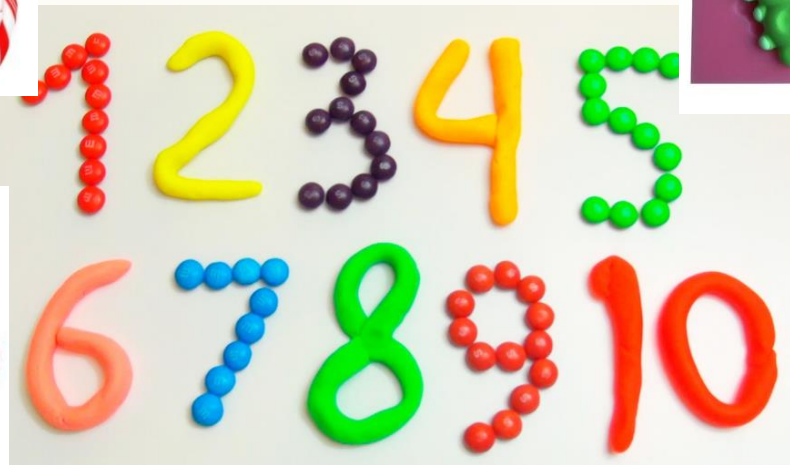


# Aula 21 Goodies\*



\* Goodies related to animals, plants and numbers...

# World Scientists' Warning of a Climate Emergency

WILLIAM J. RIPPLE, CHRISTOPHER WOLF, THOMAS M. NEWSOME, PHOEBE BARNARD, WILLIAM R. MOOMAW, AND 11,258 SCIENTIST SIGNATORIES FROM 153 COUNTRIES (LIST IN SUPPLEMENTAL FILE S1)

**S**cientists have a moral obligation to clearly warn humanity of any catastrophic threat and to “tell it like it is.” On the basis of this obligation and the graphical indicators presented below, we declare, with more than 11,000 scientist signatories from around the world, clearly and unequivocally that planet Earth is facing a climate emergency.

as actual climatic impacts (figure 2). We use only relevant data sets that are clear, understandable, systematically collected for at least the last 5 years, and updated at least annually.

The climate crisis is closely linked to excessive consumption of the wealthy lifestyle. The most affluent countries are mainly responsible for the historical GHG emissions and generally

forest loss in Brazil's Amazon has now started to increase again (figure 1g). Consumption of solar and wind energy has increased 373% per decade, but in 2018, it was still 28 times smaller than fossil fuel consumption (combined gas, coal, oil; figure 1h). As of 2018, approximately 14.0% of global GHG emissions were covered by carbon pricing (figure 1m), but

Scientists have a moral obligation to clearly warn humanity of any catastrophic threat and to “tell it like it is.” On the basis of this obligation and the graphical indicators presented below, we declare, with more than 11,000 scientist signatories from around the world, clearly and unequivocally that **planet Earth is facing a climate emergency.**

<https://doi.org/10.1093/biosci/biz088>

# Gestão de Páginas

- Ecologia Numérica
  - Ecologia Numérica(Tecnologias de Infor
  - Teóricas
- Práticas
  - PDFs
- Outros Recursos

+ Criar

## PDFs

Página **Ficheiros 6** Permissões Link

Adicionar Ficheiro

#	Nome
1	Numerical Ecology with R <i>Borcardetal2001EcologyUseR.pdf</i>
2	The R Book.pdf
3	Publication bias: What are the challenges and can they be overcome? <i>jpn-37-149.pdf</i>
4	The mismatch between current statistical practice and doctoral training in ecology <i>Touchon&amp;McCoy_2016_MismatchStatsPractice&amp;TeachingEcology.pdf</i>
5	Quasi-poisson vs. negative binomial regression: how should we model overdispersed count data? <i>Hoef&amp;Boveng2007.pdf</i>

6	World Scientists' Warning of a Climate Emergency <i>biz088.pdf</i>
---	---

# Ecology for the Masses

Making ecological science accessible

[HOME](#) [ABOUT ECOMASS](#) [THE SCICOMM HUB](#) [THE PODCAST](#) [ARCHIVE](#) [CONTACT](#)

## THE MODERN BIOLOGIST'S CHALLENGE: DATA MANAGEMENT

Posted on [November 25, 2019](#) by [Stefan Vriend](#) | [Leave a comment](#)



### FOLLOW BLOG

Click below to follow. You'll receive a notification when there's new content, and updates every now and again.

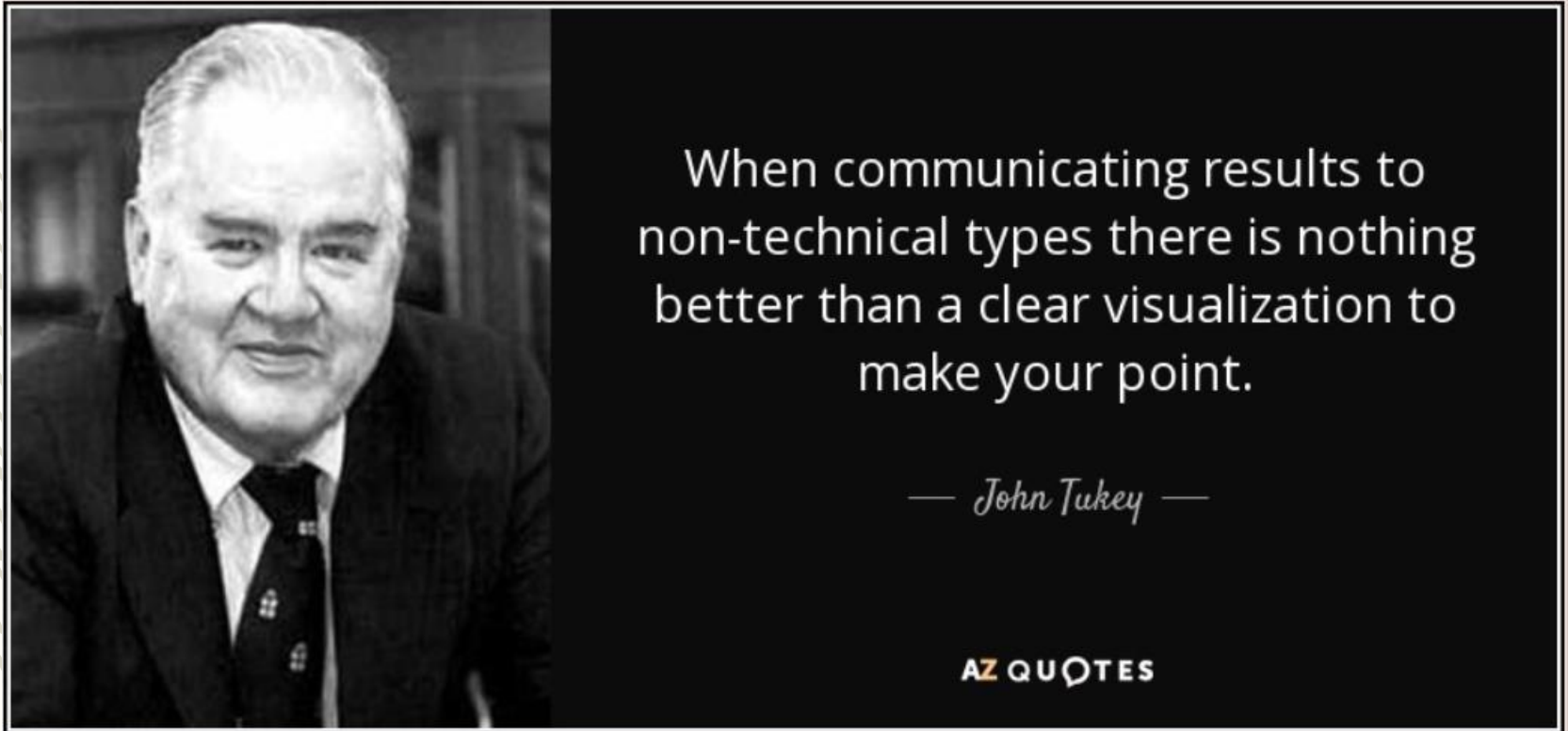
Enter your email address

Follow

*"...Modern biologists often do most of their most integral work not deep in a forest, but sitting behind a laptop while fuelling their caffeine addiction..."*

<https://ecologyforthemasses.com/2019/11/25/the-modern-biologists-challenge-data-management/>





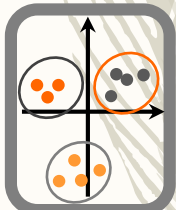
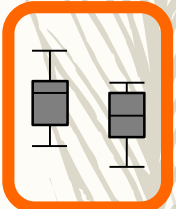
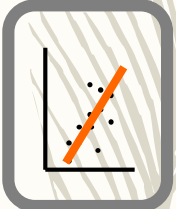
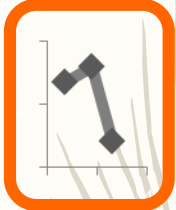
<https://www.azquotes.com/quote/593306>

# ecologia numérica

---

Introdução à análise multivariada

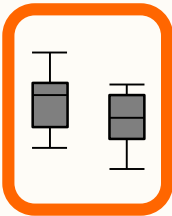
classificação hierárquica  
e não hierárquica



Ao contrário do que tínhamos até agora:

1. uma ou 2 amostras - testes de comparações
2. uma variável que queremos prever em função de outra(s) – modelos de regressão

Na análise multivariada temos muitas variáveis cujas relações de interdependência queremos explicar.



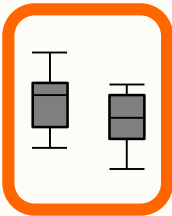
## introdução à análise multivariada

---

- Como analisar dados multivariados quando temos várias variáveis de interesse?

