Internship progress

Abdourahmane Diallo

2024-02-20

1 Plan

- Setting
- Database import
- Database exploration
- Earthworms data
- Soil data extraction
- Climate data extraction
- To do next

2 Setting

2.1 Packages

- Code

2.2 Functions

- Code

3 Database import

- Import of database **LandWorm_dataset_site.xlsx** version **V1.5** (february 12, 2024)

▶ Code

- The database contains **7919** rows and **601** columns
- Convert columns to correct format

▶ Code

4 Database exploration

- CR = Completion rate

4.1 Complete columns

▶ Code

	Variables	CR
74	ID	100%

▶ Code

```
[1] "AgrInnov_P001_2013_NA_NA" "AgrInnov_P002_2013_NA_NA"
[3] "AgrInnov_P003_2013_NA_NA" "AgrInnov_P005_2013_NA_NA"
[5] "AgrInnov_P006_2013_NA_NA" "AgrInnov_P007_2013_NA_NA"
```

4.2 Non-complete columns

	Variables	CR
75	Protocole	99.9%
1	Programme	96.7%
2	Annee	96.7%
4	ID_Site	96.7%
8	gps_x	93.2%
9	gps_y	93.2%
77	Total_AB	92.9%
79	AB_AD	85.7%

	Variables	CR
80	AB_JV	85.7%
81	AB_SA	85.7%
98	AB_Aporrectodea_rosea	83.4%
83	AB_Allolobophora_chlorotica_chlorotica	81.2%
12	clcm_lvl2	81.1%
144	BM_Aporrectodea_rosea	81%
117	AB_Lumbricus_terrestris	79.2%
108	AB_Lumbricus_castaneus	78.8%
120	AB_Octolasion_cyaneum	78.3%
166	BM_Octolasion_cyaneum	77.9%

	Variables	CR
88	AB_Aporrectodea_caliginosa	77%
11	clcm_lvl1	76.8%
92	AB_Aporrectodea_giardi	76.4%
86	AB_Allolobophora_icterica	76.2%
154	BM_Lumbricus_castaneus	76%
132	BM_Allolobophora_icterica	75.1%
126	AB_Satchellius_mammalis	74.8%
138	BM_Aporrectodea_giardi	74.7%
163	BM_Lumbricus_terrestris	74.6%
172	BM_Satchellius_mammalis	73.9%

	Variables	CR
13	clcm_lvl3	73.2%
97	AB_Aporrectodea_nocturna	72.3%
95	AB_Aporrectodea_longa_longa	71.1%
143	BM_Aporrectodea_nocturna	70.9%
119	AB_Murchieona_muldali	70.4%
165	BM_Murchieona_muldali	70.2%
129	BM_Allolobophora_chlorotica_chlorotica	70%
90	AB_Aporrectodea_caliginosa_meridionalis	68.9%
134	BM_Aporrectodea_caliginosa	68.7%
115	AB_Lumbricus_rubellus_rubellus	67.5%

	Variables	CR
136	BM_Aporrectodea_caliginosa_meridionalis	67.5%
141	BM_Aporrectodea_longa_longa	67.3%
161	BM_Lumbricus_rubellus_rubellus	66.8%
114	AB_Lumbricus_rubellus_castanoides	64.7%
160	BM_Lumbricus_rubellus_castanoides	64.5%
116	AB_Lumbricus_sp	64%
162	BM_Lumbricus_sp	63.4%
100	AB_Aporrectodea_sp	61.5%
106	AB_indéterminable	61.1%
146	BM_Aporrectodea_sp	61.1%

	Variables	CR
152	BM_indéterminable	61.1%
122	AB_Octolasion_sp	60.4%
168	BM_Octolasion_sp	60.4%
112	AB_Lumbricus_friendi	60.2%
111	AB_Lumbricus_festivus	58.2%
157	BM_Lumbricus_festivus	58.1%
91	AB_Aporrectodea_cupulifera	57.9%
137	BM_Aporrectodea_cupulifera	57.9%
158	BM_Lumbricus_friendi	57.8%
85	AB_Allolobophora_chlorotica_postepheba	56.8%

	Variables	CR
176	AB_Eiseniella_tetraedra	56.6%
131	BM_Allolobophora_chlorotica_postepheba	56.5%
32	clay	55.8%
180	BM_Eiseniella_tetraedra	55.4%
76	Code_Parcelle	54.8%
110	AB_Lumbricus_centralis	54.1%
103	AB_Dendrobaena_sp	54%
149	BM_Dendrobaena_sp	54%
156	BM_Lumbricus_centralis	53.4%
26	fine_sand	49.9%

	Variables	CR
27	coarse_sand	49.9%
29	fine_silt	49.9%
30	coarse_silt	49.9%
16	ph_eau	48.5%
78	BM_tot	48.5%
93	AB_Aporrectodea_indéterminable	48.5%
139	BM_Aporrectodea_indéterminable	48.3%
121	AB_Octolasion_lacteum_lacteum	48.1%
167	BM_Octolasion_lacteum_lacteum	48.1%
105	AB_Eisenia_fetida	48%

	Variables	CR
151	BM_Eisenia_fetida	47.9%
109	AB_Lumbricus_castaneus_disjonctus	47.3%
155	BM_Lumbricus_castaneus_disjonctus	47.1%
21	om	46.8%
240	AB_Microscolex_phosphoreus	46.7%
242	BM_Microscolex_phosphoreus	46.7%
96	AB_Aporrectodea_longa_ripicola	43.7%
142	BM_Aporrectodea_longa_ripicola	43.6%
187	AB_Eisenia_andrei	43.4%
194	BM_Eisenia_andrei	43.4%

	Variables	CR
118	AB_Microscolex_dubius	42.4%
164	BM_Microscolex_dubius	42.4%
19	n_tot	40.9%
18	c_org	40.1%
94	AB_Aporrectodea_longa	38.9%
140	BM_Aporrectodea_longa	38.7%
87	AB_Allolobophora_sp	36.5%
133	BM_Allolobophora_sp	36.3%
186	AB_Dendrodrilus_rubidus	35.8%
193	BM_Dendrodrilus_rubidus	35.8%

	Variables	CR
308	AB_Lumbricus_friendi_lineatus	35.5%
315	BM_Lumbricus_friendi_lineatus	35.4%
128	AB_Scherotheca_sp	34.3%
174	BM_Scherotheca_sp	34.3%
239	AB_Aporrectodea_limicola	34.1%
241	BM_Aporrectodea_limicola	34.1%
125	AB_Prosellodrilus_indéterminable	33.9%
171	BM_Prosellodrilus_indéterminable	33.8%
309	AB_Octodrilus_complanatus	32.9%
204	AB_Scherotheca_savignyi_indéterminable	32.8%

	Variables	CR
208	BM_Scherotheca_savignyi_indéterminable	32.8%
316	BM_Octodrilus_complanatus	32.5%
101	AB_Aporrectodea_trapezoides	31.9%
14	land_cover_detail	31.7%
147	BM_Aporrectodea_trapezoides	31.7%
197	AB_Eisenia_veneta	31.2%
199	BM_Eisenia_veneta	31.2%
107	AB_Indéterminable	31%
153	BM_Indéterminable	31%
472	AB_Allolobophora_burgondiae	30.6%

	Variables	CR
10	Altitude	30.5%
474	BM_Allolobophora_burgondiae	30.4%
123	AB_Prosellodrilus_amplisetosus_amplisetosus	30%
169	BM_Prosellodrilus_amplisetosus_amplisetosus	30%
473	AB_Scherotheca_savignyi_savignyi	30%
475	BM_Scherotheca_savignyi_savignyi	30%
185	AB_Dendrobaena_octaedra	29.2%
192	BM_Dendrobaena_octaedra	28.9%
198	AB_Proctodrilus_antipai	28.5%
200	BM_Proctodrilus_antipai	28.3%

	Variables	CR
102	AB_Aporrectodea_tuberculata	28.1%
148	BM_Aporrectodea_tuberculata	28.1%
335	AB_Bimastos_eiseni	26.9%
370	BM_Bimastos_eiseni	26.8%
219	AB_Lumbricus_rubellus_friendoides	26.4%
477	AB_Scherotheca_aquitana	26.4%
479	BM_Scherotheca_aquitana	26.4%
232	BM_Lumbricus_rubellus_friendoides	26.3%
462	AB_Eisenia_fetida_indéterminable	25.2%
464	BM_Eisenia_fetida_indéterminable	25.2%

	Variables	CR
28	sand	23.1%
31	silt	23.1%
33	type_tillage	21.4%
476	AB_Aporrectodea_rubra_acidicola	20.3%
478	BM_Aporrectodea_rubra_acidicola	20.3%
5	Modalite	19.8%
113	AB_Lumbricus_meliboeus	17.7%
159	BM_Lumbricus_meliboeus	17.7%
6	Bloc	17.3%
3	Date_Prelevement	17%

	Variables	CR
104	AB_Dendrodrilus_rubidus_subrubicundus	17%
150	BM_Dendrodrilus_rubidus_subrubicundus	17%
89	AB_Aporrectodea_caliginosa_indéterminable	16.8%
135	BM_Aporrectodea_caliginosa_indéterminable	16.3%
184	AB_Dendrobaena_attemsi	16.2%
191	BM_Dendrobaena_attemsi	16.2%
189	AB_Proctodrilus_antipai_antipai	15.8%
196	BM_Proctodrilus_antipai_antipai	15.8%
468	AB_Prosellodrilus_occidentalis_occidentalis	15.1%
471	BM_Prosellodrilus_occidentalis_occidentalis	15.1%

	Variables	CR
82	AB_STAD_X	14.9%
480	AB_Hemigastrodrilus_monicae	14.9%
484	AB_Prosellodrilus_praticola	14.9%
486	BM_Hemigastrodrilus_monicae	14.9%
490	BM_Prosellodrilus_praticola	14.9%
38	fertilisation	14.8%
217	AB_Prosellodrilus_fragilis_fragilis	14.7%
230	BM_Prosellodrilus_fragilis_fragilis	14.7%
7	postal_code	14.5%
124	AB_Prosellodrilus_fragilis_indéterminable	13.4%

	Variables	CR
170	BM_Prosellodrilus_fragilis_indéterminable	13.4%
20	c/n	13%
183	AB_Avelona_ligra	12.2%
190	BM_Avelona_ligra	12.1%
17	c_tot	11.3%
483	AB_Prosellodrilus_amplisetosus	10.8%
489	BM_Prosellodrilus_amplisetosus	10.8%
372	AB_Allolobophora.chlorotica	10.7%
373	AB_Allolobophora.cupulifera	10.7%
374	AB_Allolobophoridella.eiseni	10.7%

	Variables	CR
375	AB_Aporrectodea	10.7%
376	AB_Aporrectodea.caliginosa	10.7%
377	AB_Aporrectodea.giardi	10.7%
378	AB_Aporrectodea.icterica	10.7%
379	AB_Aporrectodea.longa	10.7%
380	AB_Aporrectodea.rosea	10.7%
381	AB_Aporrectodea.rubra	10.7%
382	AB_Aporrectodea.terrestris	10.7%
383	AB_Aporrectodea.trapezoides	10.7%
384	AB_Bimastos.eiseni	10.7%

