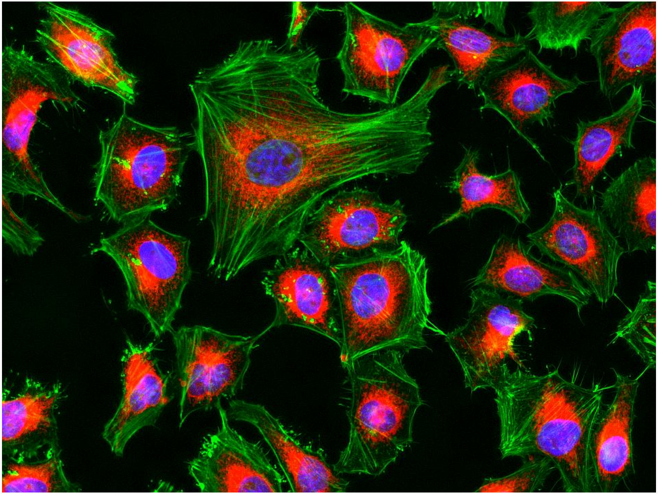
Lidar equipped drones can produce high-resolution maps.



Source: Noonan *et al.* (2021)

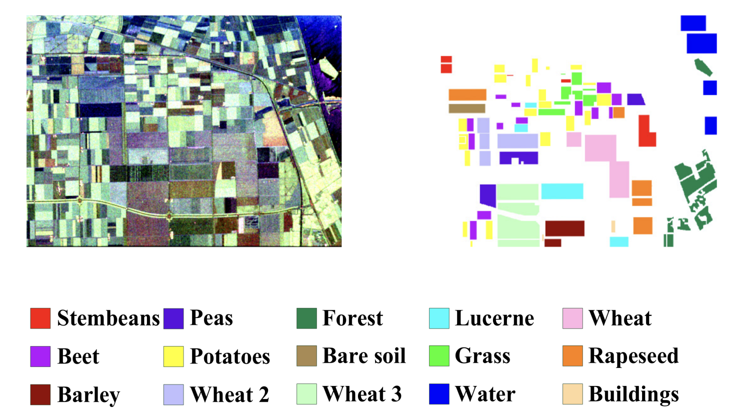
Spatial data collection cont.

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.

THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

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?

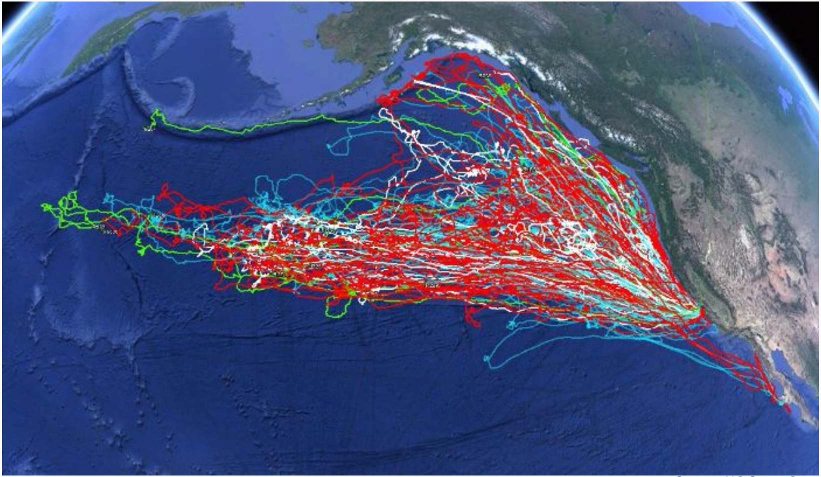
John Snow, cholera deaths

Sometimes plotting the data is enough to learn something new.

