

Vizualizace dat

Vizualizační formy prakticky

2. 12. 2021

Vizualizační metafory

- je jich mnoho
- některé se osvědčily
- některé se ještě hledají
- některé jsou doménově specifické
- některé vyvolávají kontroverze

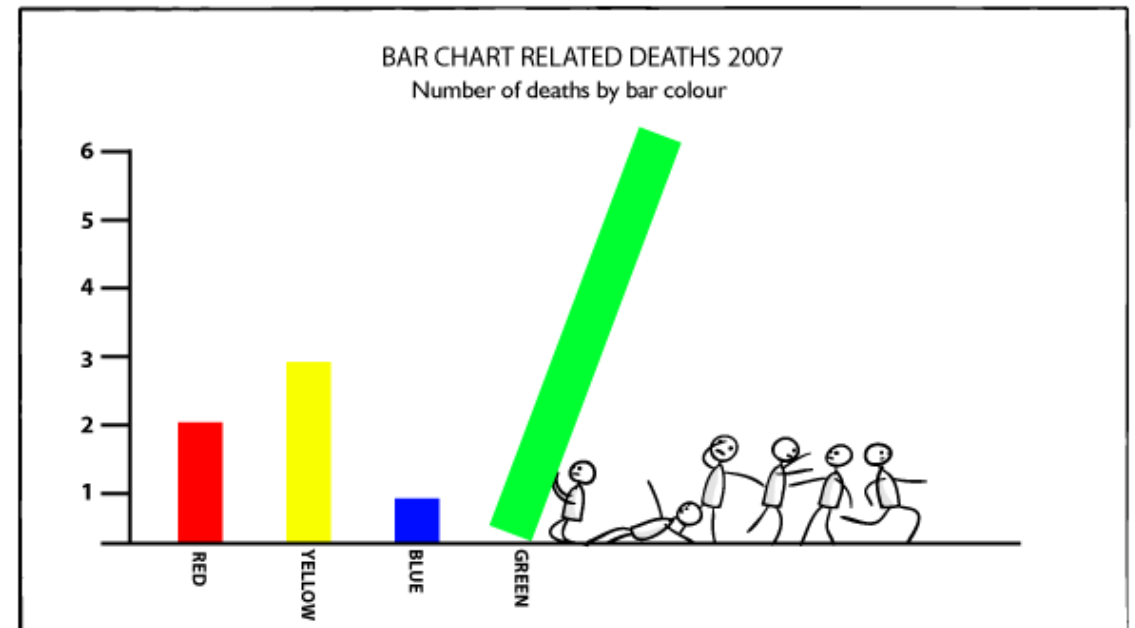


tradiční metafory

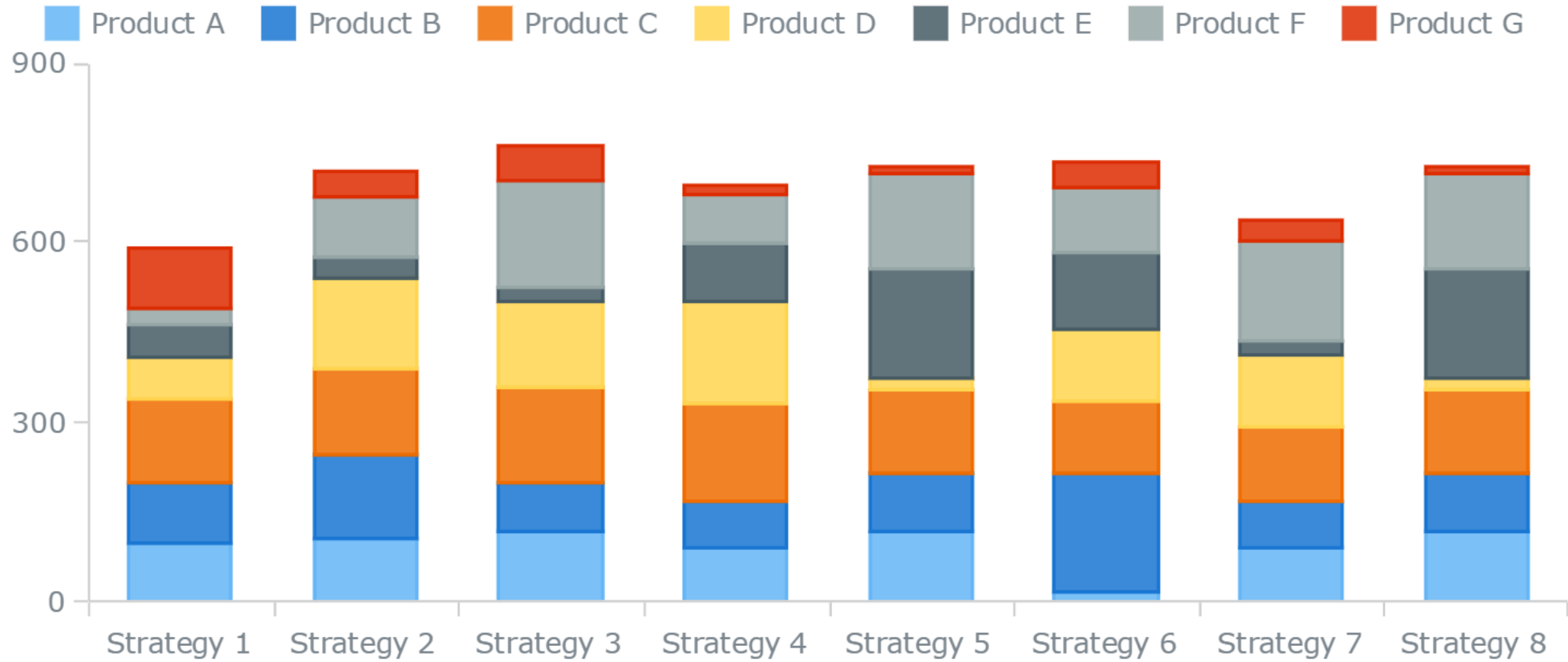
Sloupcové grafy (a jeho varianty)



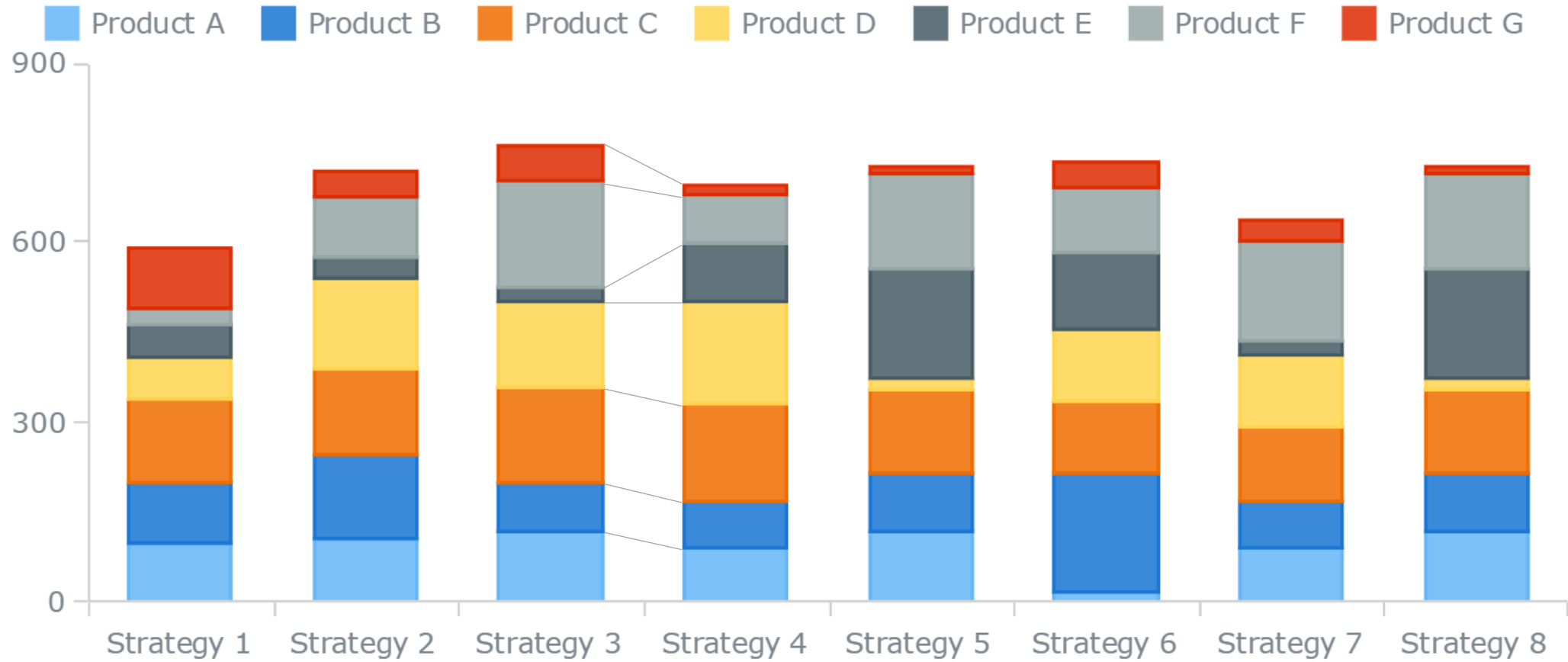
- rozložení souboru dat skrze sloupce
- délka mapuje hodnoty
- sloupcový klasický
pruhový graf
- skládaný sloupcový
100% skládaný sloupcový



Compare sales strategy

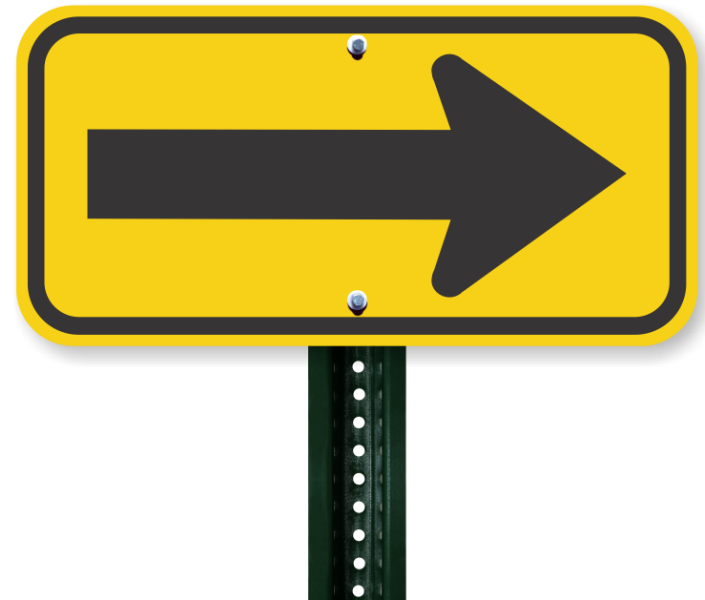


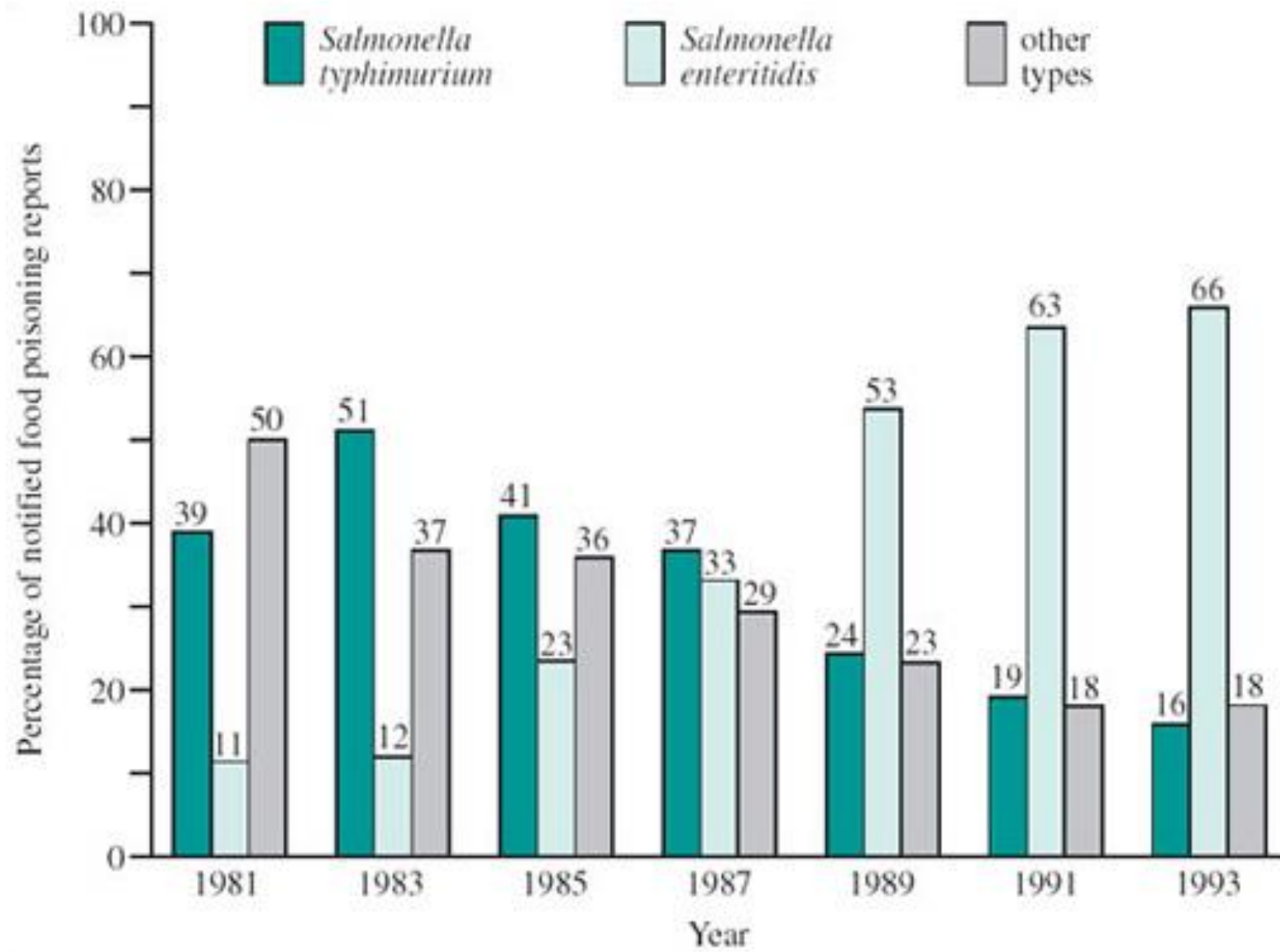
Compare sales strategy



Pár tipů ke sloupcům

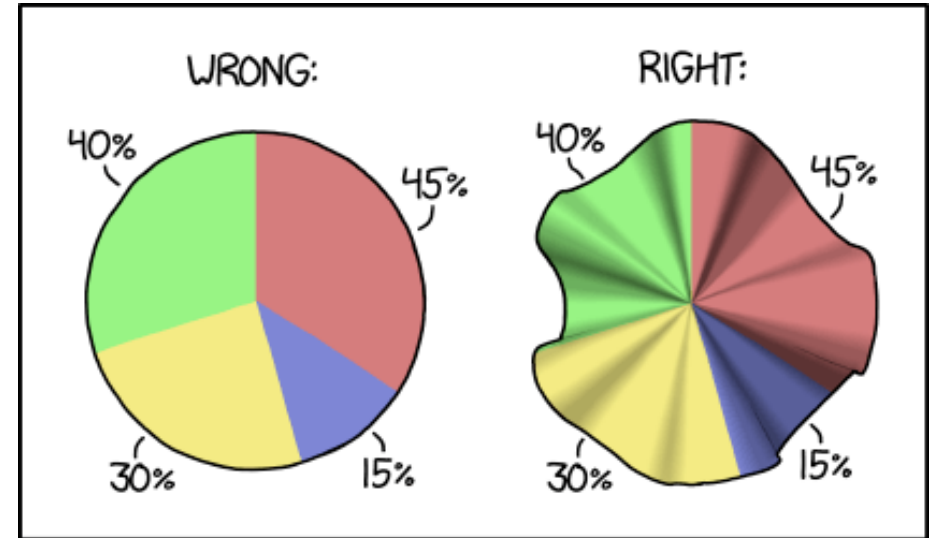
- mezery mezi sloupci
- příliš dlouhá legenda
- přirozená posloupnost osy (2, 4, 6...)
- pozitivní a negativní prostory
- ...





Koláčové grafy

- kontroverzní forma
- obrovská výhoda: 100 %
- má své velké limity
- někdo na něj nedá dopustit
- hádáme se o něm již přes sto let
- [článek o historii koláčového grafu](#)

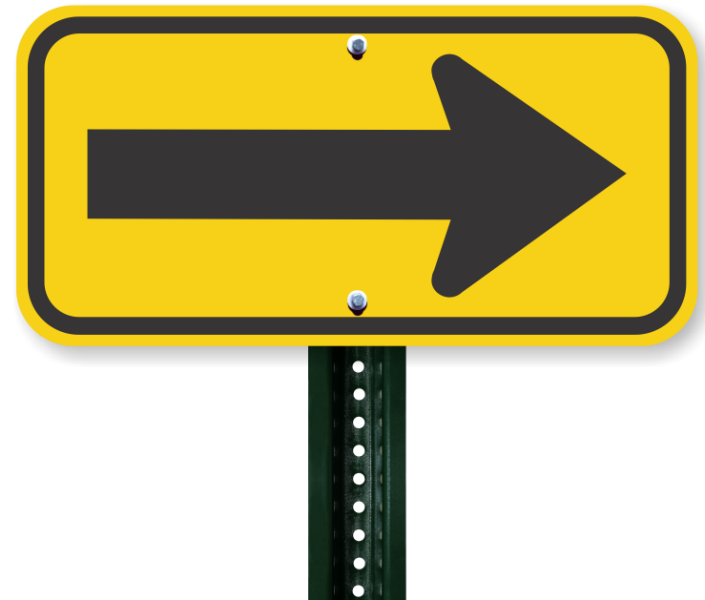


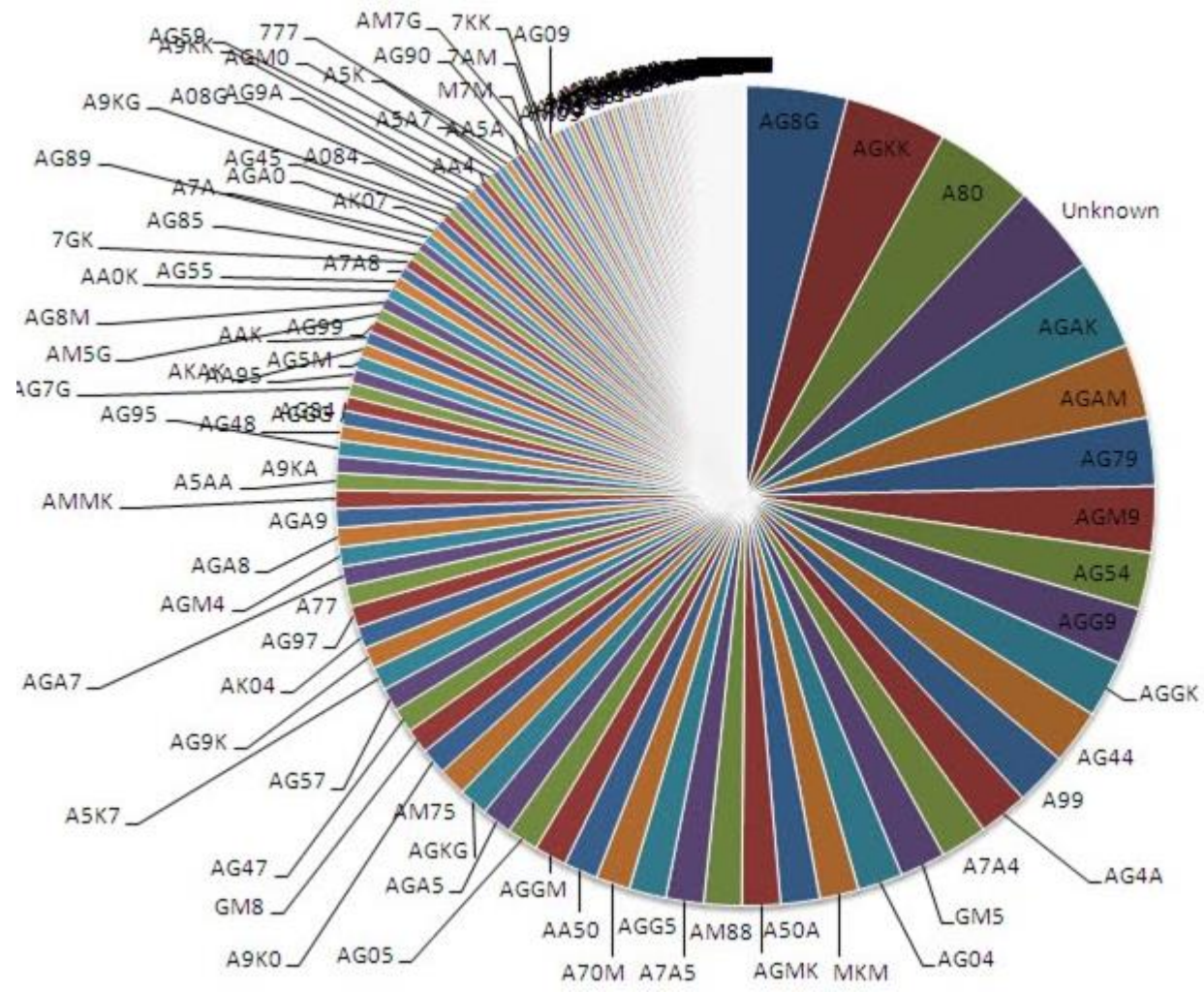
HOW TO MAKE A PIE CHART IF YOUR PERCENTAGES DON'T ADD UP TO 100



Pár tipů ke koláčům

- omezený počet výsečí
- „rozložení“ nevyužívat
- doporučená posloupnost výsečí
- ...