	Variables	CR
385	AB_Dendrobaena	10.7%
386	AB_Dendrobaena.alpina	10.7%
387	AB_Dendrobaena.alpina.juvenile	10.7%
388	AB_Dendrobaena.mammalis	10.7%
389	AB_Dendrobaena.octaedra	10.7%
390	AB_Dendrodrilus.rubidus	10.7%
391	AB_Dendrodrilus.subrubicundus	10.7%
392	AB_Eiseniella.tetraedra.	10.7%
393	AB_Lumbricidae	10.7%
394	AB_Lumbricus	10.7%

	Variables	CR
395	AB_Lumbricus.castaneus	10.7%
396	AB_Lumbricus.centralis	10.7%
397	AB_Lumbricus.festivus	10.7%
398	AB_Lumbricus.friendi.friendi	10.7%
399	AB_Lumbricus.meliboeus	10.7%
400	AB_Lumbricus.rubellus.castanoides	10.7%
401	AB_Lumbricus.rubellus.rubellus	10.7%
402	AB_Lumbricus.terrestris	10.7%
403	AB_Microscolex	10.7%
404	AB_Microscolex.dubius	10.7%

	Variables	CR
405	AB_Microscolex.phosphoreus	10.7%
406	AB_Murchieona.minuscula	10.7%
407	AB_Octolasion	10.7%
408	AB_Octolasion.cyaneum	10.7%
409	AB_Octolasion.lacteum	10.7%
410	AB_Octolasion.tyrtaeum	10.7%
411	AB_Oligochaeta	10.7%
412	AB_Prosellodrilus	10.7%
413	AB_Satchellius.mammalis	10.7%
414	AB_Scherotheca	10.7%

	Variables	CR
415	AB_Scherotheca.gigas	10.7%
416	AB_Scherotheca.rhodana	10.7%
418	BM_Allolobophora.cupulifera	10.7%
419	BM_Allolobophoridella.eiseni	10.7%
420	BM_Aporrectodea	10.7%
422	BM_Aporrectodea.giardi	10.7%
426	BM_Aporrectodea.rubra	10.7%
427	BM_Aporrectodea.terrestris	10.7%
428	BM_Aporrectodea.trapezoides	10.7%
429	BM_Bimastos.eiseni	10.7%

	Variables	CR
430	BM_Dendrobaena	10.7%
431	BM_Dendrobaena.alpina	10.7%
432	BM_Dendrobaena.alpina.juvenile	10.7%
433	BM_Dendrobaena.mammalis	10.7%
434	BM_Dendrobaena.octaedra	10.7%
436	BM_Dendrodrilus.subrubicundus	10.7%
437	BM_Eiseniella.tetraedra.	10.7%
441	BM_Lumbricus.centralis	10.7%
442	BM_Lumbricus.festivus	10.7%
443	BM_Lumbricus.friendi.friendi	10.7%

	Variables	CR
444	BM_Lumbricus.meliboeus	10.7%
448	BM_Microscolex	10.7%
449	BM_Microscolex.dubius	10.7%
450	BM_Microscolex.phosphoreus	10.7%
451	BM_Murchieona.minuscula	10.7%
452	BM_Octolasion	10.7%
453	BM_Octolasion.cyaneum	10.7%
454	BM_Octolasion.lacteum	10.7%
455	BM_Octolasion.tyrtaeum	10.7%
456	BM_Oligochaeta	10.7%

	Variables	CR
457	BM_Prosellodrilus	10.7%
458	BM_Satchellius.mammalis	10.7%
459	BM_Scherotheca	10.7%
460	BM_Scherotheca.gigas	10.7%
461	BM_Scherotheca.rhodana	10.7%
424	BM_Aporrectodea.longa	10.6%
435	BM_Dendrodrilus.rubidus	10.6%
439	BM_Lumbricus	10.6%
445	BM_Lumbricus.rubellus.castanoides	10.5%
481	AB_Octodrilus_indéterminable	10.5%

	Variables	CR
482	AB_PAX	10.5%
485	AB_Scherotheca_Porotheca	10.5%
487	BM_Octodrilus_indéterminable	10.5%
488	BM_PAX	10.5%
491	BM_Scherotheca_Porotheca	10.5%
440	BM_Lumbricus.castaneus	10.4%
446	BM_Lumbricus.rubellus.rubellus	10.4%
46	herbicide_freq	10.3%
423	BM_Aporrectodea.icterica	10.2%
237	AB_Haplotaxis_sp	10%

	Variables	CR
238	BM_Haplotaxis_sp	10%
447	BM_Lumbricus.terrestris	9.9%
417	BM_Allolobophora.chlorotica	9.8%
425	BM_Aporrectodea.rosea	9.2%
23	cu_EDTA	8.7%
44	fungicide_freq	8.6%
421	BM_Aporrectodea.caliginosa	7.8%
45	insecticide_freq	7.7%
22	cu_tot	7.6%
41	ferti_orga_product	7.6%

	Variables	CR
253	AB_Dendrobaena_cognettii	7.1%
254	BM_Dendrobaena_cognettii	7.1%
356	AB_Pheretima_indéterminable	7.1%
363	BM_Pheretima_indéterminable	7.1%
68	herbage_use	6.8%
15	ph_kcl	6.4%
34	tillage_depth	5.8%
353	AB_Dendrobaena_hortensis	5.8%
360	BM_Dendrobaena_hortensis	5.8%
438	BM_Lumbricidae	5.7%

	Variables	CR
39	ferti_min_product	5.6%
47	molluscicide_freq	5.6%
84	AB_Allolobophora_chlorotica_indéterminable	5.6%
130	BM_Allolobophora_chlorotica_indéterminable	5.6%
355	AB_Flabellodrilus_bartolii	5.6%
362	BM_Flabellodrilus_bartolii	5.6%
178	AB_Scherotheca_rhodana	5.5%
182	BM_Scherotheca_rhodana	5.5%
305	AB_Ethnodrilus_lydiae	5%
311	AB_Zophoscolex_graffi	5%

	Variables	CR
312	BM_Ethnodrilus_lydiae	5%
318	BM_Zophoscolex_graffi	5%
466	AB_Microscolex_sp	4.9%
469	BM_Microscolex_sp	4.9%
350	AB_Aporrectodea_georgii	4.8%
367	BM_Aporrectodea_georgii	4.8%
58	rotation_plant_div	4.7%
243	AB_Aporrectodea_balisa	4.7%
244	BM_Aporrectodea_balisa	4.7%
57	crop_rotation_yr	4.6%

	Variables	CR
467	AB_Pheritima_Diffringens	4.6%
470	BM_Pheritima_Diffringens	4.6%
73	grassland_type	4.5%
333	AB_Scherotheca_portonana	4.5%
334	AB_Scherotheca_brevisella	4.5%
337	AB_Dendrobaena_byblica	4.5%
341	AB_Lumbricus_klarae	4.4%
345	AB_Ethnodrilus_aveli	4.4%
492	AB_Adult	4.4%
493	AB_cocon	4.4%

	Variables	CR
494	AB_indéterminé	4.4%
495	AB_Juvenile	4.4%
496	AB_Sub.adult	4.4%
497	AB_Aporrectodea_arverna	4.4%
498	AB_Aporrectodea_caliginosa_alternisetosus	4.4%
499	AB_Aporrectodea_caliginosa_caliginosa	4.4%
500	AB_Aporrectodea_ripicola	4.4%
501	AB_Aporrectodea_rubra_rubra	4.4%
502	AB_Aporrectodea_voconca	4.4%
503	AB_Boucheona_corbierensis	4.4%

	Variables	CR
504	AB_Ethnodrilus_setusmonsanus	4.4%
505	AB_Flabellodrilus_luberonensis	4.4%
506	AB_Gatesona_chaetophora	4.4%
507	AB_Gatesona_lablacherensis	4.4%
508	AB_Haplotaxis_gordioides	4.4%
509	AB_Helodrilus_oculatus	4.4%
510	AB_Indéterminé	4.4%
511	AB_Kritodrilus_calarensis	4.4%
512	AB_Lucquesia_tiginosa	4.4%
513	AB_Lumbricus_bouchei	4.4%

	Variables	CR
514	AB_Lumbricus_castaneus_castaneus	4.4%
515	AB_Lumbricus_improvisus	4.4%
516	AB_Lumbricus_rubellus_castaneoides	4.4%
517	AB_Murchieona_minuscula	4.4%
518	AB_Octolasium_lacteum	4.4%
519	AB_Octolasium_tyrtaeum	4.4%
520	AB_Panoniona_satchelli	4.4%
521	AB_Proctodrilus_tuberculatus	4.4%
522	AB_Prosellodrilus_alatus	4.4%
523	AB_Prosellodrilus_fragilis	4.4%

	Variables	CR
524	AB_Prosellodrilus_polythecosus	4.4%
525	AB_Prosellodrilus_pyrenaicus_pyrenaicus	4.4%
526	AB_Scherotheca_altarocca	4.4%
527	AB_Scherotheca_aquitania	4.4%
528	AB_Scherotheca_betharramensis	4.4%
529	AB_Scherotheca_capcorsana	4.4%
530	AB_Scherotheca_dugesi	4.4%
531	AB_Scherotheca_gigas_gigas	4.4%
532	AB_Scherotheca_minor_minor	4.4%
533	AB_Scherotheca_monspessulensis_idica	4.4%

	Variables	CR
534	AB_Scherotheca_monspessulensis_monspessulensis	4.4%
535	AB_Scherotheca_sanaryensis	4.4%
536	AB_Scherotheca_savignyi	4.4%
537	AB_Scherotheca_sp.	4.4%
538	AB_Scherotheca_trezencensis	4.4%
539	AB_Vignysa_callasensis	4.4%
540	AB_Vignysa_teres	4.4%
541	AB_Vosgesia_zicsii	4.4%
542	AB_Zophoscolex_atlanticus	4.4%
543	AB_Zophoscolex_micellus	4.4%

	Variables	CR
544	BM_Aporrectodea_arverna	4.4%
545	BM_Aporrectodea_caliginosa_alternisetosus	4.4%
546	BM_Aporrectodea_caliginosa_caliginosa	4.4%
547	BM_Aporrectodea_ripicola	4.4%
548	BM_Aporrectodea_rubra_rubra	4.4%
549	BM_Aporrectodea_voconca	4.4%
550	BM_Boucheona_corbierensis	4.4%
551	BM_Dendrobaena_byblica	4.4%
552	BM_Ethnodrilus_aveli	4.4%
553	BM_Ethnodrilus_setusmonsanus	4.4%

	Variables	CR
554	BM_Flabellodrilus_luberonensis	4.4%
555	BM_Gatesona_chaetophora	4.4%
556	BM_Gatesona_lablacherensis	4.4%
557	BM_Haplotaxis_gordioides	4.4%
558	BM_Helodrilus_oculatus	4.4%
559	BM_Indéterminé	4.4%
560	BM_Kritodrilus_calarensis	4.4%
561	BM_Lucquesia_tiginosa	4.4%
562	BM_Lumbricus_bouchei	4.4%
563	BM_Lumbricus_castaneus_castaneus	4.4%