O exemplo mais comum em dados ecológicos serão valores de abundâncias de espécies em múltiplos locais, para os quais também poderemos ter, ou não, variáveis ambientais



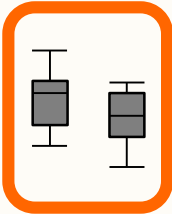


## Principais objectivos

Identificar padrões em conjuntos de dados complexos e de grande dimensionalidade;

Sumariar a informação;

**Estabelecer** e/ou **confirmar** hipóteses – regra geral as metodologias multivariadas são descritivas e exploratórias por natureza, pelo que a primeiro objectivo é mais simples de atingir



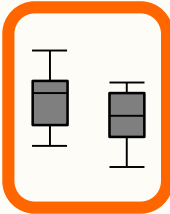
# introdução à análise multivariada

Exemplo de dados multivariados:

Densidades de espécies em várias estações de amostragem

59 estações x 58 spp.

estac	A ang	A imp	A lat	A min	A obs	A pre	A reg	A tho	B boc	B boo	B lut	C car	C con	C lab	C lyr
Inv 1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	5.965	0.000	0.000	0.000	0.000	0.000	0.000
Inv 2	0.000	0.000	0.000	0.000	0.000	0.000	2.540	0.000	3.778	0.000	0.000	0.000	0.000	0.000	0.000
Inv 3	0.000	0.000	0.000	0.000	0.000	0.000	0.261	0.000	0.585	0.000	0.000	0.000	0.000	0.000	0.000
Inv 4	0.000	0.000	0.000	0.000	0.000	0.000	0.330	0.000	0.521	0.000	0.000	0.000	0.000	0.000	0.000
Inv 5	0.000	0.000	0.000	0.000	0.000	0.000	0.121	0.000	0.000	0.000	0.000	0.023	0.000	0.000	0.000
Inv 6	0.000	0.000	0.000	0.000	0.000	0.000	0.149	0.000	0.000	0.000	0.000	0.000	0.091	0.000	0.000
Inv 7	0.000	0.000	0.000	0.000	0.000	0.000	0.583	0.000	0.000	0.000	0.000	0.000	0.000	0.052	0.000
Inv 8	0.000	0.000	0.000	0.000	0.000	0.000	0.066	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Inv 9	0.000	0.000	0.000	0.029	0.000	0.000	0.316	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Inv 10	0.000	0.000	0.000	0.000	0.000	0.000	0.089	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Inv 11	0.000	0.000	0.000	0.098	0.000	0.033	0.030	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000
Inv 12	0.000	0.000	0.000	0.000	0.000	0.187	0.623	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Inv 13	0.030	0.000	0.000	0.000	0.000	0.000	1.219	0.000	0.000	0.000	0.000	0.000	0.113	0.000	0.000
Inv 14	0.000	0.000	0.000	0.000	0.000	0.000	0.982	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Inv 15	0.000	0.000	0.000	0.064	0.000	0.000	0.185	0.000	0.000	0.000	0.000	0.000	0.032	0.000	0.000
Inv 16	0.030	0.000	0.000	0.021	0.000	0.000	0.611	0.000	0.000	0.000	0.000	0.000	0.064	0.000	0.000
Inv 17	0.063	0.000	0.000	0.000	0.000	0.000	0.308	0.000	0.000	0.000	0.000	0.000	0.028	0.000	0.000
Inv 18	0.000	0.000	0.000	0.000	0.000	0.000	0.060	0.000	0.000	0.000	0.000	0.000	0.036	0.000	0.000
Inv 19	0.000	0.000	0.000	0.000	0.000	0.000	0.485	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Inv 20	0.000	0.000	0.000	0.000	0.000	0.000	0.913	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Inv 21	0.000	0.000	0.000	0.267	0.000	0.000	0.024	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Inv 22	0.000	0.000	0.000	0.024	0.000	0.024	0.054	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Inv 23	0.000	0.000	0.000	0.269	0.000	0.000	0.077	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Inv 25	0.000	0.013	0.064	0.000	0.000	0.000	0.299	0.013	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Inv 26	0.000	0.000	0.036	0.000	0.000	0.000	1.401	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Inv 27	0.000	0.000	0.273	0.000	0.000	0.000	3.066	0.000	0.000	0.044	0.117	0.000	0.011	0.000	0.064
Inv 28	0.000	0.000	1.465	0.000	0.000	0.329	0.604	0.000	0.000	0.023	0.231	0.000	0.000	0.000	2.270
Inv 29	0.000	0.000	0.356	0.000	0.000	0.075	0.365	0.000	0.000	0.000	0.244	0.000	0.000	0.000	1.271
Inv 30	0.000	0.000	0.239	0.000	0.000	0.084	0.000	0.000	0.000	0.016	0.144	0.000	0.000	0.000	1.452
Ver 1	0.000	0.000	0.000	0.000	0.000	0.000	1.013	0.000	0.224	0.000	0.000	0.000	0.000	0.000	0.000
Ver 2	0.000	0.000	0.000	0.000	0.000	0.000	1.240	0.000	9.675	0.000	0.000	0.000	0.000	0.000	0.000
Ver 3	0.000	0.000	0.000	0.000	0.000	0.000	0.431	0.000	0.082	0.000	0.000	0.000	0.000	0.000	0.000
Ver 4	0.000	0.000	0.000	0.000	0.000	0.000	0.355	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ver 5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ver 6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.042	0.000	0.000
Ver 7	0.000	0.000	0.000	0.000	0.000	0.000	1.330	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ver 8	0.000	0.000	0.000	0.000	0.000	0.000	0.422	0.000	0.000	0.000	0.000	0.000	0.046	0.048	0.000
Ver 9	0.000	0.000	0.000	0.000	0.000	0.000	0.245	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ver 10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ver 11	0.000	0.000	0.000	0.000	0.000	0.000	0.534	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000
Ver 12	0.000	0.000	0.000	0.000	0.000	0.000	1.065	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ver 13	0.000	0.000	0.000	0.000	0.000	0.000	2.400	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ver 14	0.000	0.000	0.000	0.000	0.000	0.047	1.835	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ver 15	0.000	0.000	0.000	0.000	0.000	0.000	0.092	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ver 16	0.000	0.000	0.000	0.000	0.000	0.000	0.040	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ver 17	0.000	0.000	0.000	0.000	0.000	0.048	0.000	0.000	0.000	0.000	0.000	0.000	0.095	0.000	0.238
Ver 18	0.054	0.000	0.000	0.000	0.000	0.109	2.716	0.000	0.000	0.000	0.000	0.000	0.096	0.326	0.193
Ver 19	0.000	0.000	0.000	0.000	0.000	0.000	6.914	0.000	0.000	0.000	0.000	0.000	0.115	0.000	0.000
Ver 20	0.000	0.000	0.000	0.000	0.000	0.000	0.221	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ver 21	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ver 22	0.000	0.000	0.000	0.000	0.000	0.000	0.112	0.000	0.000	0.000	0.000	0.000	0.091	0.000	0.000
Ver 23	0.000	0.000	0.000	0.000	0.000	2.952	1.566	0.000	0.000	0.000	0.000	0.000	0.000	0.056	0.000
Ver 24	0.000	0.000	0.000	0.000	0.000	0.000	0.358	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.654
Ver 25	0.000	0.000	0.407	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ver 26	0.000	0.000	2.052	0.000	0.000	0.000	0.000	0.000	0.000	0.060	0.000	0.000	0.000	0.000	0.000
Ver 27	0.000	0.000	2.979	0.000	0.190	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ver 28	0.000	0.000	2.743	0.000	0.187	0.000	0.000	0.000	0.000	0.312	0.312	0.000	0.000	0.000	12.904



