

# ETC5512: Wild Caught Data

Week 6

## Modelling Great Barrier Reef data

Lecturer: *Patricia Menendez*

Department of Econometrics and Business Statistics

✉ ETC5512.Clayton-x@monash.edu

29th April 2020



An aerial photograph of the Great Barrier Reef, showing the intricate patterns of the coral reefs in shallow, turquoise waters. The reef extends from the foreground towards the horizon under a bright blue sky with scattered white clouds.

*Great Barrier Reef  
Marine Park Authority*

# Two monitoring programs in the GBR

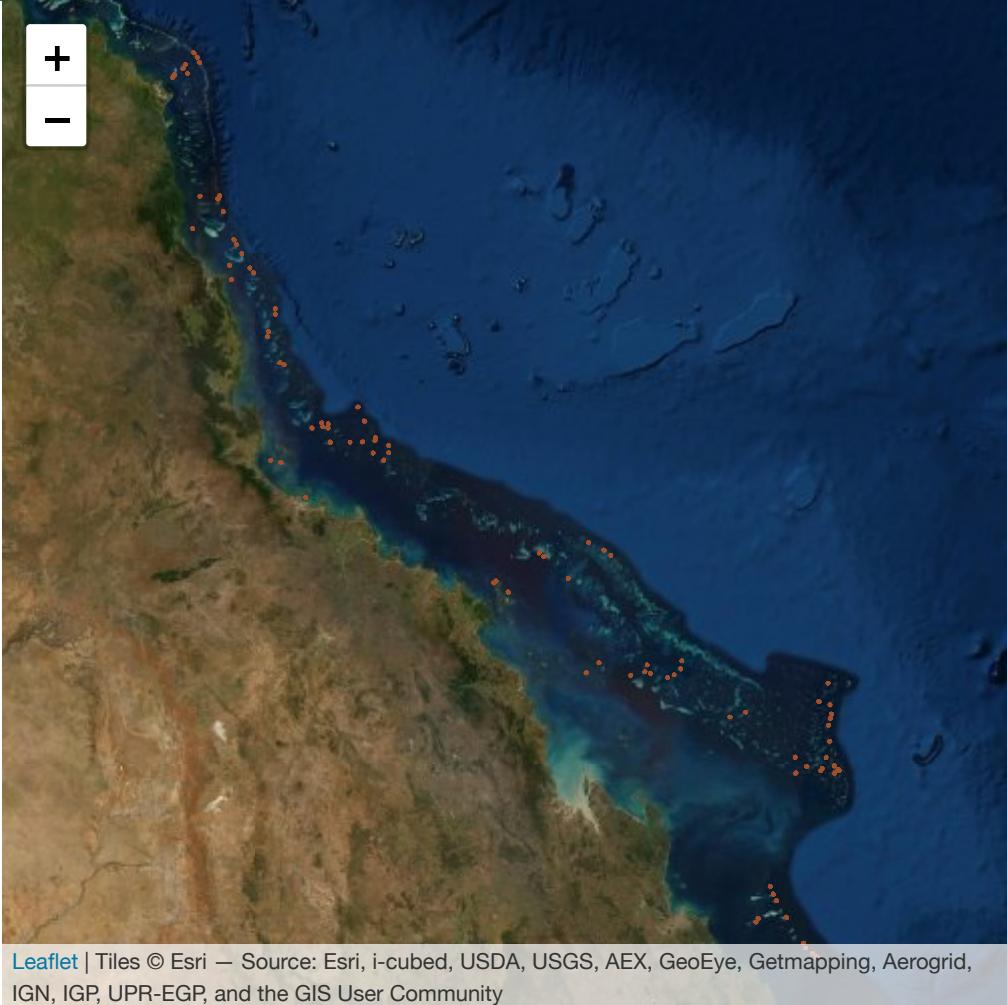


LTMP: For over 30 years, AIMS has been surveying the health of 47 midshore and offshore reefs across the Great Barrier Reef region. The Long-term Monitoring Program represents the longest continuous record of change in reef communities.



MMP: Inshore reefs (those that can be reached from shore by a small boat) are vulnerable to more threats than those further from shore. 32 inshore reefs are monitored under the Great Barrier Reef Marine Park Authority's Marine Monitoring Program.

# Learnt to use leaflet for mapping



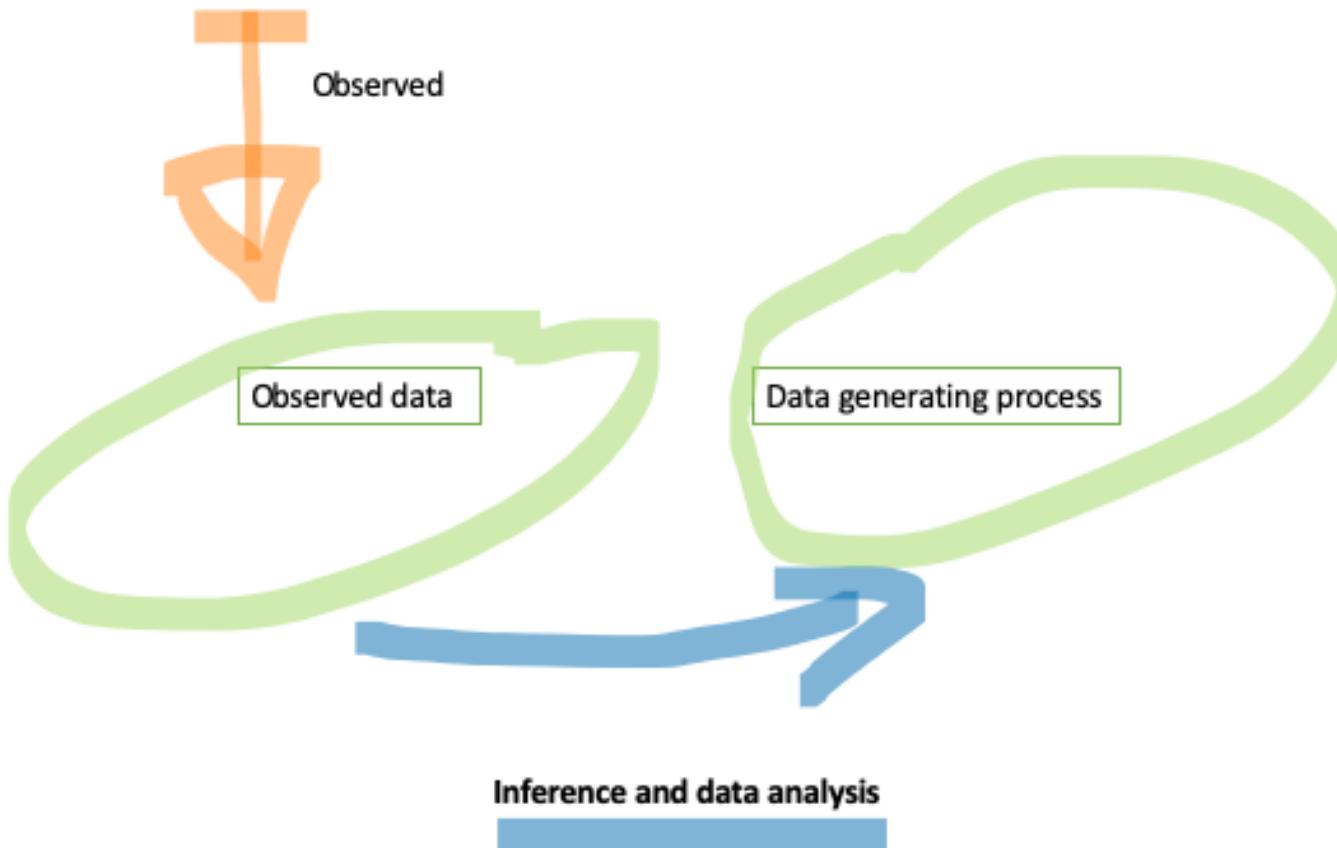
Leaflet | Tiles © Esri — Source: Esri, i-cubed, USDA, USGS, AEX, GeoEye, Getmapping, Aerogrid, IGN, IGP, UPR-EGP, and the GIS User Community

# Learnt to create maps using shape files

# This week

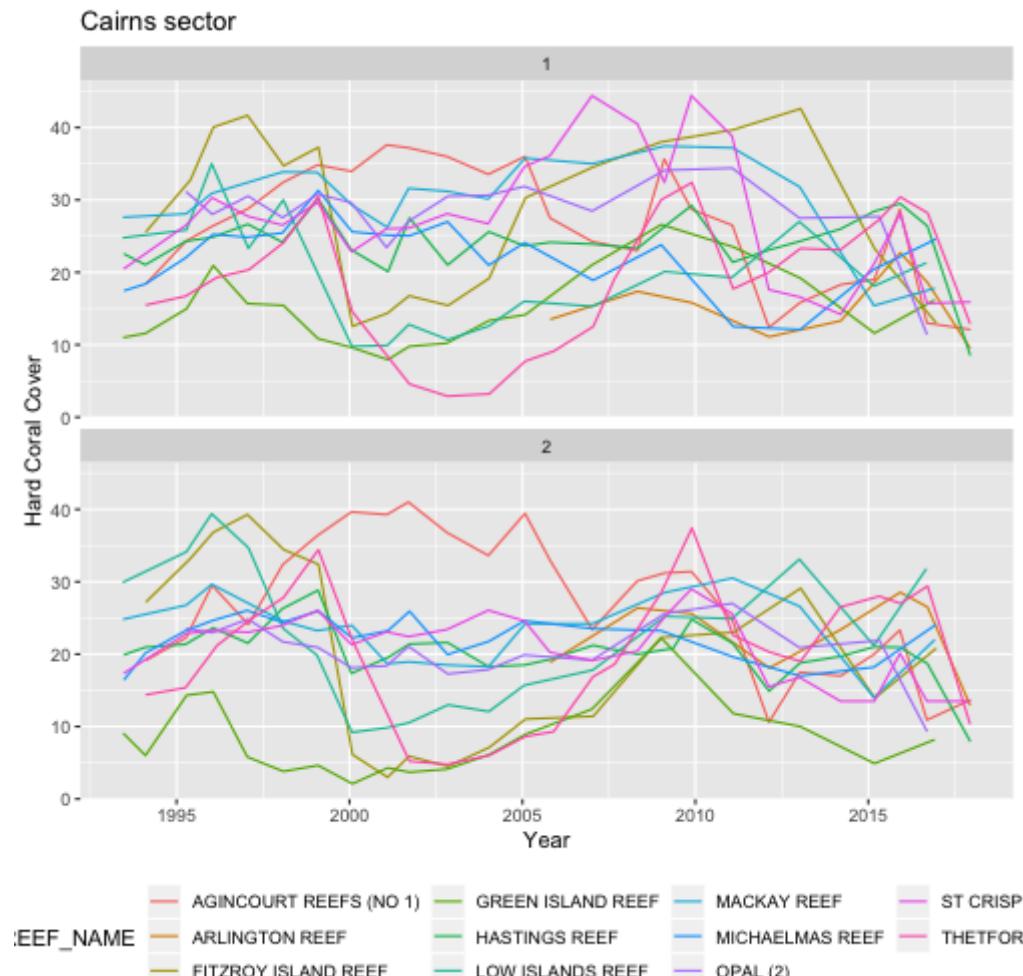
- 📊 Simple inference
- 📊 Understand variability
- 📊 Visualization
- 📊 Linear regression

# Data and inference



# Time series: measurements over time

## Hard coral cover for Cairns sector for sites 1 and 2



# What is a time series?

 Observations taken over time.