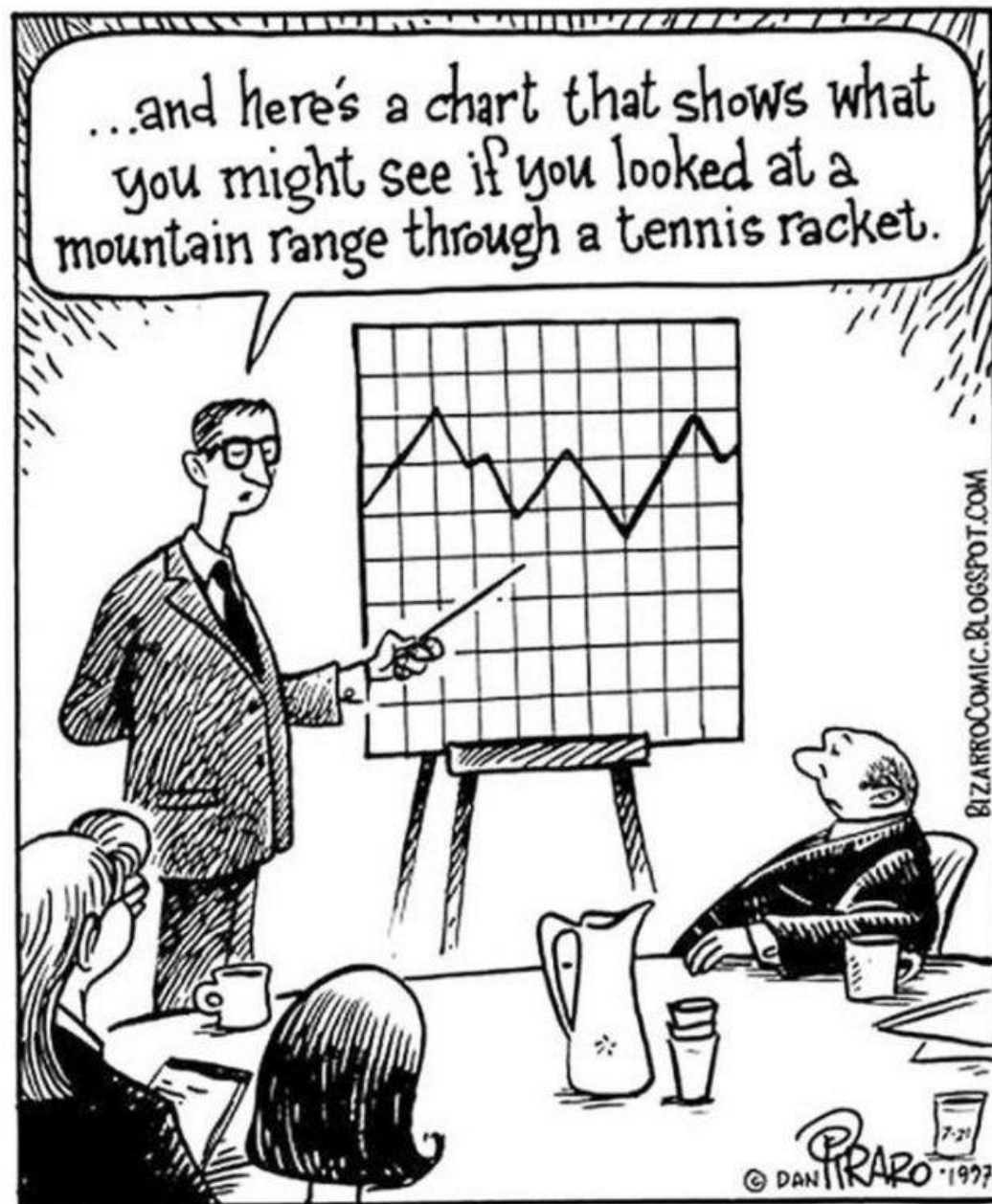
Koblihové grafy

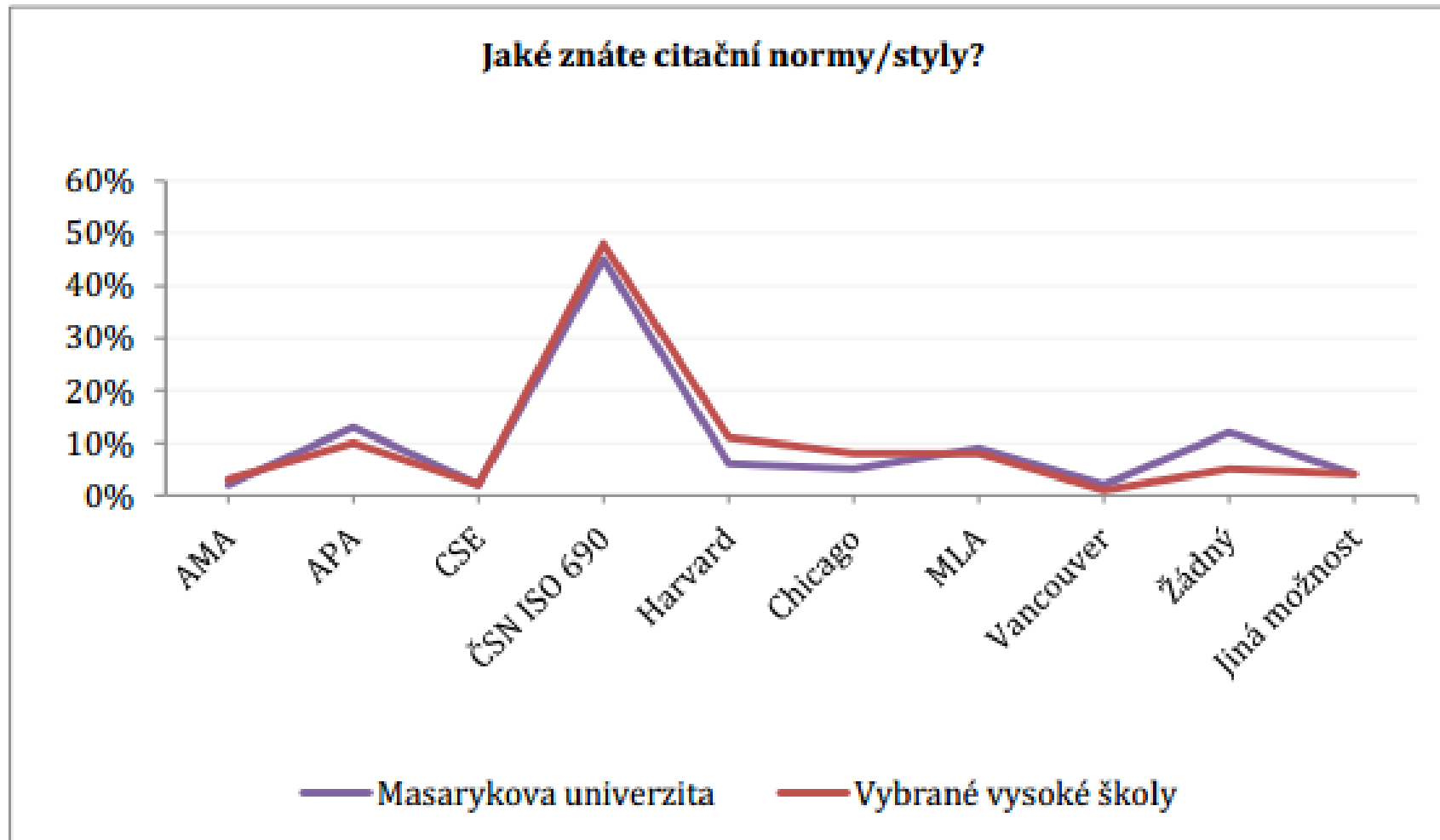
- americký donut
- problematická forma ve většině případů
- střed znemožňuje určité formy čtení



Spojnicové grafy

- trendy v čase
- pouze pro spojitě hodnoty!
- nespojují kategorické hodnoty

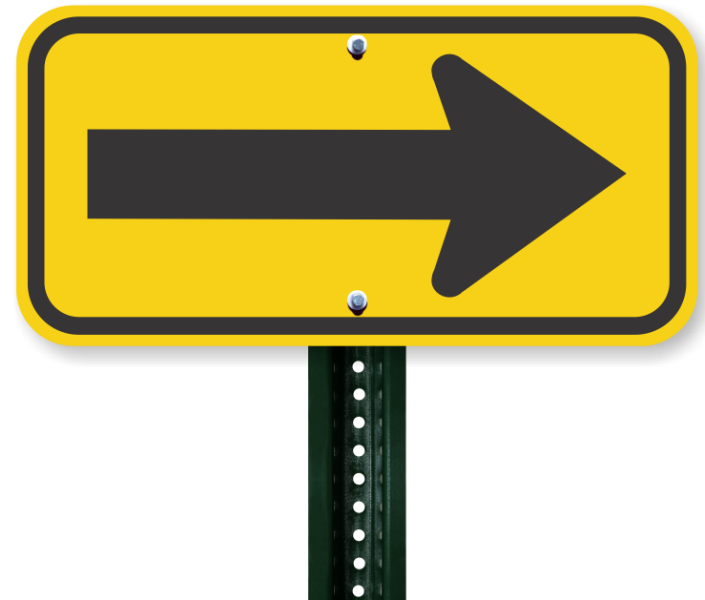




Graf č. 39: Znalost citačních norem - Masarykova univerzita, vybrané vysoké školy

Pár tipů ke spojnicím

- limit počtu linií
- síla spojnic
- ...



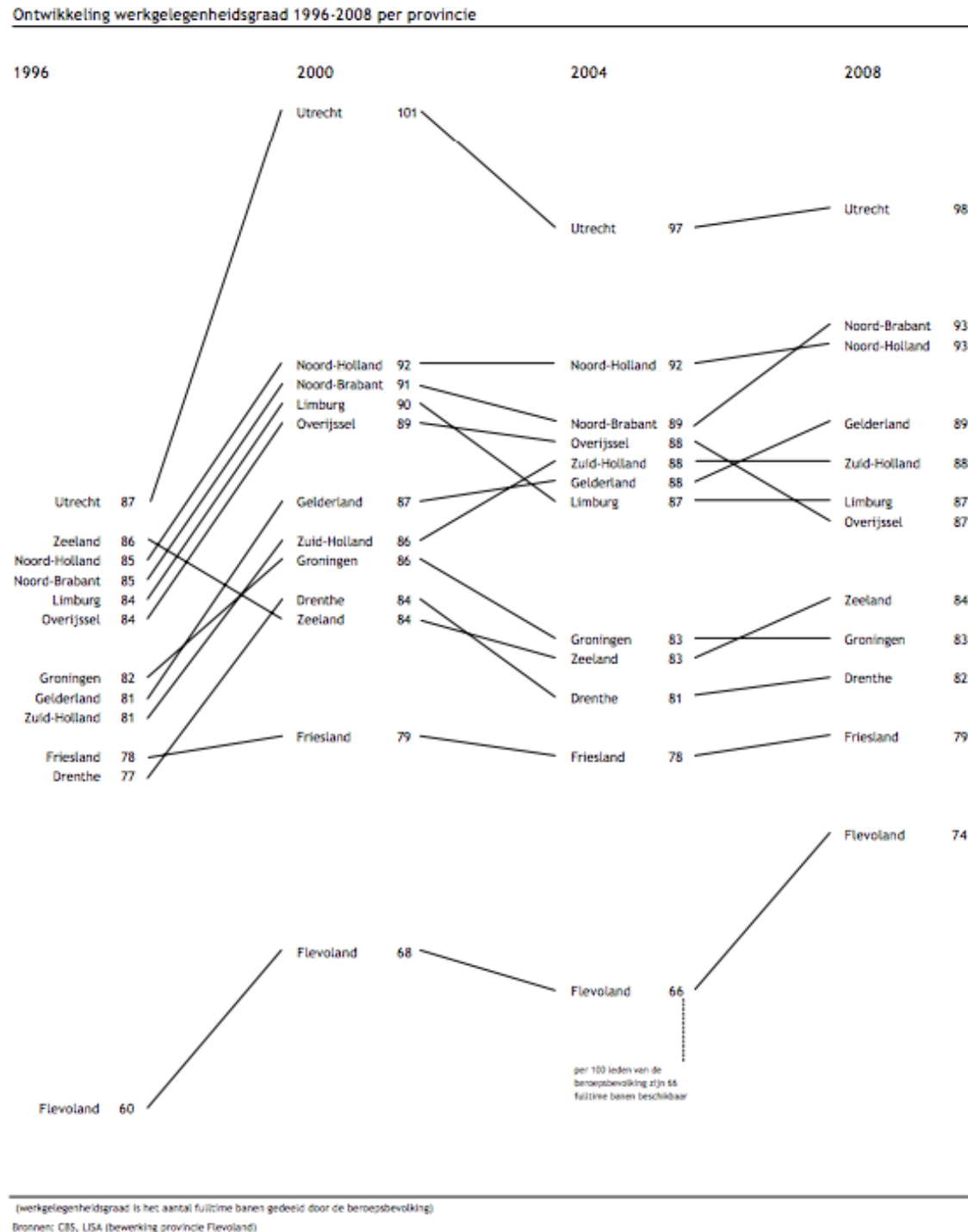
variacie na spojnice

Slopegraph

- v podstatě osekáný spojnicový
- velkým propagátorem je Tufte
- dává smysl pro určité případy
- pouze dvě hodnoty na ose x



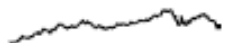
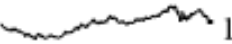
Current Receipts of Government as a Percentage of Gross Domestic Product, 1970 and 1979



Sparkline



- v Excelu jako *minigrafy*
- velmi malé grafy většinou bez os a legendy
- původně pouze spojnicové
- umožňují rychlou analýzu trendů
- doplňují tabulková data
- [zase Tufte](#)

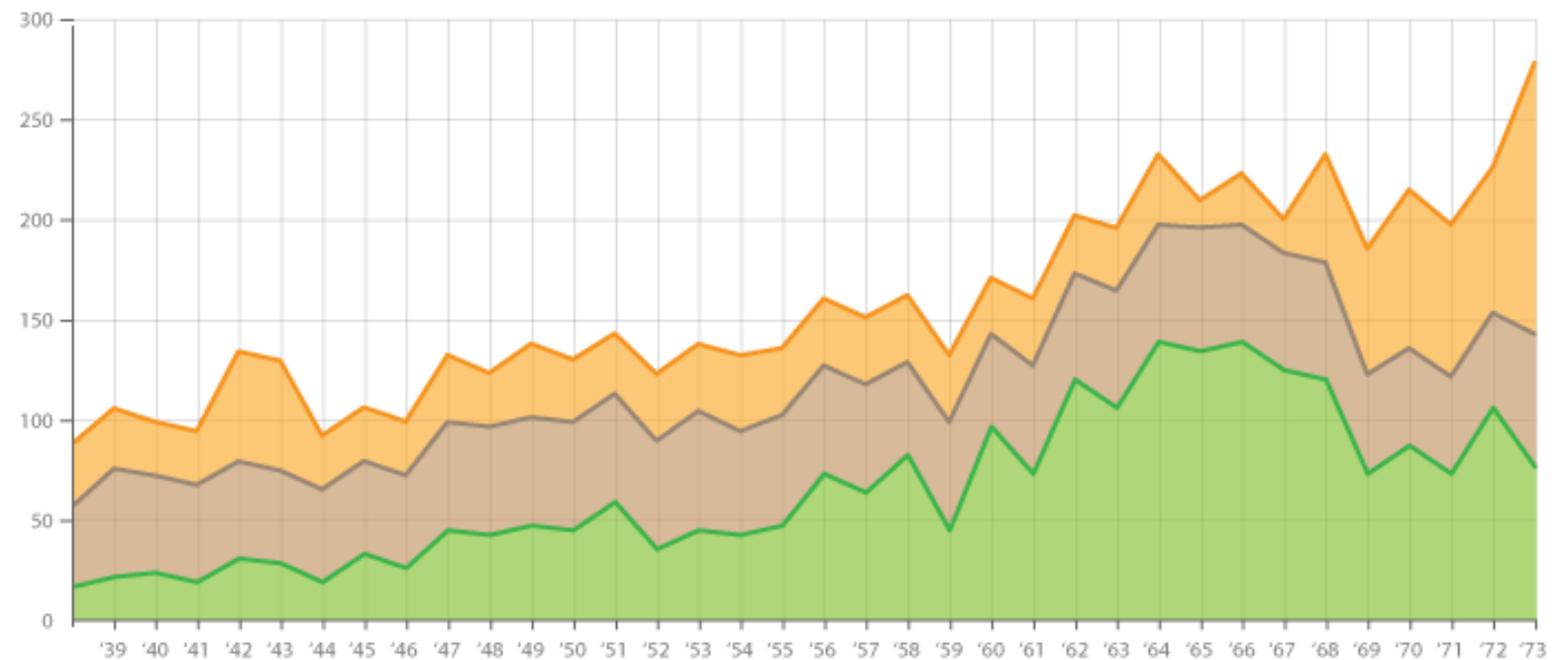
Iris, Tante Lisbeths Nichte, blickte von ihrer Zeitung auf. Sie hat ein gutes Zahlengedächtnis. „Als der Euro eingeführt wurde, sind wir bei knapp 90 Cent  gestartet“, sagte sie. Stimmt. Zur Einführung des Euro als Bargeld am 1.1.2002 musste man 0,88 Dollar für einen Euro bezahlen. Und nochmals drei Jahre früher, zur Einführung des Euro als Buchgeld am 01.01.1999, stand der Euro bei 1,17. Seitdem sieht die Entwicklung so aus: 1,17  1,23.