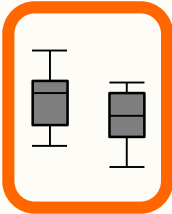
# introdução à análise multivariada



Dados das Var.  
ambientais

59 estações x  
12 var

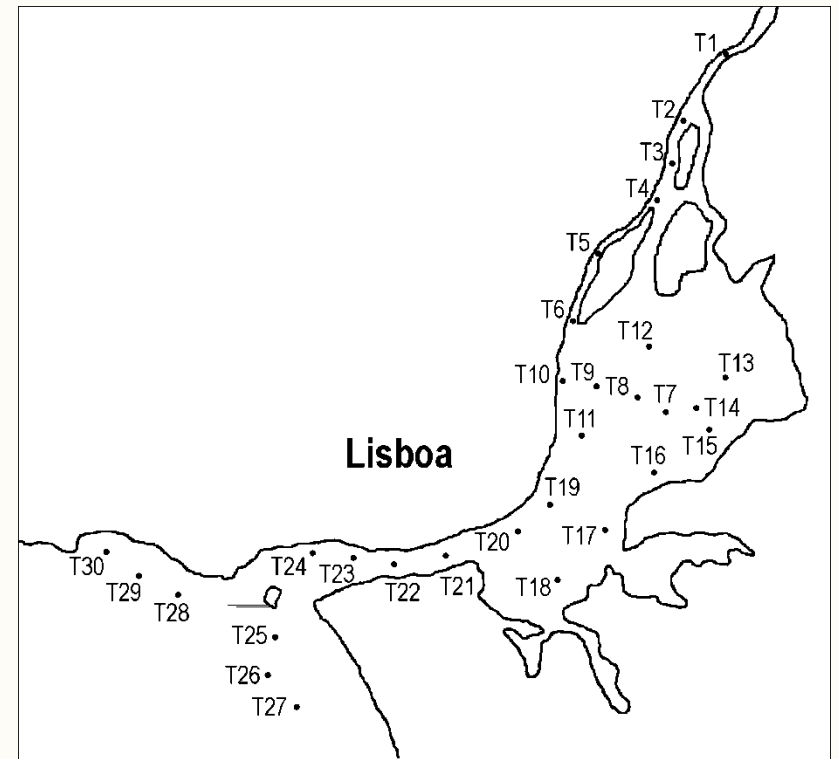
estac	prof	transp	temp	OD	sal	MO	clorofila	feofitina	vasa	areia fina	areia med	cascalho
Inv 1	10.945	31.667	16.946	7.720	0.791	2.344	7.460	6.196	0.413	3.364	81.046	15.416
Inv 2	5.608	35.000	17.006	6.863	4.558	2.668	4.967	5.879	1.141	54.634	44.298	0.009
Inv 3	5.309	35.000	16.944	6.730	8.163	2.508	3.864	4.154	1.745	19.986	67.714	10.524
Inv 4	3.669	35.000	16.954	6.498	8.281	2.122	3.297	8.127	92.113	6.428	1.422	0.148
Inv 5	3.720	25.000	16.310	7.837	8.457	4.534	2.911	5.491	96.244	2.773	0.882	0.055
Inv 6	4.718	30.000	16.361	6.451	15.233	6.587	2.400	4.702	66.960	26.652	6.167	0.991
Inv 7	3.172	50.000	13.591	7.693	19.053	4.383	0.964	1.808	93.234	6.658	0.304	0.049
Inv 8	3.851	47.500	14.538	6.668	20.734	4.329	1.611	4.030				
Inv 9	6.585	41.667	14.608	6.643	20.622	5.252	2.143	3.167	67.662	4.255	28.886	1.111
Inv 10	4.272	17.500	13.983	7.370	17.841	6.266	2.043	4.312	81.845	8.412	2.415	7.492
Inv 11	4.984	40.000	14.822	6.658	20.230	4.686	1.021	2.467	90.362	2.911	5.232	1.556
Inv 12	4.391	42.500	16.314	6.369	26.621	5.371	2.108	3.037	63.959	32.660	3.095	0.248
Inv 13	5.933	55.000	14.407	6.539	23.139	0.000	0.157	0.527	47.800	40.112	12.124	0.116
Inv 14	9.900	65.000	14.352	6.270	24.559	2.788	0.064	0.526	21.297	74.636	4.000	0.142
Inv 15	11.350	35.000	16.371	6.374	27.348	0.000	2.245	3.057	46.906	32.356	16.828	4.112
Inv 16	10.200	40.000	14.020	7.780	20.640	2.299	0.193	1.107	54.915	14.037	23.470	7.669
Inv 17	6.700	45.000	14.230	7.853	20.546	1.607	0.623	1.418	28.555	5.446	2.062	63.661
Inv 18	7.300	47.500	14.064	7.883	20.398	2.363	0.299	0.755	19.409	14.529	43.248	22.911
Inv 19	13.139	72.500	13.870	7.265	26.060	5.053	0.964	1.844	40.537	43.881	15.514	0.077
Inv 20	14.197	80.000	13.863	7.168	26.064	5.547	0.584	1.392	91.315	8.518	0.202	0.009
Inv 21	13.025	70.000	13.764	10.396	15.482	5.310	1.039	4.718	44.442	42.492	3.598	0.331
Inv 22	10.912	55.000	13.883	9.349	24.571	6.252	0.593	4.239	80.958	8.600	8.747	1.716
Inv 23	11.234	35.000	13.738	6.783	26.765	5.734	0.775	5.423	69.507	27.493	2.410	0.684
Inv 25	9.904	445.000	15.013	7.573	33.479	4.266	1.488	0.520	1.270	46.007	52.769	0.187
Inv 26	9.644	292.500	14.818	7.460	31.021	3.073	1.507	0.581	1.952	67.664	30.077	0.287
Inv 27	12.960	200.550	15.022	7.754	33.383	3.498	0.971	1.222	1.177	30.944	67.973	0.104
Inv 28	17.250	240.000	14.931	7.586	33.203	1.911	1.258	1.010	2.131	89.226	6.032	2.493
Inv 29	15.709	340.000	14.882	7.261	33.119	3.807	1.269	1.491	3.934	87.823	7.001	1.182
Inv 30	17.706	350.000	15.128	8.873	33.310	2.956	1.287	1.102	2.517	86.175	11.074	0.243
Ver 1	10.201	55.000	22.464	6.378	7.744	4.062	16.925	10.477	7.985	26.217	54.496	11.616
Ver 2	3.337	60.000	22.544	6.839	11.095	3.808	22.470	10.679	6.335	41.038	51.983	0.708
Ver 3	2.713	67.500	22.404	6.730	15.250	4.471	22.404	9.854	1.081	6.731	82.892	9.519
Ver 4	2.052	42.500	22.258	6.460	18.114	5.221	24.139	10.835	84.109	10.989	4.703	0.259
Ver 5	1.939	45.000	22.007	6.043	22.561	6.058	26.801	10.008	72.300	7.457	19.876	1.045
Ver 6	2.457	40.000	21.903	5.535	25.360	7.462	25.345	12.363	98.847	0.865	0.432	0.038
Ver 7	3.261	75.000	20.836	6.406	30.503	4.105	8.526	3.456	95.630	4.119	0.168	0.009
Ver 8	4.217	60.000	21.126	6.313	30.076	3.616	10.424	5.845	17.001	22.741	49.045	11.123
Ver 9	6.412	60.000	21.653	6.496	28.714	5.048	12.446	6.461	28.745	38.942	30.967	1.164
Ver 10	6.059	30.000	21.561	6.054	29.150	2.769	8.537	7.154	94.867	4.735	0.427	0.076
Ver 11	6.270		21.448	6.275	29.035	3.950	12.717	6.826	22.207	32.806	31.434	13.489
Ver 12	4.431	77.500	21.935	7.164	28.443	3.073	16.136	5.499	22.735	72.745	4.096	0.620
Ver 13	7.098	67.500	20.374	6.208	30.430	5.907	11.116	10.081	79.932	16.104	3.694	0.201
Ver 14	9.603	80.000	20.014	5.626	31.230	6.028	8.861	7.909	99.085	0.951	0.027	0.046
Ver 15	8.740	50.000	20.391	5.690	30.649	6.249	14.407	15.980	64.017	33.891	1.595	0.480
Ver 16	11.932	50.000	20.446	5.613	30.638	6.111	13.039	11.507	51.523	5.394	31.672	11.532
Ver 17	8.447	40.000	20.353	5.611	30.629	5.672	12.441	11.183	25.958	24.978	27.125	21.890
Ver 18	7.793	50.000	20.676	5.530	30.720	5.368	3.047	4.624	72.759	22.683	3.026	1.305
Ver 19	12.625	100.000	18.995	7.008	32.803				26.484	48.084	24.464	1.194
Ver 20	18.020	105.000	17.478	6.805	34.023	4.091	2.886	0.000	22.542	58.153	11.989	7.663
Ver 21	14.732	150.000	18.113	6.635	34.235	0.000	4.100	0.000	79.160	19.087	1.300	0.062
Ver 22	13.724	70.000	18.668	4.936	33.136	0.000	2.854	1.499	87.394	11.140	1.364	0.111
Ver 23	4.683	75.000	18.325	4.766	33.435	0.000	2.562	0.742	82.010	16.020	1.669	0.218
Ver 24	12.599	145.000	18.261	4.803	33.426	0.000	2.406	0.257	14.530	17.347	25.410	42.818
Ver 25	9.074	290.000	17.843	3.833	34.598				1.149	11.384	87.377	0.100
Ver 26	9.977	340.000	17.690	3.853	34.663				1.940	40.636	57.424	0.009
Ver 27	12.383	340.000	17.785	3.925	34.610				2.245	57.194	40.383	0.169
Ver 28	17.340	400.000	17.090	3.625	34.808				3.732	91.462	4.045	0.596
Ver 29	15.800		16.938	3.570	34.893				8.031	75.070	16.560	0.418
Ver 30	17.490		17.190	3.808	34.785				4.653	86.657	7.145	1.458



## introdução à análise multivariada

Amostragens em 30 estações, em duas épocas: arrasto de vara e medição das variáveis ambientais (prof, T, Sal, transp, OD, MO susp, clorof, feofitina, % vasa, % areia fina, % areia média e % cascalho);

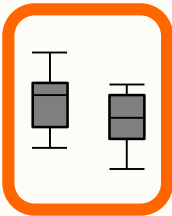
Estimativas da densidade.



# Potenciais objetivos

1. Agrupar (agrupamento ou classificação) ou organizar (ordenação) locais mais semelhantes entre si em função das espécies presentes
2. Identificar padrões de abundância de grupos de espécies em relação às variáveis ambientais consideradas



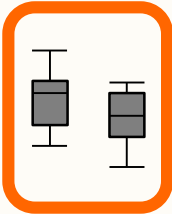


## introdução à análise multivariada

# Matriz de dados

- $x_{jk}$  = measurement of the  $k^{th}$  variable on the  $j^{th}$  entity.

	Variable 1	Variable 2	...	Variable $k$	...	Variable $p$
Item 1:	$x_{11}$	$x_{12}$	...	$x_{1k}$	...	$x_{1p}$
Item 2:	$x_{21}$	$x_{22}$	...	$x_{2k}$	...	$x_{2p}$
⋮	⋮	⋮		⋮		⋮
Item $j$ :	$x_{j1}$	$x_{j2}$	...	$x_{jk}$	...	$x_{jp}$
⋮	⋮	⋮		⋮		⋮
Item $n$ :	$x_{n1}$	$x_{n2}$	...	$x_{nk}$	...	$x_{np}$

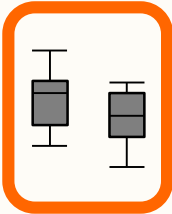


## introdução à análise multivariada

---

# Matriz de dados

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2k} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{j1} & x_{j2} & \dots & x_{jk} & \dots & x_{jp} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} & \dots & x_{np} \end{bmatrix}$$



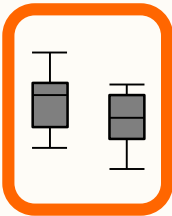
# Estatísticas descritivas multivariadas: média

- For the  $k^{th}$  variable, the sample mean is:

$$\bar{x}_k = \frac{1}{n} \sum_{j=1}^n x_{jk}$$

- An array of the means for all  $p$  variables then looks like this (which we will come to know as the mean vector):

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \\ \bar{x}_4 \end{bmatrix}$$



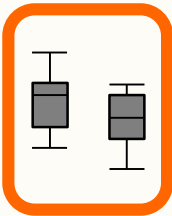
# Estatísticas descritivas multivariadas: variância e covariância

- For the  $k^{th}$  variable, the sample variance is:

$$s_k^2 = s_{kk} = \frac{1}{n} \sum_{j=1}^n (x_{jk} - \bar{x}_k)^2$$

- Note the “kk” subscript, this will be important because the equation that produces the variance for a single variable is a derivation of the equation of the covariance for a pair of variables.
- Also note the division by  $n$ . Reasons for this will become apparent in the near future.
- For a pair of variables,  $i$  and  $k$ , the sample covariance is:

$$s_{ik} = \frac{1}{n} \sum_{j=1}^n (x_{ji} - \bar{x}_i)(x_{jk} - \bar{x}_k)$$



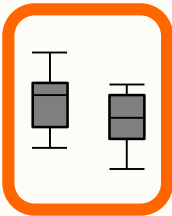
## introdução à análise multivariada

# Estatísticas descritivas multivariadas: Matriz de variância - covariância

$$\mathbf{S}_n = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1p} \\ s_{21} & s_{22} & \dots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \dots & s_{pp} \end{bmatrix}$$

```
> round(cov(iris[,1:4]),2)
      Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length      0.69      -0.04        1.27         0.52
Sepal.Width       -0.04       0.19       -0.33        -0.12
Petal.Length      1.27      -0.33         3.12         1.30
Petal.Width       0.52      -0.12         1.30         0.58
```





