

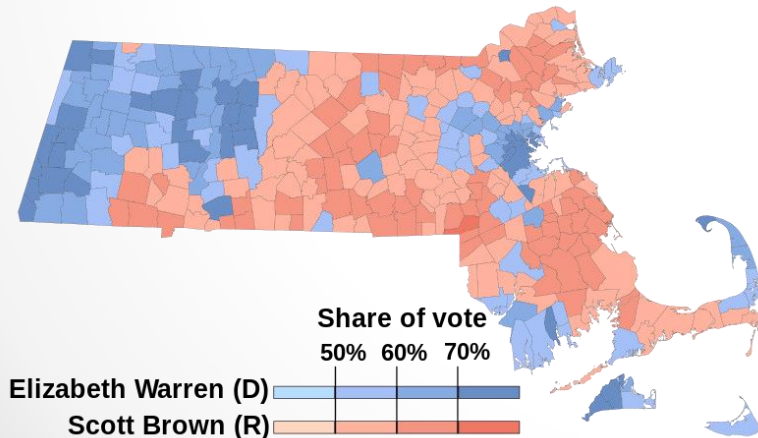
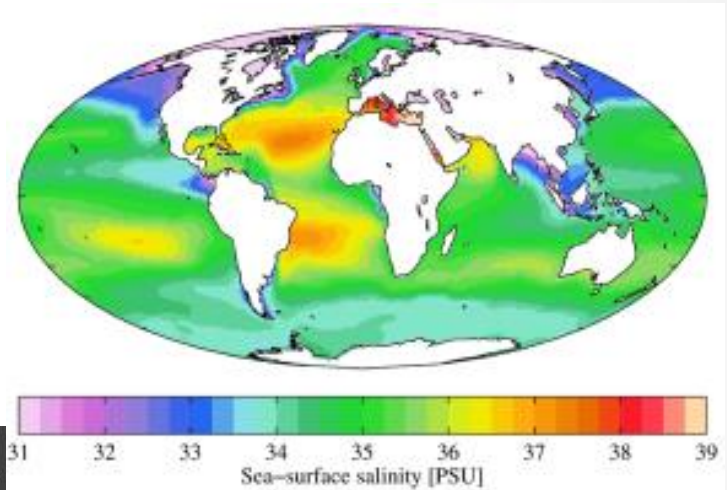
Cluster Heat Map

PA055 Vizualizace komplexních dat

Karel Sedlář

Heat Map

- = forma grafické reprezentace dat, kde jsou hodnoty kódovány do barevné stupnice



- everyone.
- Archive your thoughts.
- Why the heck not?

Hot Blogs Today

1. [Lorelle on WordPress](#)
2. [churumuri](#)
3. [Seobleizer - Microsoft Geek Blogger](#)
4. [Pomme & Kelly](#)
5. [My Journey to Macintosh](#)
6. [Sam Ideas, Thoughts, Programming](#)
7. [El Blog Oficial de Carlos Sicilia](#)
8. [Qwerty Maniac - The Typo Killer](#)
9. [Football Videos](#)
10. [Nosy Snoop](#)
11. [It's a Definite Maybe](#)

More of the [best blogs on WordPress.com](#).

Password:

