

Vizualizace dat

Vizualizační formy II.

11. 12. 2024

Vizualizace *multivariate* dat

- *více než tři proměnné*
 - *dimenzionalita jako vizualizační problém*
 - vizualizace vícerozměrných dat je náročná
 - většina metod je cenná, ale mají své nevýhody
-
- *radar chart*
 - *rovnoběžné souřadnice*
 - *chernoff faces*
 - *a mnohé další...*

KALISNI-LISTEK-SIRKA

KALISNI-LISTEK-DELKA

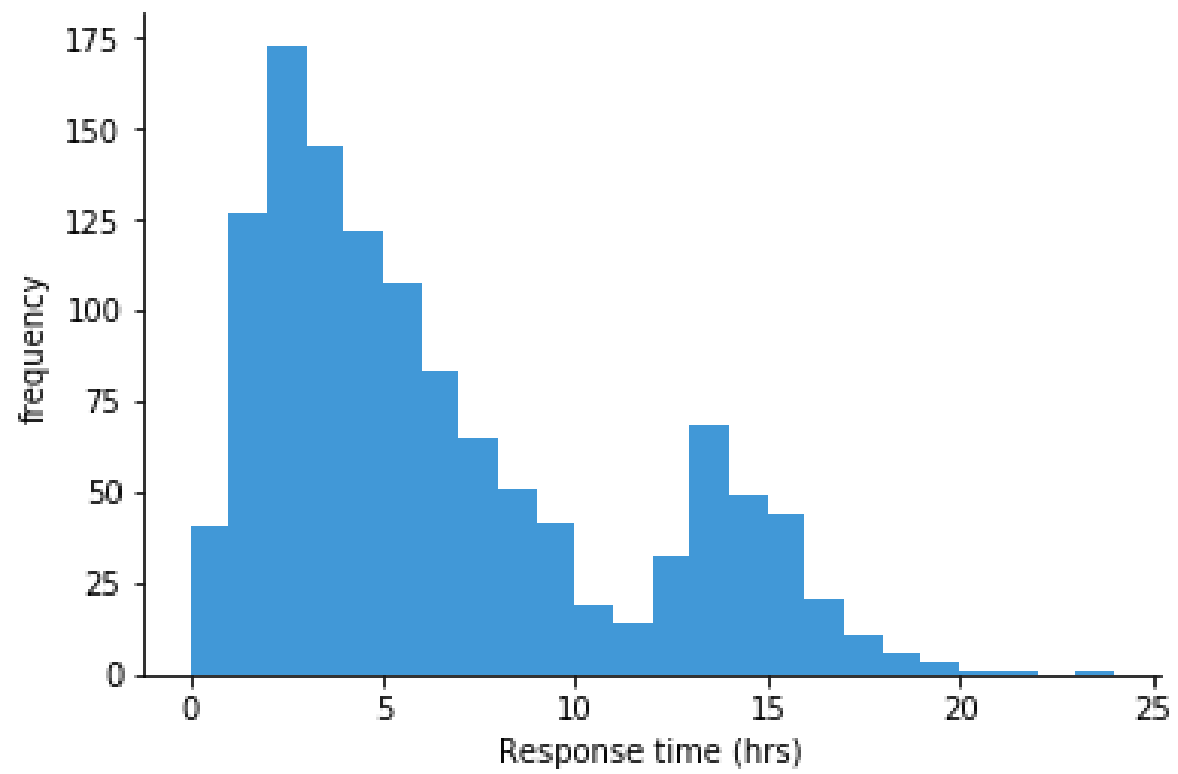
OKVETNI-LISTEK-SIRKA

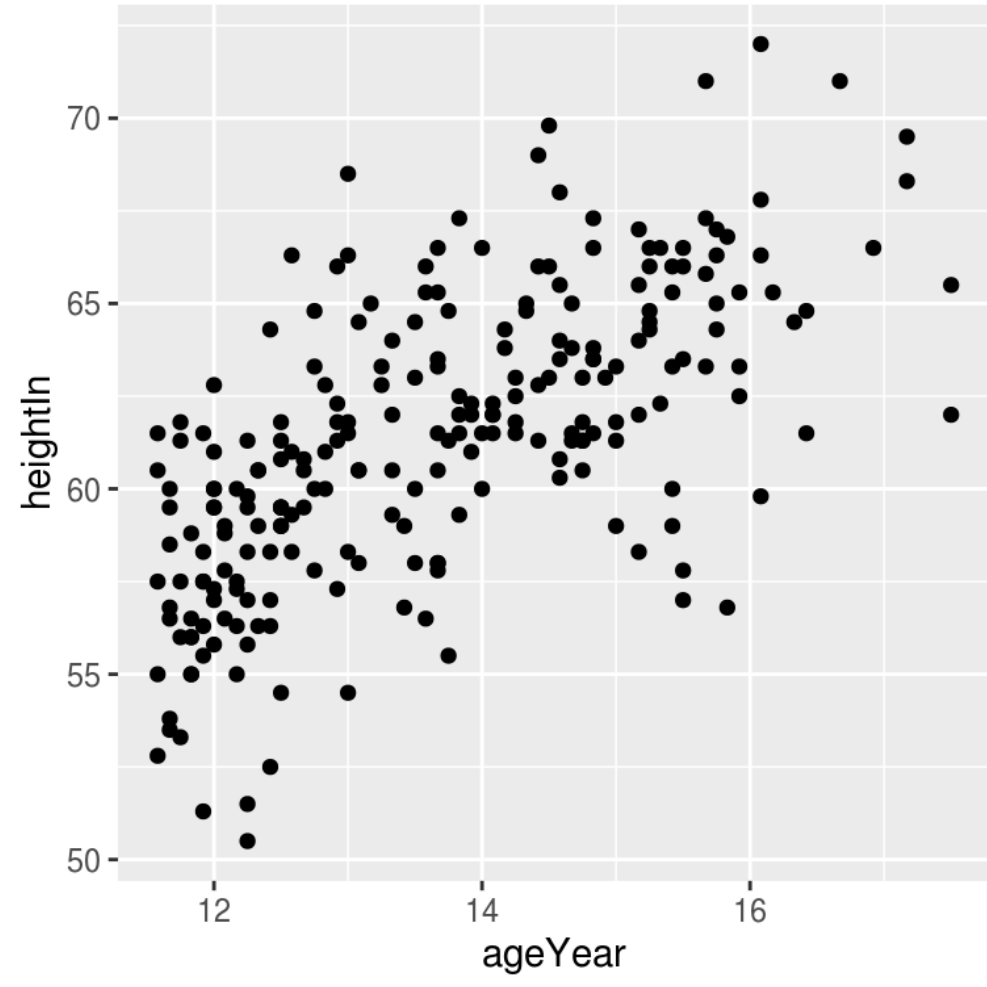
OKVETNI-LISTEK-DELKA

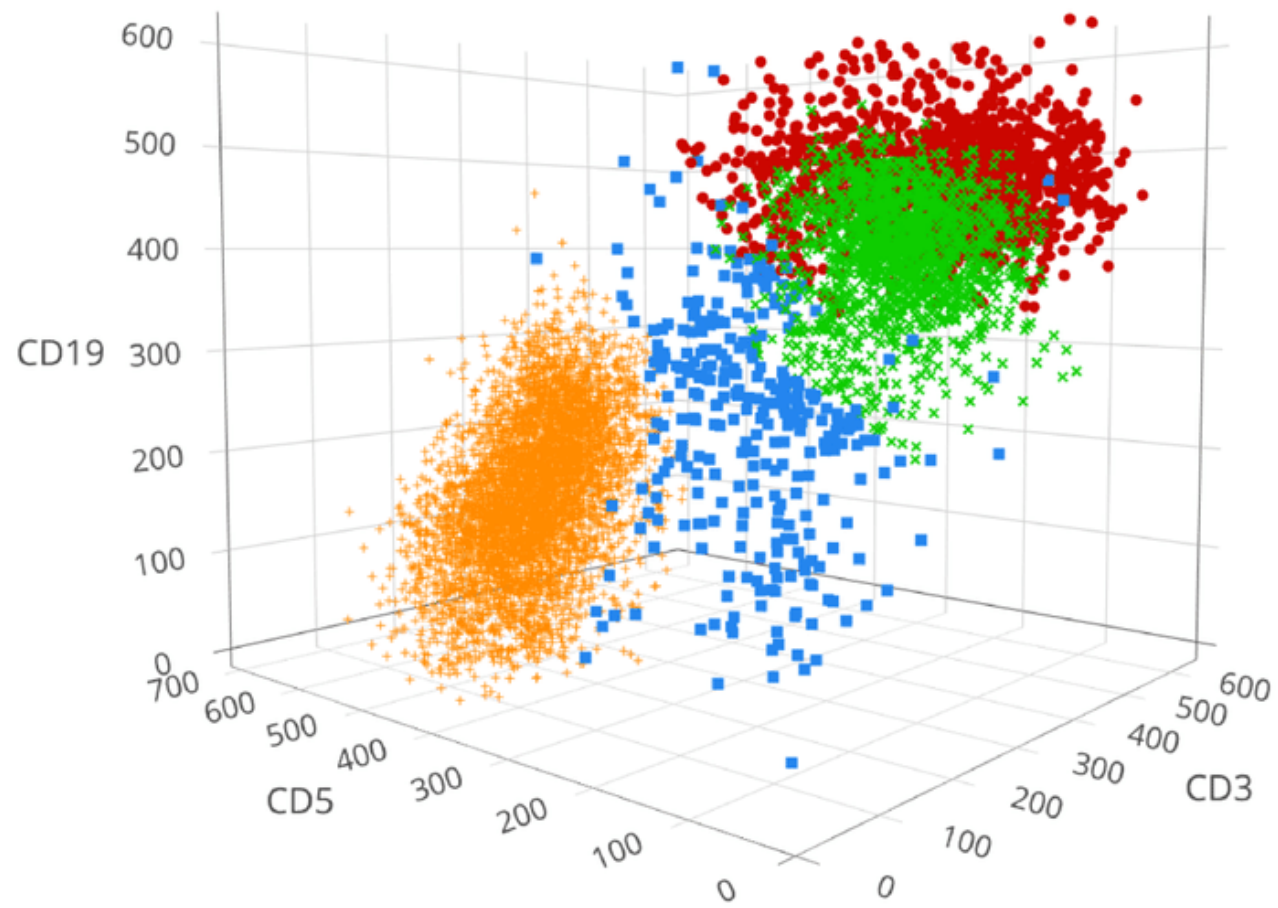
ODRUDA-ROSTLINY

50	33	14	2	Setosa
64	28	56	22	Virginica
65	28	46	15	Versicolor
67	31	56	24	Virginica
63	28	51	15	Virginica
46	34	14	3	Setosa
69	31	51	23	Virginica
62	22	45	15	Versicolor
59	32	48	18	Versicolor
46	36	10	2	Setosa
61	30	46	14	Versicolor
60	27	51	16	Versicolor
65	30	52	20	Virginica
56	25	39	11	Versicolor
65	30	55	18	Virginica
58	27	51	19	Virginica
68	32	59	23	Virginica
51	33	17	5	Setosa
57	28	45	13	Versicolor
62	34	54	23	Virginica
77	38	67	22	Virginica
63	33	47	16	Versicolor
67	33	57	25	Virginica
76	30	66	21	Virginica
49	25	45	17	Virginica
55	35	13	2	Setosa
67	30	52	23	Virginica
70	32	47	14	Versicolor
64	32	45	15	Versicolor
61	28	40	13	Versicolor
48	31	16	2	Setosa
59	30	51	18	Virginica
55	24	38	11	Versicolor
63	25	50	19	Virginica
64	32	53	23	Virginica
52	34	14	2	Setosa
49	36	14	1	Setosa
54	30	45	15	Versicolor
79	38	64	20	Virginica
44	32	13	2	Setosa
67	33	57	21	Virginica
50	35	16	6	Setosa
58	26	40	12	Versicolor
44	30	13	2	Setosa
77	28	67	20	Virginica
63	27	49	18	Virginica
47	32	16	2	Setosa
55	26	44	12	Versicolor

rosliny.csv



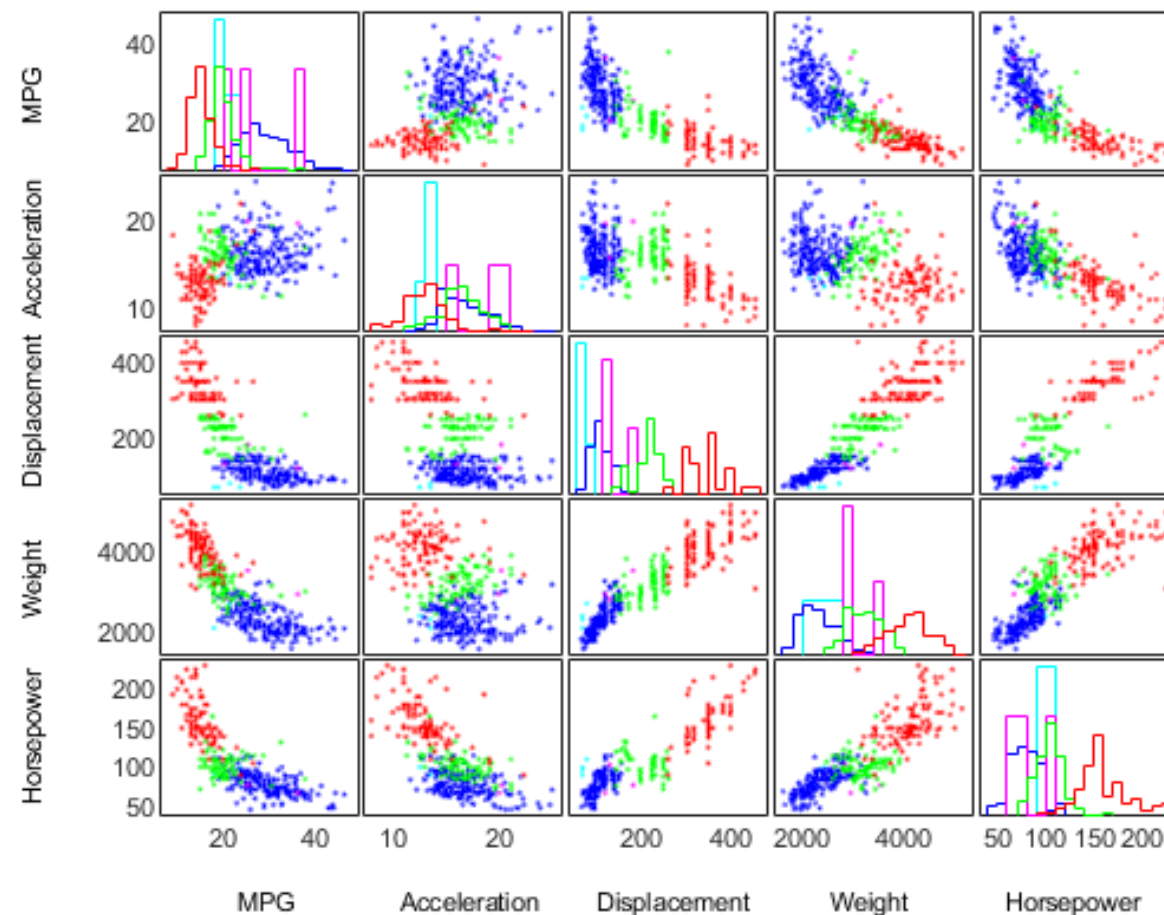




geometrické metody

vícerozměrný korelační diagram

- *multivariate scatter plot*
- matice korelačních diagramů
- čím více proměnných, tím více řádků
- *zahlcenost*
- *překrývání*

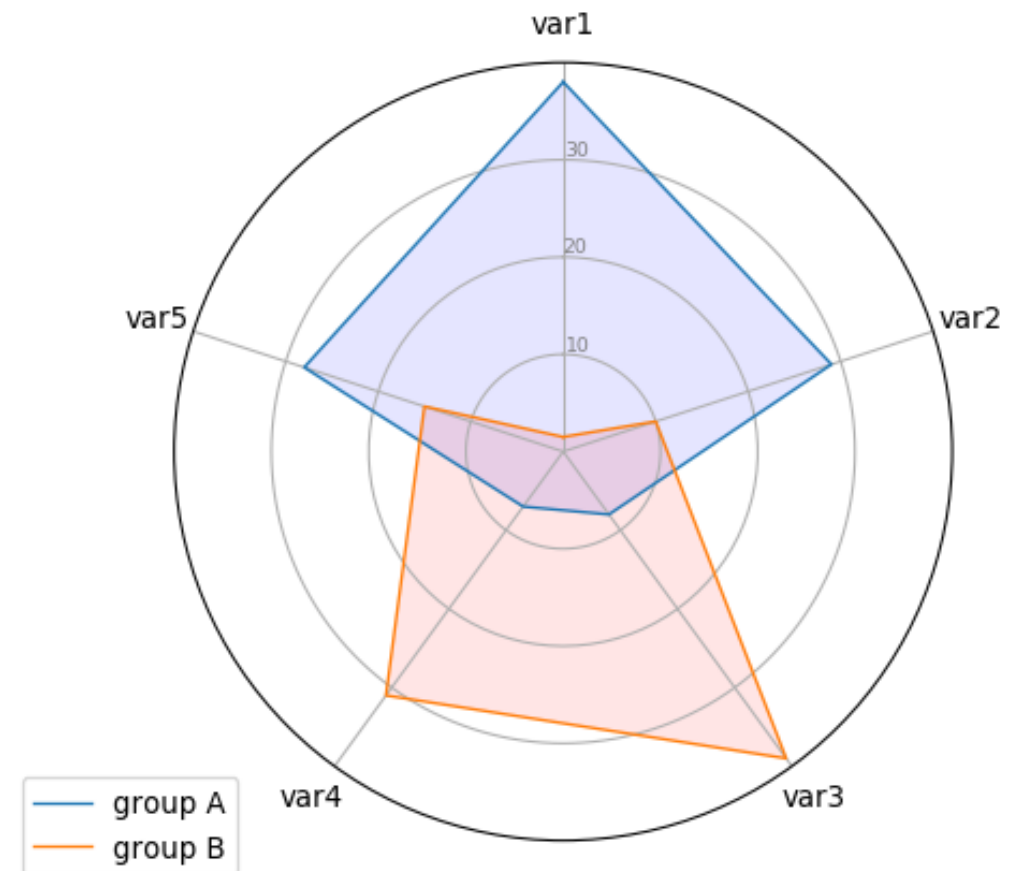


radar chart

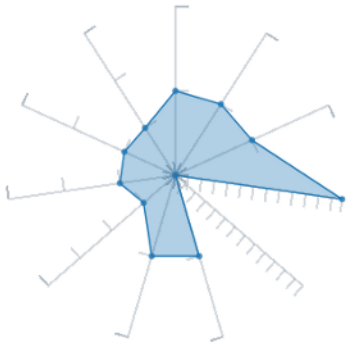
- *relativně těžko se čte*
- počet proměnných = počet paprsků
- na pořadí os záleží



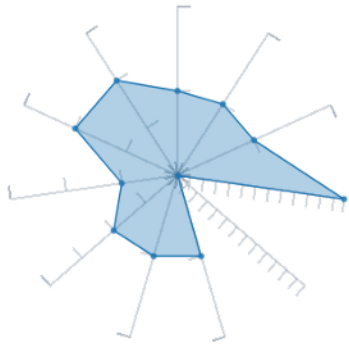
RAWGraphs



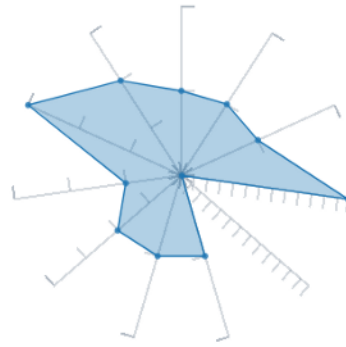
star plot



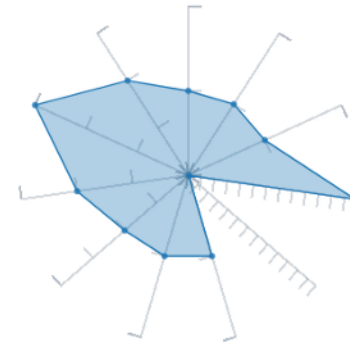
TRIE NODE



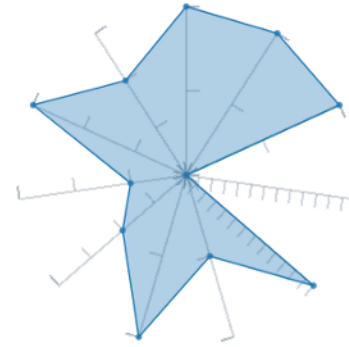
DATAPAGE (SORTED)



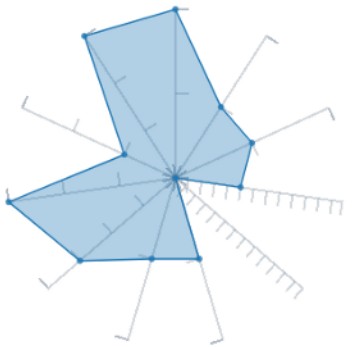
DATAPAGE



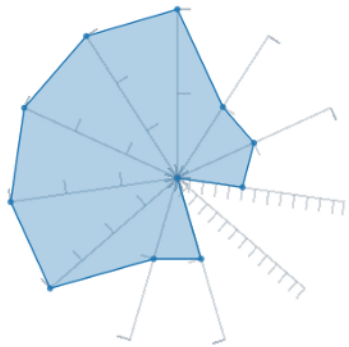
DATAPAGE (COMPR.)



