ReadMe file

[Stewart, Kerry](https://researchdata.reading.ac.uk/view/creators/Stewart%3D3AKerry%3D3A%3D3A.html) (2023): Data supporting the article 'Functional diversity metrics can perform well with highly incomplete datasets'. University of Reading. Dataset. <https://doi.org/10.17864/1947.000481>

Copyright 2023 Kerry Stewart. This dataset is licenced under a Creative Commons Attribution 4.0 International Licence: <https://creativecommons.org/licenses/by/4.0/>

For data and scripts supporting ‘*Functional diversity metrics can perform well with highly incomplete datasets’*.

Kerry Stewart, Carlos P. Carmona, Chris Clements, Chris Venditti, Joseph A. Tobias, Manuela González-Suárez

Please send any questions to kerrysmith189@gmail.com.

# Data

## **Datasets**

## *Birds*

Datasets for analysis of functional diversity response to missing data in birds. The datasets were made by removing data from AVONET data (Tobias et al., 2021; available for use under a [CC by 4.0](https://creativecommons.org/licenses/by/4.0/legalcode) license), and in some cases imputing missing trait data. These datasets were the foundation for the results produced in the main text and were used for Supplementary information section 2 (dimensionality comparison, data handling comparison, and dataset comparison) alongside *Crocs* and *Plants*.

## *Crocs*

Datasets for analysis of functional diversity response to missing data in crocodiles. The datasets were made by removing data from Griffith et al. (2022; available for use under a [CC by 4.0](https://creativecommons.org/licenses/by/4.0/legalcode) license) and in some cases imputing missing trait data. Used for dataset comparison in Supplementary information section 2.

## *Plants*

Datasets for analysis of functional diversity response to missing data in plants. The datasets were made by removing data from Díaz al. (2022; available for use under a [CC by 4.0](https://creativecommons.org/licenses/by/4.0/legalcode) license) and in some cases imputing missing trait data. Used for dataset comparison in Supplementary information section 2.

## *Tyrannidae*

Datasets for comparing accuracy of imputation methods in Tyrannidae. The datasets were made by selecting Tyrannidae species from AVONET data (Tobias et al., 2021; available for use under a [CC by 4.0](https://creativecommons.org/licenses/by/4.0/legalcode) license), removing data and imputing missing trait data with a range of imputation methods. These datasets are relevant for results presented in Supplementary Information section 8.

Naming structure of datasets:

Data handling

**cc**: complete case analysis

**pf**: missForest (with phylogenetic information)

mf: missForest (without phylogenetic information, Tyrannidae only)

ph: Phylopars (Tyrannidae only)

mi: MICE (Tyrannidae only)

**cc\_r10\_third\_1.csv**

Bias

**r**: random

**l**: low

**m**: medium

**h**: high

Missingness

**10**: 10% of species have missing data

..

**70**: 70% of species have missing data

Number of traits with missing data

**third:** birds- 4 out of 11 traits have data removed (for species with missing data)

Crocs- 2 out of 5 traits have data removed

Plants – 2 out of 6 traits have data removed

**twothirds**: seven out of 11 traits have data removed (birds only)

Iteration

**1:** first iteration

…

**50**: 50th iteration

## **Results**

In all results files functional diversity was calculated with distance-based Rao (dis\_cg), convex hull (con.hull), trait probability density richness (tpd\_ric) and trait probability density divergence (tpd\_div).

## *imp\_comparison\_results\_tyrannidae.csv*

Functional diversity of all datasets in *Tyrannidae*, with deviation from functional diversity of the full dataset. Relevant for supplementary information section 7.

## *results\_birds\_calc\_PCA\_2dim.csv*

Functional diversity of all datasets in *Birds*, with deviation from functional diversity of the full dataset. Functional diversity was calculated with two dimensions (two principal components). Relevant for supplementary information section 2.

