

Vizualizace dat

Grafické vizualizační nástroje



Jiří Hynek, VUT FIT UIFS
ihynek@fit.vut.cz

Dělení dle účelu

- **hodnota v rozsahu:** gauge, bullet graph
- **porovnání, ranking:** bar chart, spider/radar, wordcloud, parallel, ...
- **distribuce hodnot:** histogram, boxplot, density, violin, ...
- **korelace:** scatter plot, heatmap, bubble, ...
- **vývoj (v čase):** line plot, sparkline, candlestick chart, area, stacked area, , ...
- **část celku, hierarchie:** pie/donut, stacked bars, treemap, circular packing, dendrogram, ...
- **relace, flow:** Sankey, chord, edge bundling, network, ...
- **mapy, geovizualizace:** choropleth, map with markers (bubble map), connection, cartogram, ...

<https://www.d3-graph-gallery.com>

<https://datavizcatalogue.com>

Dělení dle typu dat a dimenzí

- **numerická data** (kvantitativní)
- **kategorická data** (kvalitativní)
- **kombinovaná data**
- **časové řady**
- **sítě** (grafy – uzly a spojení)
- **mapy**

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

Hodnota v rozsahu

Porovnání, ranking

Distribuce hodnot

Korelace

Vývoj

Část celku, hierarchie

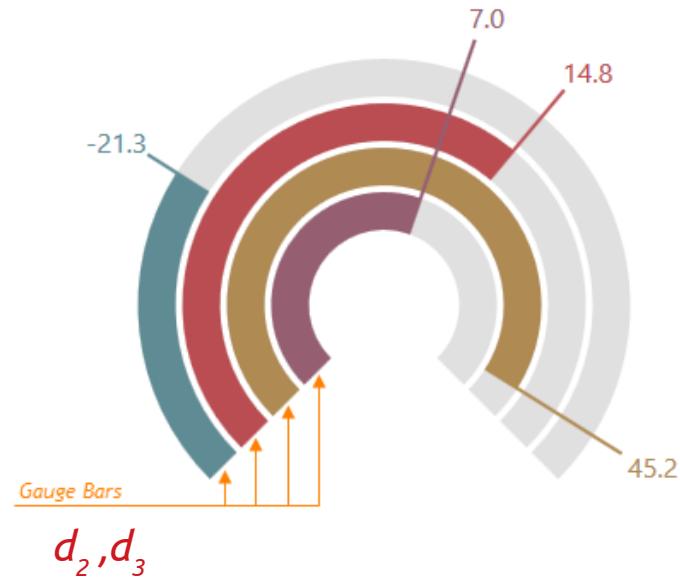
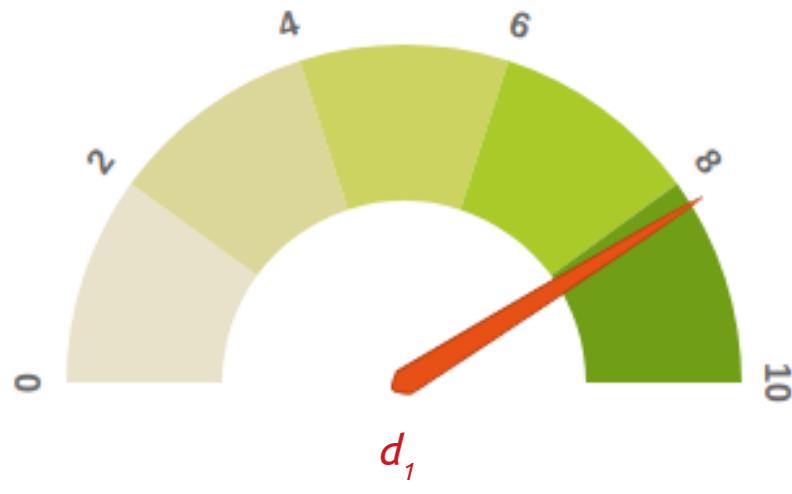
Relace, flow

Mapy, geovizualizace

Hodnota v rozsahu

Měřidlo

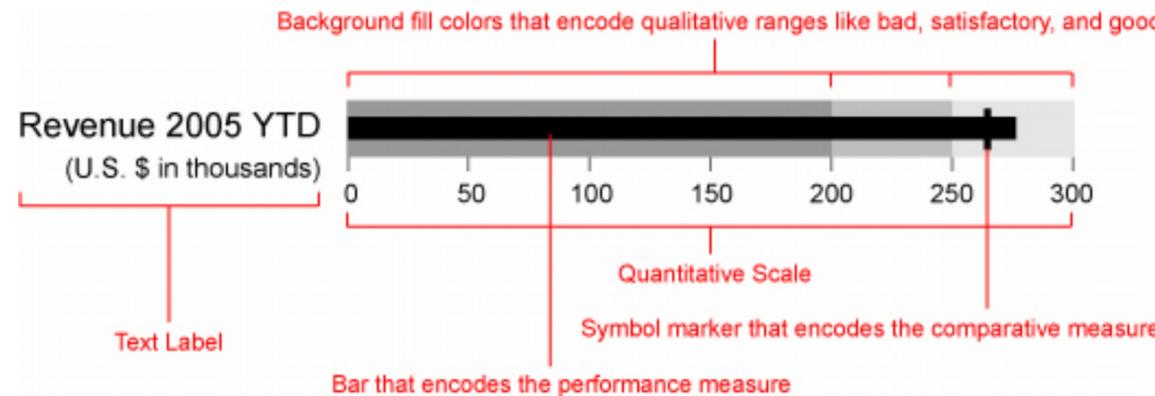
- *gauge*
- zobrazení n hodnot v rozsahu



- d_1 : numerická – hodnota z rozsahu
- $(+d_2$: kategorická – více proměnných z d_1)
- $(+d_3$: skupiny kategorií – např pomocí barev)

Bullet graph

- zobrazení hodnot v rozsahu, lze dobře skládat vedle sebe

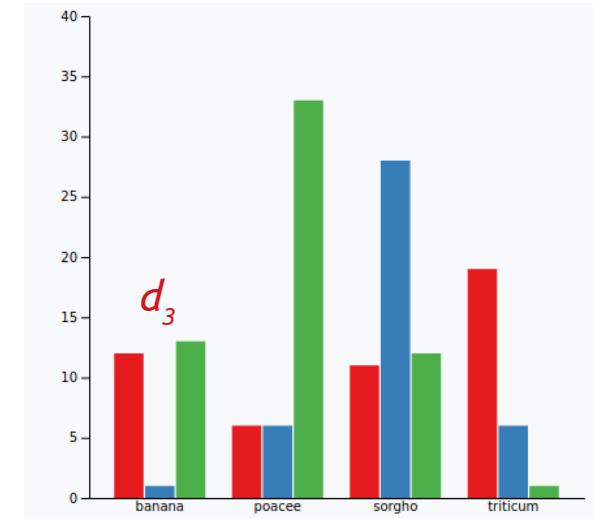
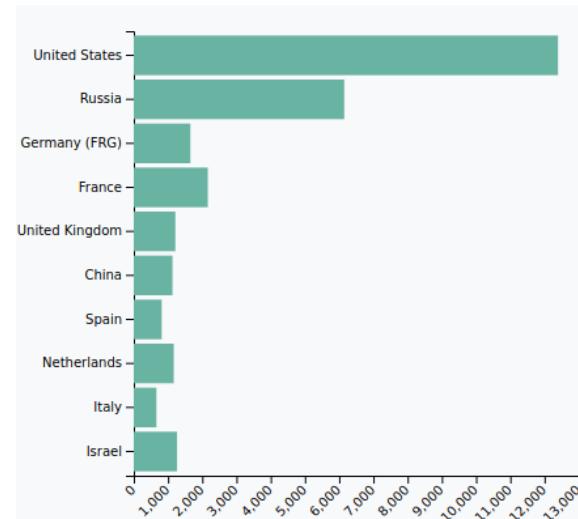
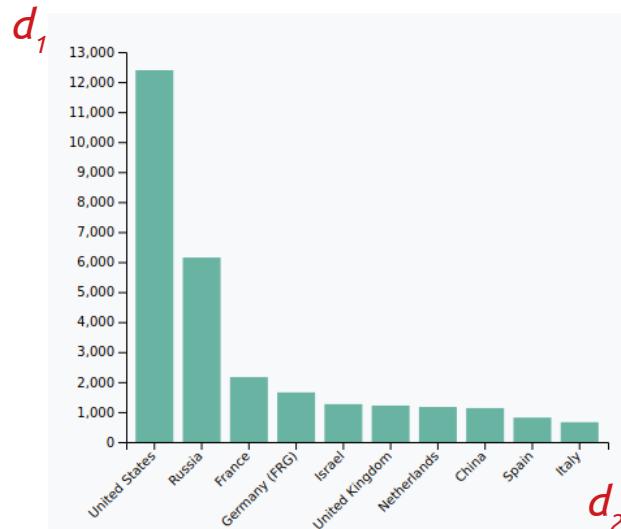


Hodnota v rozsahu
Porovnání, ranking
Distribuce hodnot
Korelace
Vývoj
Část celku, hierarchie
Relace, flow
Mapy, geovizualizace

Porovnání, ranking

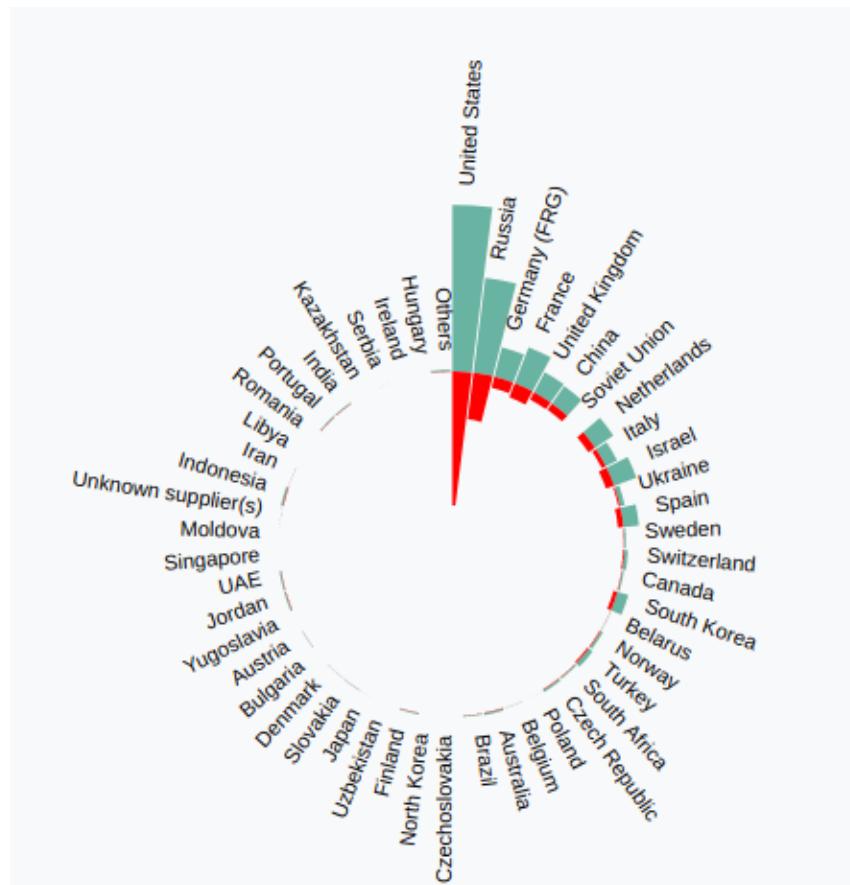
Sloupcový graf

- *bar chart / barplot*, porovnává hodnoty

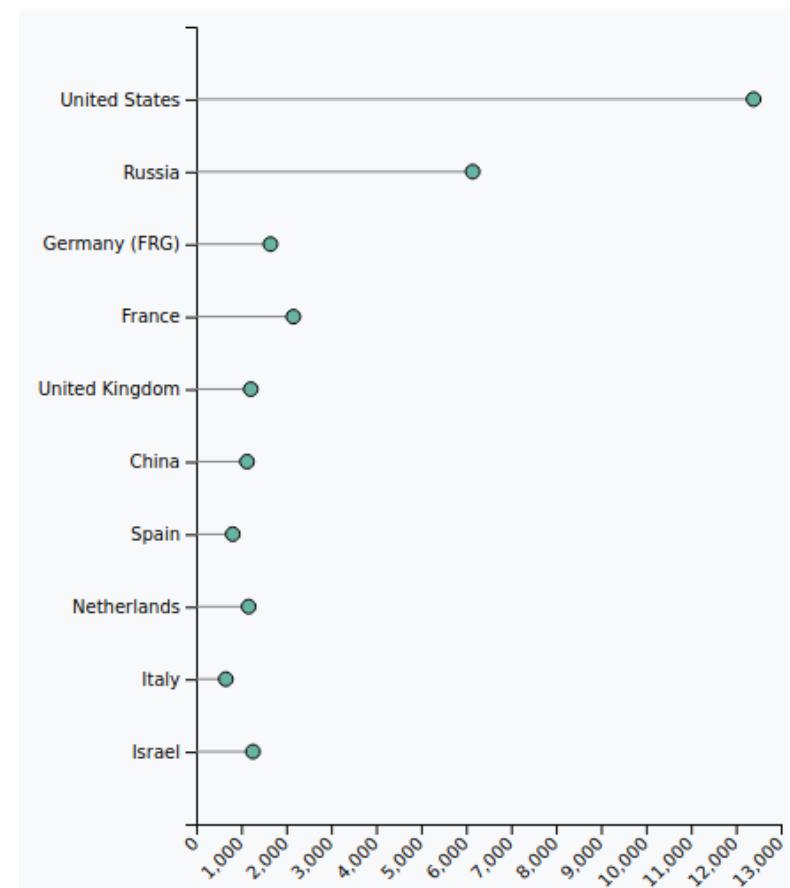


- d_1 : numerická – hodnota
- d_2 : kategorická – zpravidla ordinální – nezáleží na pořadí
- $(+d_3, \dots)$: kategorická – více sad sloupců, např. barevně odlišených)

(Sloupcový graf - varianty)



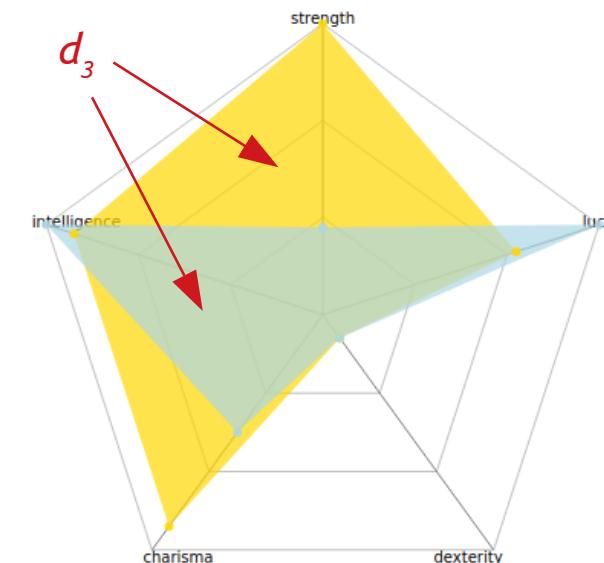
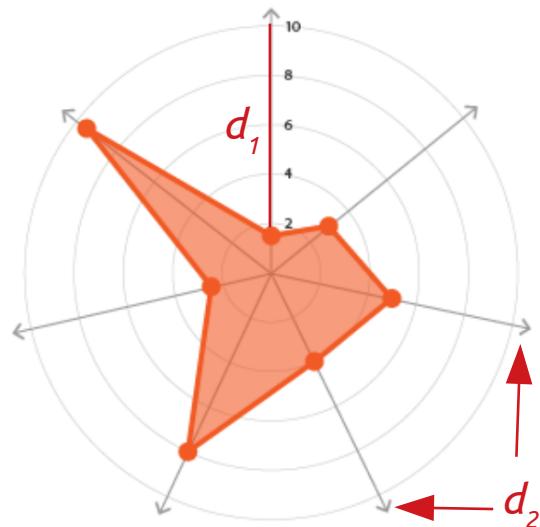
radial/circular bar chart



lollipop chart

Paprskový graf

- **radaorvý (radar) / pavučinový (spider) chart, porovnává hodnoty, případně relace hodnot**



- d_1 : numerická – hodnota
- d_2 : kategorická – sada kategorií vytváří polygon
- $(+d_3, \dots)$: kategorická – více polygonů, např. barevně odlišených)

Paprskový graf - příklad

- polygony mohou posloužit pro porovnávání

Points (G+A)

table chart [§ Glossary](#)



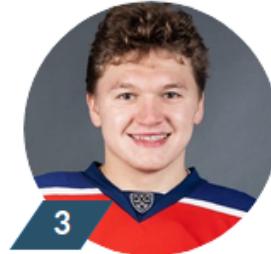
Shipachyov Vadim

№ 87, Dynamo Msk



Jaskin Dmitrij

№ 23, Dynamo Msk



Kaprizov Kirill

№ 97, CSKA



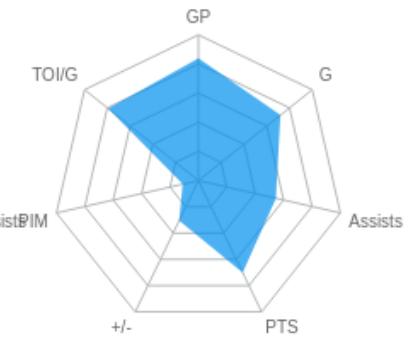
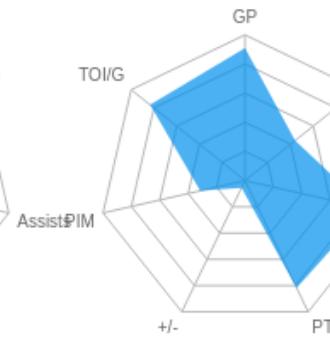
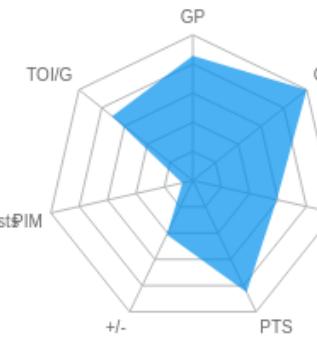
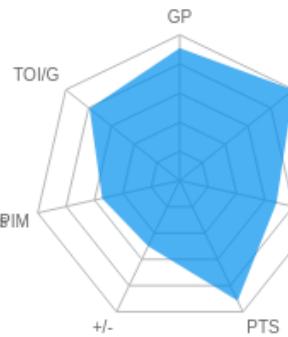
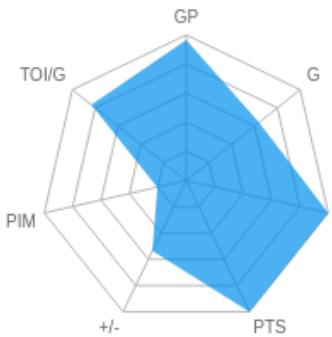
Omark Linus

