Brief methological explanation and preliminary results

Here we summarise all the critical steps conducted to analise bee preferences.

Filtering criteria

1) Filter records above 1985 and 1987 (this is decided based on the GIS data available) for Europe and USA respectively.
2) Filter by unique capture event.
3) Filter by minimum of 3 decimals on coordinates.
4) Filter number of levels per species (minimum sample size, 100 occurrences).
5) Filter by wide geographical distribution.

Land use classification

1) Europe

Link with land European use classes

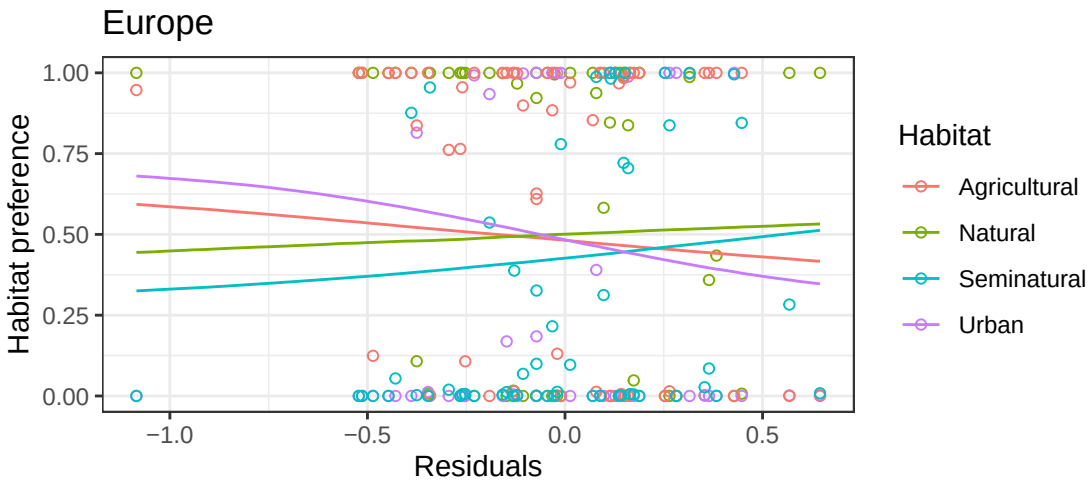
https://land.copernicus.eu/user-corner/technical-library/corine-land-cover-nomenclature-guidelines/html/index-clc-512.html

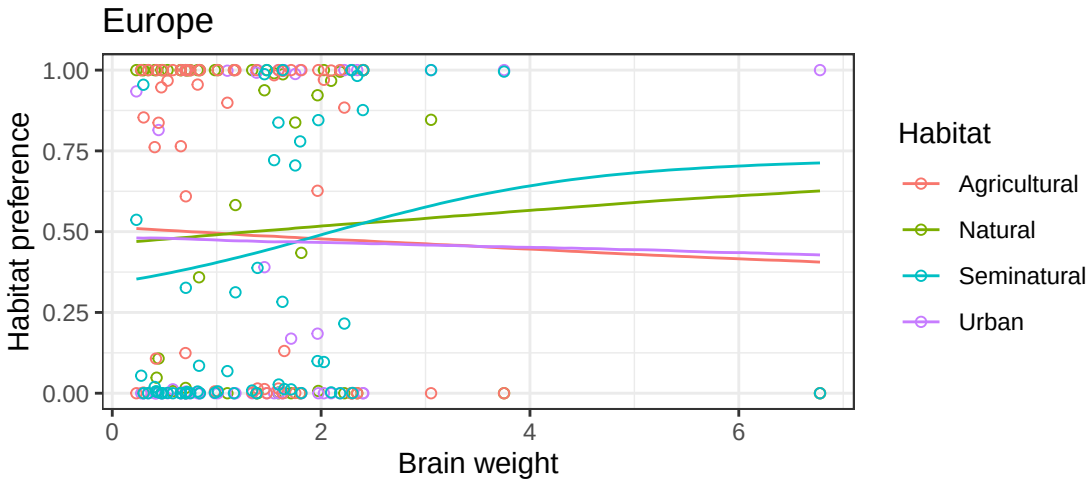
Land_use	n_rows	Cover_names
Industrial or commercial units	127500	Urban
Permanently irrigated land	71330	Seminatural
Annual crops associated with permanent crops	56860	Agricultural
Land principally occupied by agriculture, with significant areas of natural vegetation	30602	Seminatural
Coniferous forest	17812	Natural
Agro-forestry areas	14249	Seminatural
Road and rail networks and associated land	11380	Seminatural
Non-irrigated arable land	11097	Agricultural
Natural grasslands	9370	Natural
Mixed forest	9036	Natural
Sport and leisure facilities	6659	Seminatural
Moors and heathland	3148	Natural
Sclerophyllous vegetation	2877	Natural
Coastal lagoons	2733	Discard
Port areas	2522	Urban
Fruit trees and berry plantations	2252	Agricultural
Peat bogs	2207	Natural
Discontinuous urban fabric	2029	Seminatural
Water bodies	1846	Discard
Beaches, dunes, sands	1459	Discard
Dump sites	1411	Discard
Olive groves	1296	Agricultural
Salt marshes	1247	Natural
Green urban areas	973	Urban
Bare rocks	752	Discard
Water courses	632	Discard
Sea and ocean	572	Discard
Mineral extraction sites	492	Discard
Airports	487	Urban
Construction sites	305	Urban
Salines	208	Seminatural
Burnt areas	35	Discard
Estuaries	22	Discard
Sparsely vegetated areas	10	Natural

