**Data dictionary**

This data dictionary is a document that describes the basic organization of the tables in our project. This document contains a list of variables in the tables as well as the assigned variable names and a description of each variable. The data dictionary is used primarily for data analysis.

**1. Table name ‘lat’. Record count = 534,810,844**

This table is used to store the distribution (number of cells) of each species (including the original and new species) within a latitude band(one degree).

|  |  |  |
| --- | --- | --- |
| Variable name | Variable type | Values/notes |
| ID | Numeric | identity |
| LAT | Numeric | [-55, 84] 139 |
| COUNT\_LAT | Numeric | Number of the cells occupied by the species in year X |
| SPECIES\_ID | Varchar | Species id within a simulation |
| YEAR | Numeric | [100, 1200000] by =100 |
| sp\_id | Numeric | Seed id |
| niche\_breadth | Varchar | large/narrow |
| dispersal\_ability | Varchar | good/poor |
| speciation\_year | Numeric | 10000 only |
| extinction\_threshold | Numeric | 0/5 |

**2. Table name ‘lon’. Removed**

This table is used to store the distribution (number of cells) of each species (including the original and new species) within a longitude band(one degree).

|  |  |  |
| --- | --- | --- |
| Variable name | Variable type | Values/notes |
| ID | Numeric | identity |
| LON | Numeric | [-180, 180] 360 |
| COUNT\_LON | Numeric | Number of the cells occupied by the species in year X |
| SPECIES\_ID | Varchar | Species id within a simulation |
| YEAR | Numeric | [100, 1200000] by =100 |
| sp\_id | Numeric | Seed id |
| niche\_breadth | Varchar | large/narrow |
| dispersal\_ability | Varchar | good/poor |
| speciation\_year | Numeric | 10000 only |
| extinction\_threshold | Numeric | 0/5 |

**3. Table name ‘biome’. Record count = 150,500,124**

This table is used to store the distribution (number of cells) of each species (including the original and new species) within a biome. We have 11 different type of biomes. The original reference has 9 biomes. I split the desert in to 3 parts by the annual temperature. See figure below.



|  |  |  |
| --- | --- | --- |
| Variable name | Variable type | Values/notes |
| ID | Numeric | identity |
| BIOME | Numeric | [1, 11] |
| COUNT\_BIOME | Numeric | Number of the cells occupied by the species in year X |
| SPECIES\_ID | Varchar | Species id within a simulation |
| YEAR | Numeric | [100, 1200000] by =100 |
| sp\_id | Numeric | Seed id |
| niche\_breadth | Varchar | large/narrow |
| dispersal\_ability | Varchar | good/poor |
| speciation\_year | Numeric | 10000 only |
| extinction\_threshold | Numeric | 0/5 |

**4. Table name ‘****Temperate\_Tropic’. Record count = 55,193,460**

This table is used to store the distribution (number of cells) of each species (including the original and new species) within temperate or tropic. (18 degree annual temperature). 0 is temperate. 1 is tropic.