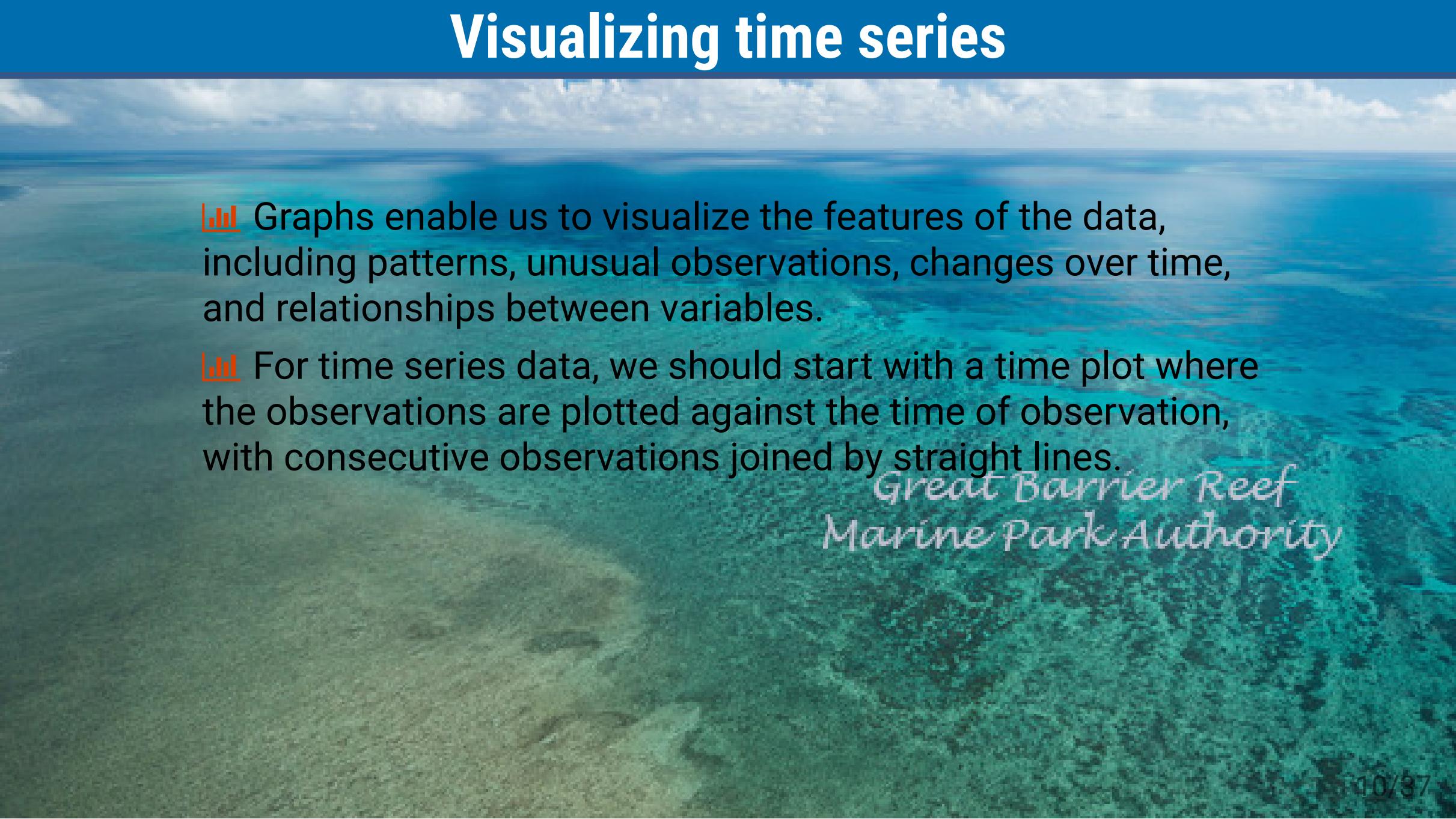
 More formally:

**i**

A time series is a set of time-ordered observations measuring a quantitative characteristic of a process taken at consecutive periods or points of timeline (equally or unequally spaced).

 Typically in this kind of data there is a time dependency.  
**Why? What does that mean?**

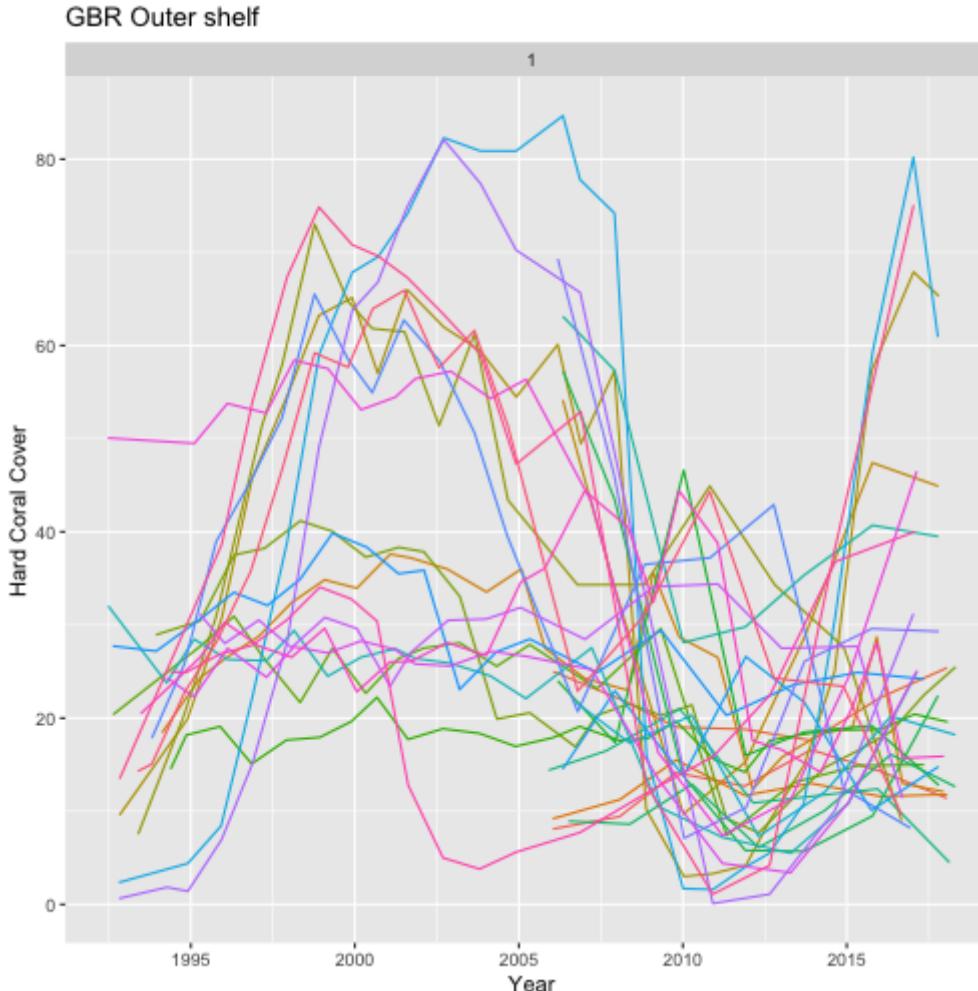
# Visualizing time series

- 
- An aerial photograph of the Great Barrier Reef, showing a vast expanse of turquoise and teal waters with intricate patterns of coral reefs and sandbars stretching towards the horizon under a bright blue sky.
- Graphs enable us to visualize the features of the data, including patterns, unusual observations, changes over time, and relationships between variables.
  - For time series data, we should start with a time plot where the observations are plotted against the time of observation, with consecutive observations joined by straight lines.

*Great Barrier Reef  
Marine Park Authority*

# Choosing the right plot is crucial!

Question: Is this an informative figure?



# Time series patterns



Trend: A trend exists when there is a long-term increase or decrease in the data. It does not have to be linear. Sometimes we refer to a trend as “changing direction”, when it might go from an increasing trend to a decreasing trend. [Hyndman and Athanasopoulos, 2016](#)

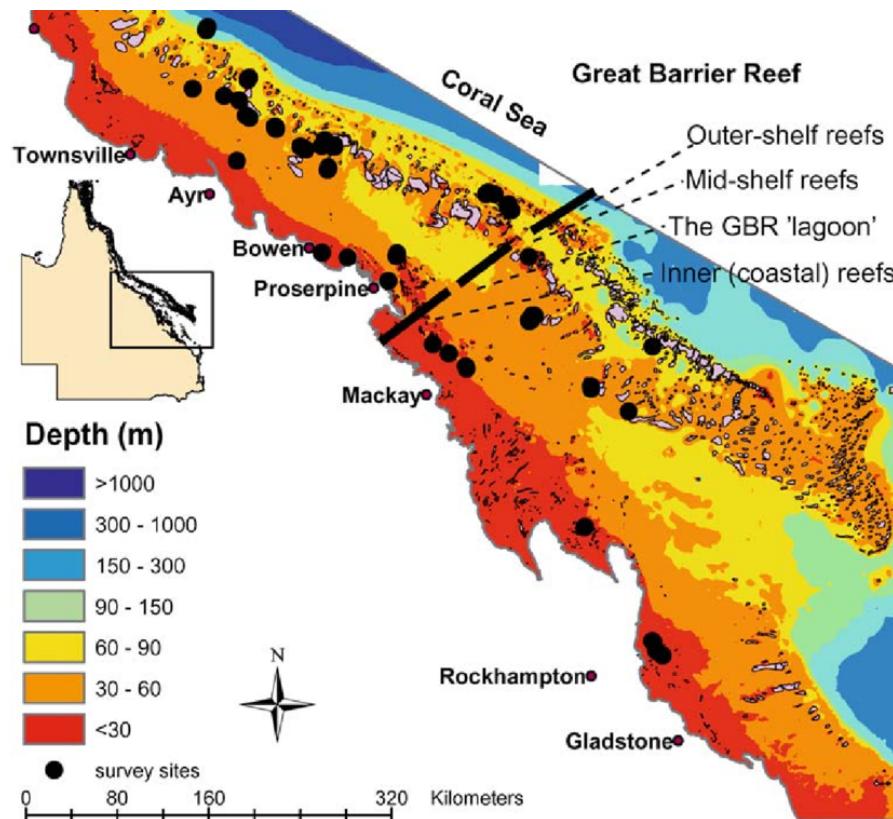


There are other patterns in time series such as seasonal and cycles.

