# Raw Data Manipulation

Author: Joshua K. Pickering

Date Initiated: 2025-01-29 | Last Updated: 2025-02-27

## Breeding Bird Territories

Breeding bird territoriy data of the Long Point Breeding Bird Census (LPBBC) project.

1. Original dataset “1965-2021 BBC Stats\_MSWB\_29 Mar\_ 2022.xlsx” is maintained as a Microsoft Excel file, saved in RProject file ‘Data\_Analysis.RProject’. Data was not manipulated but spreadsheet was subsequently copied and saved as a flat CSV file; “1965-2021 BBC Stats\_MSWB\_29 Mar\_ 2022.csv”.

Data manipulations in Microsoft Excel csv files are as follows:

1. Original data CSV file copied for manipulation in a new file to evaluate a subset of the data (i.e., 1991-2021), and is saved as a flat CSV file; “bird\_spterritories\_abund.csv”
2. Provide consistent text formatting that is font ‘Aptos’, size 11, not bolded.
3. A value ‘2906’ has been added to the line (i.e., Tamarack Slough | 1991 | 8.75 | Middle | 2017 | YSFL | + | 8 | 23.42 | M. Iles), that does not contain a value in the column “Orig Line” that subsequently follows 2905, and precedes 2908.
4. Data from 1965-1990 was removed from the dataset to accommodate research interests in data from 1991-2021, this included the removal of rows identified as 1-349 as noted in the column labelled “Orig Line”.
5. The column labelled “Stage” was also removed from the dataset due to limited research interest and potential for error in oversimplification of complex successional stage definitions (Bradstreet & Pickering, 2022).
6. Added new blank column for column “B” and labelled this column “sitecode”. Added all site codes as an identified in column “sitecode” for each site (e.g., identifier “BGGR” for site “Bluegrass – Milkweed Grassland”) for consistency and efficiency in data analyses from up-to-date sources (Bradstreet & Pickering, 2022; Pickering et al., 2024). Subsequently removed column ‘C’ labelled as “Full Site Name” to limit confusion and remove long format site names.
7. Update of all site sizes according to new measurements (Bradstreet & Pickering, 2022; Pickering et al., 2024), in column ‘D’, labelled “Area (ha) in year”. Subsequently relabelled this column ‘D’ as “sitesize\_ha”
8. Rename column ‘E’ from “Year” to lowercase ‘y’ as “year” for consistency of column labels.
9. Rename column ‘F’ from “Species” to lowercase ‘s’ as “species” for consistency of column labels.
10. Add new column ‘G’ and label as “speciescode”. Subsequently update speciescode column with contemporary 4-letter species codes for bird species (e.g., HOWR, House Wren, is now NHWR, Northern House Wren) and remove column ‘F’ labelled as “species” but keep column labelled “speciescode”, which becomes the new column ‘F’.
11. Rename column ‘G’ from “# Territories” to lowercase ‘t’ and remove ‘#’ to result in the column labelled as “territories” for consistency of column labels.
12. Replace all values ‘+’ in column ‘G’ labelled as “territories” to reflect the value associated with a rare territory as ‘0.25’ (Bradstreet et al., 1991). There was 76 replacements within the dataset.
13. Increase the number of decimals for column ‘G’, labelled as “territories” to 2 decimal places to include all integer values appropriate for data analyses.
14. Created new column ‘A’ and label as “uniqueID”. Populate column with a unique sample id using formula “=F2&"\_"&C2&"\_"&G2” to create a standard identifier that has year, site code (4-letter), and species code (e.g. 1991\_BGGR\_TRES).
15. Subsequently remove column ‘B’ labelled as “Orig Line” to limit confusion and remove a duplication of a unique row identifier.
16. Remove column ‘C’ labelled as “Year at size”, as not necessary for data analyses.
17. Remove columns ‘G’, ‘H’, ‘I’, ‘J’, ‘K’, ‘L’, ‘M’, ‘N’, and ‘O’, labelled as “Number of visits”, “Census hours”, “Author1”, “Author2”, “Author3”, “Author4”, “Author5”, “Author6”, and “Notes” to reduce unnecessary meta data for data analyses.
18. Column headers are as follows:

Column A - uniqueID

Column B - speciescode

Column C - year

Column D - sitesize\_ha

Column E - sitecode

Column F - territories

To conduct analyses with entire presence/absence data, complete autofill of zeros for all 100 species observed over time within the LPBBC dataset for all sites during sampled years. New dataset is maintained as CSV file, saved as “bird\_spterritories\_presabs.csv”.