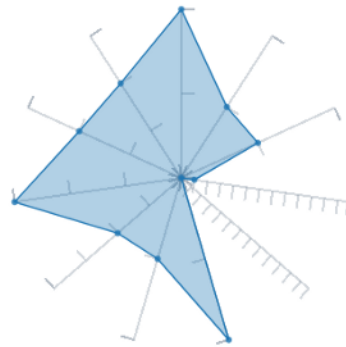
LINKED LIST



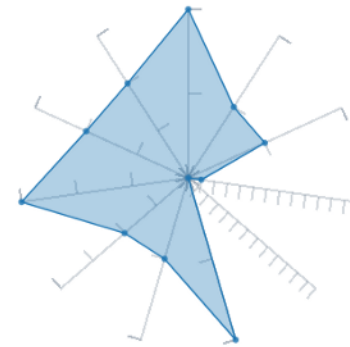
RANGE PARTITIONING



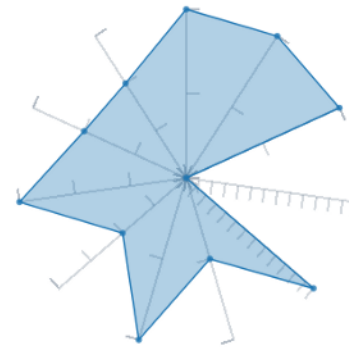
HASH PARTITIONING



B+TREE NODE



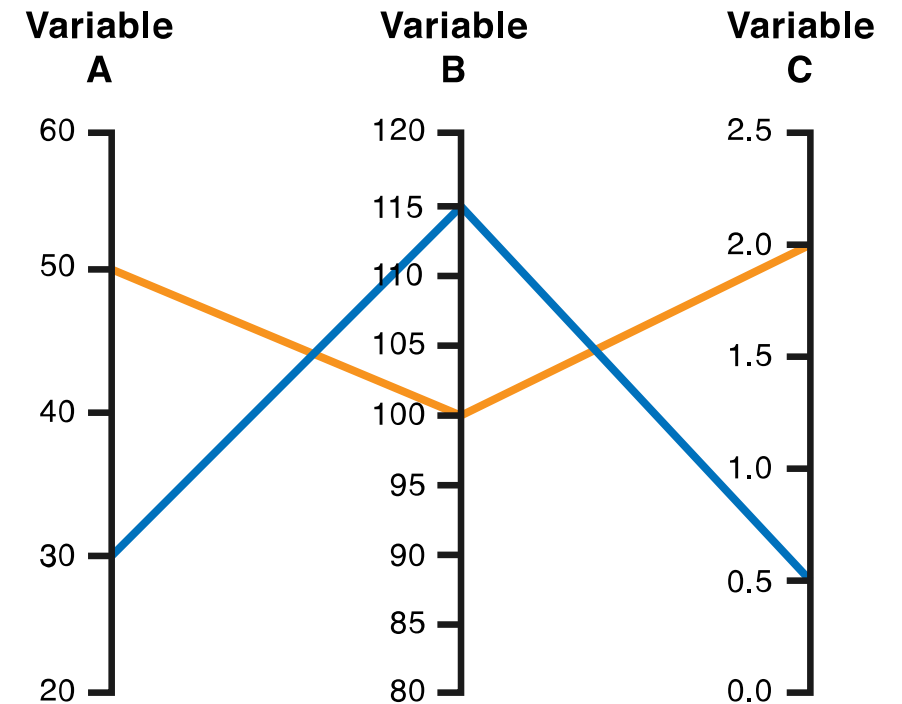
B-TREE NODE



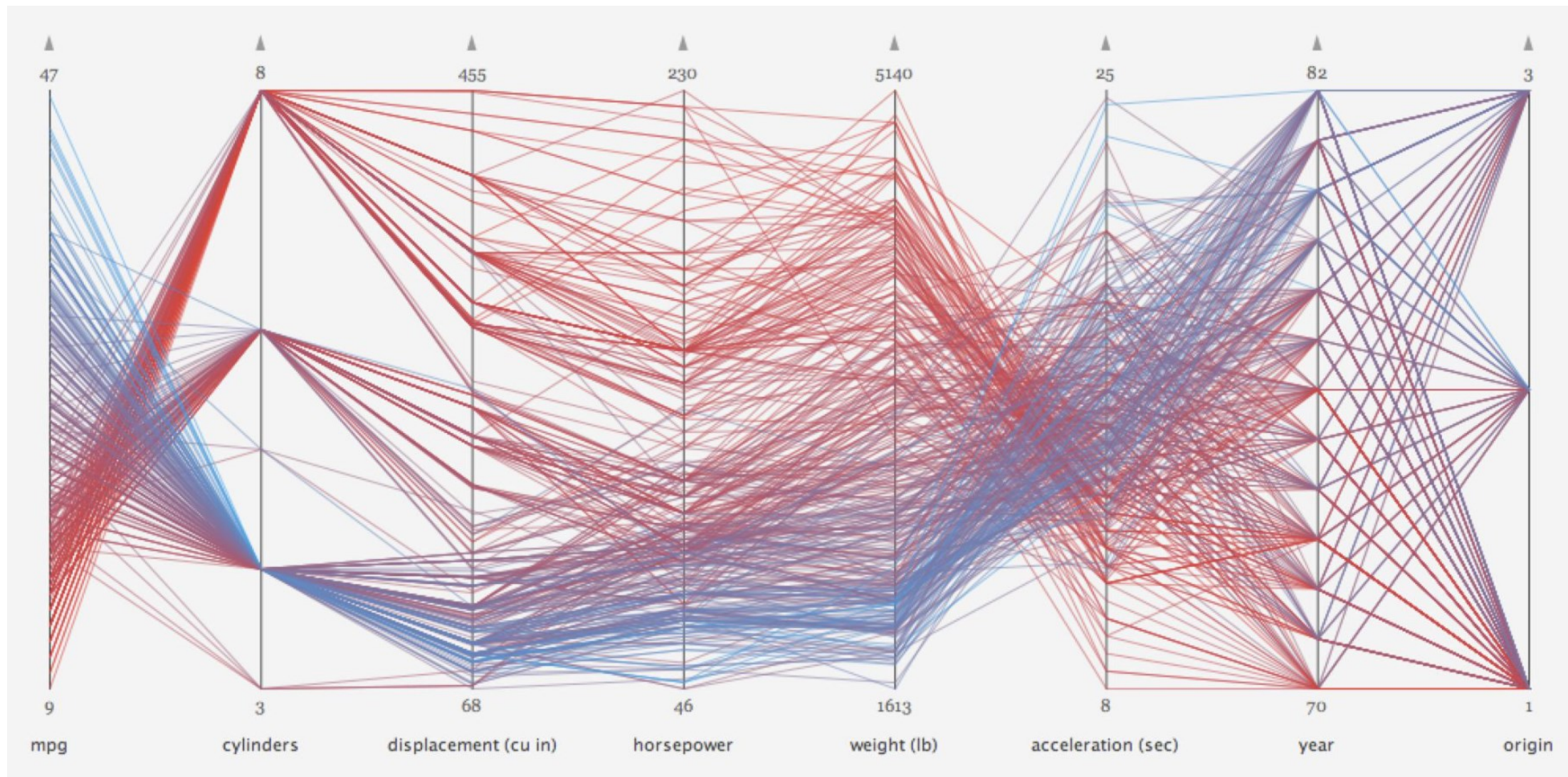
SKIP LIST

parallel coordinates

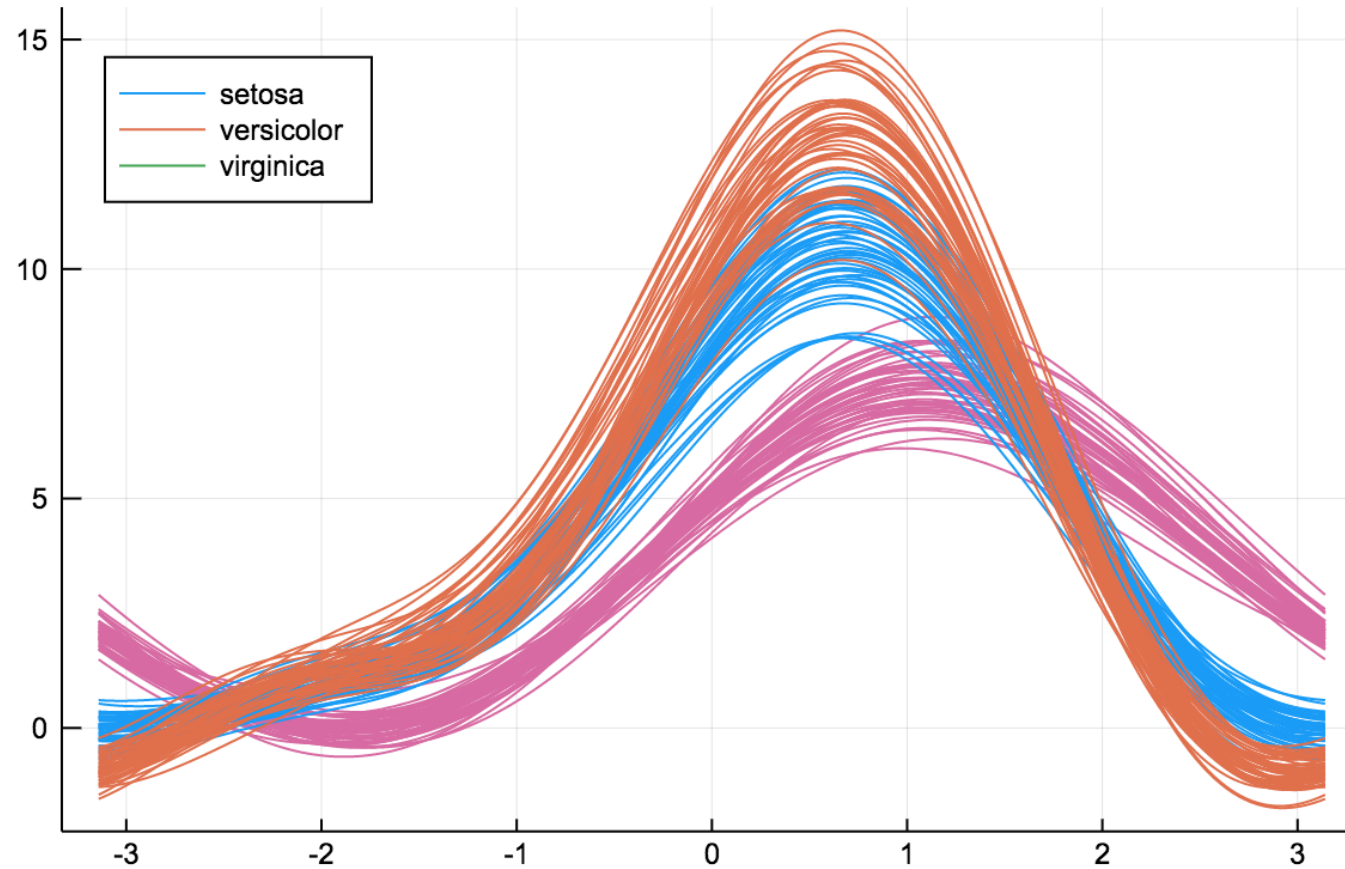
- *rovnoběžné souřadnice*
- vícerozměrná data
- jedna položka s více proměnnými
- dobré k odhalení vztahů mezi p
- **onboarding potřeba**
- na pořadí os záleží



Data			
	Variable A	Variable B	Variable C
Item 1	50	100	2.0
Item 2	30	115	0.5



andrews plot

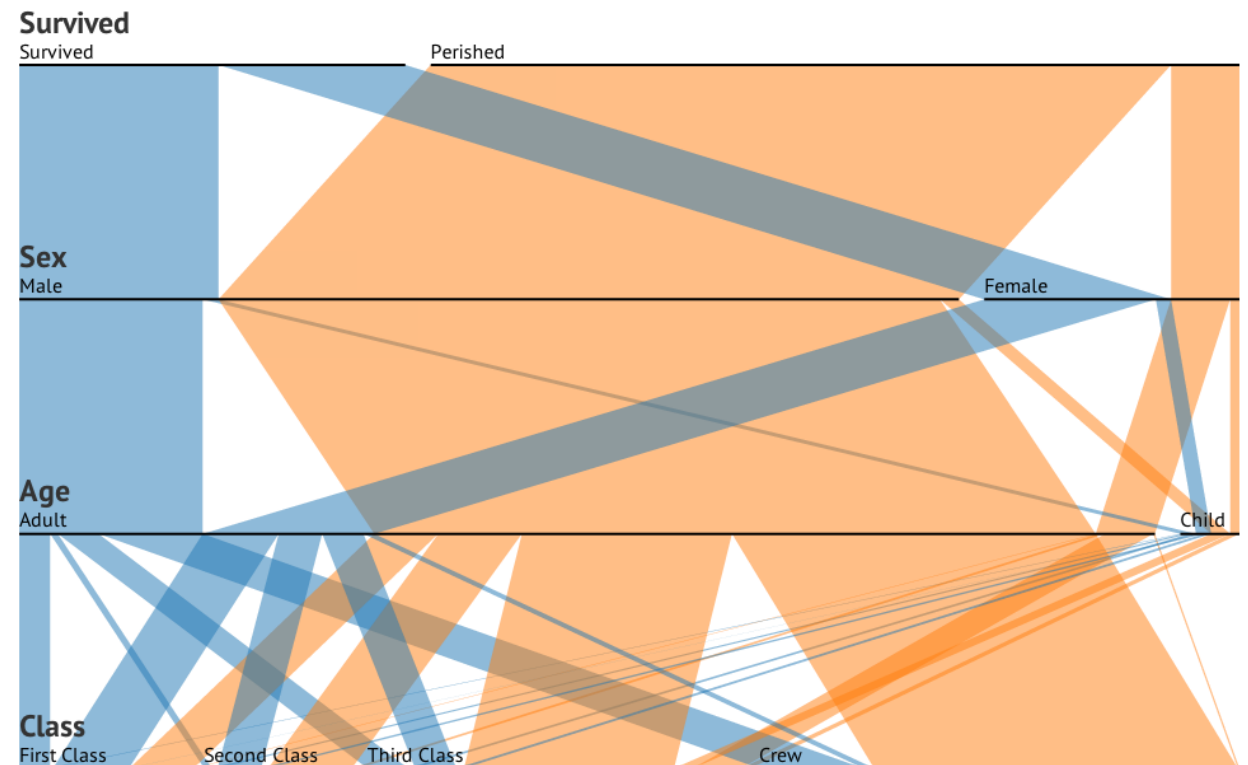
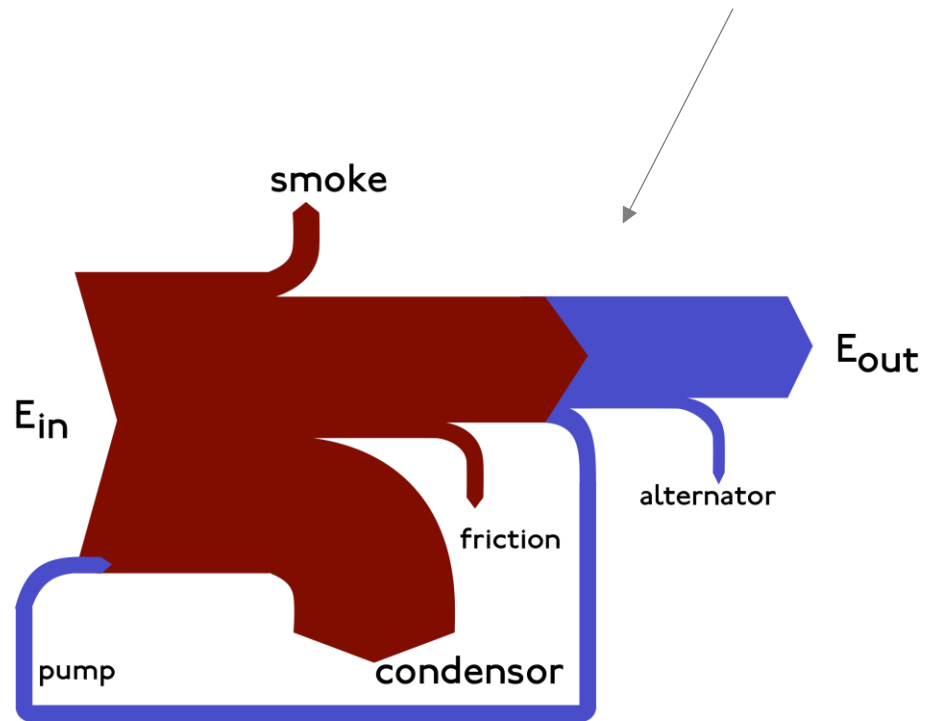


