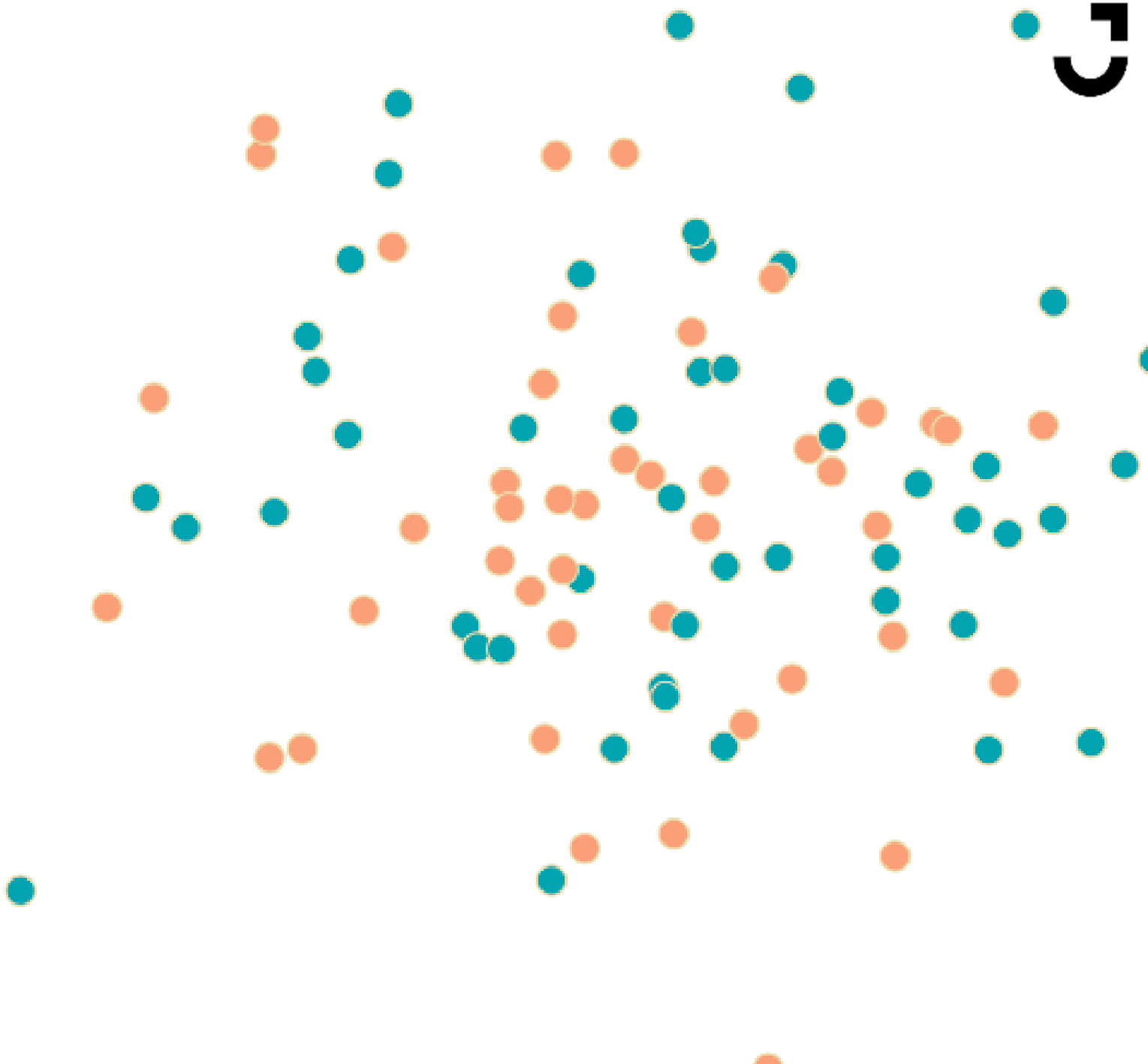
What if you have so much data it’s hard to make sense of it?

Greg Slack & Matthew Whittle

Jacobs



Outline

- The problem
  - 100s of spreadsheets
  - Using the mean but struggling to rationalise the outliers
- The solution
  - Talk about the statistics we use
  - Visualisation techniques
  - Make sense of large datasets

Why do we collect (so much) acoustic monitoring data?

- Detect presence of different species of bats
- Quantify bat activity levels at each site and use this to make inferences about their [relative] abundance
- Determine why bats are using the site
- Identify areas/periods with relatively high bat activity
- Evaluate the importance of the population
- Ecological Impact Assessment

How much should we collect?

Identifying Presence or Peak

Mathews et. al. (2016) show:

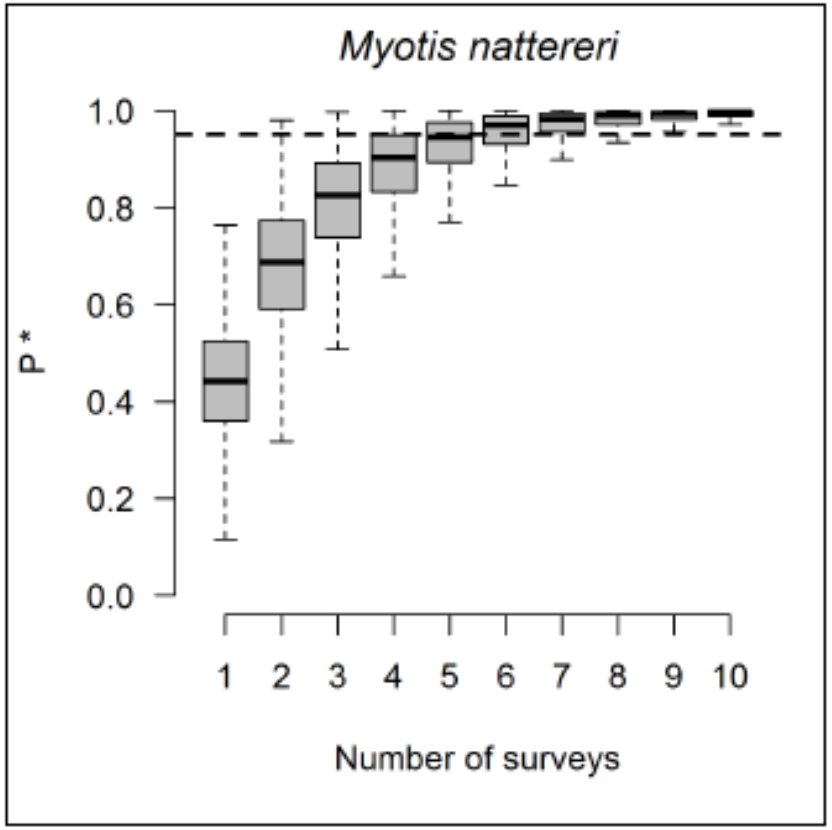
- 10 nights to detect presence;
- 16 nights to detect peak count.

Scott & Altringham (2014) identified:

- Up to 10 woodland transect surveys to identify presence

Current Guidance

- BCT Guidance: up to 5 nights per month, April – October;
- Wind Farm Guidance – min 10 nights per season.