	Variables	CR
564	BM_Lumbricus_improvisus	4.4%
565	BM_Lumbricus_klarae	4.4%
566	BM_Lumbricus_rubellus_castaneoides	4.4%
567	BM_Murchieona_minuscula	4.4%
568	BM_Octolasium_lacteum	4.4%
569	BM_Octolasium_tyrtaeum	4.4%
570	BM_Panoniona_satchelli	4.4%
571	BM_Proctodrilus_tuberculatus	4.4%
572	BM_Prosellodrilus_alatus	4.4%
573	BM_Prosellodrilus_fragilis	4.4%

	Variables	CR
574	BM_Prosellodrilus_polythecosus	4.4%
575	BM_Prosellodrilus_pyrenaicus_pyrenaicus	4.4%
576	BM_Scherotheca_altarocca	4.4%
577	BM_Scherotheca_aquitania	4.4%
578	BM_Scherotheca_betharramensis	4.4%
579	BM_Scherotheca_brevisella	4.4%
580	BM_Scherotheca_capcorsana	4.4%
581	BM_Scherotheca_dugesi	4.4%
582	BM_Scherotheca_gigas_gigas	4.4%
583	BM_Scherotheca_minor_minor	4.4%

	Variables	CR
584	BM_Scherotheca_monspessulensis_idica	4.4%
585	BM_Scherotheca_monspessulensis_monspessulensis	4.4%
586	BM_Scherotheca_portonana	4.4%
587	BM_Scherotheca_sanaryensis	4.4%
588	BM_Scherotheca_savignyi	4.4%
589	BM_Scherotheca_sp.	4.4%
590	BM_Scherotheca_trezencensis	4.4%
591	BM_Vignysa_callasensis	4.4%
592	BM_Vignysa_teres	4.4%
593	BM_Vosgesia_zicsii	4.4%

	Variables	CR
594	BM_Zophoscolex_atlanticus	4.4%
595	BM_Zophoscolex_micellus	4.4%
188	AB_Octolasion_lacteum	4%
54	total_tfi	3.9%
42	ferti_orga_qtty	3.8%
35	tillage_frequency_intra	3.7%
195	BM_Octolasion_lacteum	3.6%
463	AB_Lumbricus_rubellus_indéterminable	3.6%
465	BM_Lumbricus_rubellus_indéterminable	3.6%
48	nematicide_freq	3.2%

	Variables	CR
51	tfi_herbicide	3.1%
209	Parcelle	3%
236	AB_tot	3%
201	AB_Allolobophora_muldali	2.9%
202	AB_Dendrobaena_mammalis	2.9%
203	AB_Lumbricus_herculeus	2.9%
205	BM_Allolobophora_muldali	2.9%
206	BM_Dendrobaena_mammalis	2.9%
207	BM_Lumbricus_herculeus	2.9%
99	AB_Aporrectodea_rubra	2.8%

	Variables	CR
127	AB_Scherotheca_dinoscolex	2.8%
145	BM_Aporrectodea_rubra	2.8%
173	BM_Scherotheca_dinoscolex	2.8%
60	rotation_grassland	2.4%
71	animal_loading	2.2%
40	ferti_min_qtty	2.1%
70	herb_age	2%
61	crop_residues_management	1.9%
251	AB_Ethnodrilus_zajonci	1.5%
252	BM_Ethnodrilus_zajonci	1.5%

	Variables	CR
259	AB_A.X	1.5%
269	AB_L.A.X	1.5%
271	AB_LR.X	1.5%
272	AB_N.A.X	1.5%
276	BM_A.X	1.5%
286	BM_L.A.X	1.5%
288	BM_LR.X	1.5%
289	BM_N.A.X	1.5%
55	mecanical_weed_control	1.3%
72	trampling_nature	1.3%

	Variables	CR
59	intercrop_div	1.2%
211	AB_Allolobophora_chlorotica	1.2%
293	AB_EA	1.2%
294	AB_EN.X	1.2%
295	AB_EP.X	1.2%
296	AB_L.E.X	1.2%
297	AB_NCTU	1.2%
298	AB_OT	1.2%
299	BM_EA	1.2%
300	BM_EN.X	1.2%

	Variables	CR
301	BM_EP.X	1.2%
302	BM_L.E.X	1.2%
303	BM_NCTU	1.2%
304	BM_OT	1.2%
352	AB_Dendrobaena_alpina_zeugochaeta	1.2%
354	AB_Eisenia_sp	1.2%
357	AB_Prosellodrilus_pyrenaicus	1.2%
358	AB_Scherotheca_nivicola	1.2%
359	BM_Dendrobaena_alpina_zeugochaeta	1.2%
361	BM_Eisenia_sp	1.2%

	Variables	CR
364	BM_Prosellodrilus_pyrenaicus	1.2%
365	BM_Scherotheca_nivicola	1.2%
50	tfi_insecticide	1.1%
69	mowing_frequency_yr	1.1%
175	AB_Aporrectodea_nocturna_nocturna_cistercianus	1.1%
177	AB_Scherotheca_mifuga	1.1%
179	BM_Aporrectodea_nocturna_nocturna_cistercianus	1.1%
181	BM_Scherotheca_mifuga	1.1%
213	AB_Aporrectodea_longa/giardi	1.1%
226	BM_Aporrectodea_longa/giardi	1.1%

	Variables	CR
210	AB_Aporrectodea_icterica	0.8%
36	tillage_frequency_inter	0.7%
215	AB_Lumbricus_friendi/centralis	0.7%
223	BM_Aporrectodea_icterica	0.7%
228	BM_Lumbricus_friendi/centralis	0.7%
224	BM_Allolobophora_chlorotica	0.6%
306	AB_Hemigastrodrilus_monicae_magnus	0.6%
307	AB_Hormogaster_praetiosa	0.6%
310	AB_Scherotheca_corsicana_corsicana	0.6%
313	BM_Hemigastrodrilus_monicae_magnus	0.6%

	Variables	CR
314	BM_Hormogaster_praetiosa	0.6%
317	BM_Scherotheca_corsicana_corsicana	0.6%
600	AB_Octolasion_lacteum_gracile	0.6%
601	BM_Octolasion_lacteum_gracile	0.6%
25	soil_humidity	0.5%
24	soil_temperature	0.4%
260	AB_ACC.X	0.4%
261	AB_An.X	0.4%
262	AB_D.X	0.4%
263	AB_endo.X	0.4%

	Variables	CR
264	AB_Ep.An.X	0.4%
265	AB_Hormogaster.A.X	0.4%
266	AB_Hormogaster.End.X	0.4%
267	AB_Hormogaster.X	0.4%
268	AB_id1961mt1	0.4%
270	AB_L.X	0.4%
273	AB_N.X	0.4%
274	AB_Octodrilus_lisseansis	0.4%
275	AB_P.X	0.4%
277	BM_ACC.X	0.4%

	Variables	CR
278	BM_An.X	0.4%
279	BM_D.X	0.4%
280	BM_endo.X	0.4%
281	BM_Ep.An.X	0.4%
282	BM_Hormogaster.A.X	0.4%
283	BM_Hormogaster.End.X	0.4%
284	BM_Hormogaster.X	0.4%
285	BM_id1961mt1	0.4%
287	BM_L.X	0.4%
290	BM_N.X	0.4%

	Variables	CR
291	BM_Octodrilus_lisseansis	0.4%
292	BM_P.X	0.4%
326	AB_Lumbricus_rubellus	0.4%
366	AB_Panoniona_leoni	0.4%
368	BM_Panoniona_leoni	0.4%
212	AB_A._muldali/rosea	0.3%
225	BM_A._muldali/rosea	0.3%
255	AB_Ep.X1	0.3%
256	AB_Ep.X2	0.3%

4.3 Focus on GPS coordinates

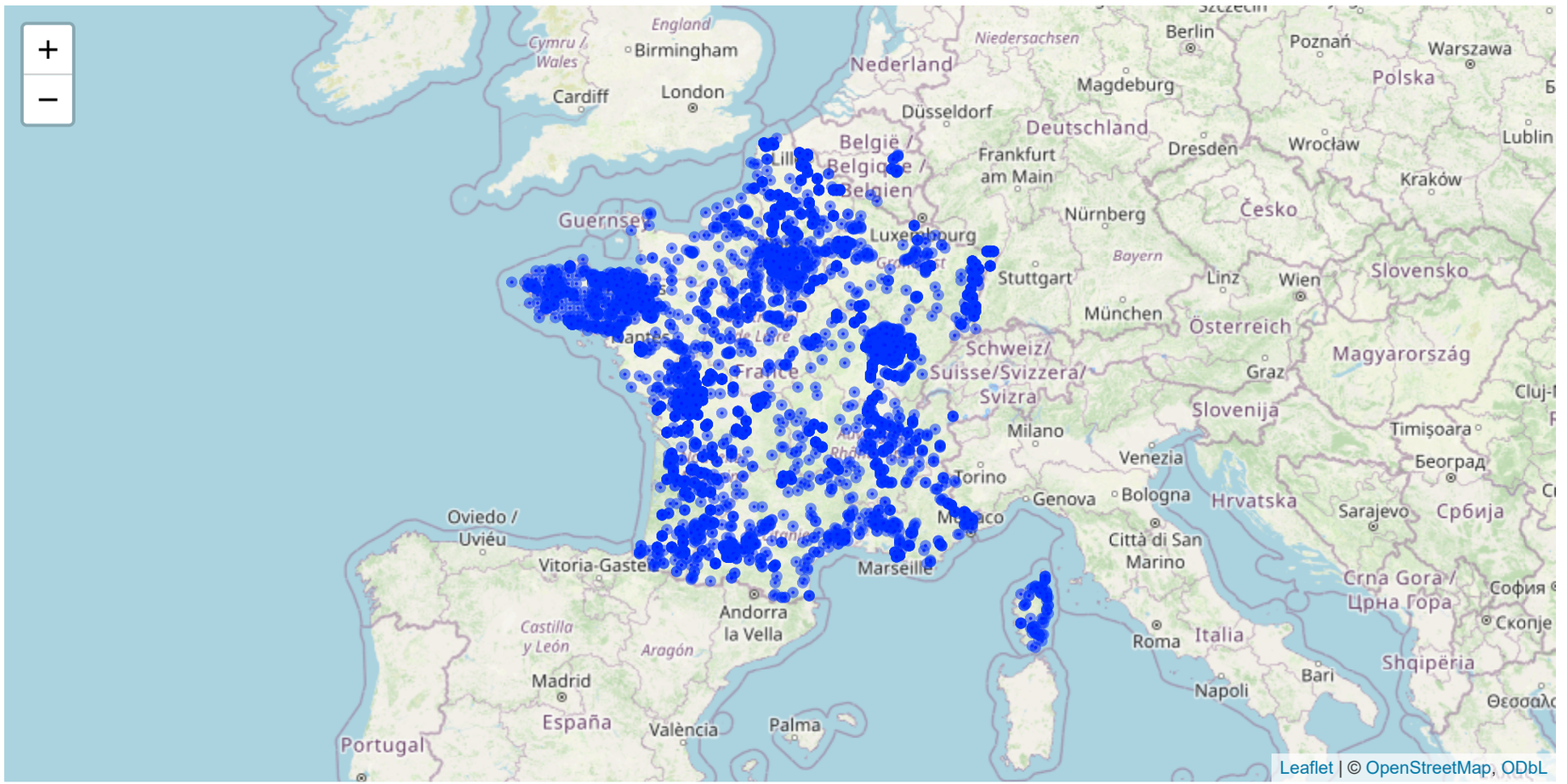
- There is **537** NA (CR = 93.2%) in **GPS_X**
- There is **540** NA (CR = 93.2%) in **GPS_Y**

▶ Code

- We delete the *NA* lines in the GPS coordinates
- The database therefore changes from **7919** to **7378** observations.