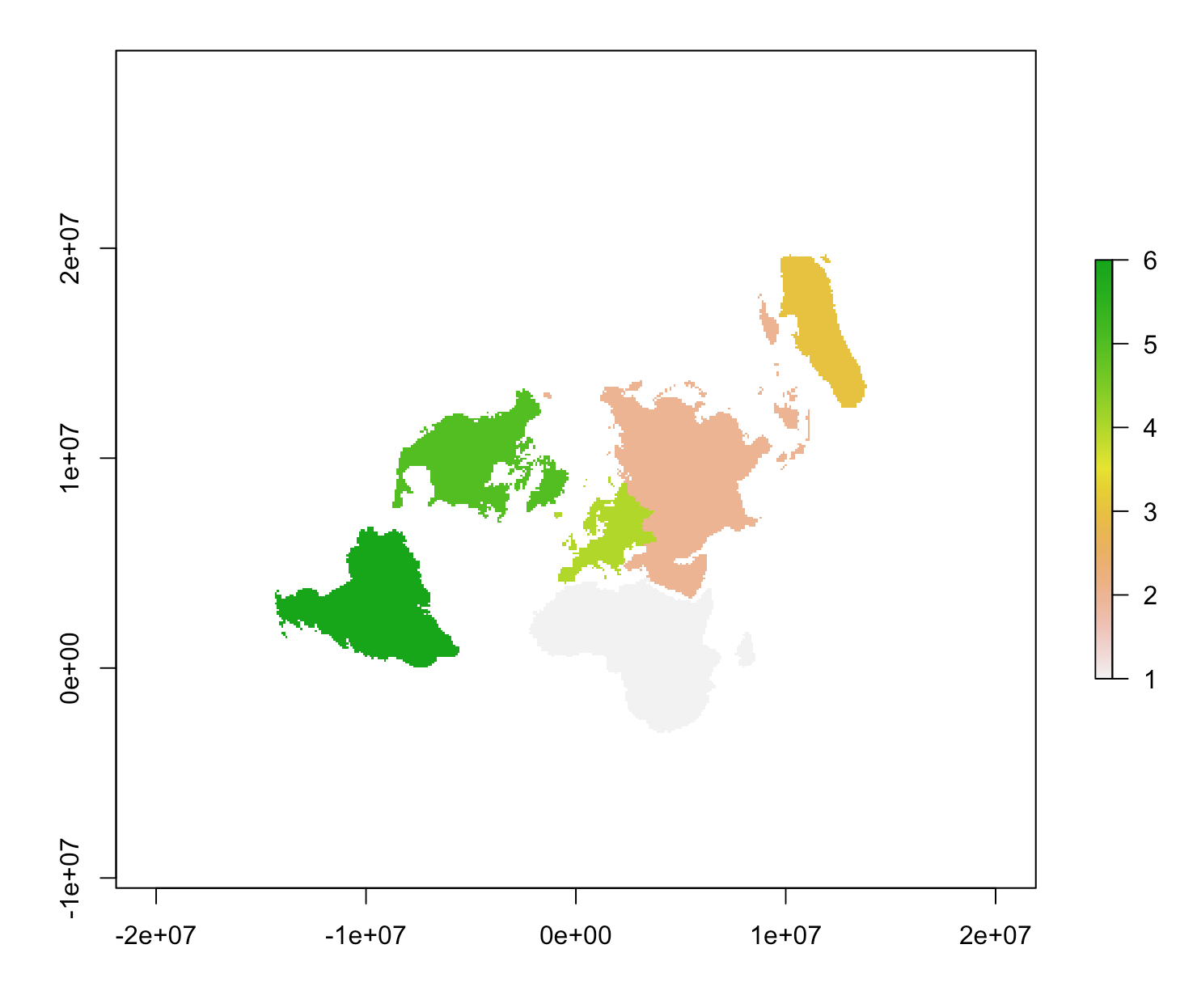
|  |  |  |
| --- | --- | --- |
| Variable name | Variable type | Values/notes |
| ID | Numeric | identity |
| Temperate\_Tropic | Numeric | [0, 1] |
| COUNT\_Temperate\_Tropic | Numeric | Number of the cells occupied by the species in year X |
| SPECIES\_ID | Varchar | Species id within a simulation |
| YEAR | Numeric | [100, 1200000] by =100 |
| sp\_id | Numeric | Seed id |
| niche\_breadth | Varchar | large/narrow |
| dispersal\_ability | Varchar | good/poor |
| speciation\_year | Numeric | 10000 only |
| extinction\_threshold | Numeric | 0/5 |