Remember me

WordPress.com News Blog

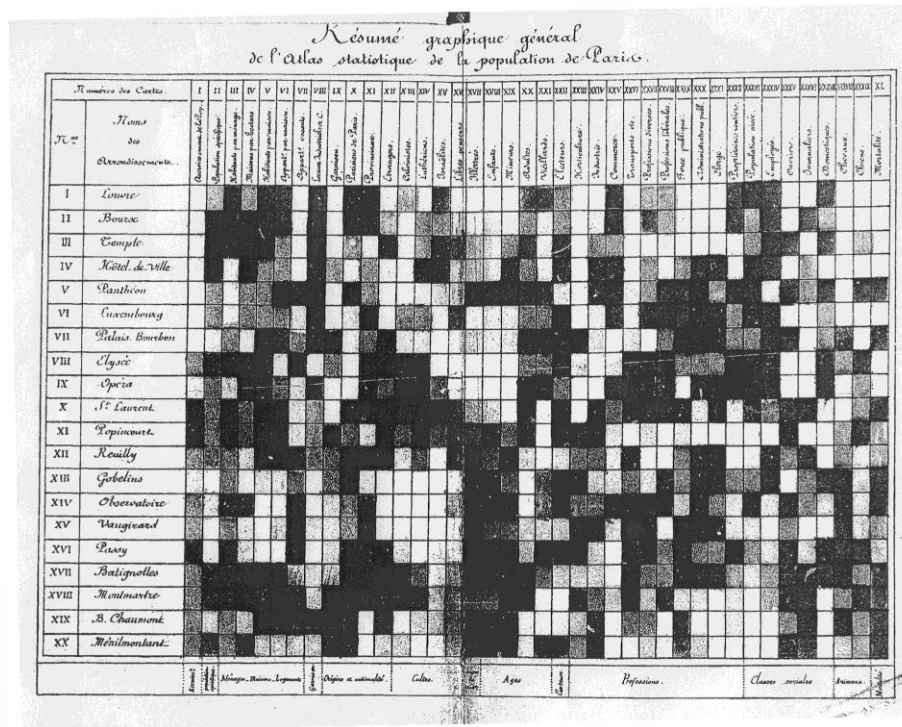
- [Blog Stats Widget](#)
- [Unlimited Blogs](#)
- [Widgets: It's a Plugin!](#)
- [Check Your Spelling](#)
- [Tag Feeds and Paging](#)
- [Andreas04 and Thirteen](#)
- [Login Errors](#)

Metody kláastrování

- nemá omezení
- nejčastěji hierarchické aglomerativní (complete, single, average linkage, wardova metoda)
- možno použít ty nehierarchické divizivní (k-means clustering)

Historie

- 1873 Loua: Shaded matrix (stínovaná matice)
 - vytvořena ručně, matice bez přerovnání
 - zachycuje složení obyvatelstva 20 různých částí Paříže



Historie

- 1914 Brinton: Permuting matrix (zpřeházená matice)
 - upraveno pouze pořadí řádků
- ~1950 vizualizace asociačních matic

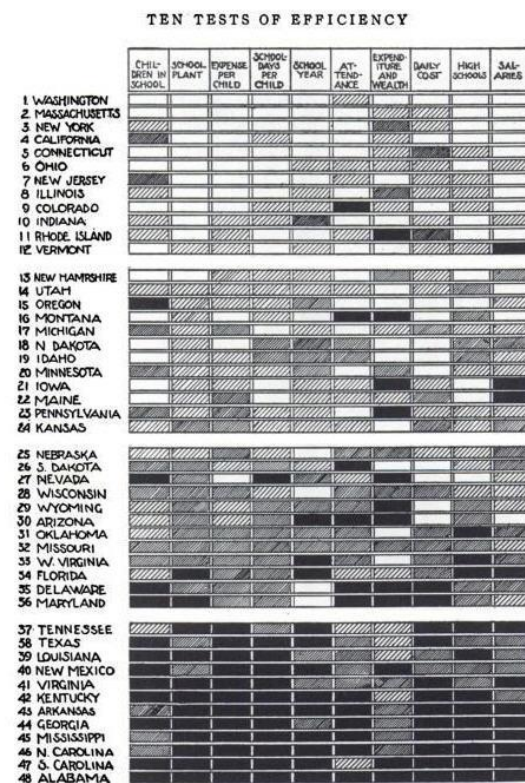


Fig. 33. Rank of States in Each of Ten Educational Features, 1910. White Indicates that the State Ranks in the Highest 12 of the 48, Light Shading that it Ranks in Second 12, Dark Shading that it Ranks in Third 12, and Black that it Ranks in Lowest 12



Microarrays

- typické využití v bioinformatice při zpracování genové exprese
- Eisen, M., Spellman, P., Brown, P., and Botstein, D. (1998), Cluster analysis and display of genome-wide expression patterns, " Proceedings of the National Academy of Sciences", 95, 14863-14868.
- červeno-zelená stupnice
- příklad:
 - sledování genové exprese sinic při vystavení silnému zdroji světla
 - k-means clustering (nehierarchická metoda -> není dendrogram)
 - výrazně zelená skupina genů -> protein kódující geny uplatňující se při tvorbě Phycobilisome, tj. systém antének pro sběr světla
 - software: Cluster + TreeView

Vizualizace vztahů

- nástavba vícerozměrných statistických analýz (PCA, PCoA, CoA, FA...)
- vyjádření korelací rozsáhlých souborů dat
- zpravidla jiná než červeno-zelená stupnice
- příklad:
 - vizualizace korelací složení mikroflóry ryb s genovou expresí genů rezistence
 - metagenomická analýza 16S
 - ampliconové sekvenování na Roche 454