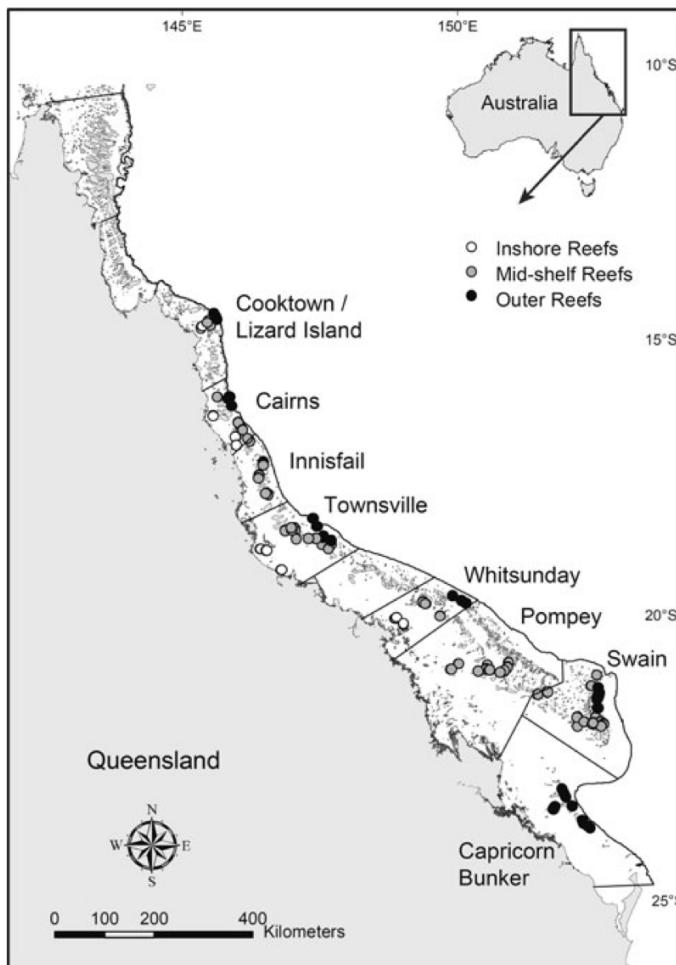


In this lecture we are focusing on trends.

# Positions in the shelf



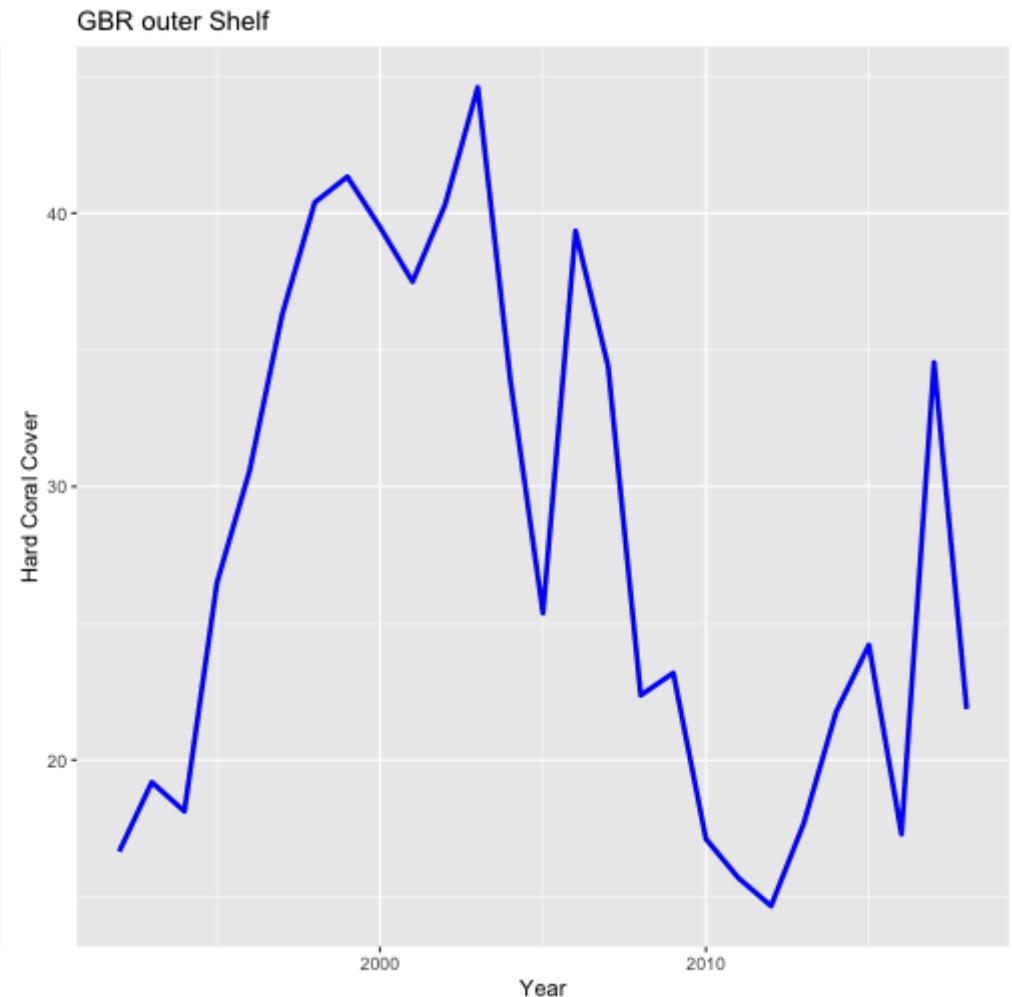
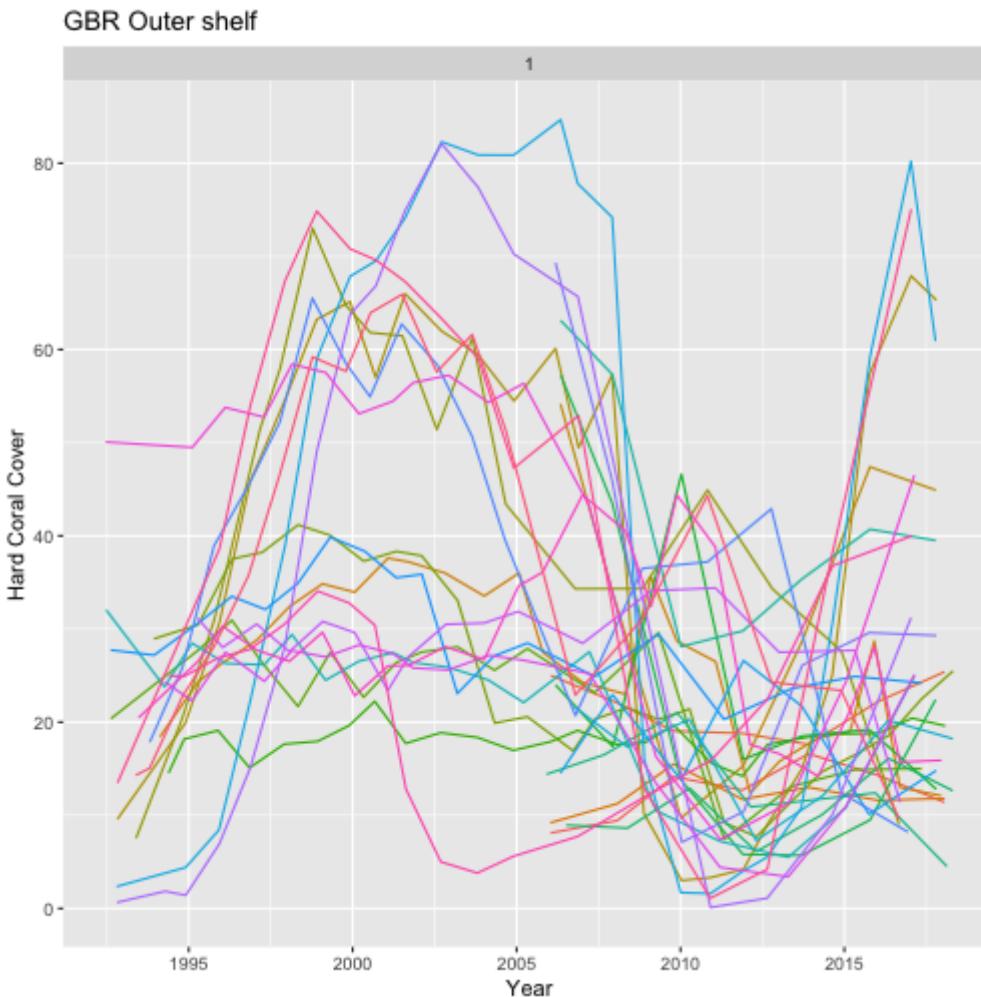
# Positions in the sectors



**Fig. 1** Map of the Great Barrier Reef showing survey reefs (circles) coded by shelf position (unfilled symbols inshore, shaded mid-shelf, filled outer). Sectors are demarcated by solid lines