4.4 Cartography

▶ Code



- We delete points outside France (**22**)
- The database therefore changes from **7378** to **7356** observations.

4.5 Focus on years

- Cleaning the Annee column

▶ Code

- CR of Annee = **96.7%** (43 levels)

▶ Code

	Numbers
1821	1
1960	1

	Numbers
1978	1
1982	1
1983	1
1984	3
1986	1
1988	1
1989	2
1990	19
1991	23
1992	23

	Numbers
1993	15
1994	29
1995	7
1996	19
1997	18
1998	20
1999	37
2000	25
2001	11
2002	24

	Numbers
2003	18
2004	10
2005	59
2006	73
2007	102
2008	47
2009	100
2010	410
2011	105
2012	253

	Numbers
2013	307
2014	735
2015	273
2016	852
2017	360
2018	493
2019	593
2020	407
2021	1215
2022	515

	Numbers
2023	145
NA's	2

- We remove all the years before **1990** and the NA

▶ Code

	Numbers
1990	19
1991	23
1992	23
1993	15

	Numbers
1994	29
1995	7
1996	19
1997	18
1998	20
1999	37
2000	25
2001	11
2002	24
2003	18

	Numbers
2004	10
2005	59
2006	73
2007	102
2008	47
2009	100
2010	410
2011	105
2012	253
2013	307

	Numbers
2014	735
2015	273
2016	852
2017	360
2018	493
2019	593
2020	407
2021	1215
2022	515
2023	145

4.6 Table land use & protocol (start)

- clcm_lvl1 & protocol

▶ Code

	AITC_HS	AITCTM	F	F_HS	FHS	hand sorting	HS
1_Naturel	0	74	0	40	0	0	164
2_Agricole	193	153	0	825	0	0	191
3_Artificialisé	18	0	51	5	0	0	861
Agricole	0	0	0	95	0	0	0

	AITC_HS	AITCTM	F	F_HS	FHS	hand sorting	HS
NA	0	0	0	2	46	346	14

- clcm_lvl2 & protocol

▶ Code

	AITC_HS	AITCTM	F	F_HS	FHS
11_Naturel_fermé	0	8	0	33	0
12	0	0	0	0	0
12_Naturel_ouvert	0	66	0	7	0

	AITC_HS	AITCTM	F	F_HS	FHS
13_Zone_humide	0	0	0	0	0
14	0	0	0	0	0
21	0	0	0	0	0
21_Agricole_ouvert	136	137	0	766	0
22	0	0	0	0	0
22_Agricole_boisé	57	16	0	59	0
23	0	0	0	0	0
23_Bord_de_champs	0	0	0	0	0
31	0	0	0	0	0

	AITC_HS	AITCTM	F	F_HS	FHS
31_Espace_vert_boisé	4	0	0	0	0
32	0	0	0	0	0
32_Espace_vert_ouvert	2	0	51	0	0
33	0	0	0	0	0
33_Espace_vert_ornemental	0	0	0	0	0
34_Espace_cultivé_urbain	0	0	0	0	0
35_Bordure_de_voirie	10	0	0	0	0
37_Zone_de_chantier	0	0	0	0	0
38_Friche_industrielle	2	0	0	5	0

	AITC_HS	AITCTM	F	F_HS	FHS
39_Zone_humide_artificielle	0	0	0	0	0
41	0	0	0	0	0
Agroforestry	0	0	0	4	0
Arableland	0	0	0	44	0
Forest	0	0	0	2	0
Grassland	0	0	0	45	0
NA	0	0	0	2	46

- clcm_lvl3 & protocol

▶ Code

	AITC_HS	AITCTM
111_Forêt_de_feuillus	0	2
112_Forêt_de_conifères	0	0
113_Forêt_mixte	0	0
114_Bois	0	0
114c_Bois_ou_forêt_contaminé	0	0
115_Autre	0	6
120_Végétation_clairsemée	0	66
121_Prairie_naturelle_&_Paturage	0	0
122_Lande_mésophile_et_broussailles	0	0

	AITC_HS	AITCTM
124_Plage,_dune_et_sable	0	0
130_Prairie_humide	0	0
131_Mégaphorbiaie	0	0
132_Lande_hygrophile	0	0
210_Prairie_agricole_permanente	11	0
210_Prairie_agricole_permanente?	0	0
211_Prairie_agricole_temporaire	49	0
212_Jachère	0	0
213_Bande_enherbée	12	46

	AITC_HS	AITCTM
214_Culture_annuelle	60	81
214c_Culture_annuelle	0	0
215_Interculture	0	0
216_Maraîchage_et_légume_plein_champ	0	0
216l_Légume_plein_champ	0	0
216m_Maraichage	0	0
217_Serre,_tunnel_fixe	0	0
218_Vignes_et_autres_Cultures_pérennes	0	1
219_Autre	4	9

	AITC_HS	AITCTM
220ir_Agroforesterie	0	0
220r_Agroforesterie	0	0
221_Arboriculture	0	0
221_Verger	0	0
223_Autre	57	0
230_Talus_et_bordure_enherbés	0	0
231_Prairies	0	0
232_Haie	0	0
232_Haies	0	16

	AITC_HS	AITCTM
310_Forêt_urbaine	4	0
311_Bois_urbain	0	0
312_Verger	0	0
313_Autre	0	0
320_Prairie_urbaine	0	0
321_Pelouse	0	0
321_Pelouse_urbaine	2	0
322_Terrain_récréatif	0	0
323_Autre	0	0

	AITC_HS	AITCTM
324_Prairie_aéroport	0	0
330_Massif_ornemental	0	0
340_Potager	0	0
340p_Potager-pelouse	0	0
340p_Potager-Pelouse	0	0
343_Autre	0	0
350_Bande_enherbée_contrainte	10	0
351_Massif_ornemental_contraint	0	0
352_Alignement_d'arbres_en_fosse_continue	0	0

4.7 Focus on protocols

- List of protocols available on the database (18 levels)

▶ Code

	Numbers
AITC_HS	211
AITCTM	227
F	51
F_HS	967
FHS	46

	Numbers
hand sorting	346
HS	2949
HS_16	214
HS_4	396
HS_F_16	24
HS_M_16	179
HSAITC_16	1
HSAITC_4	123
HSAITC_6.25	56
HSAITC_7.95775385	10

	Numbers
M	1166
M_HS	69
Unknown	307

- Selection of protocols: **F_HS, FHS, hand sorting, HS**

▶ Code

	Numbers
F_HS	967
FHS	46
hand sorting	346

	Numbers
HS	2949

- The database therefore changes from **7342** to **4308** observations.
- Merging levels :
  - F_HS F_HS FHS
  - HS HS hand sorting
    - Code

	Numbers
F_HS	1013
HS	3295

4.8 Focus on clcm_lvl1

- CR of clcm_lvl1 = **76.8%** (4 levels)

▶ Code

	Numbers
1_Naturel	204
2_Agricole	2735
3_Artificialisé	866
Agricole	95
NA's	408

- Merging levels :
  - 2_Agricole $=$ 2_Agricole $+$ Agricole

▶ Code

	Numbers
1_Naturel	204
2_Agricole	2830
3_Artificialisé	866
NA's	408

- For the moment, we will keep the NA of **clcm_lvl1**

4.9 Focus on clcm_lvl2

- CR of clcm_lvl2 = **81.1%** (27 levels)
- To be continued (with BDD.V2)

▶ Code

	Numbers
11_Naturel_fermé	130
12	1
12_Naturel_ouvert	54
13_Zone_humide	19

	Numbers
14	1
21	2
21_Agricole_ouvert	2625
22	7
22_Agricole_boisé	101
23	131
23_Bord_de_champs	9
31	101
31_Espace_vert_boisé	65
32	69

	Numbers
32_Espace_vert_ouvert	363
33	6
33_Espace_vert_ornemental	22
34_Espace_cultivé_urbain	257
35_Bordure_de_voirie	142
37_Zone_de_chantier	4
38_Friche_industrielle	10
39_Zone_humide_artificielle	3
41	27
Agroforestry	4

	Numbers
Arableland	44
Forest	2
Grassland	45
NA's	64

4.10 Focus on clcm_lvl3

- CR of clcm_lvl3 = **73.2%** (60 levels)
- To be continued (with BDD.V2)

▶ Code

	Numbers
111_Forêt_de_feuillus	25
112_Forêt_de_conifères	4
113_Forêt_mixte	88
114_Bois	6

	Numbers
114c_Bois_ou_forêt_contaminé	6
115_Autre	1
121_Prairie_naturelle_&_Paturage	46
122_Lande_mésophile_et_broussailles	6
124_Plage,_dune_et_sable	1
130_Prairie_humide	19
132_Lande_hygrophile	1
210_Prairie_agricole_permanente	172
210_Prairie_agricole_permanente?	1
211_Prairie_agricole_temporaire	114

	Numbers
212_Jachère	3
213_Bande_enherbée	1
214_Culture_annuelle	1286
214c_Culture_annuelle	2
215_Interculture	99
216_Maraîchage_et_légume_plein_champ	10
216l_Légume_plein_champ	78
216m_Maraichage	21
217_Serre,_tunnel_fixe	4
218_Vignes_et_autres_Cultures_pérennes	741

	Numbers
219_Autre	5
220ir_Agroforesterie	46
220r_Agroforesterie	43
221_Arboriculture	11
221_Verger	1
230_Talus_et_bordure_enherbés	7
231_Prairies	17
232_Haie	1
232_Haies	1
310_Forêt_urbaine	15

	Numbers
311_Bois_urbain	39
312_Verger	6
313_Autre	7
320_Prairie_urbaine	69
321_Pelouse	29
321_Pelouse_urbaine	209
322_Terrain_récréatif	7
323_Autre	1
324_Prairie_aéroport	44
330_Massif_ornemental	22

	Numbers
340_Potager	253
340p_Potager-pelouse	1
340p_Potager-Pelouse	1
343_Autre	4
350_Bande_enherbée_contrainte	44
351_Massif_ornemental_contraint	10
352_Alignement_d'arbres_en_fosse_continue	38
353_Arbre_en_fosse_unitaire	13
354_Rond-point	19
355_Noue_enherbée	7