Collins, J. (ed.), *Bat Surveys for Professional Ecologists: Good Practice Guidelines, 3rd edition.* The Bat Conservation Trust, London, 2016

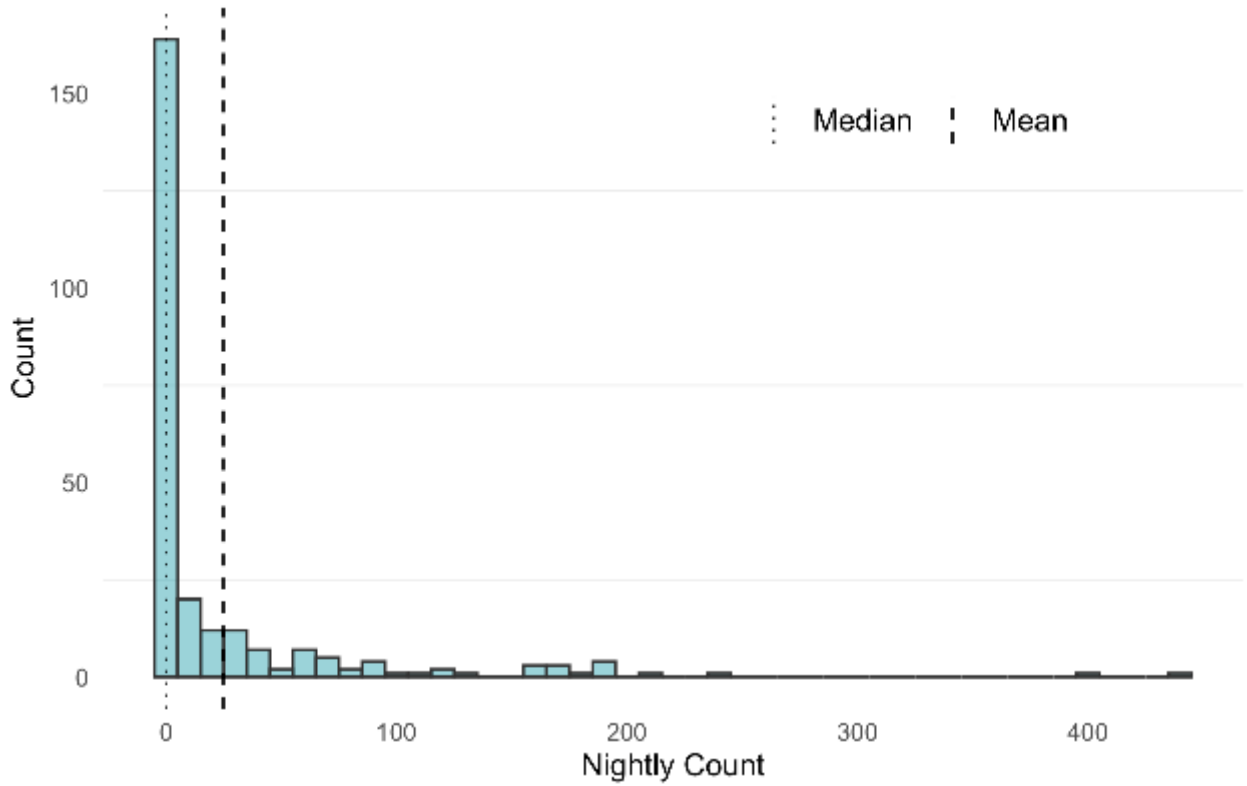
Anon (2019) Bats and onshore wind turbines:survey, assessment and mitigation. https://www.nature.scot/sites/default/files/2019-01/Bats%20and%20onshore%20wind%20turbines%20-%20survey%2C%20assessment%20and%20mitigation.pdf

Mathews F., Richardson S., Lintott P., & Hosken D. (2016) *Understanding the Risk to European Protected Species (bats) at Onshore Wind Turbine Sites to inform Risk Management. Report by University of Exeter.* Report for RenewableUK. Report for UK Department of Energy and Climate Change (DECC).

Scott C., & Altringham J. (2014) WC1015 *Developing Effective Methods for the Systematic Survailance of Bats in Woodland Habitats in the UK*. Defra http://sciencesearch.defra.gov.uk/Document.aspx?Document=12239_WC1015WoodlandBatsFinalReport.pdf

Measuring Average Activity Levels

- How should we measure average activity levels?
- The **mean** is commonly used to summarise bat activity
- Highly variable & temporally clustered
- Using only the mean can lead to improper estimates.
- Significant implications for:
  - impact assessment, and
  - targeting mitigation

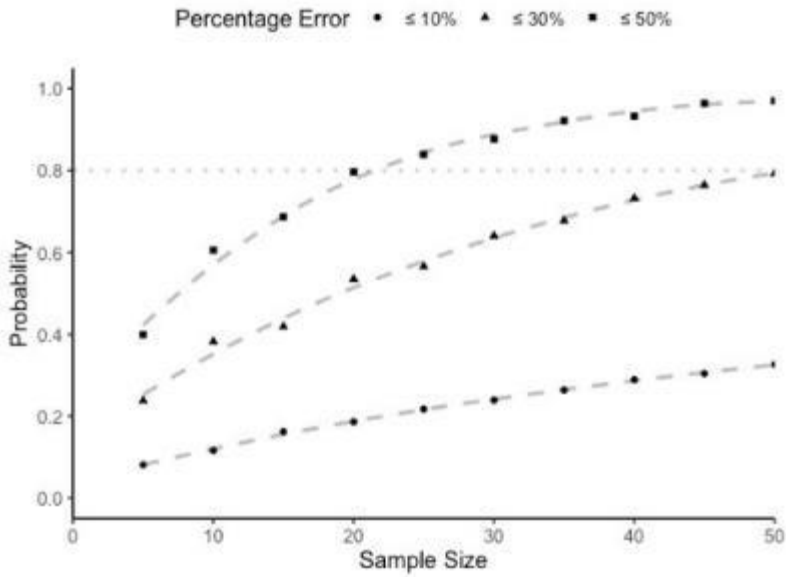


Lintott, P.R. and Mathews, F., 2018. *Basic mathematical errors may make ecological assessments unreliable*. Biodiversity and conservation, 27(1), pp.265-267.

How much should we collect?

Identifying Representative Level of Activity

- Representative mean or median?
- Neither is consistently representative where:
  - The mean activity is low
  - Variation (SD) is high
  - Even with relatively large sample sizes (20+ nights)
- 50 nights potentially required



The effect of sample size on the probability of accurately (10%, 30% & 50% error) characterising the mean activity level of common pipistrelle in spring

Slack G., & Whittle M. (2020) *What's sufficient for statics? - How much static detector data is enough?* British Island Bats. Vol 1, p84-96 https://cdn.bats.org.uk/pdf/Resources/Bat-Groups/Accessing-journals/BritishIslandsBats_VolOne_2020.pdf?mtime=20200629213941&focal=none

The more data the better, right?