odhalení struktury

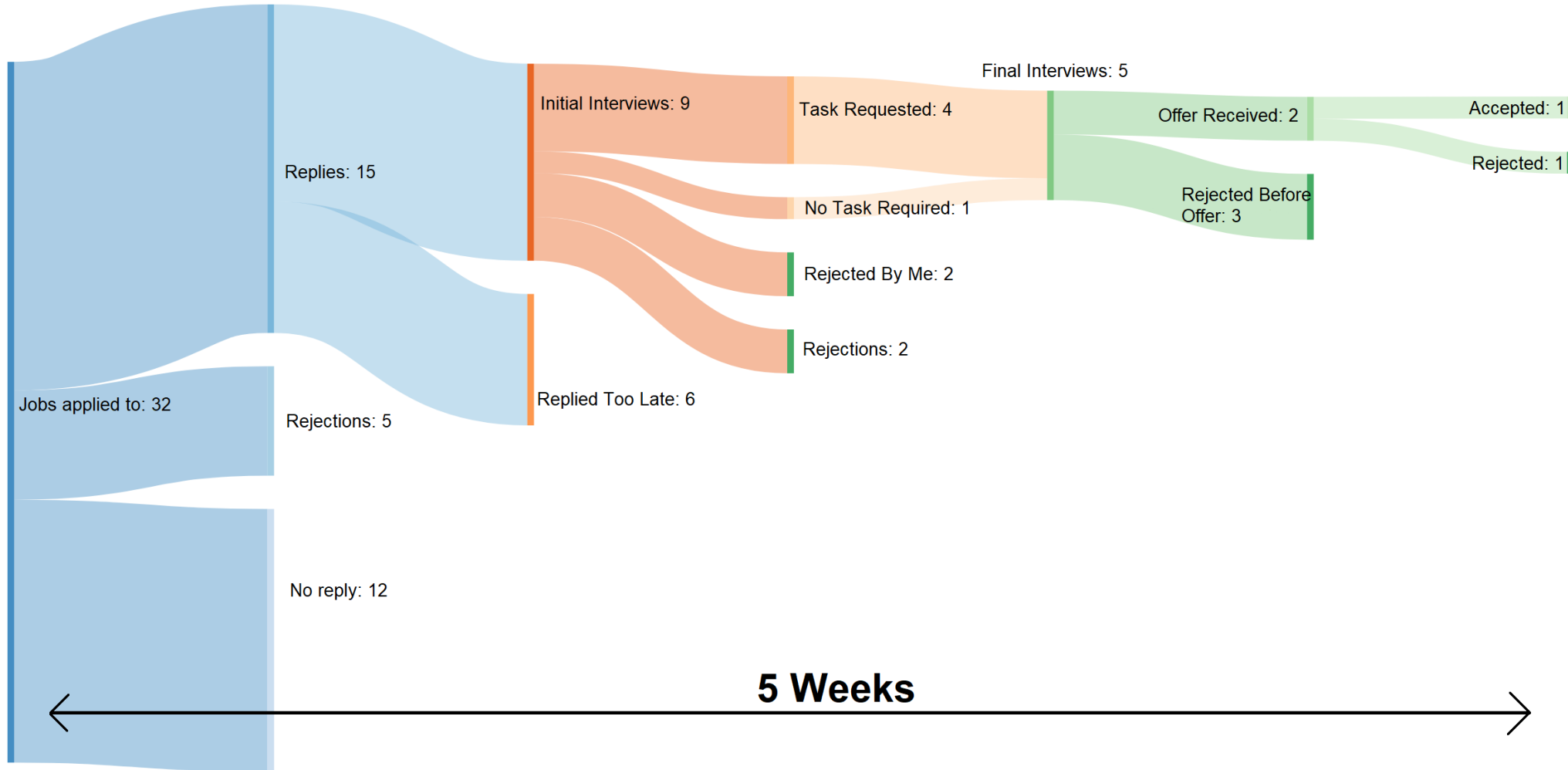
redukce dimenzionality: *mapuje všechny atributy z jednoho řádku/pozorování do funkce*

odbočka: paralell plots

- pro **kategorická** data, [Robert Kosara](#)
- jako *alluvial diagrams* (RAW)
- ještě větší odbočka: *sankey*



odbočka: sankey diagram



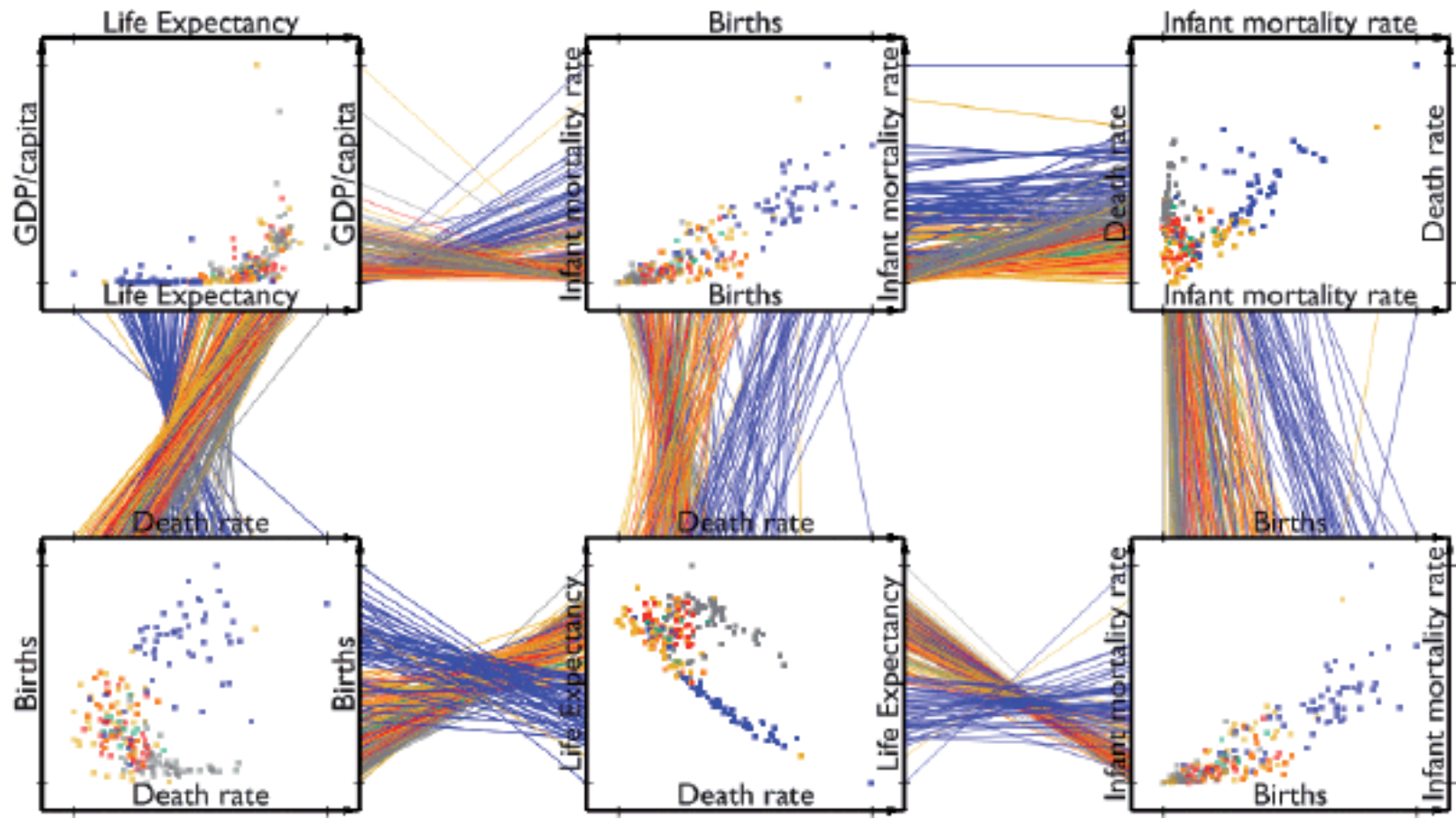
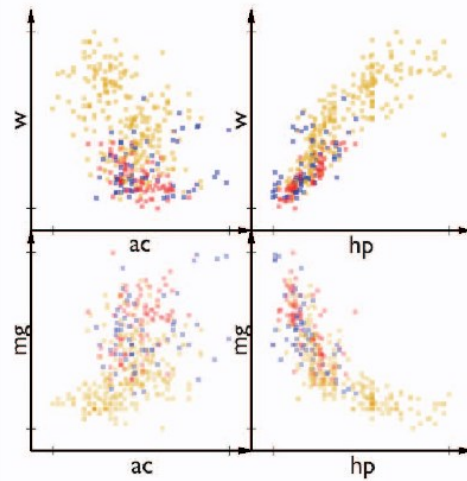
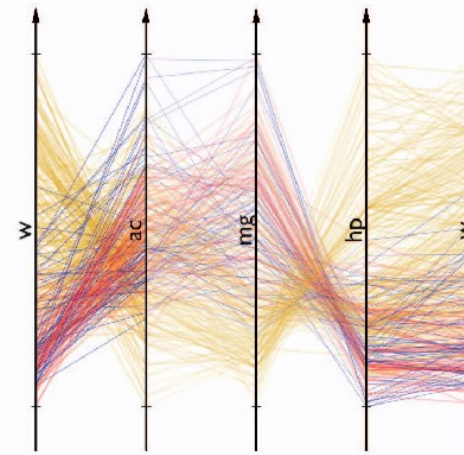


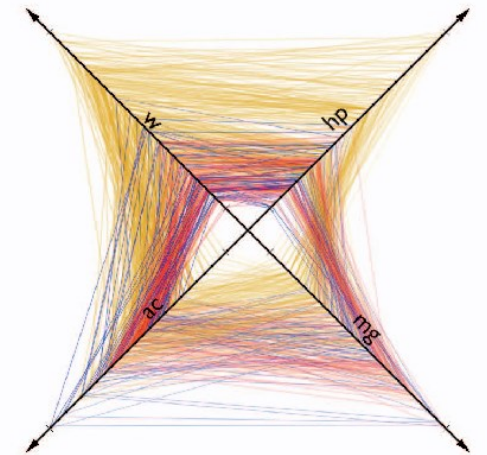
Fig. 1. Examples of methods for multivariate data visualization, showing the cars dataset [16], with ac = acceleration, mg = miles per gallon, w = weight, hp = horse power, yr = model year, or = origin, cy = cylinders, and dp = displacement. Colors denote the origin: yellow for USA; blue for Europe; and red for Japan.



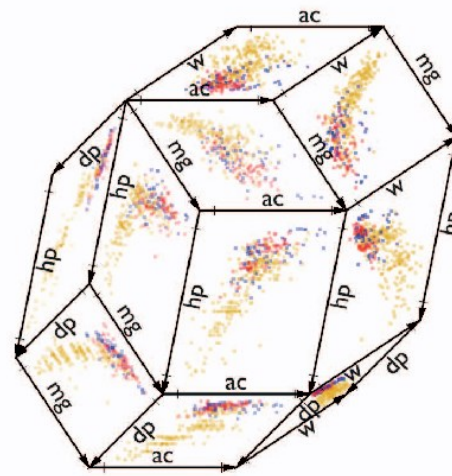
(a) scatterplots



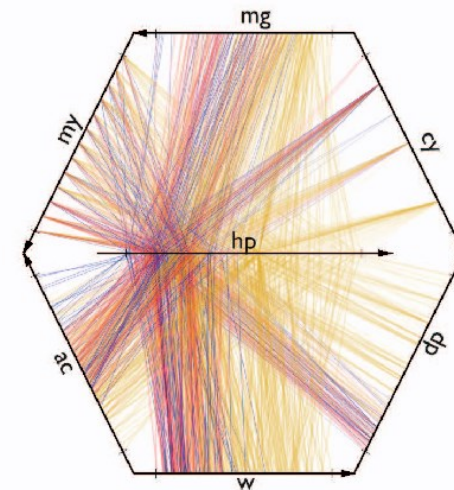
(b) Parallel Coordinates Plot



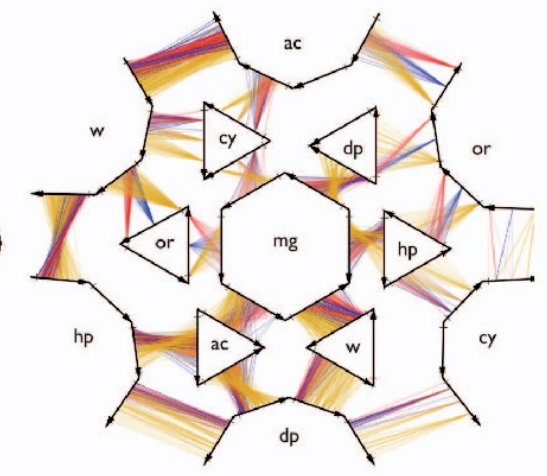
(c) radar chart



(d) Hyperbox

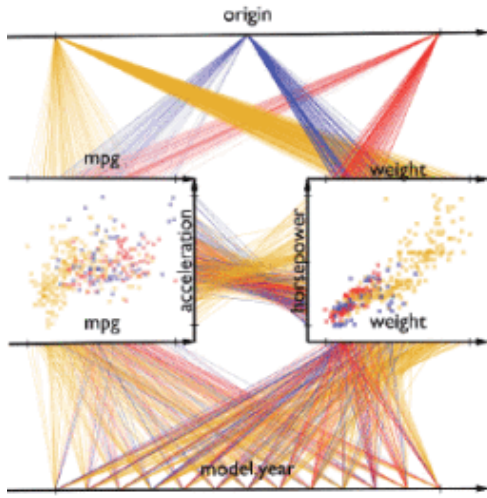
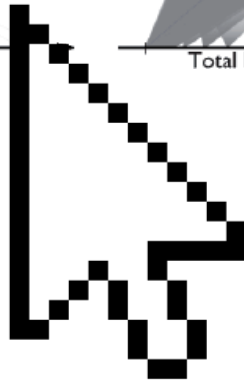
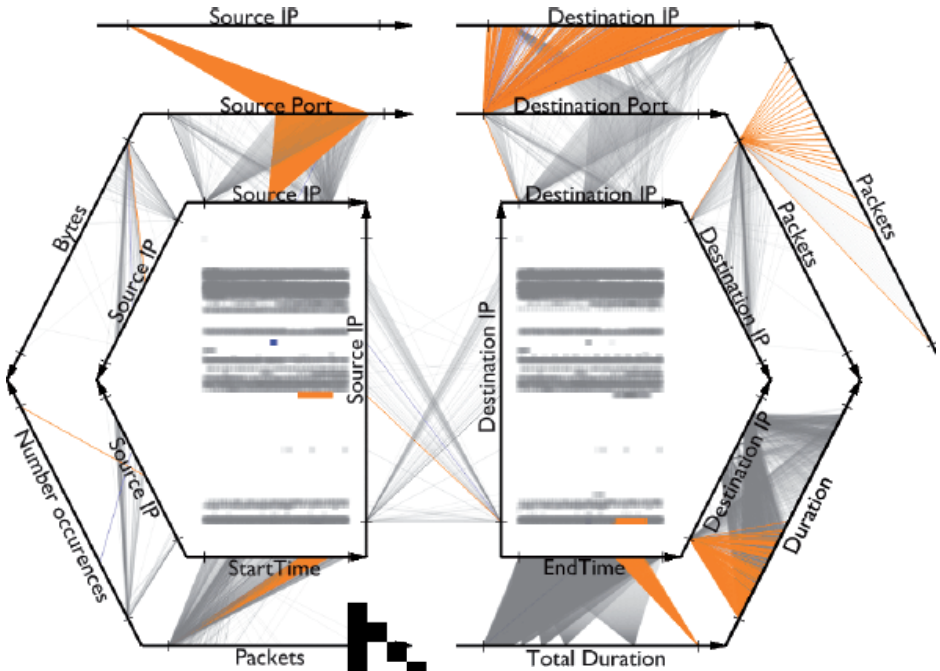


(e) Time Wheel

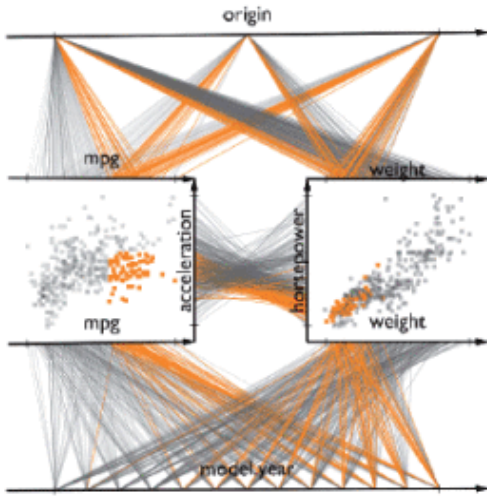


(f) Many-to-many PCP

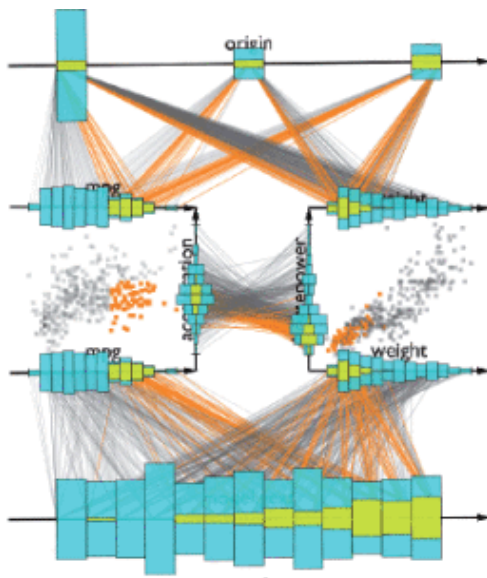
flina plots



a)

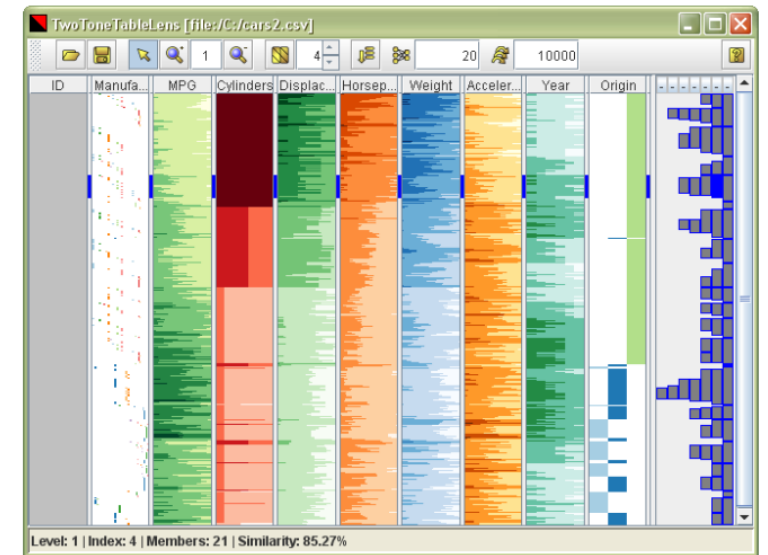
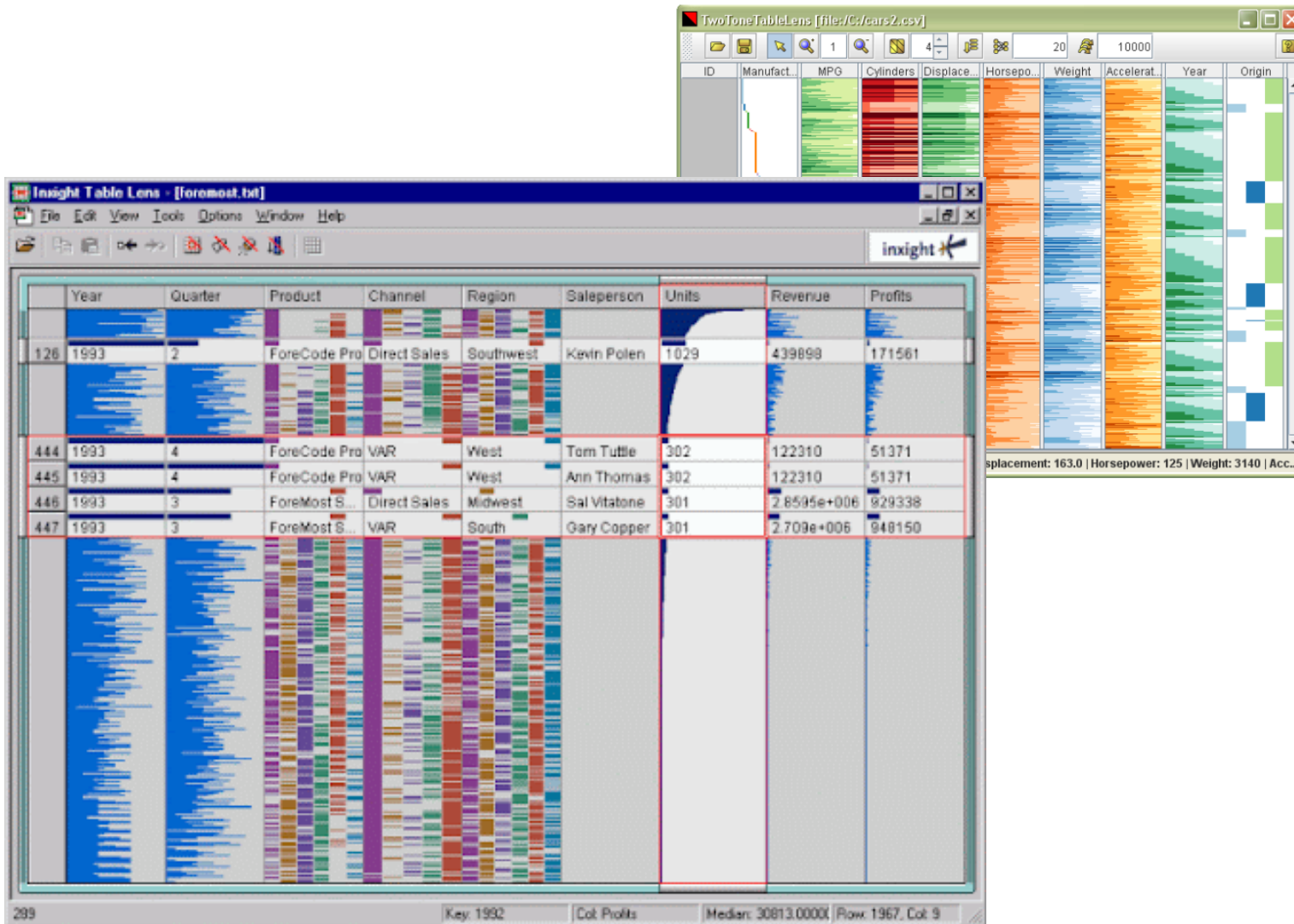