4.11 Land use selection

- **Forêt** 111_Forêt_de_feuillus + 112_Forêt_de_conifères + 113_Forêt_mixte
- **Prairie agricole p.** 210_Prairie_agricole_permanente = Prairie_agricole
- **Prairie agricole t.** 211_Prairie_agricole_temporaire
- **Culture** 214_Culture_annuelle
- **Vignoble** 218_Vignes_et_autres_Cultures_pérennes
- **Pelouse urbaine** 321_Pelouse_urbaine
- **Potager** 340_Potager

(To be continued with BDD.V2)

- Code

4.12 Land use & protocol overview

	F_HS	HS
111_Forêt_de_feuillus	9	16
112_Forêt_de_conifères	3	1
113_Forêt_mixte	11	77
114_Bois	4	2
114c_Bois_ou_forêt_contaminé	6	0
115_Autre	0	1
121_Prairie_naturelle_&_Paturage	3	43
122_Lande_mésophile_et_broussailles	3	3

	F_HS	HS
124_Plage,_dune_et_sable	1	0
130_Prairie_humide	0	19
132_Lande_hygrophile	0	1
210_Prairie_agricole_permanente	53	119
210_Prairie_agricole_permanente?	1	0
211_Prairie_agricole_temporaire	56	58
212_Jachère	1	2
213_Bande_enherbée	0	1
214_Culture_annuelle	251	1035
214c_Culture_annuelle	2	0

	F_HS	HS
215_Interculture	21	78
216_Maraîchage_et_légume_plein_champ	2	8
216l_Légume_plein_champ	0	78
216m_Maraichage	0	21
217_Serre,_tunnel_fixe	0	4
218_Vignes_et_autres_Cultures_pérennes	373	368
219_Autre	0	5
220ir_Agroforesterie	26	20
220r_Agroforesterie	22	21
221_Arboriculture	11	0

	F_HS	HS
221_Verger	0	1
230_Talus_et_bordure_enherbés	0	7
231_Prairies	0	17
232_Haie	0	1
232_Haies	0	1
310_Forêt_urbaine	0	15
311_Bois_urbain	0	39
312_Verger	0	6
313_Autre	0	7
320_Prairie_urbaine	0	69

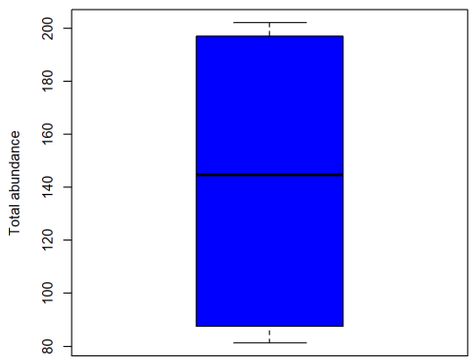
	F_HS	HS
321_Pelouse	0	29
321_Pelouse_urbaine	0	209
322_Terrain_récréatif	0	7
323_Autre	0	1
324_Prairie_aéroport	0	44
330_Massif_ornemental	0	22
340_Potager	0	253
340p_Potager-pelouse	0	1
340p_Potager-Pelouse	0	1
343_Autre	0	4

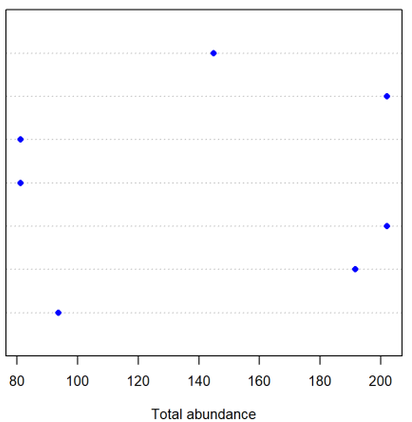
	F_HS	HS
350_Bande_enherbée_contrainte	0	44
351_Massif_ornemental_contraint	0	10
352_Alignement_d’arbres_en_fosse_continue	0	38
353_Arbre_en_fosse_unitaire	0	13
354_Rond-point	0	19
355_Noue_enherbée	0	7
356_Noue_arborée	0	10
357_Autre	0	1
370_Stock_de_terre	0	4
380_Friche_industrielle	5	5

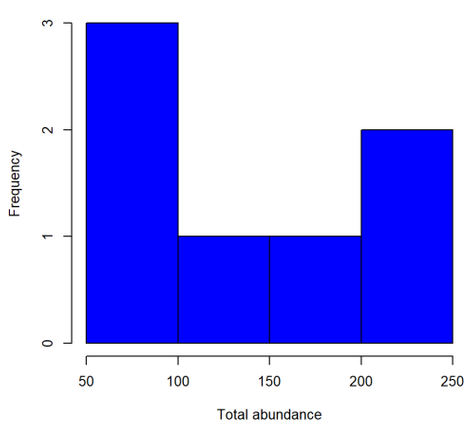
5 Earthworms data

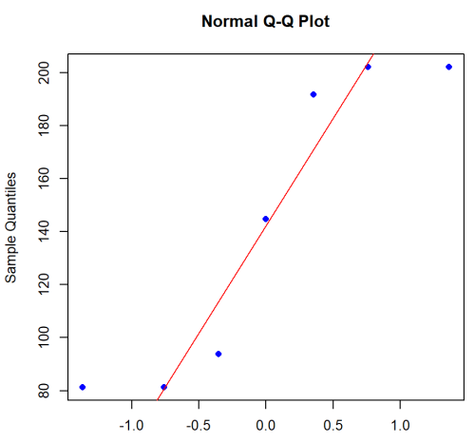
5.1 Total abundance (CR = 3%)

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  81.25   87.50  144.79  142.41  196.88  202.08    4301
```



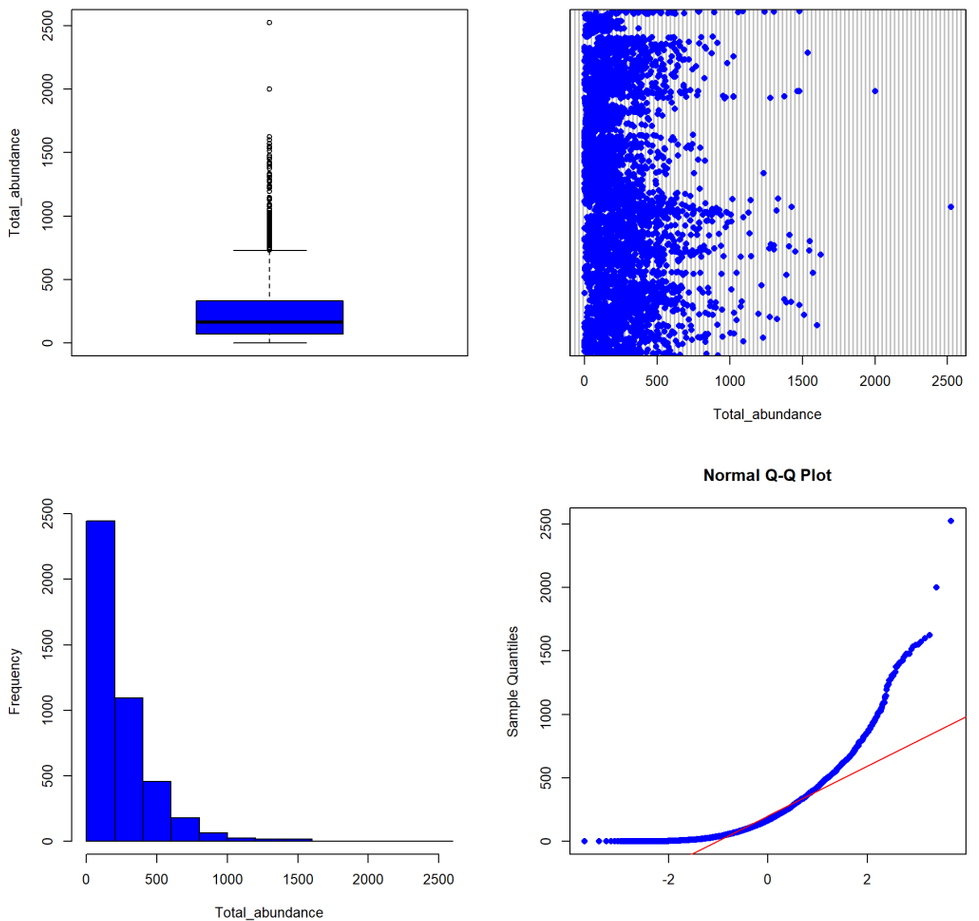






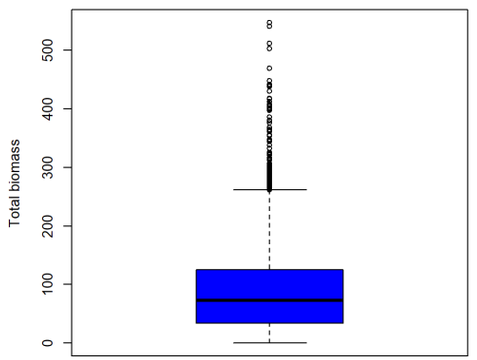
5.2 Total_AB ? (CR = 92.9%)

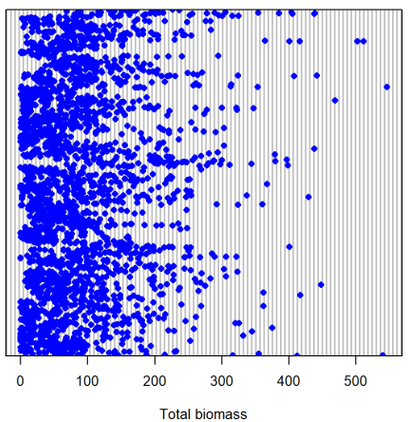
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   0.00   68.75  166.67  234.64  333.33 2525.00       7
```

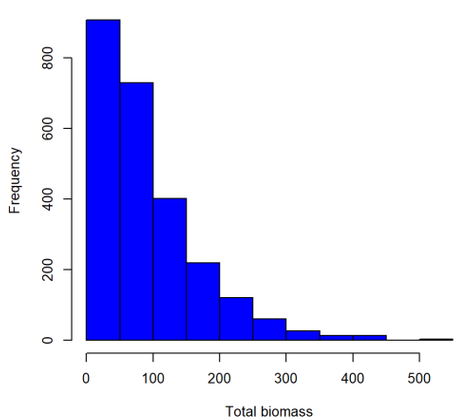


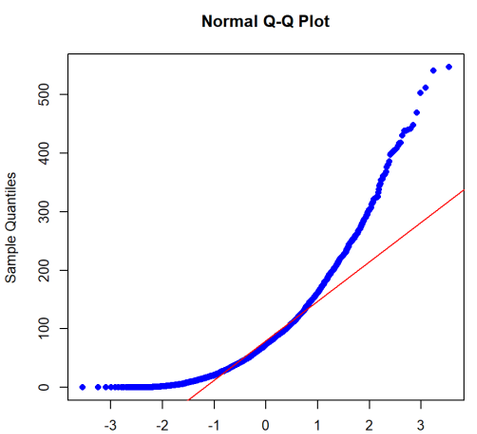
5.3 Total biomass (CR = 48.5%)