The quantity of data we collect will be driven by:

- Increased storage capacity
- Better batteries
- Cheaper & smaller statics
- Faster processers and auto ID

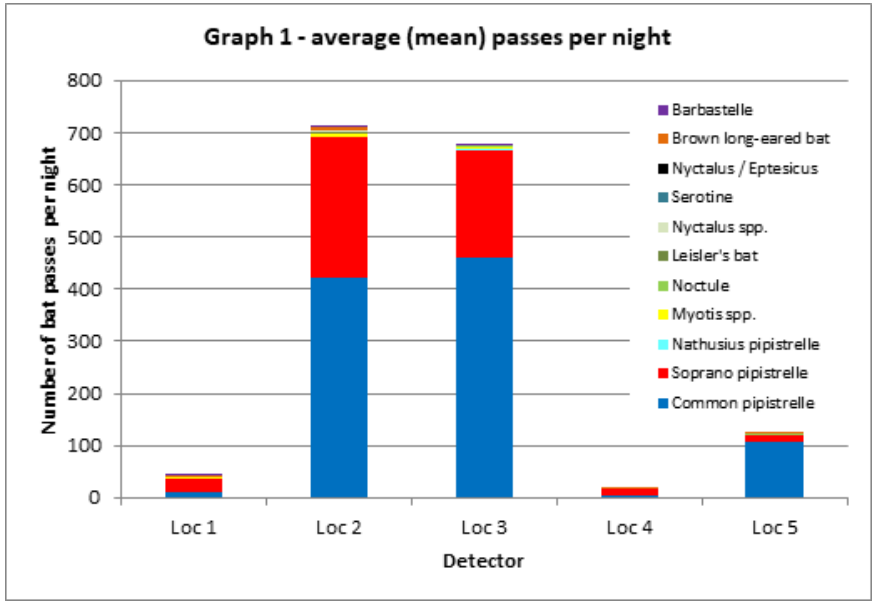
An increase is inevitable.

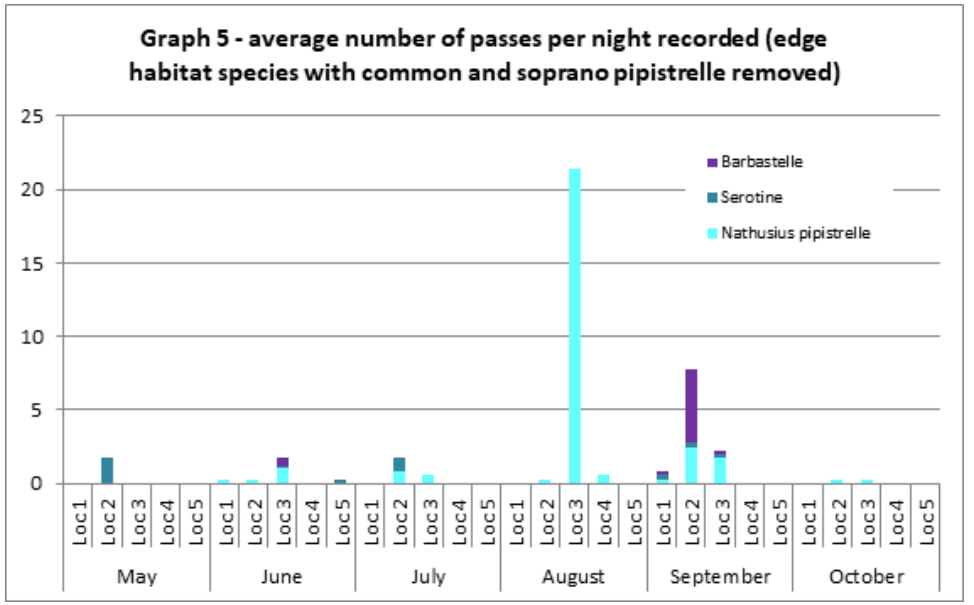
Photo courtesy of Dan Wildsmith

The more data the better, right?

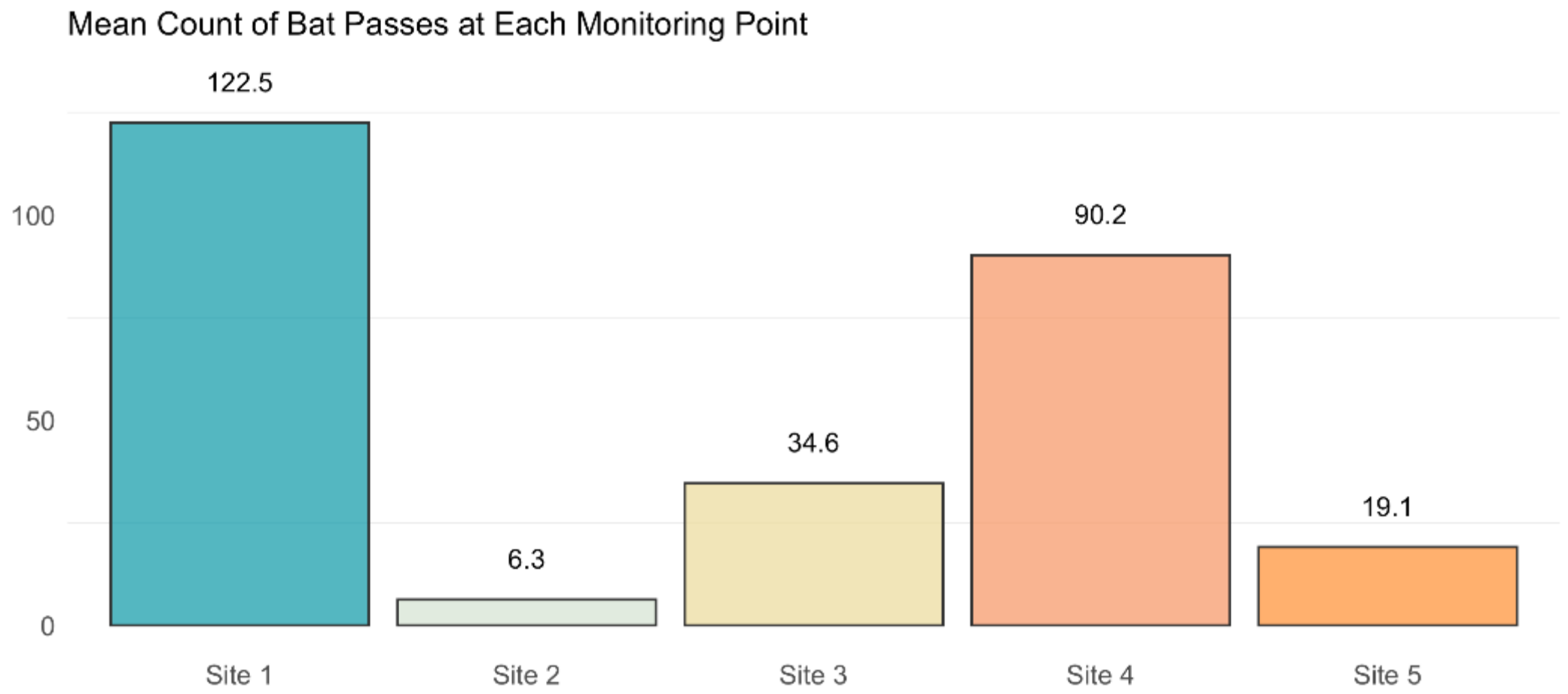
- **With more data, it becomes increasingly important to use well considered, robust methods for analysis**
- Simple visualizations are ok for few data
- Data must be managed and stored.