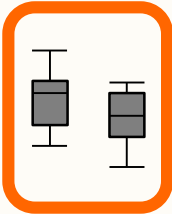
## introdução à análise multivariada

---

# Estatísticas descritivas multivariadas: Matriz de correlação

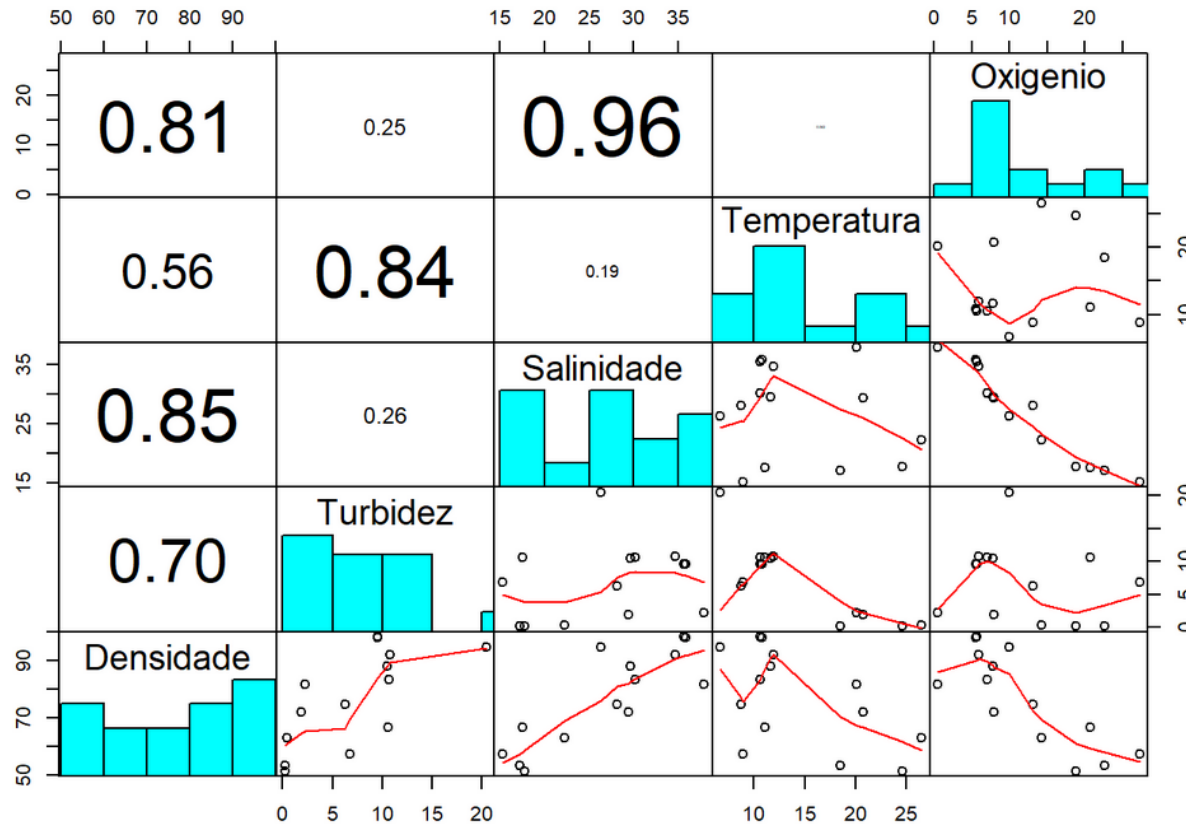
```
> round(cor(iris[,1:4]),2)
```

	Sepal.Length	Sepal.width	Petal.Length	Petal.width
Sepal.Length	1.00	-0.12	0.87	0.82
Sepal.width	-0.12	1.00	-0.43	-0.37
Petal.Length	0.87	-0.43	1.00	0.96
Petal.width	0.82	-0.37	0.96	1.00

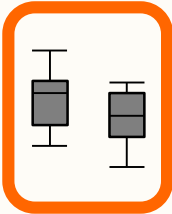


# introdução à análise multivariada

## Métodos gráficos

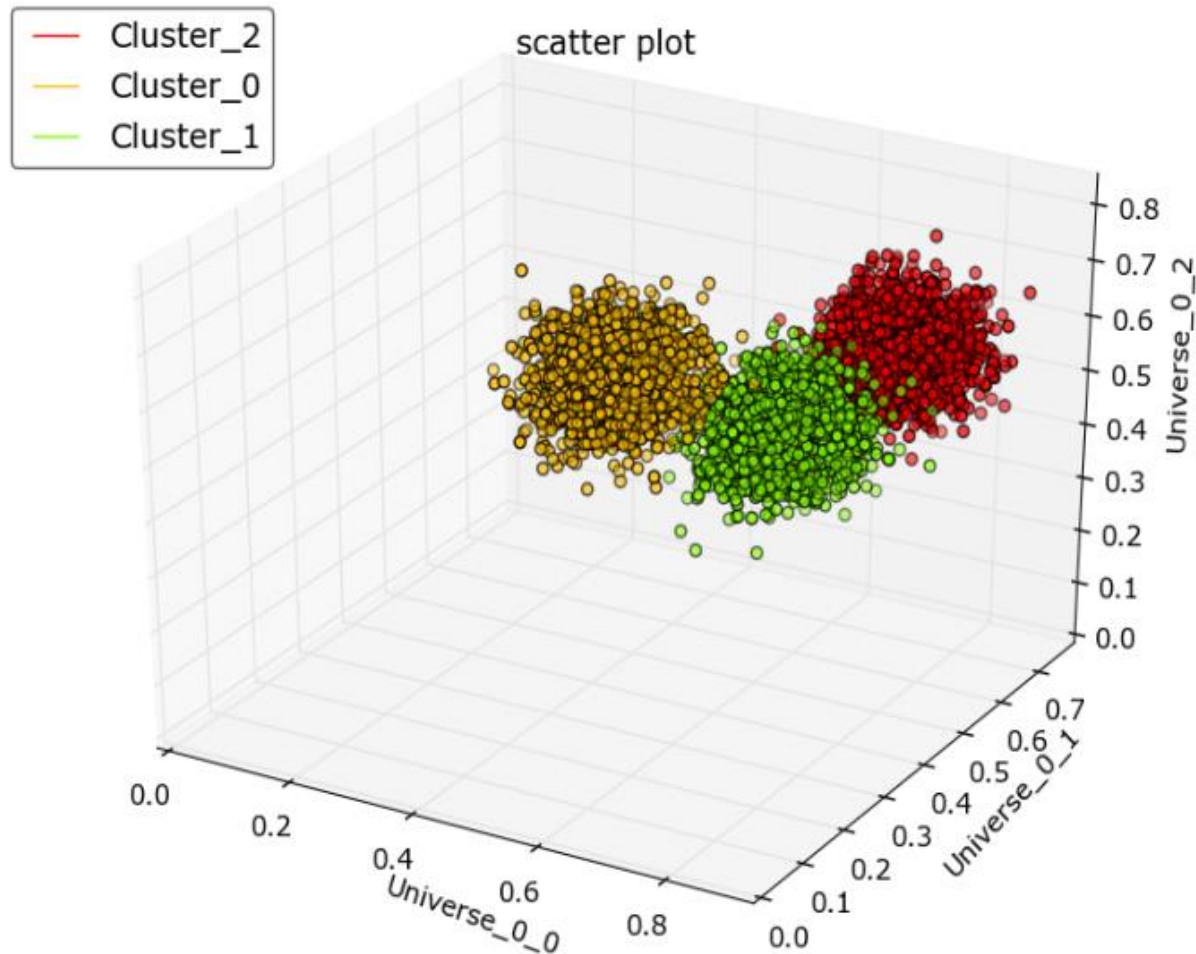


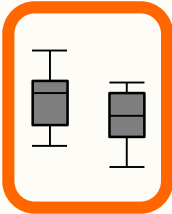
Ver código no help da função `pairs` ou na resolução da ficha FT7b