**5. Table name ‘continent’. Record count = 87,715,653**

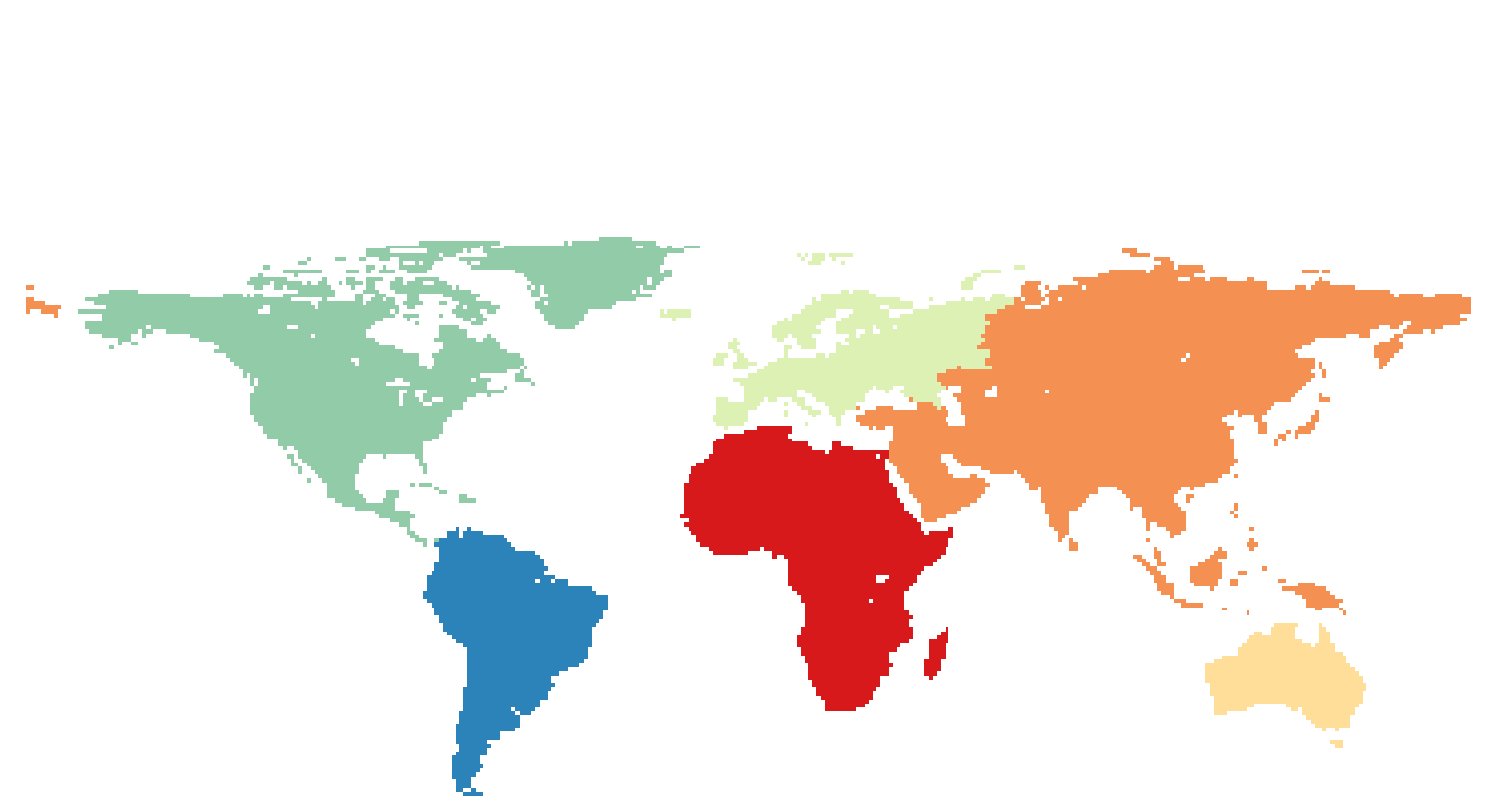
This table is used to store the distribution (number of cells) of each species (including the original and new species) within a continent.

1: Africa 2: Asia 3: Australia 4: Europe 5: North America 6: South America

|  |  |  |
| --- | --- | --- |
| Variable name | Variable type | Values/notes |
| ID | Numeric | identity |
| CONTINENT | Numeric | [1, 6] |
| COUNT\_ CONTINENT | Numeric | Number of the cells occupied by the species in year X |
| SPECIES\_ID | Varchar | Species id within a simulation |
| YEAR | Numeric | [100, 1200000] by =100 |
| sp\_id | Numeric | Seed id |
| niche\_breadth | Varchar | large/narrow |
| dispersal\_ability | Varchar | good/poor |
| speciation\_year | Numeric | 10000 only |
| extinction\_threshold | Numeric | 0/5 |