Vizualizace korelací

- OTU table

| phylum | class | 1 | 4 | 5 | 7 | 8 | 9 | 10 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 2 |
|------------|------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| Acidobact | Acidobact | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,014881 | 0 | 0 | 0 | 0 | 0 | 0,051112 | 0 | 0 | 0,028393 | 0 |
| Acidobact | Acidobact | 0 | 0 | 0 | 0 | 0,036417 | 0,014136 | 0,018352 | 0 | 0,016105 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Acidobact | Holophag | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Acidobact | Chloracid | 0 | 0,043057 | 0,067009 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Acidobact | iii1-8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Acidobact | Solibacter | 0 | 0,129171 | 0,2457 | 0,1673 | 0 | 0,014136 | 0 | 0 | 0,005368 | 0 | 0,028711 | 0 | 0,038334 | 0 | 0,013543 | 0 | 0 | 0 |
| Actinobac | Acidimicro | 0 | 0,086114 | 0,156355 | 0,030418 | 0,072833 | 0,028273 | 0 | 0,029762 | 0,016105 | 0 | 0,014355 | 0 | 0,025556 | 0 | 0 | 0,028393 | 0 | 0 |
| Actinobac | Actinobac | 1,186998 | 0,796555 | 0,625419 | 0,1673 | 0,072833 | 0,565451 | 0,165168 | 0,059524 | 0,101997 | 0,033916 | 0,301464 | 0,400802 | 0,907232 | 0,011756 | 0,027086 | 0,48268 | 1,199073 | 0 |
| Actinobac | Thermole | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,005368 | 0 | 0 | 0 | 0 | 0,023513 | 0 | 0 | 0 | 0 |
| Aquificae | Aquificae | 0 | 0 | 0 | 0,015209 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Armatimo | Armatimo | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,013626 |
| Armatimo | S1a-1H | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Bacteroidi | Bacteroidi | 0,328707 | 18,57912 | 39,51307 | 16,98859 | 3,932993 | 19,88974 | 62,96568 | 13,02083 | 13,94138 | 0,203493 | 14,41286 | 3,186373 | 46,31996 | 4,914178 | 3,927411 | 17,54685 | 0,258891 | 0 |
| Bacteroidi | Flavobact | 33,729 | 9,967707 | 1,317847 | 1,855513 | 28,58704 | 9,202714 | 10,91943 | 18,06548 | 2,098991 | 41,13956 | 24,43296 | 31,30261 | 0,523895 | 2,915589 | 6,812026 | 0,468484 | 21,73321 | 0 |
| Bacteroidi | Sphingob | 15,70489 | 6,437029 | 4,26625 | 4,395437 | 4,879825 | 2,176986 | 4,14755 | 24,28571 | 4,520077 | 12,04002 | 7,823715 | 9,699399 | 0,191669 | 1,152128 | 5,701517 | 0,212947 | 33,1244 | 0 |

- gen table

| | 1 | 4 | 5 | 7 | 8 | 9 | 10 | 12 | 13 | 14 | 15 |
|------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| cml | 2,706048 | 2,886196 | 2,538842 | 2,810607 | 2,813911 | 3,17576 | 3,31364 | 2,674493 | 3,421 | 2,485323 | 2,924236 |
| qac | 3,207208 | 3,037594 | 3,053765 | 3,535 | 3,688879 | 3,396855 | 2,892869 | 3,438975 | 3,492865 | 2,769772 | 3,397858 |
| dfr | 2,267994 | 3,027231 | 2,386467 | 2,851862 | 1,791759 | 2,741485 | 2,820783 | 2,455735 | 2,698673 | 2,412336 | 2,818697 |
| bla | 2,642622 | 2,871585 | 2,492379 | 2,945228 | 2,740195 | 3,06269 | 3,329701 | 2,4681 | 2,877512 | 2,074429 | 2,716349 |
| aacA | 2,396531 | 2,854457 | 2,543176 | 3,113738 | 2,758426 | 3,126541 | 2,900047 | 2,611539 | 2,938633 | 2,092481 | 2,644045 |
| aph | 2,594508 | 3,581573 | 3,172622 | 3,115513 | 2,827314 | 3,250374 | 3,490124 | 3,462292 | 3,554634 | 3,179303 | 3,386929 |
| ere | 2,884521 | 3,033028 | 2,984166 | 3,147165 | 2,821974 | 3,111959 | 3,35742 | 2,942331 | 2,919121 | 2,598607 | 2,921547 |
| arr | 3,73767 | 3,677439 | 3,688879 | 3,688879 | 3,688879 | 3,73767 | 3,73767 | 3,73767 | 3,688879 | 3,688879 | 3,688879 |
| qnr | 2,85532 | 3,187798 | 3,480009 | 3,238875 | 2,910719 | 3,20173 | 3,40053 | 2,671732 | 2,948641 | 2,574519 | 2,888982 |
| nit | 3,73767 | 2,803663 | 2,702367 | 3,127856 | 2,386926 | 3,068285 | 3,125663 | 2,672423 | 3,386929 | 2,140654 | 2,628646 |