introdução à análise multivariada

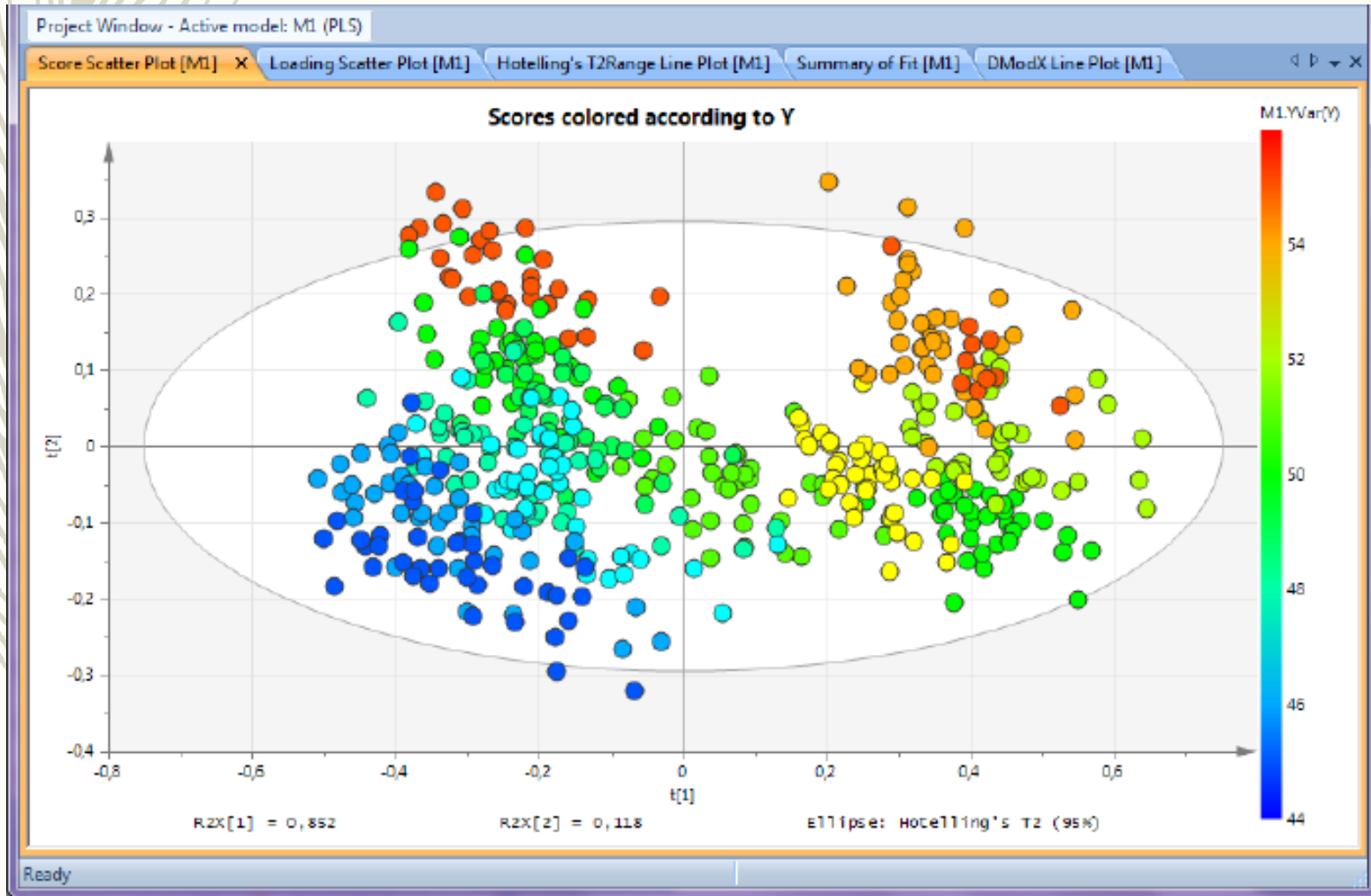
# Métodos gráficos

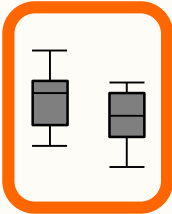




# introdução à análise multivariada

## Métodos gráficos

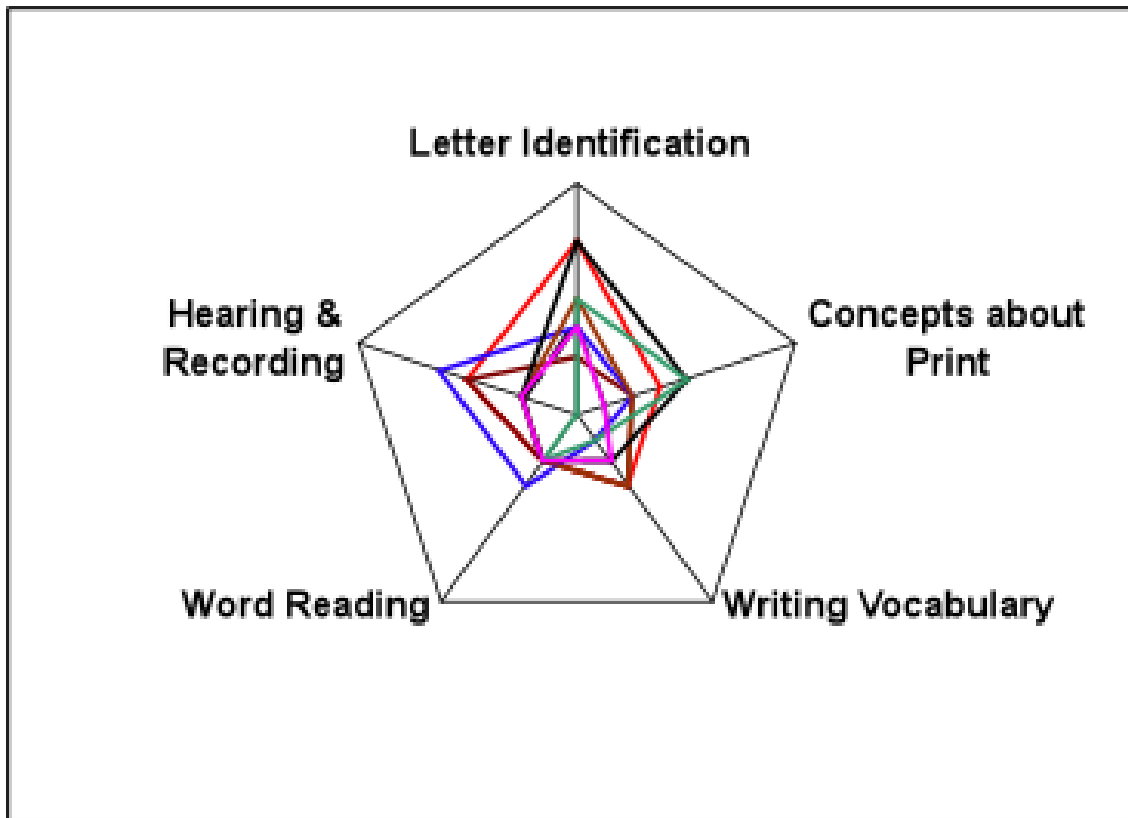




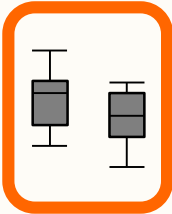
introdução à análise multivariada

---

# Métodos gráficos



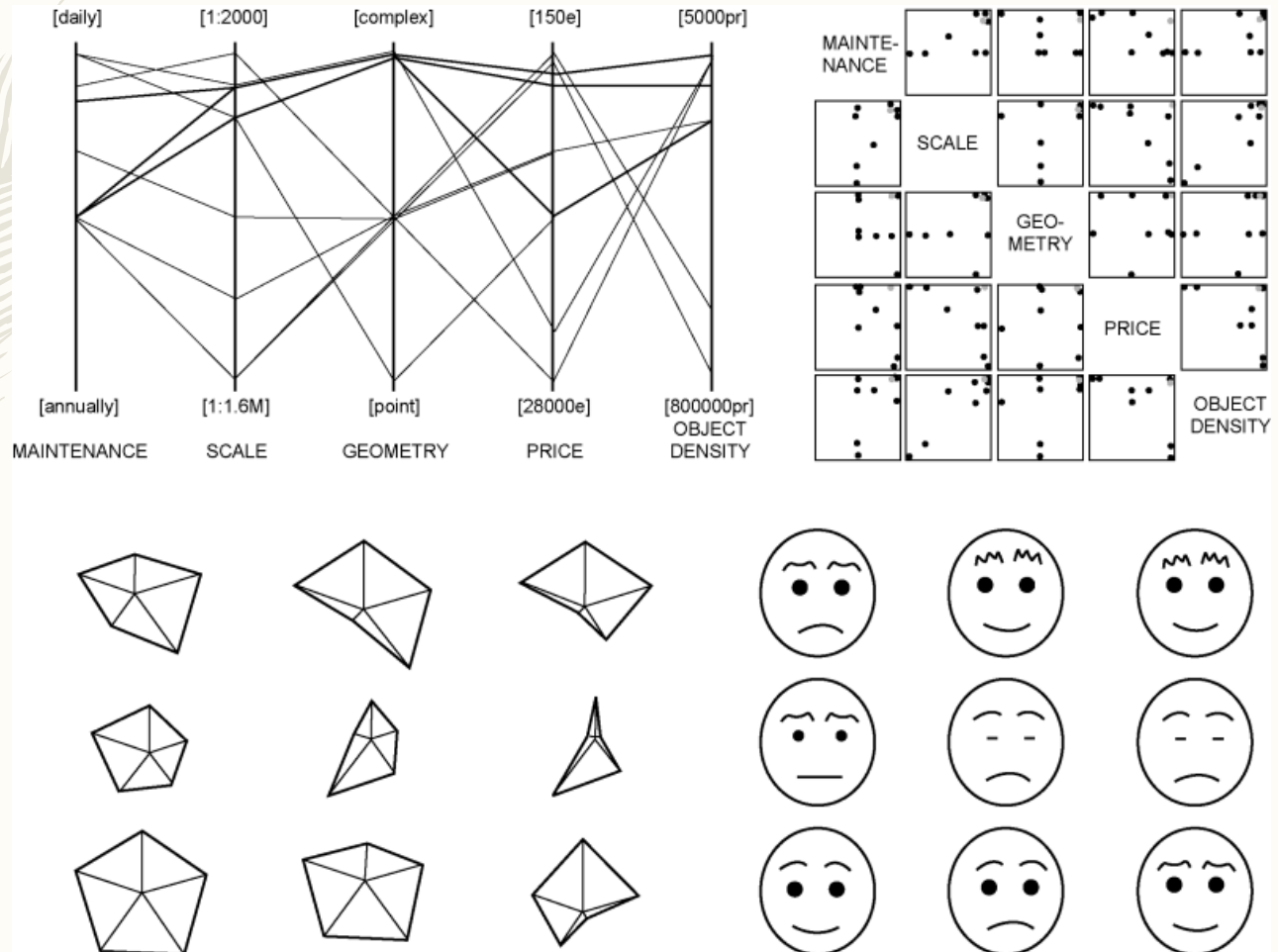


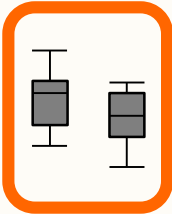


# introdução à análise multivariada

## Métodos gráficos

The metadata has been visualized as a parallel coordinates plot (upper left), a scatterplot matrix (upper right), star plots (lower left) and Chernoff faces (lower right). The axes of the star plots run clockwise in the same order as the sequence shown in the parallel coordinates plot, starting from the upright axis. The Chernoff faces express only three variables: maintenance frequency by the curvature of mouth, scale fraction by the size of eyes and price by the density of eyebrows. Evaluation of the datasets becomes easier if the user first organizes the values of each variable on the basis of their fitness for an intended use.





## Principais técnicas

Classificação (cluster analysis)

Ordenação

Hierárquica  
Não hierárquica

ACP (PCA)

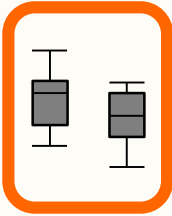
AC (CA)

ACC (CCA)

ACoP (PCoA)

EMD (MDS)

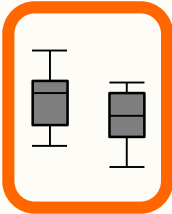
AD (DA)



## classificação

---

- Como definir grupos de observações/variáveis?
- Como explorar padrões nos dados com base em agrupamentos?



classificação

---

# Classificação

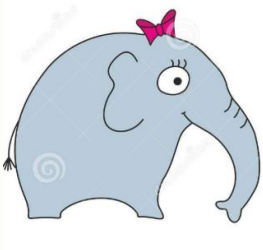
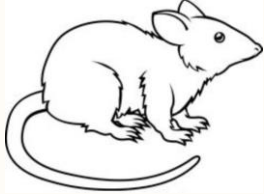
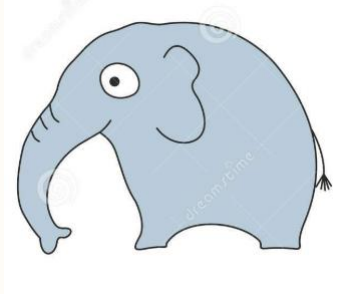
*Principais utilizações:*