№ 67, Salavat Yulaev



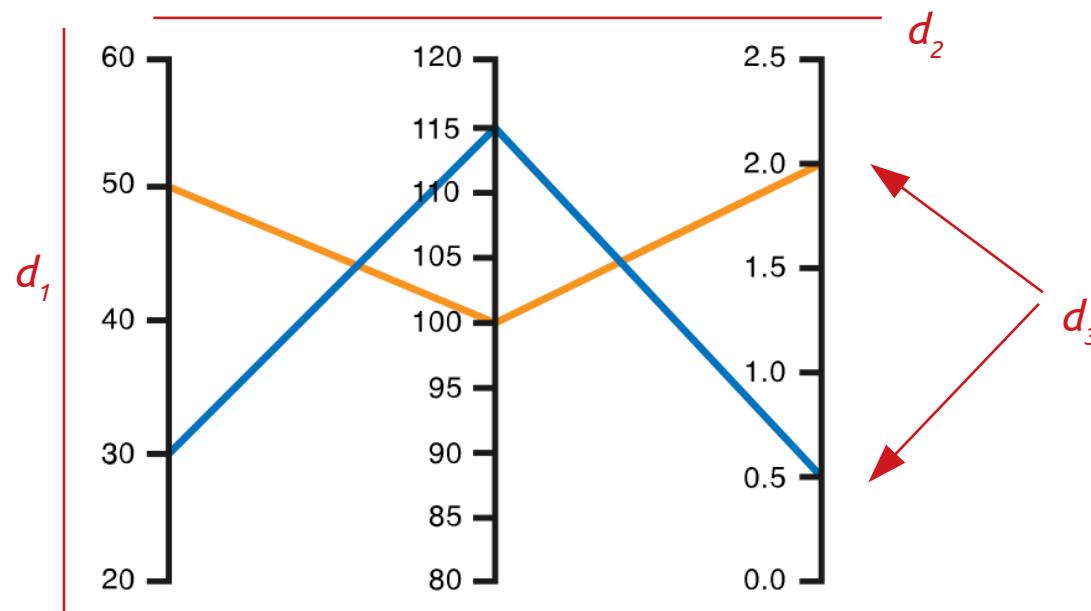
O'Neill Brian

№ 21, Jokerit

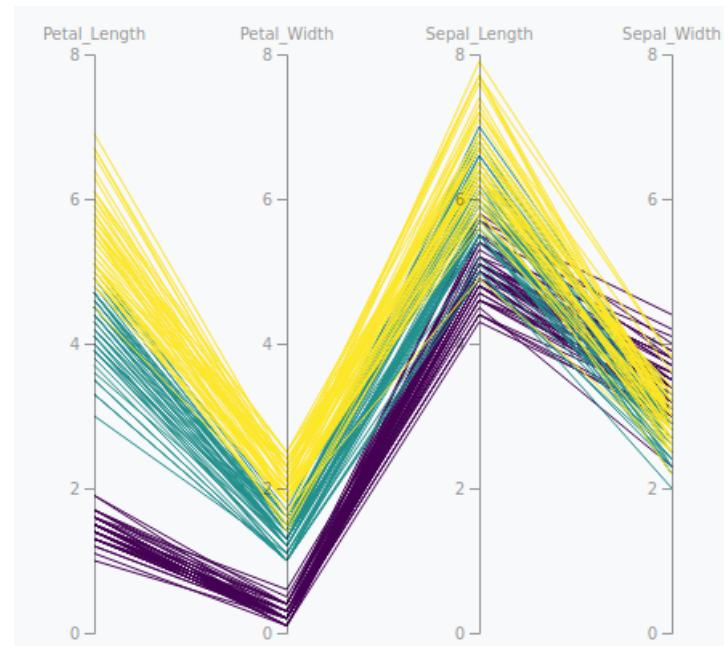


(Paralel coordinates plot)

- porovnává relace hodnot

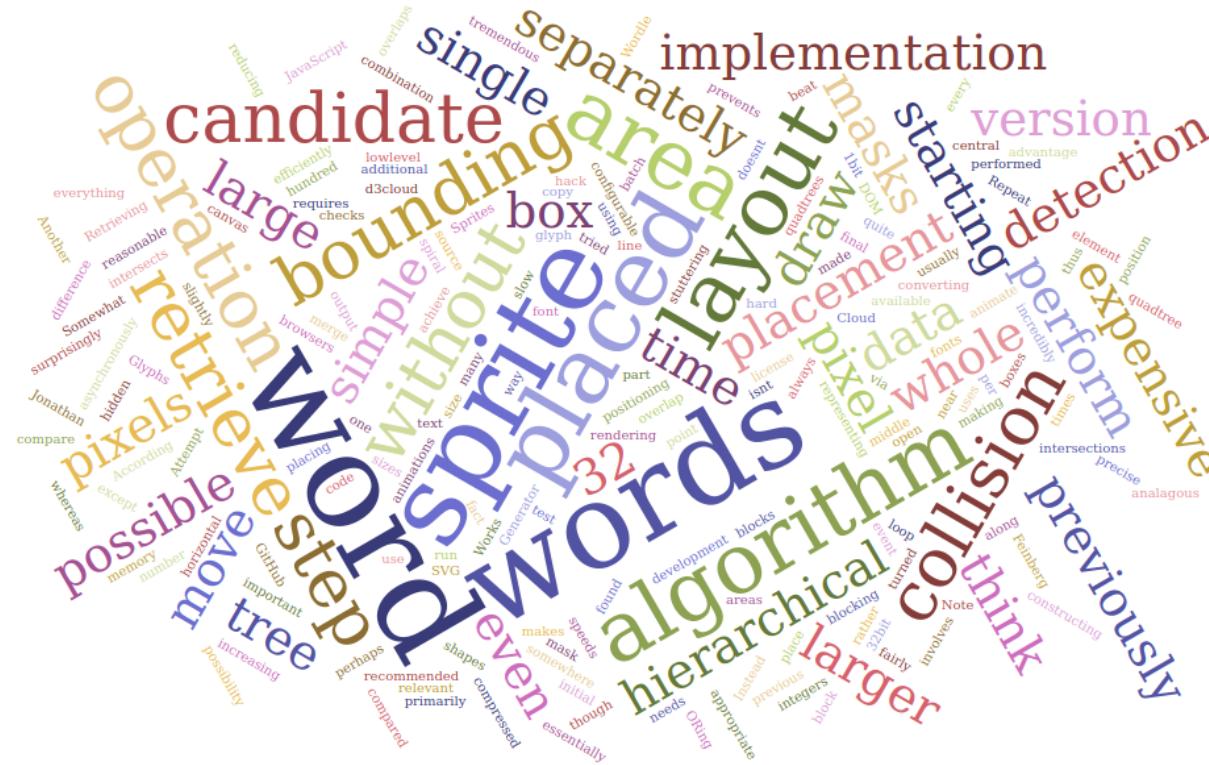


- d_1 : numerická
- d_2 : kategorická – různé proměnné
- d_3 : kategorická – např. barevné odlišení
 - $(+d_4, \dots)$: skupiny kategorií – např. barevné odlišení)
- interaktivní příklad: <http://bl.ocks.org/syntagmatic/raw/3150059/>



(Mraky slov)

- ***wordcloud***, porovnává hodnoty, klade důraz na kategorie



- d_1 : numerická – např. četnost nějaké kategorie
 - d_2 : kategorická – reprezentuje nějaký slovní výraz
 - $(+d_3, \dots : \text{skupiny kategorií} - \text{např. barevné odlišení})$

Hodnota v rozsahu
Porovnání, ranking
Distribuce hodnot

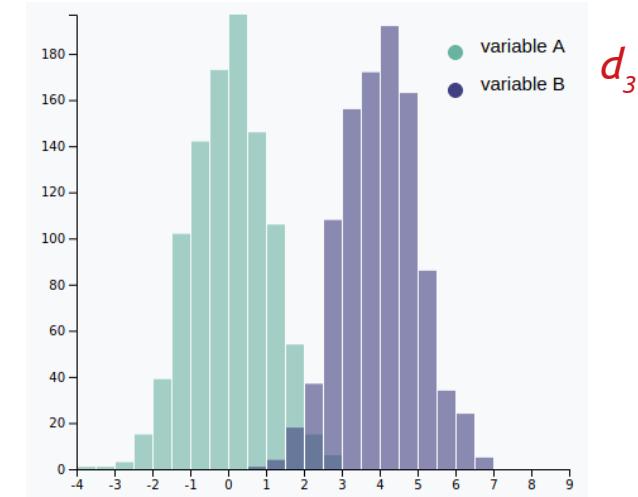
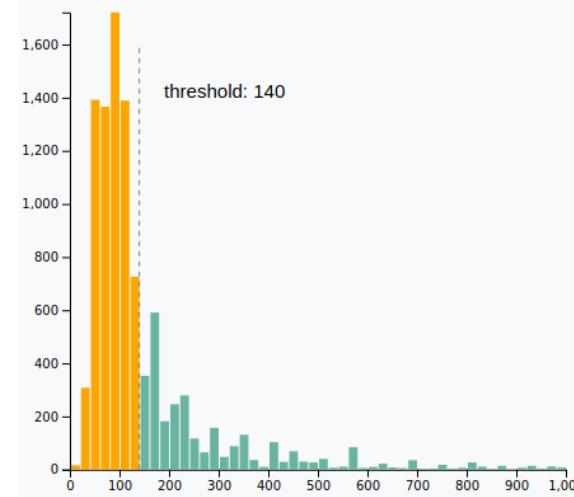
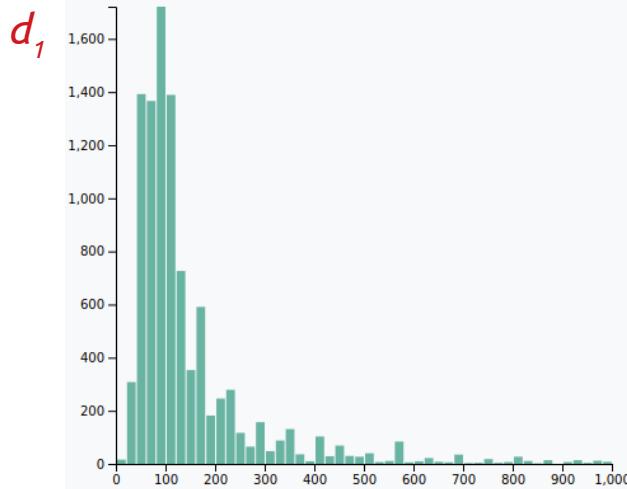
Korelace
Vývoj

Část celku, hierarchie
Relace, flow
Mapy, geovizualizace

Distribuce hodnot

Histogram

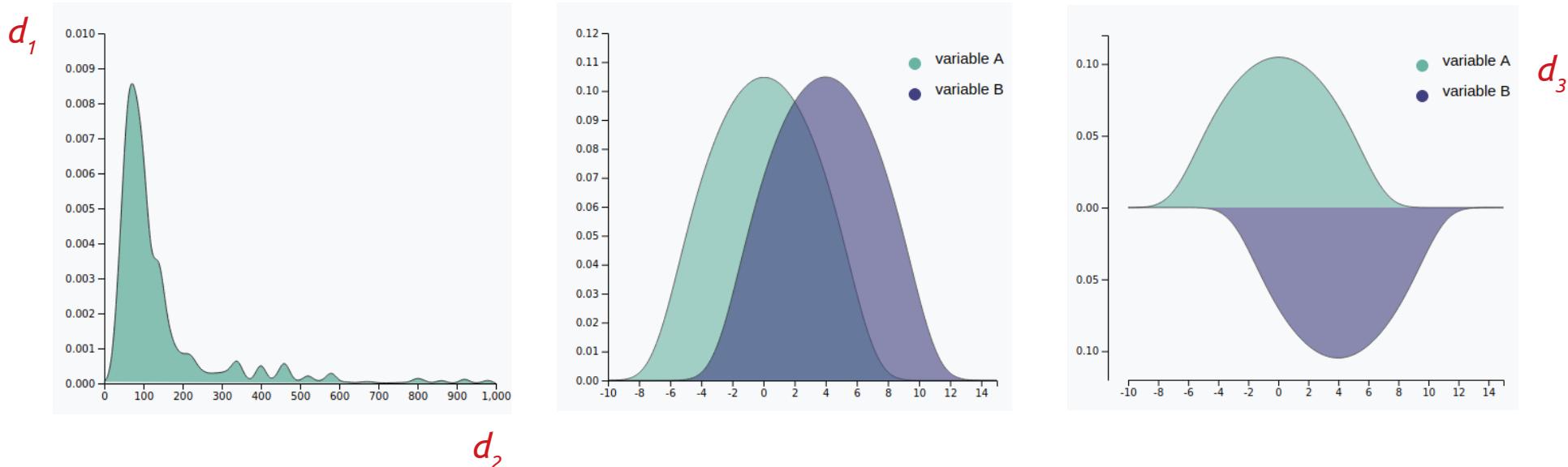
- zobrazení **distribuce hodnot** nějaké proměnné



- d_1 : numerická (spojitá dimenze) – např. četnost hodnot nějaké proměnné
- d_2 : numerická (diskrétní) – možné hodnoty proměnné
- (+ d_3 , ... : více proměnných – např. barevné odlišení)
- Vstup:** zpravidla agregovaný (**počet**) seznam hodnot

Graf hustoty

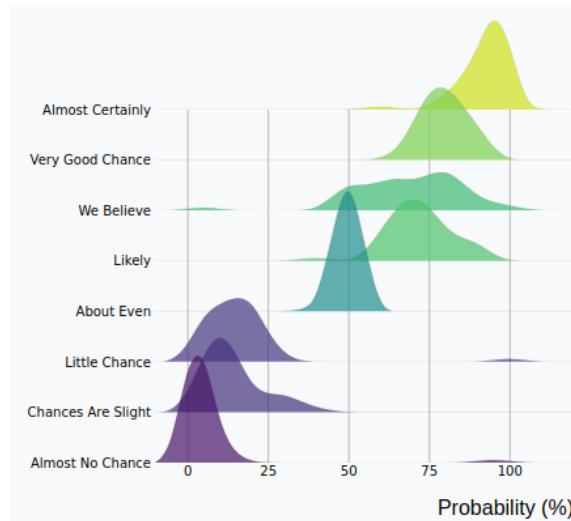
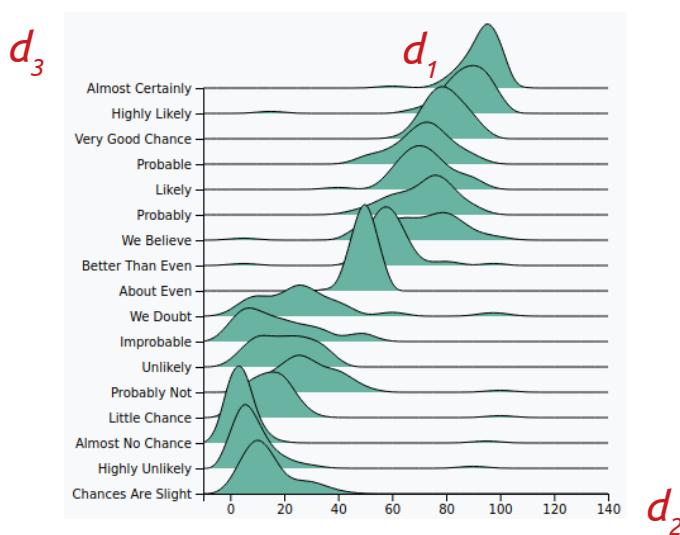
- **density chart**, zobrazení **distribuce hodnot** (hustota pravděpodobnosti, ...) nějaké proměnné/funkce



- d_1 : numerická – např. četnost nějakých hodnot proměnné
- d_2 : numerická (spojitá) – např. definiční obor hodnot nějaké funkce
- (+ d_3 : kategorická – více proměnných/funkcí, překrývajících se)
 - (+ d_4 , ... : skupiny kategorií – např. barevně odlišených)

(Ridgeline chart (*ridge ~ hřeben*))

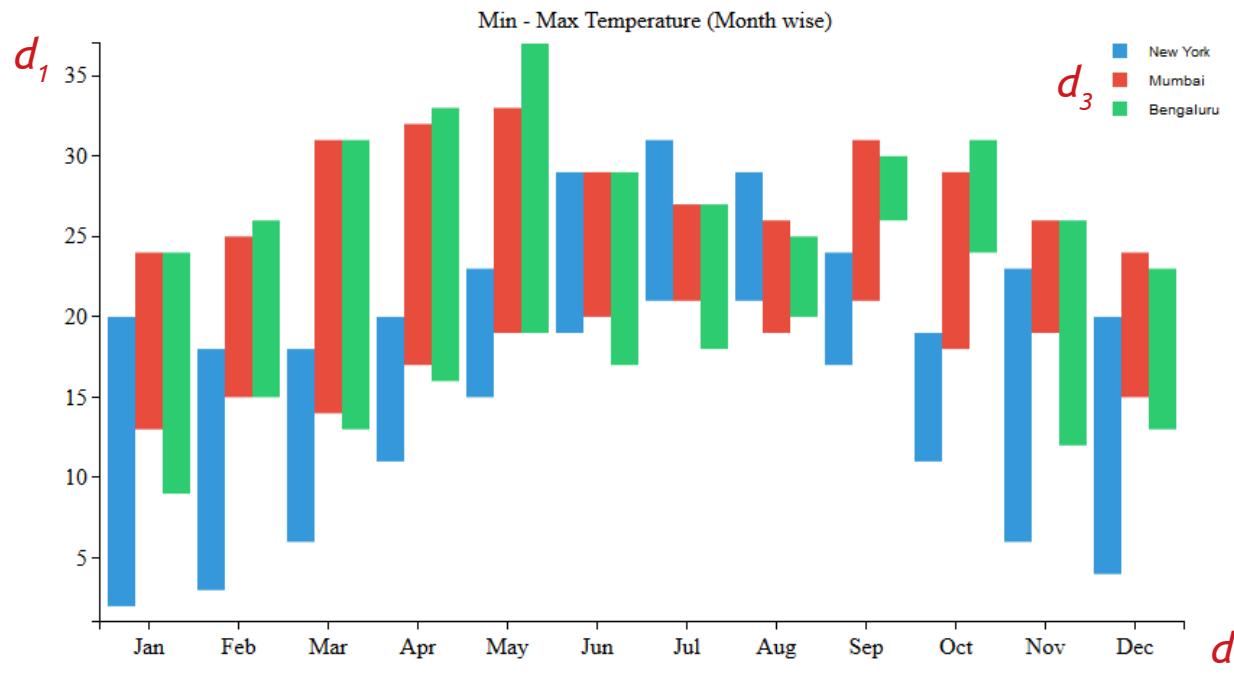
- graf skládající se z grafů hustoty, vizualizuje **distribuci skupin hodnot** a jejich **porovnání**



- d_1 : numerická – např. četnost nějakých hodnot proměnné (*agregované hodnoty*)
- d_2 : numerická – např. definiční obor hodnot nějaké funkce/proměnné
- d_3 : kategorická – více proměnných/funkcí, pod sebou
 - ($+d_4, \dots$: skupiny kategorií – např. barevné odlišení)