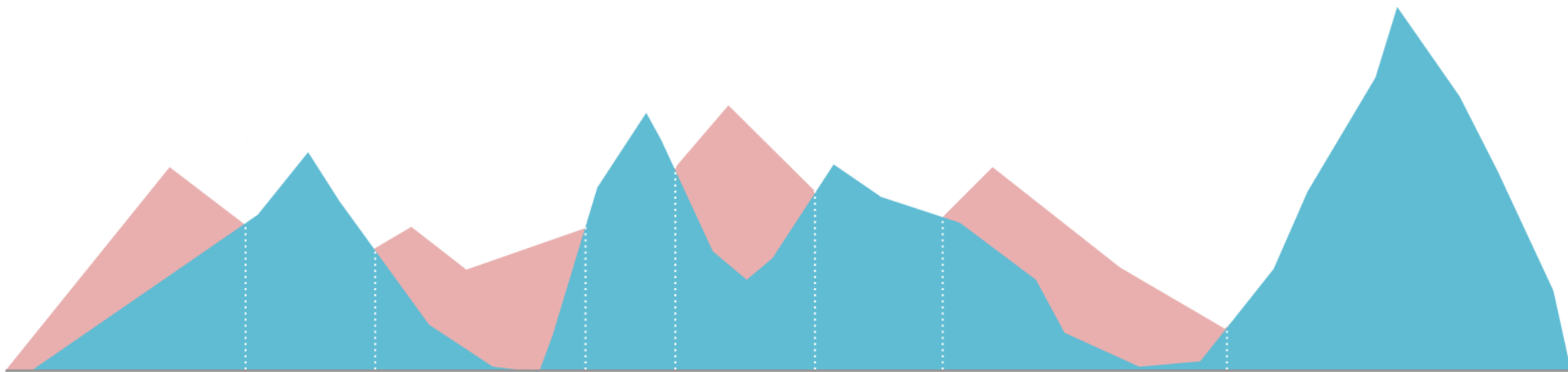
Tante Lisbeth staunte. Wenn der Euro immer mal wieder auf dem Stand von heute ankommt, mal darüber, mal darunter liegt, woher kommt dann die Hysterie? Geholfen bei unserer Betrachtung hat uns Datendichte: Statt

Plošný graf

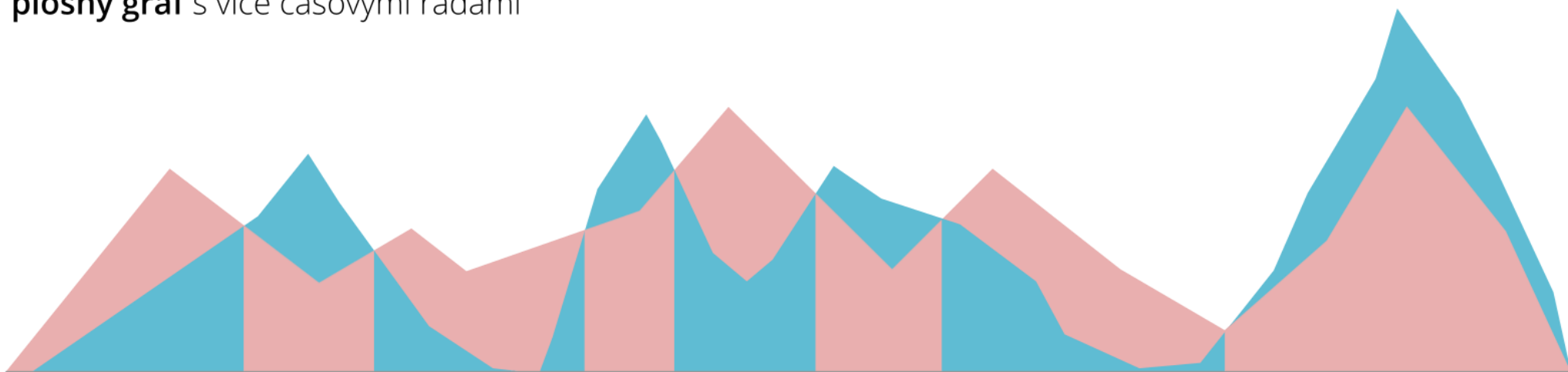


- *area chart*
- jako spojnicový
- skládaný plošný

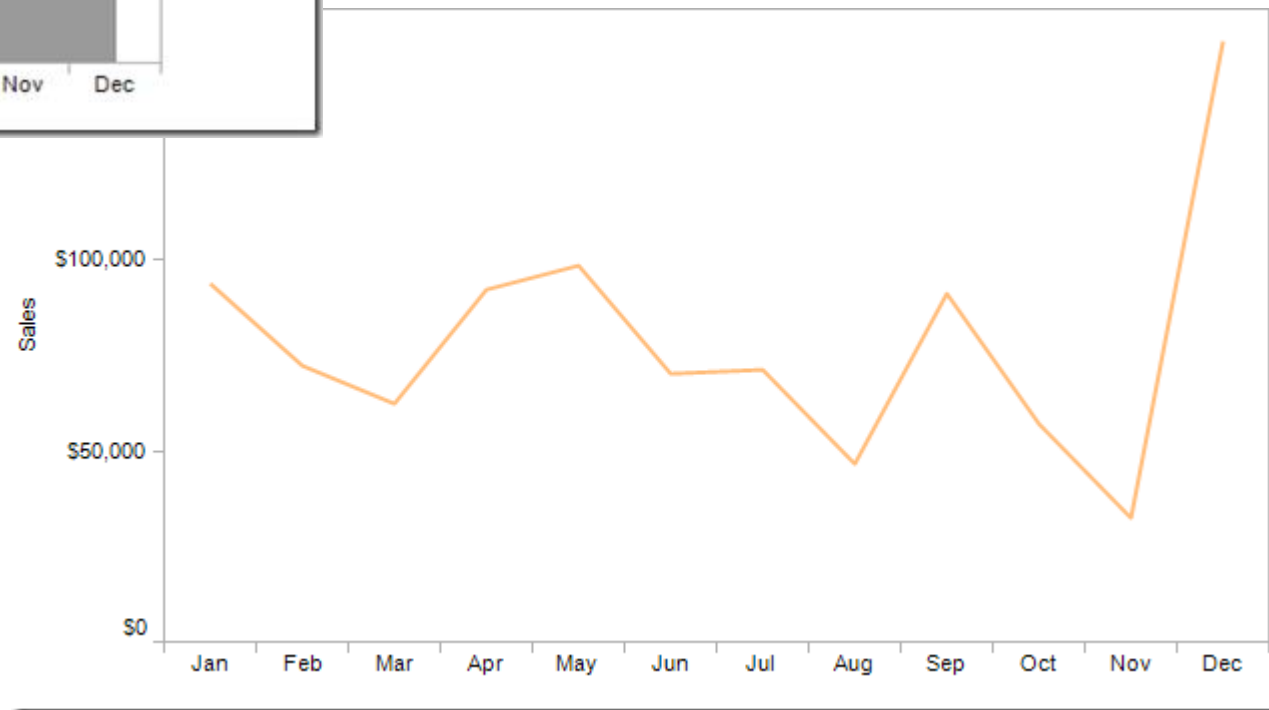
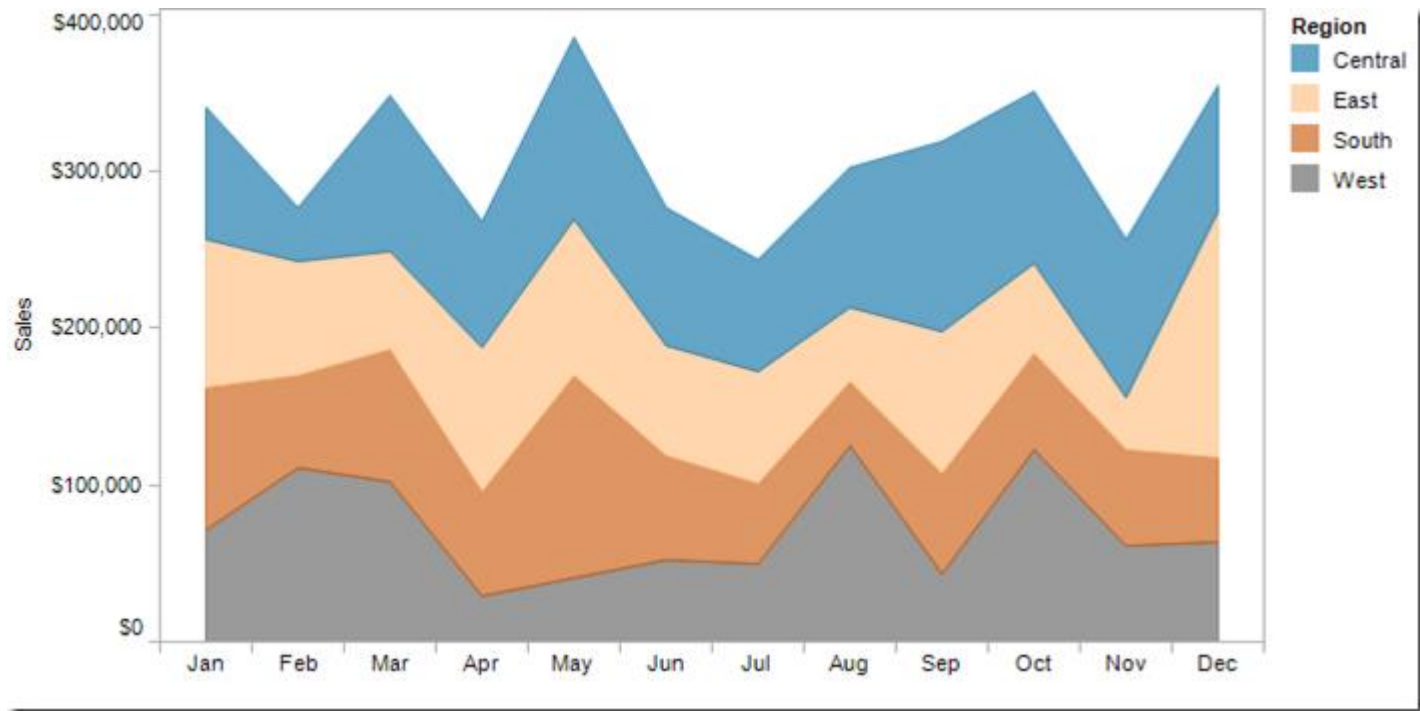




plošný graf s více časovými řadami

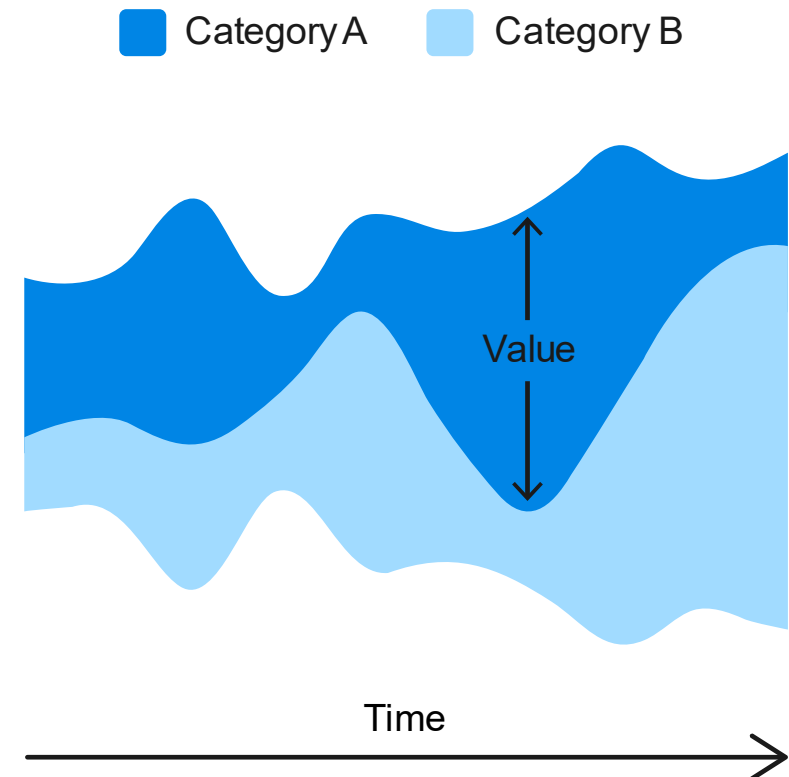


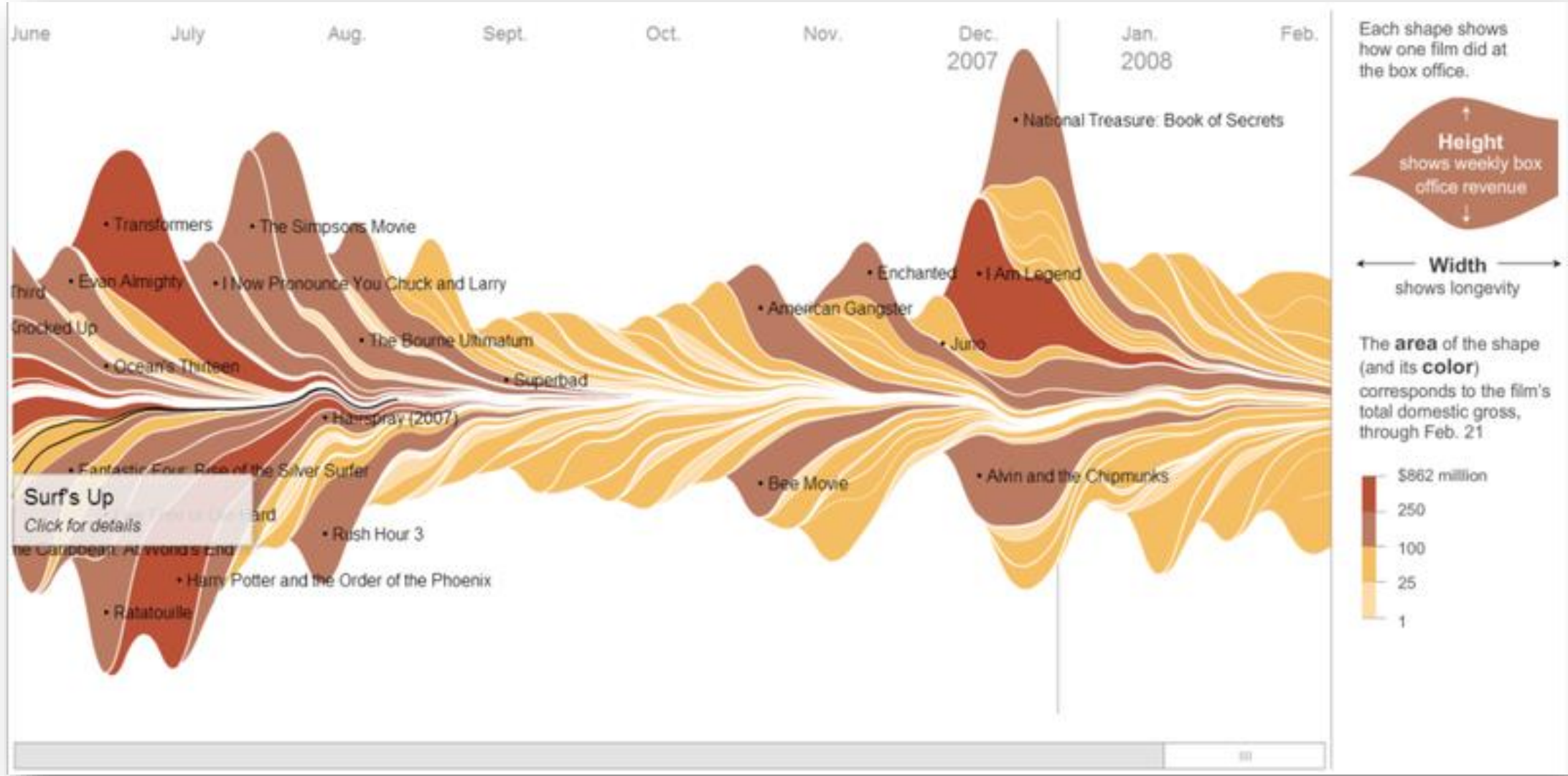
braided graf



Streamgraph

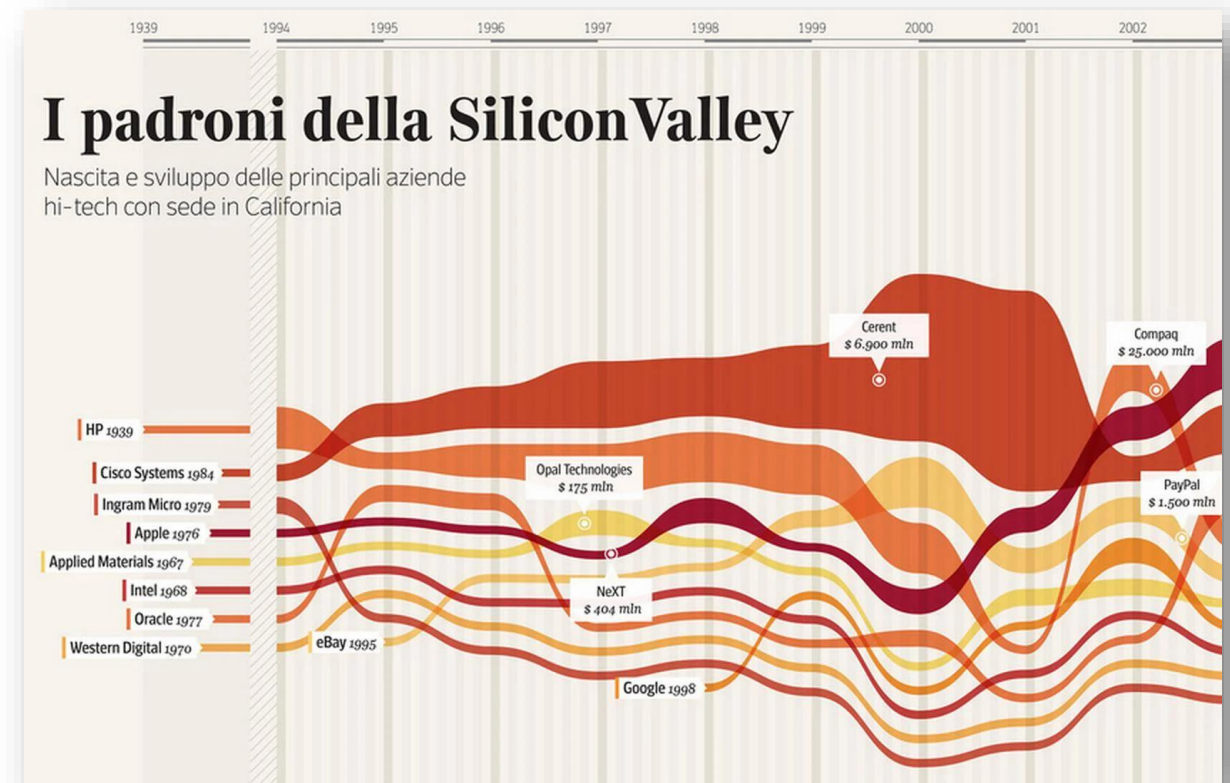
- variace skládaného plošného
- vhodný pro větší data
- rychle se zahltí
- spíše popularizační forma
- také jako [*theme river*](#)





Bumpchart

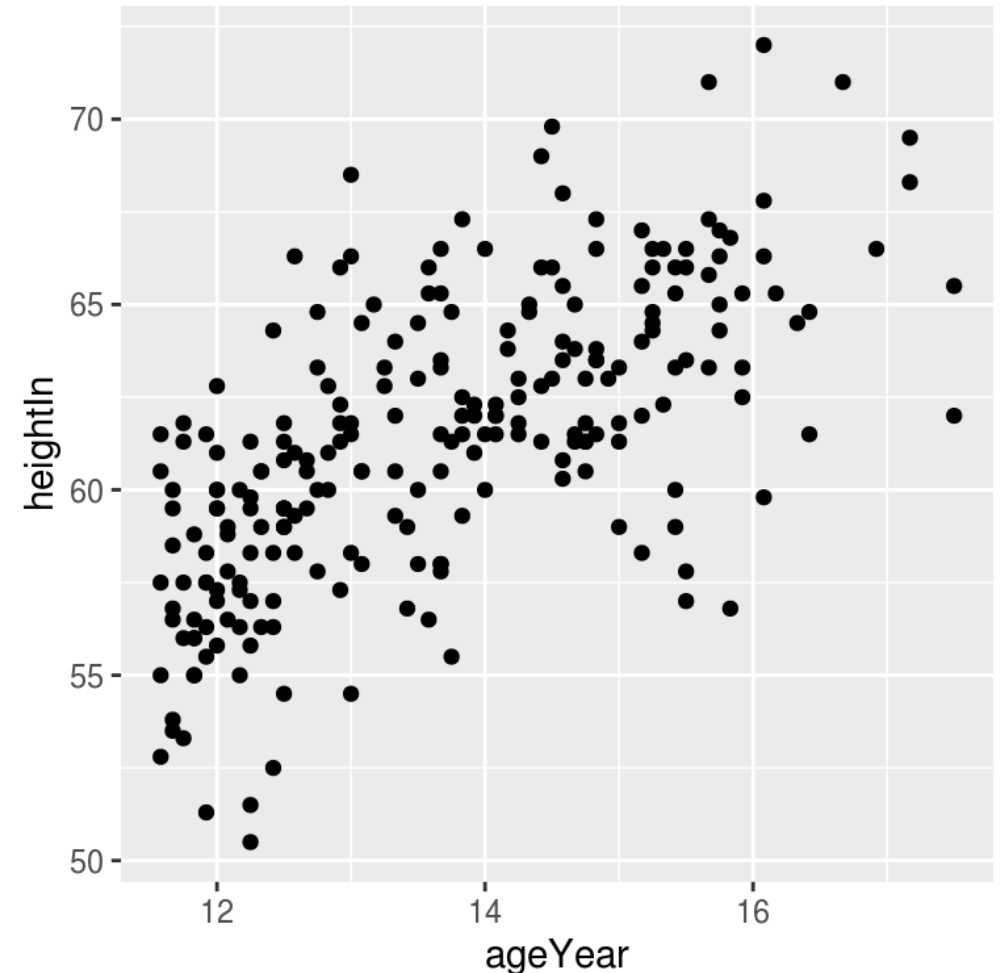
- varianta streamgraphu
- v každém řezu seřazeno
- žebříčky + kvantita