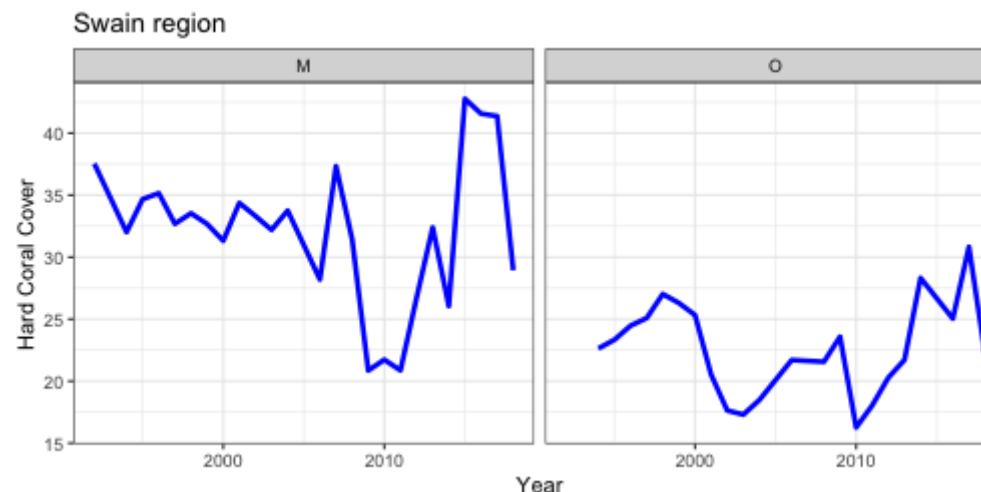
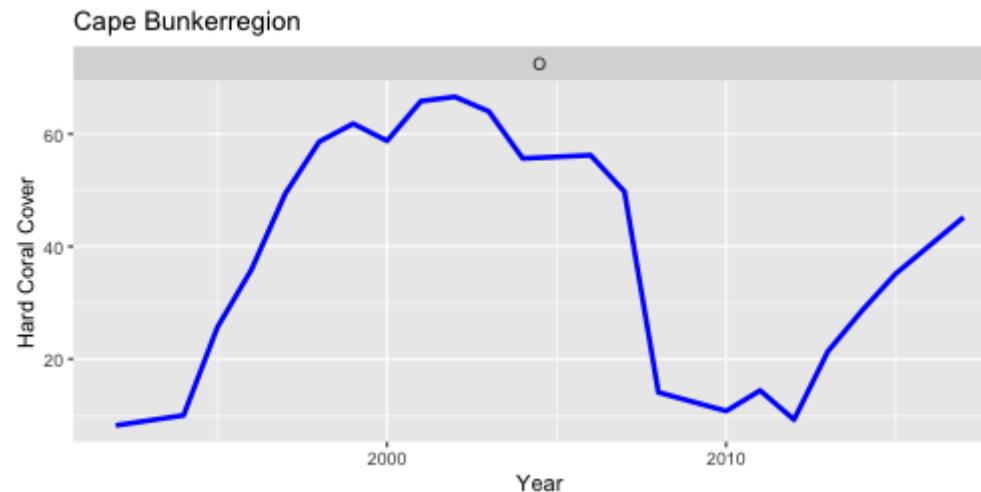
# Summarizing data

Hard coral cover condition in the outer shelf:

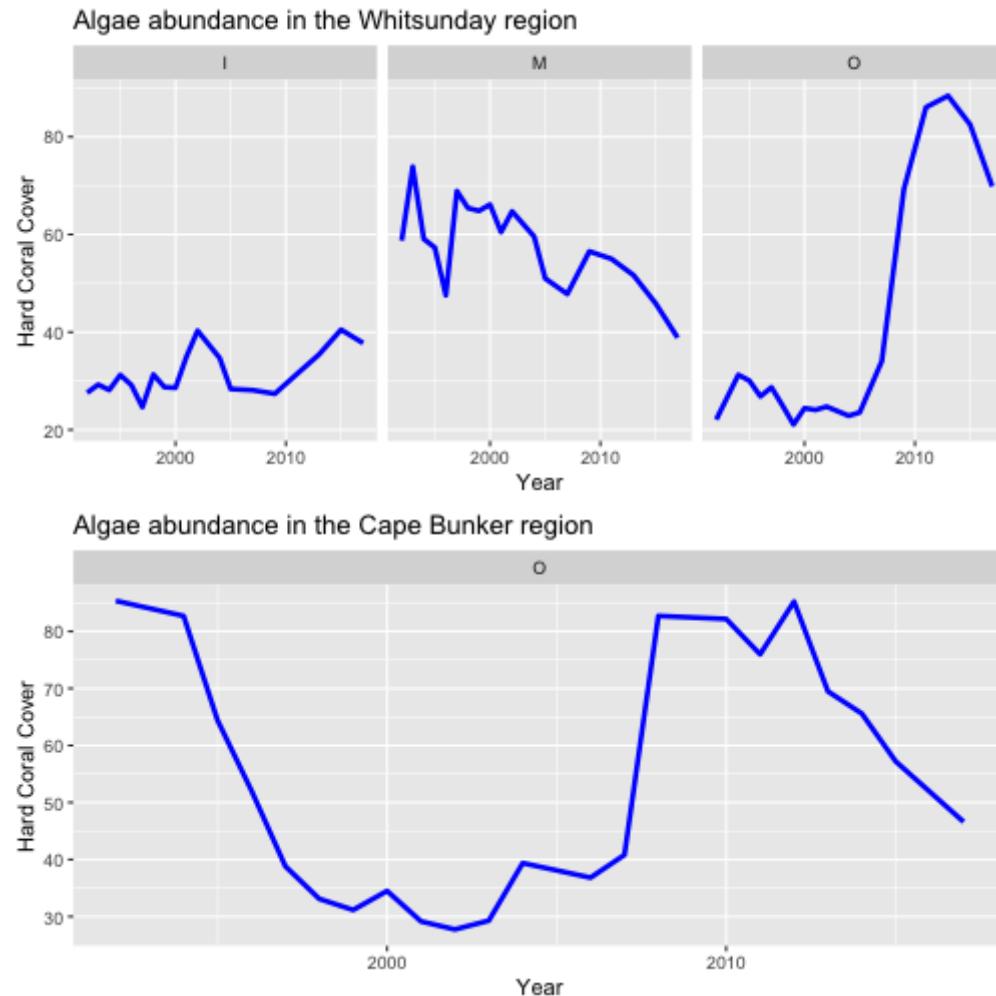


# Hard coral cover condition in the CB and WH regions

GBR condition:



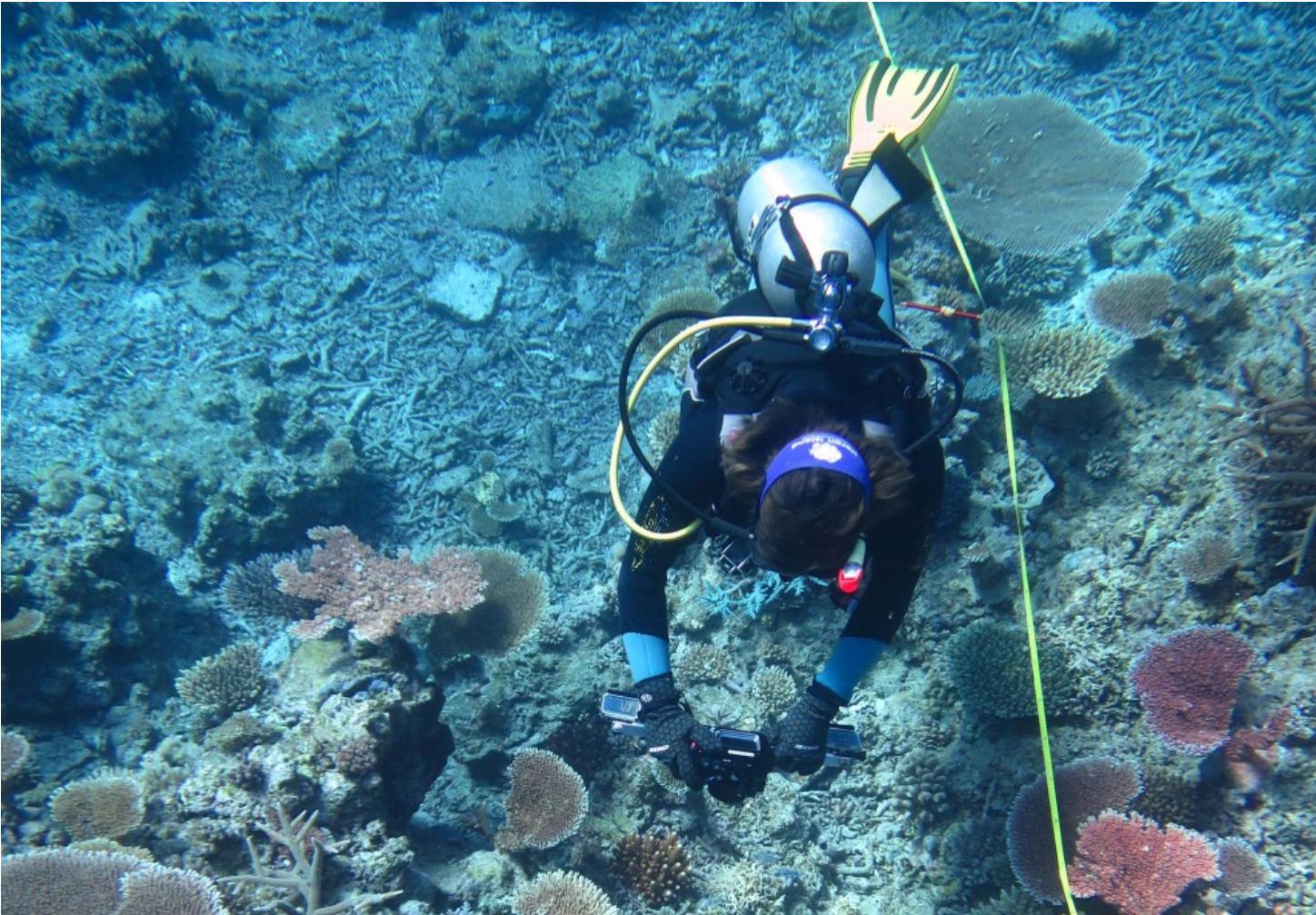
# Algae condition in the Swain and Whitsunday regions



# Focus: research question

- 📊 For your data analysis --> it is essential to have a research question or questions.
- 📊 A research question will enable you to focus on the right data and context.
- 📊 Most likely you will need to do a fair bit of data wrangling --> which is essential!
- 📊 Visualize the data first!