b)



c)

table lense

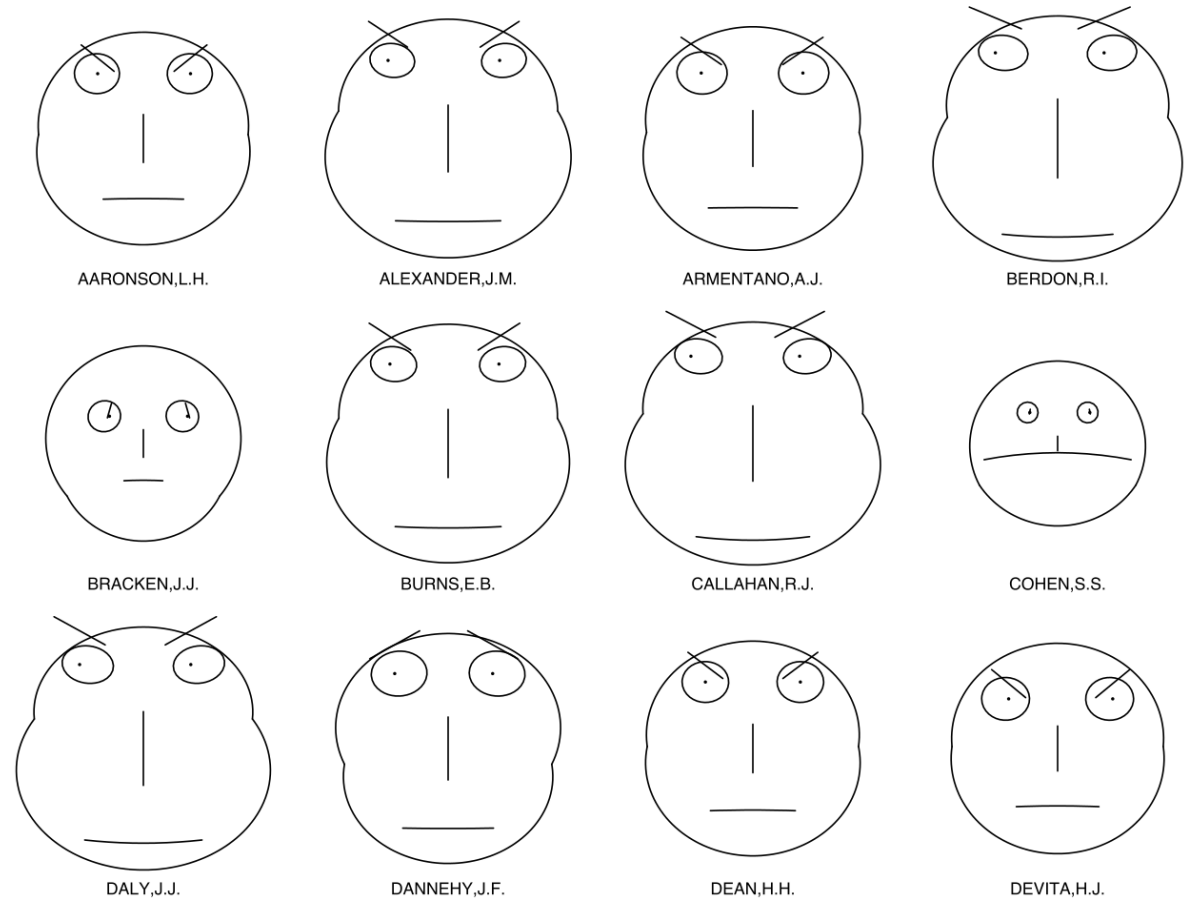


<https://inria.hal.science/hal-00656250v1>

metody kompozitních glyfů & ikon

Chernoff faces

- 1973, Herman Chernoff
- kompozitní glyfy
- mozek: *fusiform face area*
- vícerozměrná data
- velmi špatně se čtou
- v praxi se používají minimálně
- existují pokusy je vylepšit



1A. FACES FOR 87 FOSSIL SPECIMENS OF EXAMPLE 1

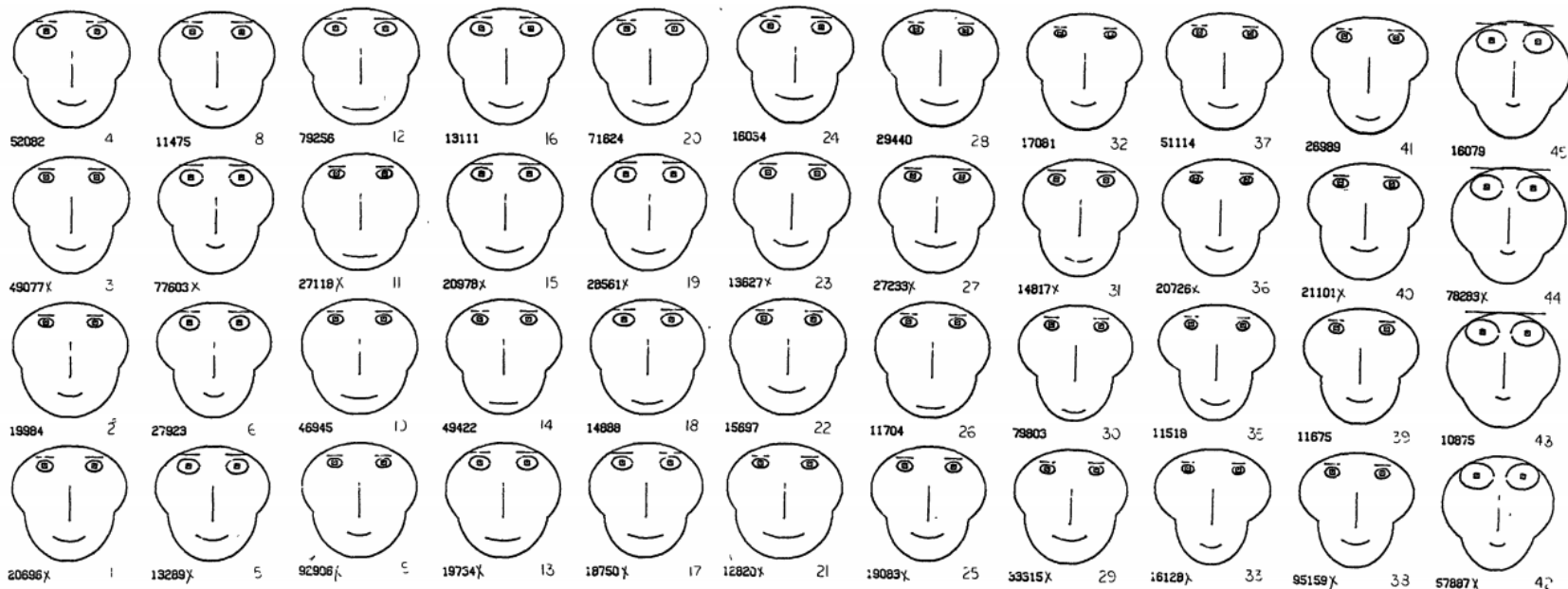
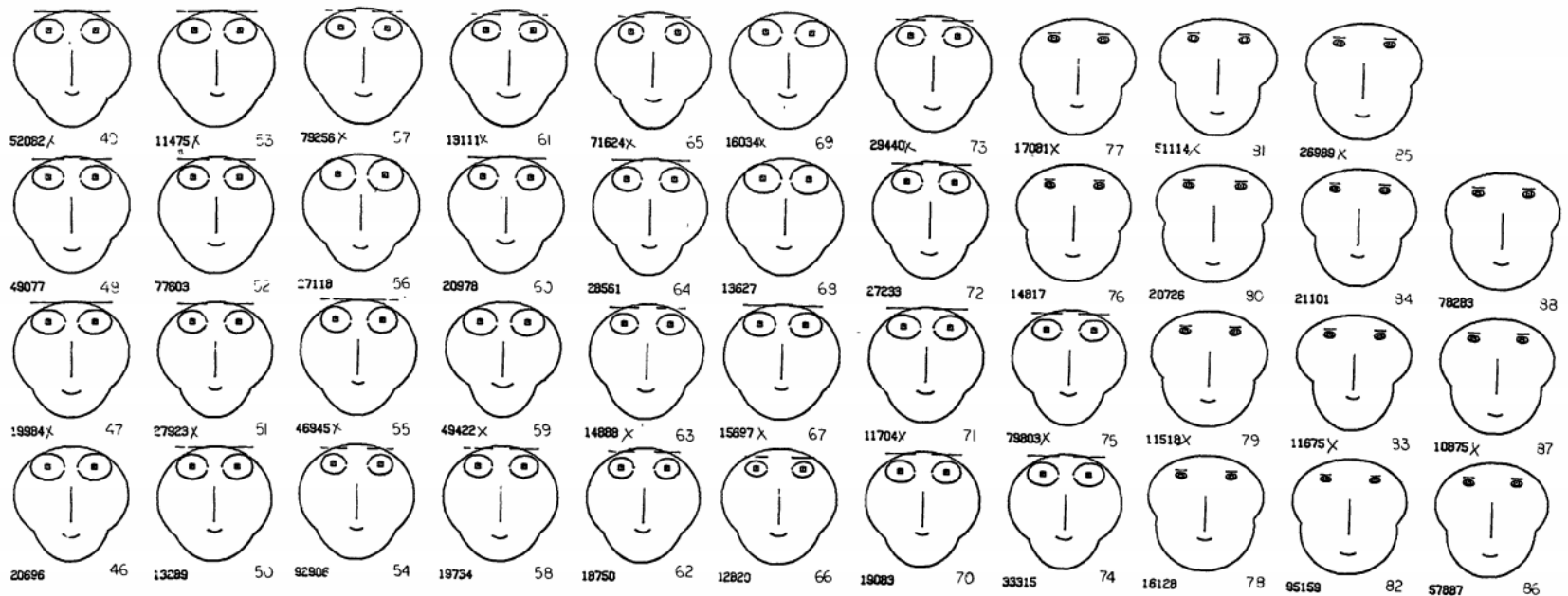


FIG. 1A



Chernoff faces

- *lidé velmi dobře čtou tváře*
- finanční situace 4 firem
- do které investujete?
- *empathic visualization algorithm* (EVA)

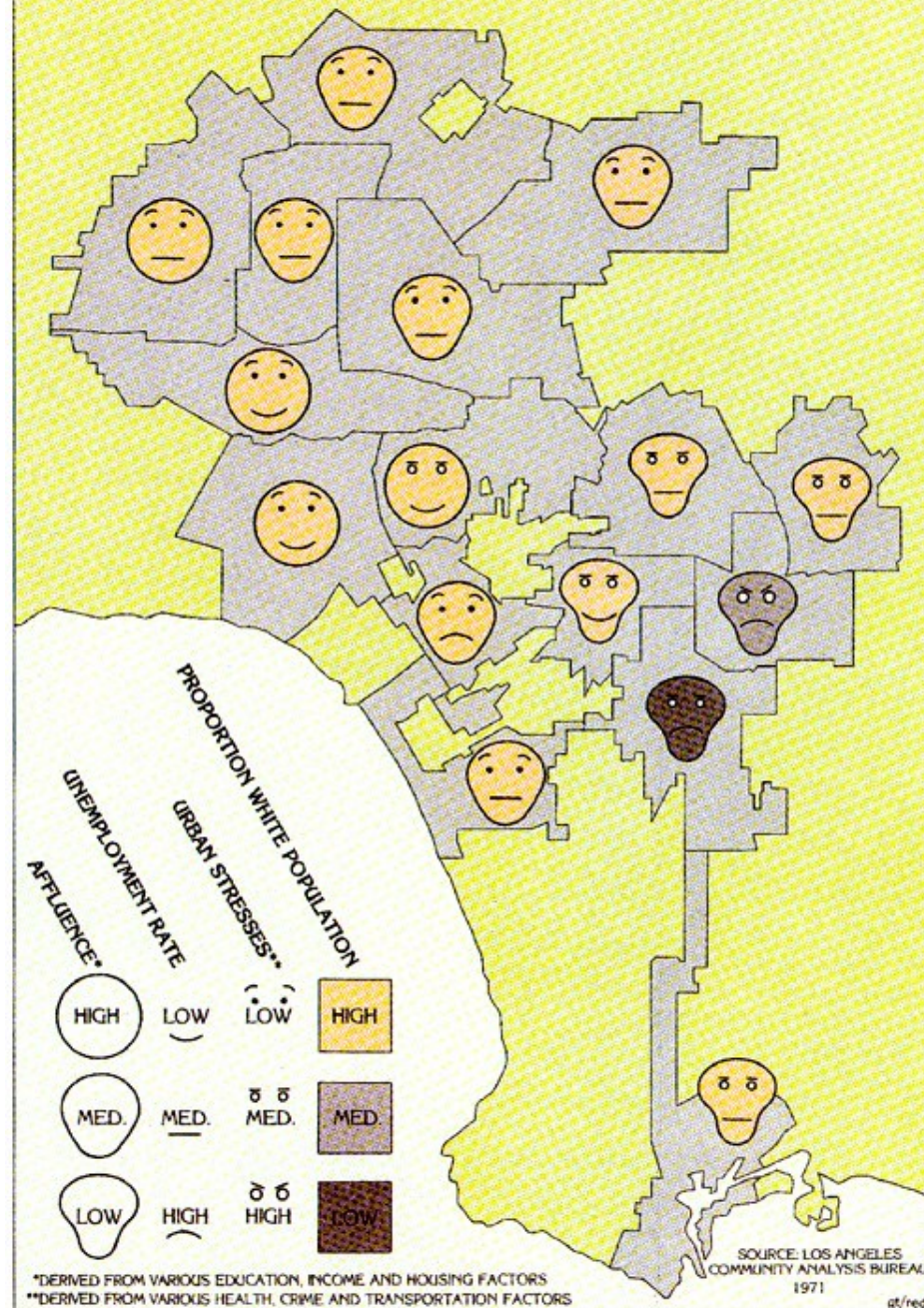


Chernoff faces

„Despite the best effort of academics to find a way to make the faces ‘work’, the usual response to Chernoff faces from the cartographic community is a combination of fascination and loathing.“ – *Martin Elmer*

- Předpozornostní povaha tváře?
- co je s tím za problémy?
- *naštvaný* může být *dobrý* v jistých ohledech
- musí to být jen obličej...?

Life in Los Angeles





star glyph

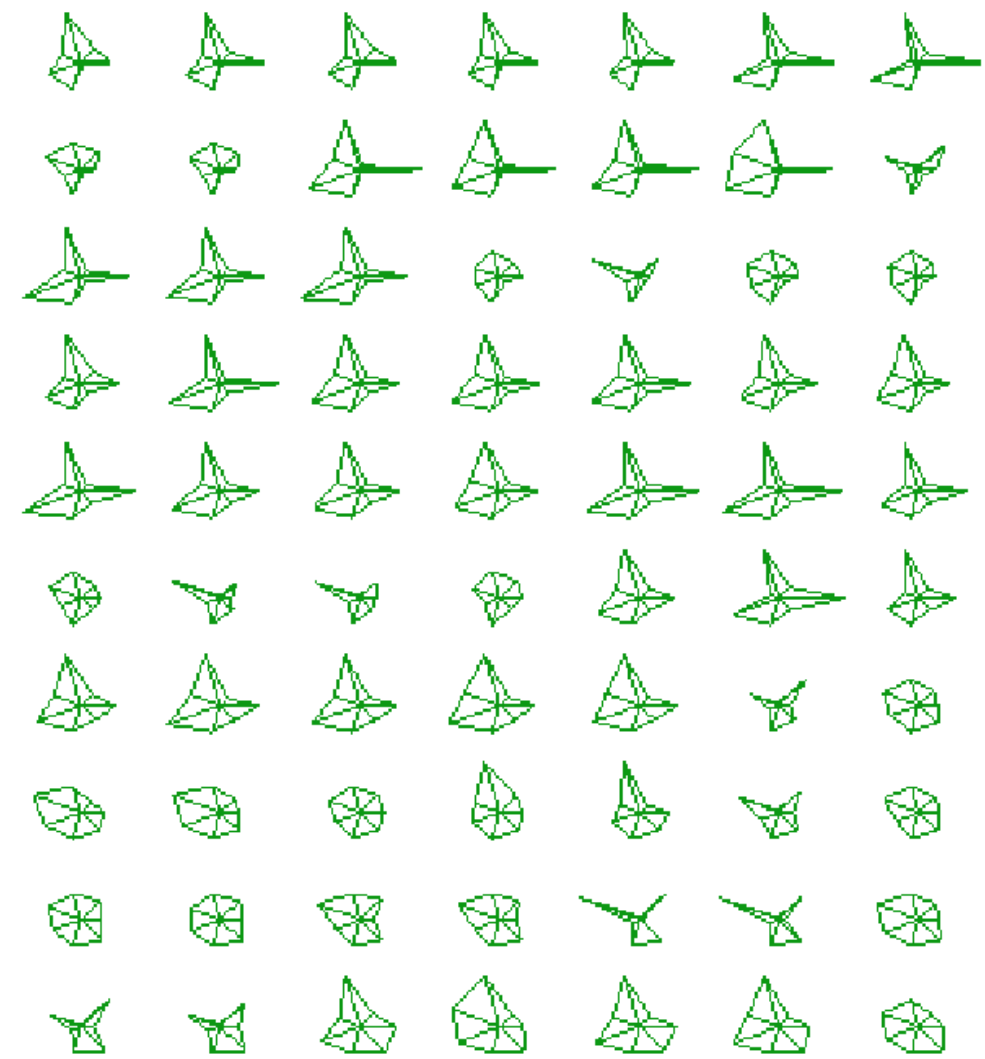
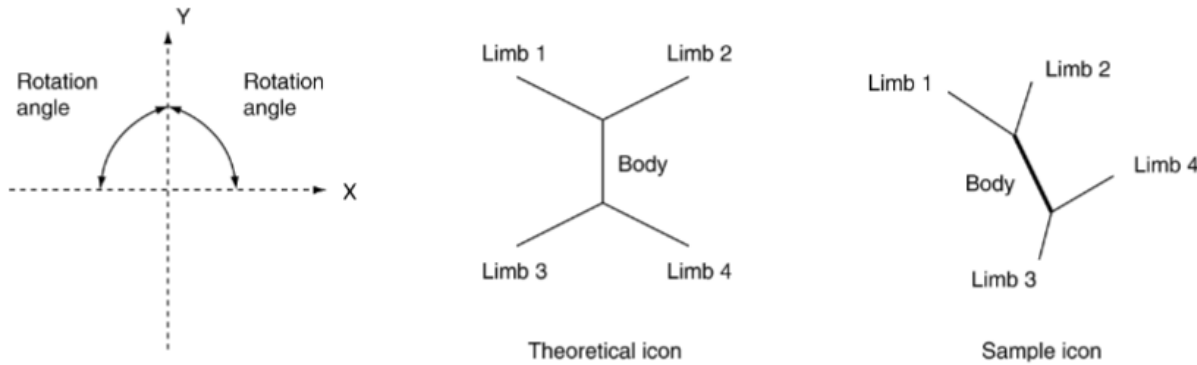
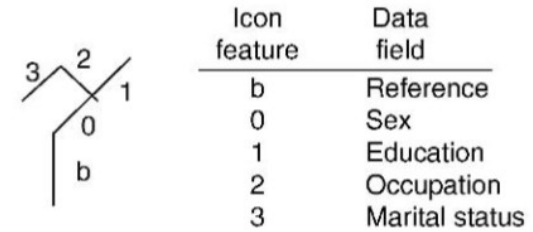


Fig. 3. Star glyphs for a subset of the cars dataset using a random dimension ordering and one based on shape analysis. More simple shapes can be seen in the second version than the first, which we believe can facilitate detection of groups of similar shapes as well as outliers.