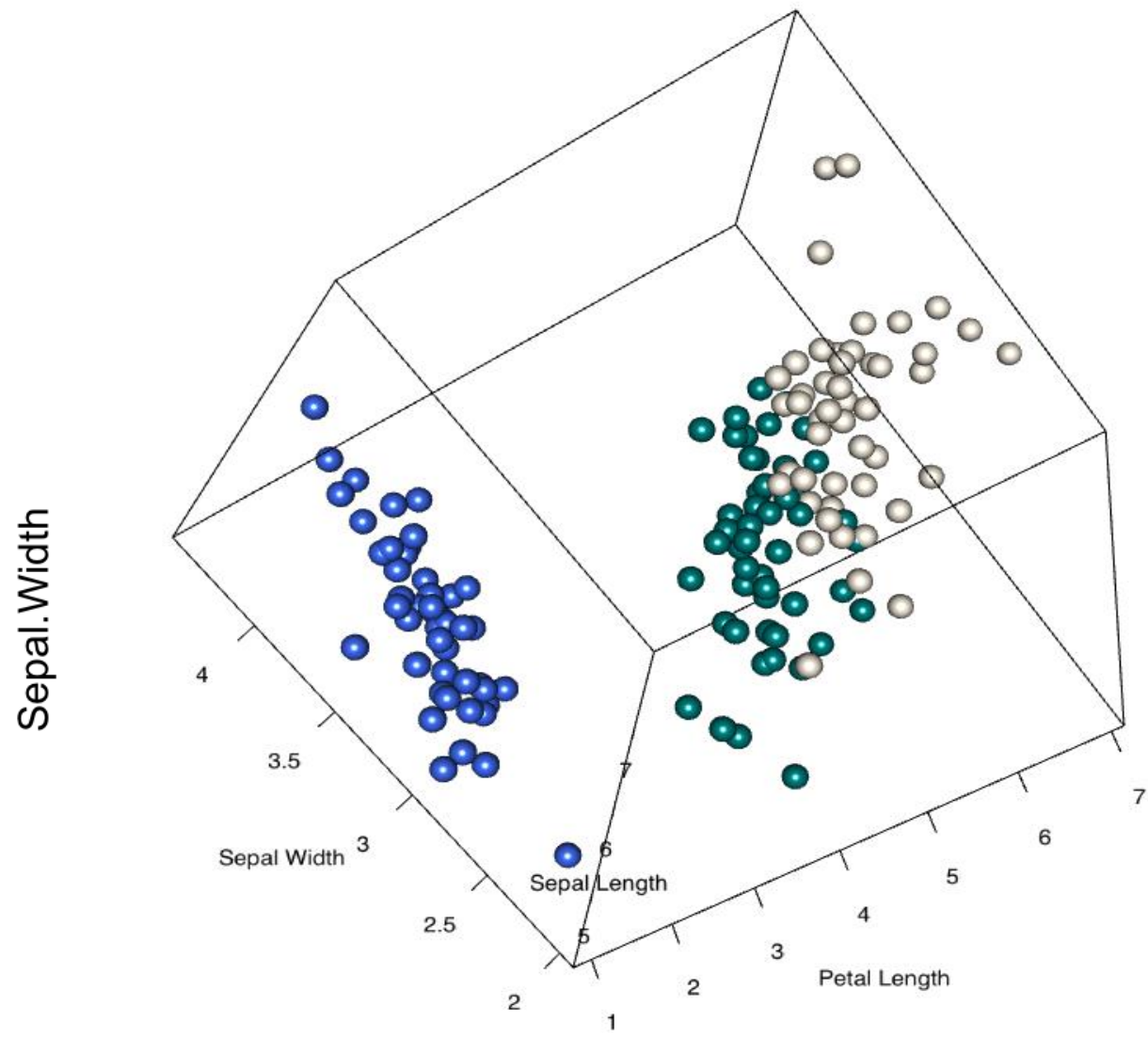
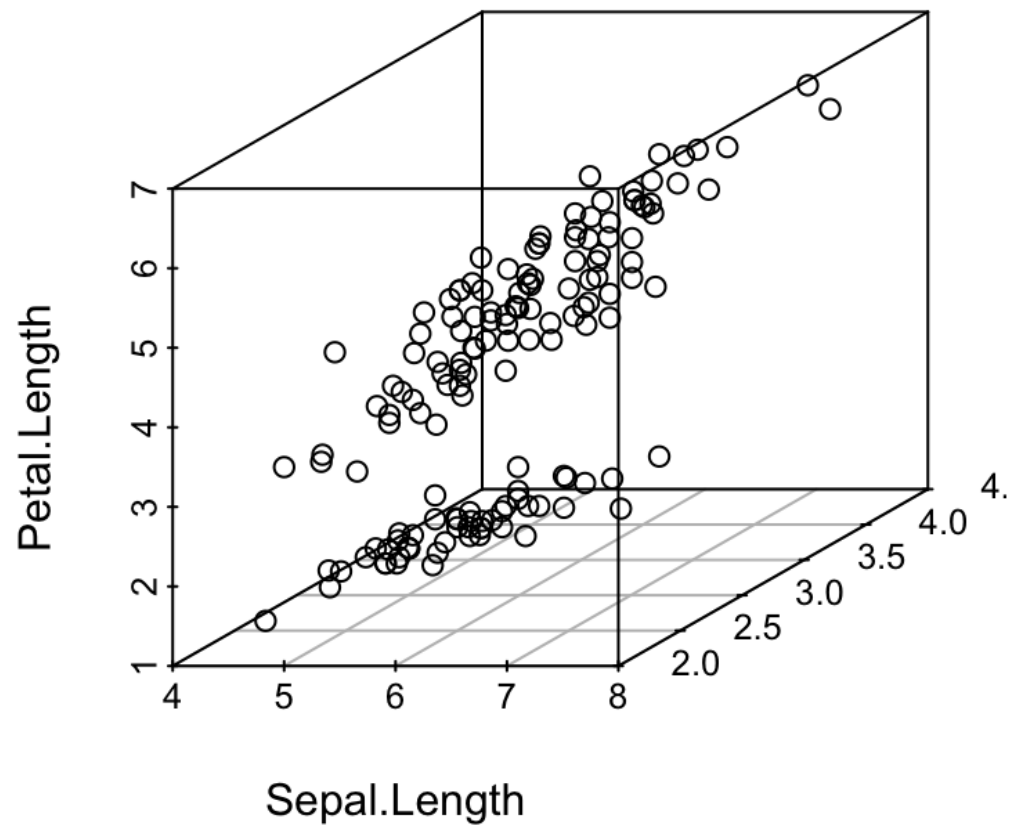


Korelační diagram

- scatterplot, bodový graf
- hodnoty dvou proměnných
- *podmíněné formátování?*
- v Excelu náročné
- *RAWGraphs*

+ velikost bodů = bubble chart

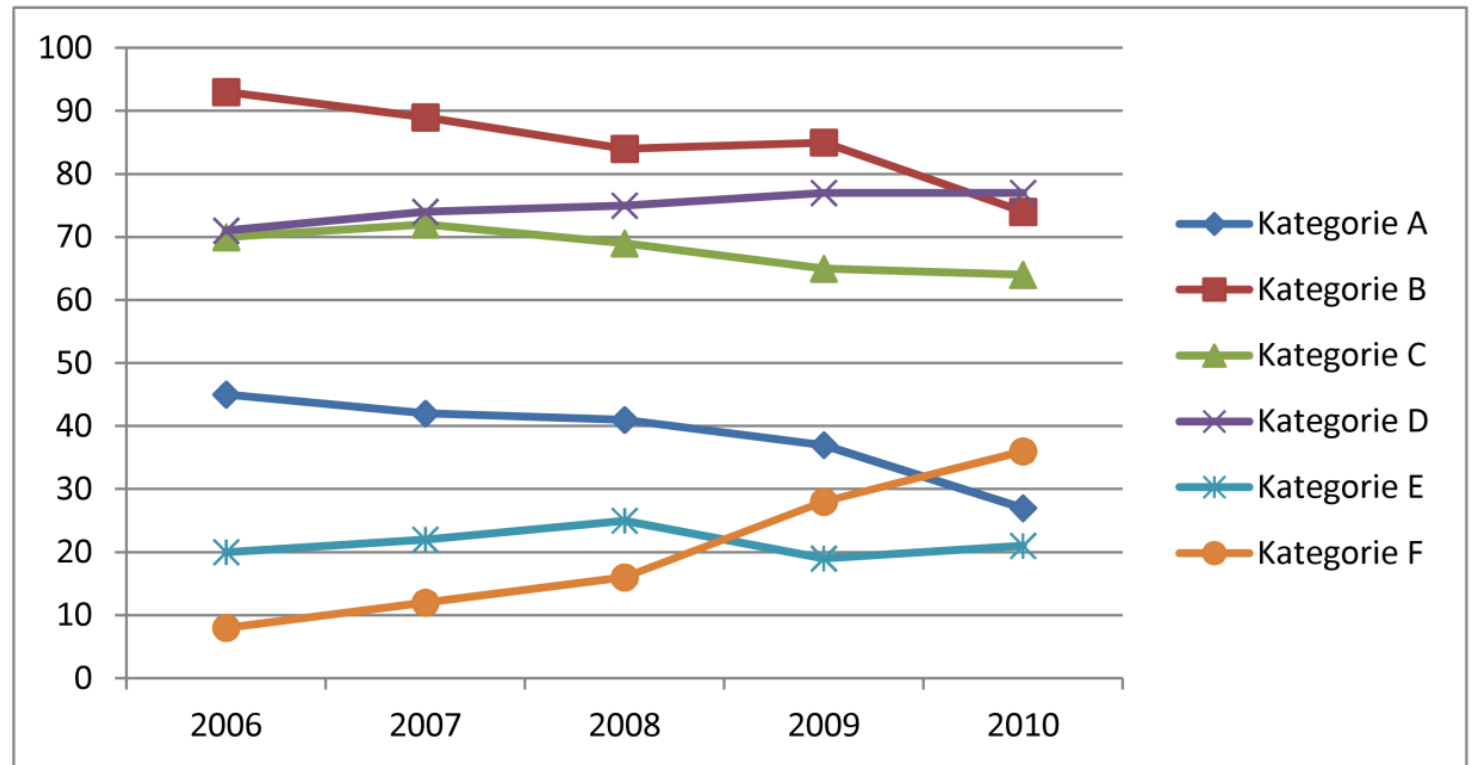


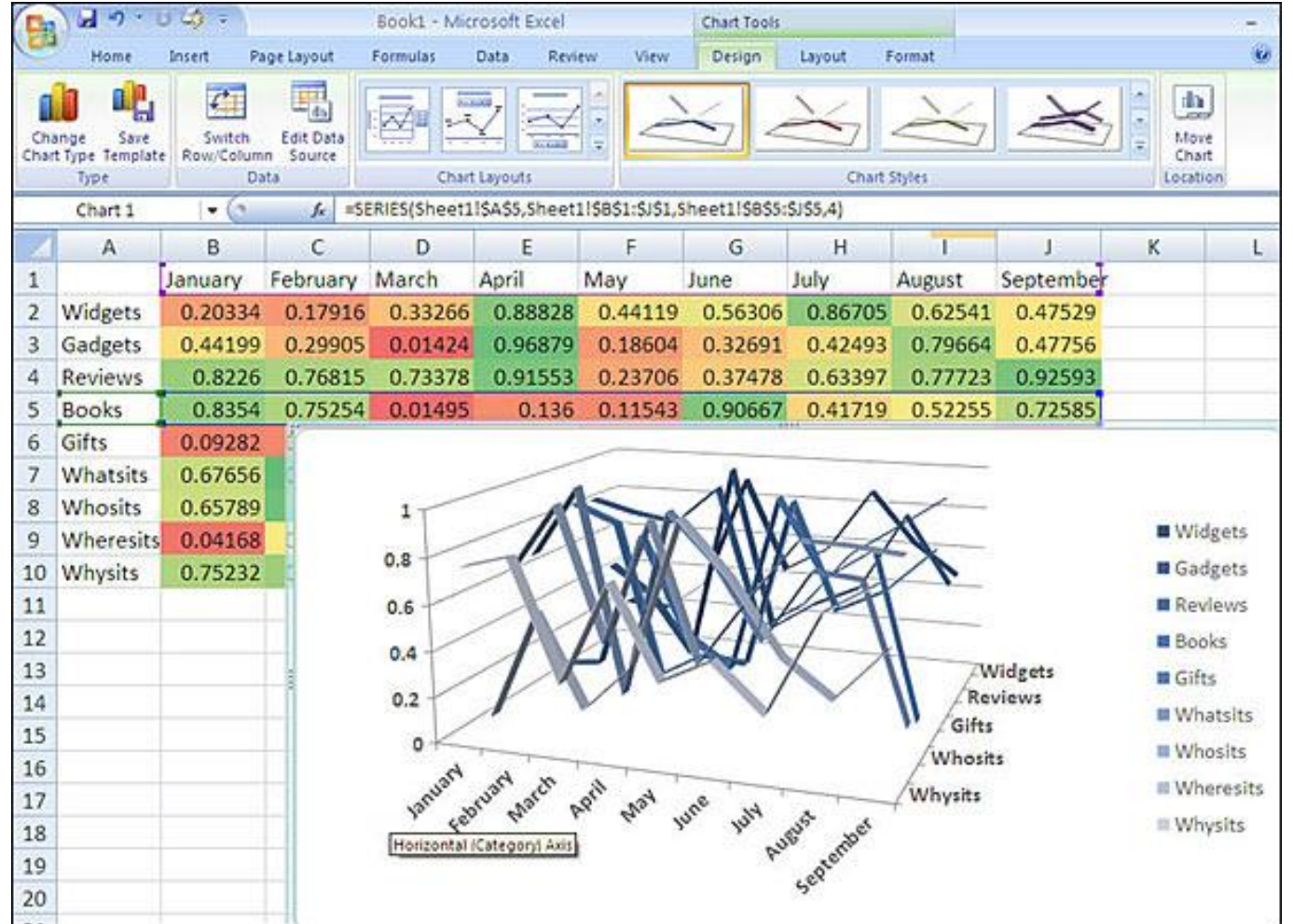
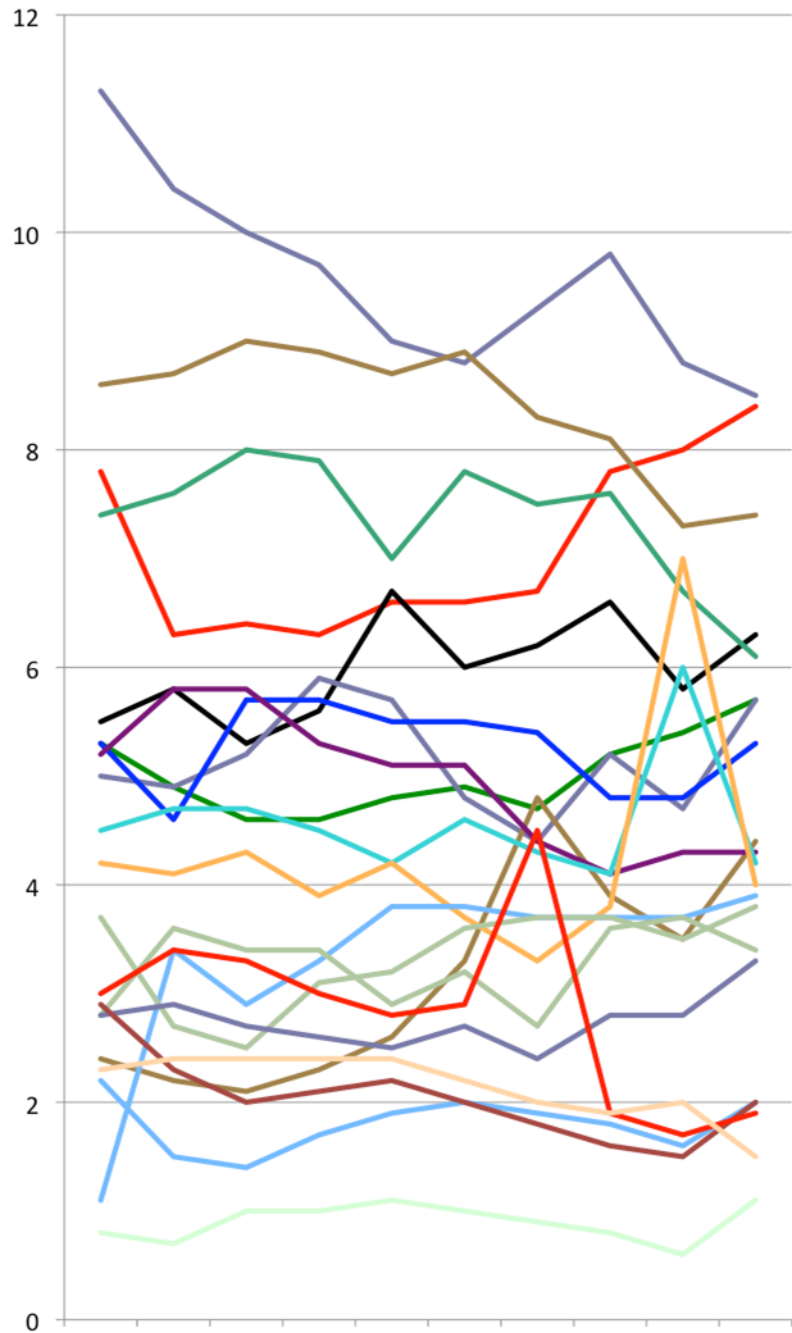


Potíže s mnoha daty

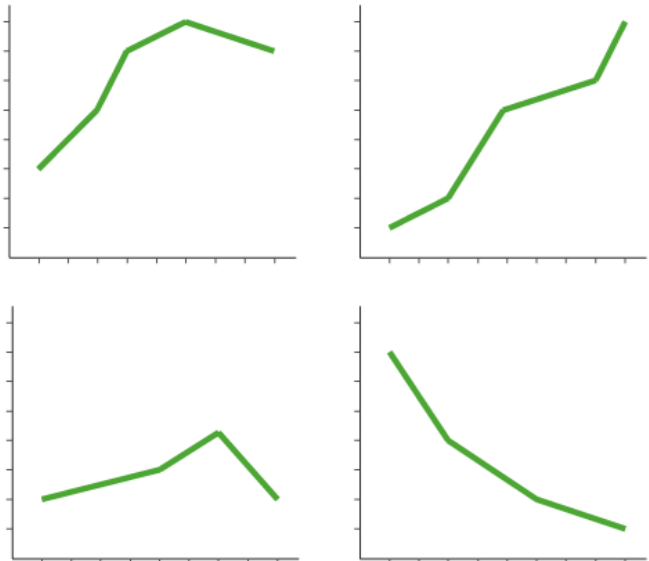
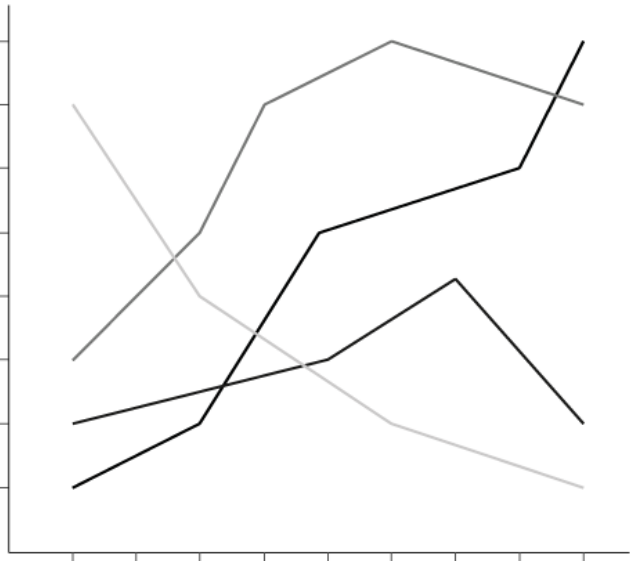
- máme mnoho dat, mnoho záznamů
 - tradiční formy na to nestačí
 - naráží na své limity
 - co dál?
-
- *small multiples*
 - specifické formy – horizon chart

	2006	2007	2008	2009	2010
Kategorie A	45	42	41	37	27
Kategorie B	93	89	84	85	74
Kategorie C	70	72	69	65	64
Kategorie D	71	74	75	77	77
Kategorie E	20	22	25	19	21
Kategorie F	8	12	16	28	36