Plovoucí sloupcový graf

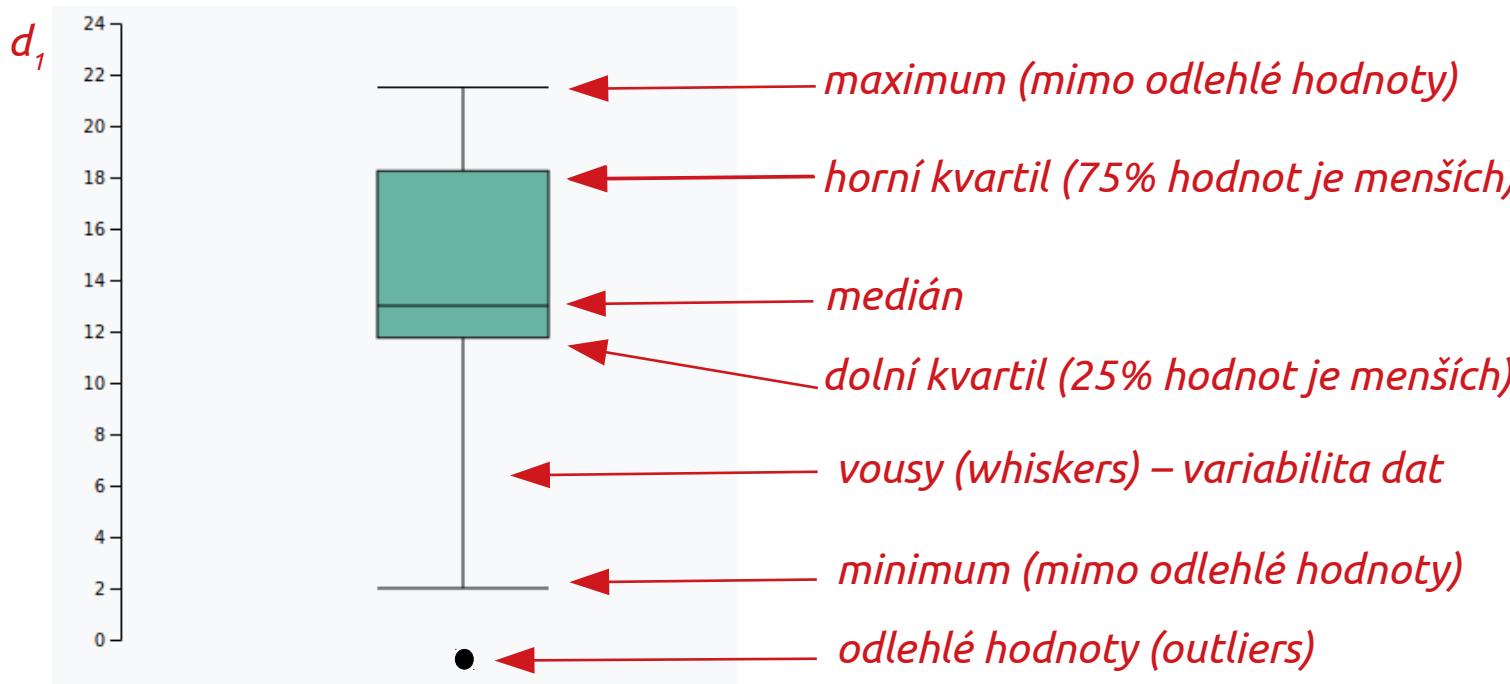
- *floating bar chart*, vizualizace a porovnávání rozsahů hodnot



- d_1 : numerická – *relace*: (min, max)
- d_2 : kategorická
- d_3 : kategorická – např. barevné odlišení
- **Vstup:** např. agregované seznamy (min , max) hodnot

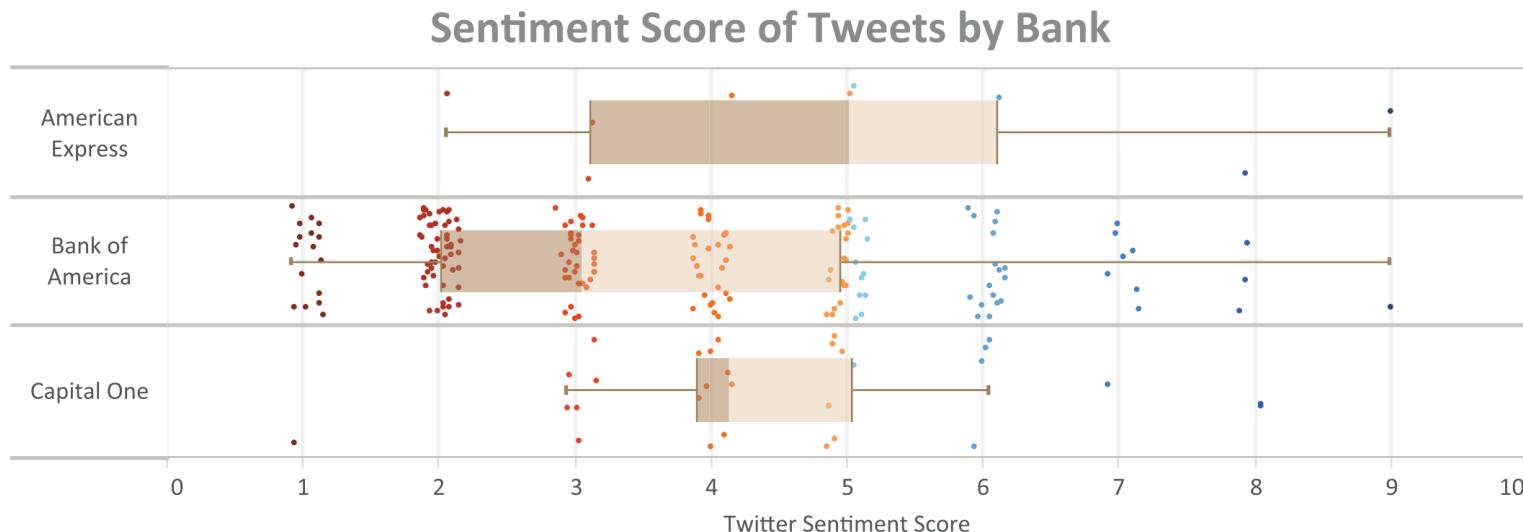
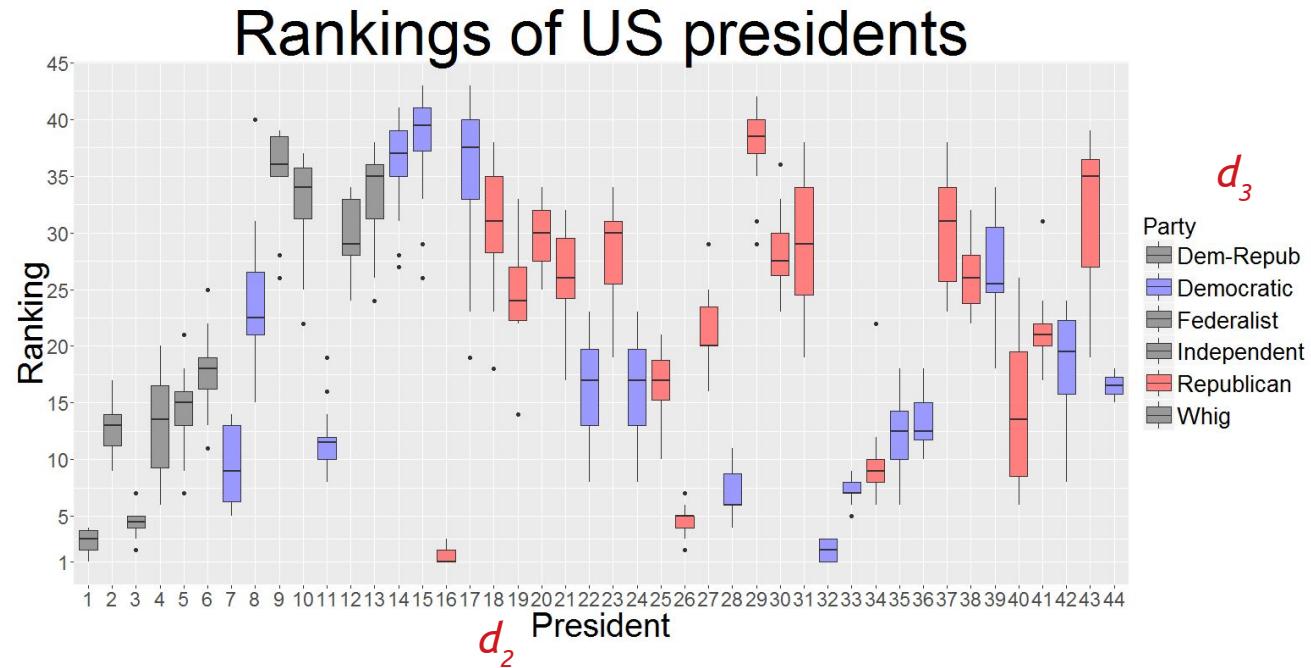
Krabicový graf

- **box plot**, vizualizuje distribuce skupin hodnot a jejich porovnání



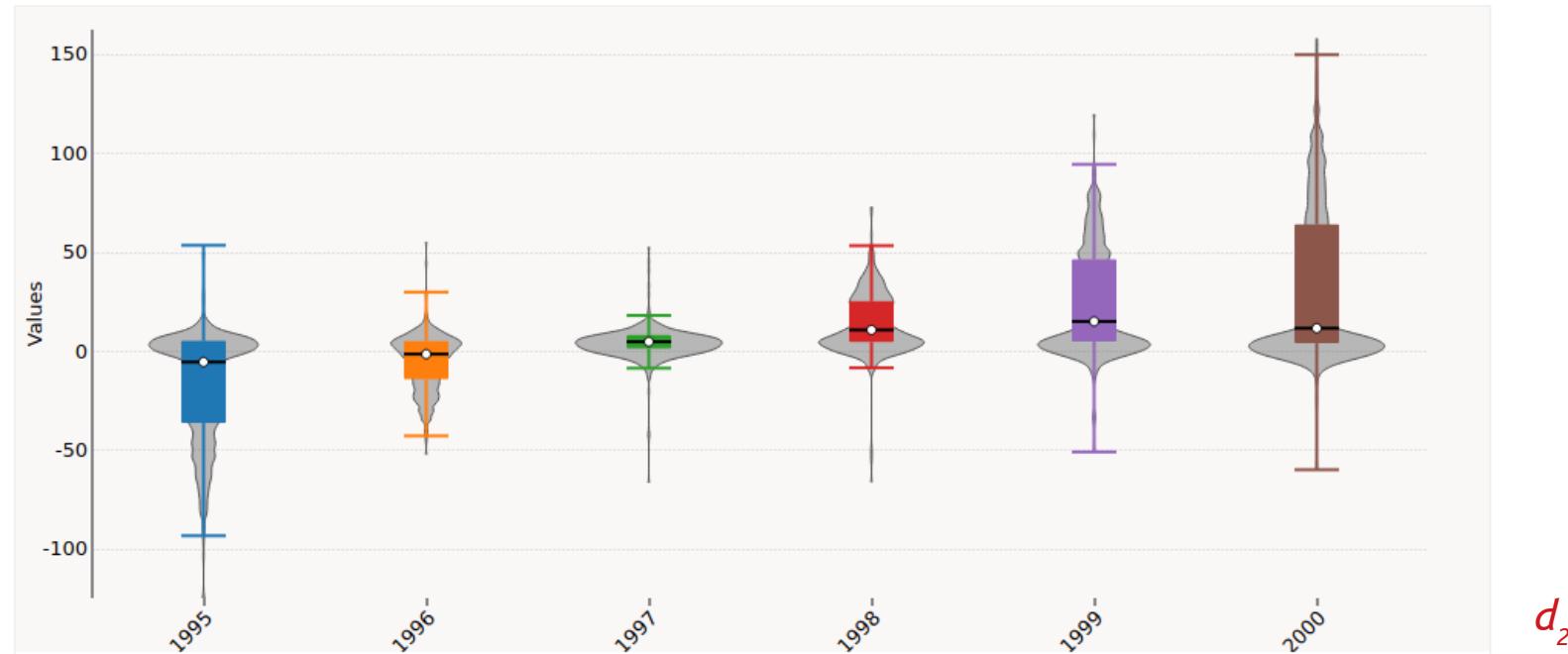
- d_1 : numerická – relace: (\min , $Q_{0,25}$, median, $Q_{0,75}$, \max)
- d_2 : kategorická
- (+ d_3 , ... : kategorická – např. barevné odlišení)
- **Vstup:** zpravidla agregované seznamy hodnot

Krabicový graf - příklady



(Violin plot („houslový“ - tvar houslí))

- kombinuje krabicový graf a graf hustoty



- d_1 : numerická – viz krabicový graf a graf hustoty
- d_2 : kategorická
- $(+d_3, \dots)$: kategorická – např. barevné odlišení
- **Vstup:** zpravidla agregované seznamy hodnot

Hodnota v rozsahu
Porovnání, ranking
Distribuce hodnot

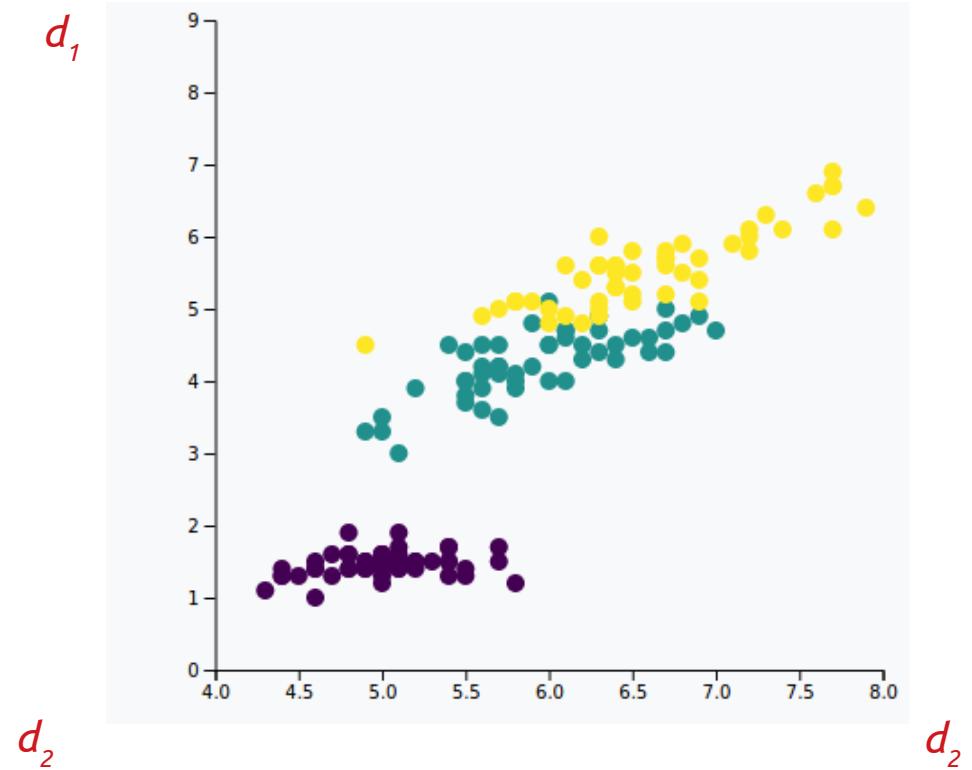
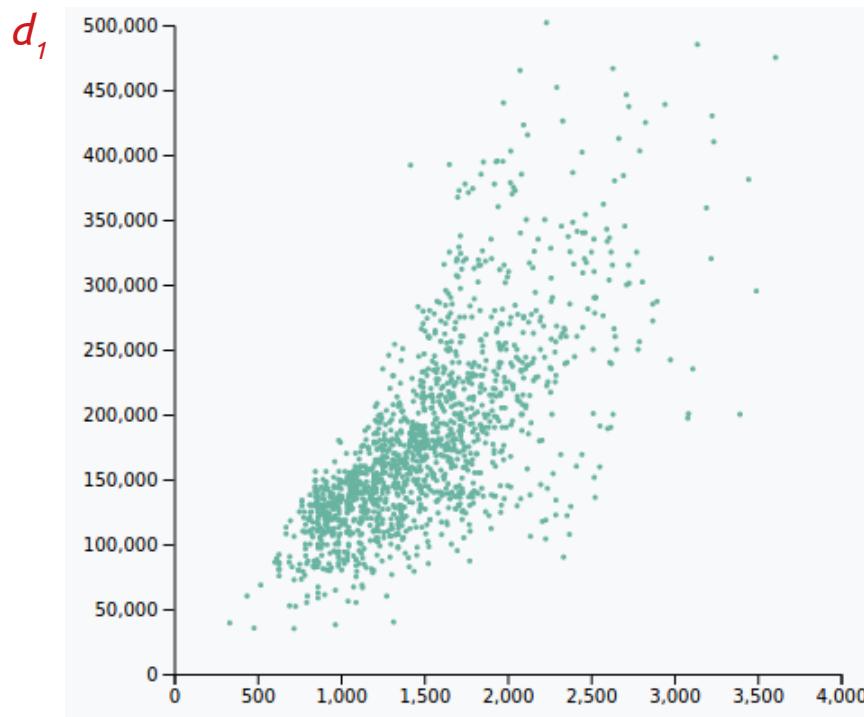
Korelace

Vývoj
Část celku, hierarchie
Relace, flow
Mapy, geovizualizace

Korelace

Bodový graf

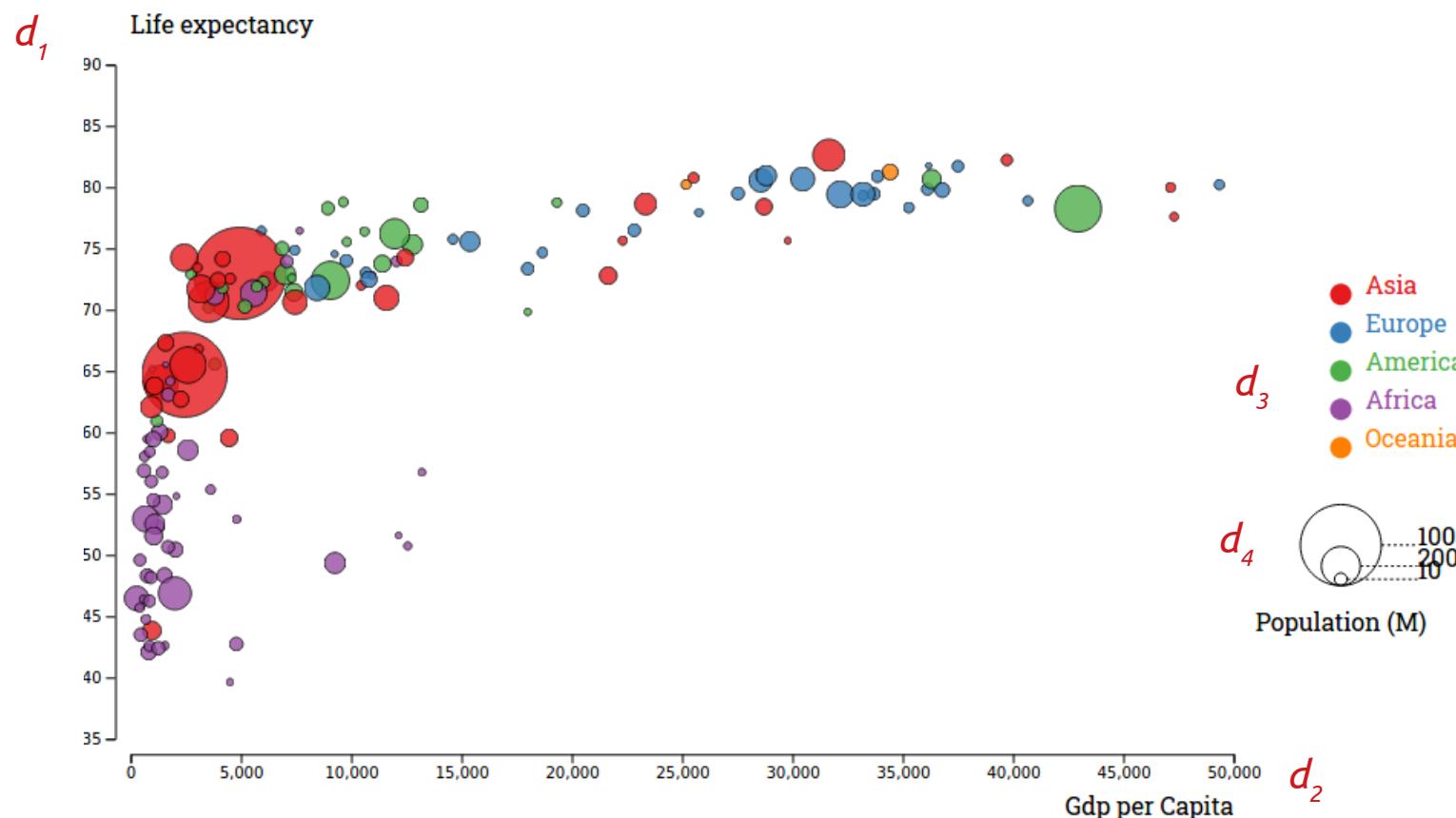
- *scatter plot*, vizualizuje korelaci dvou proměnných