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   0.00   34.31   73.00   91.74  125.27  547.01    1806
```









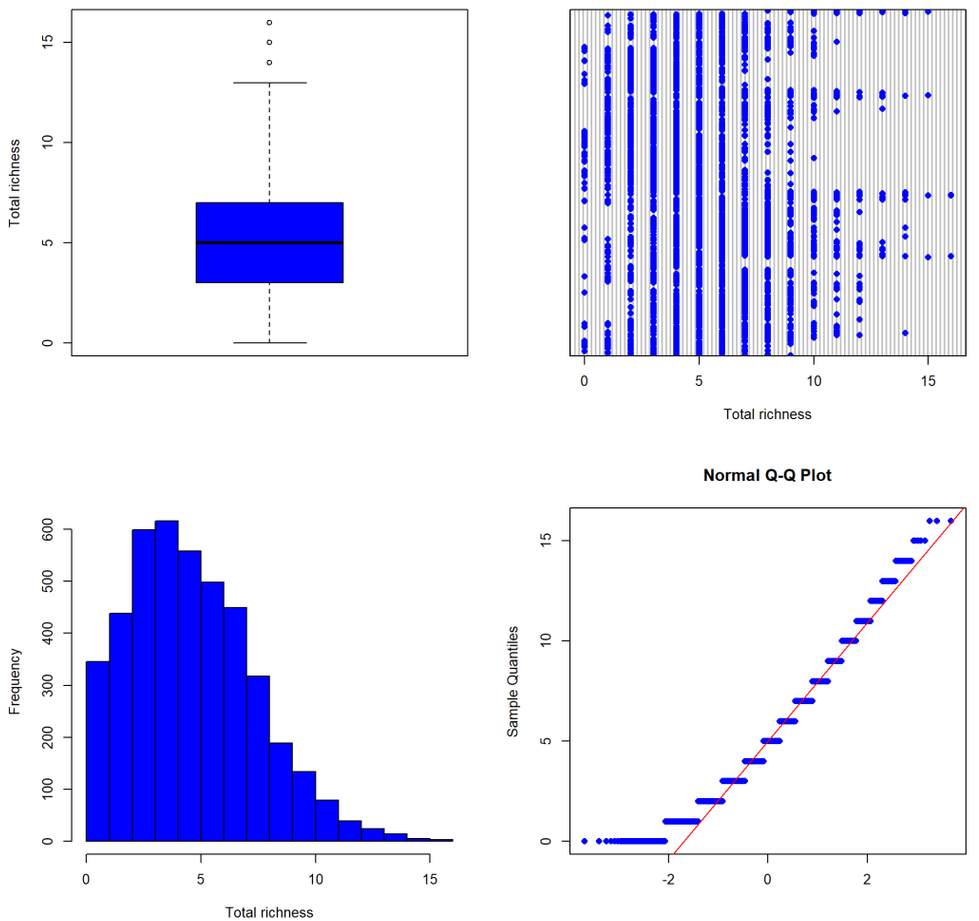
5.4 Total richness calculation method

- Removal of columns with only NA (**115**) and/or only 0
- Identify columns beginning with **AB_**
- Deletion of **AB_** columns that are not species
- Calculate richness by assigning **1** to each column if the value is different from 0 and NA
- Total richness = **1** if the plot has a value in AB and/or BM

▶ Code

5.5 Total richness

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   3.000   5.000   5.077   7.000  16.000
```



6 Soil data extraction

6.1 The source database (openlandmap)

▶ Code

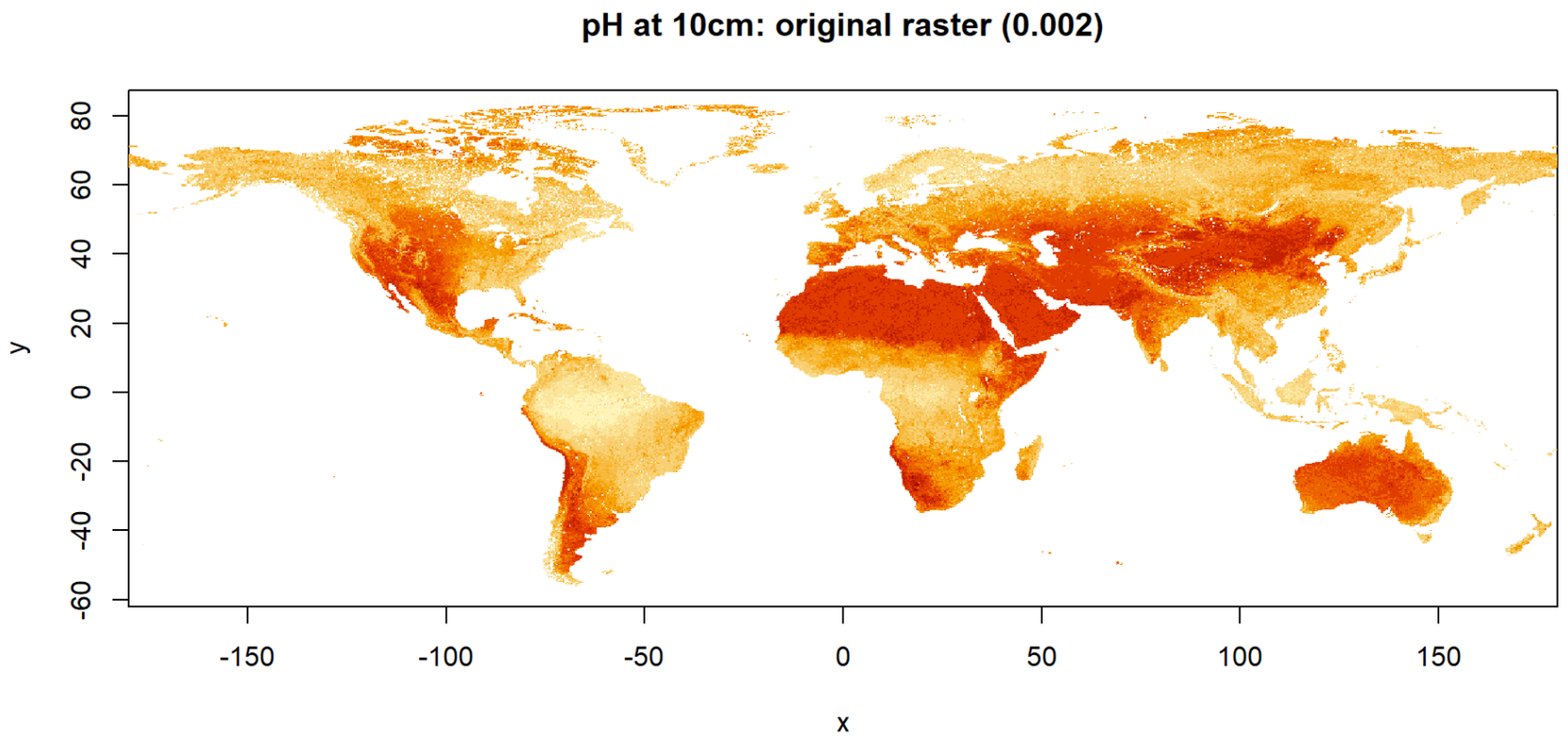
Source	DB.name	Categories	Variables	Units	Forma
Fourcade et al., 2022	OpenLandMap	soil data	pH		.tif
			Bulk density	kg / m-cube	
			Sand content	% (kg/kg)	

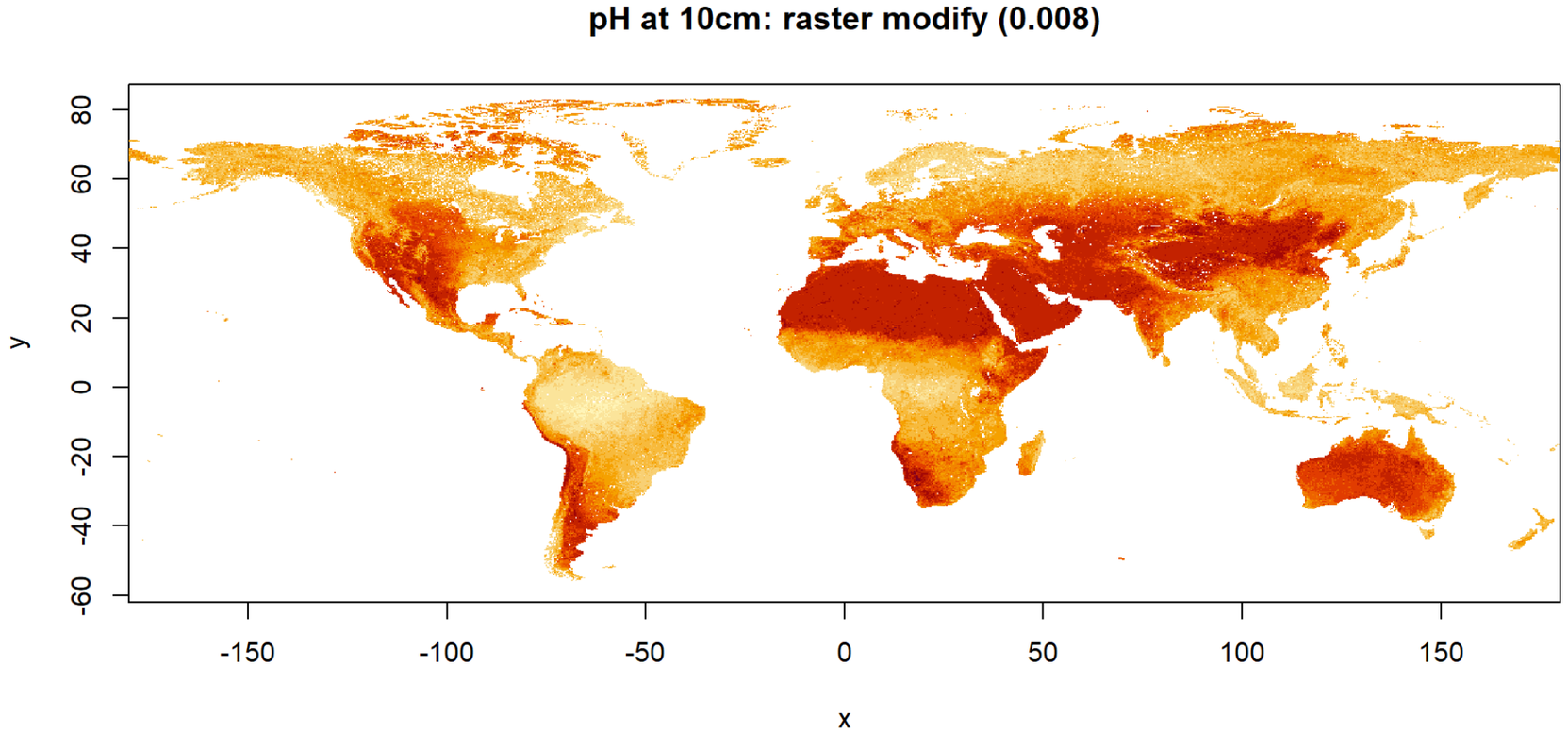
Source	DB.name	Categories	Variables	Units	Forma
			carbone content	g/kg	