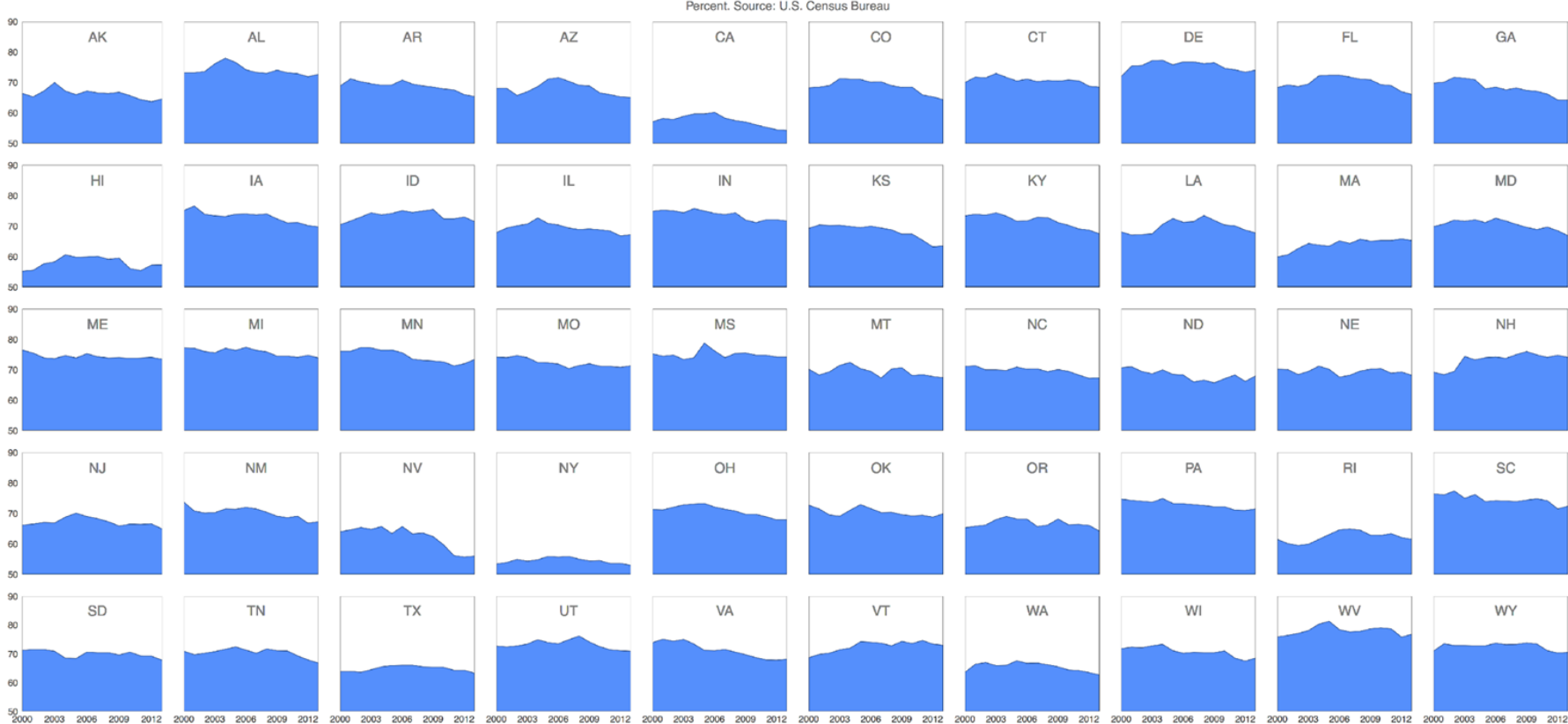


small multiples



small multiples

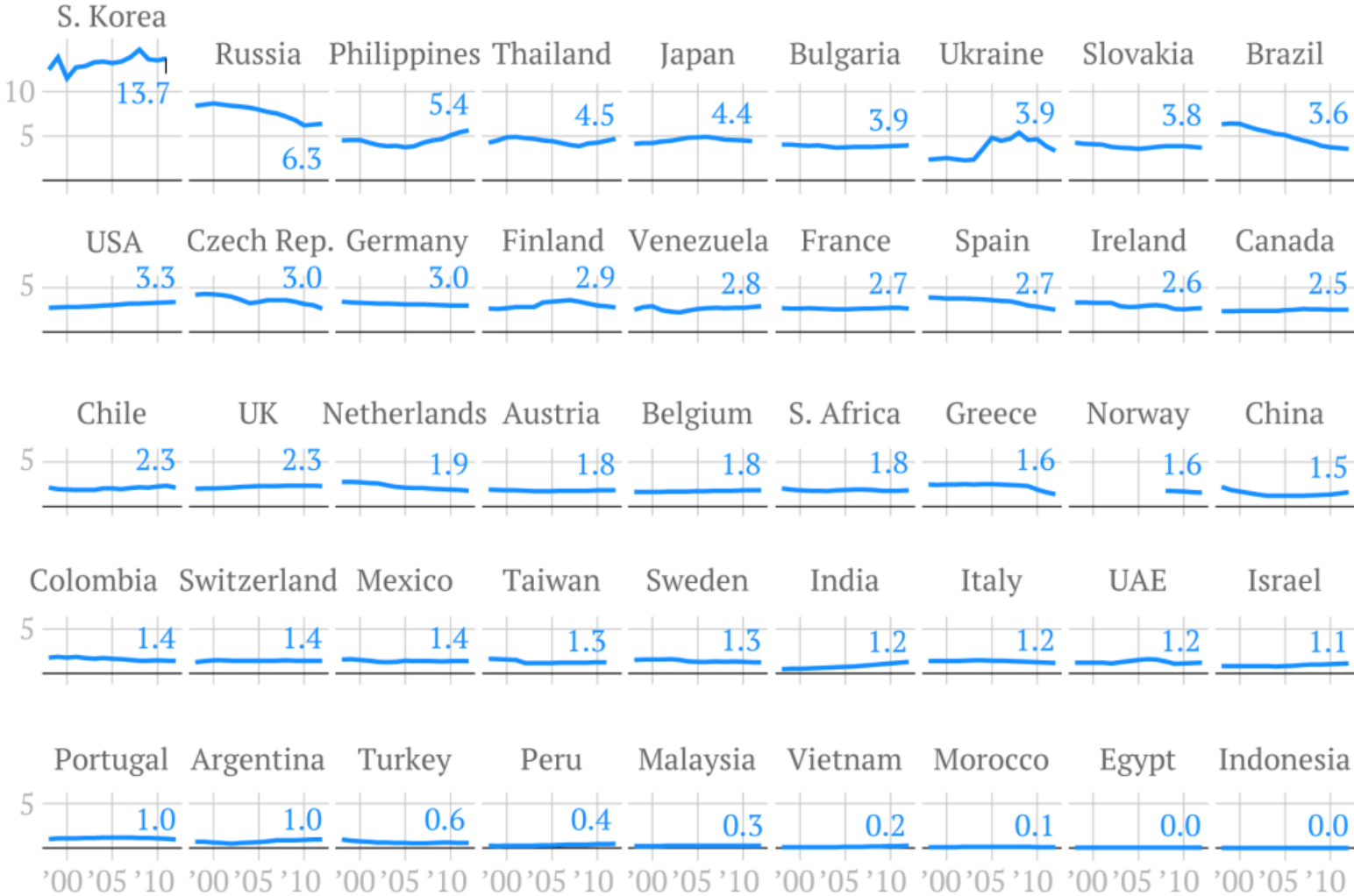
Evolution Of Homeownership Rates Across The United States, 2000–2013



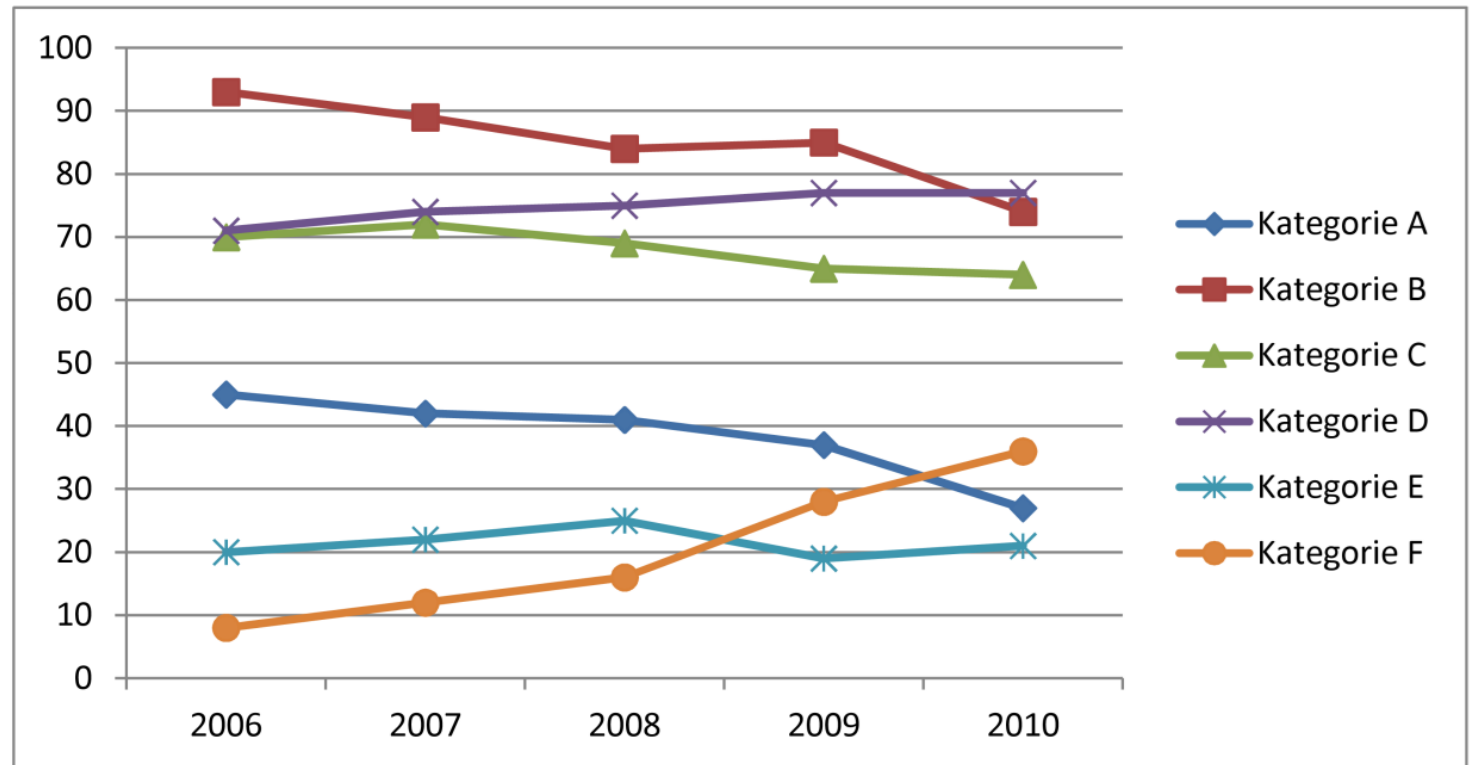
small multiples

The average amount of liquor consumed by a person of drinking age

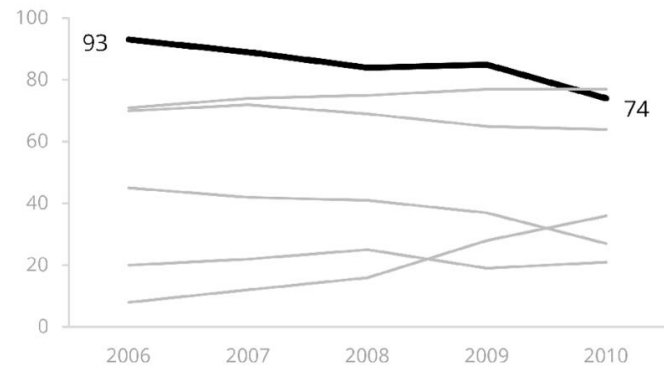
Shots per week of any spirit



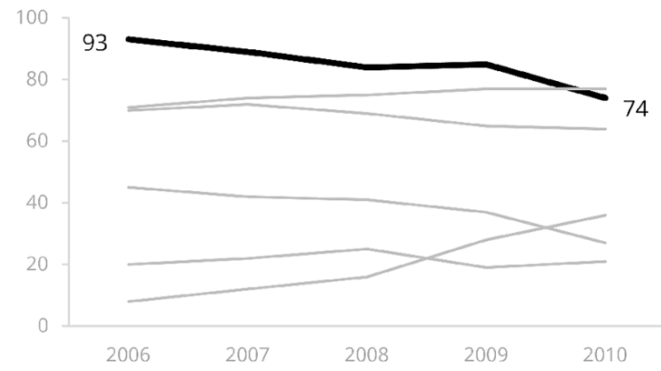
	2006	2007	2008	2009	2010
Kategorie A	45	42	41	37	27
Kategorie B	93	89	84	85	74
Kategorie C	70	72	69	65	64
Kategorie D	71	74	75	77	77
Kategorie E	20	22	25	19	21
Kategorie F	8	12	16	28	36



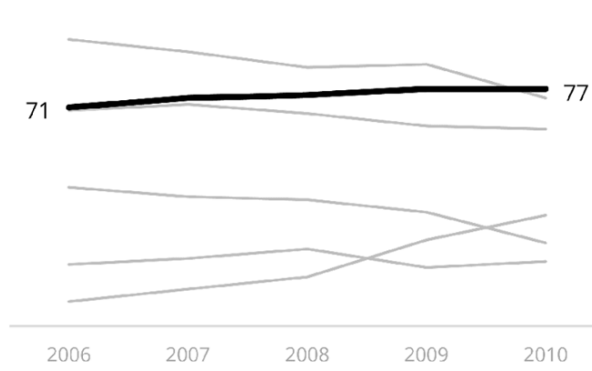
Varianta B



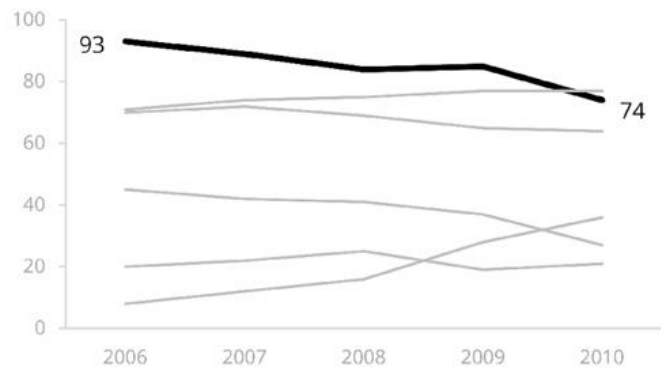
Varianta B



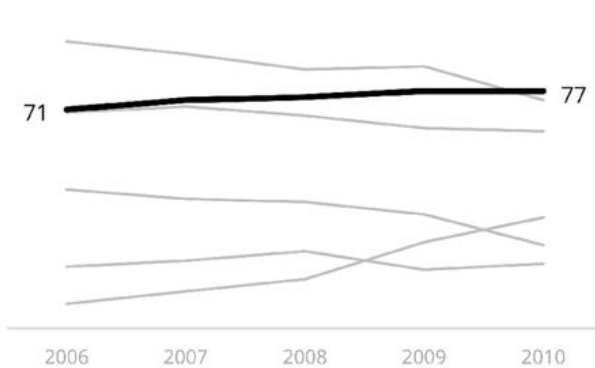
Varianta D



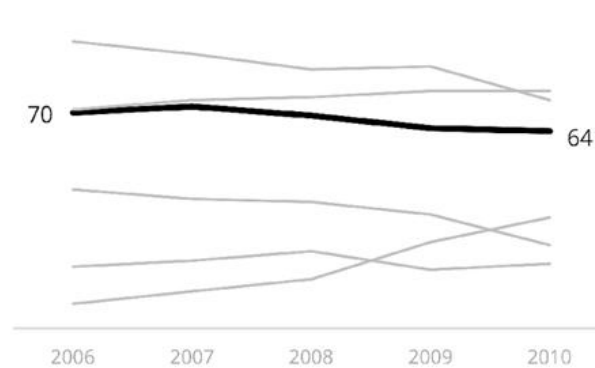
Varianta B



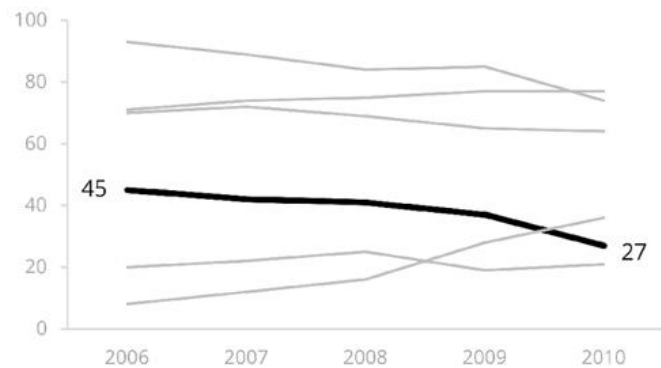
Varianta D



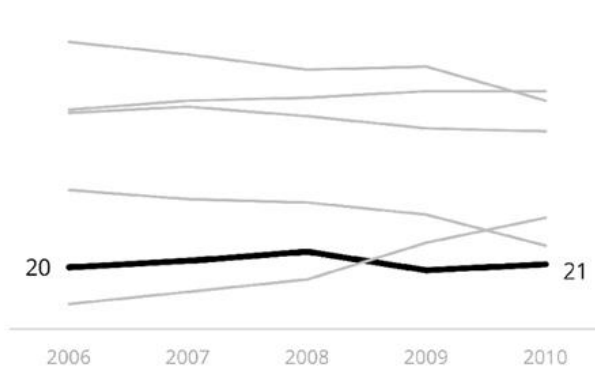
Varianta C



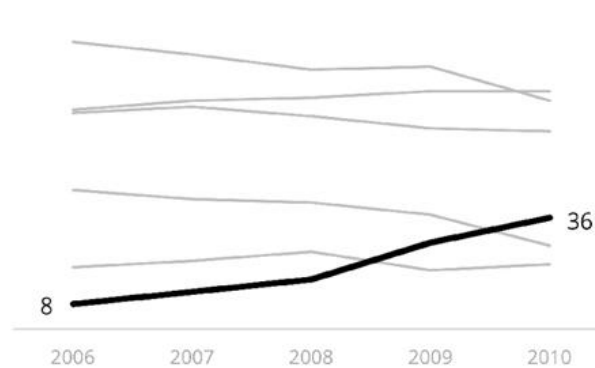
Varianta A



Varianta E

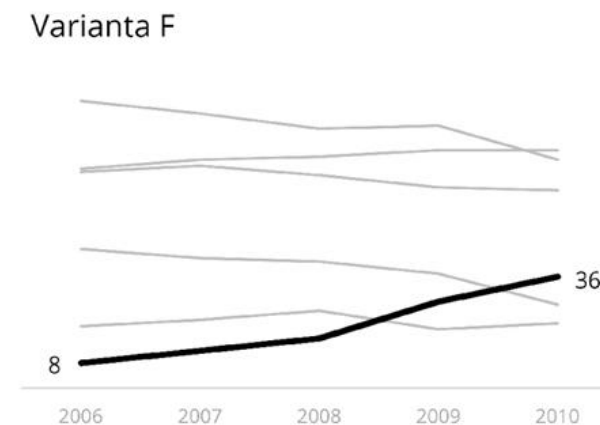
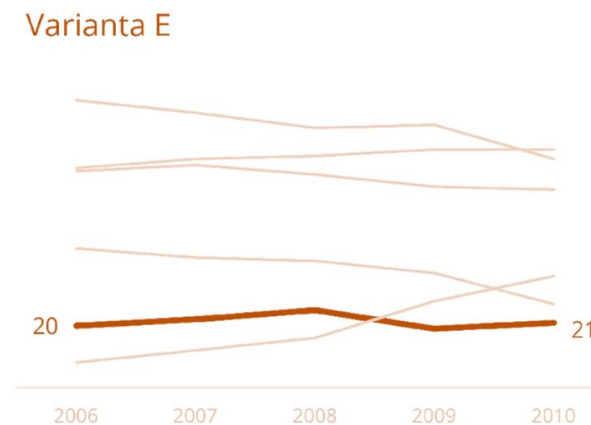
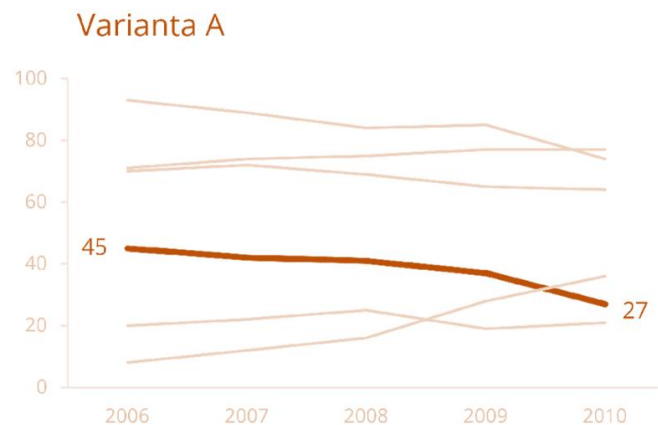
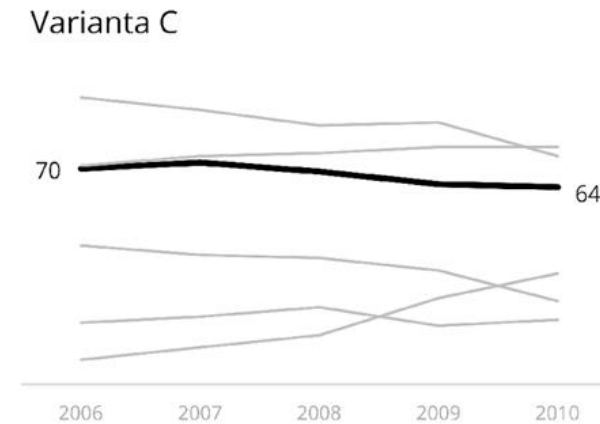
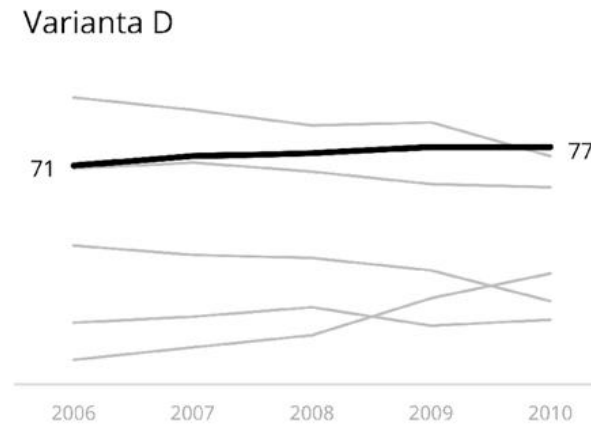
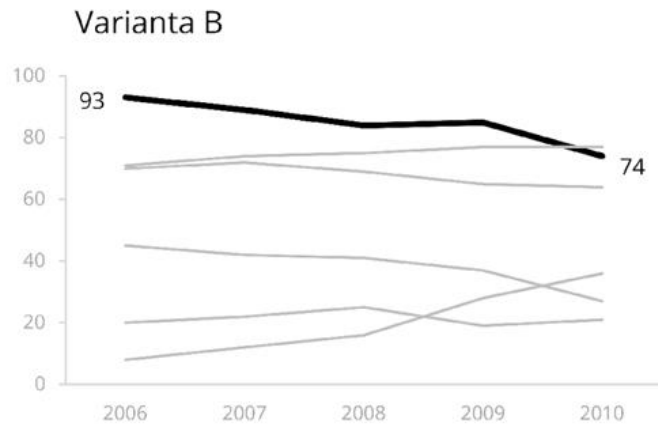