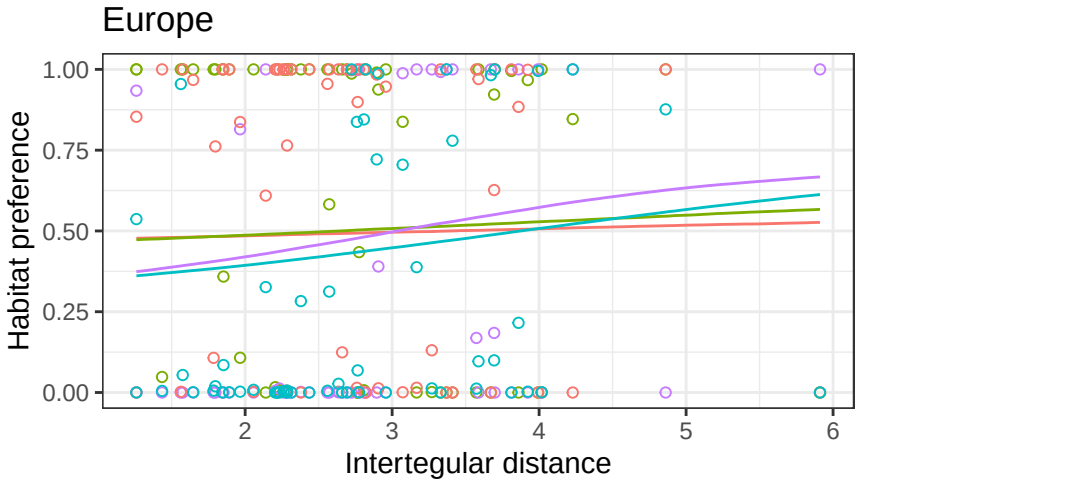
Check final levels per category:

Cover_names	Levels
Seminatural	134428
Urban	133816
Agricultural	71505
Natural	45707

Then, we calculate preferences and model the data.







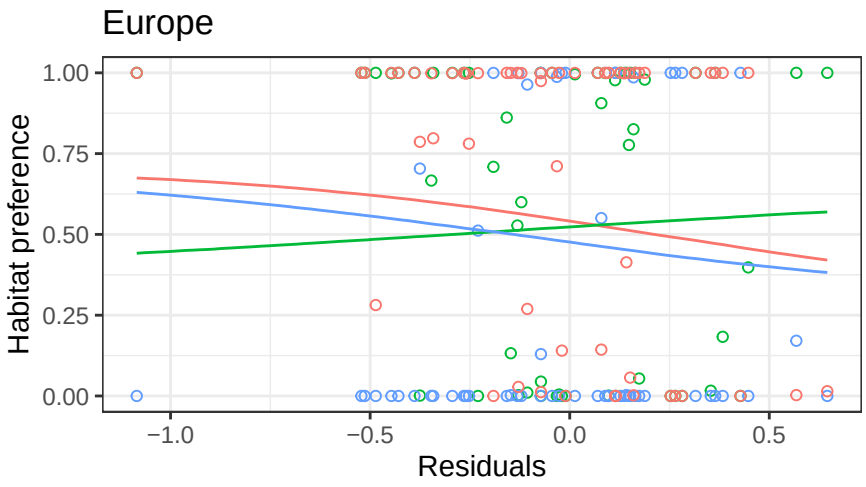
Now we make more strict land use classifications