# Example: LTMP reports



LTMP reports

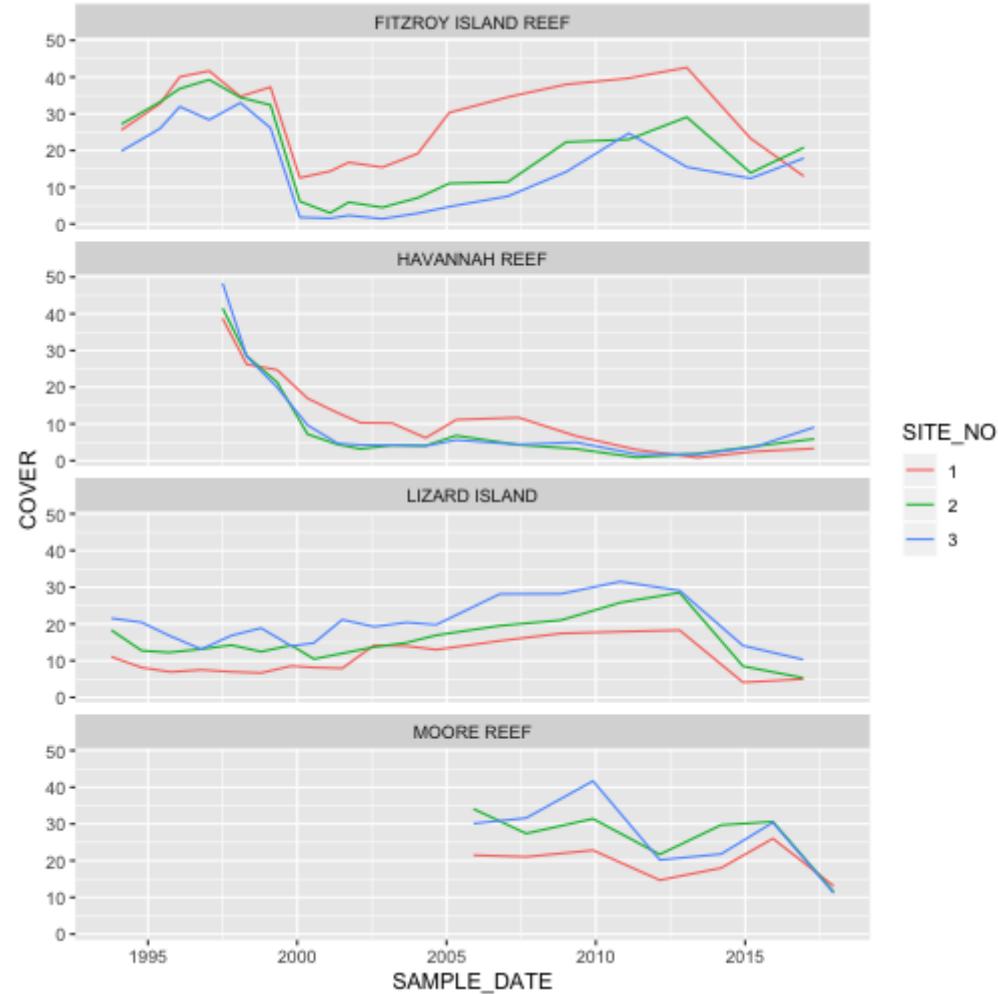
# Let's talk about sample means

Are means enough to explain the GBR condition the GBR?

- 📊 Sample mean --> summary statistic --> statistical inference.
- 📊 Do we have any information about the data variability by just studying the mean?

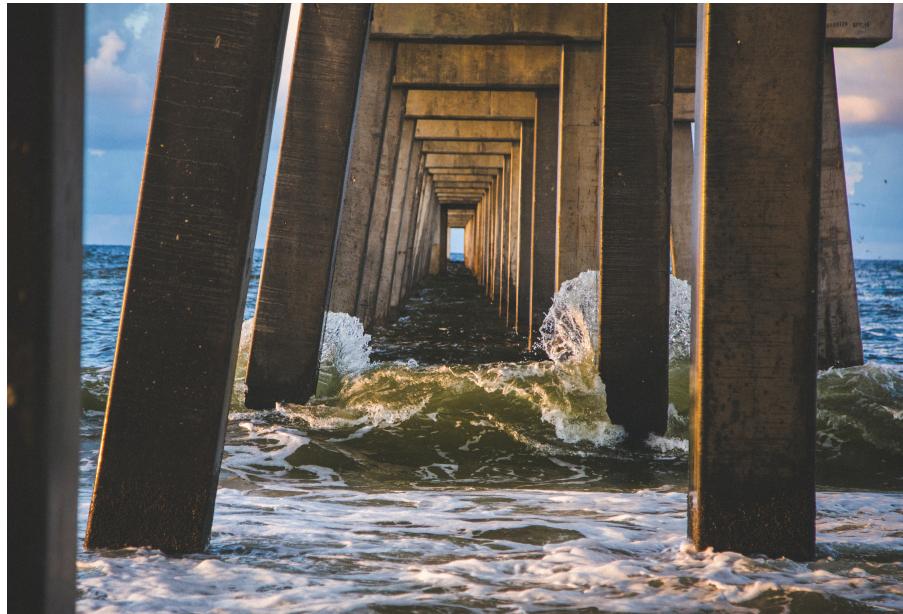
# Time series measurements: replicates

Replicates?



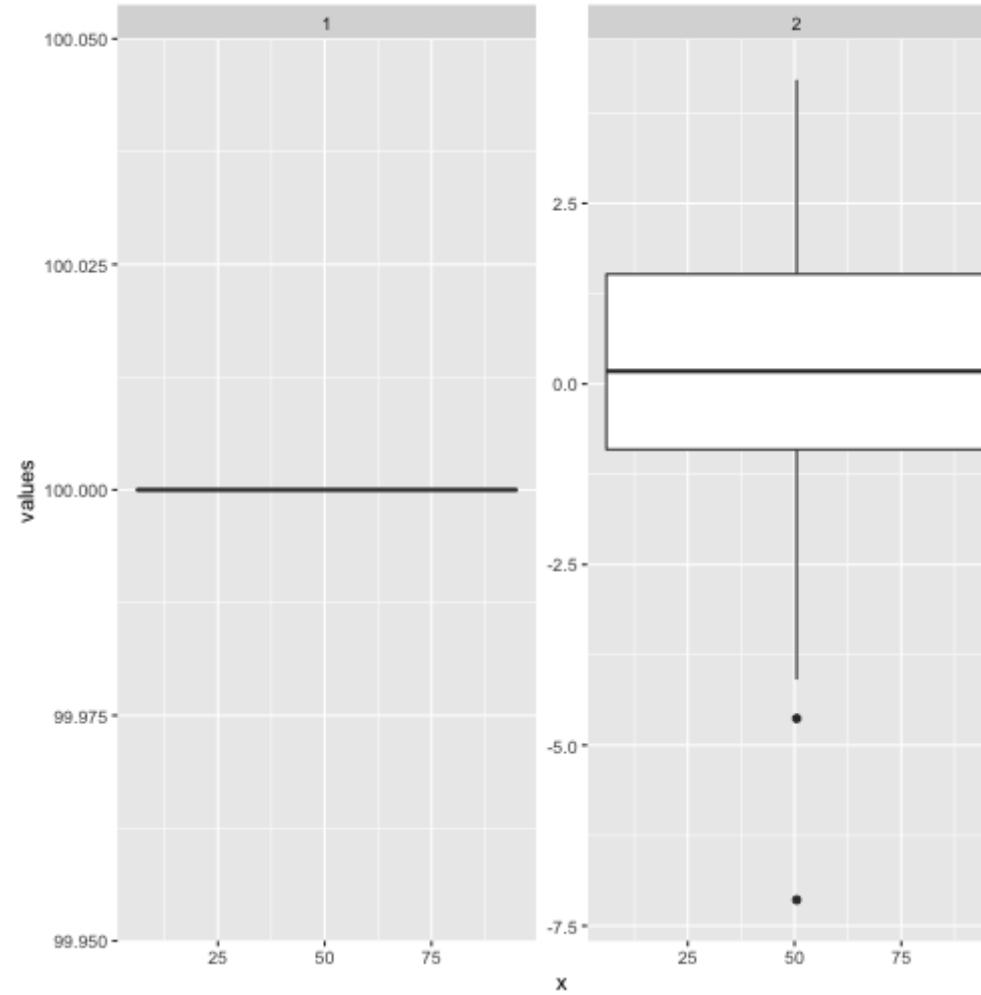
# Why do we need replicates?

📊 Replication is the repetition of an experimental condition so that the variability associated with the phenomenon can be estimated.



# Which information can we obtain from the replicates?

## Variability! BOX PLOT figure



# Variability intuition

📊 **Sample** of size 3  $x_1, x_2, x_3 \rightarrow$  sample mean  $\frac{x_1+x_2+x_3}{3}$

📊  $x_1 = 1, x_2 = 1, x_3 = 1 \rightarrow$  sample mean  $\frac{x_1+x_2+x_3}{3} = 1$

📊 **Variance** → measure of variability in my sample → 0

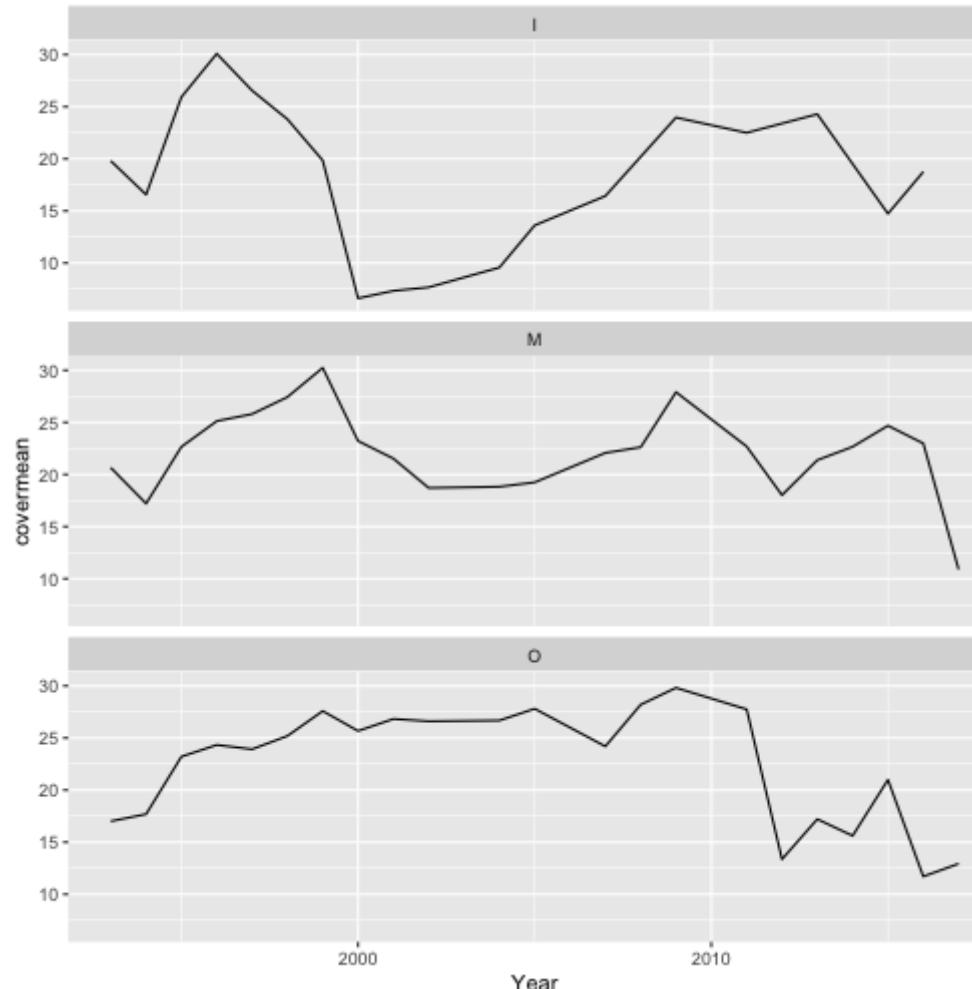
📊  $x_1 = 1, x_2 = 2, x_3 = 3 \rightarrow$  sample mean  $\frac{x_1+x_2+x_3}{3} = 2$