- d_1 : numerická – *proměnná #1*
- d_2 : numerická – *proměnná #2*
- $(+d_3, \dots)$: kategorická – *např. barevné odlišení*

Bublinový graf

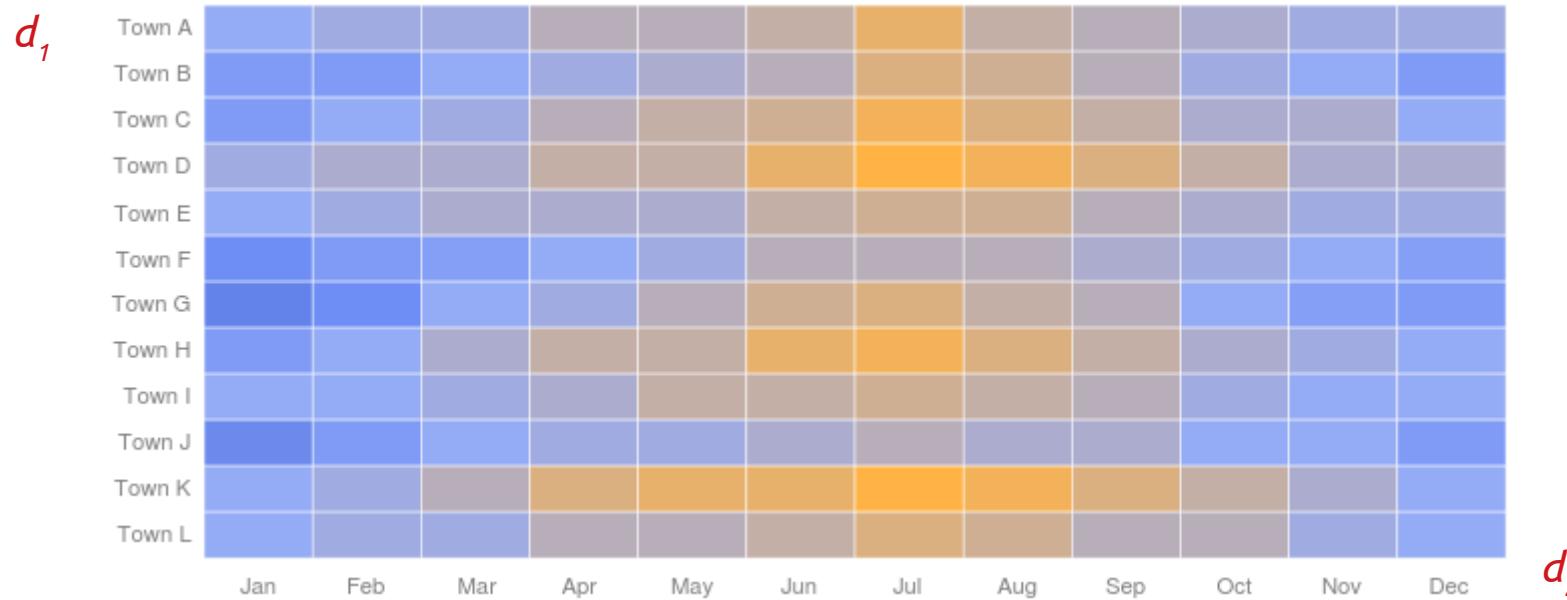
- **bubble chart**, bodový graf obsahující kruhy (*poloměr kruhu*) – přidává další dimenzi – vizualizuje vztah



- $d_{1,2,(3)}$, ... : viz bodový graf
- $+d_4, \dots$: numerická – *poloměr kruhu*

Teplotní mapa

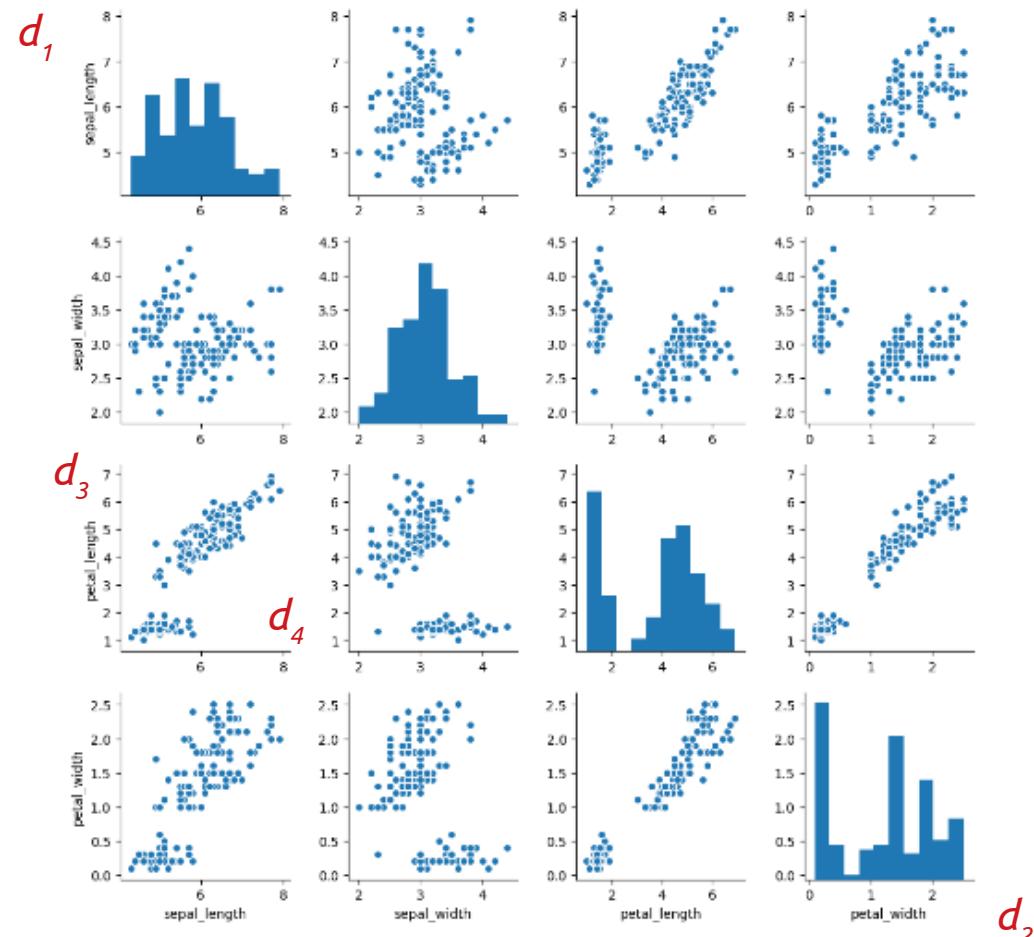
- heatmap, vizualizuje vztah mezi dvěma proměnnými



- $d_{1,2}$: numerická (diskrétní) /kategorická
- d_3 : numerická – barevná intenzita
- $(+d_4, \dots)$: kategorická – např. barevný odstín)

(Korelogram)

- *correlogram*, vizualizuje vztahy mezi množinou proměnných



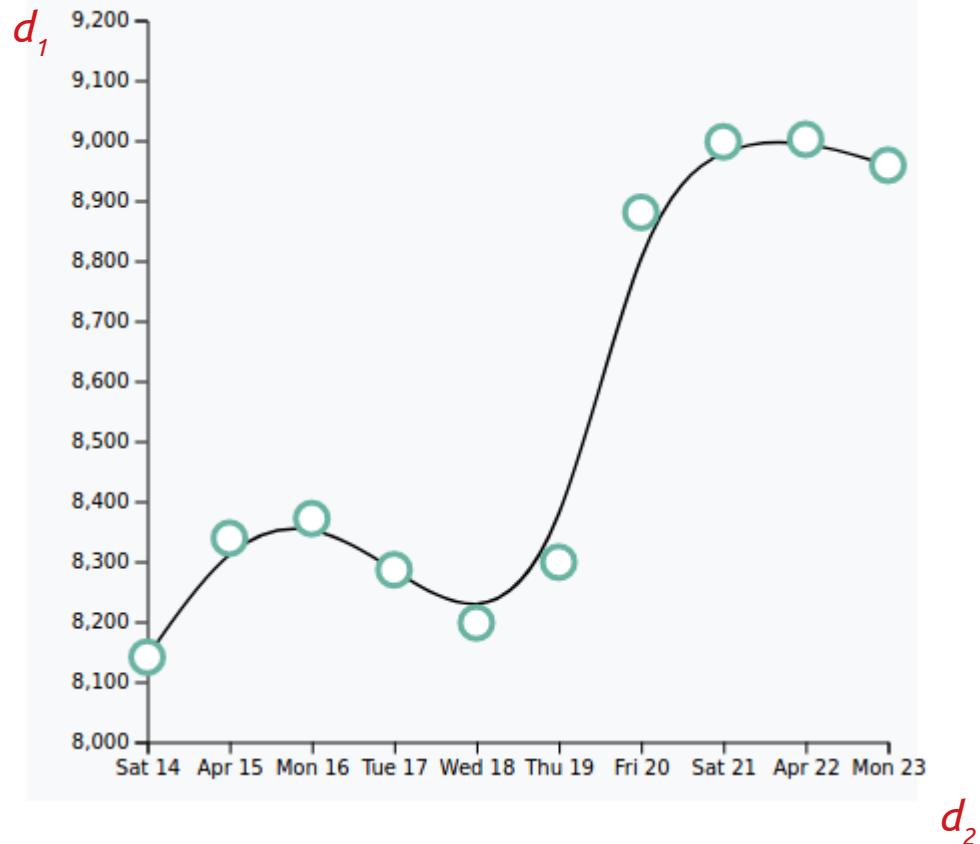
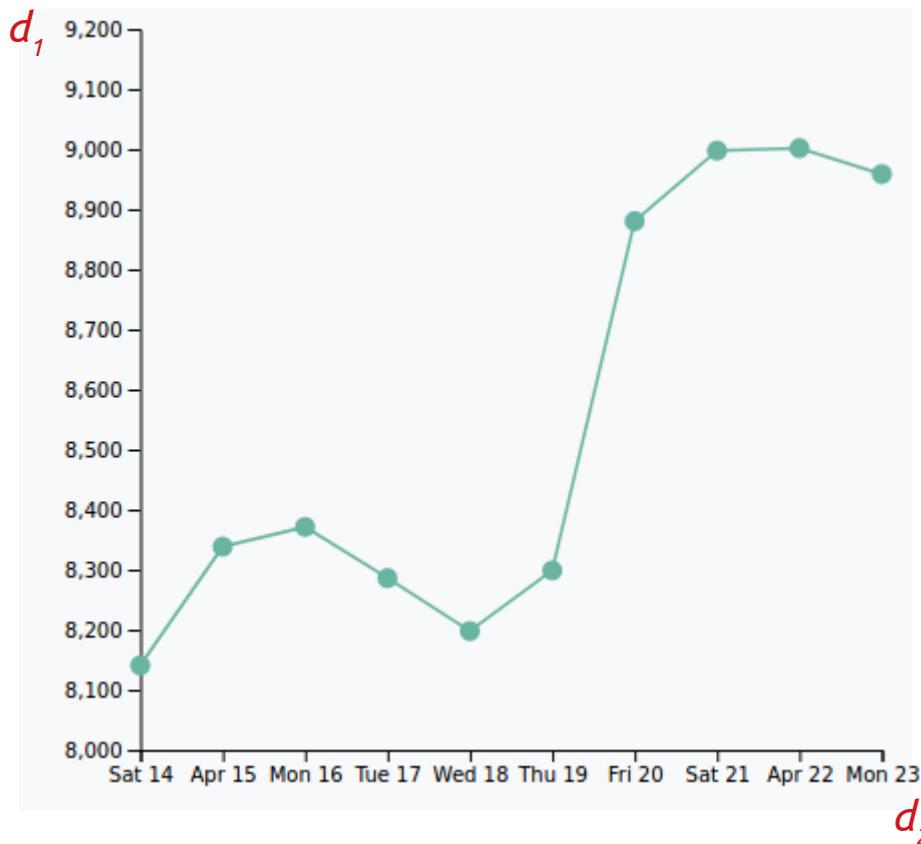
- $d_{1,2}$: kategorická – matici
- $d_{3(4)}$: numerická
- $(+d_5, \dots)$: kategorická – např. barevné odlišení

Hodnota v rozsahu
Porovnání, ranking
Distribuce hodnot
Korelace
Vývoj
Část celku, hierarchie
Relace, flow
Mapy, geovizualizace

Vývoj

Spojnicový graf

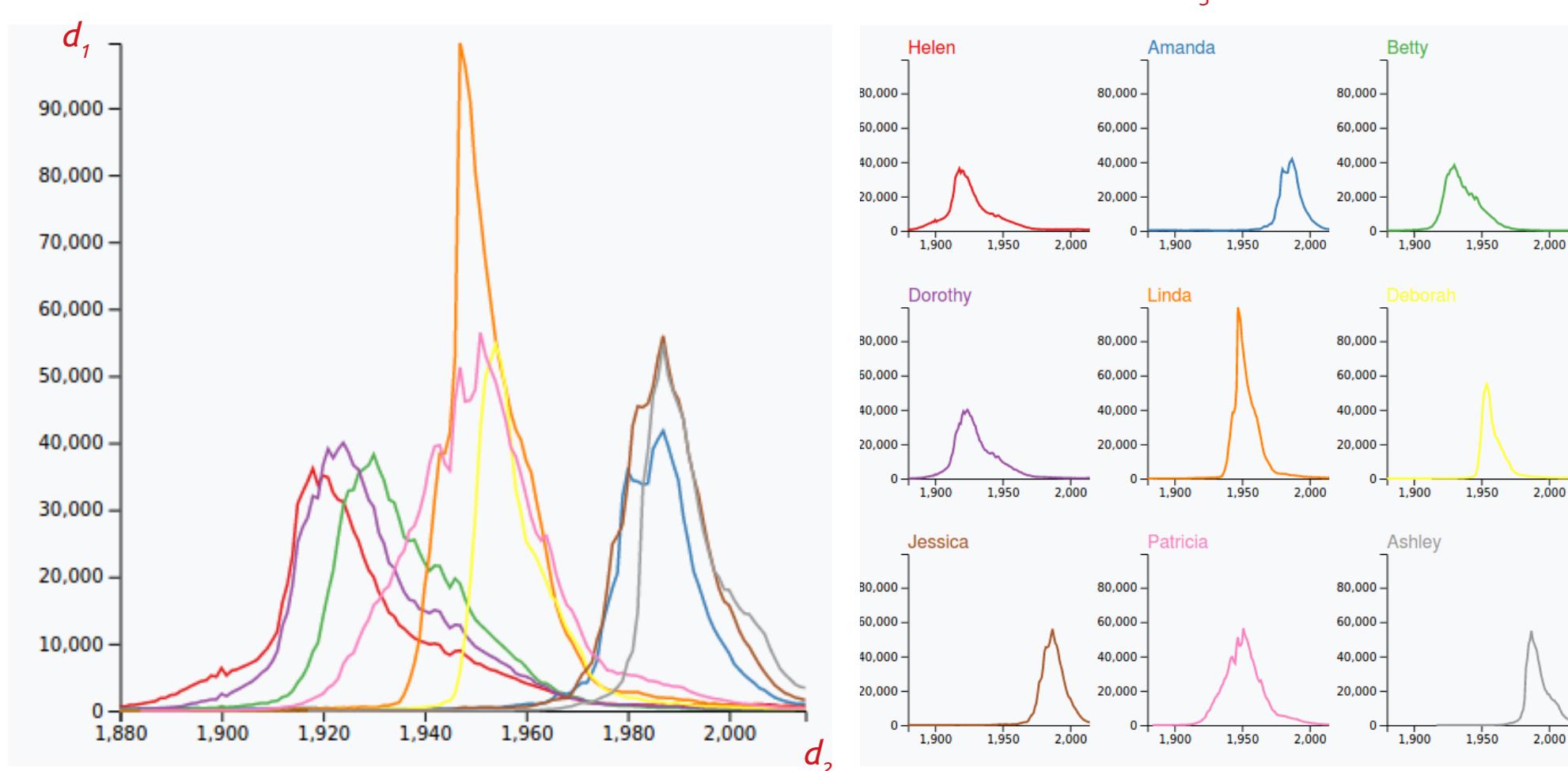
- *line chart*, spojuje body na bodovém grafu – vizualizuje vývoj



- $d_{1,2}$: numerická
- může přímo propojit body, nebo proložit přibližnou křivku

Spojnicových graf, matice

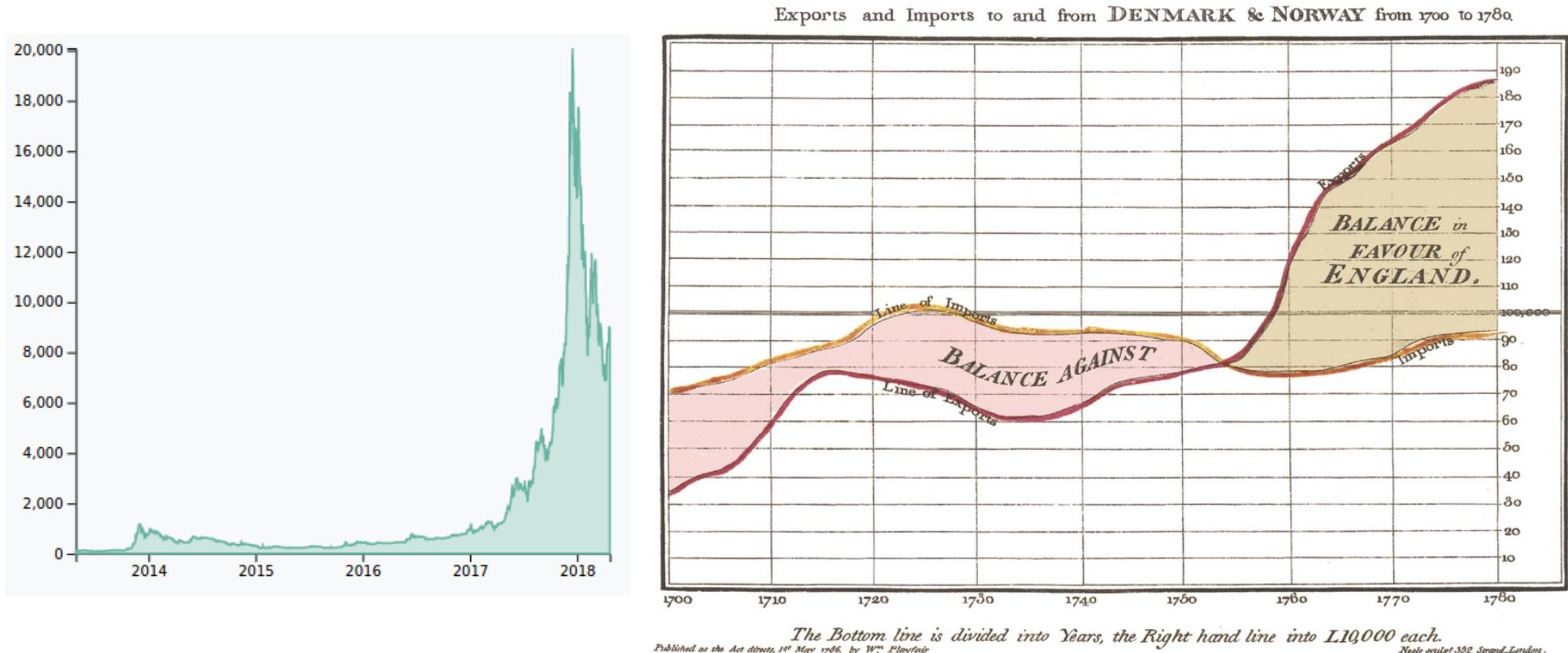
- porovnává vývoj více kategorií



- $+d_3, \dots$: kategorická – např. barevné odlišení, matice spojnicových grafů, ...