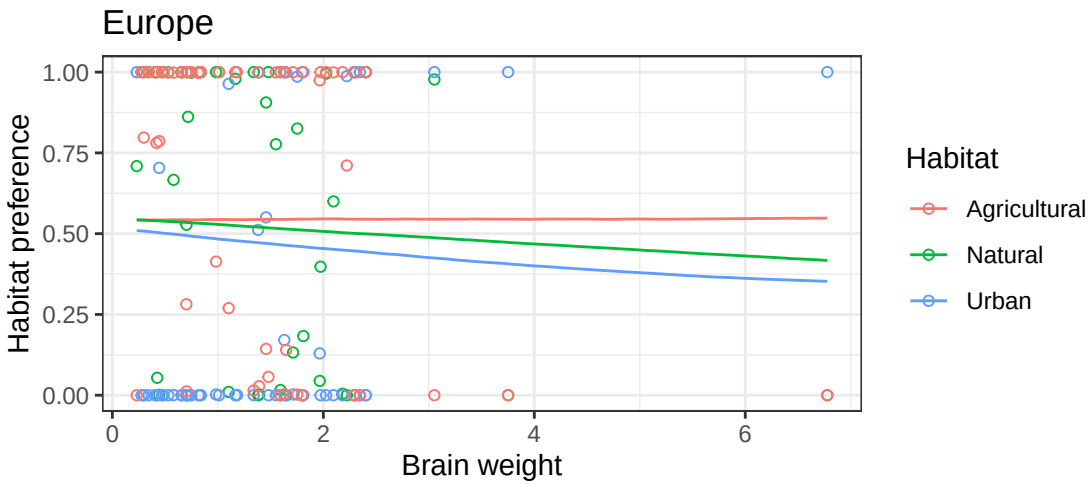
Check final levels per category:

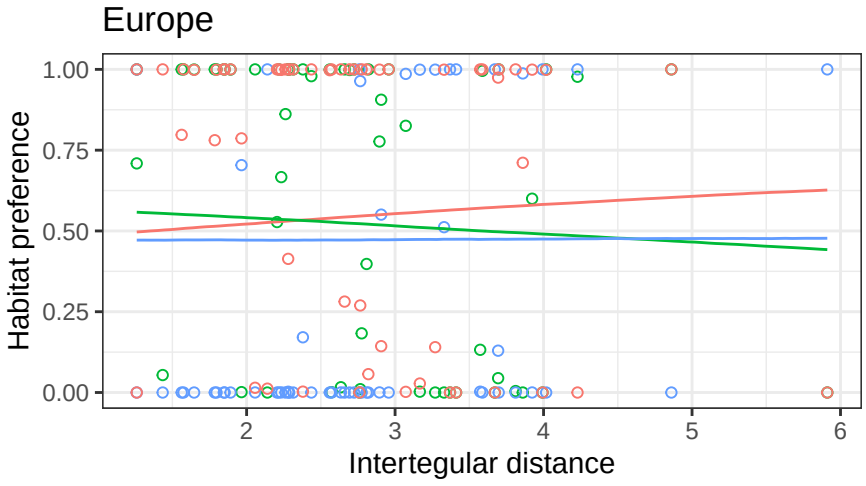
Land_use	n_rows	Cover_names
Industrial or commercial units	127500	Urban
Permanently irrigated land	71330	Agricultural
Annual crops associated with permanent crops	56860	Agricultural
Land principally occupied by agriculture, with significant areas of natural vegetation	30602	Agricultural
Coniferous forest	17812	Natural
Agro-forestry areas	14249	Discard
Road and rail networks and associated land	11380	Discard
Non-irrigated arable land	11097	Agricultural
Natural grasslands	9370	Discard
Mixed forest	9036	Natural
Sport and leisure facilities	6659	Discard
Moors and heathland	3148	Discard
Sclerophyllous vegetation	2877	Discard
Coastal lagoons	2733	Discard
Port areas	2522	Discard
Fruit trees and berry plantations	2252	Agricultural
Peat bogs	2207	Discard
Discontinuous urban fabric	2029	Discard
Water bodies	1846	Discard
Beaches, dunes, sands	1459	Discard
Dump sites	1411	Discard
Olive groves	1296	Discard
Salt marshes	1247	Discard
Green urban areas	973	Discard
Bare rocks	752	Discard
Water courses	632	Discard
Sea and ocean	572	Discard
Mineral extraction sites	492	Discard
Airports	487	Discard
Construction sites	305	Discard
Salines	208	Discard
Burnt areas	35	Discard
Estuaries	22	Discard
Sparsely vegetated areas	10	Discard