Source: Wikipedia

Elephant seal movement

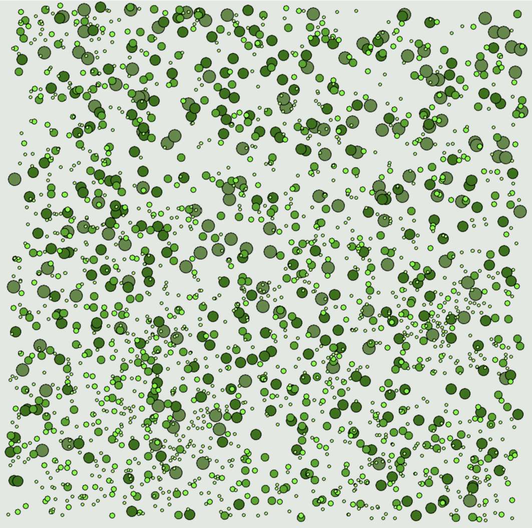
Sometimes plotting the data is enough to learn something new



Source: UC Santa Cruz

Trees in southern Washington

...but not always

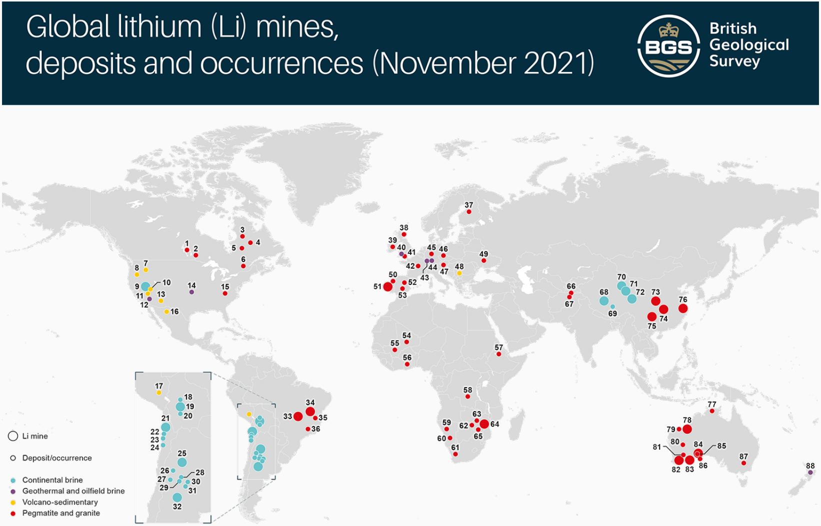


Source: Conservation Biology Institute

Predicting lithium deposits

THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

...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

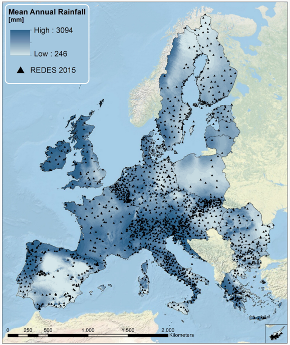
A set of formal techniques for studying processses using their spatial properties.

Spatial data

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

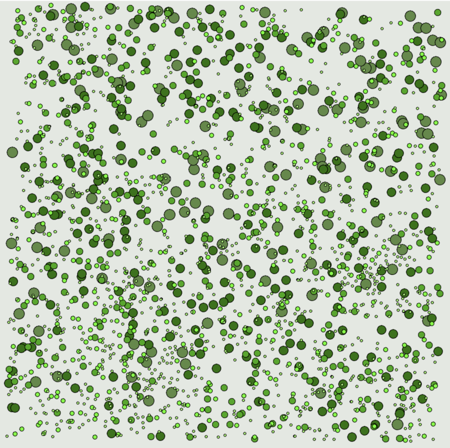
The analysis of point data requires a special set of statistical tools that fall under a general umbrella termed 'point pattern processes'.

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

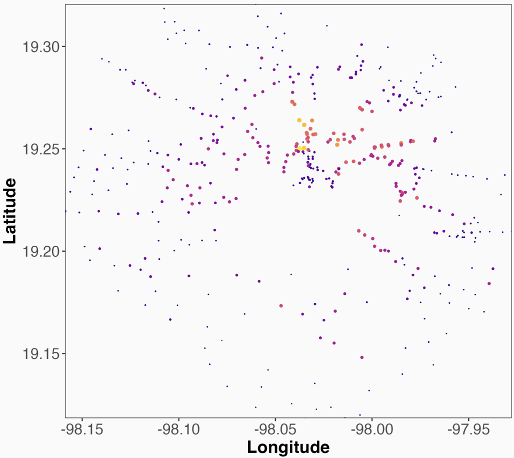
The second type of spatial data are measurements of 'things' taken across space.

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 focus on how to work with, and visualise spatial 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.