Plošný graf

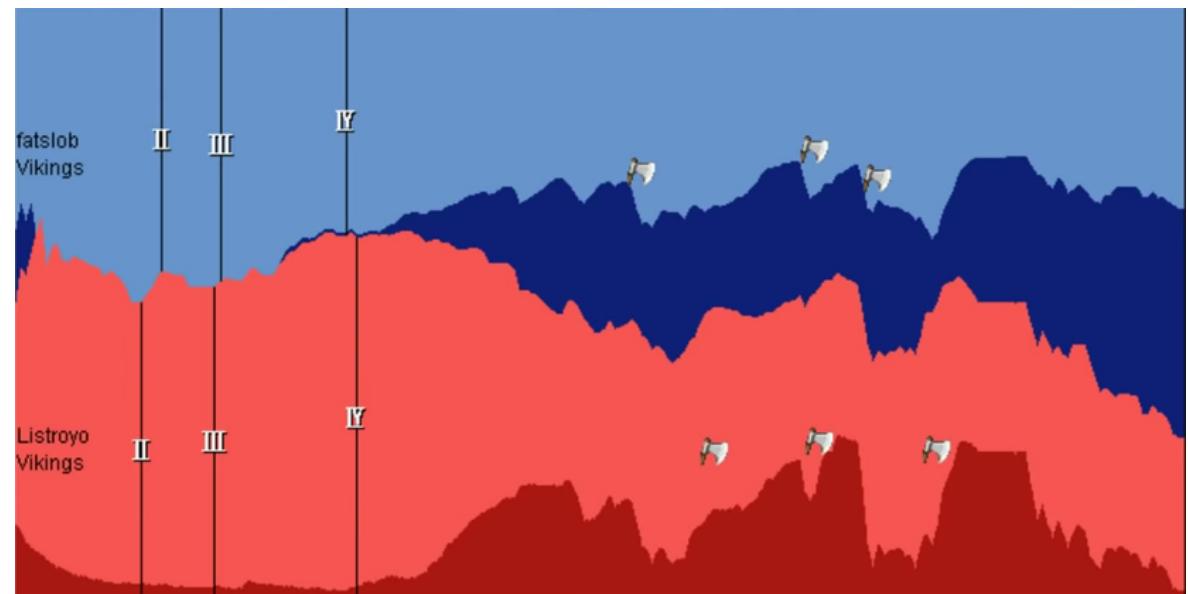
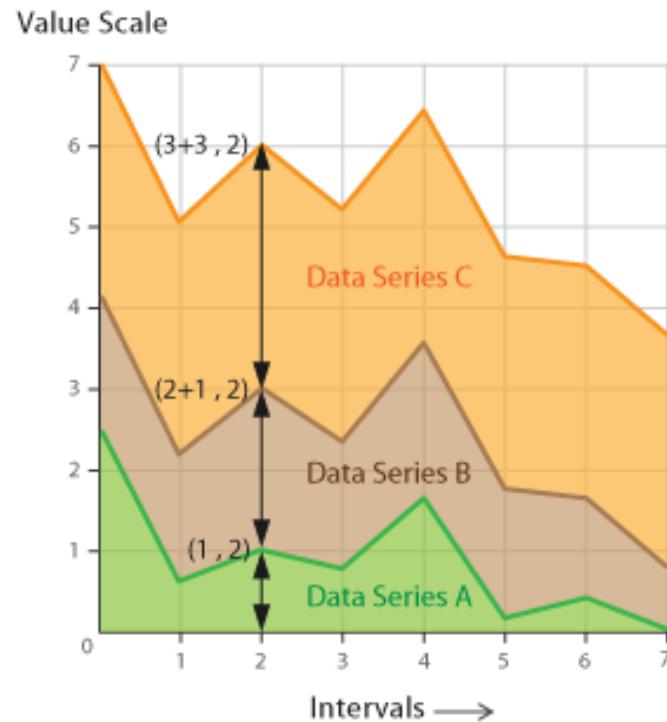
- area chart, vývoj hodnoty (např. v čase)
- zvýrazňuje plochu pod spojnicí



Vývoj importu a exportu mezi Anglií a Dánskem+Norskem mezi lety 1700 - 1780

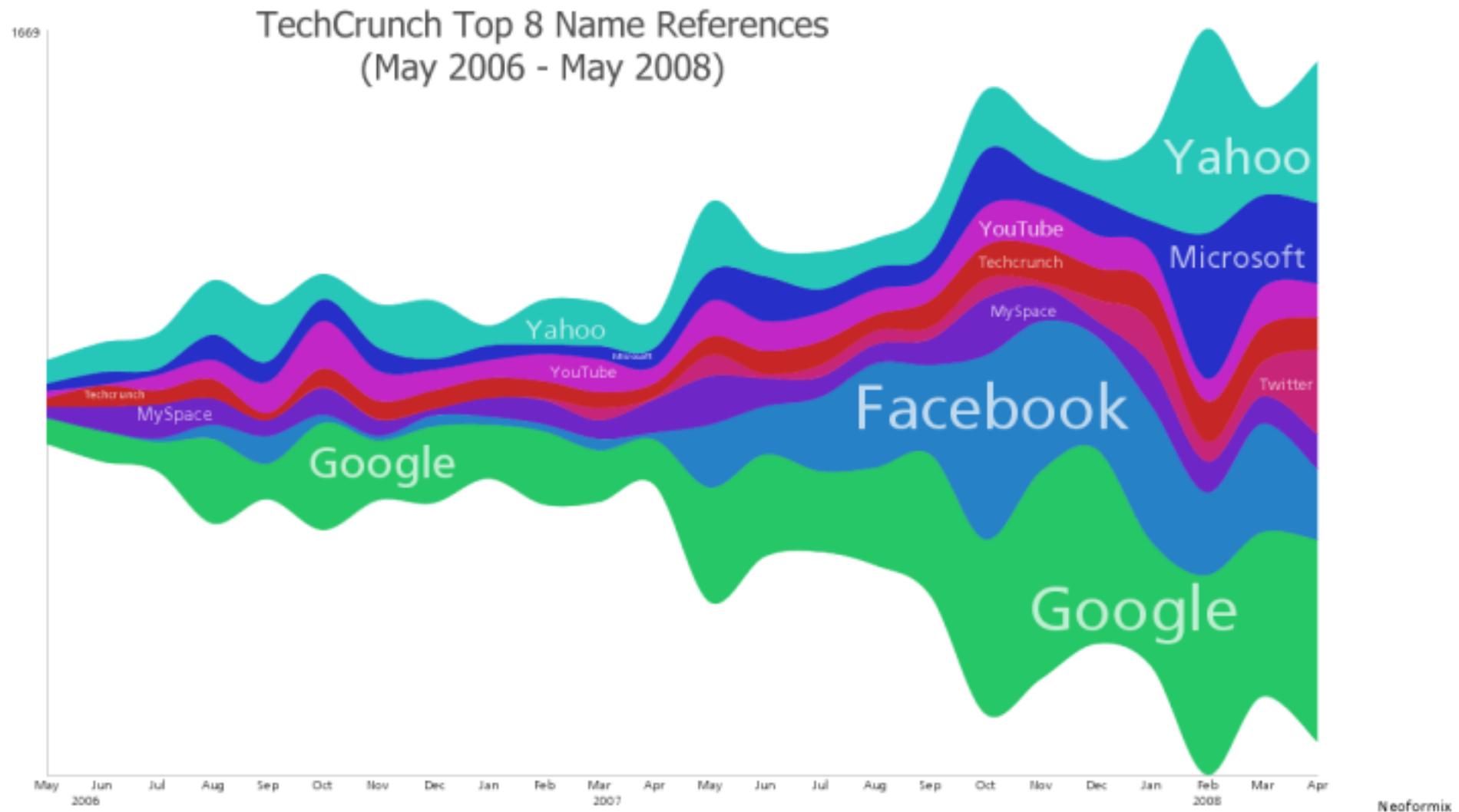
Plošný skládaný graf

- *stacked area chart*, vývoj hodnoty a porovnává mezi kategoriemi



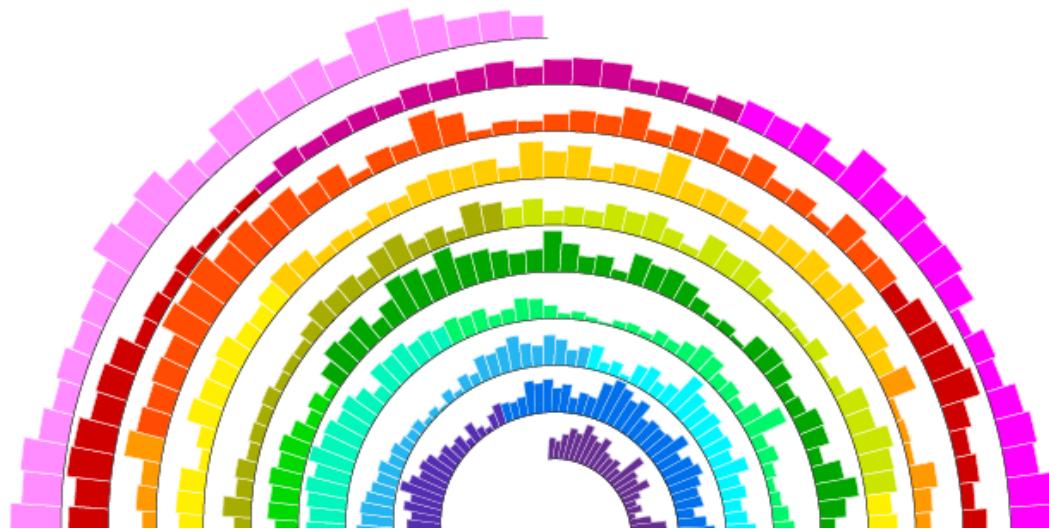
- hodnota je vizualizována výškou plochy

(Streamgraph)

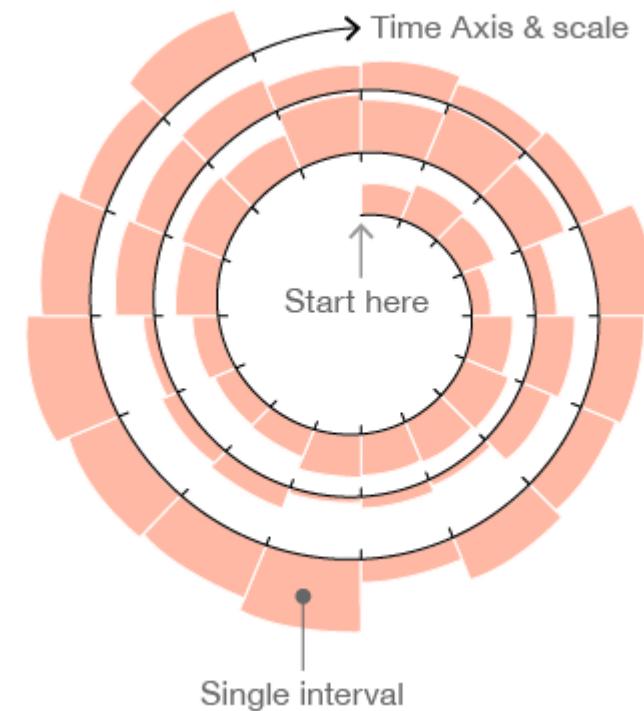


(Spirálový graf)

- *spiral plot*, hodnoty proměnné měnící se v čase
- hodnoty jsou shlukované do časových intervalů – sloupce
- časová osa je spirálová – úspora místa

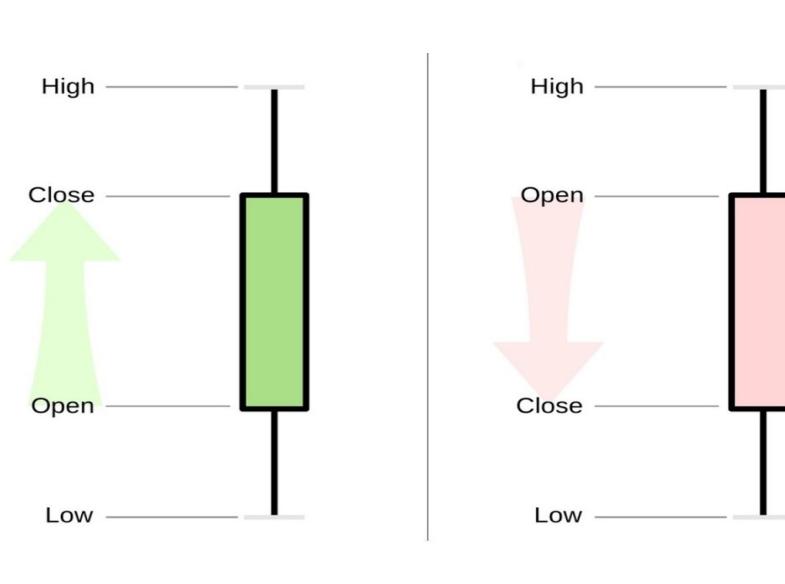


- horší porovnávání hodnot



Svíčkový graf (burzovní grafy)

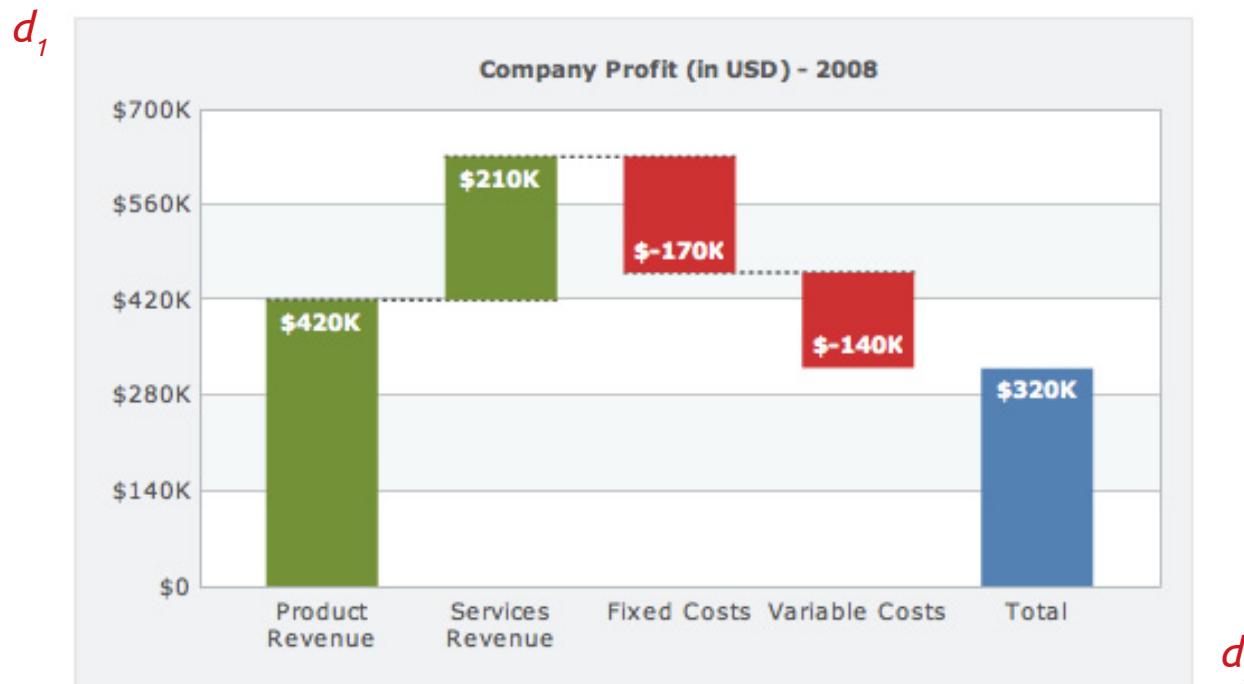
- **candlestick chart**, sledujeme vývoj proměnné v čase
- hodnoty jsou shlukované do časových intervalů - **svíčky** grafu



- d_1 : numerická – relace: (**open, close, low, high**) – agregace
- d_2 : časové intervaly (např. týdenní, denní, 4-hodinové, hodinové, minutové, ...)
- d_3 : kategorická – přírůstek/úbytek – barevné např. odlišení