Redução de dimensionalidade

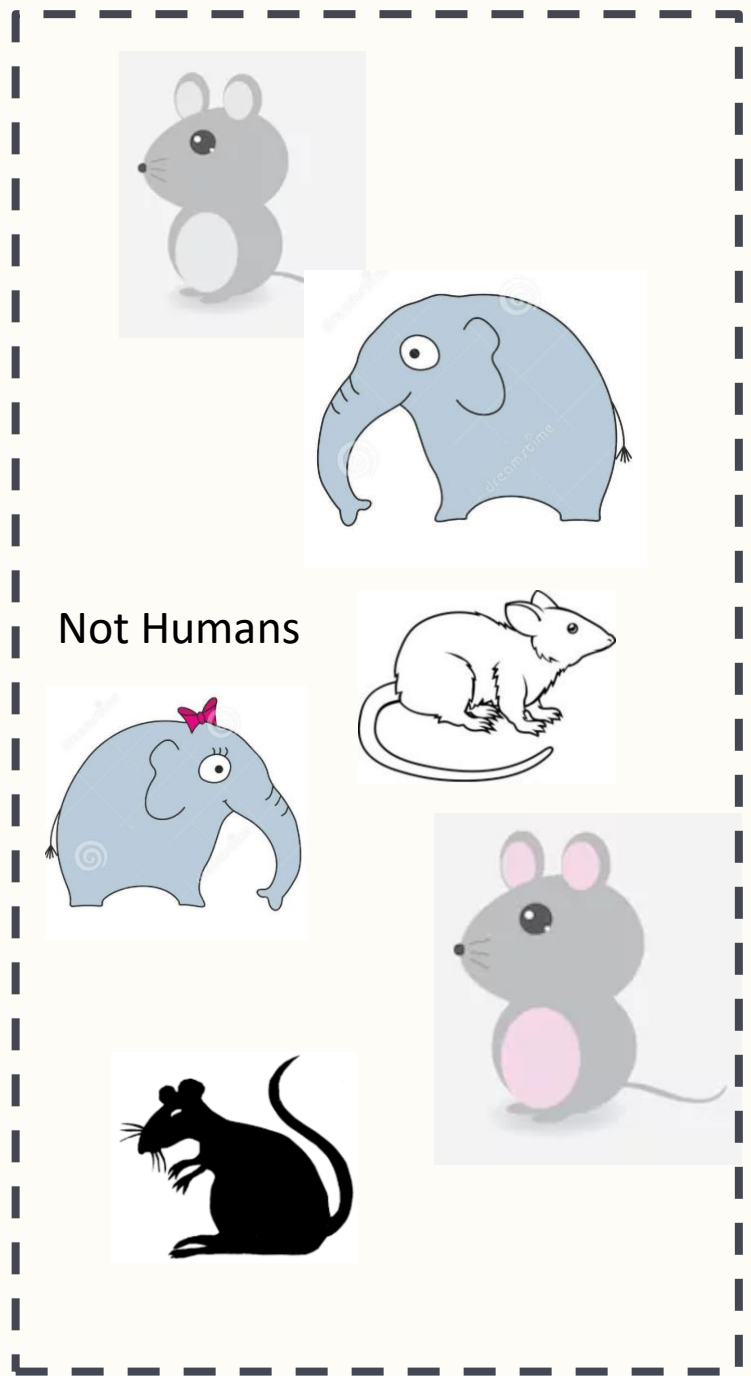
Evidenciar padrões nos dados

Avaliação da consistência de hipóteses ou tipologias

# Classificação (ou agrupamento, clustering)



# Classificação (ou agrupamento, clustering)





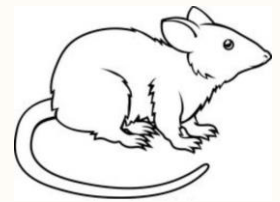
# Classificação (ou agrupamento, clustering)



Color

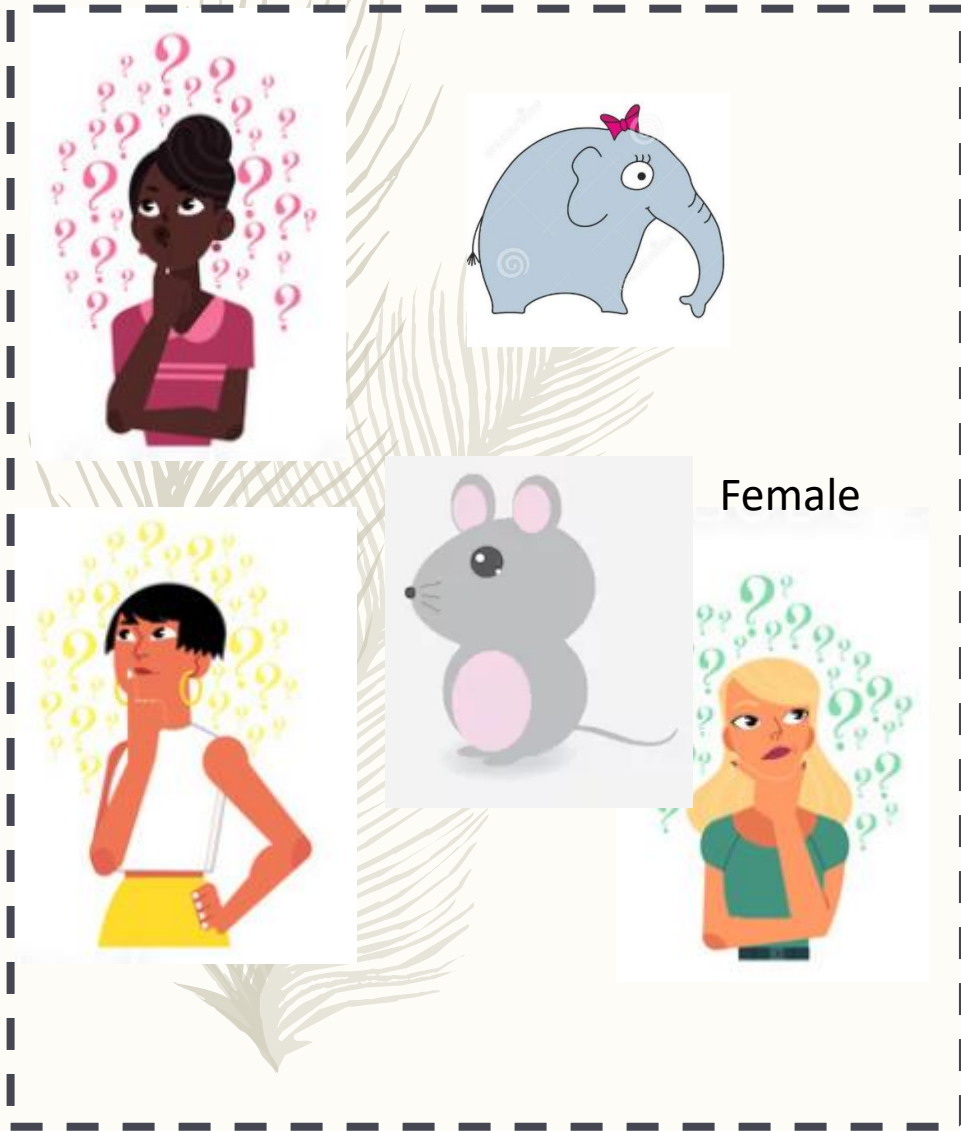


Black & White

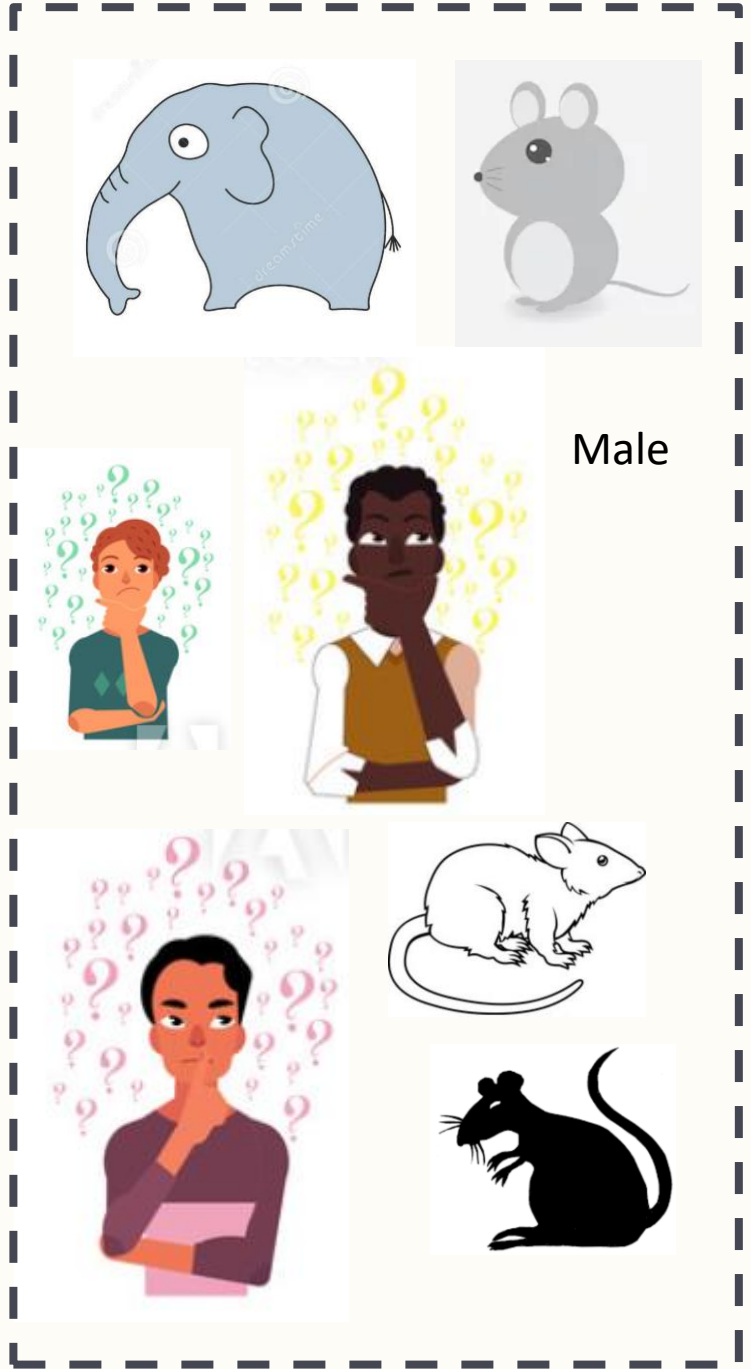




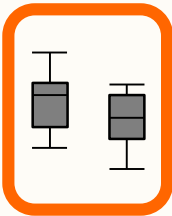
# Classificação (ou agrupamento, clustering)



Female



Male



agrupamento

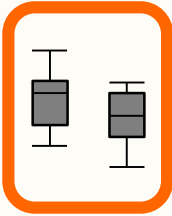
---

## Classificação

*Principais etapas da análise classificativa:*

(dependentes do objectivo a atingir, e dos critérios utilizados)

1. Selecção de medidas de semelhança
2. Selecção do algoritmo de aglomeração



agrupamento

---


## Considerações

- Abordagem exploratória vs. confirmatória;
- A selecção das variáveis utilizadas para caracterizar os objectos deve ser criteriosa;
- A inclusão de variáveis indeferenciadas ou colineares afectam negativamente a análise – é indispensável uma análise exploratória dos dados prévia.

Use R!