## *results\_birds\_calc\_PCA\_3dim.csv*

Functional diversity of all datasets in *Birds*, with deviation from functional diversity of the full dataset. Functional diversity was calculated with three dimensions (three principal components). Relevant for results presented in the main text.

## *results\_birds\_no\_PCA\_2dim.csv*

Functional diversity of all datasets in *Birds*, with deviation from functional diversity of the full dataset. Functional diversity was calculated with two dimensions (two traits, body mass and hand wing index). Relevant for supplementary information section 2.

## *results\_birds\_proj\_PCA\_3dim.csv*

Functional diversity of all datasets in *Birds*, with deviation from functional diversity of the full dataset. Functional diversity was calculated with three dimensions (three principal components) using projected principal component values, rather than calculated principal component values. Relevant for supplementary information section 8.

## *results\_crocs\_calc\_PCA\_2dim.csv*

Functional diversity of all datasets in *Crocs*, with deviation from functional diversity of the full dataset. Functional diversity was calculated with two dimensions (two principal components). Relevant for supplementary information section 2.

## *results\_plants\_calc\_PCA\_2dim.csv*

Functional diversity of all datasets in *Plants*, with deviation from functional diversity of the full dataset. Functional diversity was calculated with two dimensions (two principal components). Relevant for supplementary information section 2.

# Scripts

## *imp\_comparison\_processing\_results\_tyrannidae\_supp7.R*

Processing datasets in *Tyrannidae*, and producing figures in supplementary information section 7.

## *results\_stats\_main\_and\_supp2.R*

Statistics presented in the main text, Supplementary Information section 2 and Supplementary Information section 11 and for producing figures in Supplementary information section 2.

## *crocs.R*

Analysis on crocodile datasets including dataset creation (random and biased removal of trait data, and imputation), principal component analysis and functional diversity estimation.

## *imp\_comparison\_dataset\_creation\_tyrannidae\_supp7.R*

Create datasets in *Tyrannidae*, including random and biased removal of trait data and imputation.

## *processing\_birds\_3dim.R*

Processing datasets in *Birds* and producing *results\_birds\_calc\_PCA\_3dim.csv*.

## *plants.R*

Analysis on *Plants* datasets including dataset creation (random and biased removal of trait data, and imputation), principal component analysis and functional diversity estimation.

## *tree\_balance\_supp6.R*

For analyses in Supplementary information section 6

## *imputation\_nrmse\_supp5.R*

For calculating imputation accuracy, as presented in Supplementary information section 5.

*pca\_creation\_birds.R*

Calculating and projecting principal components based on datasets in *Birds*.

*birds\_2dims\_nopc.R*

Analysis of *Birds* datasets in two dimensions with no principal component analysis. Produces *results\_birds\_no\_PCA\_2dim.csv*, and relevant for data handling comparison in Supplementary information section 2.

*results\_birds\_3dims\_inc\_figure\_3.R*

For presenting results in *results\_birds\_calc\_PCA\_3dim.csv* and for plotting figure 3.

*dataset\_creation\_birds.R*

Production of *Birds* datasets, by removing data from the complete dataset, and in some cases imputing missing data.

*figure\_5.R*

Code for producing figure 5.

*figure\_4.R*

Code for producing figure 4.

*figure\_2.R*

Code for producing figure 2.

# References

Díaz, S., Kattge, J., Cornelissen, J., H., C. et al. (2022). The global spectrum of plant form and function: enhanced species-level trait dataset. *Scientific Data*, 9, 755. https://doi.org/10.1038/s41597-022-01774-9.

Griffith, P., Lang, J., W., Turvey, S., T. & Gumbs, R. (2022). Data from: Using functional traits to identify conservation priorities for the world's crocodylians [Data set]. *Zenodo*. https://doi.org/10.5281/zenodo.6645415.

Tobias, J. (2021). AVONET: morphological, ecological and geographical data for all birds (Tobias et al., 2022 Ecology Letters doi: https://doi.org/10.111/ele.13898). *figshare*. Dataset. https://doi.org/10.6084/m9.figshare.16586228.v5