Vodopádový (sloupcový) graf

- *waterfall bar chart*, sledujeme **vývoj** nějaké proměnné



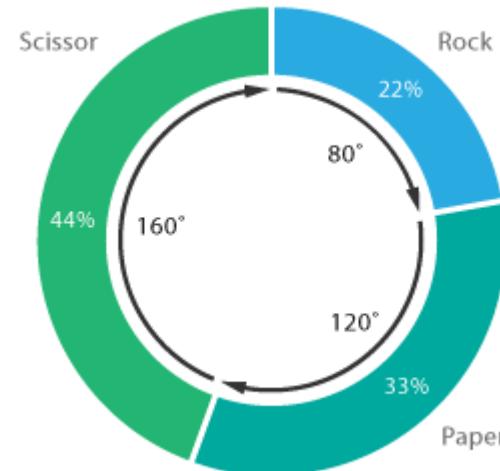
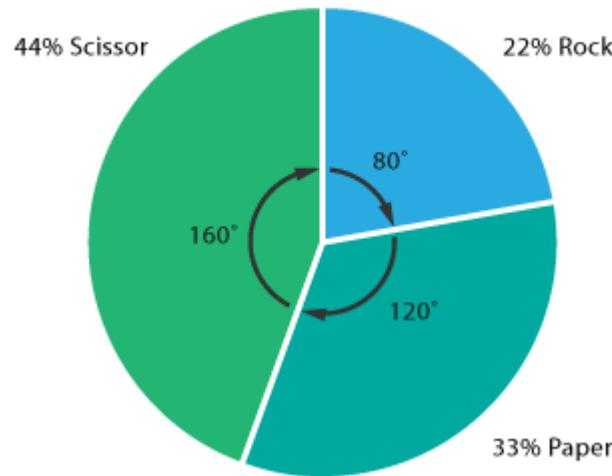
- d_1 : numerická – reprezentuje hodnoty přírůstku nebo úbytku
- d_2 : kategorická
- d_3 : kategorická – přírůstek/úbytek – barevné např. odlišení

Hodnota v rozsahu
Porovnání, ranking
Distribuce hodnot
Korelace
Vývoj
Část celku, hierarchie
Relace, flow
Mapy, geovizualizace

Část celku, hierarchie

Výšečový/prstencový graf

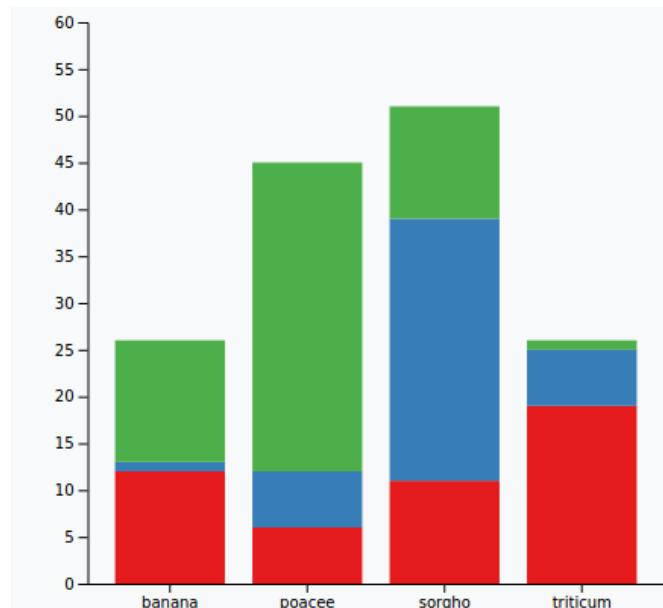
- *pie/donut chart*, porovnávání hodnot k celku



- d_1 : numerická – podíl, hodnota (podíl je vypočítán)
- d_2 : kategorická
- $(+d_3, \dots)$: skupiny kategorií – např. barevně odlišených

Skládaný sloupcový graf

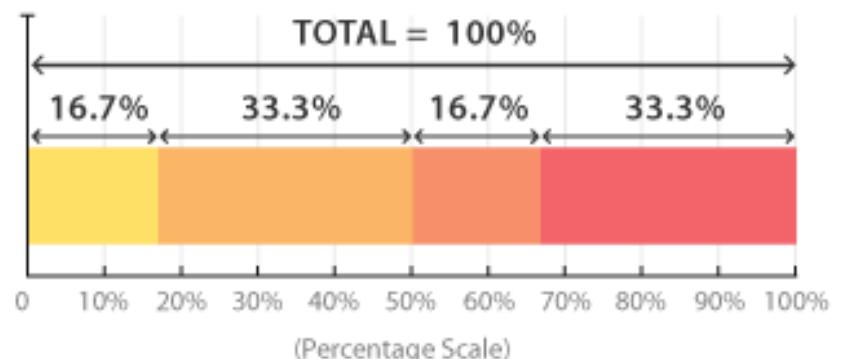
- *stacked bar chart*, porovnávání hodnot k celku a mezi sebou



Simple



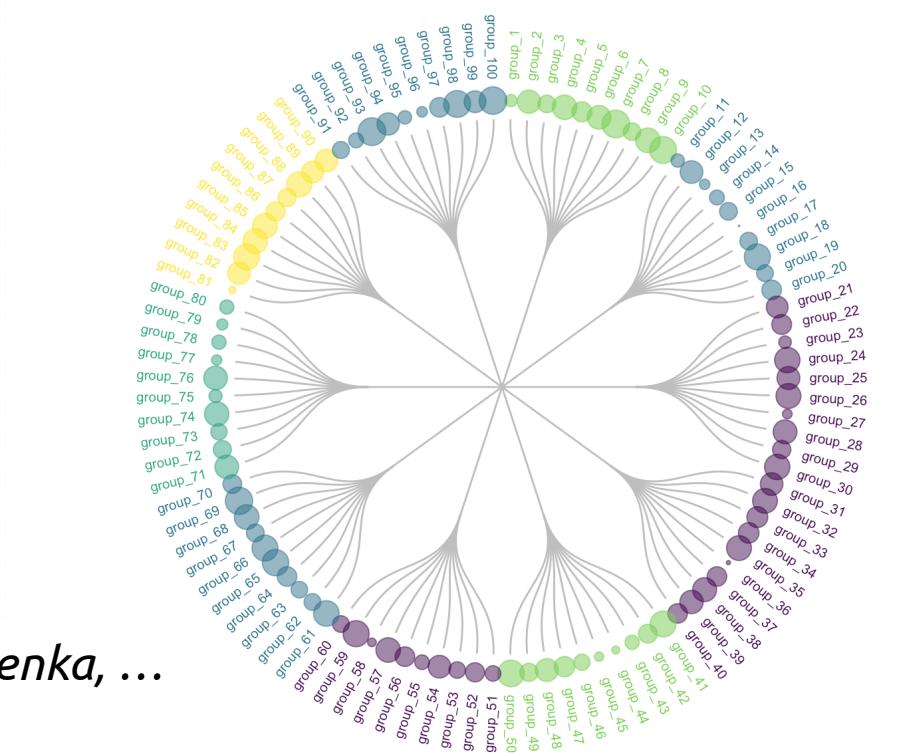
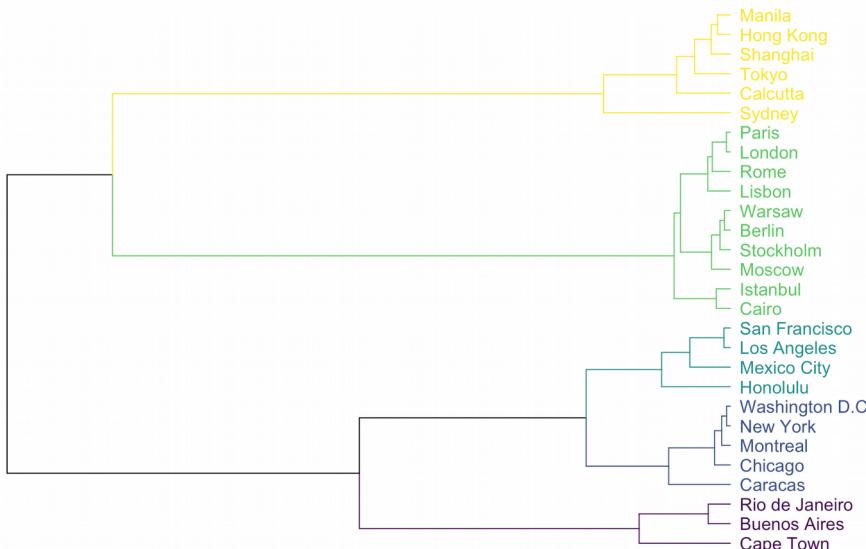
100%



- d_1 : numerická – *podíl, hodnota*
(podíl je vypočítán)
- d_2 : kategorická
- d_3 : kategorická

Dendrogram

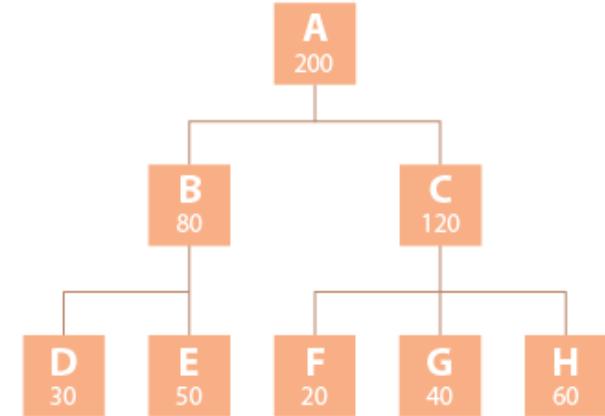
- zobrazení **hierarchie**, využívá se pro *myšlenkové mapy*
například aplikace FreeMind



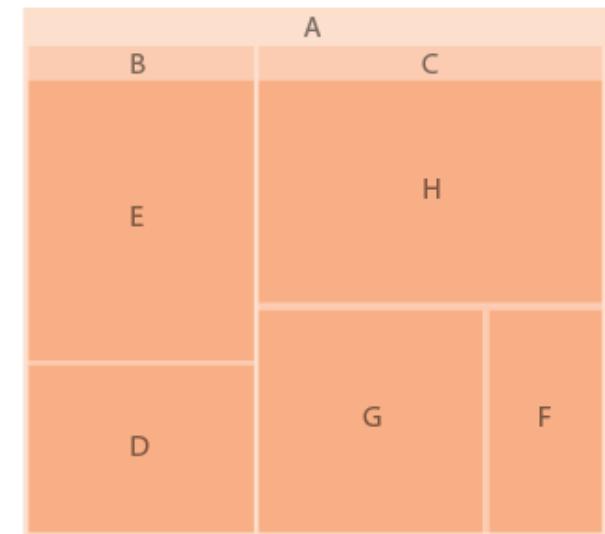
- d_1 : kvalitativní hodnota – *text, myšlenka, ...*
- d_2 : kategorická – *rodič*
- $(+d_3, \dots)$: skupiny kategorií – *např. barevně odlišených*
- další příklady: [1], [2]

Stromová mapa

- **treemap, porovnávání hierarchických hodnot k celku**
například pro vizualizaci zaplnění souborového systému

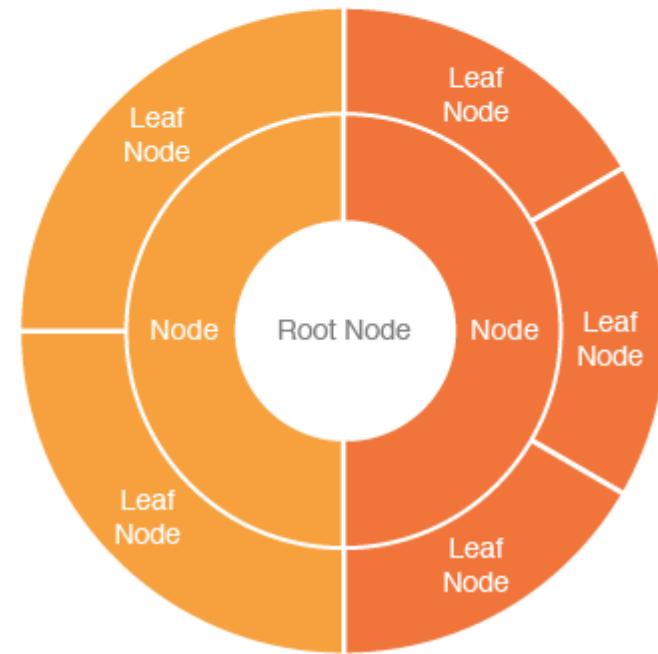
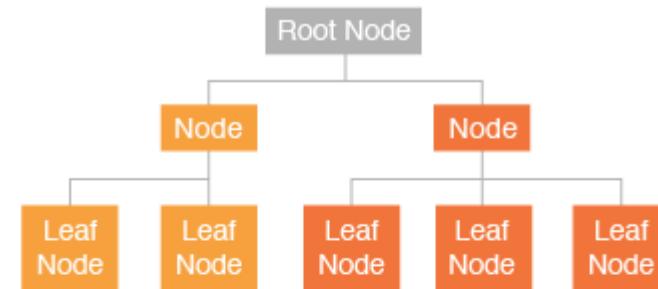


- d_1 : numerická – poměr/hodnota
- d_2 : kategorická – pojmenování
- d_3 : kategorická – rodič (hierarchie)
- $(+d_4, \dots : skupiny kategorií – např. barevně odlišených)$



(Vícevrstvý prstencový graf)

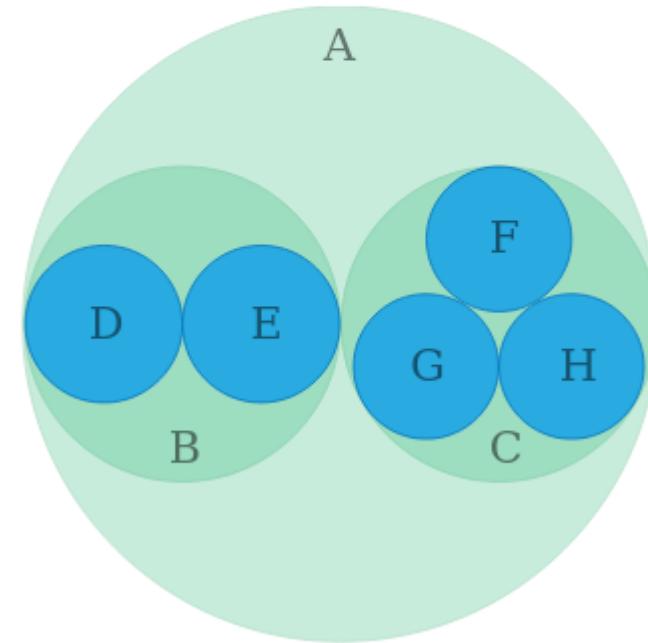
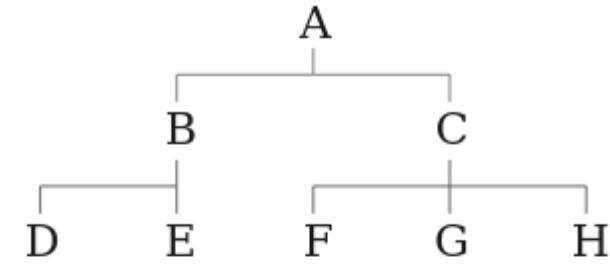
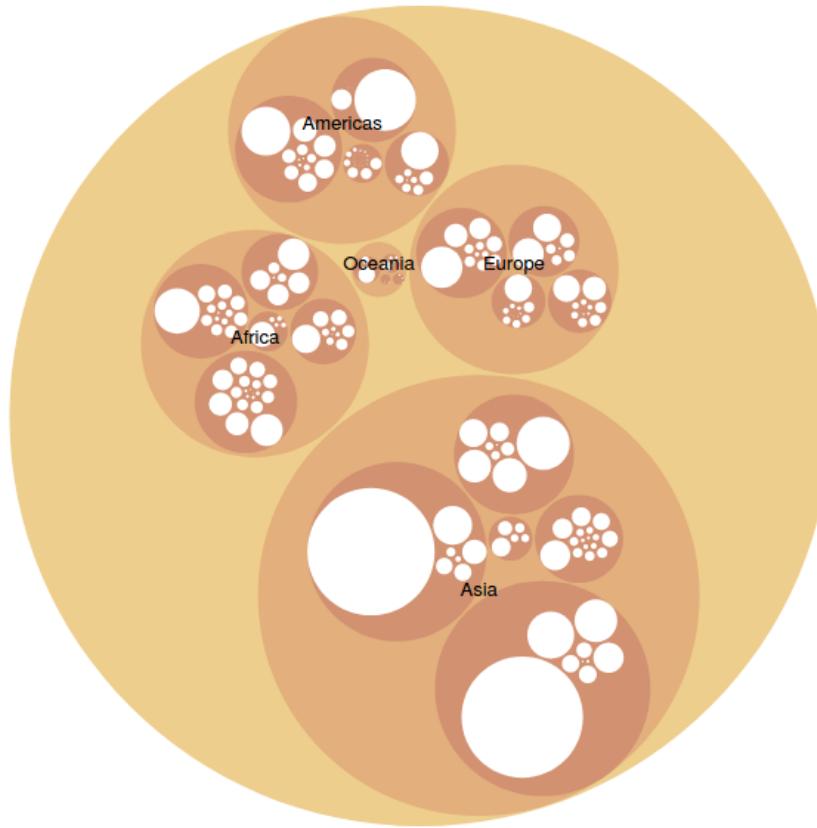
- *sunburst chart*, porovnávání **hierarchických** hodnot k celku



- d_1 : numerická – *poměr/hodnota*
- d_2 : kategorická – *pojmenování*
- d_3 : kategorická – *rodič (hierarchie)*
- $(+d_4, \dots : skupiny kategorií – např. barevně odlišených)$

(Circular packing)

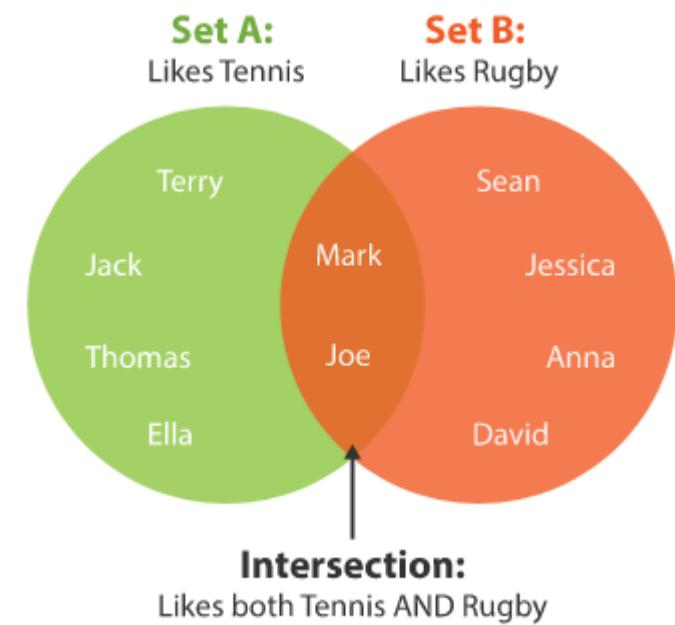
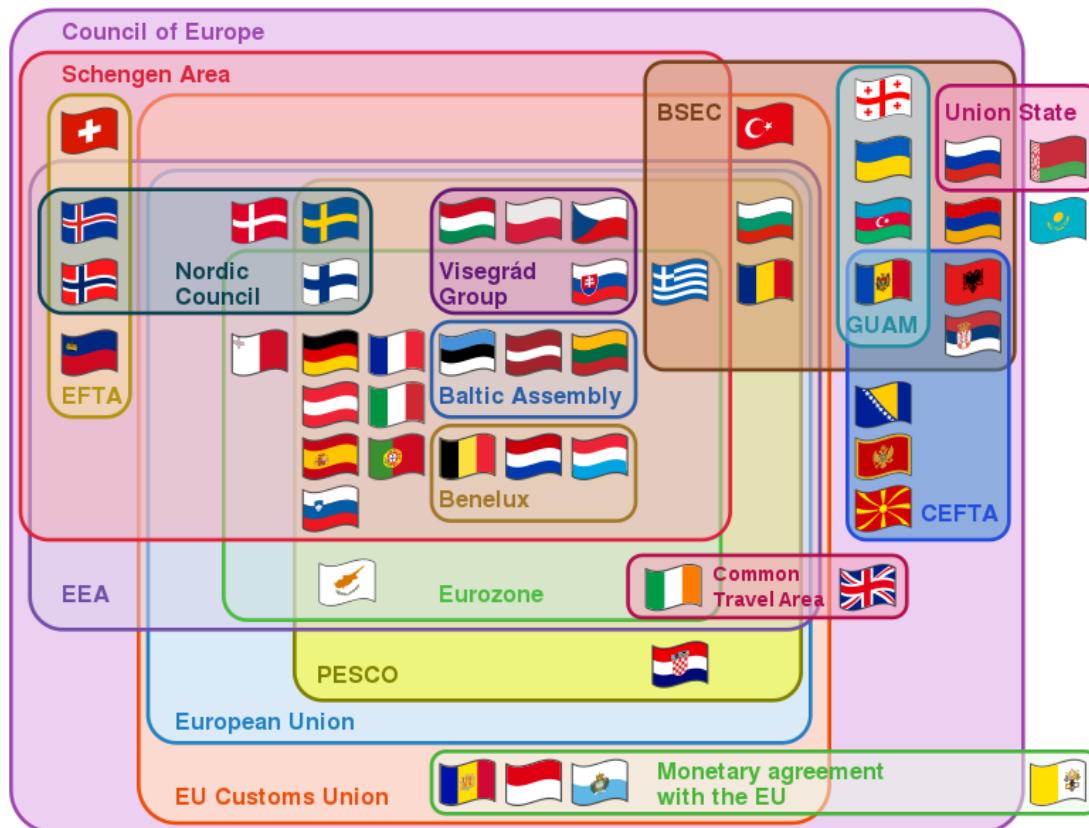
- porovnávání **hierarchických** hodnot k celku