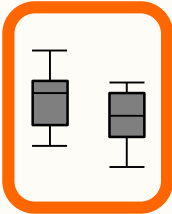
Daniel Borcard  
François Gillet  
Pierre Legendre

# Numerical Ecology with R

 Springer

Métodos heurísticos (geralmente que fazem algum sentido do ponto de vista prático mas sem grande sustentação teórica)

a **heuristic**, is any approach to problem solving, learning, or discovery that employs a practical method, not guaranteed to be optimal, perfect, logical, or rational, but instead sufficient for reaching an immediate goal



agrupamento

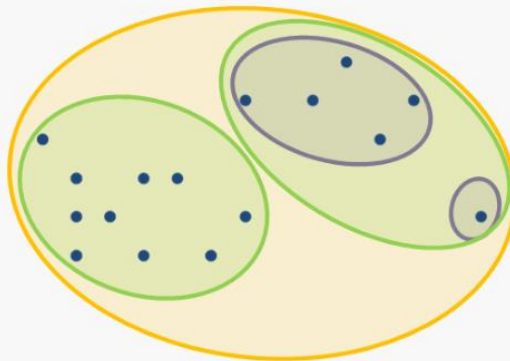
# Métodos de classificação

- Métodos hierárquicos
- Métodos não hierárquicos

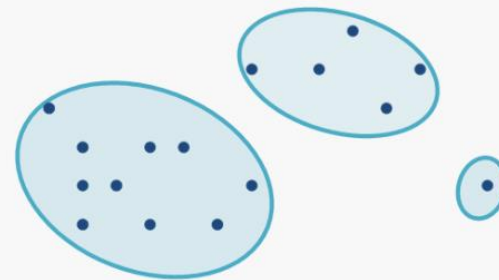
Aglomerativos

Divisivos

Hierarchical Clustering



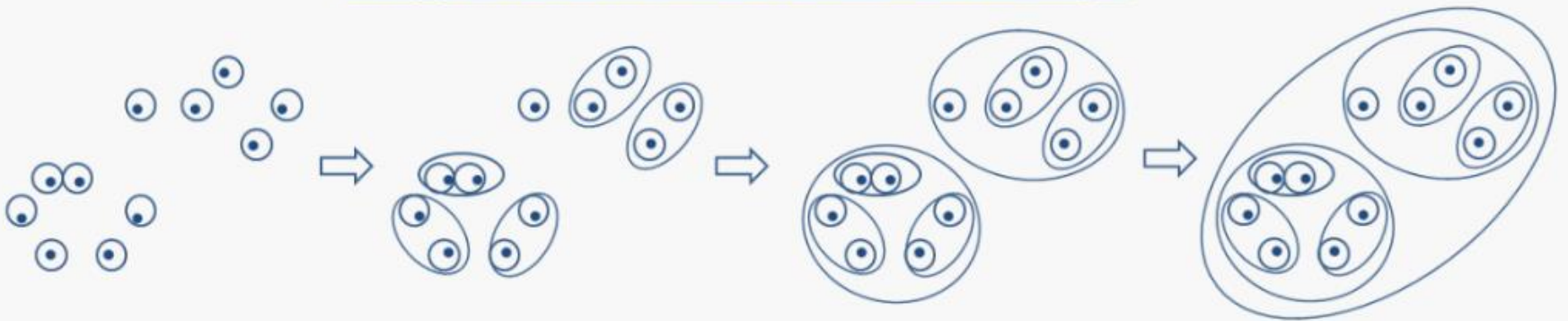
Partitional Clustering



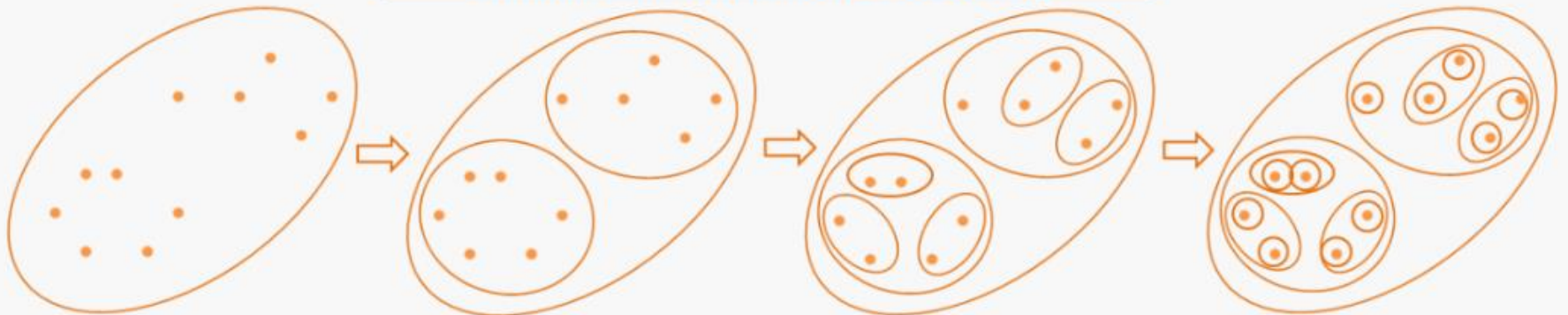


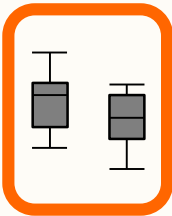
# Métodos aglomerativos vs. divisivos

Agglomerative Hierarchical Clustering



Divisive Hierarchical Clustering





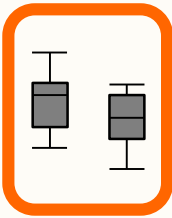
agrupamento

---

## Seleccção de medidas de semelhança

- Como medir a semelhança/dissemalhança entre objectos?
- Será conveniente efectuar transformação dos dados?





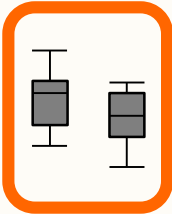
agrupamento

---

# Seleccção de medidas de semelhança

## *Tipos de medidas de semelhança/dissemelhança*

- **Medidas de correlação:** avaliam a correspondência e semelhança dos padrões (tipicamente para análises em modo R – comparar descritores)
- **Medidas de distância:** proximidade multidimensional entre os objectos (tipicamente para análises em modo Q – comparar objectos)
- **Medidas de associação:** para variáveis nominais ou ordinais, mede o grau de associação ou concordância entre pares de objectos (tipicamente para análises em modo Q – comparar objectos)



agrupamento

## Variáveis medidas em escalas intervaladas

Distância euclideana:

$$\sqrt{\sum (X_i - Y_i)^2}$$

Distância de Minkowski:

$$\left( \sum |X_i - Y_i|^p \right)^{\frac{1}{p}}$$

Distância euclideana quadrada:

$$\sum (X_i - Y_i)^2$$

Distância potência:

$$\left( \sum |X_i - Y_i|^p \right)^{\frac{1}{r}}$$

Distância de Chebychev  
("maximum"):

$$\max_i |X_i - Y_i|$$

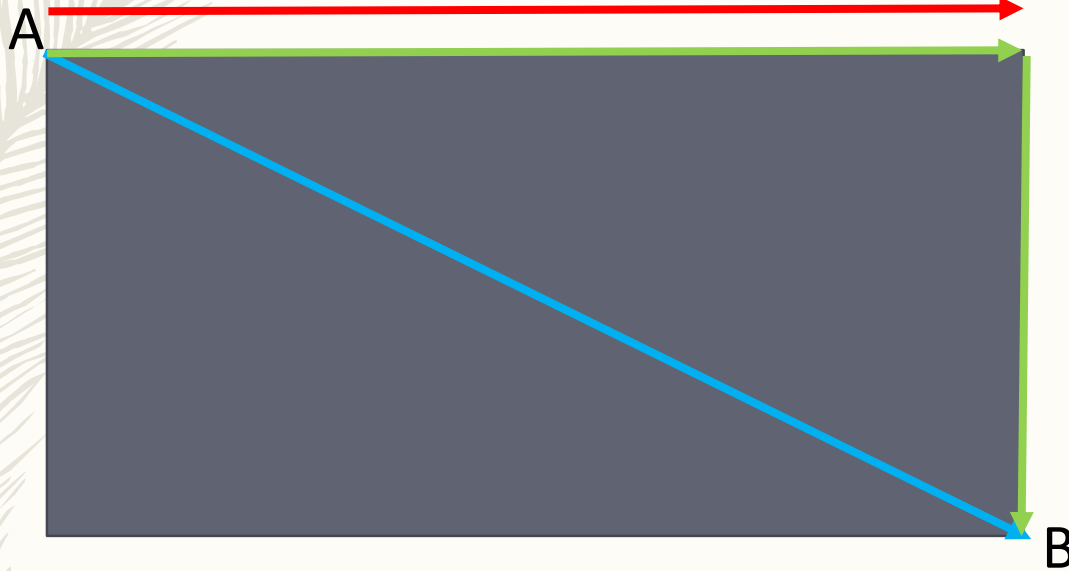
Distância de City-Block  
(ou Manhattan):

$$\sum |X_i - Y_i|$$

Distância  
Euclideana

Distância  
Chebychev

Distância de City-Block  
(ou Manhattan):



Função `dist` – por default calcula as distâncias entre as linhas de um objecto!

```
> set.seed(123)
> x <- matrix(rnorm(30), nrow = 5)
> x
      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
[1,] -0.56047565  1.7150650  1.2240818  1.7869131 -1.0678237 -1.6866933
[2,] -0.23017749  0.4609162  0.3598138  0.4978505 -0.2179749  0.8377870
[3,]  1.55870831 -1.2650612  0.4007715 -1.9666172 -1.0260044  0.1533731
[4,]  0.07050839 -0.6868529  0.1106827  0.7013559 -0.7288912 -1.1381369
[5,]  0.12928774 -0.4456620 -0.5558411 -0.4727914 -0.6250393  1.2538149
> dist(x,diag=TRUE,method="euclidean")
      1      2      3      4      5
1 0.000000
2 3.344529 0.000000
3 5.614848 3.657284 0.000000
4 3.000205 2.382557 3.392281 0.000000
5 4.718215 1.752270 2.689429 2.759841 0.000000
>
> dist(x,diag=TRUE,method="euclidean")^2
      1      2      3      4      5
1 0.000000
2 11.185872 0.000000
3 31.526516 13.375723 0.000000
4  9.001232  5.676578 11.507570 0.000000
5 22.261551  3.070449  7.233027  7.616723 0.000000
```

```
> dist(x,diag=TRUE,method="minkowski",p=2)
```

	1	2	3	4	5
1	0.000000				
2	3.344529	0.000000			
3	5.614848	3.657284	0.000000		
4	3.000205	2.382557	3.392281	0.000000	
5	4.718215	1.752270	2.689429	2.759841	0.000000

```
>
```

```
> dist(x,diag=TRUE,method="manhattan")
```

	1	2	3	4	5
1	0.000000				
2	7.112107	0.000000			
3	11.558037	7.512732	0.000000		
4	6.119347	4.387932	6.613093	0.000000	
5	10.273410	3.975432	6.200665	4.636445	0.000000

```
>
```

```
> dist(x,diag=TRUE,method="maximum")
```

	1	2	3	4	5
1	0.0000000				
2	2.5244804	0.0000000			
3	3.7535303	2.4644676	0.0000000		
4	2.4019178	1.9759240	2.6679731	0.0000000	
5	2.9405082	0.9706419	1.4938257	2.3919519	0.0000000

# Uma nota sobre a relação entre distâncias

distância de Minkowski com  $p=2$   
=  
distancia potencia com  $p=2$  e  $r=2$   
=  
distância Euclideana !

```
> p=2; r=2; sum(abs(x[1,]-x[2,])^p)^(1/r)
[1] 3.344529
>
> p=3; r=1.5; sum(abs(x[1,]-x[2,])^p)^(1/r)
[1] 7.731666
```

Distância de Minkowski:

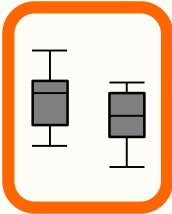
$$\left( \sum |X_i - Y_i|^p \right)^{\frac{1}{p}}$$

