Analysing field data collected in the tropical rainforests of Sabah, Borneo

SE_25 Data Science module check-in

What question are you trying to answer with the study?

What are the characteristic conditions in which certain species tend to appear more frequently?

Are certain species more sensitive to human activity?

Are **differences in land use** influence biodiversity?

Is there a **measurable decline** in biodiversity near sites with increased human activity?

How to analyse soundscape as a whole?

How to detect broad scale changes?

Why is this an interesting question?

- supporting current efforts (research project Continuous bio-acoustic monitoring)
- once a machine learning algorithm is set up to identify species from audio recordings, looking at spatial distribution could give real insights
- discovering patterns could lead to more focused research

Describing the data

- **Point_count_recordings:** tabular data describing the recorded audio files
- **Point_count_data:** visual and/or audio observations (of species) of two research assistants. These two tables can be linked by the "Point_count_ID" column and therefore locate the corresponding audio segment to each observation.
- **Locations:** Additional information about each recording location

Preliminary results

	Site	Site_type	Total nr of observations	Total nr of recordings	observation / recording
0	C Matrix	salvage_logged_forest	1314	84	15.642857
2	C10 621	selectively_logged_forest	1196	98	12.204082
3	D Matrix	salvage_logged_forest	1190	102	11.666667
1	OP Belian	oil_palm	1250	120	10.416667
4	OP3 843	oil_palm	1089	114	9.552632
7	VJR 2	old_growth_forest	878	97	9.051546
8	E1 648	selectively_logged_forest	839	96	8.739583
6	VJR 1	old_growth_forest	889	102	8.715686
5	D100 641	selectively_logged_forest	906	109	8.311927
9	E100 edge	selectively_logged_forest	796	113	7.044248
10	B10	selectively_logged_forest	747	108	6.916667
12	Riparian 1	old_growth_forest	648	102	6.352941
11	B1 602	selectively_logged_forest	686	124	5.532258
14	LFE river	selectively_logged_forest	32	6	5.333333
13	Riparian 2	old_growth_forest	446	107	4.168224

Conclusions and next steps

- the data available as of now is insufficient for a comprehensive study
- gather more information, especially about the different locations
- determine the scope, what can be build up on the data available (species detection or general soundscape analysis)