1. Column headers are as follows:

Column A - uniqueID

Column B - speciescode

Column C - year

Column D - sitecode

Column E - territories

Populate a supporting file with relevant information into a species characteristics description file, saved as “bird\_spcharacteristics.csv”.

Column A - sp\_common\_name\_uppercase\_IBP

Column B - sp\_scientific\_name\_IBP

Column C - 2024\_species\_4code\_IBP

Column D - 2024\_species\_6code\_IBP

Column E - 2021\_species\_4code\_LPBBC

Column F - BOTW\_reference

Column G - sp\_weight\_g

Column H - weight\_reference

Column I - habitat\_guild\_BOTW

Column J - food\_item\_BOTW

Column K - nest\_guild\_BOTW

Column L - forage\_behavior\_BOTW

Column M - conservation\_status\_BOTW

Column N - food\_foraging\_guild\_degraaf

Column O - habitat\_foraging\_guild\_degraaf

Column P - behaviour\_foraging\_guild\_degraaf

Column Q - sp\_common\_name\_uppercase\_NHIC

Column R - sp\_scientific\_name\_NHIC

Column S - sp\_common\_name\_lowercase\_NHIC

Column T - sp\_common\_name\_noun\_NHIC

Column U - sp\_common\_name\_sentence\_NHIC

Column V - sp\_author\_NHIC

Column W - sp\_french\_name\_NHIC

Column X - sp\_common\_name\_AOS

Column Y - sp\_french\_name\_AOS

Column Z - sp\_order\_AOS

Column AA - sp\_family\_AOS

Column AB - sp\_subfamily\_name\_AOS

Column AC - sp\_genus\_AOS

Column AD - sp\_scientific\_name\_AOS

Column AE - status\_introduced\_AOS

Column AF - sp\_common\_name\_PFN

Column AG - sp\_Ojibwe\_name\_PFN

Column AH - sp\_name\_meaning\_PFN

Column AI - sp\_latin\_name\_PFN

Column AJ - Breeding.Biome

Column AK - Winter.Biome

Column AL - Family

Column AM - bird.group

Column AN - Migrate

Column AO - AI

Column AP - native

Column AQ - popest

Column AR - popestlci

Column AS - popestuci

Column AT - first\_year\_popest

Column AU - last\_year\_popest

Column AV - Pop.source

Column AW - Trajectory\_data\_source

Column AX - Trajectory\_firstyear

Column AY - Trajectory\_lastyear

Column AZ - Loss\_med

Column BA - Loss\_lci

Column BB - Loss\_uci

Column BC - Loss\_lqrt

Column BD - Loss\_uqrt

Column BE - PopUsCa

Column BF - PopLC95

Column BG - PopUC95

Column BH - TimeAdj.meanlog

Column BI - TimeAdj.sdlog

Column BJ - Distance Adj.

Column BK - Pair Adj.