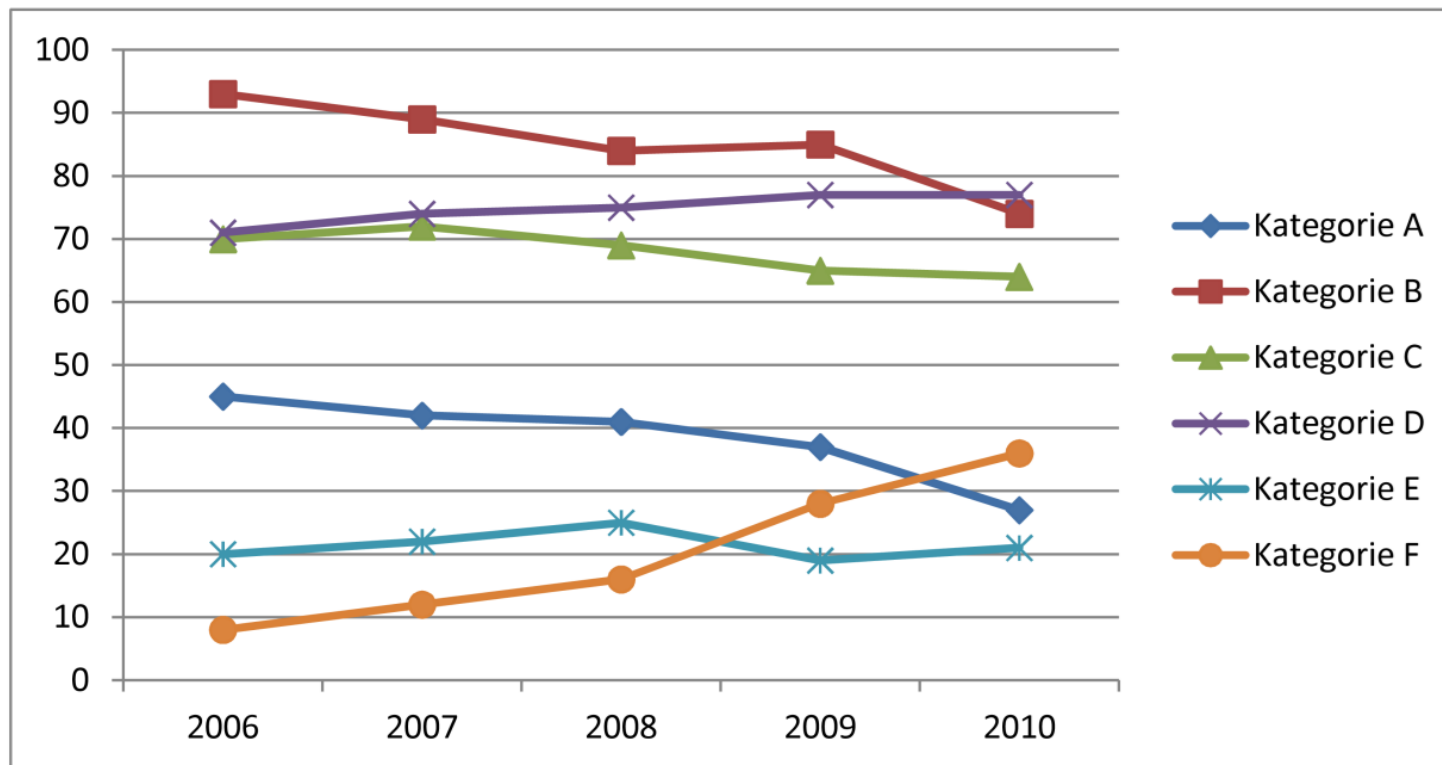
Varianta F



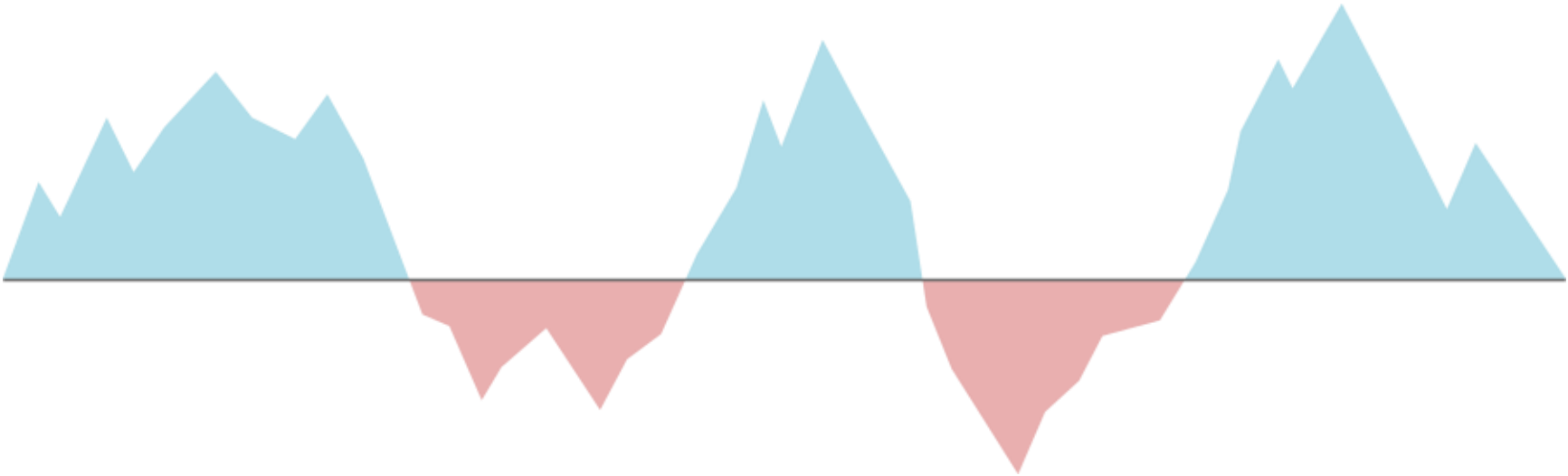
Varianty A a E nejsou ekonomicky perspektivní

Zatímco varianta F, kterou jsme po spuštění výroby považovali za neperspektivní, na trhu nalézá odbyt, výrobu dlouhodobě průměrných variant A a E doporučujeme zastavit.

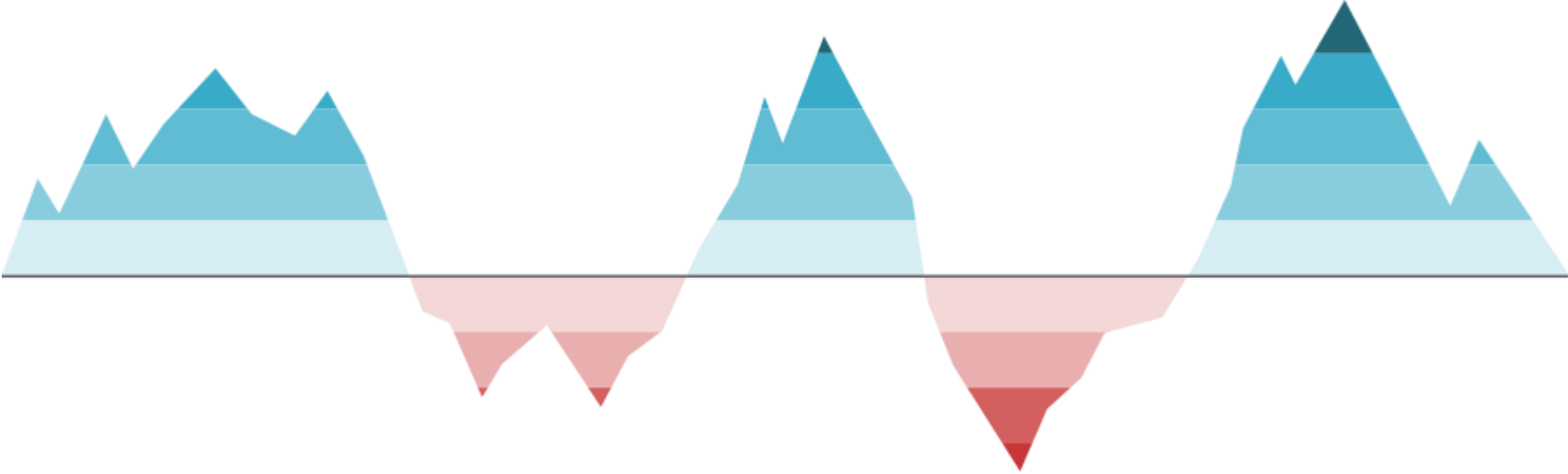




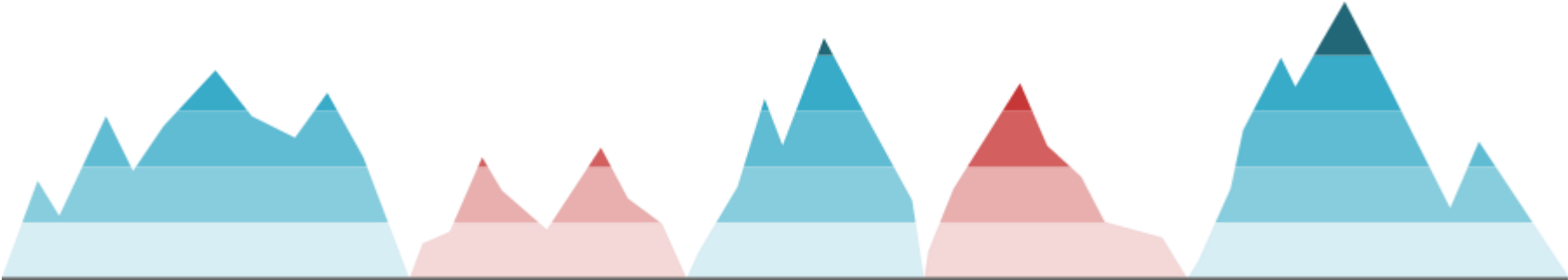
horizon chart



horizon chart



horizon chart



horizon chart



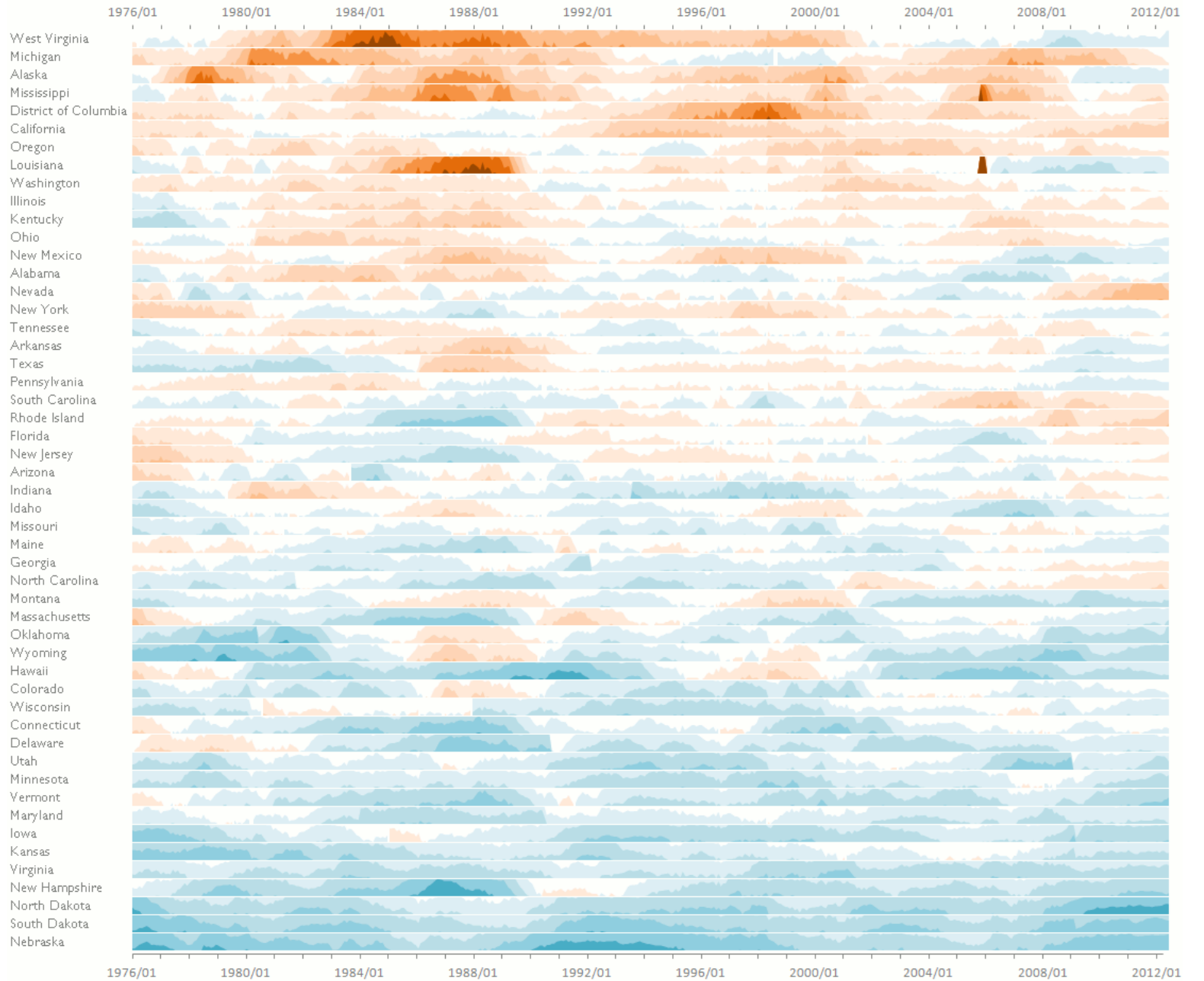
zrcadlené negativní hodnoty



offset negativní hodnoty

horizon chart

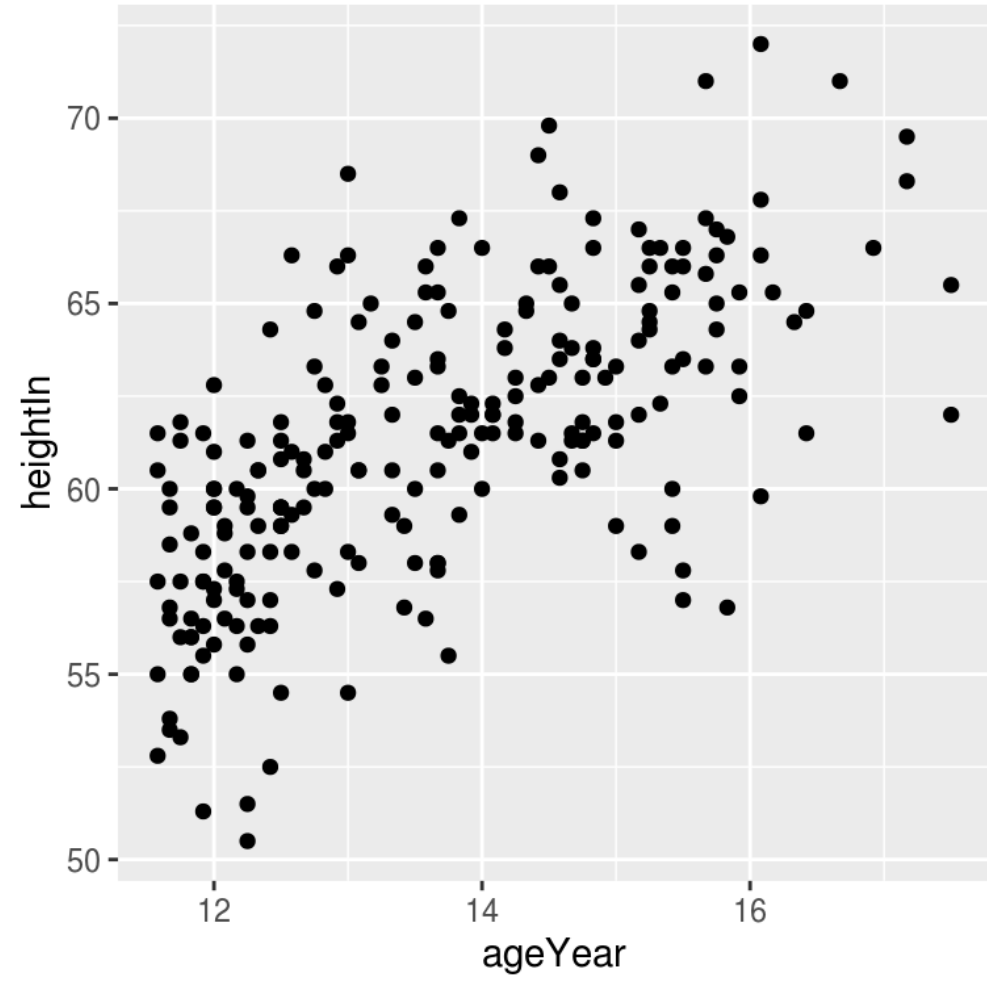
Unemployment Rate: Differences to the National Average



Vizualizace vícerozměrných dat

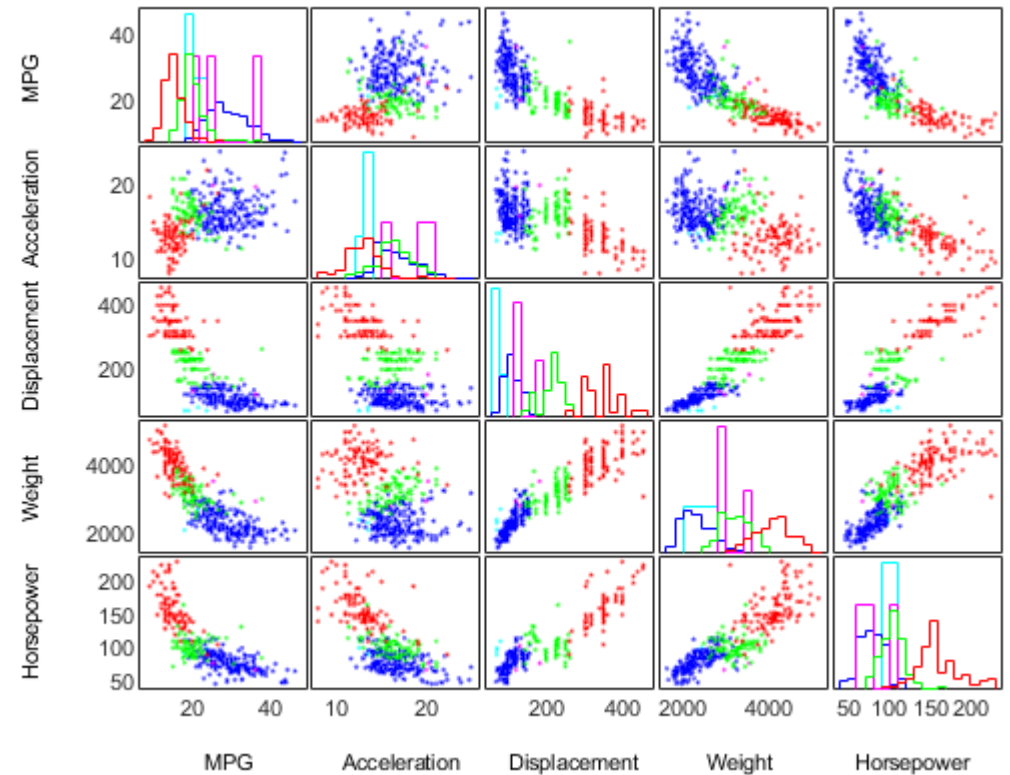
- vizualizace vícerozměrných dat je velmi náročná
- většina metod je cenná, ale mají své nevýhody