Development Size	Monitoring Locations	Nights of Monitoring	Results Spreadsheets
Medium	10	350	50
Large	50	1,750	350

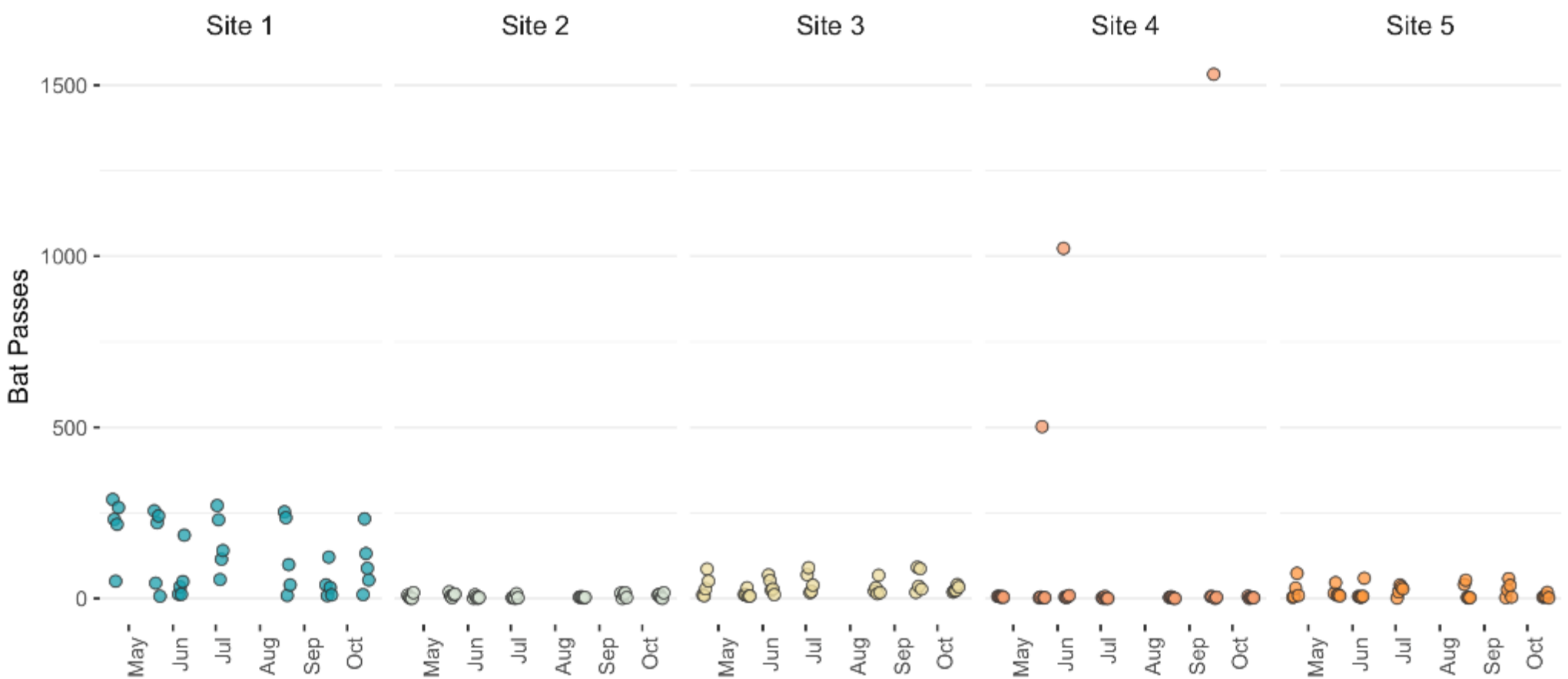




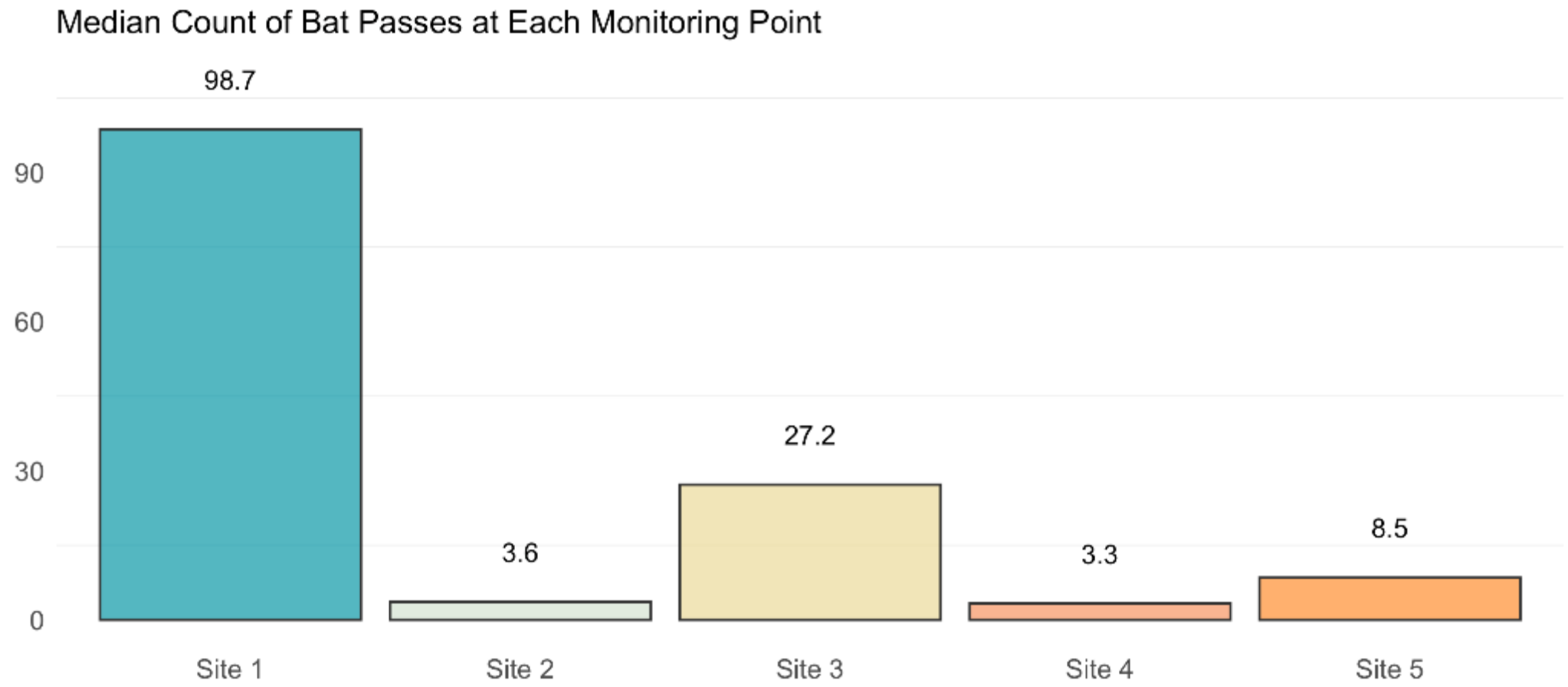
Which site has the highest & second highest value for bats?