Distância potência:

$$\left( \sum |X_i - Y_i|^p \right)^{\frac{1}{r}}$$

Distância Euclideana:

$$\sqrt{\sum (X_i - Y_i)^2}$$



agrupamento

---

## Dados de contagens

Distância do Qui-quadrado:

$$\sqrt{\sum \frac{(X_i - E(X_i))^2}{E(X_i)} + \sum \frac{(Y_i - E(Y_i))^2}{E(Y_i)}}$$



```
> set.seed(123)
> x <- matrix(rpois(30,2), nrow = 5)
> x
```

```
      [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    1    0    5    4    4    3
[2,]    3    2    2    1    3    2
[3,]    2    4    3    0    2    2
[4,]    4    2    2    1    6    1
[5,]    4    2    0    5    2    1
```

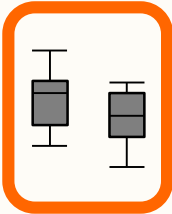
distance is in  
`library(anaLogue)`

```
>
> as.matrix(distance(x, method = "chi.distance"))
      1      2      3      4      5
1 0.000000 0.7857024 1.0782705 0.9305595 1.0809102
2 0.7857024 0.0000000 0.5516414 0.4140697 0.8774957
3 1.0782705 0.5516414 0.0000000 0.8145777 1.2308826
4 0.9305595 0.4140697 0.8145777 0.0000000 0.9551025
5 1.0809102 0.8774957 1.2308826 0.9551025 0.0000000
attr(,"method")
[1] "chi.distance"
attr(,"class")
[1] "distance" "matrix"
attr(,"type")
[1] "symmetric"
> as.matrix(distance(x, method = "chi.square"))
      1      2      3      4      5
1 0.000000 2.535463 3.114482 2.878492 3.252350
2 2.535463 0.000000 1.505545 1.214986 2.311462
3 3.114482 1.505545 0.000000 2.206052 3.109126
4 2.878492 1.214986 2.206052 0.000000 2.581989
5 3.252350 2.311462 3.109126 2.581989 0.000000
```

$$\sqrt{\sum \frac{(X_i - E(X_i))^2}{E(X_i)} + \sum \frac{(Y_i - E(Y_i))^2}{E(Y_i)}}$$

```
sqrt(  
sum(((x[,1]-mean(x[,1]))^2)/mean(x[,1])))  
+  
sum(((x[,2]-mean(x[,2]))^2)/mean(x[,2])))  
[1] 2.535463
```

```
> as.matrix(distance(x, method = "chi.square"))  
      1      2      3      4      5  
1 0.000000 2.535463 3.114482 2.878492 3.252350
```



agrupamento

# Dados binários (medidas de associação)

Russel & Rao:

$$\frac{a}{a + b + c + d}$$

Jaccard:

$$\frac{a}{a + b + c}$$

Dice:

$$\frac{2a}{2a + b + c}$$

Pensar... o que me interessa  
realçar...  
ou  
ignorar!

	P (1)	A (0)
P (1)	a	b
A (0)	c	d

library(vegan)

```
vegdist(x, method="bray", binary=FALSE, diag=FALSE, upper=FALSE,  
        na.rm = FALSE, ...)
```

## Arguments

**x** Community data matrix.

**method** Dissimilarity index, partial match to "manhattan", "euclidean", "canberra", "clark", "bray", "kulczynski", "jaccard", "gower", "altGower", "morisita", "horn", "mountford", "raup", "binomial", "chao", "cao" or "mahalanobis".

```
> dist(x,diag=TRUE,method="manhattan")  
      1      2      3      4      5  
1 0.000000  
2 7.112107 0.000000  
3 11.558037 7.512732 0.000000  
4 6.119347 4.387932 6.613093 0.000000  
5 10.273410 3.975432 6.200665 4.636445 0.000000  
> vegdist(x,diag=TRUE,method="manhattan")  
      1      2      3      4      5  
1 0.000000  
2 7.112107 0.000000  
3 11.558037 7.512732 0.000000  
4 6.119347 4.387932 6.613093 0.000000  
5 10.273410 3.975432 6.200665 4.636445 0.000000
```

```
> presabs=matrix(sample(0:1,size=80,replace=TRUE),ncol=8,nrow=10,dimnames=list(sites=paste0("Site",1:10),sps=paste0("Sp",1:8)))
```

```
>
```

```
> presabs
```

```
      sps
sites sp1 sp2 sp3 sp4 sp5 sp6 sp7 sp8
Site1  1  0  0  0  0  1  1  0
Site2  1  1  0  0  1  1  1  0
Site3  1  1  1  1  1  1  1  1
Site4  0  1  0  1  0  1  1  1
Site5  1  1  1  1  1  1  1  0
Site6  0  0  0  1  1  1  1  1
Site7  1  1  0  0  0  0  1  1
Site8  0  0  0  1  0  0  1  1
Site9  0  0  1  1  0  0  1  0
Site10 0  1  1  1  0  1  0  0
```

```
> round(vegdist(presabs,diag=TRUE,method="jaccard"),2)
```

```
      site1 site2 site3 site4 site5 site6 site7 site8 site9 site10
Site1  0.00
Site2  0.40  0.00
Site3  0.62  0.38  0.00
Site4  0.67  0.57  0.38  0.00
Site5  0.57  0.29  0.12  0.50  0.00
Site6  0.67  0.57  0.38  0.33  0.50  0.00
Site7  0.60  0.50  0.50  0.50  0.62  0.71  0.00
Site8  0.80  0.86  0.62  0.40  0.75  0.40  0.60  0.00
Site9  0.80  0.86  0.62  0.67  0.57  0.67  0.83  0.50  0.00
Site10 0.83  0.71  0.50  0.50  0.43  0.71  0.86  0.83  0.60  0.00
```

```

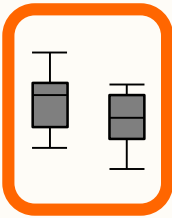
> dist(x=presabs, method = "Russel",diag=TRUE)
      Site1 Site2 Site3 Site4 Site5 Site6 Site7 Site8 Site9 Site10
Site1  0.000
Site2  0.625 0.000
Site3  0.625 0.375 0.000
Site4  0.750 0.625 0.375 0.000
Site5  0.625 0.375 0.125 0.500 0.000
Site6  0.750 0.625 0.375 0.500 0.500 0.000
Site7  0.750 0.625 0.500 0.625 0.625 0.750 0.000
Site8  0.875 0.875 0.625 0.625 0.750 0.625 0.750 0.000
Site9  0.875 0.875 0.625 0.750 0.625 0.750 0.875 0.750 0.000
Site10 0.875 0.750 0.500 0.625 0.500 0.750 0.875 0.875 0.750 0.000
> round(dist(x=presabs, method = "Jaccard",diag=TRUE),2)
      Site1 Site2 Site3 Site4 Site5 Site6 Site7 Site8 Site9 Site10
Site1  0.00
Site2  0.40  0.00
Site3  0.62  0.38  0.00
Site4  0.67  0.57  0.38  0.00
Site5  0.57  0.29  0.12  0.50  0.00
Site6  0.67  0.57  0.38  0.33  0.50  0.00
Site7  0.60  0.50  0.50  0.50  0.62  0.71  0.00
Site8  0.80  0.86  0.62  0.40  0.75  0.40  0.60  0.00
Site9  0.80  0.86  0.62  0.67  0.57  0.67  0.83  0.50  0.00
Site10 0.83  0.71  0.50  0.50  0.43  0.71  0.86  0.83  0.60  0.00
> round(dist(x=presabs, method = "Dice",diag=TRUE),2)
      Site1 Site2 Site3 Site4 Site5 Site6 Site7 Site8 Site9 Site10
Site1  0.00
Site2  0.25  0.00
Site3  0.45  0.23  0.00
Site4  0.50  0.40  0.23  0.00
Site5  0.40  0.17  0.07  0.33  0.00
Site6  0.50  0.40  0.23  0.20  0.33  0.00
Site7  0.43  0.33  0.33  0.33  0.45  0.56  0.00
Site8  0.67  0.75  0.45  0.25  0.60  0.25  0.43  0.00
Site9  0.67  0.75  0.45  0.50  0.40  0.50  0.71  0.33  0.00
Site10 0.71  0.56  0.33  0.33  0.27  0.56  0.75  0.71  0.43  0.00

```

dist in

library(proxy)





agrupamento

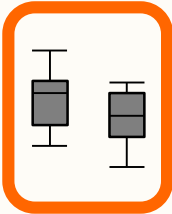
---

# Seleccção de medidas de semelhança

*Tipos de medidas de semelhança/dissemelhança*

- **Medidas de correlação:** avaliam a correspondência e semelhança dos padrões;
- **Medidas de distância:** proximidade multidimensional entre os objectos;
- **Medidas de associação:** para variáveis nominais ou ordinais, mede o grau de associação ou concordância entre pares de objectos.





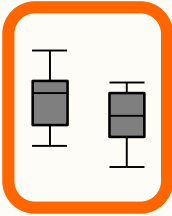
agrupamento

---

# Seleção de medidas de semelhança

## *Transformação dos dados*

- Muitas medidas de distância são particularmente sensíveis a diferenças entre escalas e magnitude dos valores das variáveis;
- Transformação mais frequente é a redução e centragem (Z scores) – dá igual peso a todas as variáveis;
- Transformação em relação às médias das linhas da matriz permite avaliar a importância relativa das várias variáveis – identificação de padrões.



agrupamento

---

# Seleccção de algoritmo de aglomeração

*Métodos de aglomeração (métodos hierárquicos)*

Ligação simples – distância ao vizinho mais próximo (*single linkage*);

Ligação completa – distância ao vizinho mais distante (*complete linkage*);

Método das médias (*average linkage*);

Método da mínima variância (*Ward's method*);

Método dos centróides (*centroid method*).

distância

algoritmo de aglomeração

`hclust {stats}`

R Documentation

# Hierarchical Clustering

## Description

Hierarchical cluster analysis on a set of dissimilarities and methods for analyzing it.

## Usage

```
hclust(d, method = "complete", members = NULL)
```

```
## S3 method for class 'hclust'  
plot(x, labels = NULL, hang = 0.1, check = TRUE,  
      axes = TRUE, frame.plot = FALSE, ann = TRUE,  
      main = "Cluster Dendrogram",  
      sub = NULL, xlab = NULL, ylab = "Height", ...)
```

## Arguments

<code>d</code>	a dissimilarity structure as produced by <code>dist</code> .
<code>method</code>	the agglomeration method to be used. This should be (an unambiguous abbreviation of) one of "ward.D", "ward.D2", "single", "complete", "average" (= UPGMA), "mcquitty" (= WPGMA), "median" (= WPGMC) or "centroid" (= UPGMC).

```
habitats <- read.csv("DataTP9habitats123.csv", sep=";")
```

```
library(cluster)
```

```
teste<-hclust(dist(habitats[,-1],method="manhattan"),method="average")
```

```
par(mfrow=c(1,1))
```

```
plot(teste,labels=habitats[,1])
```

A primeira coluna são os *labels* dos sitios

