

Final Project – Birds Biodiversity Temporal Trends

Overview

As the concluding project for the applied statistics course, you will conduct an independent analysis of the long-term monitoring data collected in the Birds Biodiversity programme. Your goal is to quantify how biodiversity indicators have evolved over time, provide sound statistical uncertainty assessments, and highlight species-specific stories that emerge from the dataset.

To promote genuine data-driven thinking, this assignment specifies what questions must be addressed without prescribing how you should answer them. You are free to select the statistical techniques, models, and visual encodings you feel are most appropriate, provided that you justify your choices.

This project introduces a new dataset that you have not analysed previously in the course: the multi-year Birds Biodiversity monitoring records stored in `projects/birds-biodiversity/data/raw/Observations_2012-2025.xlsx`. You may reuse any helper modules in the repository (for example `src/data_io.py`) or build your own tooling if it better suits your workflow.

Core Deliverables

Your final submission must include:

1. Technical report (Markdown or PDF) explaining your approach, the indicators you studied, and the conclusions you reached. The narrative should be self-contained and understandable to a statistically literate audience.
2. Notebook(s) or scripts that implement your analysis reproducibly. Structure your code so that another analyst can re-run it from a clean checkout.
3. Figures and tables that support your findings (export them to `projects/birds-biodiversity/figures` and `projects/birds-biodiversity/results` respectively). Label axes clearly and include units.

Key Questions to Address

1. Dataset Familiarisation and Descriptive Analysis

- Provide an overview of the dataset structure: table dimensions, key columns, time coverage, and unique identifiers (transects, observers, species, etc.).
- Produce descriptive statistics that quantify the distribution of relevant variables (counts, observer effort, and environmental descriptors such as weather conditions recorded during surveys). Pay special attention to how observation effort varies across time and space, since it influences the reliability and interpretation of downstream results. Use both tabular summaries and visualisations.
- Discuss any notable data quality considerations (missing values, outliers, anomalous periods) that may influence subsequent analyses.

2. Multi-Year Indicator Trends

- Select at least three biodiversity or sampling indicators that you believe are informative (e.g., spatial coverage, species diversity, density, observer effort). Justify each choice.
- For every indicator, compute annual estimates over the full 2014–2025 horizon and provide confidence intervals that reflect estimation uncertainty (you decide on the method and confidence level).
- Quantify and interpret temporal trends. You may use linear models, generalized models, smoothing, or any other defensible approach, but explain why it suits the indicator's behaviour.

3. Species-Level Evolution

- Identify a subset of bird species that you consider interesting (e.g., high abundance, conservation concern, distinctive trends).
- For each selected species, analyse how its recorded presence or counts have changed over time. Present confidence intervals and comment on the limitations of your estimation procedure.
- Discuss potential ecological or operational reasons behind the patterns you observe (based on external knowledge or plausible hypotheses).

4. Synthesis and Recommendations

- Summarise the main insights you discovered about the monitoring programme. Highlight converging evidence across indicators as well as any contradictions.

- Outline actionable recommendations for future data collection or biodiversity management. These can include methodological improvements, additional indicators to monitor, or follow-up questions that deserve investigation.
- Reflect on limitations: data quality, modelling assumptions, sensitivity to methodological choices, etc.

Guidance (Non-Prescriptive)

Expectations

- The goal is not to produce breakthrough ecological findings, but to demonstrate statistical creativity and rigour in how you analyse the data.
- You are encouraged to propose novel indicators or visualisations, provided that you justify them and implement them correctly.
- Actionable biodiversity management recommendations are welcome but not mandatory; you will not be penalised if no clear actions emerge.

Incorrect conclusions caused by flawed assumptions, improper methodology, or coding mistakes will be penalised heavily, so validate every step carefully.

You will only receive the raw dataset and this assignment brief. Feel free to rely on standard statistical libraries and to create your own helper code as needed. AI assistance is permitted, but you are responsible for every line of code you submit.

- Visual outputs should be polished and professional. If warnings remain in the execution logs, explain why they could not or should not be suppressed.
- Ensure that code is well commented and easy to follow so that reviewers can understand your reasoning without guesswork.
- Confidence intervals can be obtained via analytical formulas, resampling, Bayesian intervals, or other defensible methods. Argue why your choice is appropriate for each indicator/species.
- When modelling trends, check diagnostics to ensure your assumptions are reasonable (residual plots, robustness checks, alternative specifications, etc.).
- Keep visualisations clear and focused. A small number of high-quality figures is preferable to a large set of redundant plots.

Suggested Workflow

1. Familiarise yourself with the data: explore structure, missingness, notable events, and potential covariates (transects, observers, weather conditions).
2. Define indicators and species scope: document why each metric or species was chosen before diving into coding.
3. Implement the analysis: structure your notebooks/scripts into reusable sections (data loading, indicator computation, inference, visualisation).
4. Validate results: sanity-check intermediate outputs and ensure reproducibility from a clean run.
5. Prepare the report and artefacts: polish the narrative, export figures/tables, and verify that file paths point to the expected directories.

Evaluation Criteria

- Statistical soundness (confidence intervals and trends are computed appropriately, methods are justified).
- Insight and interpretation (discussion goes beyond describing plots to making meaningful inferences).
- Clarity and reproducibility (code runs from scratch, report is well structured, figures are informative).
- Creativity and initiative (evidence of thoughtful indicator selection, methodological experimentation, or novel species/observers insights).

Submission

- Package all materials (report, code, notebooks, figures, tables, README) into a single ZIP archive named `FinalProject_Lastname1_Lastname2_Lastname3.zip`.
- Email the archive to `stephane.rivaud@universite-paris-saclay.fr` before 6 November, 23:59 (Paris time). Late submissions will not be accepted under any circumstances.
- Include a README at the root of the archive describing the folder contents and reproduction instructions. Provide environment details, entry points, and any runtime considerations.
- The report may be a standalone PDF that references generated outputs, or the entire study may be documented directly within notebook(s). Choose the format that best showcases your analysis.

- Each group should list all members and student IDs in both the README and the report/notebook.
- An oral defence will be held between 7 and 14 November. Each group will present for 10 minutes followed by a 5-minute Q&A.

Good luck, and use this project to demonstrate the full breadth of statistical skills you've acquired throughout the course.