Which site has the highest & second highest value for bats?

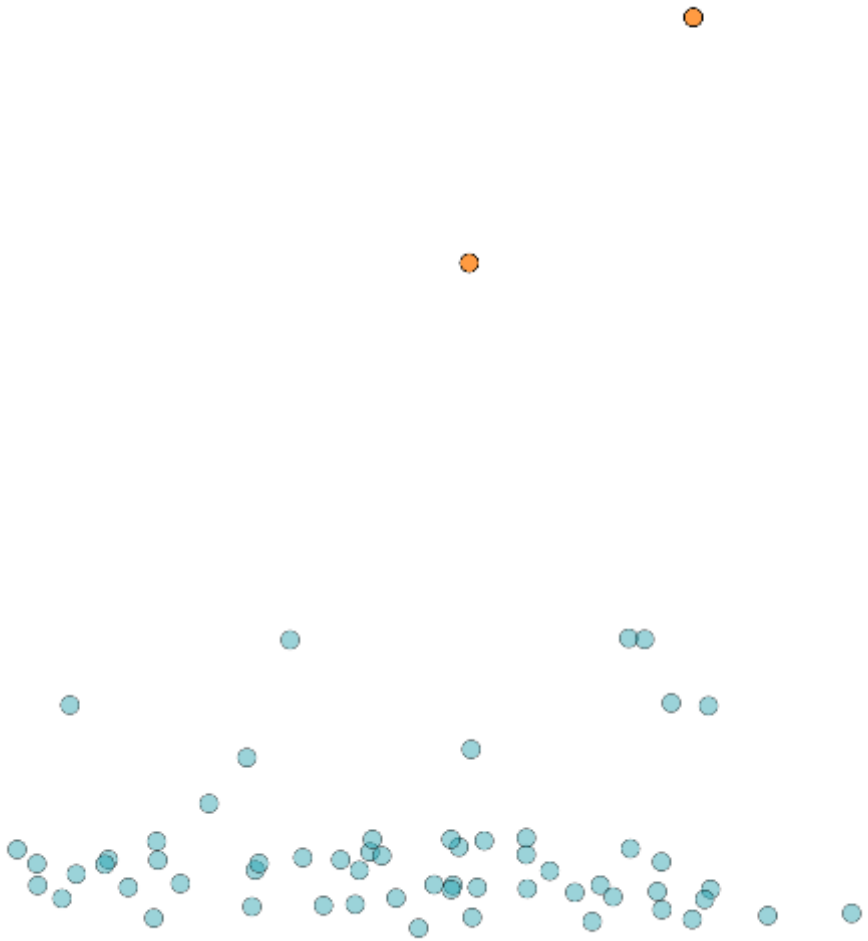


Which site has the highest & second highest value for bats?



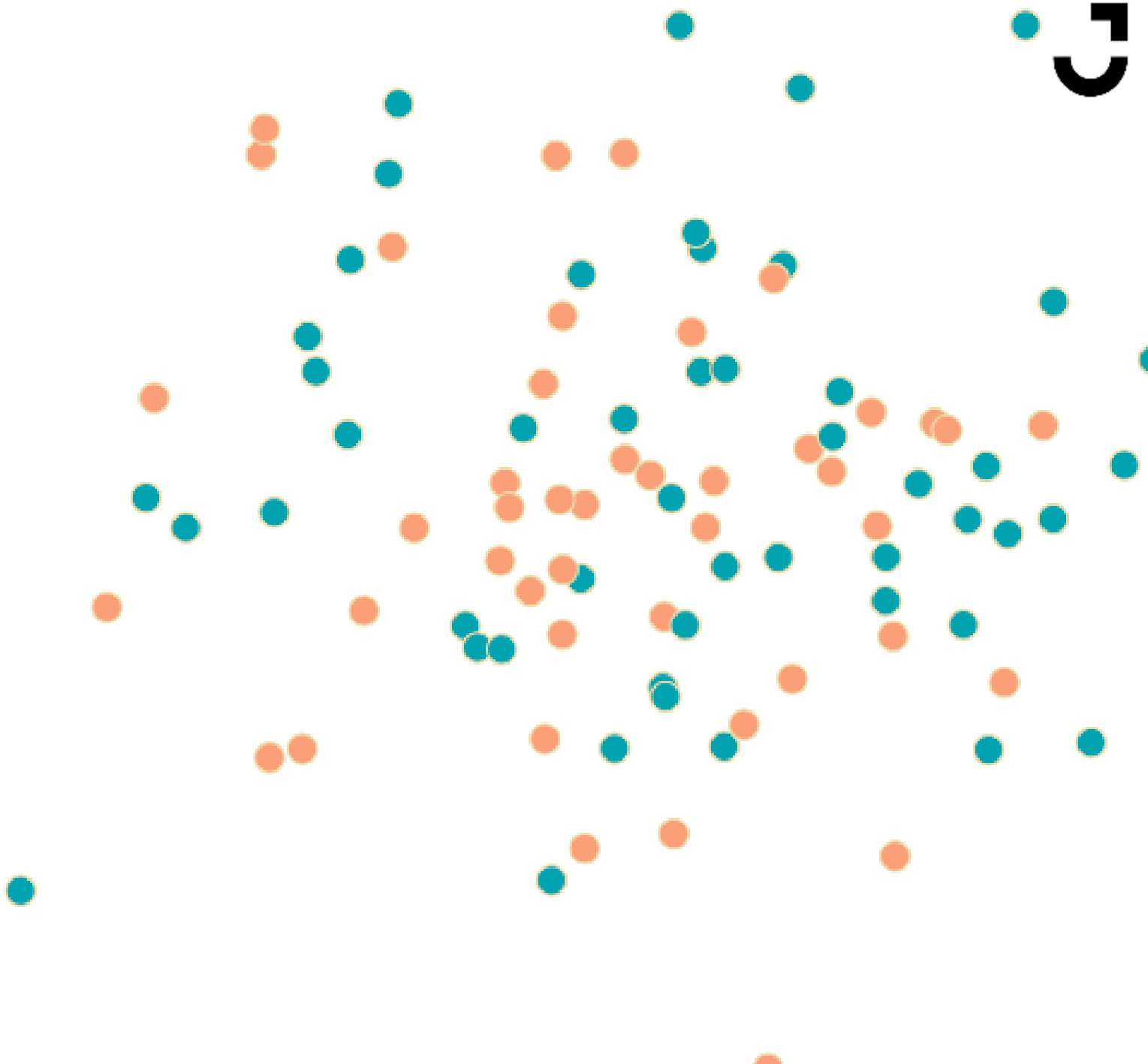
Do outliers matter?