📊 **Variance** → measure of variability in my sample → 1 (`var()`)

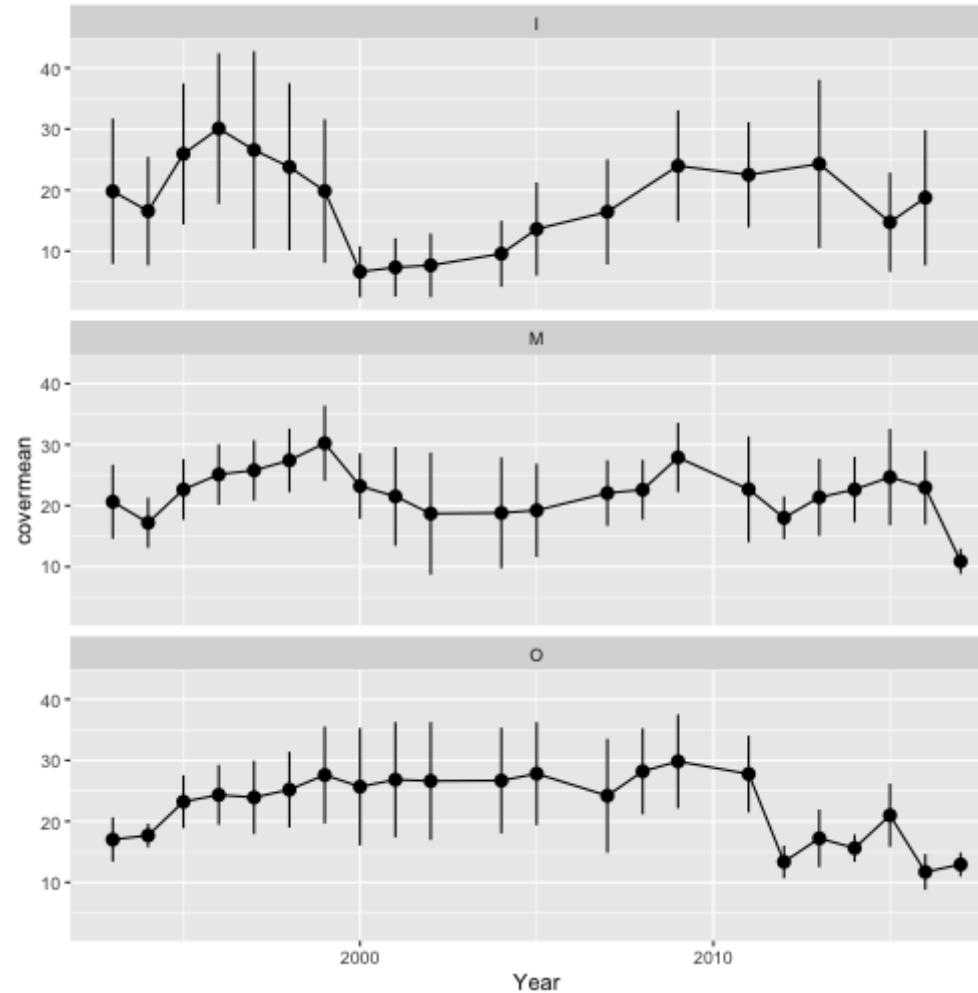
📊 The **standard deviation** (often called `sd` --> `sd()`) is the square root of the variance and measures of variability reported in the same units as our data.

# Overall mean trend for the CA sector

Considering 3 sites, here we display the overall mean trend together with a measure of variability.