Osprhone Osprhone Cyprinida Pantodon Poeciliida Poeciliida Callichthy Cobitidae Osprhone Cichlidae Callichthy

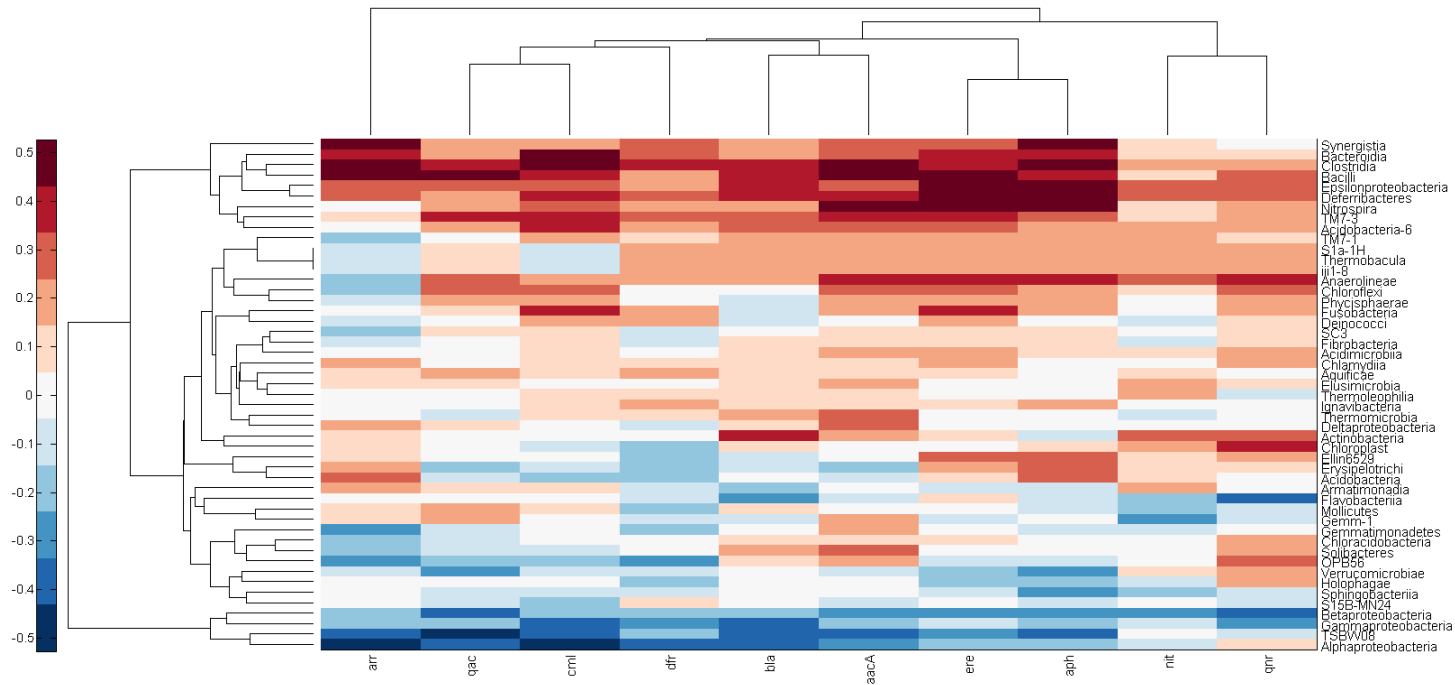
Vizualizace korelací

- Spearmanův korelační koeficient
 - negativní výsledek Shapiro-Wilks testu na normalitu dat v jednotlivých skupinách