Supporting information files are identified and referenced as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Original file name | File type | Description | Reference | Revised file name |
| aaw1313\_data\_s1 | xlsx | Supplementary material for bird species descriptions and continental trends from Rosenberg et al. (2019). | Rosenberg, K. V., Dokter, A. M., Blancher, P. J., Sauer, J. R., Smith, A. C., Smith, P. A., Stanton, J. C., Panjabi, A., Helft, L., Parr, M., & Marra, P. P. (2019). Decline of the North American avifauna. Science, 366(6461), 120–124. <https://doi.org/10.1126/science.aaw1313> | RegionalBirdPopulations\_Rosenberg et al\_2019 |
| BBC BIRDS\_Characteristics\_BI | xlsx |  | Long Point Bird Observatory, 2022 | LocalBirdCharacteristics\_LPBO\_2022 |
| NACC\_list\_species | csv |  | Chesser, R. T., S. M. Billerman, K. J. Burns, C. Cicero, J. L. Dunn, B. E. Hernández-Baños, R. A. Jiménez, Oscar Johnson, A. W. Kratter, N. A. Mason, P. C. Rasmussen, and J. V. Remsen, Jr. 2024. Check-list of North American Birds (online). American Ornithological Society. <https://checklist.americanornithology.org/taxa/> | NorthAmericanBirdSpeciesList\_AOS\_2024 |
| Ontario\_Species\_list | xlsx |  | Natural Heritage Information Center (NHIC). (2024, November 26). Ontario Species list. Ministry of Natural Resources. <https://www.ontario.ca/page/get-natural-heritage-information> | OntarioBirdSpeciesList\_NHIC\_2024 |
|  |  |  |  |  |

Data manipulations in R script files are as follows:

To conduct analyses related to community composition regarding species foraging and nesting guilds, join reference files to territory data files.

## Woody Stem Counts

Woody stem data of the Long Point Breeding Bird Census (LPBBC) project.

1. Original dataset “Appendix\_3\_Woody\_Stems\_Dataset\_1991-2021\_Long\_Point\_Vegetation\_Monitoring\_March\_2022.xlsx” is maintained as a Microsoft Excel file, saved as an RProject file ‘phd\_longpoint\_breedingbirds’. Data was not manipulated but spreadsheet was subsequently copied and saved as a flat CSV file; “1965-2021 BBC Stats\_MSWB\_29 Mar\_ 2022.csv”.

Data manipulations in the Microsoft Excel csv file are as follows:

1. Original data CSV file copied for manipulation in a new file to evaluate a subset of the data (i.e., 1991-2021), and is saved as a flat CSV file; “bird\_spterritories\_abund.csv”
2. Provide consistent text formatting that is font ‘Aptos’, size 11, not bolded.
3. Columns A, B, C & D, labelled “Timeline\_line” “Stage\_code”, “Stage\_name”, “Group\_name” were removed from the dataset due to limited research interest and potential for error in oversimplification of complex successional stage definitions (Bradstreet & Pickering, 2022).
4. Renamed new column ‘A’ without uppercase letters or special characters to “sitecode”. Subsequently removed column ‘B’ labelled as “Site\_name” to limit confusion and remove long format site names.
5. Columns C, D, F, G & H, labelled “Site\_codexPost”, “Date”, “Year”, “Month”, “Day”, “YearXSite\_codeXPost” were removed to not duplicate sample identifiers.
6. Column ‘C’, labelled “Year” was shifted two columns left to represent column ‘A’ and follow prior data layouts for breeding bird territory data.
7. New column ‘C’ labelled “Post” was renamed with all lowercase letters; “post”.
8. Column ‘D’ labelled “Species\_code” was renamed “species\_7code\_NHIC” to follow naming format as used for breeding bird territory data.
9. Created new column ‘A’ and label as “veg\_uniqueID” and populate column with a unique sample id using formula “=F2&"\_"&C2&"\_"&G2” to create a standard identifier that has 4 components; year, site code (4-letter), post, and species code (7-letter). Example is: 1991\_BGGR\_D11\_PRUVIRG).
10. Remove columns ‘K’, ‘L’, ‘M’, & ‘N’, that are labelled; “S-VT\_stemcount”, “Belowbrowse\_stemcount”, “Abovebrowse\_stemcount”, and “S-Sapling\_stemcount
11. Rename columns ‘F’, ‘G’, ‘H’, ‘I’, & ‘J’, that are labelled “Small\_stemcount”, “Medium\_stemcountTall\_stemcount”, “VeryTall\_stemcount”, and “Sapling\_stemcount” to “s\_wdystemcount”, “m\_wdystemcount”, “t\_wdystemcount”, “vt\_wdystemcount”, and “sap\_wdystemcount”, respectively.
12. Rename columns ‘K’ & ‘L’, labelled as “DBH\_measured\_sapling\_m”, and “Height\_calculated\_sapling\_m”, to “sap\_dbh\_m”, and “sap\_height\_m”
13. Remove columns ‘M’, ‘N’, ‘O’, ‘P’, ‘Q’, ‘R’, ‘S’, ‘T’, & ‘U’, that are labelled; “Small\_IRV\_m³”, “Medium\_IRV\_m³”, “Tall\_IRV\_m³”, “VeryTall\_IRV\_m³”, “Sapling\_IRV\_m³”, “S-VT\_IRV\_m³”, “Below\_browse\_IRV”, “Abovebrowse\_IRV\_m³”, and “S-Sapling\_IRV\_\_m³”.
14. Delete all descriptive species characteristics columns ‘M’, ‘N’, ‘O’, ‘P’, ‘Q’, ‘R’, ‘S’, ‘T’, ‘U’, ‘V’, ‘W’, ‘X’, ‘Y’, ‘Z’, ‘AA’, ‘AB’, … ‘BG’ that are labelled as “Species\_scientific”, “Exotic/Native\_status”, “Identified\_synonyms”, “SPECIES\_ELEMENT\_ID ELCODE”, “Species\_scientific”, “AUTHOR”, “ENGLISH\_COMMON\_NAME”, “SUBNATIONAL\_SYNONYMS”, “SUB\_OTHER\_COM\_NAME\_ALL PROVINCIALLY\_TRACKED”, “S\_RANK”, “COSEWIC\_STATUS”, “SARA\_STATUS”, “N\_RANK”, “EXOTIC\_STATUS”, “RESTRICTED\_SPECIES”, “TAXON\_GROUP\_HIGHER”, “TAXON\_GROUP\_LOWER”, “KINGDOM”, “PHYLUM”, “CLASS”, “ORDER”, “FAMILY”, “GENUS”, “S\_RANK\_CHANGE\_DATE”, “S\_RANK\_REVIEW\_DATE”, “S\_RANK\_REASONS”, “SUBNATIONAL\_TAXONOMY\_COMMENTS”, “DISTRIBUTION\_COMMENTS”, “EO\_TRACKING\_COMMENTS”, “COEFF\_CONSERVATISM”, “COEFF\_WETNESS”, “ELEMENT\_GLOBAL\_ID”, “GLOBAL\_SCIENTIFIC\_NAME”, “GNAME\_REF”, “NAME\_IN\_CONCEPT\_REF”, “CONCEPT\_REF\_CODE”, “GLOBAL\_SYNONYMS”, “GLOBAL\_ENGL\_NAME”, “GLOB\_OTHER\_COM\_NAME\_ALL”, “ELEMENT\_NATIONAL\_ID”, “NATIONAL\_SCIENTIFIC\_NAME NATIONAL\_SYNONYMS”, “NATIONAL\_ENGL\_NAME”, and “NATL\_OTHER\_COM\_NAME\_ALL”.
15. Populate all BLANK data with zeros, totalled 10,762 replacements for data in columns ‘F’, ‘G’, ‘H’, ‘I’, ‘J’, ‘K’, and ‘L’, labelled as “s\_wdystemcount”, “m\_wdystemcount”, “t\_wdystemcount”, “vt\_wdystemcount”, “sap\_wdystemcount”, “sap\_dbh\_m”, and “sap\_height\_m”.
16. Sum small, medium, and tall stem counts for all samples and label column “within\_browse”.
17. Sum very tall and sapling stem counts for all samples and label column “above\_browse”.
18. Column headers are as follows:

Column A - veg\_uniqueID

Column B - year

Column C - sitecode

Column D - post

Column E - species\_7code\_NHIC

Column F - s\_wdystemcount

Column G - m\_wdystemcount

Column H - t\_wdystemcount

Column I - vt\_wdystemcount

Column J - sap\_wdystemcount

Column K - sap\_dbh\_m

Column L - sap\_height\_m

Column M - within\_browse

Column N - above\_browse