Cover_names	Levels
Agricultural	172141
Natural	26848
Urban	127500

And we model the data with this categories:


Europe
Habitat preference
1.00
0.75
0.50
0.25
0.00
−1.0
−0.5
0.0
0.5
Residuals


Europe
Habitat preference
1.00
0.75
0.50
0.25
0.00
0
2
4
6
Brain weight
Habitat
Agricultural
Natural
Urban


Europe
Habitat preference
1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
Intertegular distance

2) USA

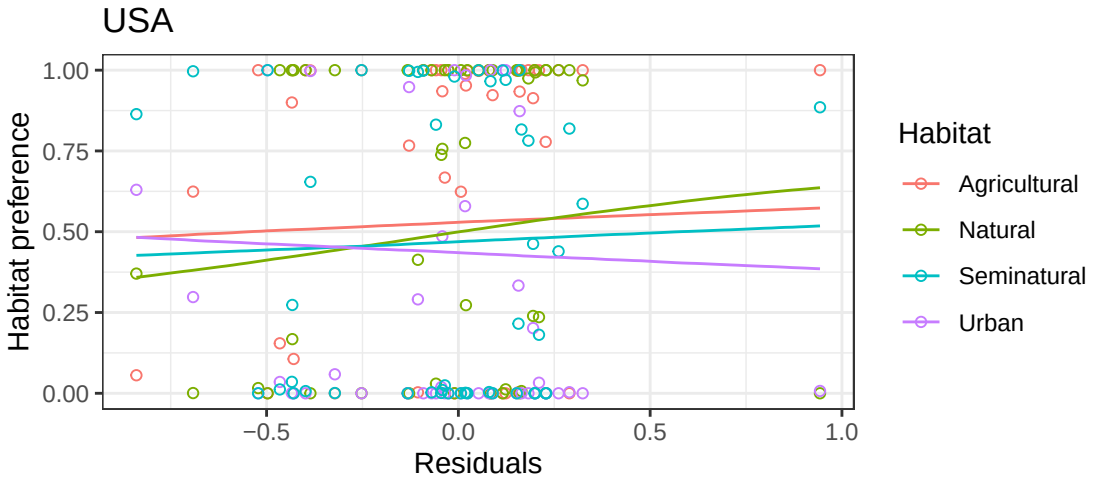
United States land use classification with total number of occurrences within cover type in our dataset