| | cml | qac | dfr | bla | aacA | aph | ere | arr | qnr | nit |
|------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| Acidobact | -0,14824 | -0,06959 | -0,23788 | -0,00647 | -0,0815 | 0,256372 | 6,65E-02 | 0,241812 | -0,04012 | 0,115134 |
| Acidobact | 0,411574 | 0,203215 | 0,147806 | 0,245252 | 0,282081 | 0,175468 | 0,299474 | 0,025483 | 0,177784 | 0,210584 |
| Holophag | 0,026324 | 0,015794 | -0,15276 | -0,00526 | -0,04738 | -0,1632 | -0,15267 | -0,0372 | 0,152756 | -0,07898 |
| Chloracid | -0,02473 | -0,07909 | 0,014783 | 0,069774 | 0,139547 | -0,04697 | 0,051034 | -0,22594 | 0,219886 | 0,019488 |
| iii1-8 | -0,05791 | 0,142148 | 0,237035 | 0,184271 | 0,163211 | 0,19479 | 0,205319 | -0,09034 | 0,163291 | 0,205341 |
| Solibacter | -0,06621 | -0,10535 | -0,03965 | 0,162526 | 0,250883 | -0,00467 | -0,03436 | -0,16429 | 0,20986 | 0,038673 |
| Acidimicro | 0,057835 | 0,027128 | -0,04719 | 0,098665 | 0,181992 | 0,105904 | 0,184043 | -0,00471 | 0,185902 | 0,07257 |
| Actinobac | 0,010205 | -0,03078 | 0,044478 | 0,368164 | 0,195875 | -0,09336 | 0,078324 | 0,080568 | 0,243408 | 0,283799 |
| Thermole | 0,140753 | 0,027489 | 0,062 | 0,065748 | 0,113264 | 0,008124 | -0,03193 | 0,044528 | -0,05869 | 0,146996 |
| Aquificae | 0,068443 | 0,184266 | 0,205431 | 0,121092 | 0,142152 | -0,00526 | 0,142144 | 0,116911 | 0,026337 | 0,100038 |
| Armatimo | 0,089503 | 0,078971 | -0,07901 | -0,20533 | -0,00526 | -0,13161 | -0,07897 | 0,21788 | 0,015802 | 0,236932 |
| S1a-1H | -0,05791 | 0,142148 | 0,237035 | 0,184271 | 0,163211 | 0,19479 | 0,205319 | -0,09034 | 0,163291 | 0,205341 |
| Bacteroid | 0,477661 | 0,233369 | 0,316207 | 0,233429 | 0,284729 | 0,378461 | 0,366844 | 0,363448 | 0,102759 | 0,07177 |
| Flavobact | 0,022689 | -0,0108 | -0,05235 | -0,30961 | -0,09532 | -0,05753 | 0,121255 | 0,016436 | -0,37786 | -0,17968 |
| Sphingob | -0,11643 | -0,0527 | -0,08103 | 0,045215 | -0,0273 | -0,25564 | -0,11778 | 0,005534 | -0,11893 | -0,15091 |
| Chloropla | -0,06291 | -0,00596 | -0,17376 | 0,075046 | 0,006936 | 0,049381 | 0,03551 | 0,116563 | 0,375413 | 0,220433 |
| S15B-MN2 | -0,15718 | -0,08927 | 0,101757 | -0,04225 | -0,12474 | -0,08879 | -0,02407 | -0,00925 | -0,12855 | 0,026146 |
| Deferriba | 0,377526 | 0,235164 | 0,256675 | 0,357105 | 0,37126 | 0,526636 | 0,484027 | 0,32793 | 0,299602 | 0,320262 |
| Deinococc | 0,229306 | 0,035214 | 0,183534 | -0,08425 | 0,045861 | 0,03378 | 0,21701 | -0,10973 | 0,079067 | -0,05999 |
| Elusimicro | -0,02007 | 0,101264 | 0,020859 | 0,143592 | 0,158954 | 0,044047 | 0,018967 | 0,074129 | 0,052383 | 0,179813 |
| Fibrobact | 0,100033 | 0,026324 | -0,06848 | 0,110562 | 0,100033 | 0,100027 | 0,100027 | -0,10097 | 0,057942 | -0,05792 |

- nepřehledná tabulka, není vidět informace o vztazích

Vizualizace korelací



○ software: Matlab, R

A to je vše přátelé!