- Ecological data often contain extreme values (or *outliers*), especially bat activity data
- Do these values contain relevant ecological information?
- Yes - might represent the use of a site during **favourable conditions that only occur occasionally**.
  - Insect hatch,
  - low wind,
  - warm nights
- Comparisons should be based upon bat activity peak, average and variance



Benhadi-Marín, J., 2018. **A conceptual framework to deal with outliers in ecology**. *Biodiversity and Conservation*, *27*(12), pp.3295-3300. Available from https://link.springer.com/article/10.1007/s10531-018-1602-2

Solution



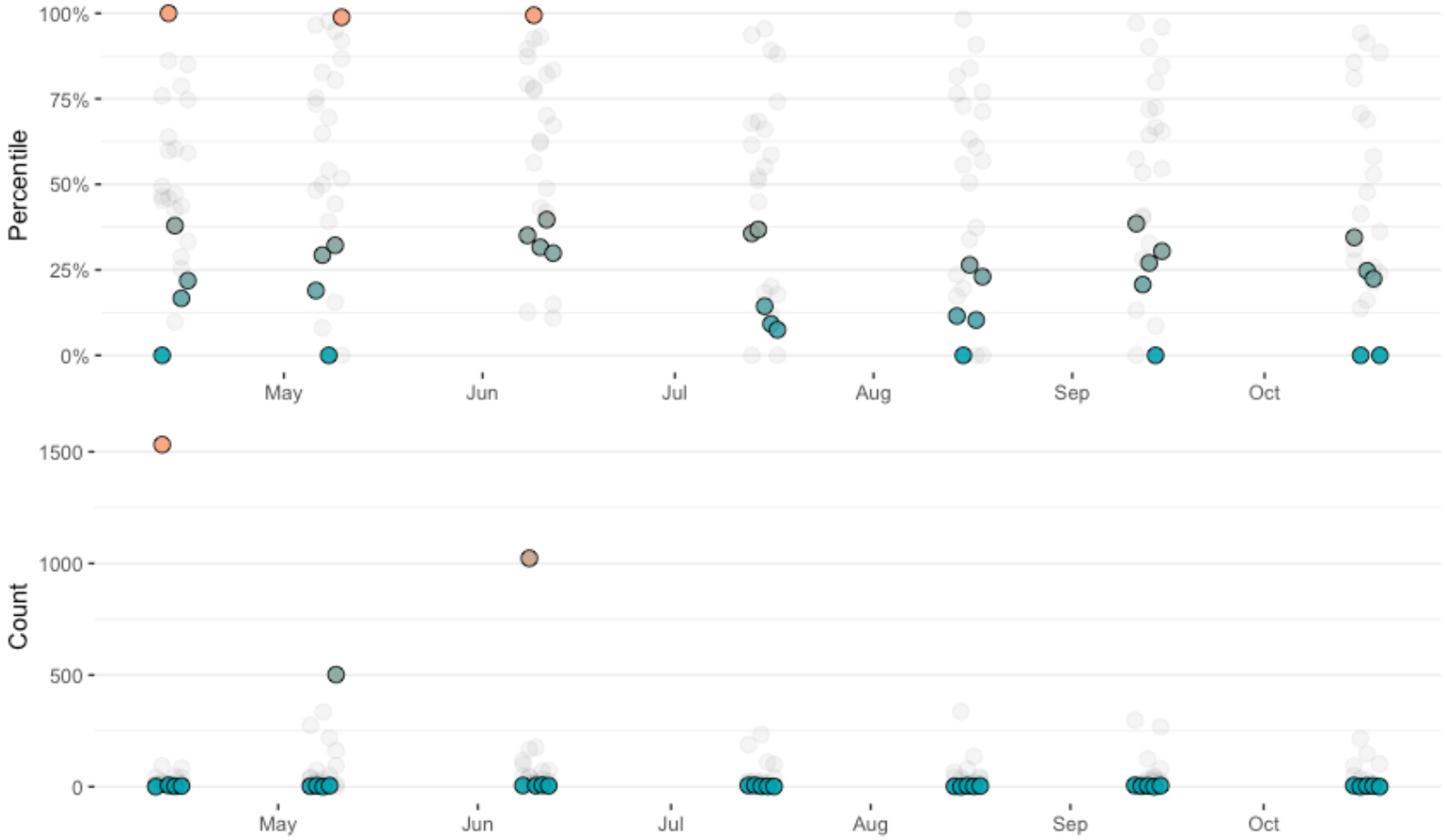
Evaluating Bat Activity

Table A7.1 Descriptive statistics for common and soprano pipistrelle passes per night.

Statistic	Common pipistrelle	Soprano pipistrelle
Number of records/nights	1,942	1,942
Mean	164.31	42.02
Median	37	5
Standard deviation	359.86	158.73
25% quartile	6	0
75% quartile	136	24
Maximum	3,815	2,426
Minimum	0	0

- Use a range of summary statistics
- Quantify bat activity in relative terms – using quartiles and percentiles
- Harness the power of Percentile Plots