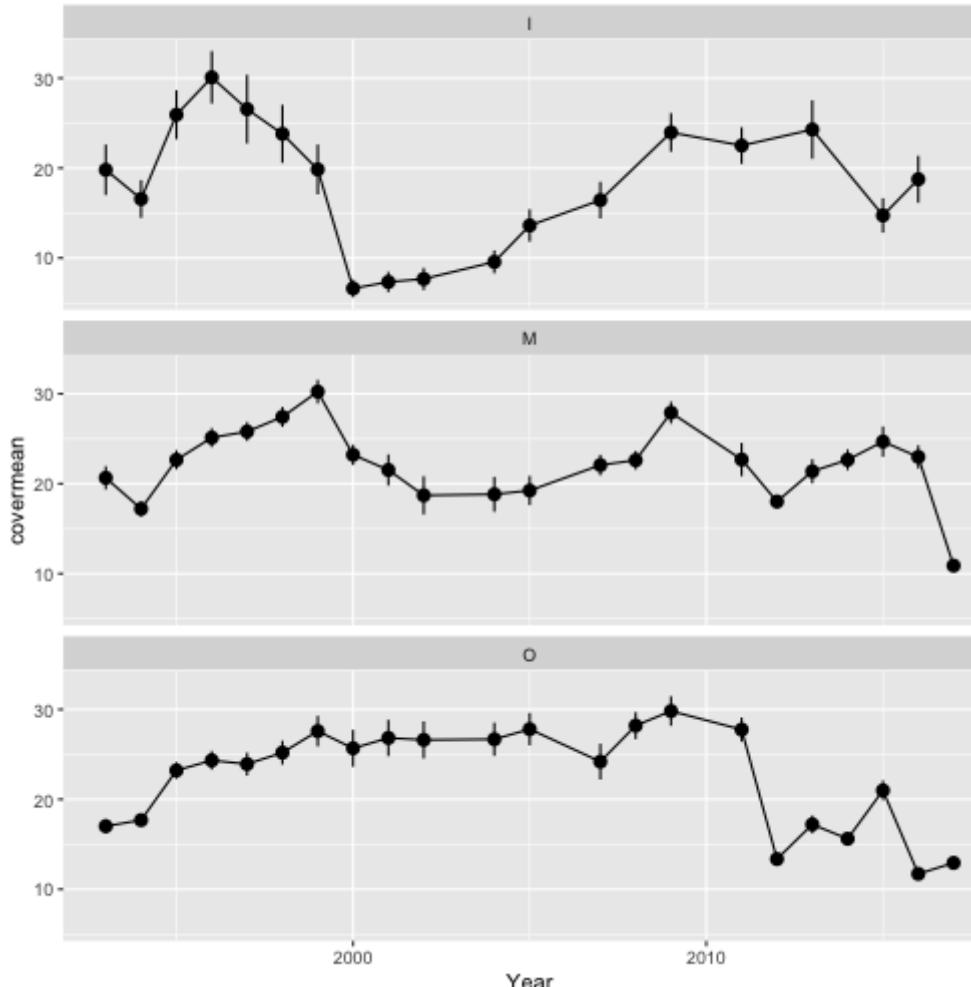


# Overall trend for CA sector with uncertainty

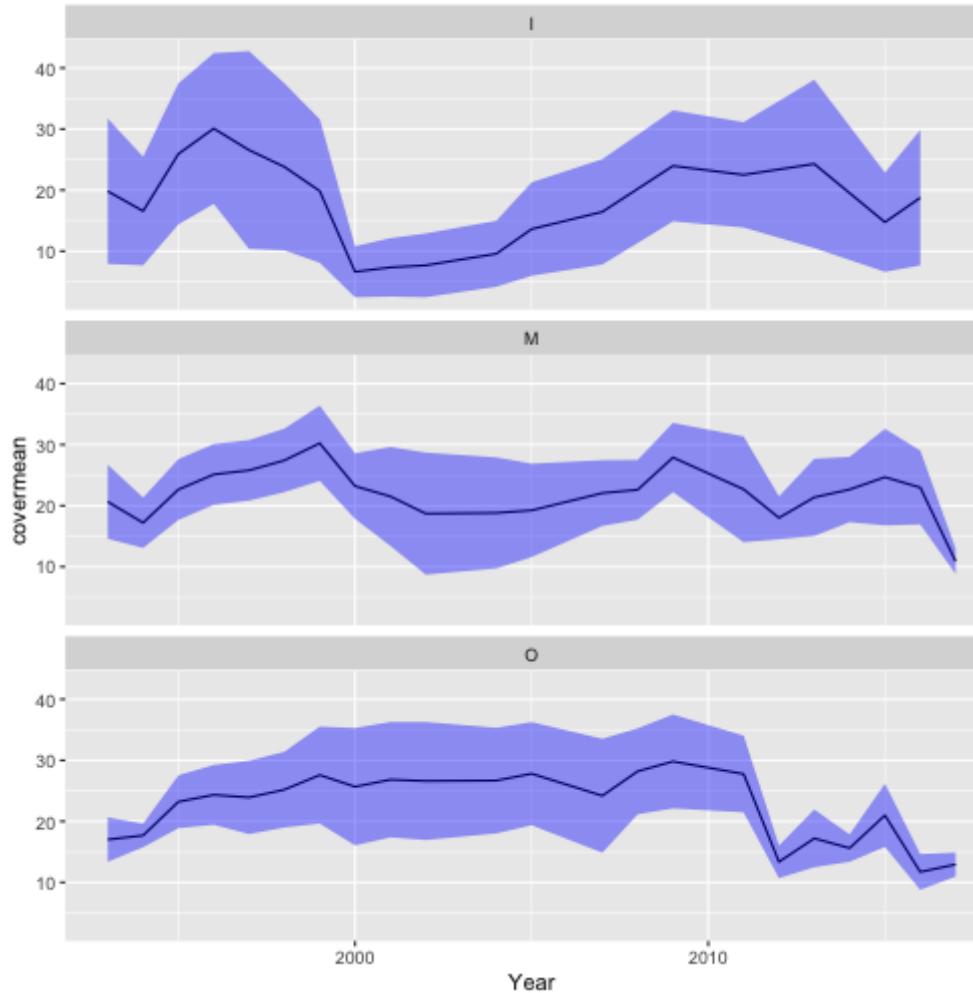


# Is there anything missing in the calculations?

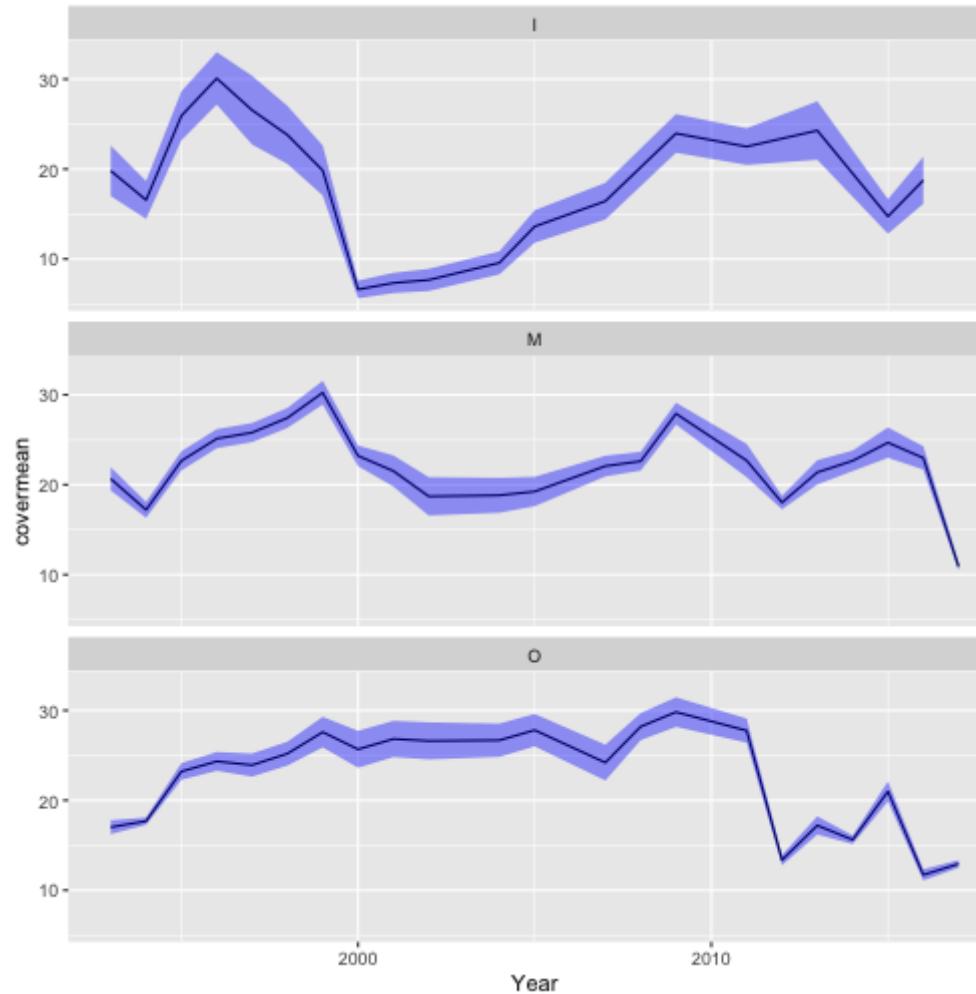
Standard errors  $\rightarrow sd/\sqrt{n}$  with  $n$  being the number of samples used to calculate the mean.



# Overall trend for CA sector with variability

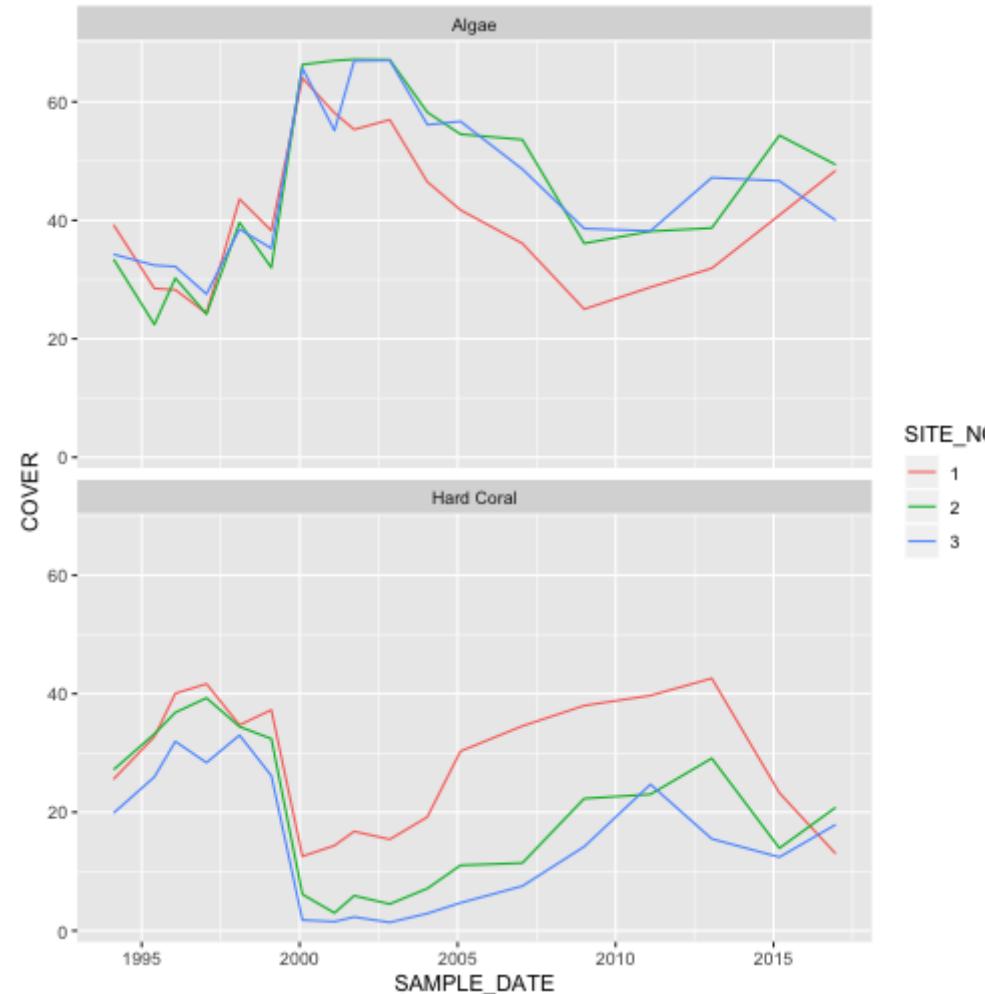


# Overall trend for CA sector: variability standard errors



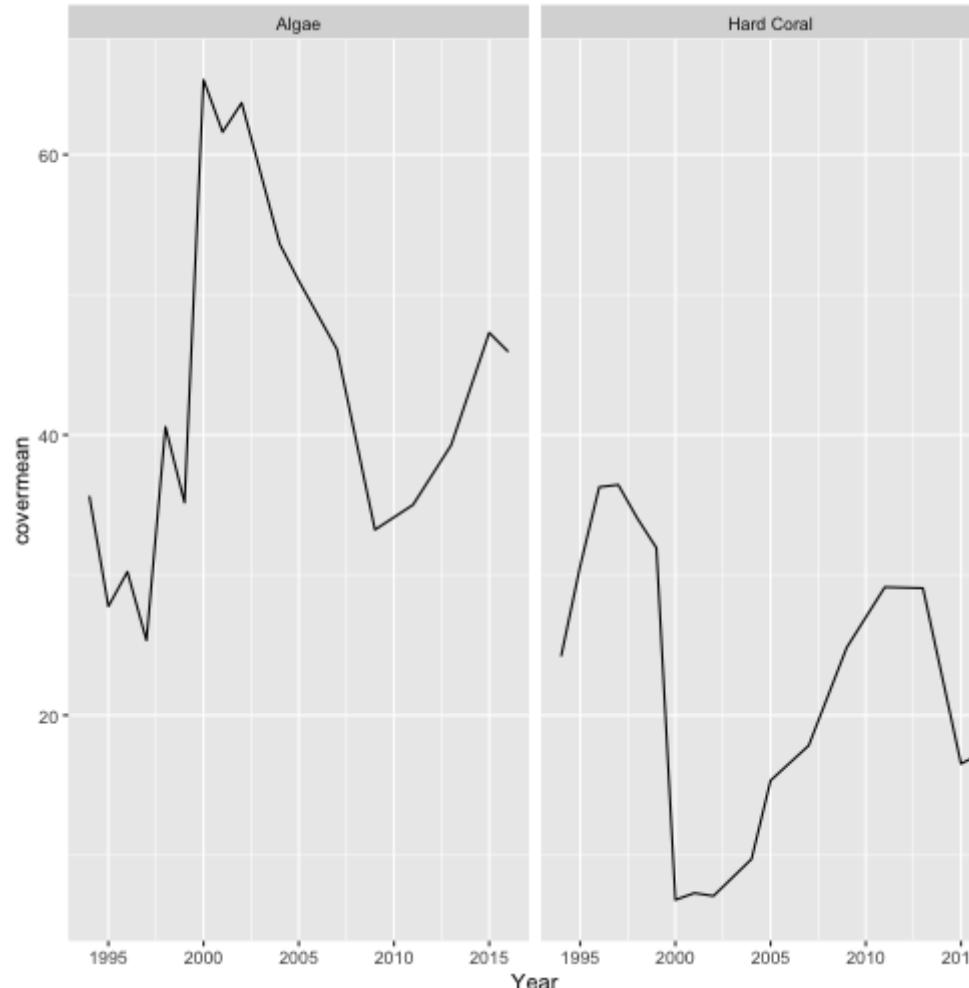
# Remember: Using the right visualizations