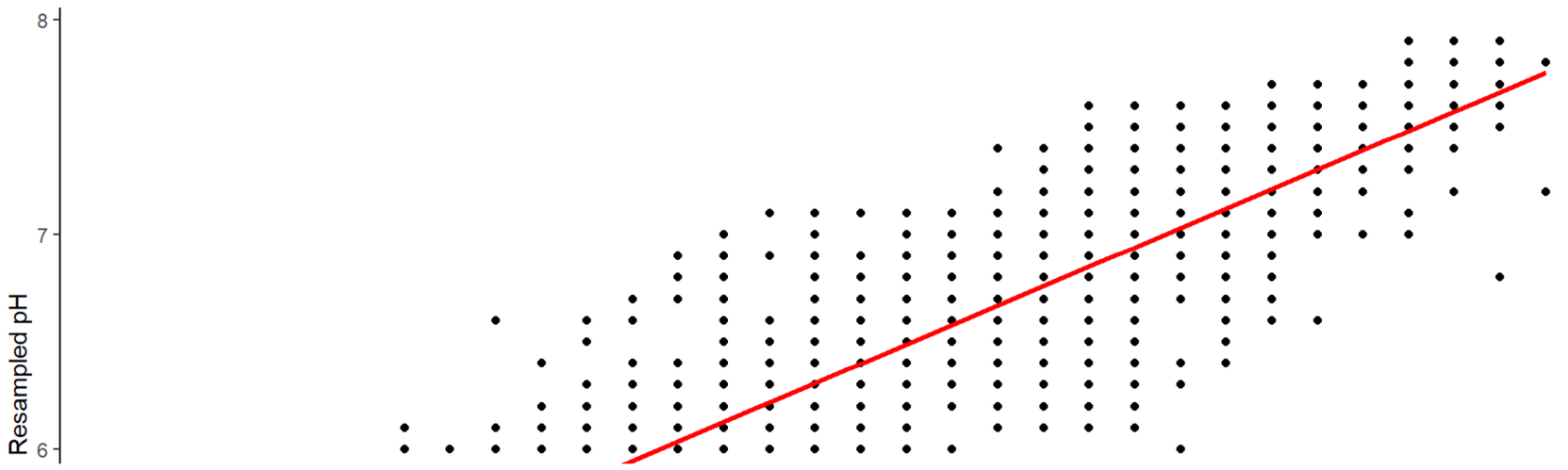
- Average values between surface (0 cm) and 30 cm depth

6.2 Changing the resolution

- Long compilation time in R
- GDAL module with the resampleAlg = bilinear method
- Resolution = 0.0083 = 30 arc-second ~ 1km

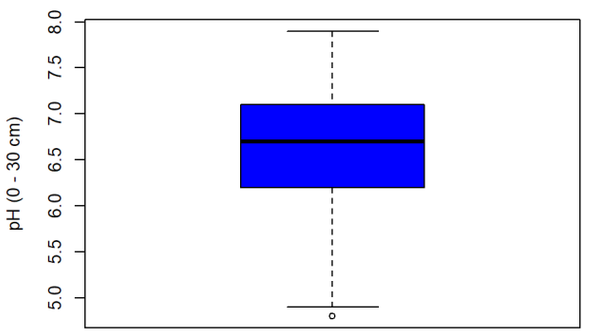


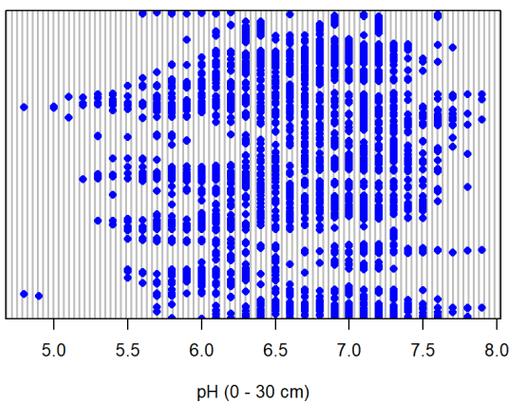

pH at 10cm: raster modify (0.008)
y
80
60
40
20
0
-20
-40
-60
-150
-100
-50
0
50
100
150
x

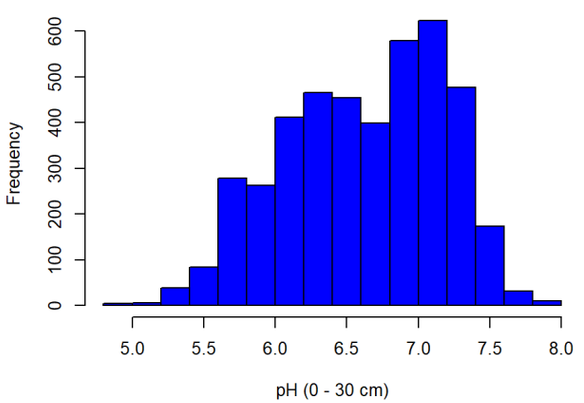

Resampled pH
8
7
6

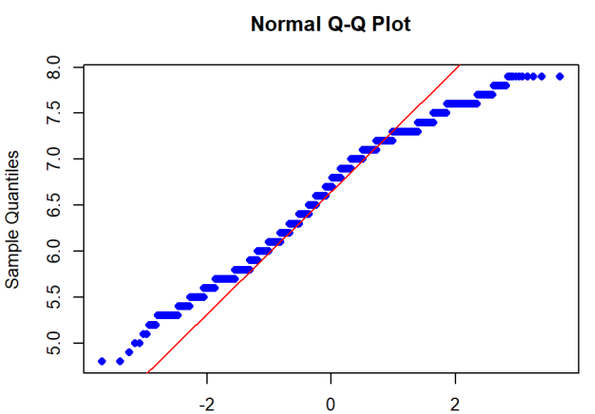
6.3 pH

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  4.800   6.200   6.700   6.669   7.100   7.900      10
```



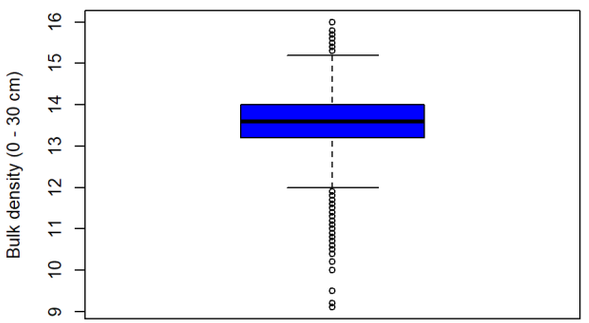


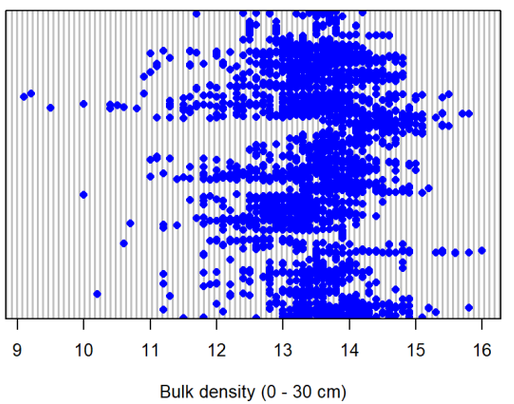


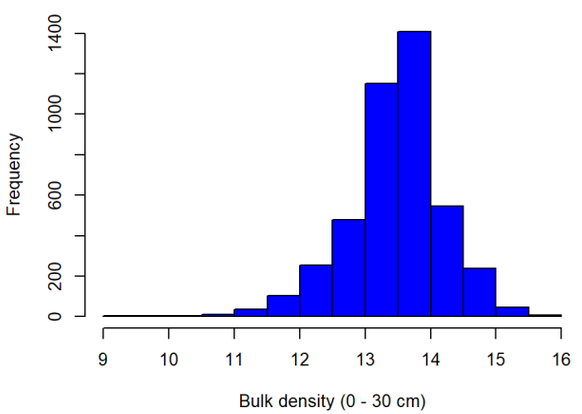


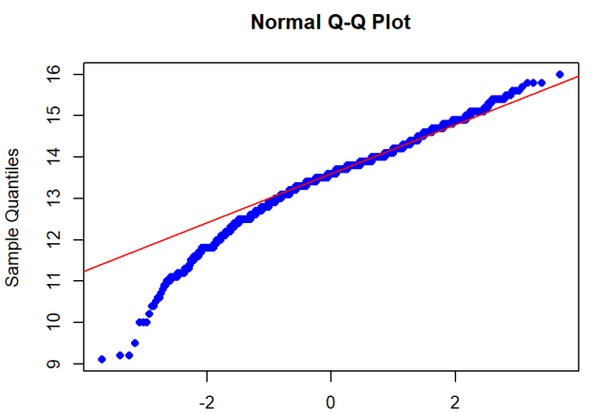
6.4 Bulk density (kg / m-cube)

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   9.10   13.20   13.60   13.52   14.00   16.00      10
```



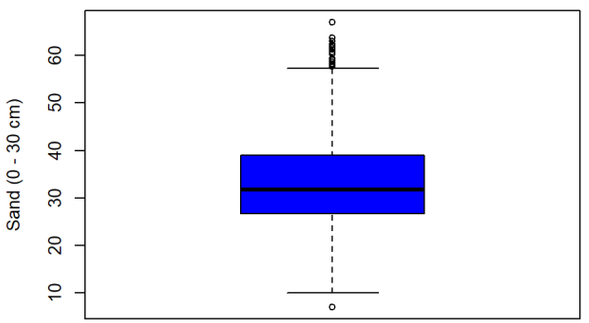


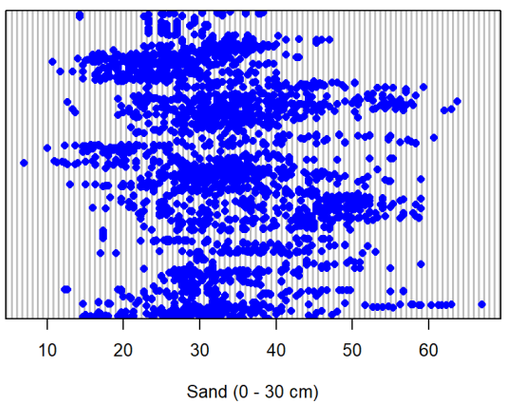


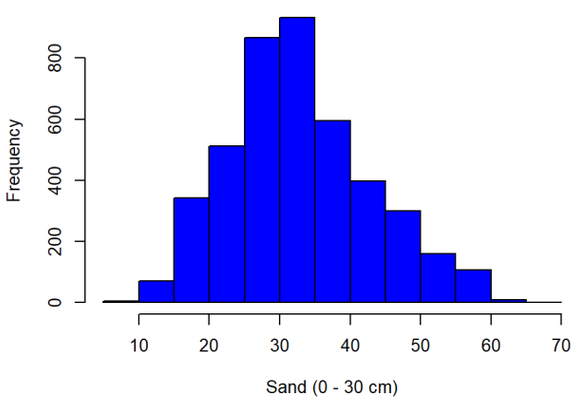


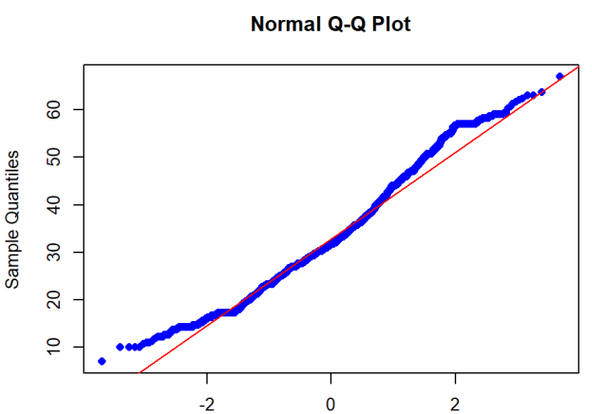
6.5 Sand content (% kg/kg)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
7.00	26.70	31.70	32.96	39.00	67.00	10