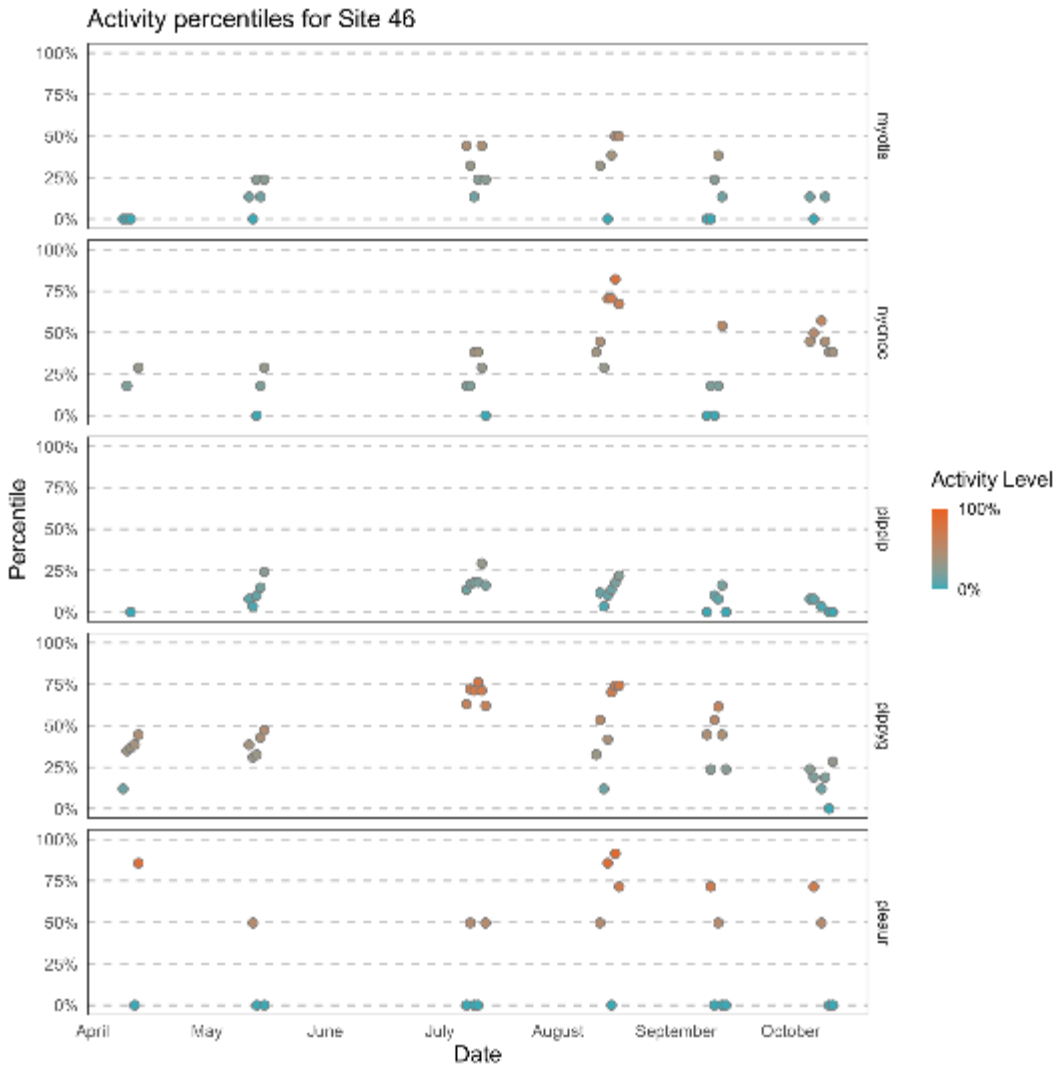
Nightly count of bat passes at Site 4

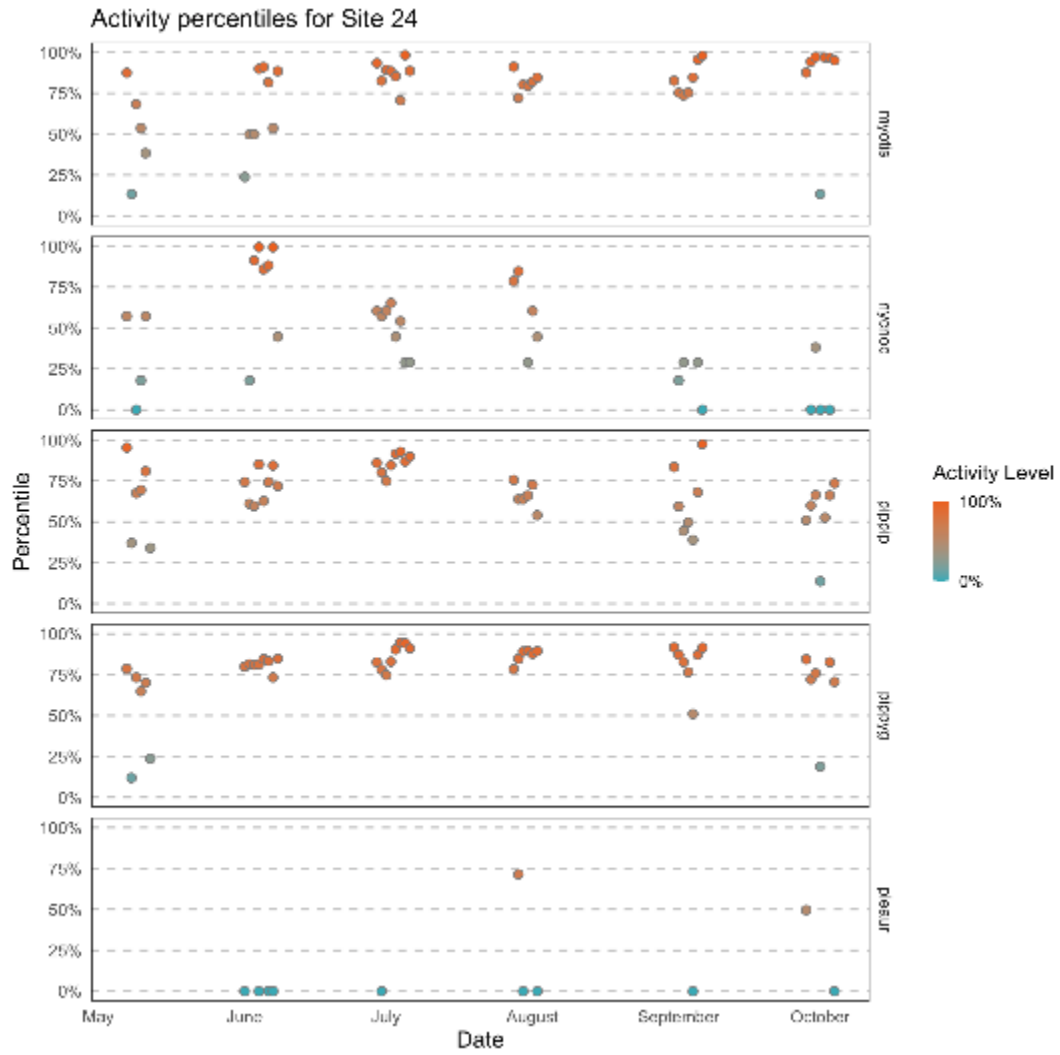


Percentile Plot: The raw count of bat passes are plotted on the bottom pane and the percentiles are plotted on the top pane.