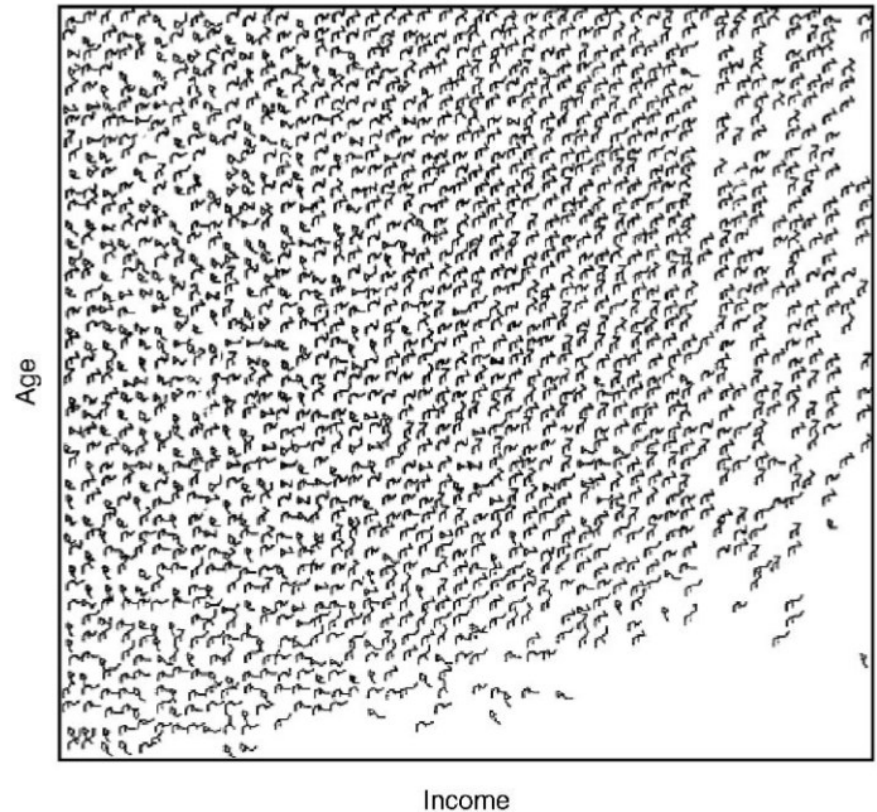
stick figures

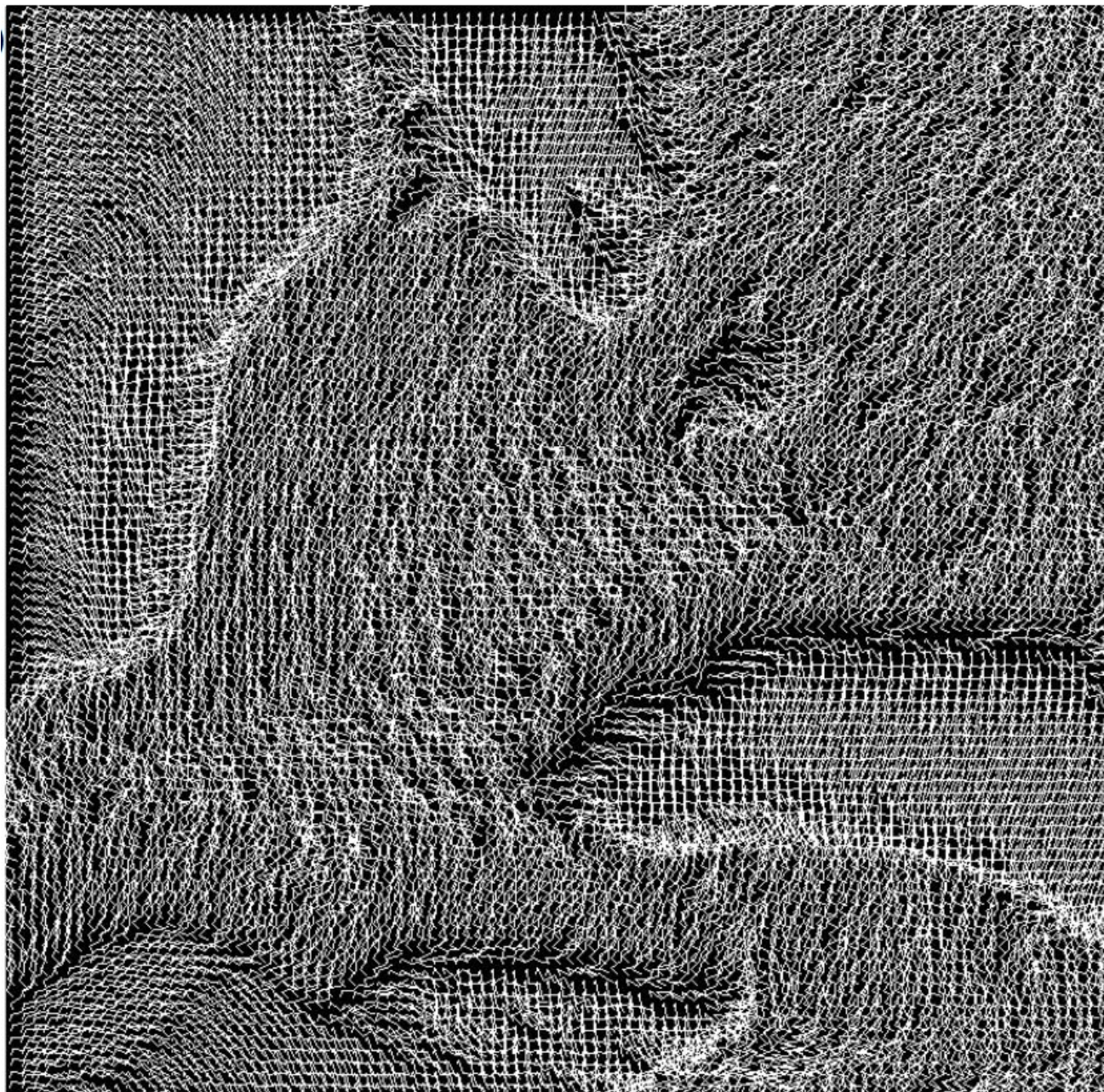


Iconic Displays. Figure 2. One member of the stick-figure icon family.



	Low job married	High job married	Low job single	High job single
Male low education				
Male high education				
Female low education				
Female high education				

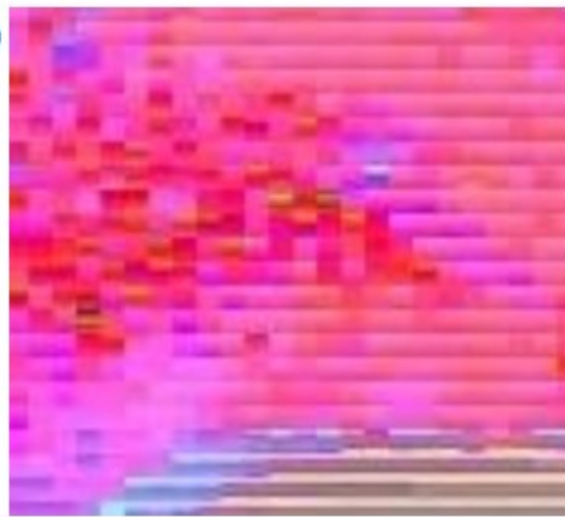




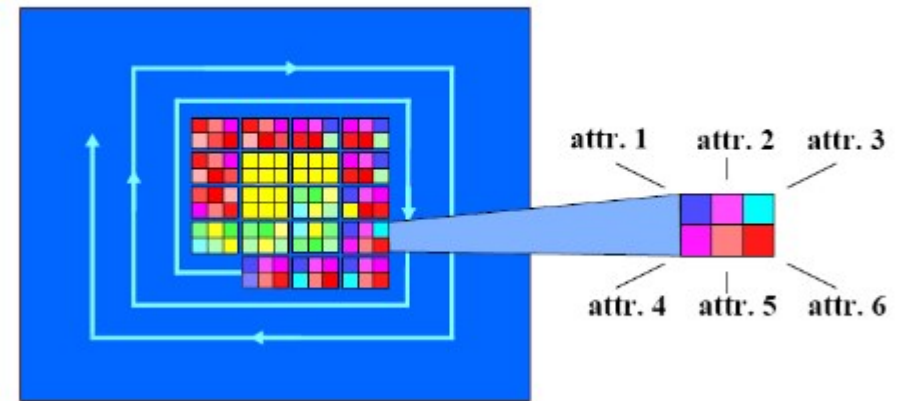
(a)



(b)

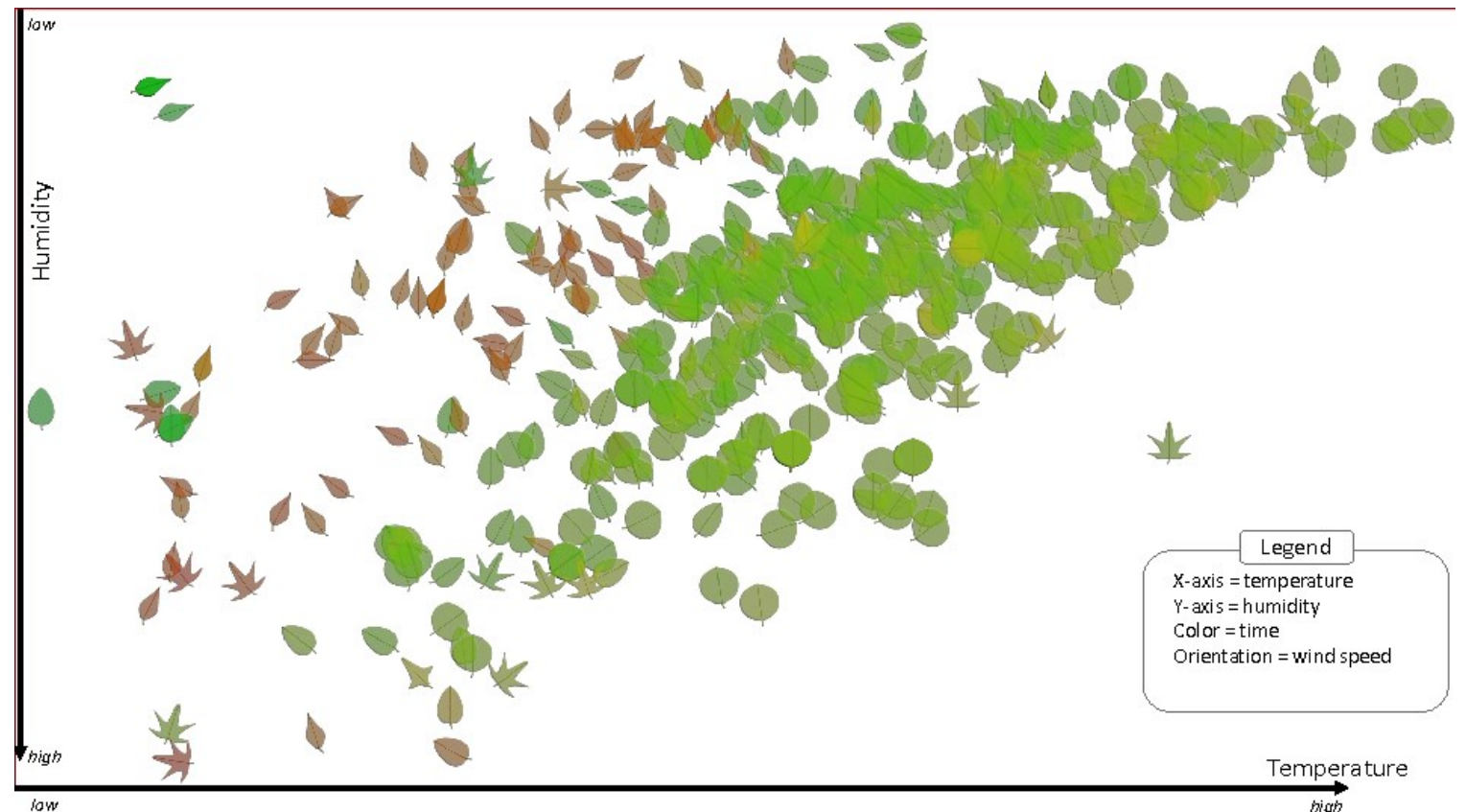


(c)

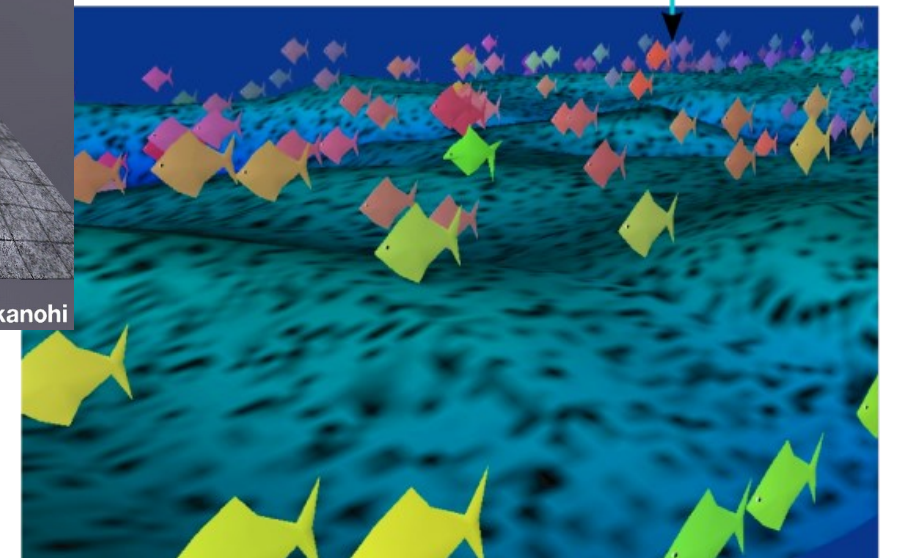
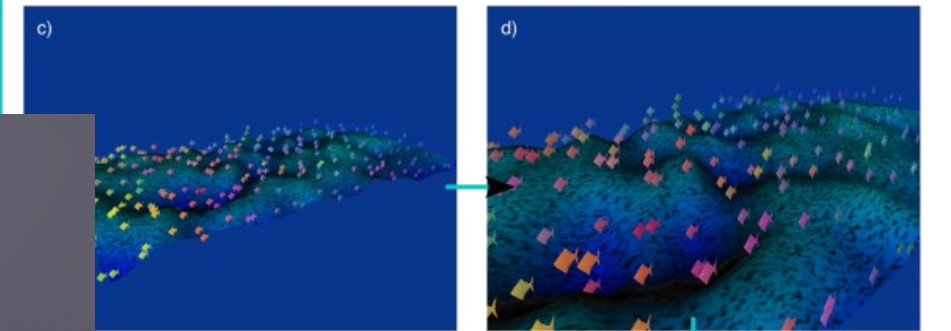
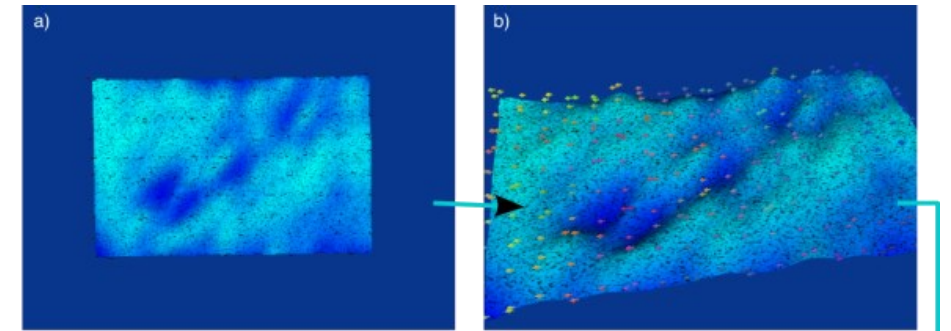
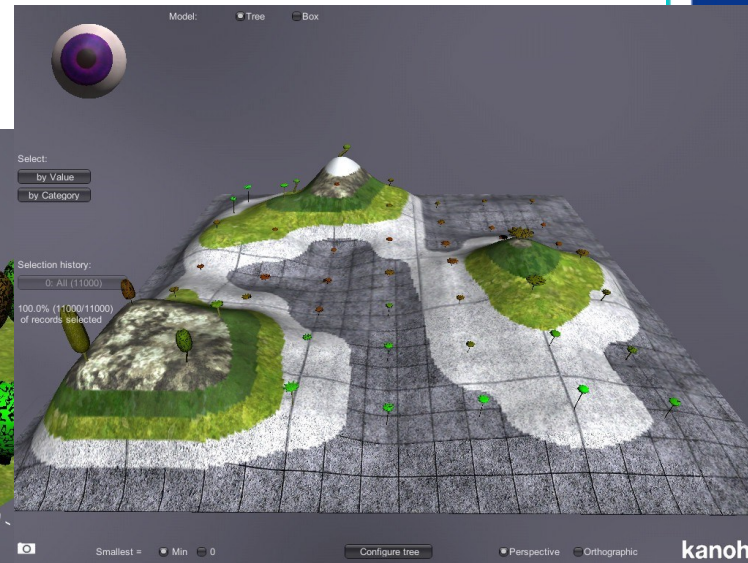
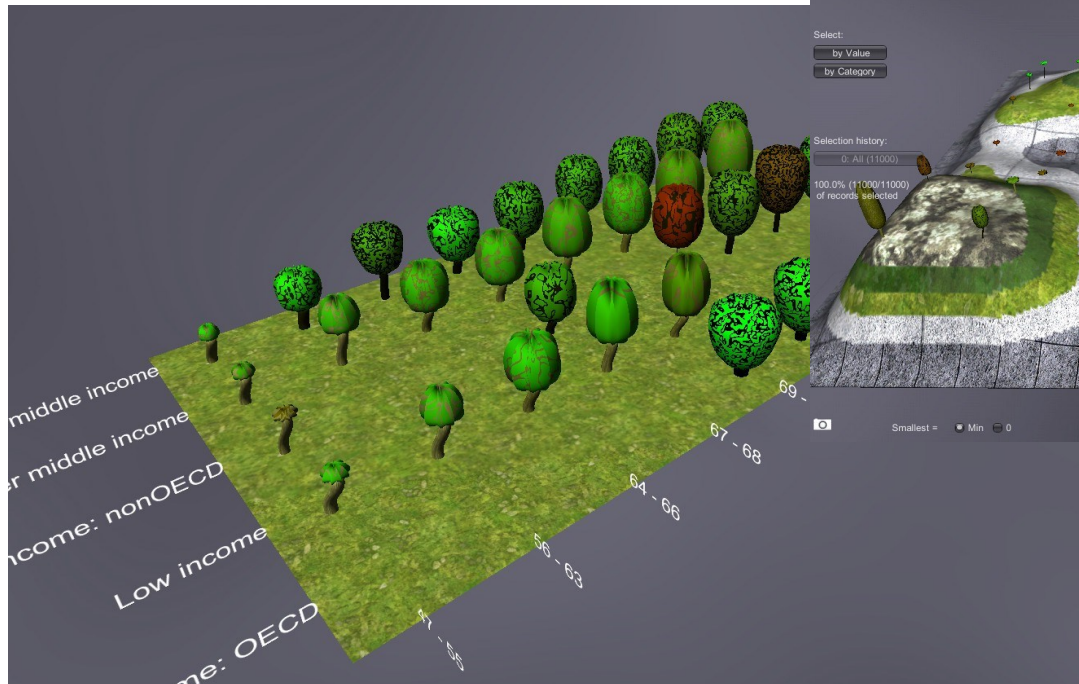


různé další kompozitní glyfy

- třeba enviromentální metafor
- glyfy jako storytelling
- vnášení emocí
- Leaf Glyphs