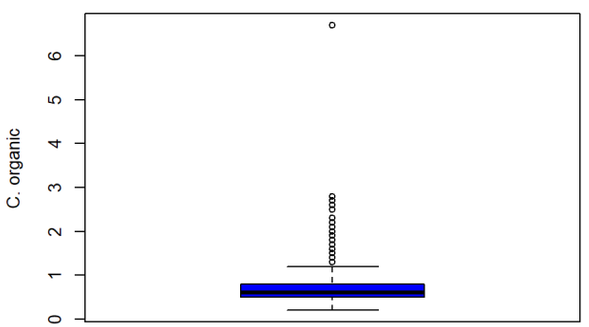


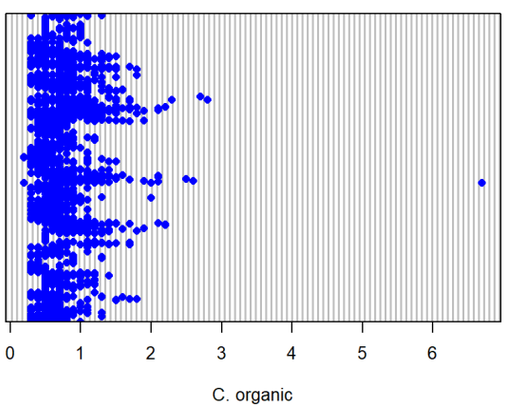


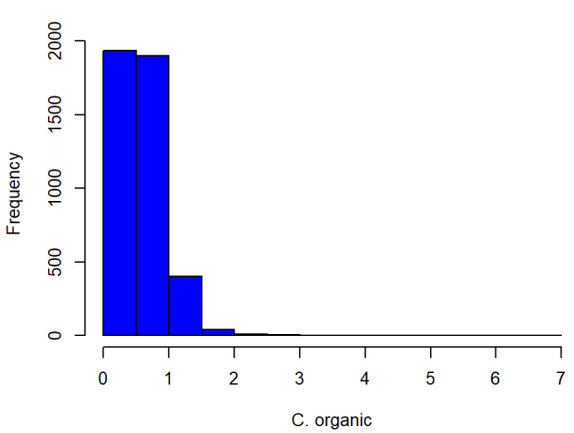


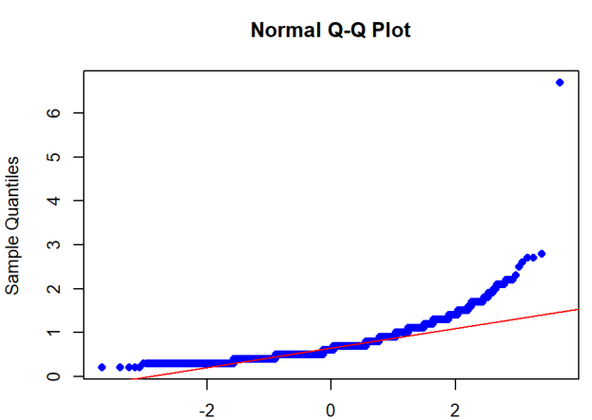
6.6 Soil organic carbone (g/kg)

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 0.2000  0.5000  0.6000  0.6792  0.8000  6.7000      10
```









7 Climate data extraction

7.1 The source database (CHELSA V2)

▶ Code

Source	DB.name	Categories	Variables	Units	Formats
Fourcade et al., 2022	CHELSA V2.1	climate	bio1	°C	.tif
			…	…	
			bio19	kg/m^2	

7.2 Extraction method

- Link recovery (see file link .tif)
- Extracting variable names
- Uses of the **extraction()** function
- Convert columns to correct format and unit
- Adding variables to the LANDWORM database

▶ Code

7.3 List of variables

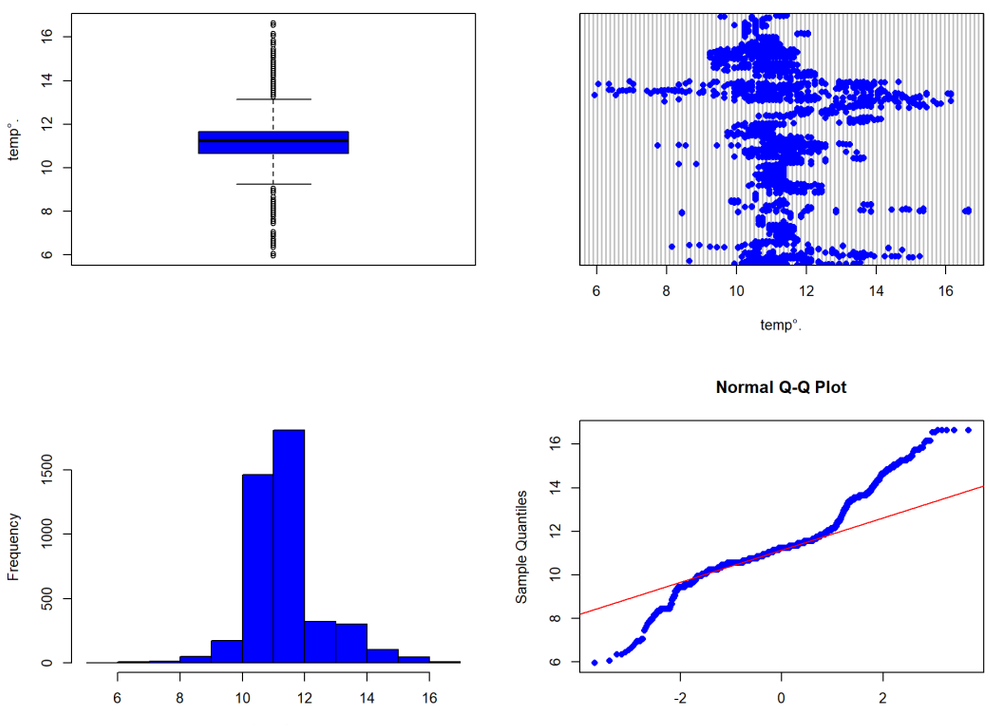
Variable description

```
      ID                 gps_x             gps_y             ai        
 Length:7378        Min.   :-5.0496   Min.   :41.44   Min.   :0.5901  
 Class :character   1st Qu.:-0.3535   1st Qu.:46.17   1st Qu.:0.9428  
 Mode  :character   Median : 2.0985   Median :47.92   Median :1.0278  
                    Mean   : 1.9577   Mean   :47.26   Mean   :1.0878  
                    3rd Qu.: 4.1679   3rd Qu.:48.76   3rd Qu.:1.1373  
                    Max.   : 9.5213   Max.   :50.98   Max.   :3.4735  
                                                      NA's   :7       
     bio10            bio11             bio12            bio13      
 Min.   : 8.35   Min.   :-10.150   Min.   : 563.7   Min.   : 61.80  
 1st Qu.:17.55   1st Qu.:  3.550   1st Qu.: 738.5   1st Qu.: 78.20  
 Median :18.45   Median :  4.250   Median : 816.2   Median : 88.30  
 Mean   :18.61   Mean   :  4.613   Mean   : 847.3   Mean   : 93.17  
 3rd Qu.:19.55   3rd Qu.:  5.850   3rd Qu.: 898.8   3rd Qu.:102.70  
 Max.   :24.75   Max.   : 10.950   Max.   :2158.3   Max.   :239.90  
```

7.4 Temperature

- Average annual air temperature (°C) = bio1

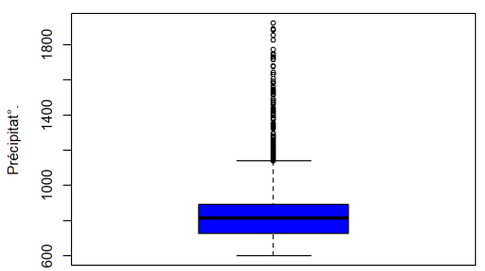
```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
5.95   10.65   11.25   11.36   11.65   16.65
```

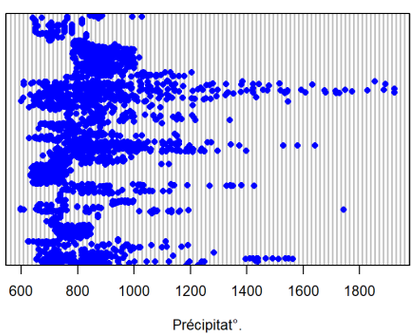


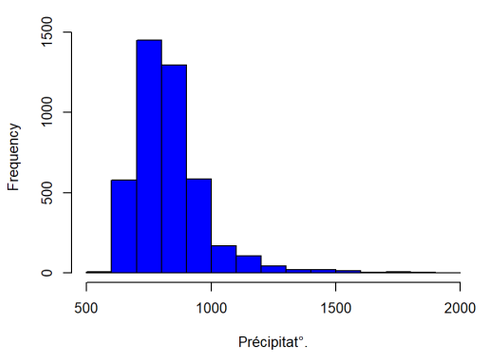
7.5 Precipitation

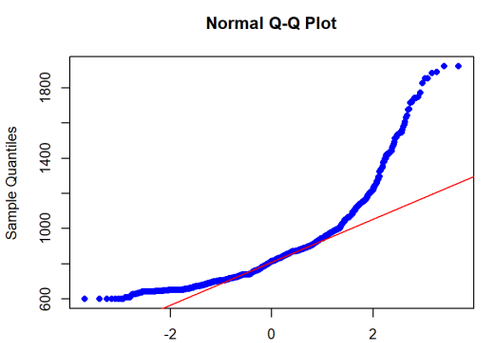
- Annual precipitation (kg/m²) = bio12

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  598.8   726.1   814.2   833.7   891.3  1924.0
```









8 Questions

- Comment gérer la répétition temporelle des parcelles ?
  - Avec répétition : **4308** observations
  - Sans répétition : **3269** observations
- Liens des données du sol (sable, argile et limon) de data.gouv.fr ?

9 To do next

- Extraction des prédicteurs : CEC, Limon, argile, évapotranspiration, paysage,…
- Analyse exploratoire : test de corrélation, VIF, ACP
- Sélection des variables
- Modélisation (GLM, GAM, RF, ANN)
- Validation croisée
- Rédaction, protocol ODMAP