Continents for R10



Continents for R11

**6. Table name ‘distribution’. Record count = 40,230,244**

This table is used to store the distribution (number of cells) of each species (including the original and new species) within temperate or tropic. (18 degree annual temperature). 0 is temperate. 1 is tropic.

|  |  |  |
| --- | --- | --- |
| Variable name | Variable type | Values/notes |
| ID | Numeric | identity |
| COUNT\_ DISTRIBUTION | Numeric | Number of the cells occupied by the species in year X |
| SPECIES\_ID | Varchar | Species id within a simulation |
| YEAR | Numeric | [100, 1200000] by =100 |
| sp\_id | Numeric | Seed id |
| niche\_breadth | Varchar | large/narrow |
| dispersal\_ability | Varchar | good/poor |
| speciation\_year | Numeric | 10000 only |
| extinction\_threshold | Numeric | 0/5 |

**7. Table name ‘****niche\_breadth’. Record count = 40,238,164**

This table is used to store niche breadth (FN and RN) of every species appeared in the simulations.

|  |  |  |
| --- | --- | --- |
| Variable name | Variable type | Values/notes |
| ID | Numeric | identity |
| FN\_MIN\_TEMP\_MIN | Numeric | niche breadth of a given species |
| FN\_MIN\_TEMP\_MAX | Numeric | niche breadth of a given species |
| FN\_MAX\_TEMP\_MIN | Numeric | niche breadth of a given species |
| FN\_MAX\_TEMP\_MAX | Numeric | niche breadth of a given species |
| FN\_MAX\_PREC\_MIN | Numeric | niche breadth of a given species |
| FN\_ MAX\_PREC\_MIN | Numeric | niche breadth of a given species |
| RN\_MIN\_TEMP\_MIN | Numeric | niche breadth of a given species |
| RN\_MIN\_TEMP\_MAX | Numeric | niche breadth of a given species |
| RN\_MAX\_TEMP\_MIN | Numeric | niche breadth of a given species |
| RN\_MAX\_TEMP\_MAX | Numeric | niche breadth of a given species |
| RN\_MAX\_PREC\_MIN | Numeric | niche breadth of a given species |
| RN\_ MAX\_PREC\_MIN | Numeric | niche breadth of a given species |
| SPECIES\_ID | Varchar | Species id within a simulation |
| YEAR | Numeric | [100, 1200000] by =100 |
| sp\_id | Numeric | Seed id |
| niche\_breadth | Varchar | large/narrow |
| dispersal\_ability | Varchar | good/poor |
| speciation\_year | Numeric | 10000 only |
| extinction\_threshold | Numeric | 0/5 |

**8. Table name ‘speciation\_extinction\_in\_combined\_environments\_X’.**

This table is used to store the number of speciation and extinction events per environmental combination (10 degree min temp by 10 degree max temp by 10 mm prec, need to double check).

359,187,091 speciation\_extinction\_in\_combined\_environments\_convex\_10\_0

350,519,642 speciation\_extinction\_in\_combined\_environments\_convex\_10\_5

152,904,665 speciation\_extinction\_in\_combined\_environments\_origin\_10\_0

100,881,831 speciation\_extinction\_in\_combined\_environments\_origin\_10\_5

152,232,499 speciation\_extinction\_in\_combined\_environments\_town\_10\_0

149,619,816 speciation\_extinction\_in\_combined\_environments\_town\_10\_5

|  |  |  |
| --- | --- | --- |
| Variable name | Variable type | Values/notes |
| ID | Numeric | identity |
| Band\_T\_MIN | Numeric | band of minimal temperature |
| Band\_T\_MAX | Numeric | band of maximal temperature |
| Band\_P\_MAX | Numeric | band of maximal precipitation |
| Value | Numeric | number of events |
| SE | Varchar | S or E. S=speciation E=Extinction |
| sp\_id | Numeric | Seed id |
| niche\_breadth | Varchar | large/narrow |
| dispersal\_ability | Varchar | good/poor |
| speciation\_year | Numeric | 10000 only |
| extinction\_threshold | Numeric | 0/5 |