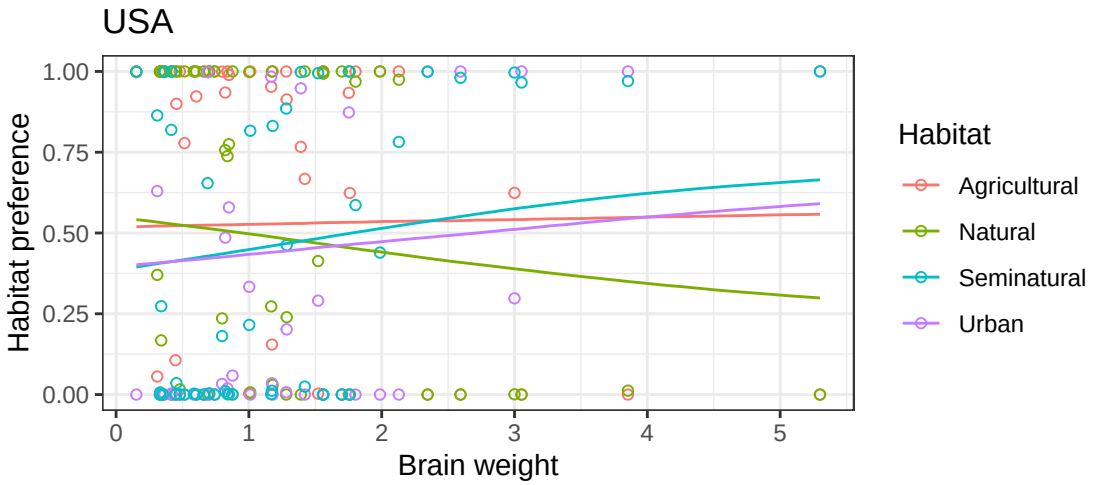
Land_use	n_rows	Cover_names
Developed, Open Space	9814	Seminatural
Developed, Low Intensity	8067	Seminatural
Developed, Medium Intensity	7821	Urban
Hay/Pasture	6931	Agricultural
Deciduous Forest	6629	Natural
Developed, High Intensity	3479	Urban
Mixed Forest	2636	Natural
Cultivated Crops	2287	Agricultural
Woody Wetlands	2218	Natural
Shrub/Scrub	1139	Natural
Evergreen Forest	1026	Natural
Herbaceous	928	Natural
Open Water	762	Discard
Emergent Herbaceous Wetlands	595	Natural
Barren Land	264	Discard

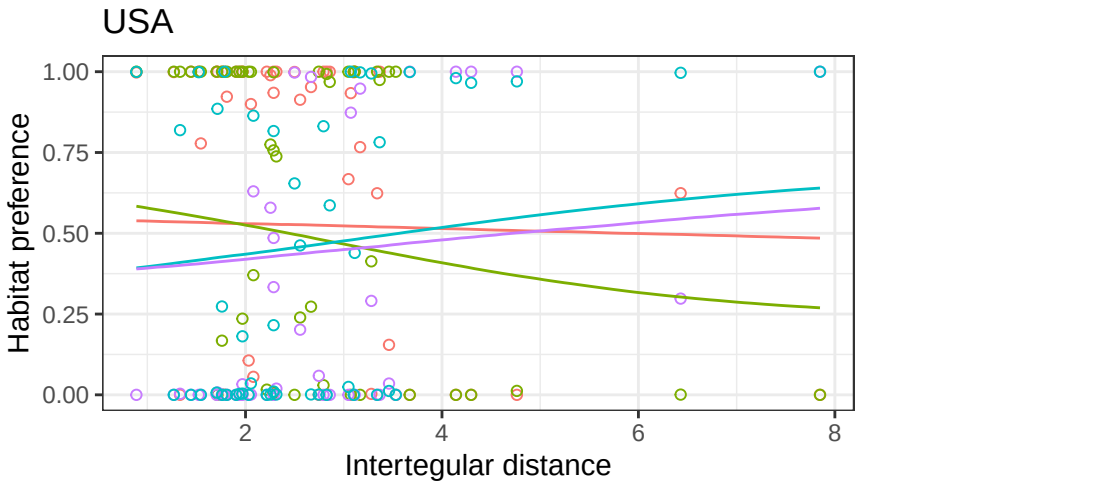
Check final levels per category:

cover.names	n_rows
Seminatural	17881
Natural	15171
Urban	11300
Agricultural	9218

Then, we calculate preferences and model the data.







Note: The models have been run with non-informative priors from the brms function but some work on priors can be done... I have explored a bit and the results were quite similar overall but reviewers seem to don't like this approach as one of the goals of Bayesian is to use the prior knowledge of our variables to improve our predictions.