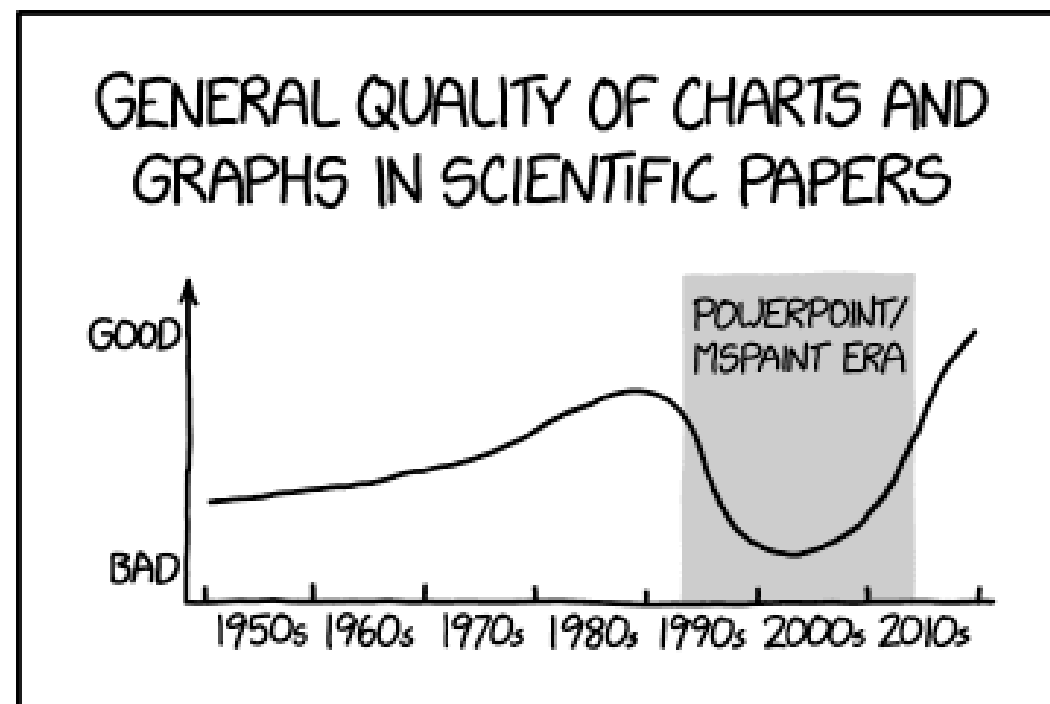
- metaforické 3D glyfy
- cesty do VR
- REEFSOME / curios.IT



vědecká komunikace

Další grafy ve vědě

- *histogram*
- *boxplot*
- *violin plot*

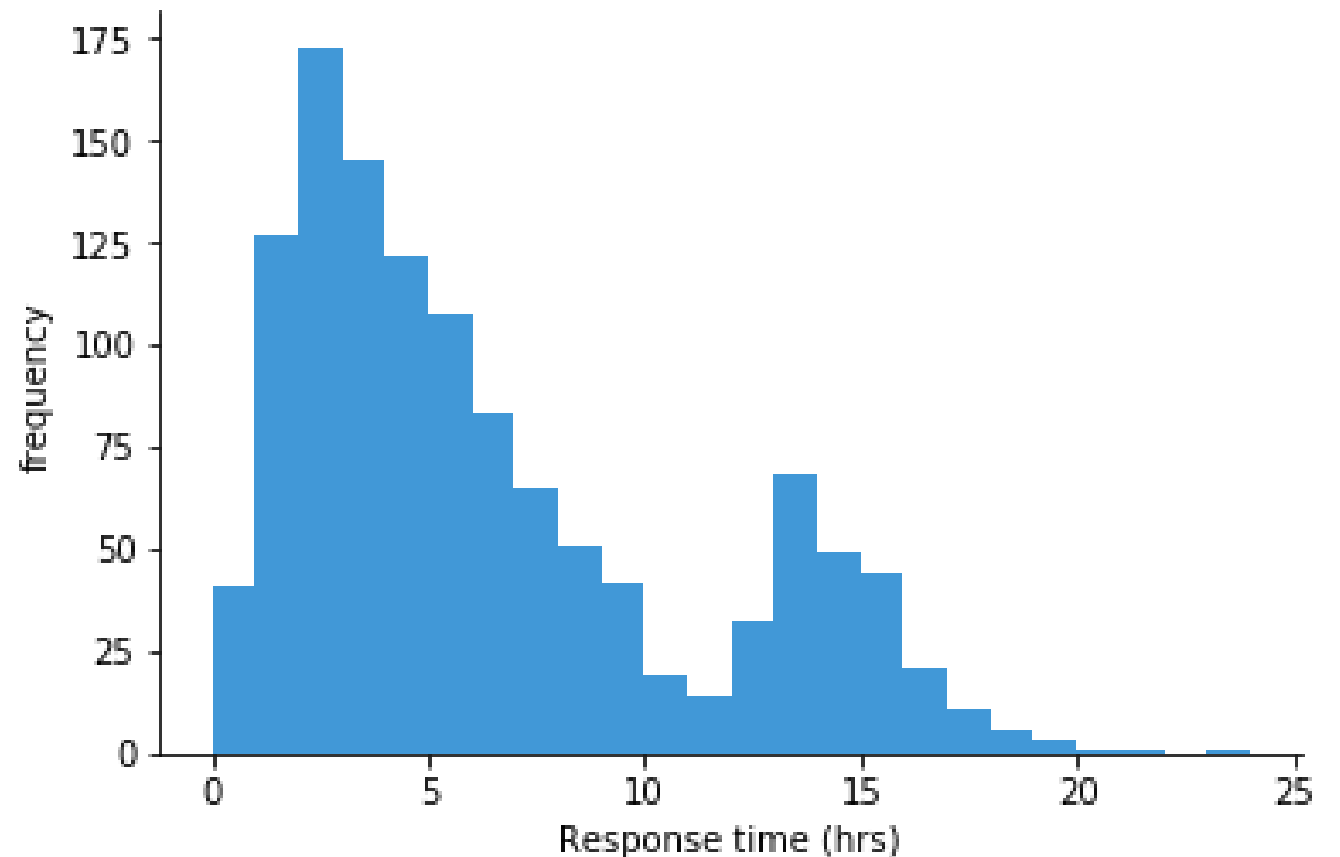


histogram

- vypadá jako sloupcový graf
- vizualizuje distribuci dat
- na ose x je interval
- na ose y četnost
- v čem je rozdíl?

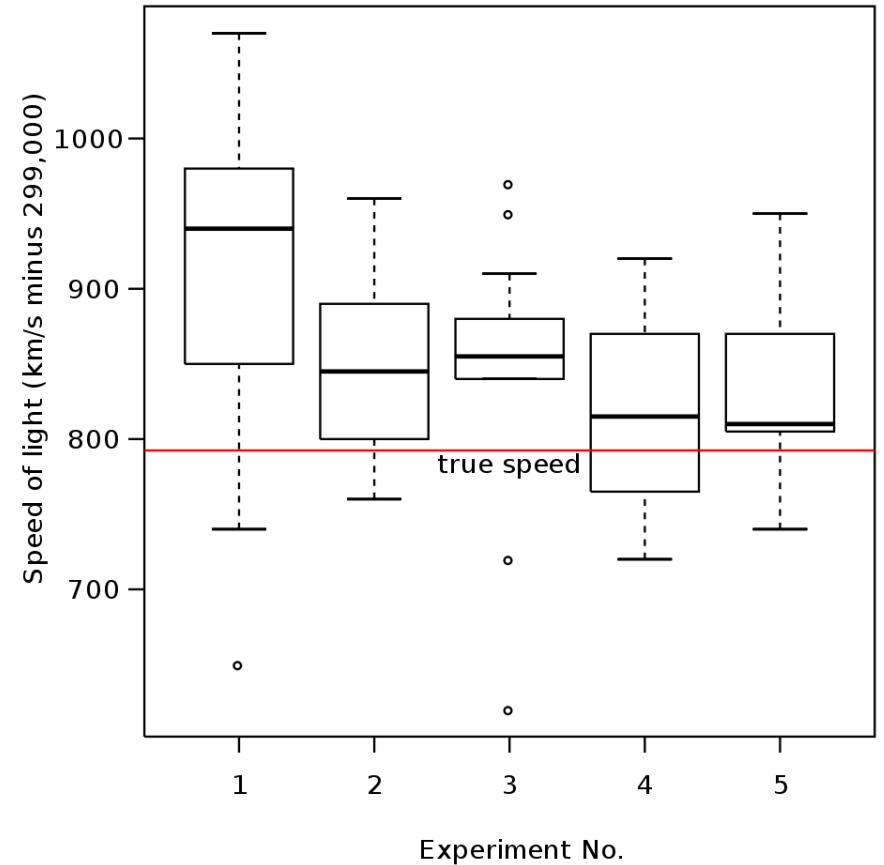
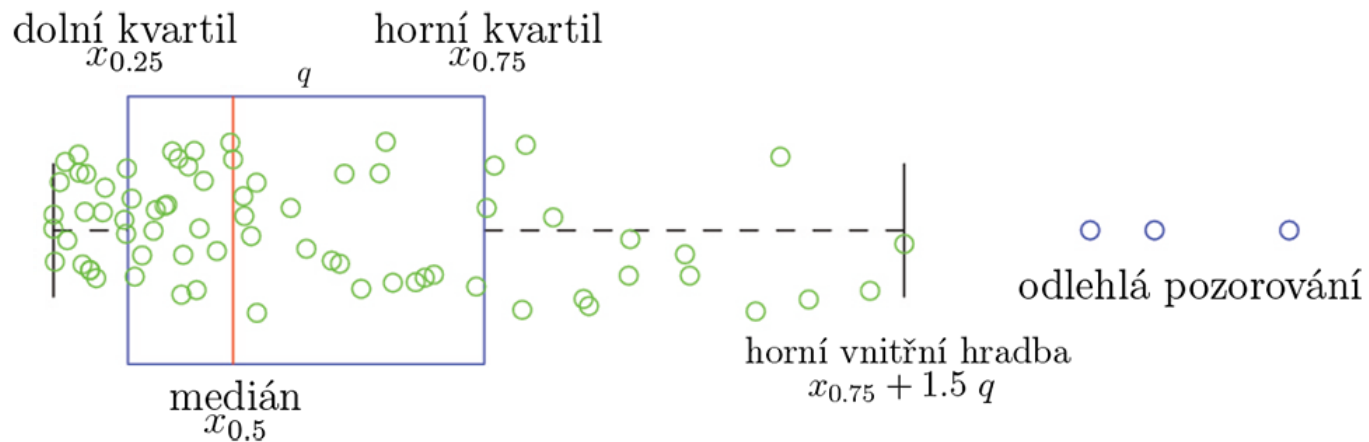


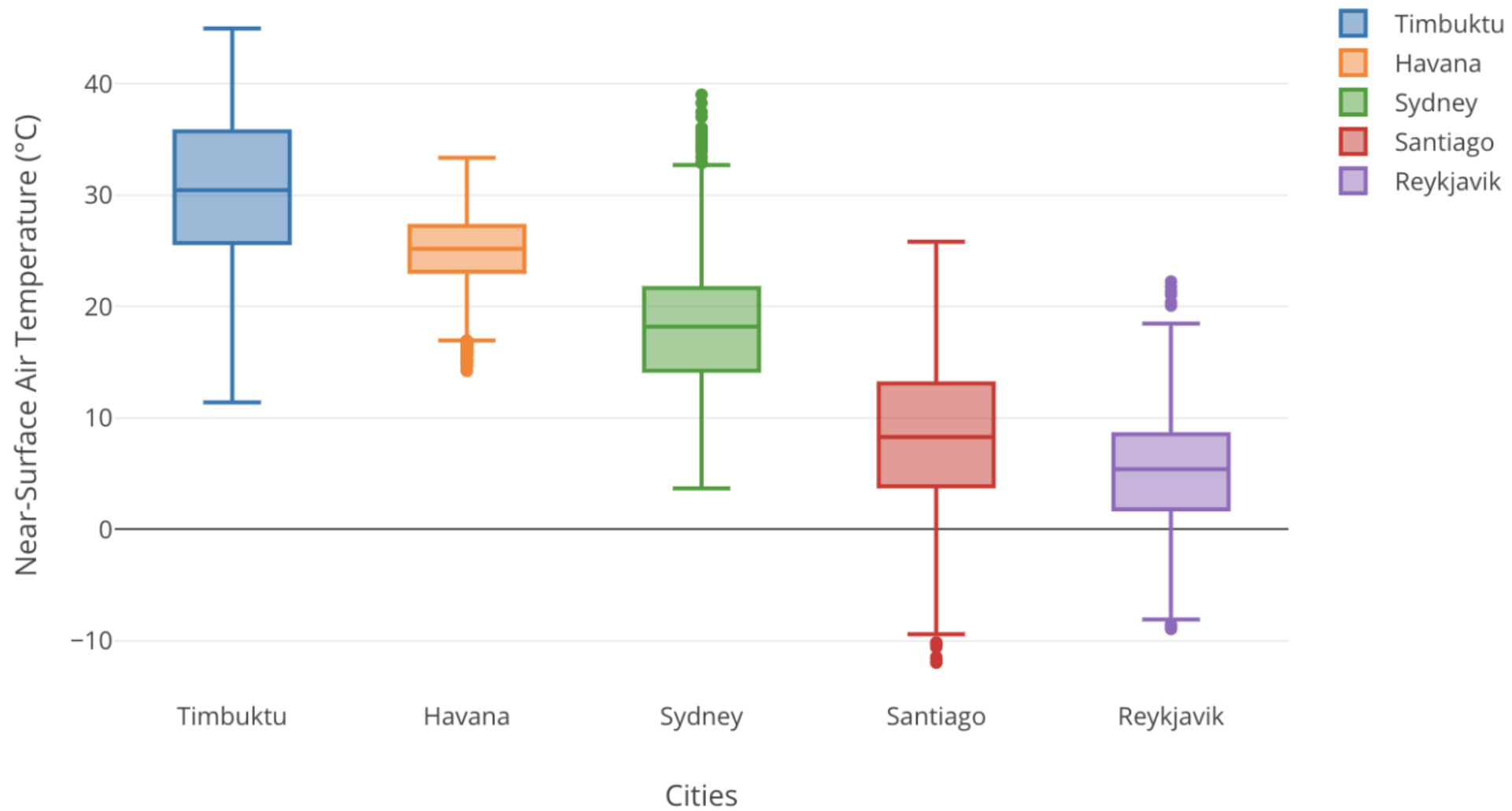
RAWGraphs



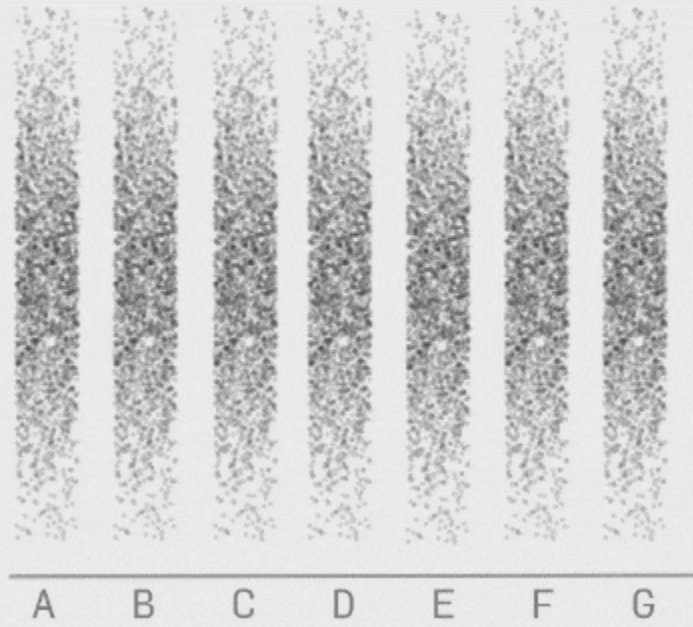
boxplot

- krabicový graf
- *box-and-whisker plot*
- zkoumání více sad dat
- *např. několik experimentů*