- podobný princip jako treemap

Venův diagram

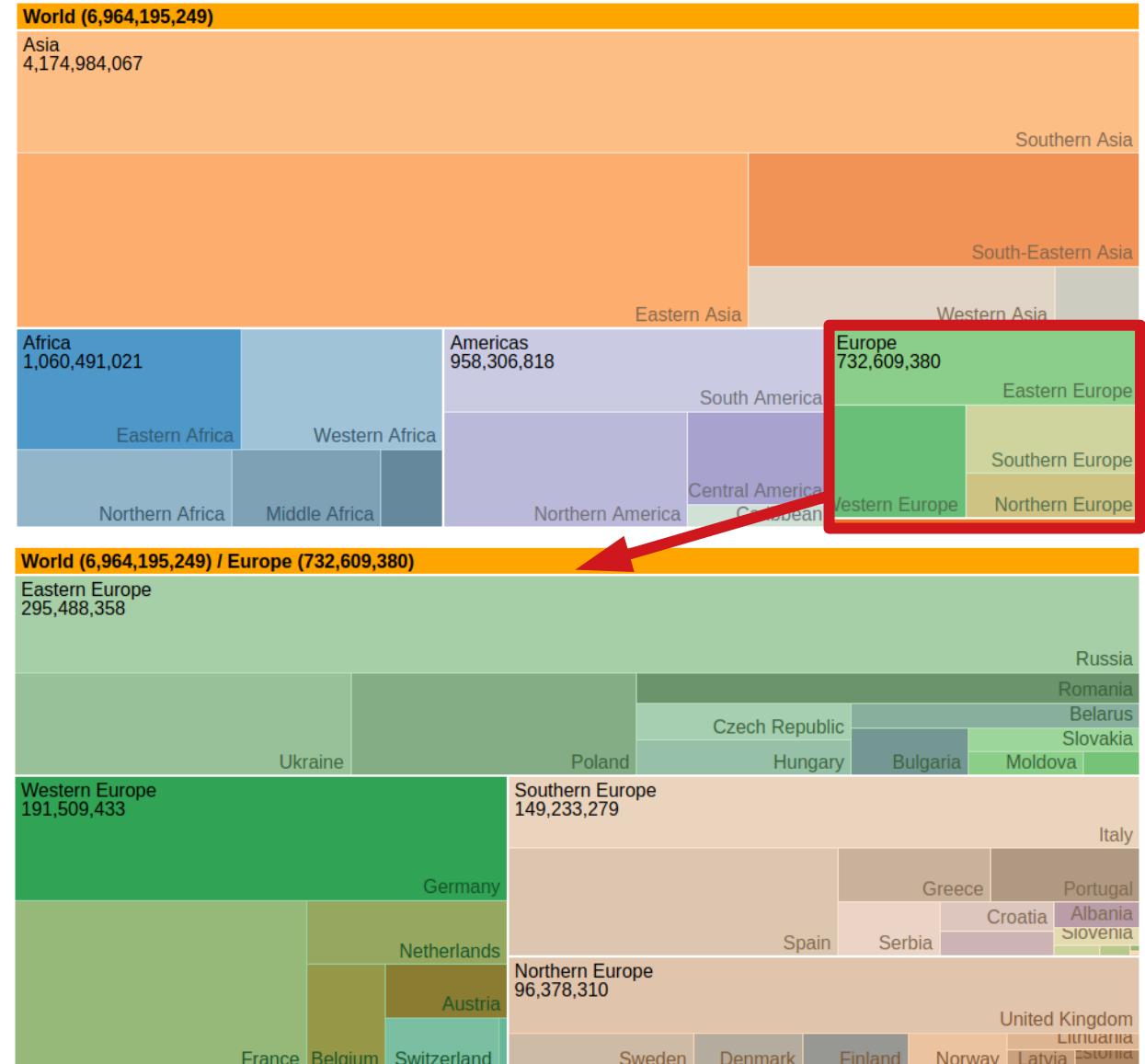
- *Venn diagram*, vizualizace příslušnosti prvků v množině



- d_1 : kvalitativní – *text, ...*
- d_2 : kategorická – *rodič (množina)*
- d_3 : skupiny kategorií – *barva, ikony, ...*

Zanořování

- využití akce ***drill-down*** pro zobrazení dalších dimenzí
- vyžaduje interakci
- uživatel nevidí všechny data pohromadě
- výhodné pro vizualizaci hierarchie

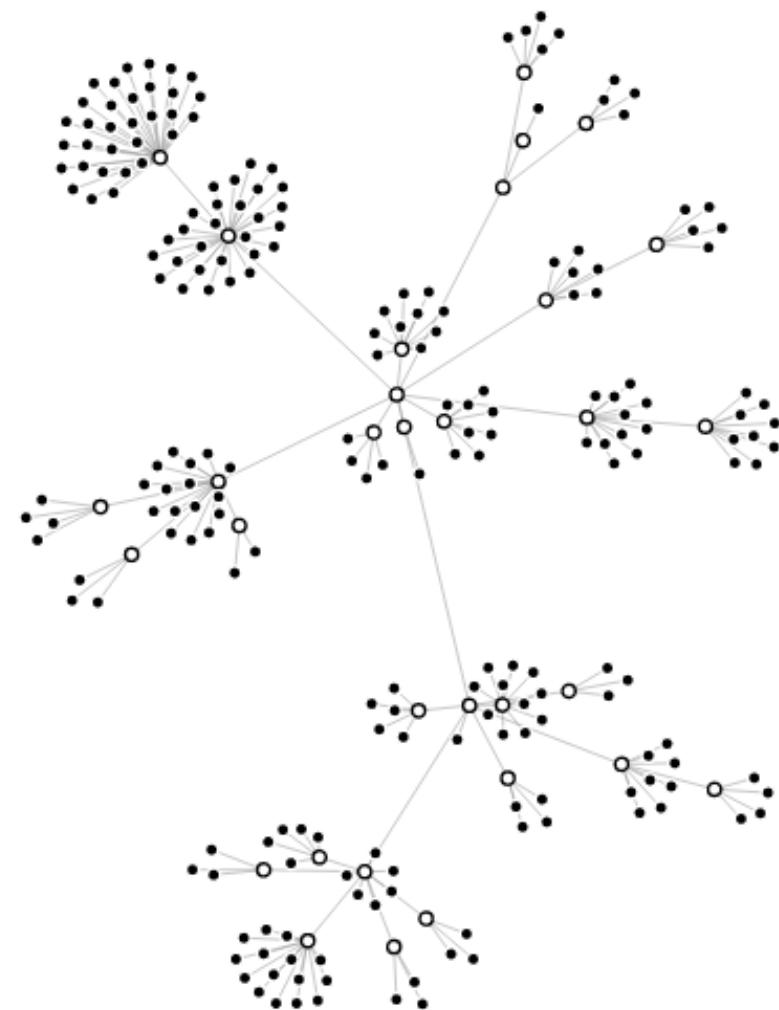
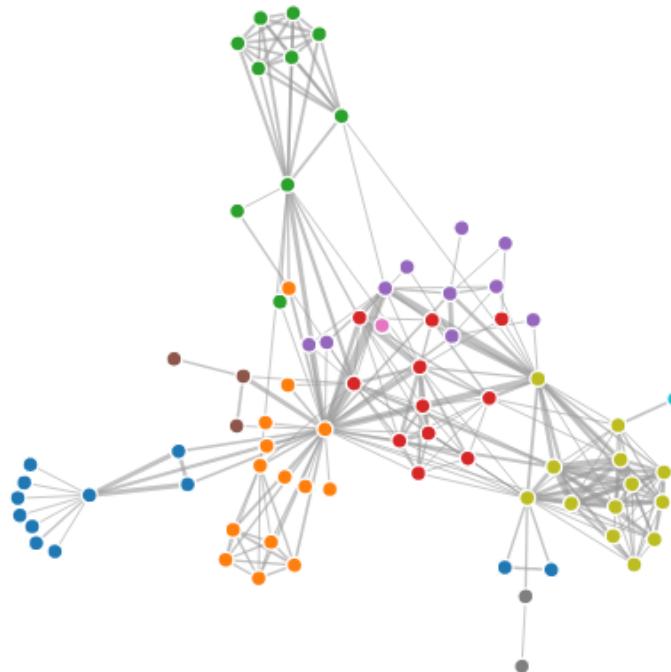


Hodnota v rozsahu
Porovnání, ranking
Distribuce hodnot
Korelace
Vývoj
Část celku, hierarchie
Relace, flow
Mapy, geovizualizace

Relace, flow

Sít

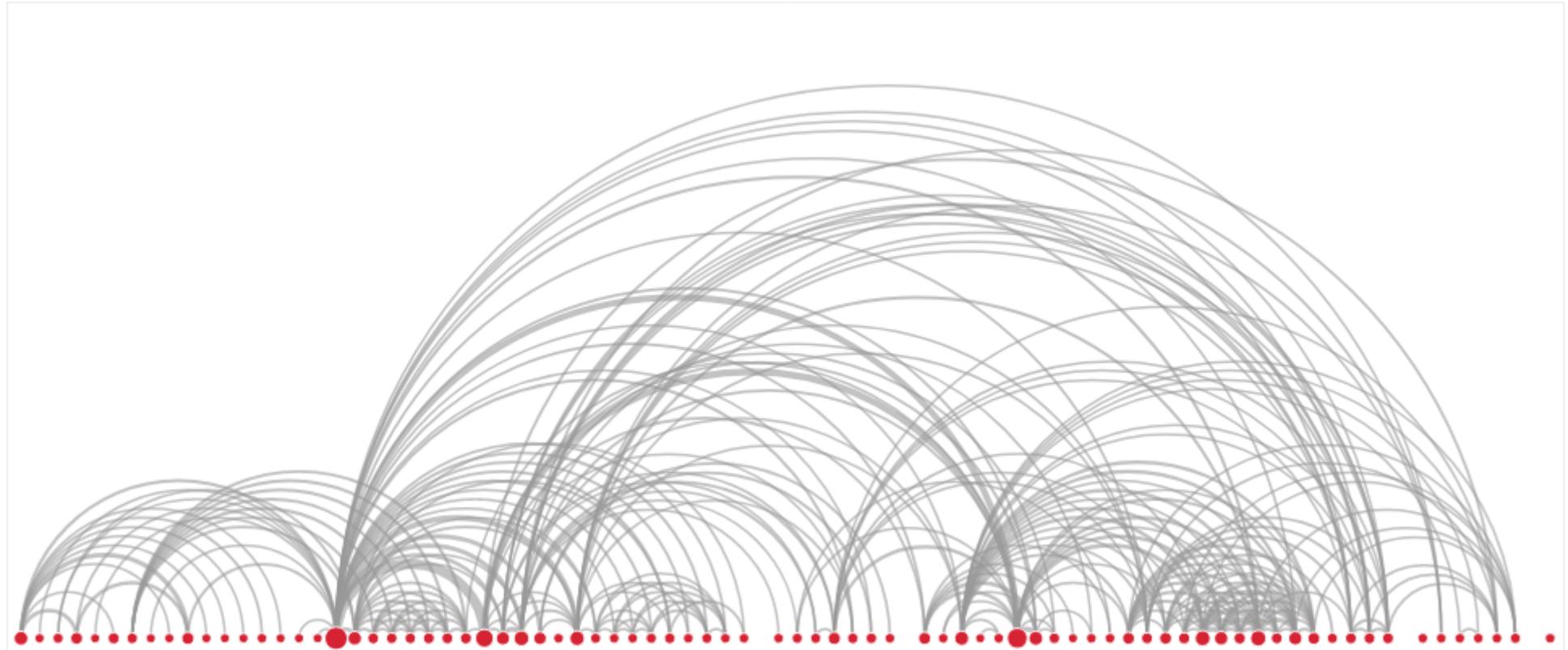
- **network**, vizualizuje vztahy mezi uzly



-
- d_1 : zdroj (kategorická)
- d_2 : cíl (kategorická)
- $(+d_3, \dots)$: skupiny kategorií
 - např. barevně odlišených

(Arc diagram)

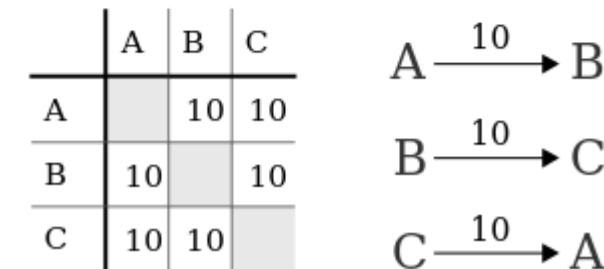
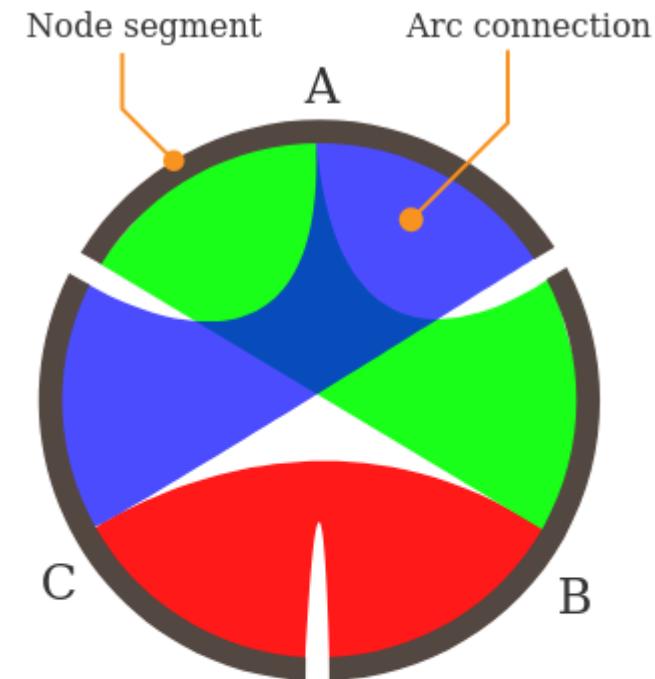
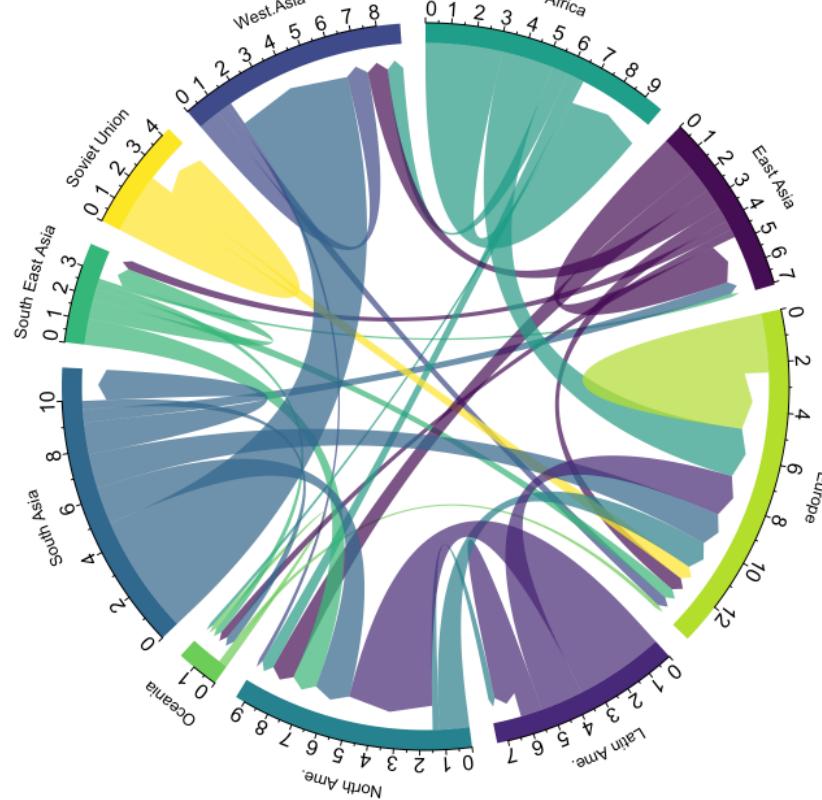
- vizualizuje vztahy mezi uzly