- *radar chart*
- *rovnoběžné souřadnice*
- *chernoff faces*

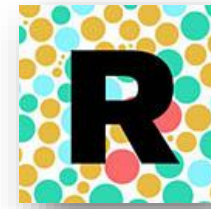


Vícerozměrný korelační diagram

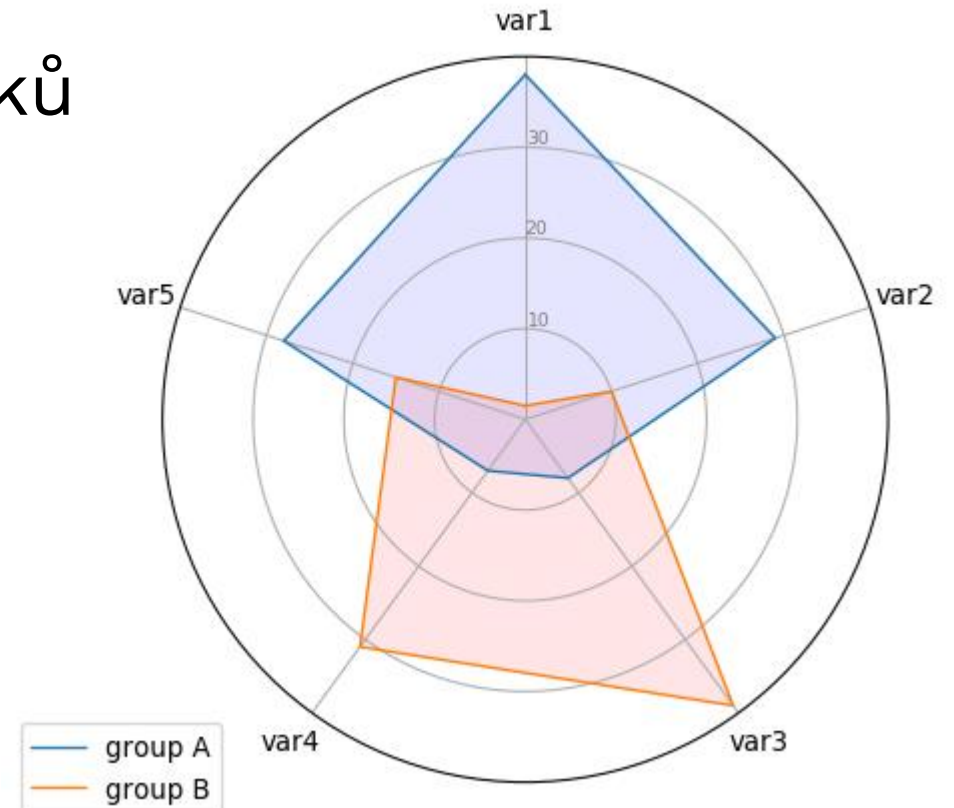
- *multivariate scatter plot*
- matice korelačních diagramů
- čím více proměnných,
tím více řádků



Radar chart

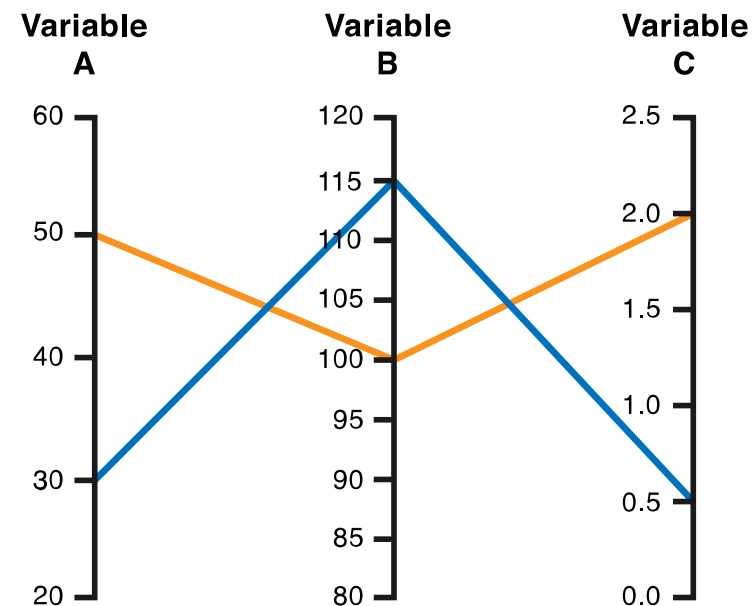


- relativně těžko se čte
- počet proměnných = počet paprsků



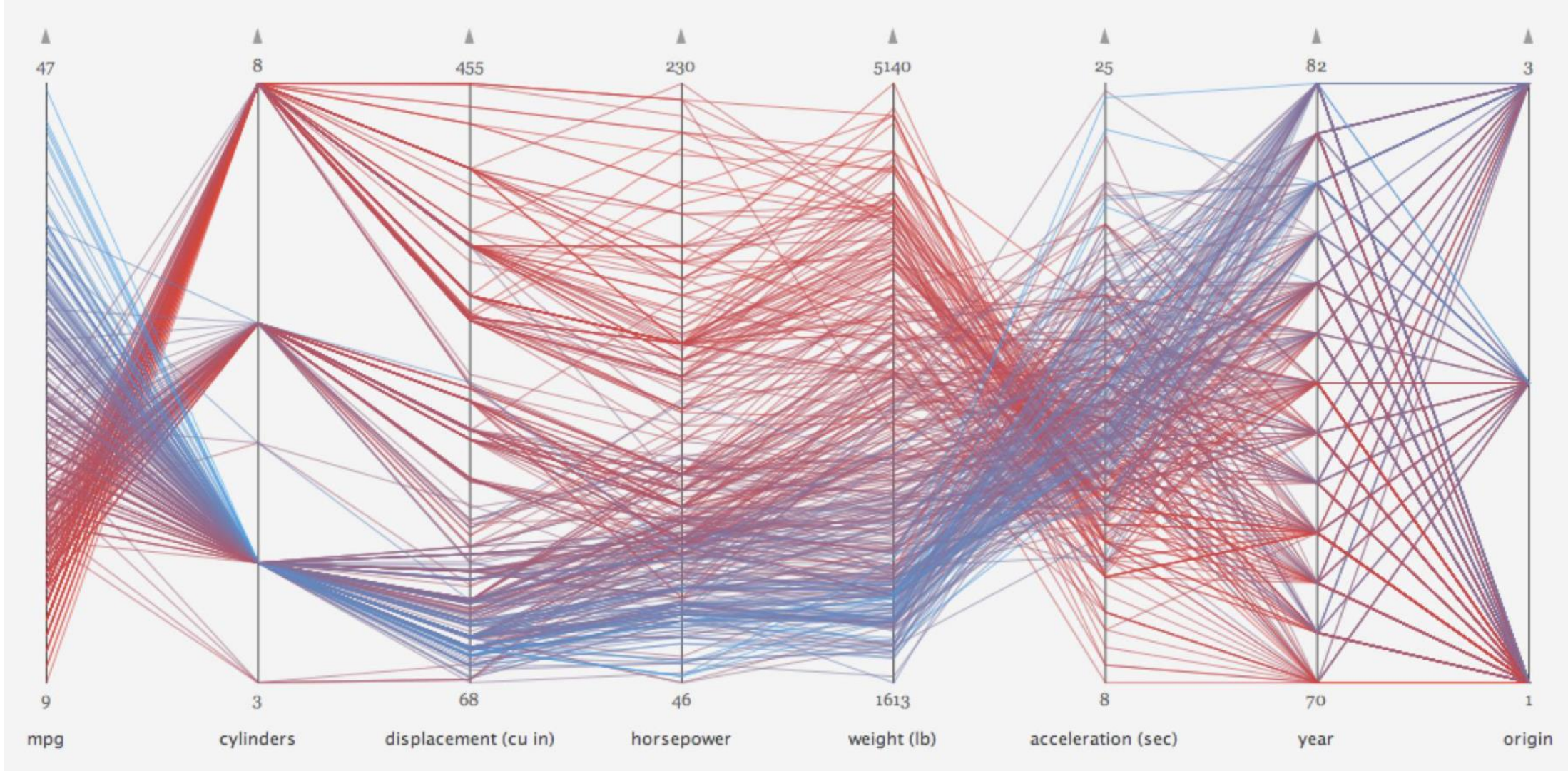
Parallel coordinates

- *rovnoběžné souřadnice*
- vícerozměrná data
- jedna položka s více proměnnými
- dobré k odhalení vztahů mezi p
- obtížné čtení



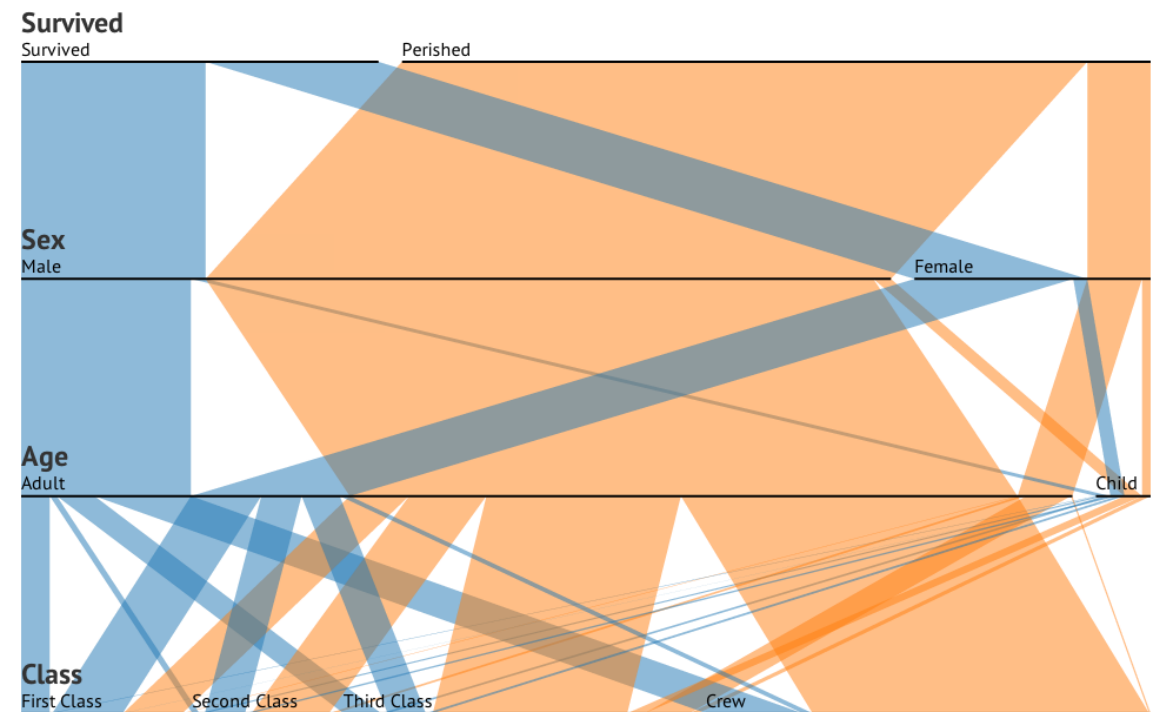
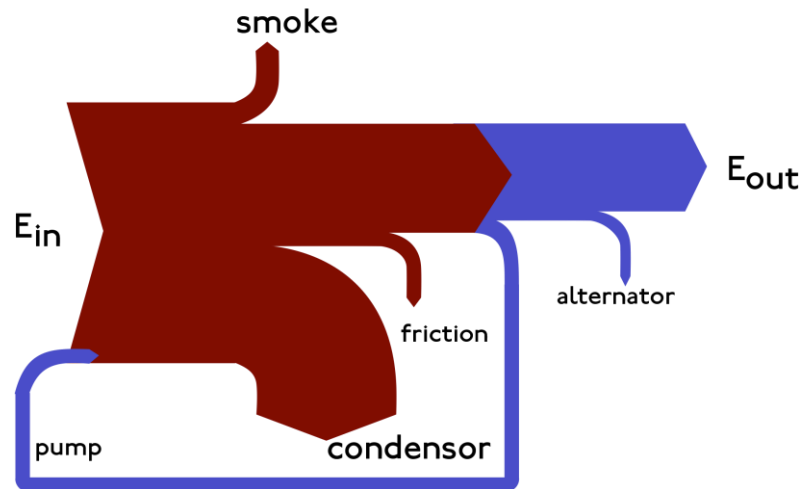
Data			
	Variable A	Variable B	Variable C
Item 1	50	100	2.0
Item 2	30	115	0.5

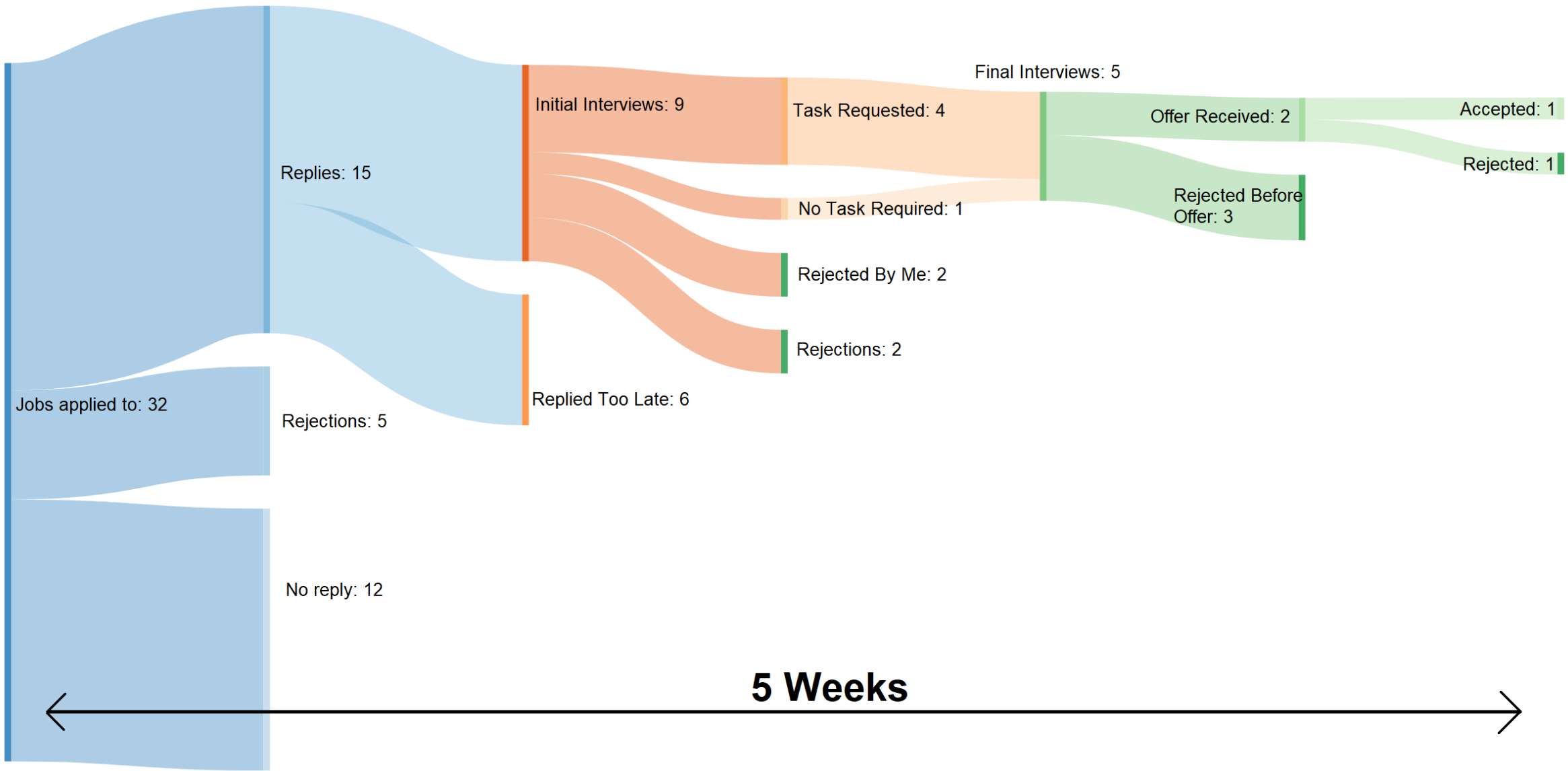




Parallel sets

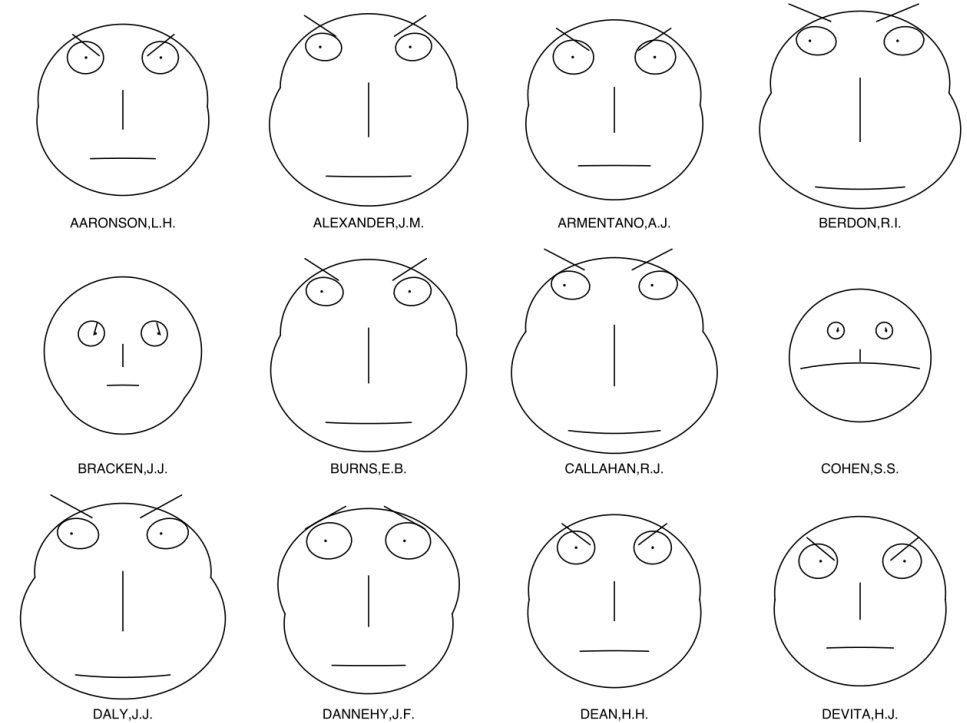
- pro kategorická data, Robert Kosara
- <https://eagereyes.org/parallel-sets>
- alluvial diagrams (RAW)
- *sankey*





Chernoffovy tváře

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



Chernoffovy tváře

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



Chernoffovy tváře

„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 *dobrá* v jistých ohledech
- [musí to být jen obličej...](#)?