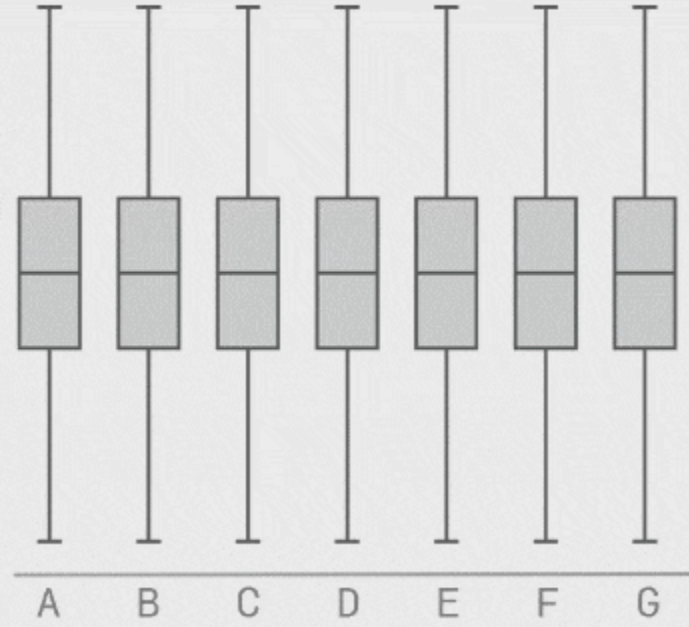




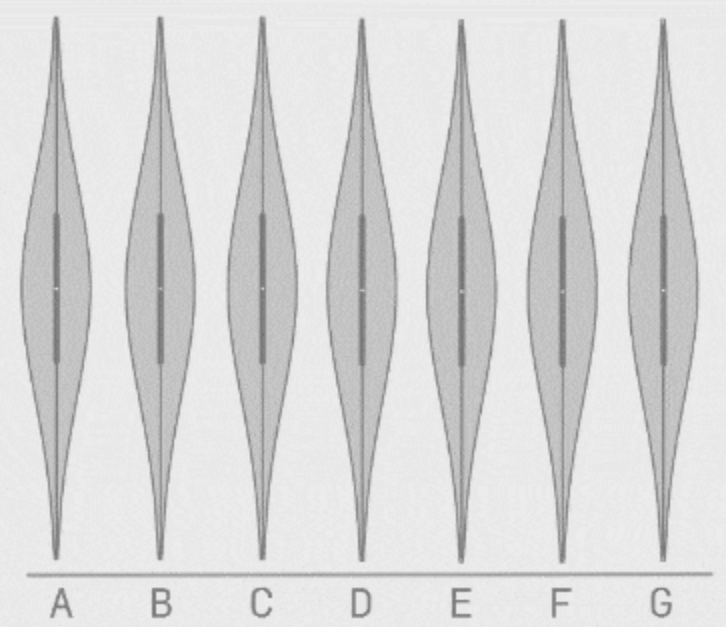
Raw Data

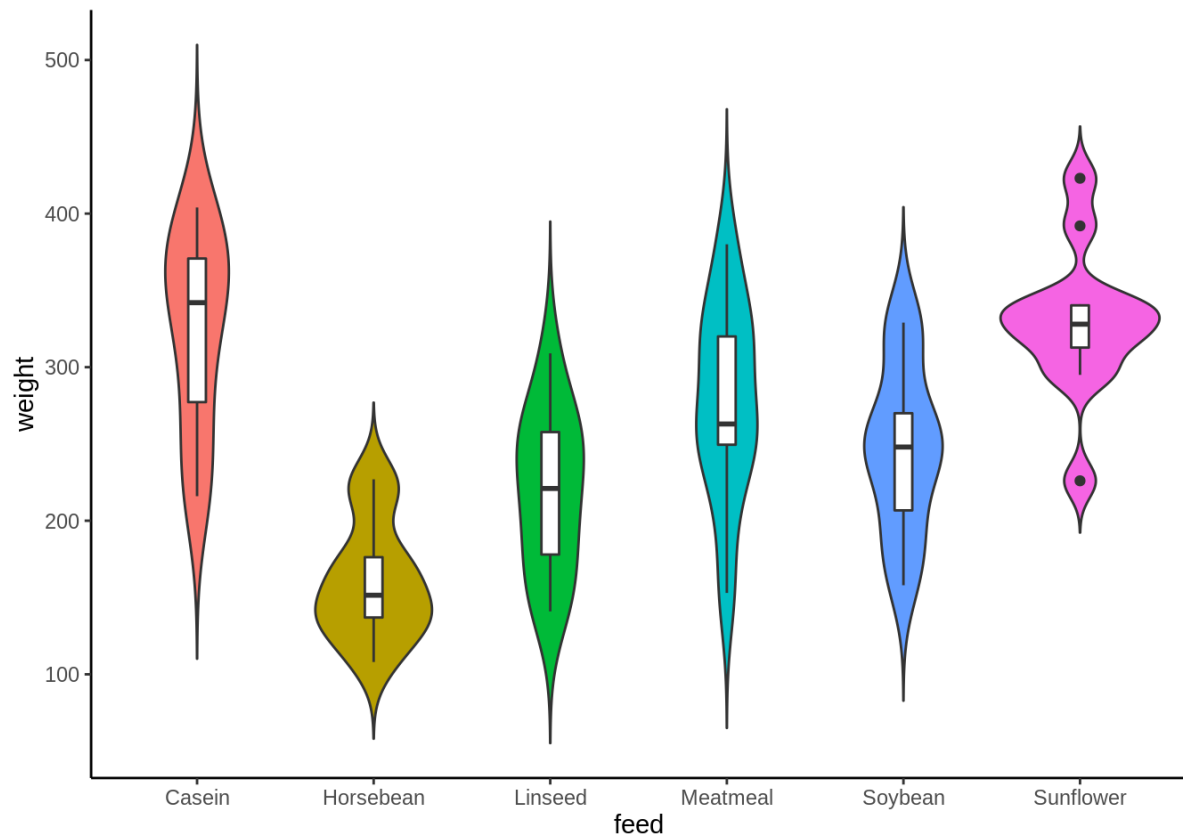


Box-plot of the Data



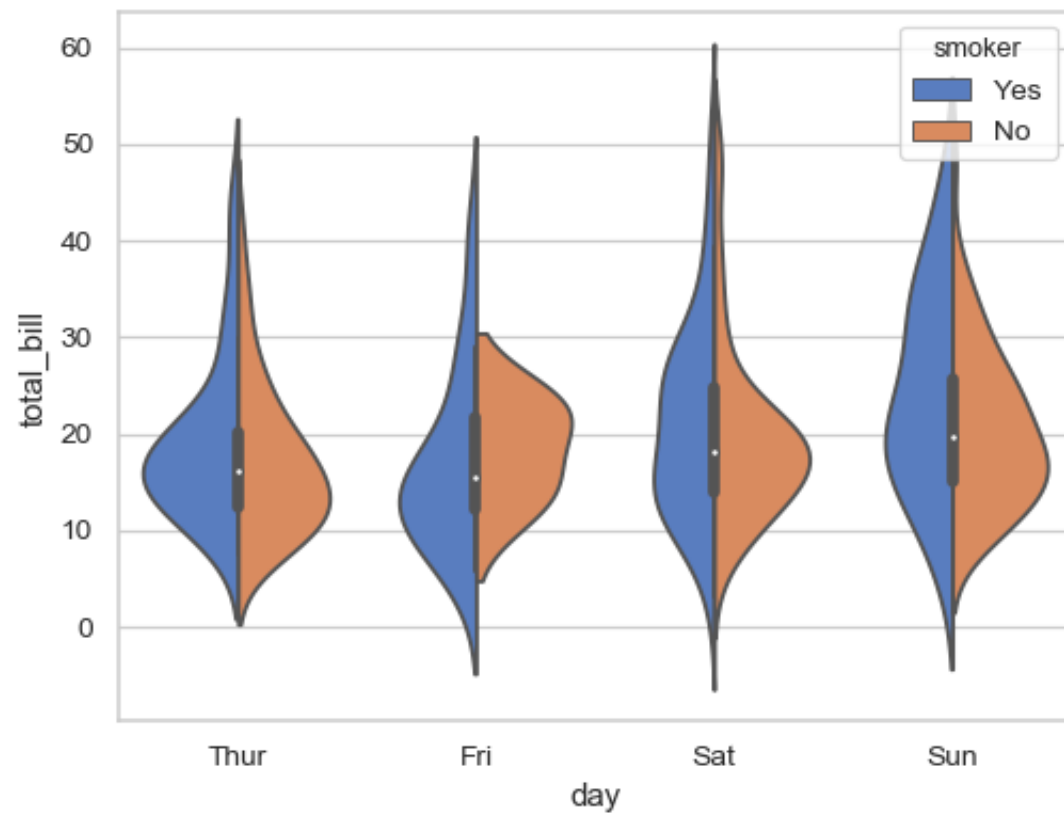
Violin-plot of the Data





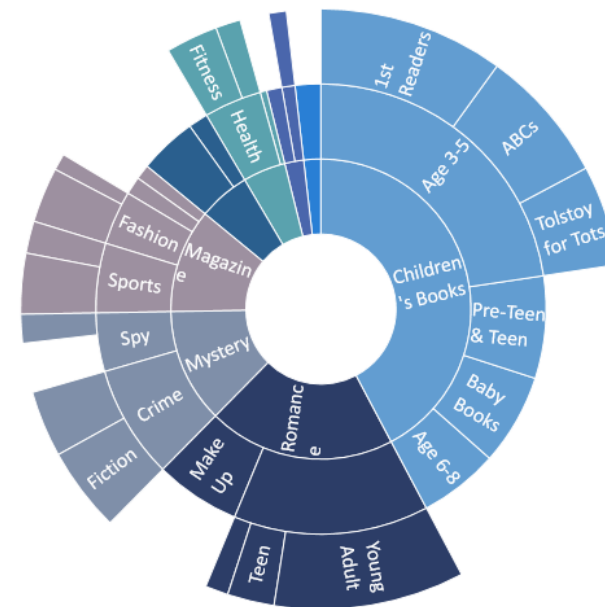
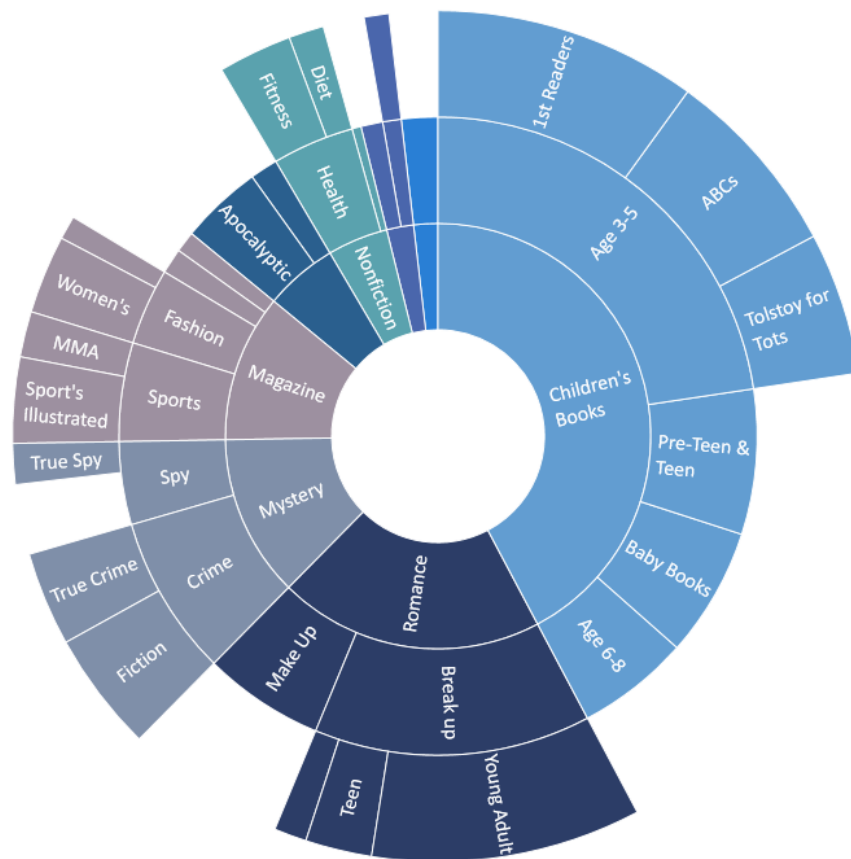
violin plot

split violin plot



Hierarchická data

- treemapy
- sunburst
- dendrogram



treemap

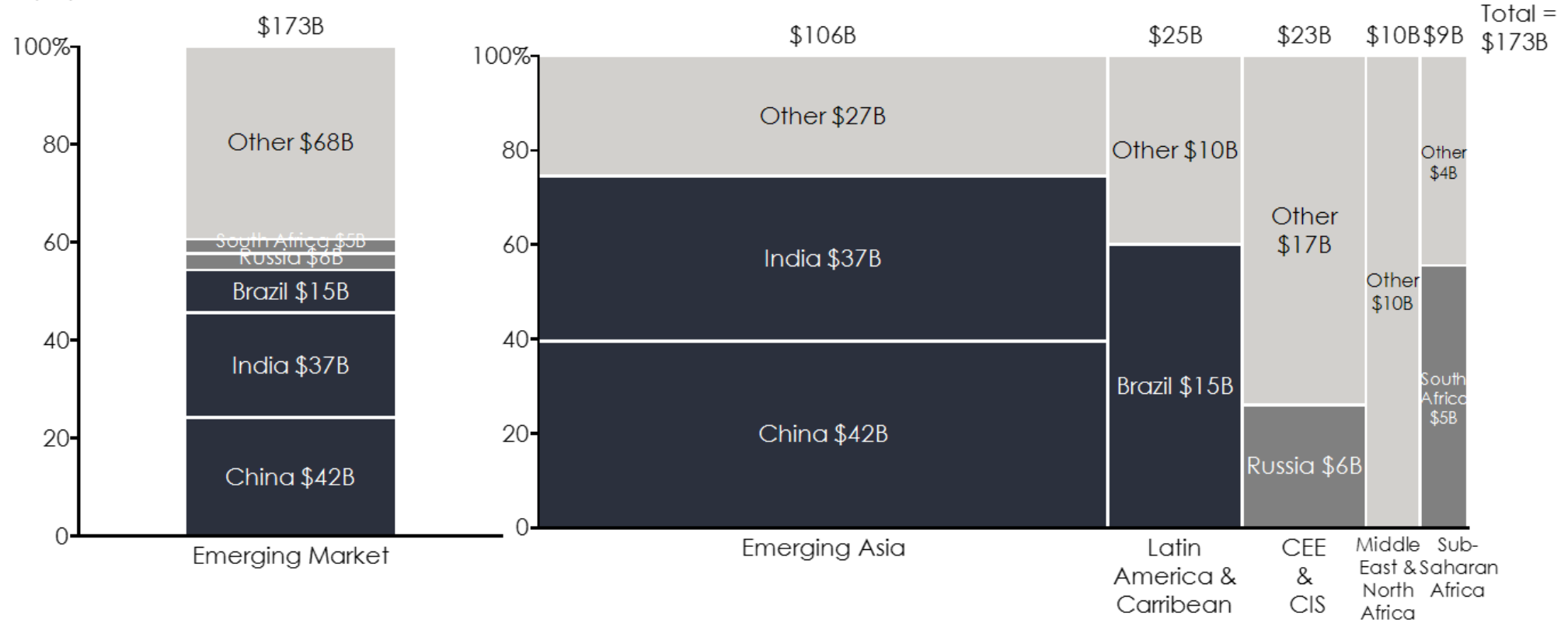
- Ben Shneiderman
- hierarchická data
- např. rozpočet
- využití úložiště
- souborová struktura
- <https://windirstat.net/>



RAWGraphs

mosaic marimekko chart

Emerging Market Private Equity Investments



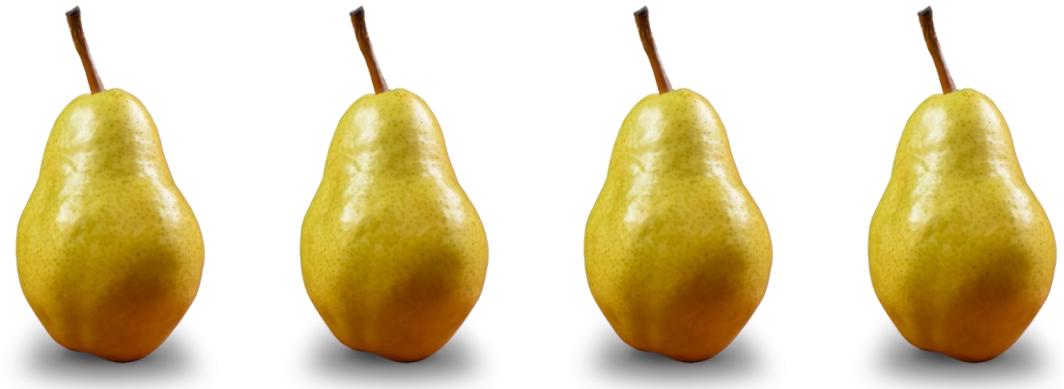
jakou formu mám použít?

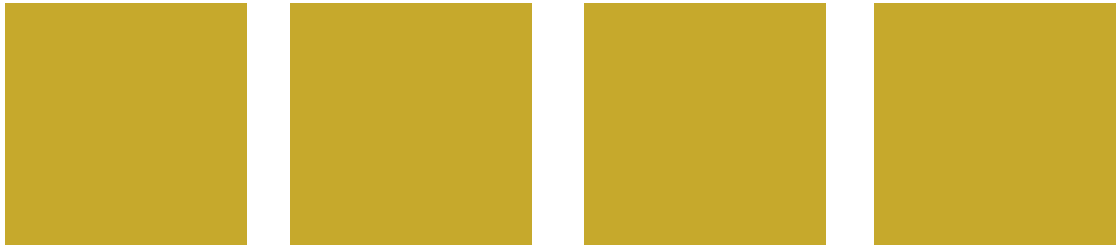
- 5 Na diskotéce na mýtině je 96 účastníků. Třetina jsou rusalky, zbytek jsou hejkalové a bludičky, přičemž hejkalů je o 14 více než bludiček. **Jaké je složení účastníků diskotéky?** (Znázorni figurálním diagramem.)

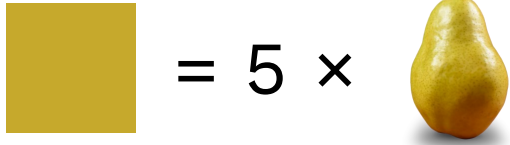
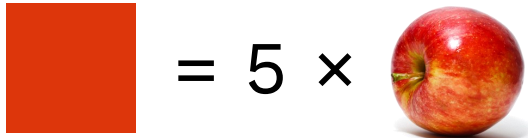


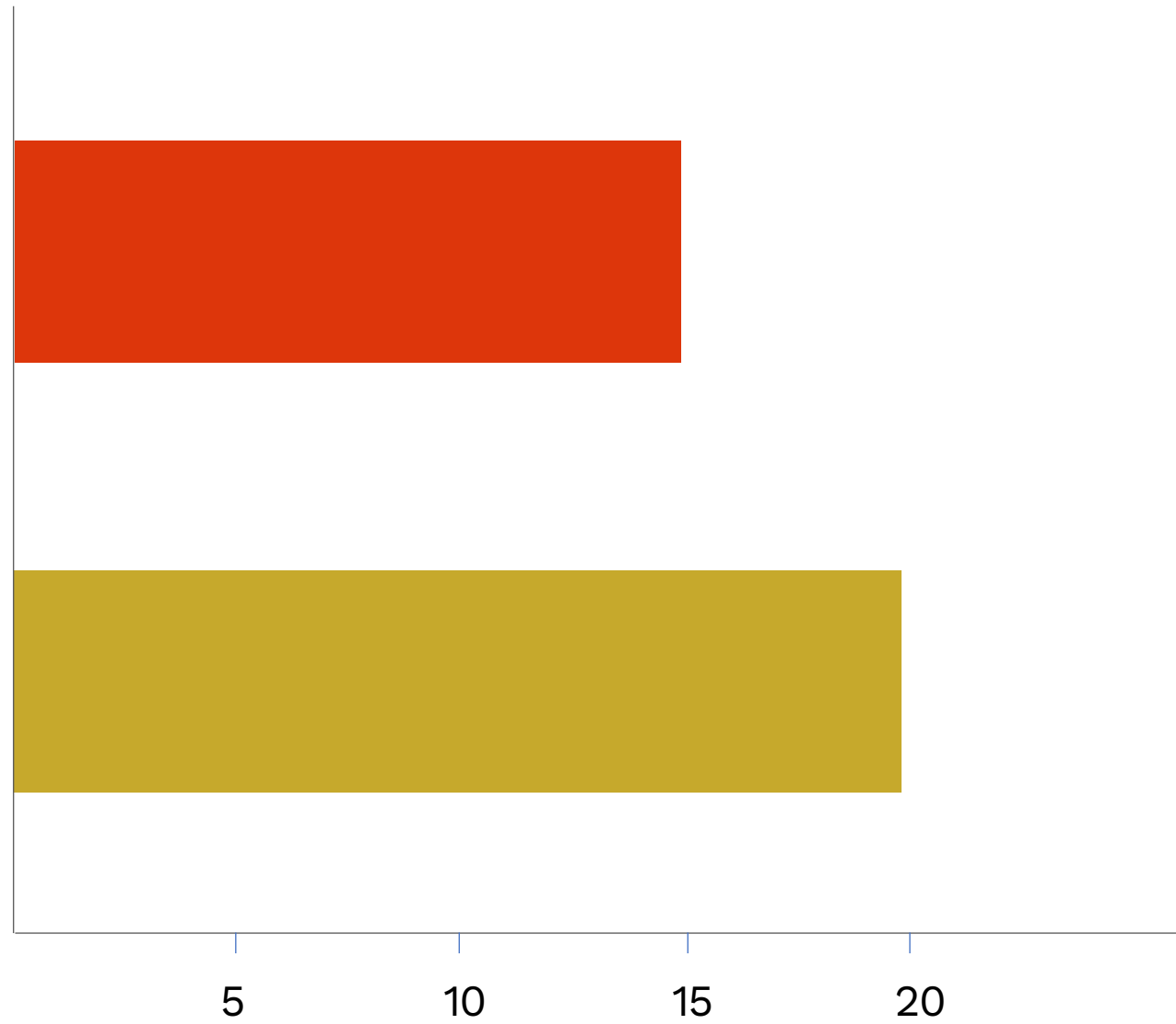






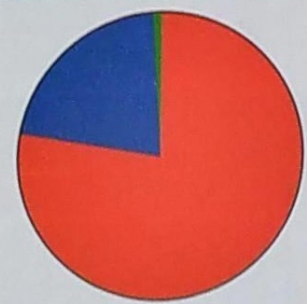






3

Na obrázku je znázorněno, z čeho se skládá vzduch.



- a) Ze kterých plynů se skládá vzduch?
- b) Kterého plynu je ve vzduchu obsaženo nejvíce?
- c) Kterých plynů je ve vzduchu nejmenší množství?