## Example: What happened in Fitzroy island?



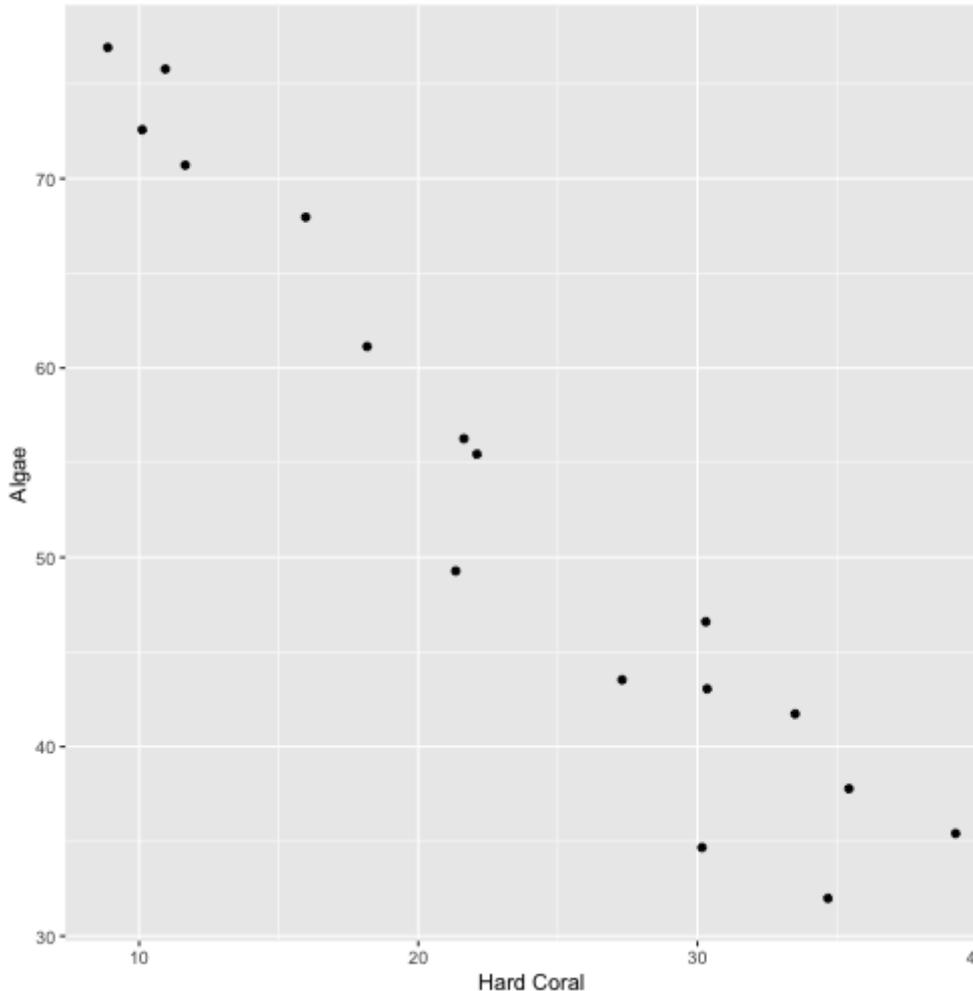
# Modelling: relationship between different variables

Here we study the relationship between algae and hard coral cover yearly means. Comments?



# Modelling: relationship between algae and hard coral cover

What can we say about this figure? Is the relation clearer here?



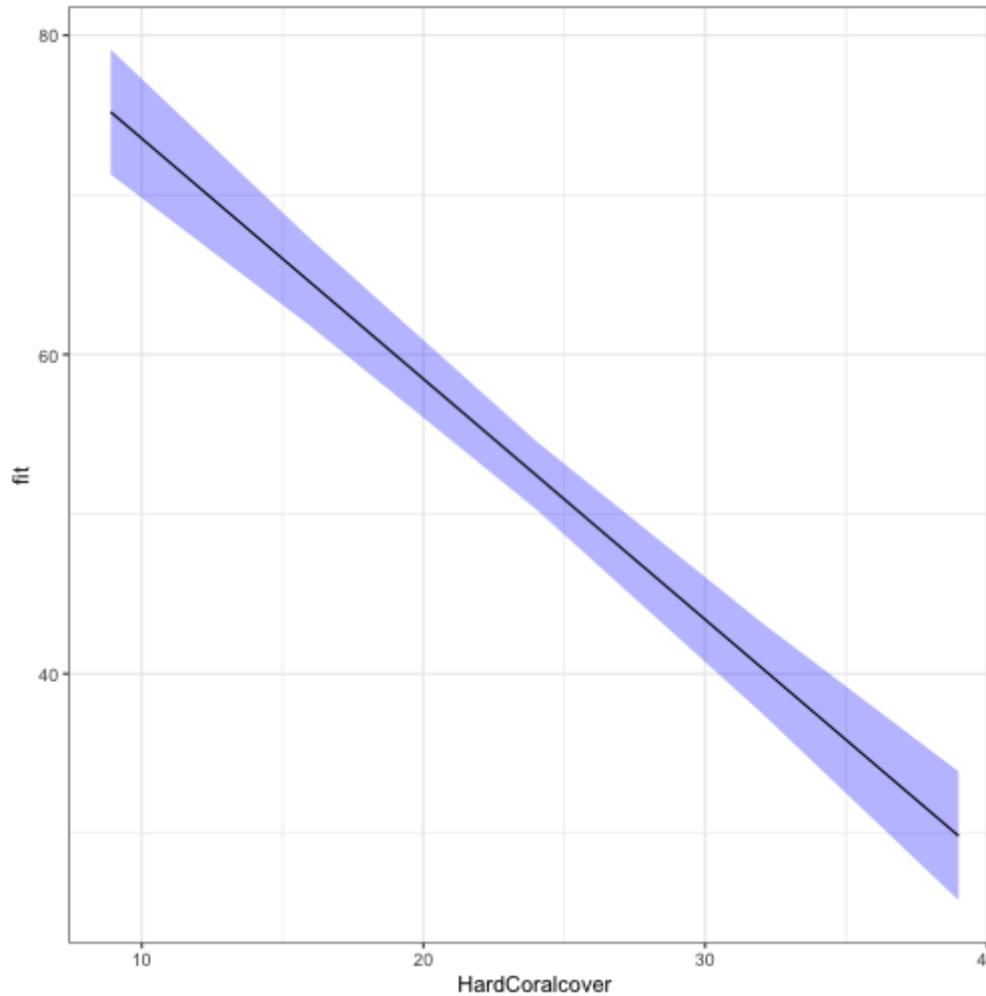
# Simple linear regression for prediction

$$Y = X \times \beta + \epsilon$$

**Using R:**

```
HardCoralcover= regdataWF$`Hard Coral`  
reg = lm(Algae ~ HardCoralcover, data = regdataWF)
```

# Simple regression for prediction



# Predicting data

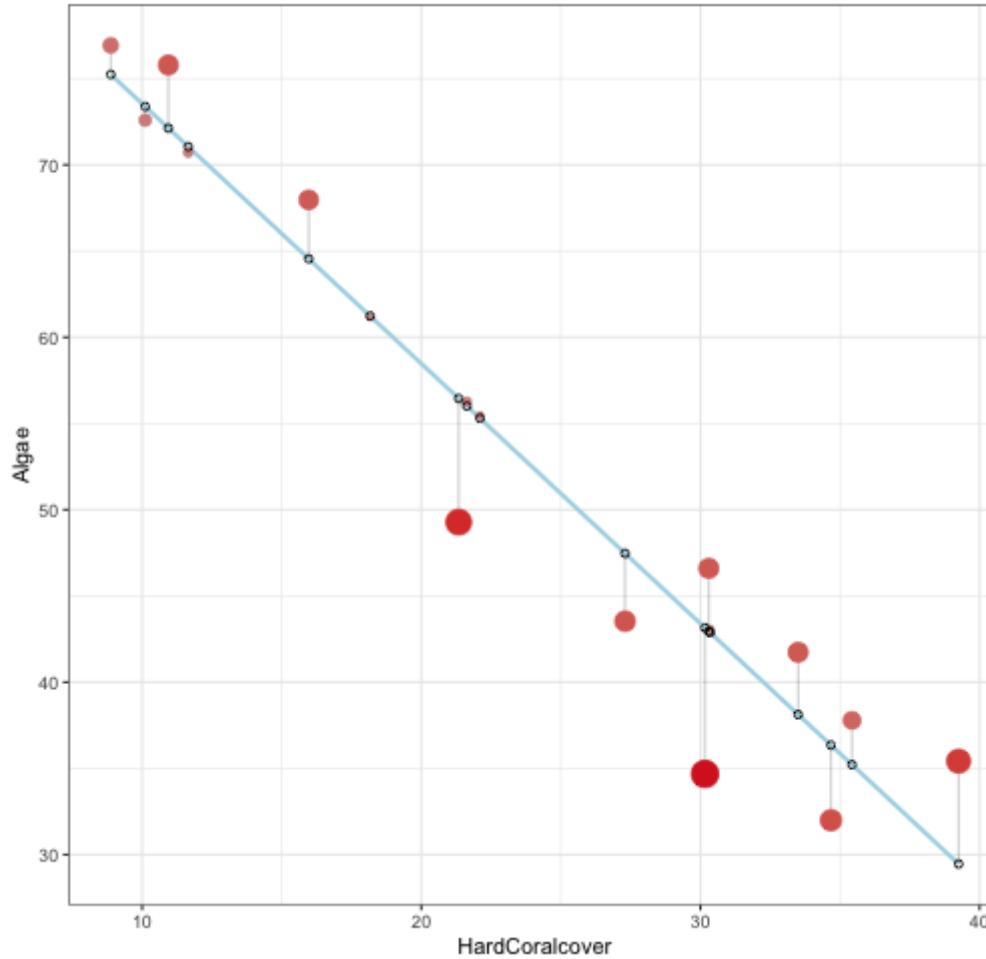
```
regdataWF$predicted <- predict(reg)
regdataWF$residuals <- residuals(reg)

head(regdataWF)

## # A tibble: 6 x 5
## # Groups:   Year [6]
##   Year Algae `Hard Coral` predicted residuals
##   <dbl> <dbl>          <dbl>      <dbl>      <dbl>
## 1 1993  34.7          30.2       43.2     -8.47
## 2 1995  32.0          34.7       36.4     -4.36
## 3 1996  35.4          39.2       29.5      5.96
## 4 1997  37.8          35.4       35.2      2.57
```

# Fit

Understanding the linear regression



# That's it!



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](#).

Lecturer: Patricia Menendez

Department of Econometrics and Business Statistics

 ETC5512.Clayton-x@monash.edu