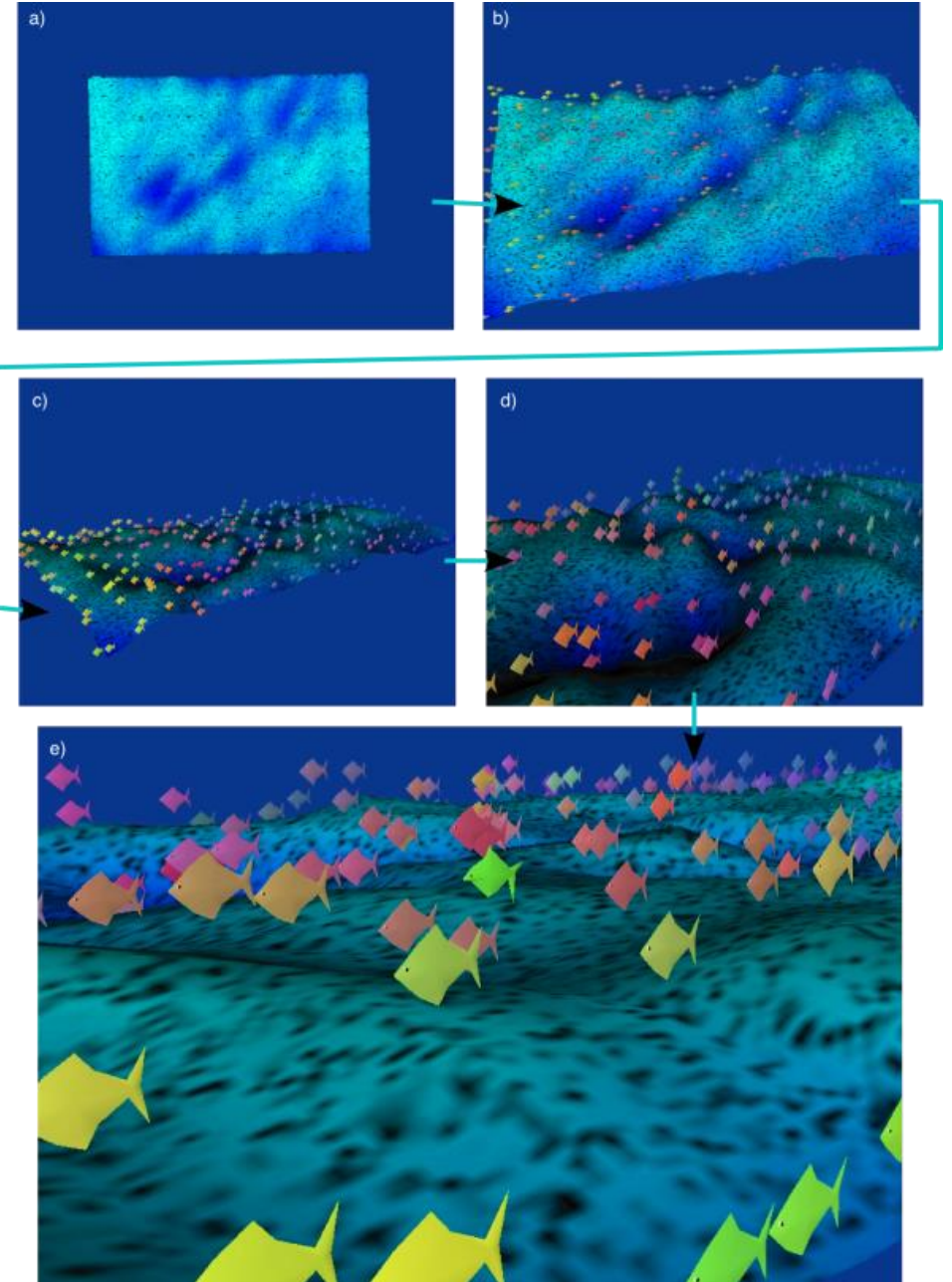
Kompozitní glyfy

- enviromentální metaforý
- [Leaf Glyphs](#)



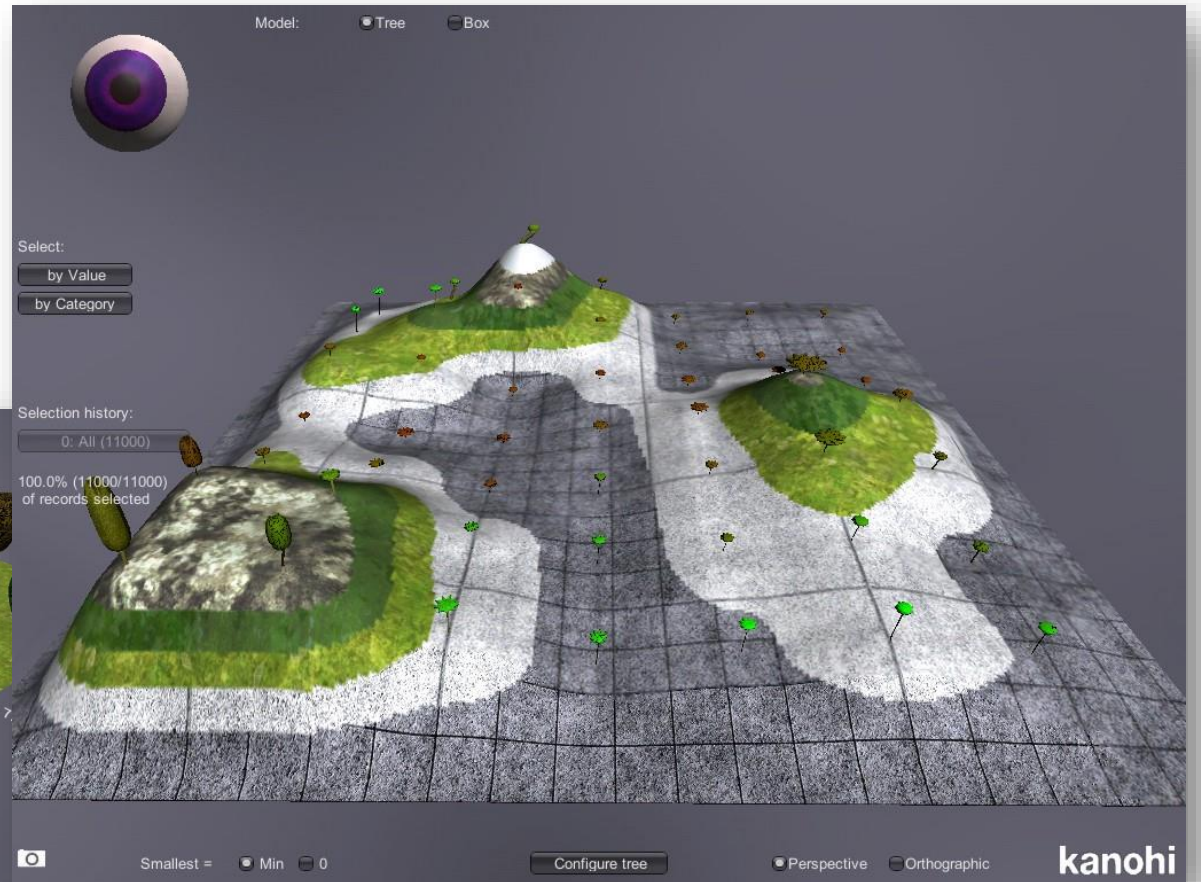
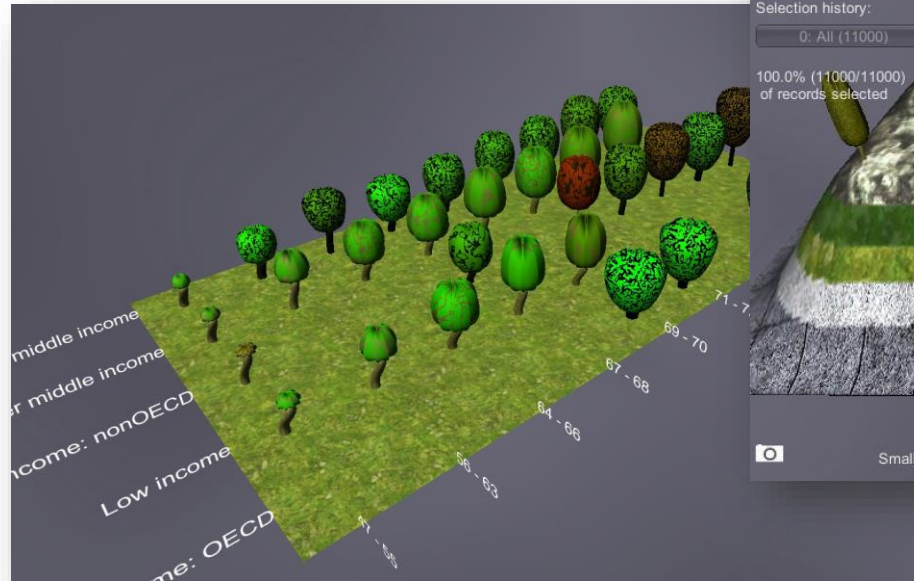
Metaforické 3D glyfy

- přidejme ještě jednu dimenzi
- proč ne, že...?
- [REEFSOME](#)



Metaforické 3D glyfy

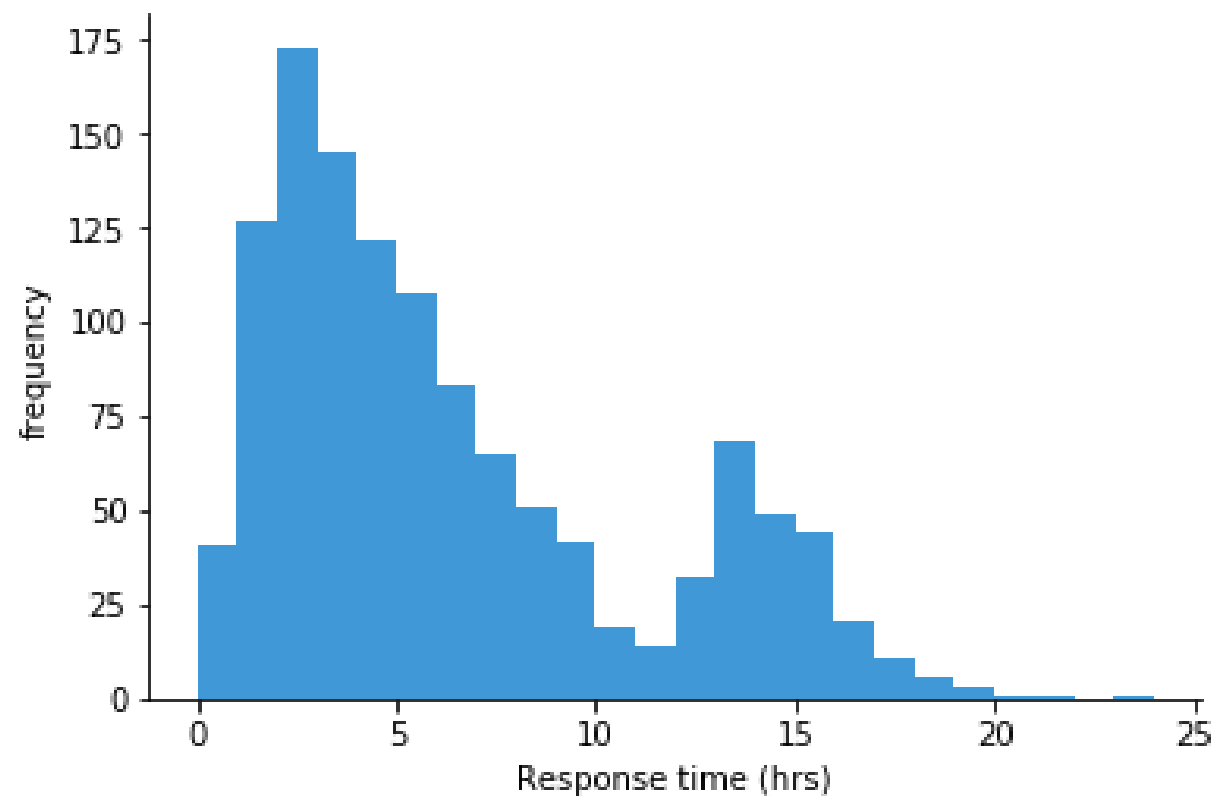
[KanoHi 3D data exploration](#)



Histogram

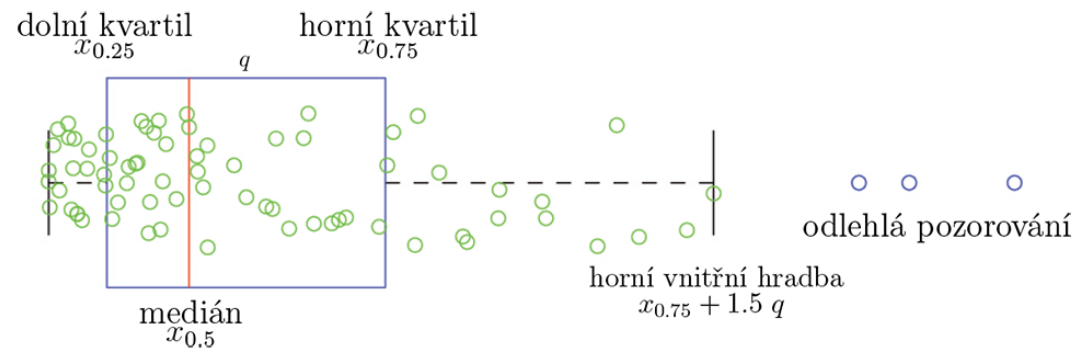
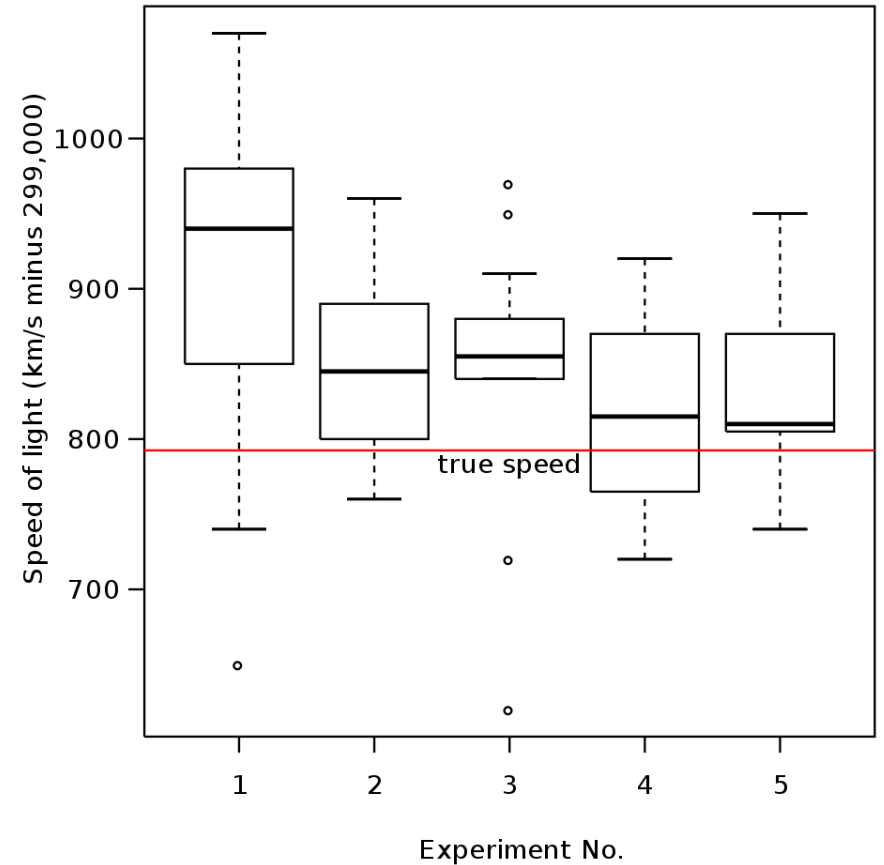


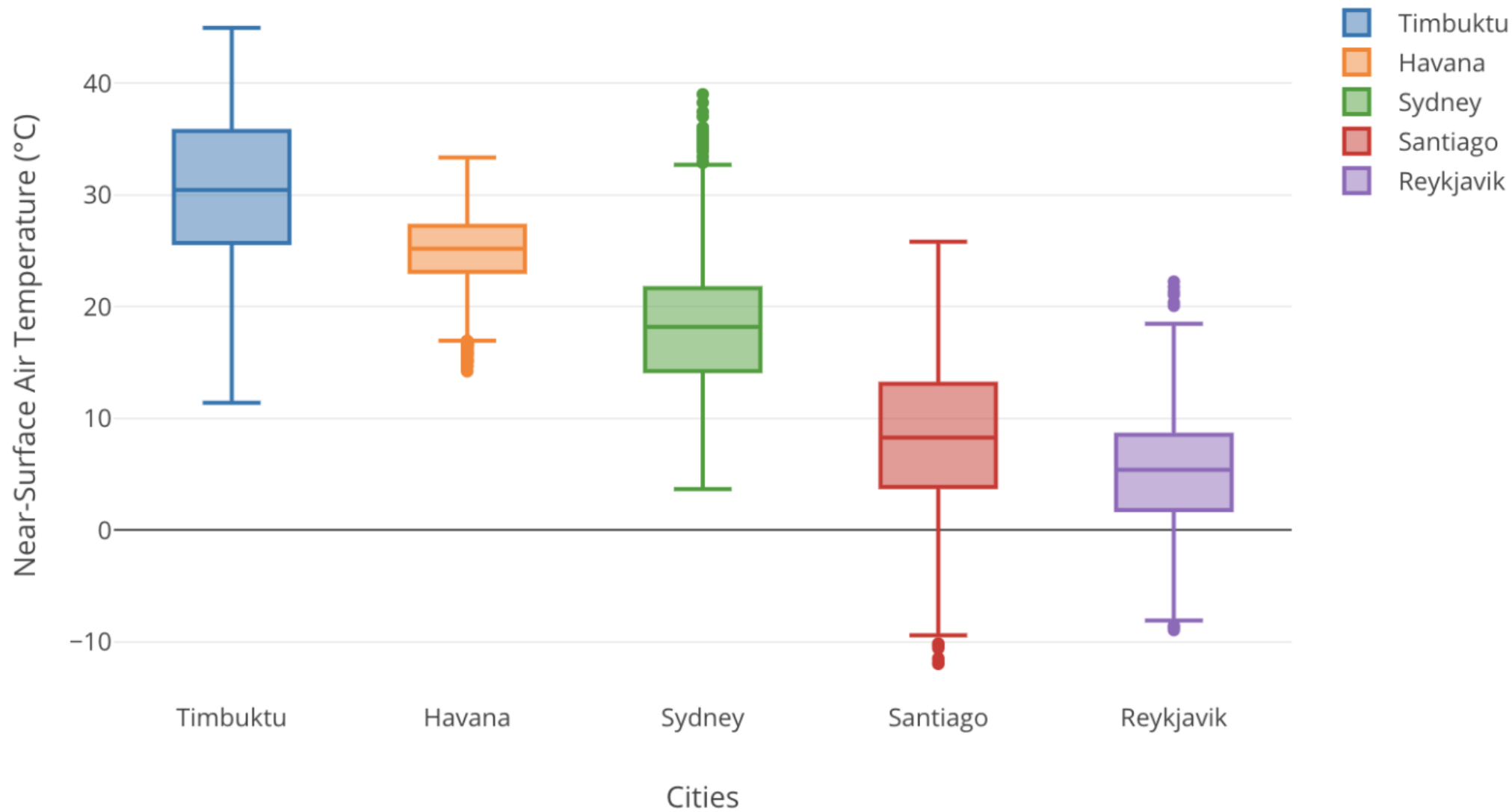
- vypadá jako sloupcový graf
- vizualizuje distribuci dat
- na ose x je interval
- na ose y četnost
- v čem je rozdíl?



Boxplot

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

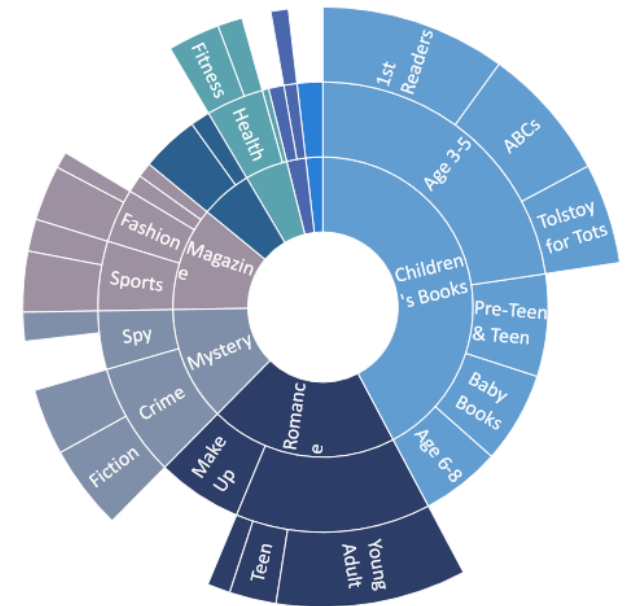
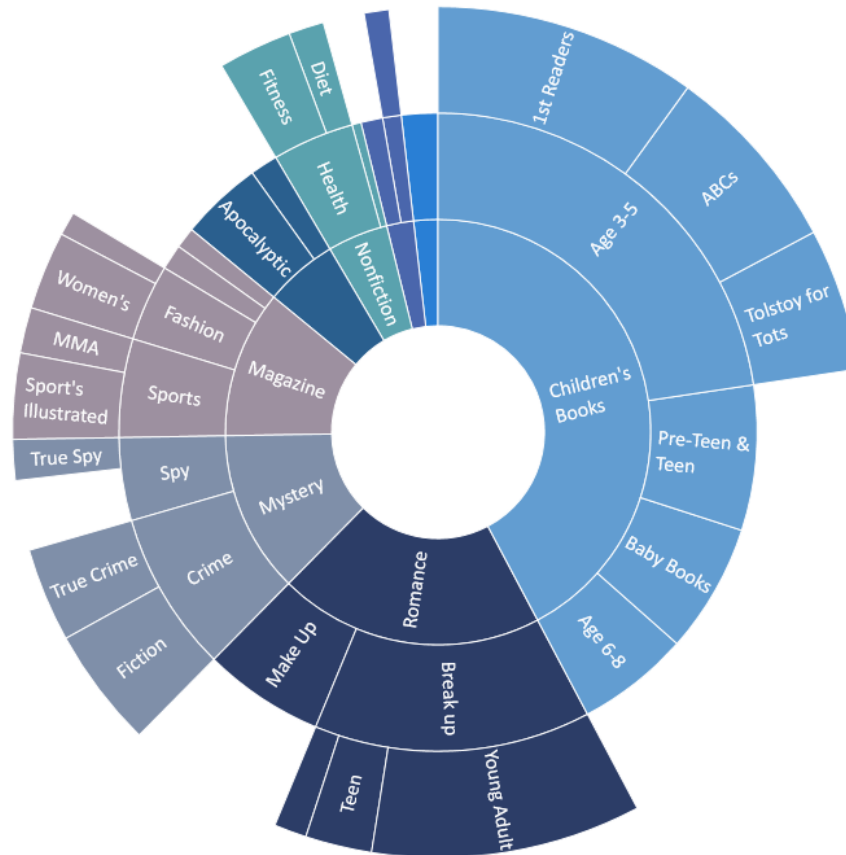




Hierarchická data



- treemapy
- sunburst
- dendrogram



Treemapy



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



Kam dál?

<https://datavizcatalogue.com/>

Charticulator – příště!