■ kyslík
■ ostatní plyny

VZDUCH

...níčkou života.

...ové a rostliny potřebují vzduch.

...nás, tvoří obal Země.

...idě, jak se můžeme vyhnout a poku...

Diagram znázorňující složení vzduchu



VZDUCH A PŮDA

Složení vzduchu:

Viš, že...

Čím výše stoupáme, tím je vzduch řídkší a hůř se nám dýchá.

... k dýchání člověku a který rybám?

...plynů.

...přibližně

...patrně.

...dy není

...telný.

...00 kilo-

1. Je k Zemi

...nazývaných



...eř tvoří plynný obal Země – **atmosféru**. Má vliv na proměny v neživé přírodě. Udržuje na Zemi teplotu, která je vhodná pro život. Většina živých organismů z ovzduší nemohla vůbec existovat.

SLOŽENÍ VZDUCHU

- dusík 78 %
- kyslík 21 %
- oxid uhličitý 0,1 %
- ostatní 0,9 %

12 století sir...

...sféře nazýváme vítr. Proudění vzduchu jsme schopni vnímat. Sílu větru můžeme být pro člověka ničícím živlem i pomocníkem.

obal Země se nazývá **ovzduší** nebo **atmosféra**. Kolem Země ho udržuje zemská přitažlivost.

Už víte, že vzduch je složený z mnoha plynů. Pro život organismů je nejdůležitější **kyslík**, pro rostliny také **oxid uhličitý**. Nejvíce je ve vzduchu **dusíku**.

Složení vzduchu

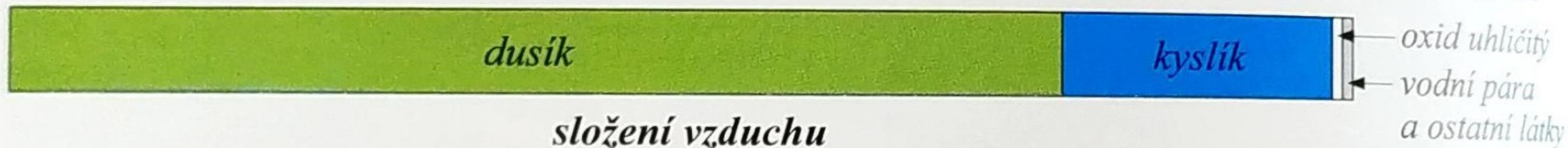
- dusík
- kyslík
- oxid uhličitý
- ostatní plyny

Jak se můžete přesvědčit o přítomnosti vzduchu?

Vysvětlíte, proč horolezci ve vysokých horách používají kyslíkové přístroje.

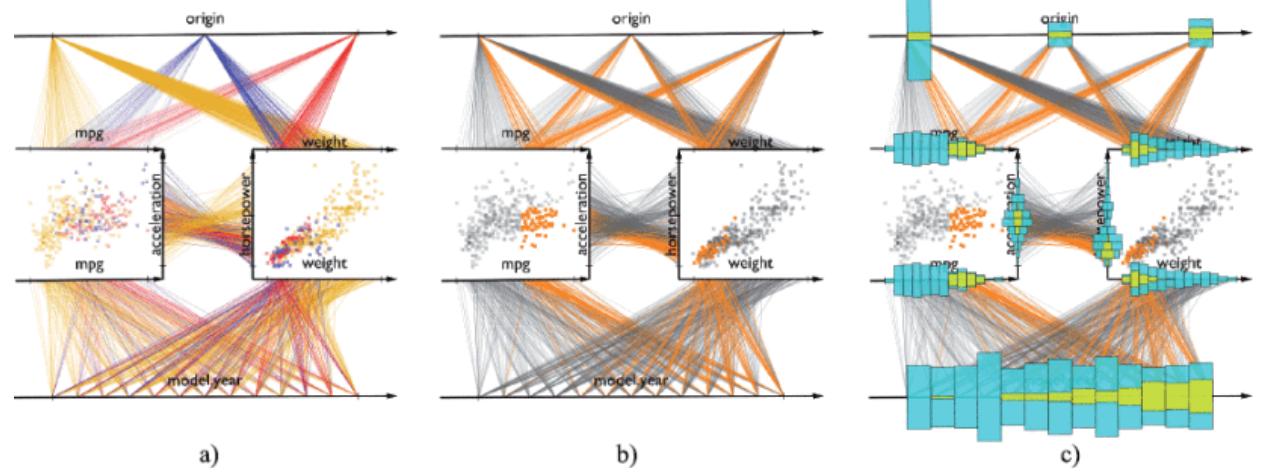
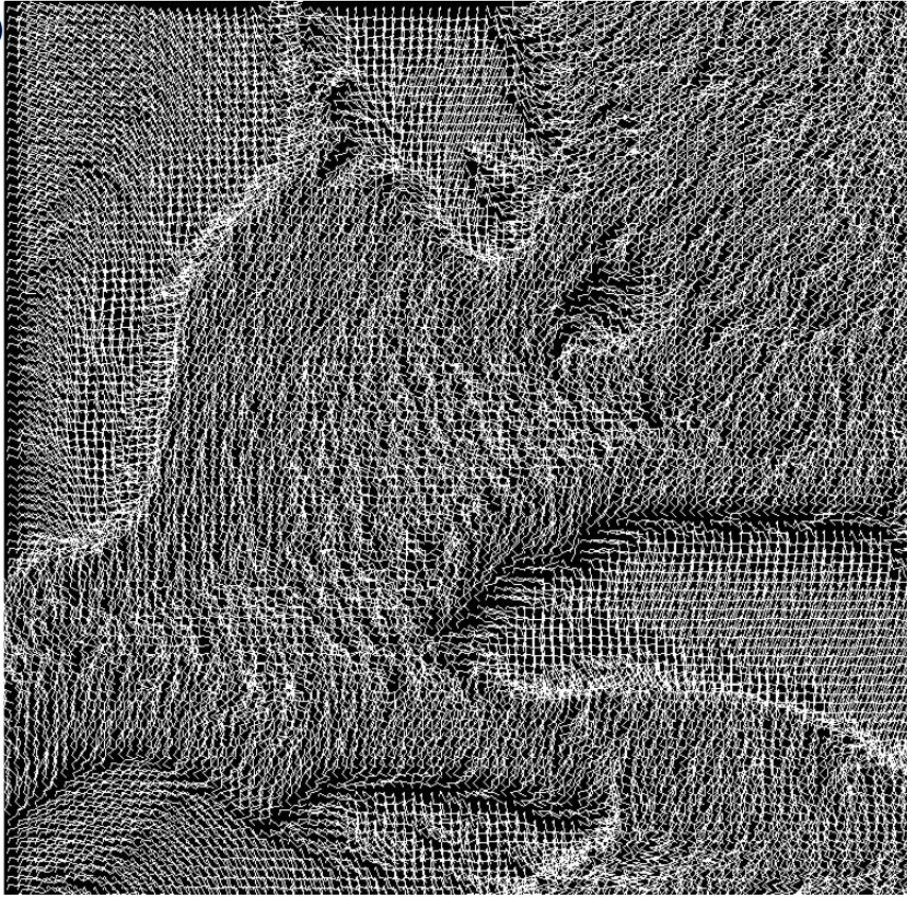
Složení vzduchu

Vzduch se skládá z několika **součástí**. Nejdůležitější z nich je **kyslík**. Ten potřebují ke svému životu zvířata, lidé i rostliny, protože ho dýchají. (Bez kyslíku bychom vydrželi žít jen několik minut.) Menší část vzduchu vyplňuje **oxid uhličitý**, který se uvolňuje při dýchání. Největší část vzduchu tvoří **dušik**.



Námět k pozorování – Jak dlouho vydržíte bez příjmu kyslíku?

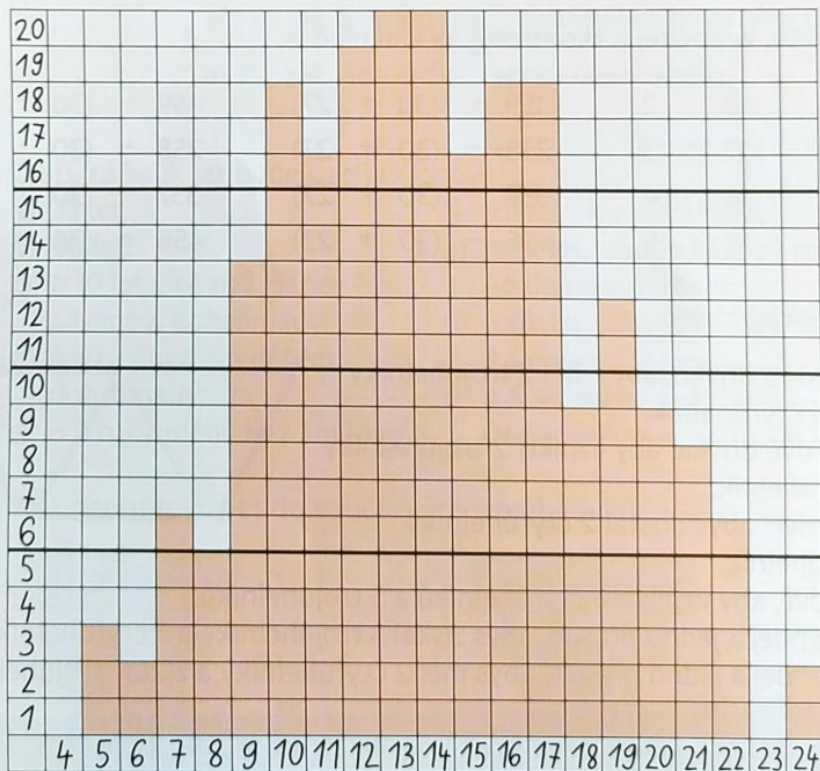
Zhluboka se nadechněte, zadržte dech. Změřte, kolik sekund (minut) vydržíte bez příjmu kyslíku.



4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
0	3	2	6	5	13	18	16	19	20	20	16	18	18	9	12	9	8	6	0	2

Vidíme, že se čísla 4 a 23 neobjevila ani jednou, číslo 5 se objevilo třikrát, číslo 6 dvakrát, ...číslo 13 dvacetkrát atd.

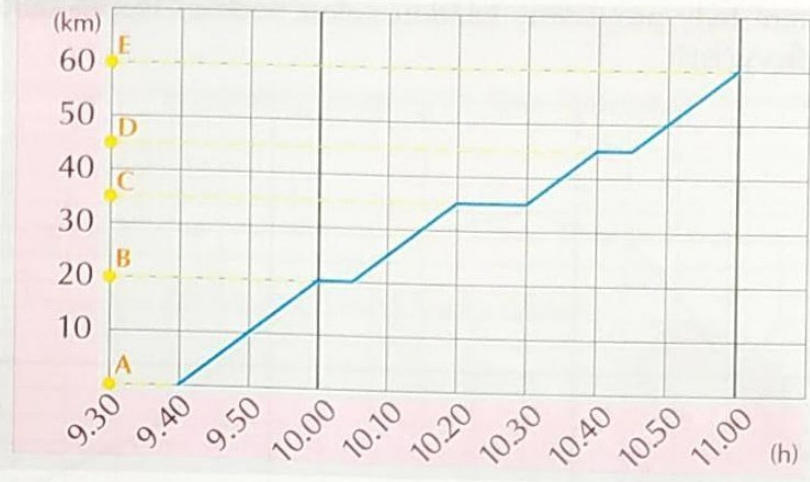
Starší bratr řekl Uršule, jak může číselnou tabulku znázornit.



Vyřeš úlohu 3
v PS na str.
32 a 33.



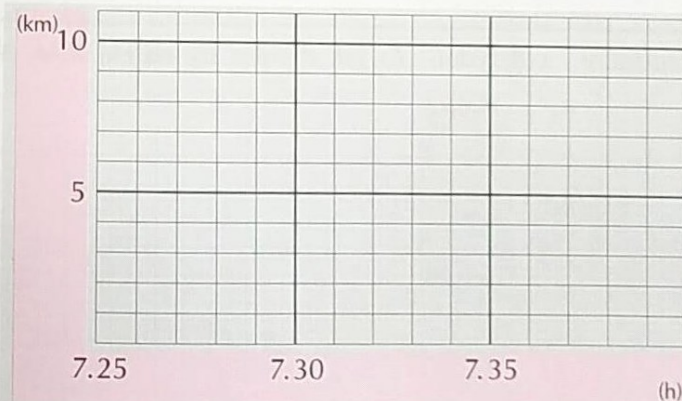
1 Zjisti z grafikonu vlaku:



	km	příjezd	odjezd
A			
B			
C			
D			
E			

- V kolik hodin vyjel vlak z místa A?
- Kolik měl na trase A–E zastávek?
- Na kterém kilometru byla první a na kterém poslední zastávka?
- Kolik km vlak ujel?
- Co ještě můžeš zjistit?
- Vypočítej, jakou průměrnou rychlostí vlak jel.
- Sestav jízdni řád tohoto vlaku.

2 Sestroj grafikon vlaku č. 23 356, jestliže na zastávku se počítá 1 minuta.



km	PŘ Os	Vlak	23 356
0	Suchdol nad Odrou		7.25
5	Hladké Životice		7.31
7	Stachovice		7.35
10	Fulnek		7.39

Jsou průměrné rychlosti vlaku mezi jednotlivými stanicemi stejné?

Absolute frequencies of types and subtypes by grades (G) and relative frequency by all types

Types and subtypes	G1	G2	G3	G4	G5	Total of types	% from all
Pictorial units	635	353	177	116	48	1329	48.15%
<i>Free-form pictographs</i>	331	122	26	29	9	517	18.73%
<i>Structured pictographs</i>	304	230	125	81	29	769	27.86%
<i>Abacus pictographs</i>		1	26	6	10	43	1.56%
Numbered lines	111	150	193	98	124	676	24.49%
<i>Regular numbered lines</i>	110	148	191	96	121	666	24.13%
<i>Thermometers</i>	1	2	2	2	3	10	0.36%
Part-whole (PW) graphics	50	143	92	125	116	526	19.06%
<i>Regular PW graphic</i>	50	132	38	45	58	323	11.70%
<i>Bar models</i>		1	18	9	9	37	1.34%
<i>Pie graphs</i>		10	36	71	49	166	6.01%
Bar graphs	2	35	11	19	67	134	4.86%
<i>Regular bar graph</i>	2	35	9	17	67	130	4.71%
<i>Histogram</i>			2	2		4	0.14%
Line graphs		1	2	6	40	49	1.78%
Point graphs		1	4	9	31	45	1.63%
<i>Dot plots</i>		1		1	8	10	0.36%
<i>Scatterplots</i>			4	8	23	35	1.27%
Specific types		1				1	0.04%
<i>Gauge graph</i>		1				1	0.04%
Totals	798	684	479	373	426	2760	100.00%

Table 2. Counts of analytical units by publishers and areas; with per 100 pages rates (sorted in descending order by total rate).

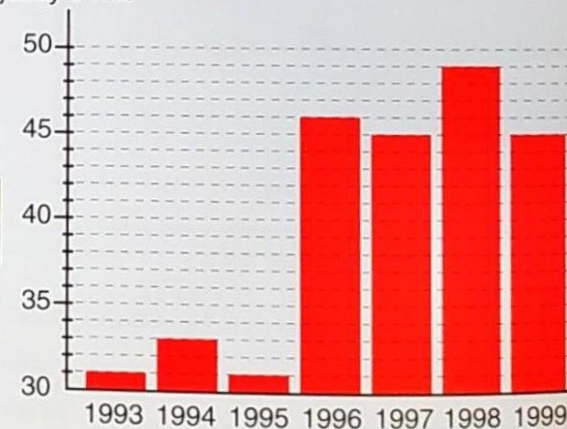
Publisher	HTW		MAT		Total	
	<i>Units</i>	<i>Per 100 p.</i>	<i>Units</i>	<i>Per 100 p.</i>	<i>Units</i>	<i>Per 100 p.</i>
TAKTIK	14	1.357	83	8.401	97	4.802
Nová škola	32	1.438	115	7.099	147	3.822
Prometheus	0	0.000	46	2.979	46	2.979
H-MAT	0	0.000	21	2.706	21	2.706
Prodos	2	0.252	43	3.415	45	2.192
FRAUS	41	2.188	41	2.137	82	2.162
ALTER	7	0.468	46	3.948	53	1.992
DIDAKTIS	6	0.713	32	2.647	38	1.854
NŠ DUHA	4	0.378	23	2.059	27	1.242
SPN	3	0.249	20	1.860	23	1.008
Total	109	1.011	470	3.709	579	2.469

kategorie úlohy

6. V grafu je znázorněn počet výjezdů českých občanů přes hranice (hodnoty jsou zaokrouhleny na miliony). Přečti z něj, kolik výjezdů se uskutečnilo v letech 1993 až 1999, a údaje zapiš do tabulky. Podle znázornění utvoř dvě slovní úlohy. (1 úloha = 1 b)

rok	1993	1994	1995	1996	1997	1998	1999
výjezdy v mil.							

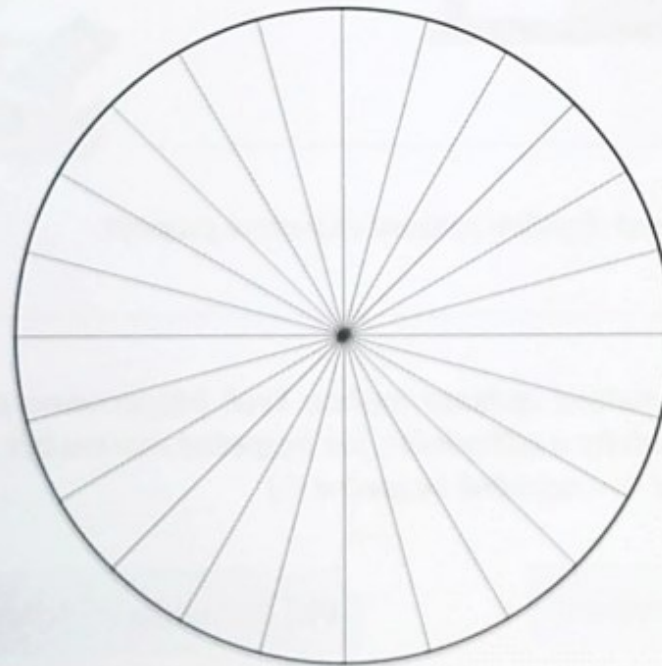
výjezdy v mil.



Zdroj: Ministerstvo pro místní rozvoj nebo www.cot.cz.

kategorie úlohy

2. Podle hesel v nabídce vybarvěte následující kruhový graf vašeho denního režimu. 24 částí tohoto grafu představuje 24 hodin jednoho všedního dne. Položku, která není součástí vašeho všedního dne, z nabídky vyškrtněte. Na pořadí údajů nezáleží.



- spánek
- pobyt ve škole
- cesta do školy a zpět
- úkoly a příprava na vyučování (včetně využití internetu ke studiu)
- jídlo (všechna jídla)
- sport, pobyt venku, kontakt s kamarády
- chat, telefonování, hry na počítači, sociální sítě
- četba
- práce v domácnosti
- jiné _____

While most employees report microaggressions, marginalized groups are disproportionately affected.

Women's personal experiences in the workplace, multiples of all men's experiences

Men

	Others get credit for my ideas	My judgment is questioned	I have been mistaken for someone more junior	I have been confused with someone else of the same race or nationality	Others comment on my appearance	Others comment on my emotional state
All women	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Women of color	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
LGBTQ+ women	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Women with disabilities	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Source: *Women in the Workplace 2023*, McKinsey & Company and LeanIn.Org

McKinsey & Company

visualization
onboarding

„A common theme during the interviews was the lack of user testing.“

[doi:10.48550/arXiv.2310.01935](https://doi.org/10.48550/arXiv.2310.01935)



Kam dál?

<https://datavizcatalogue.com/>

<https://www.data-to-viz.com/>

Charticulator & smutek

Datawrapper