Comparing Bat Activity

- Identify activity levels on a species-by-species basis



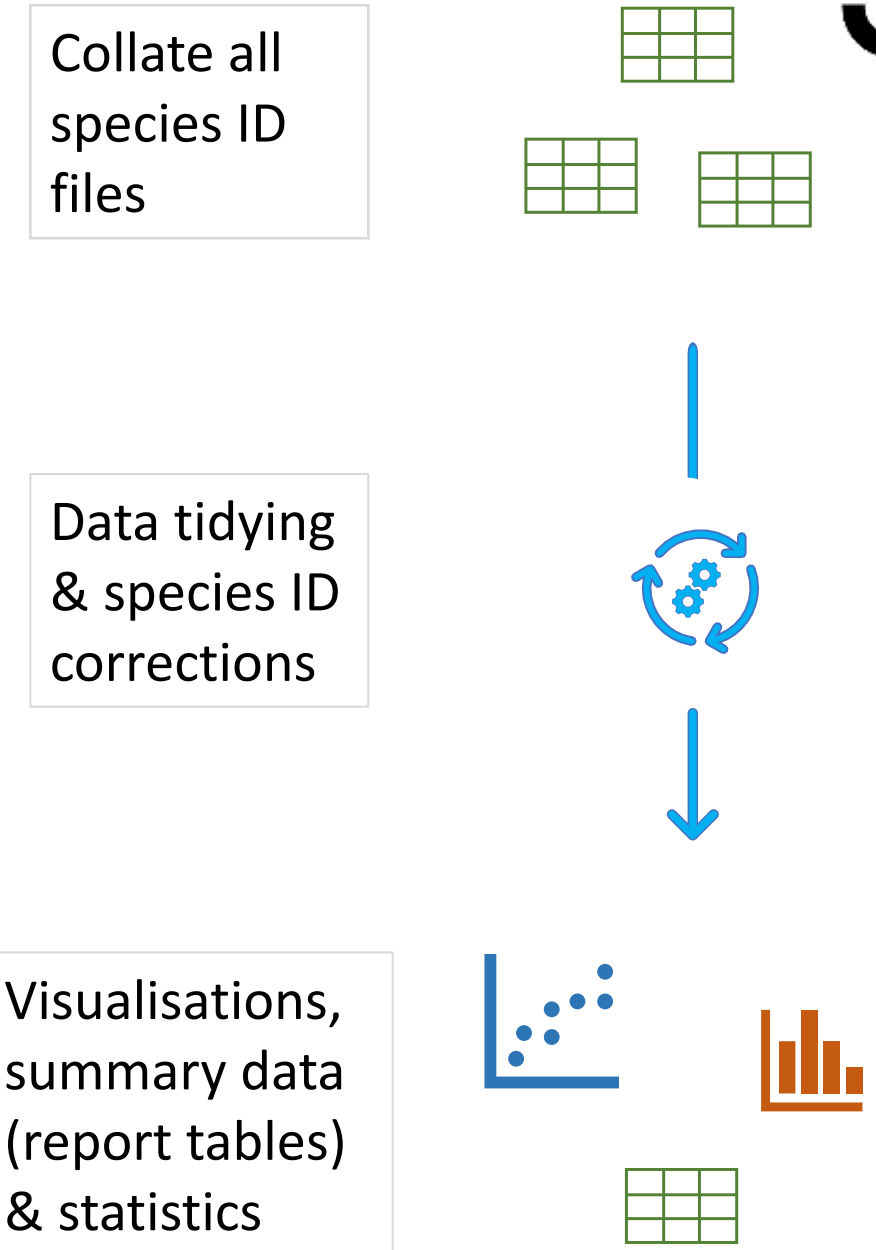


R: Critical to Adopting this Workflow

- Using R for data analysis has been critical to us adopting this workflow
- Popular programming language for data analysis
- Increased our capabilities & large time savings
- Key benefits over spreadsheets: automated, flexible and reproducible

R: Data Processing

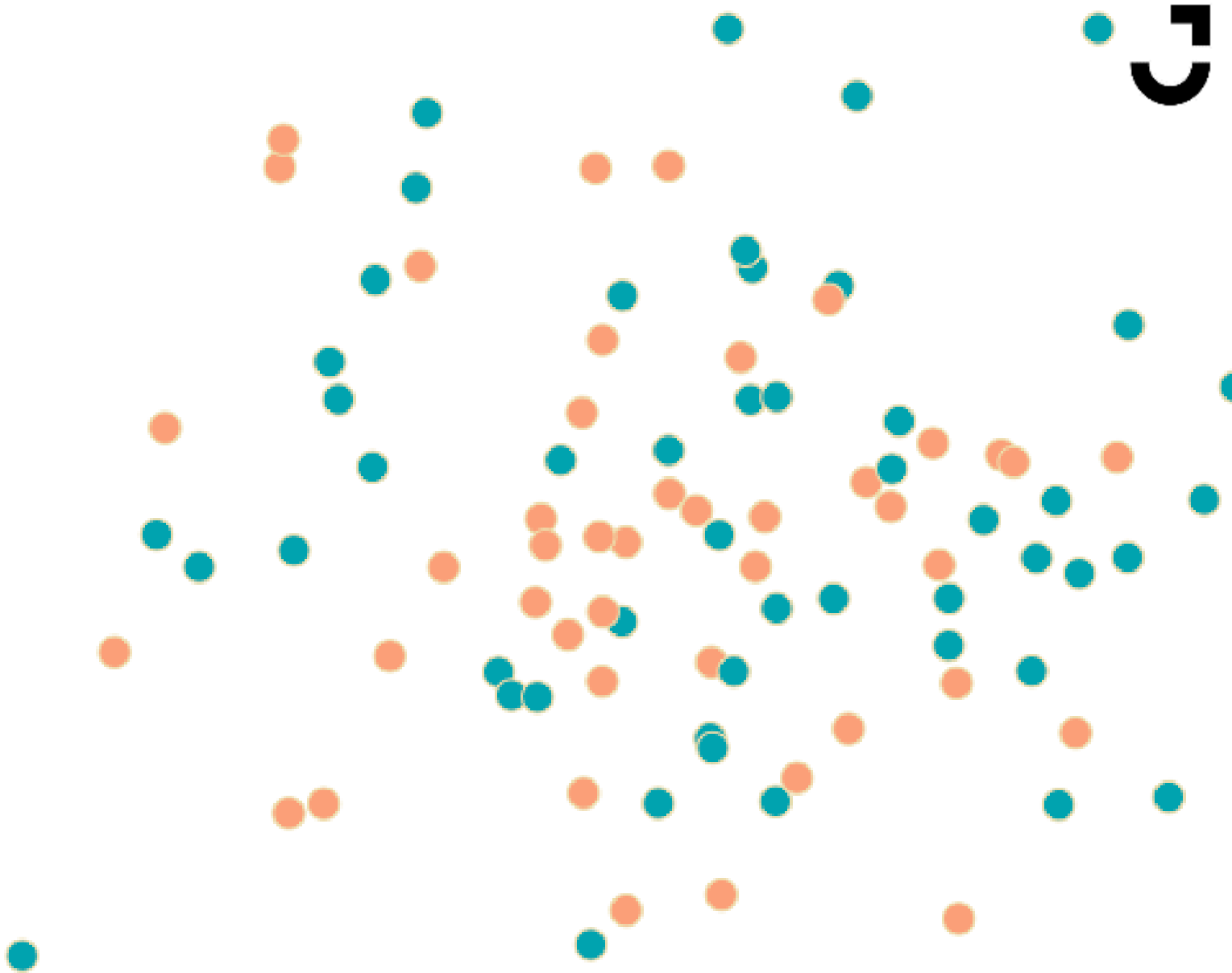
- Automated processing of acoustic monitoring data
  - Collating spreadsheets
  - Data tidying & species ID corrections
  - Statistics and visualisations
- Runs as data is collected - pick up locations with notable activity quickly.



Summary

- Ensure your survey effort is fit for purpose (are you detecting presence / absence or relative activity levels?)
- Don’t rely on one summary statistic – use a variety
- Use percentile plots to visualise nightly data
- Consider things on a species-by-species basis
- Develop an easily reproducible process

Any Questions?



Jacobs

Greg.Slack@Jacobs.com
Matthew.Whittle@Jacobs.com