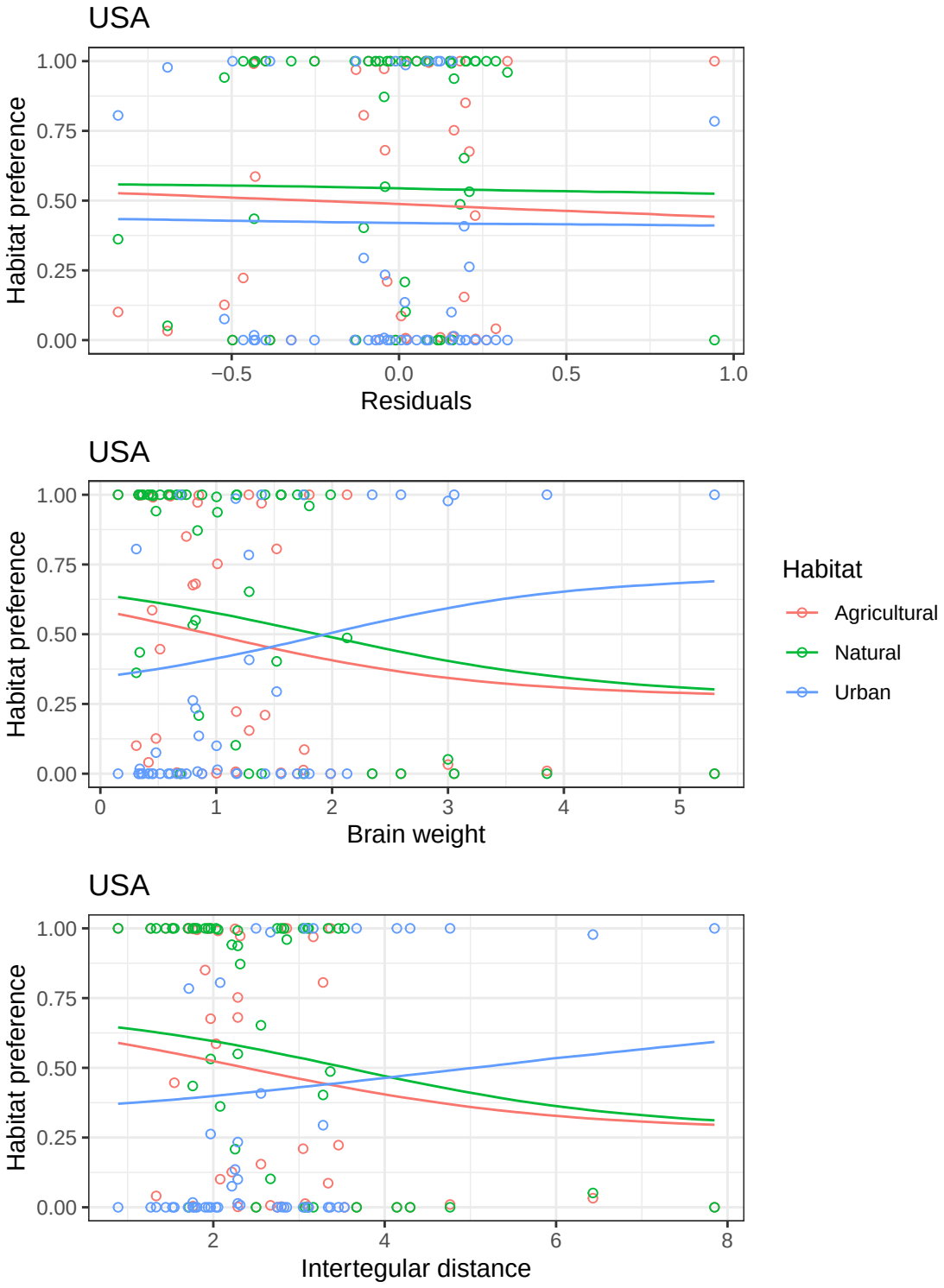
Now we make more strict categories of what is Natural, Urban and Agricultural. See below:

Land_use	n_rows	Cover_names
Developed, Open Space	9814	Discard
Developed, Low Intensity	8067	Discard
Developed, Medium Intensity	7821	Urban
Hay/Pasture	6931	Discard
Deciduous Forest	6629	Natural
Developed, High Intensity	3479	Urban
Mixed Forest	2636	Natural
Cultivated Crops	2287	Agricultural
Woody Wetlands	2218	Natural
Shrub/Scrub	1139	Natural
Evergreen Forest	1026	Natural
Herbaceous	928	Discard
Open Water	762	Discard
Emergent Herbaceous Wetlands	595	Discard
Barren Land	264	Discard

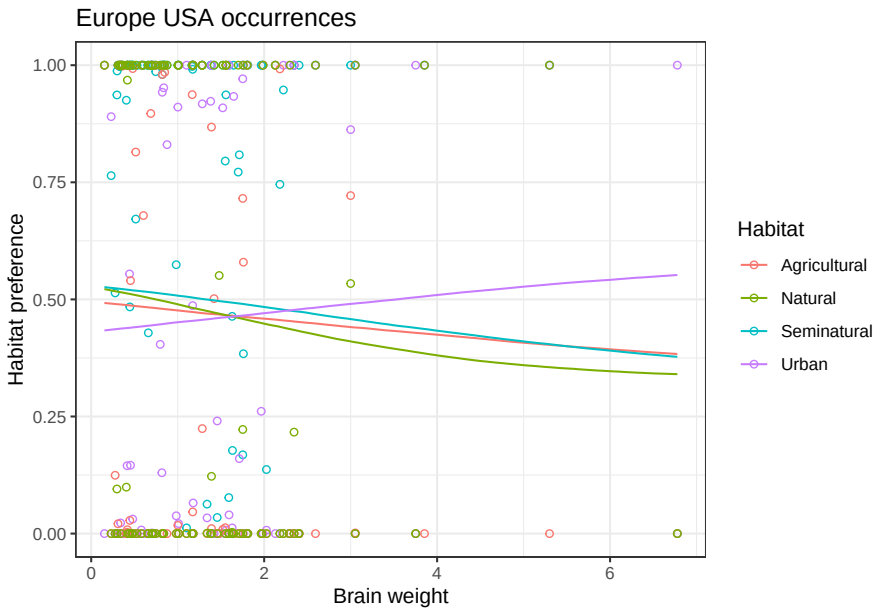
Check number of levels per category:

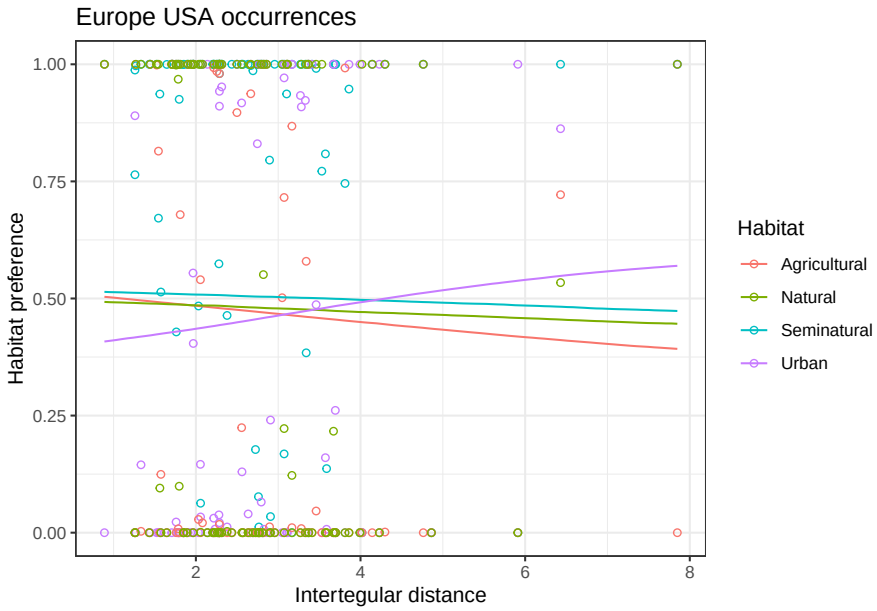
cover.names	n_rows
Natural	13648
Urban	11300
Agricultural	2287

And we model the data with this categories:



2) Finally, we consider Europe and USA together


Europe USA occurrences
Habitat preference
1.00
0.75
0.50
0.25
0.00
0
2
4
6
Brain weight
Habitat
Agricultural
Natural
Seminatural
Urban


Europe USA occurrences
Habitat preference
1.00
0.75
0.50
0.25
0.00
2
4
6
8
Intertegular distance
Habitat
Agricultural
Natural
Seminatural
Urban