- prime numbers pattern: <https://www.jasondavies.com/primos/>

Chord

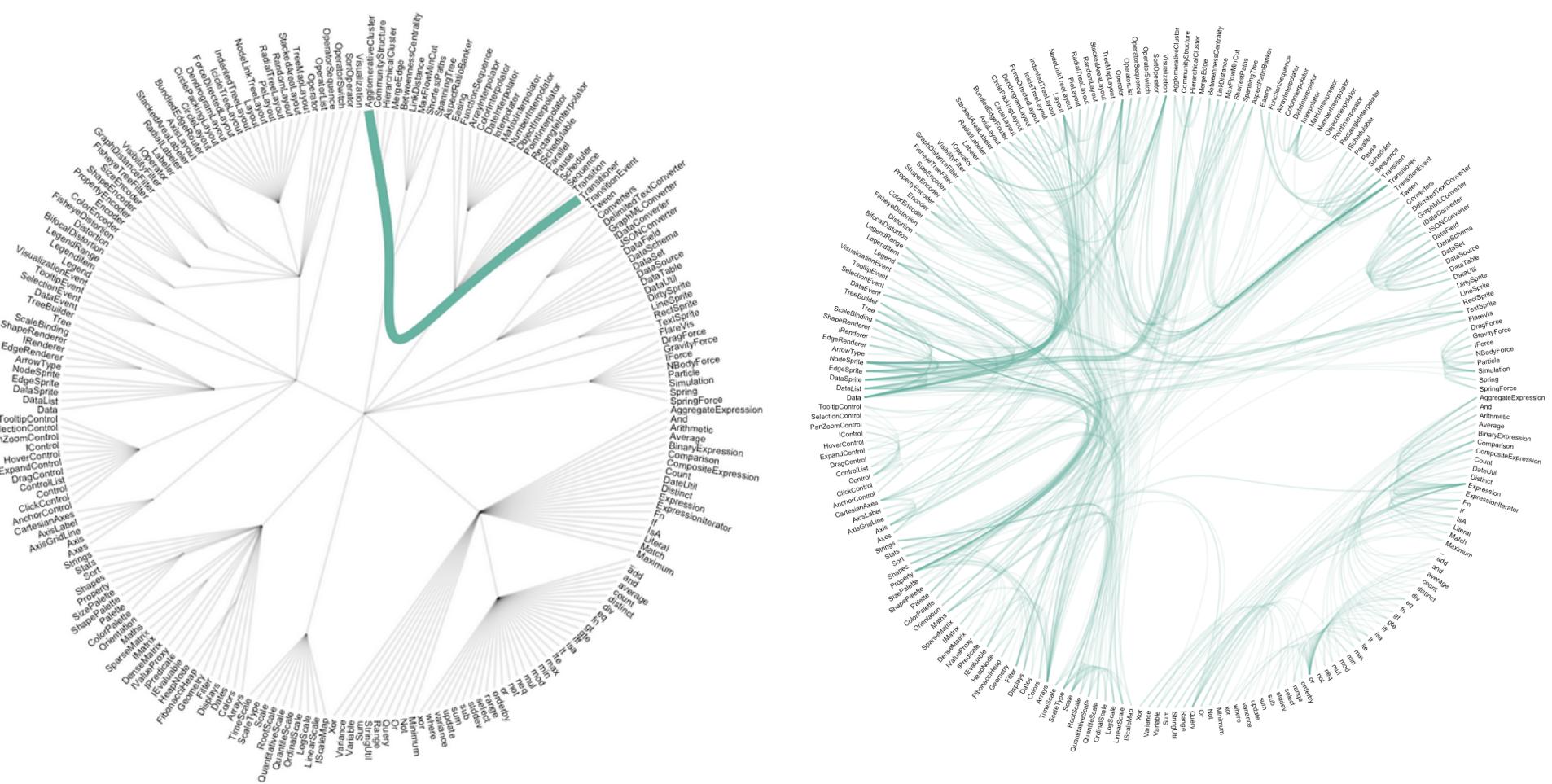
- vizualizuje vztahy mezi uzly



- $d_{1,2}$: kategorická – zdroj, cíl
- $d_{3(4)}$: hodnota (zdroje a cíle) (+ $d_{5, \dots}$: skupiny kategorií, ...)

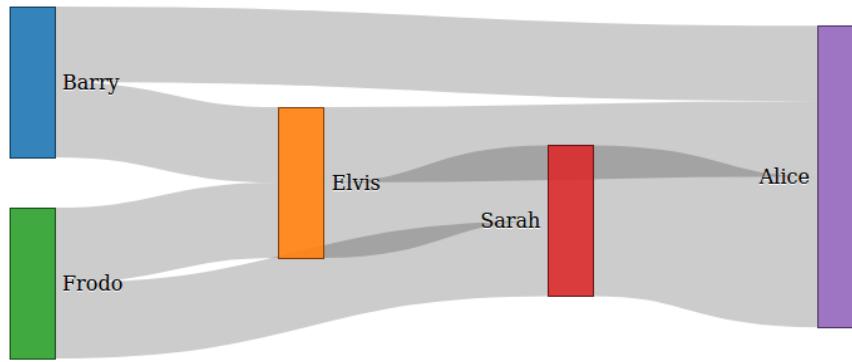
(Hierarchical Edge bundling)

- vizualizuje vztahy mezi uzly
- spojuje sousední hrany a redukuje množství průsečíků

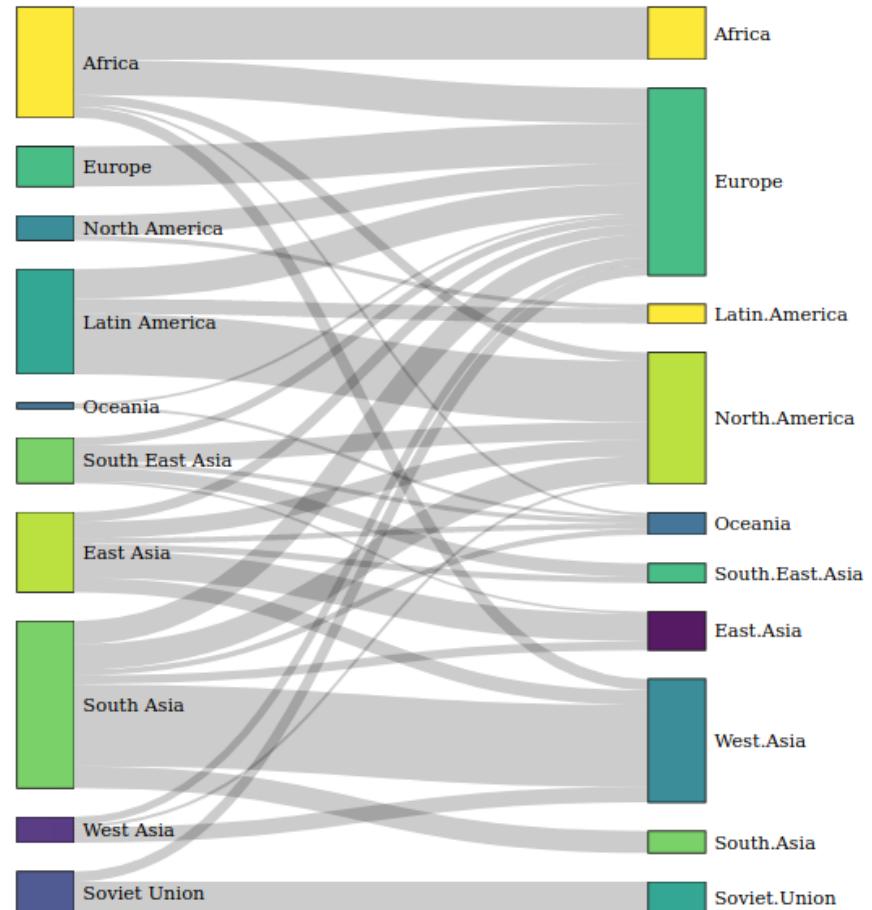


Sankey diagram

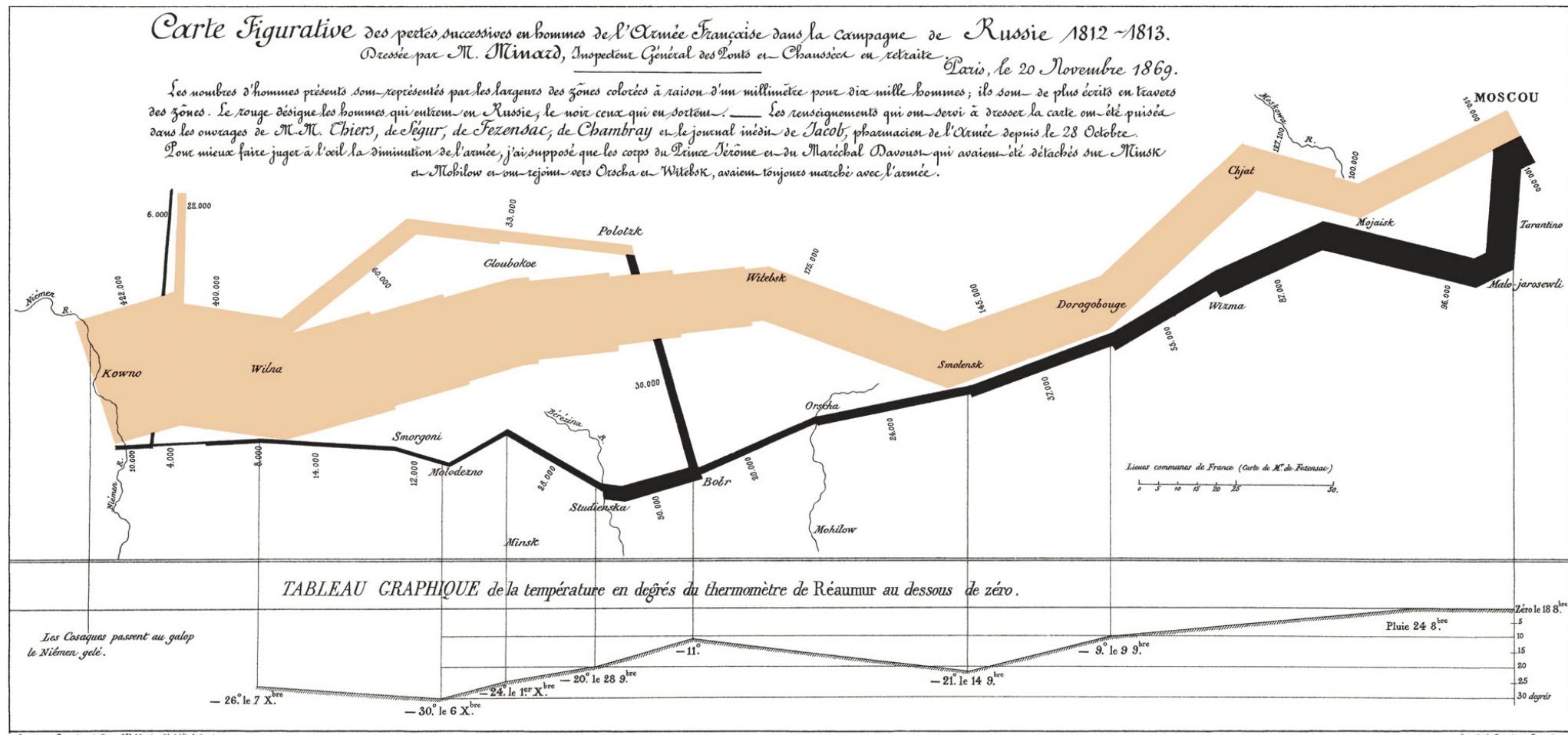
- vizualizuje vztah mezi uzly, flow



- $d_{1,2}$: kategorická – zdroj, cíl
 - $d_{3(4)}$: hodnota (zdroje a cíle)
 - $(+d_5, \dots)$: skupiny kategorií, ...)



Sankey diagram



Datová mapa ilustrující zlý osud Napoleonovy armády při tažení do Ruska

Hodnota v rozsahu
Porovnání, ranking
Distribuce hodnot
Korelace
Vývoj
Část celku, hierarchie
Relace, flow
Mapy, geovizualizace

Mapy, geovizualizace

Mapa

- vizualizuje geografická data



- $d_{1,2}$: zeměpisná šířka/délka (*latitude, longitude*)

Mapa obsahující značky

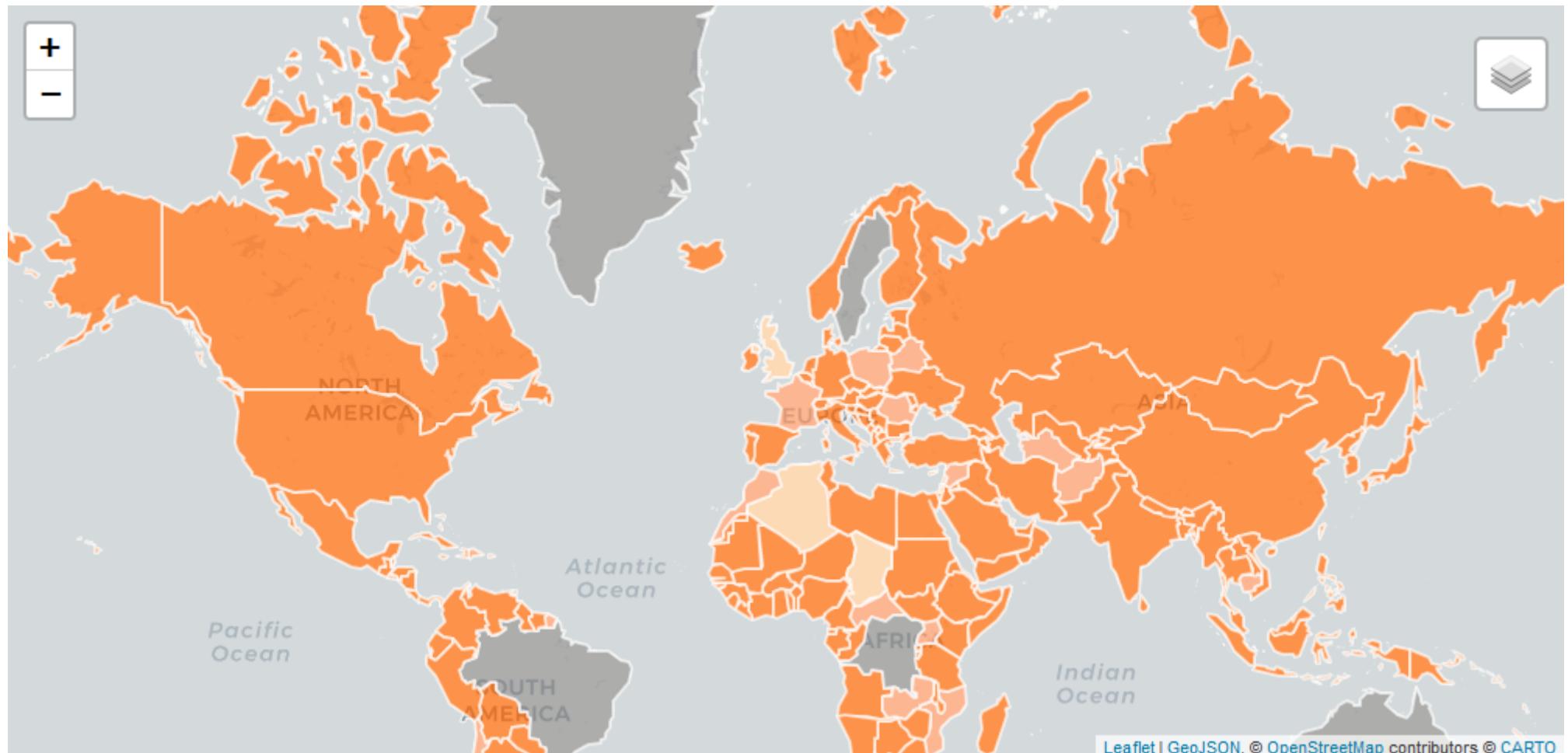
- vizualizuje body zájmu v mapě a případné hodnoty



- d_3 : kategorická – např. popis bodu zájmu apod. (+ d_4, \dots : numerická, ...)

Kartogram

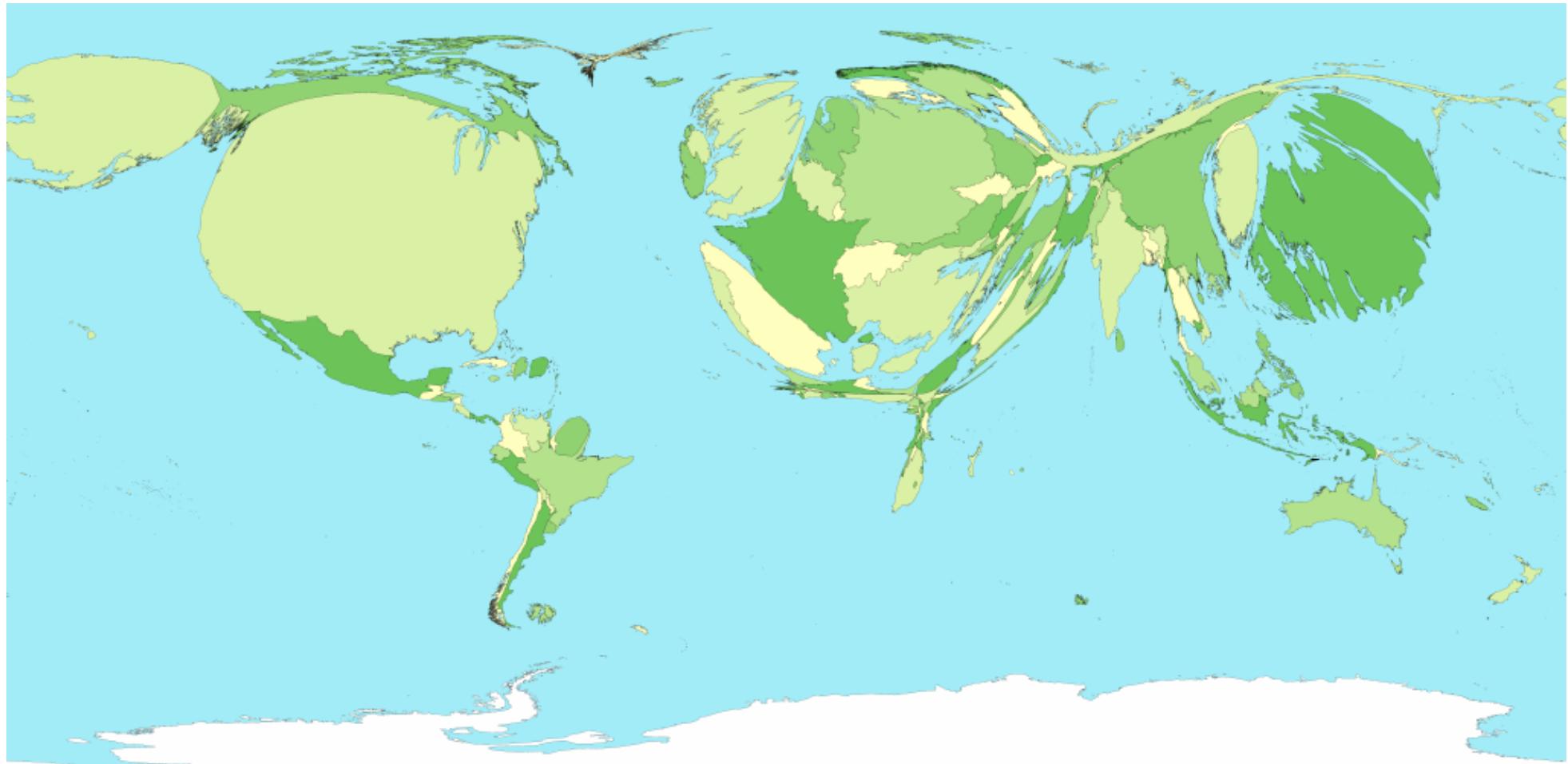
- *choropleth*, vizualizuje hodnoty vztažené k regionům



- definice polygonu – geoJson

(Plošně anamorfovaná mapa)

- *cartogram*, vizualizuje hodnoty vztažené k regionům



- polygony jsou anamorfované na základě hodnoty

Connection map

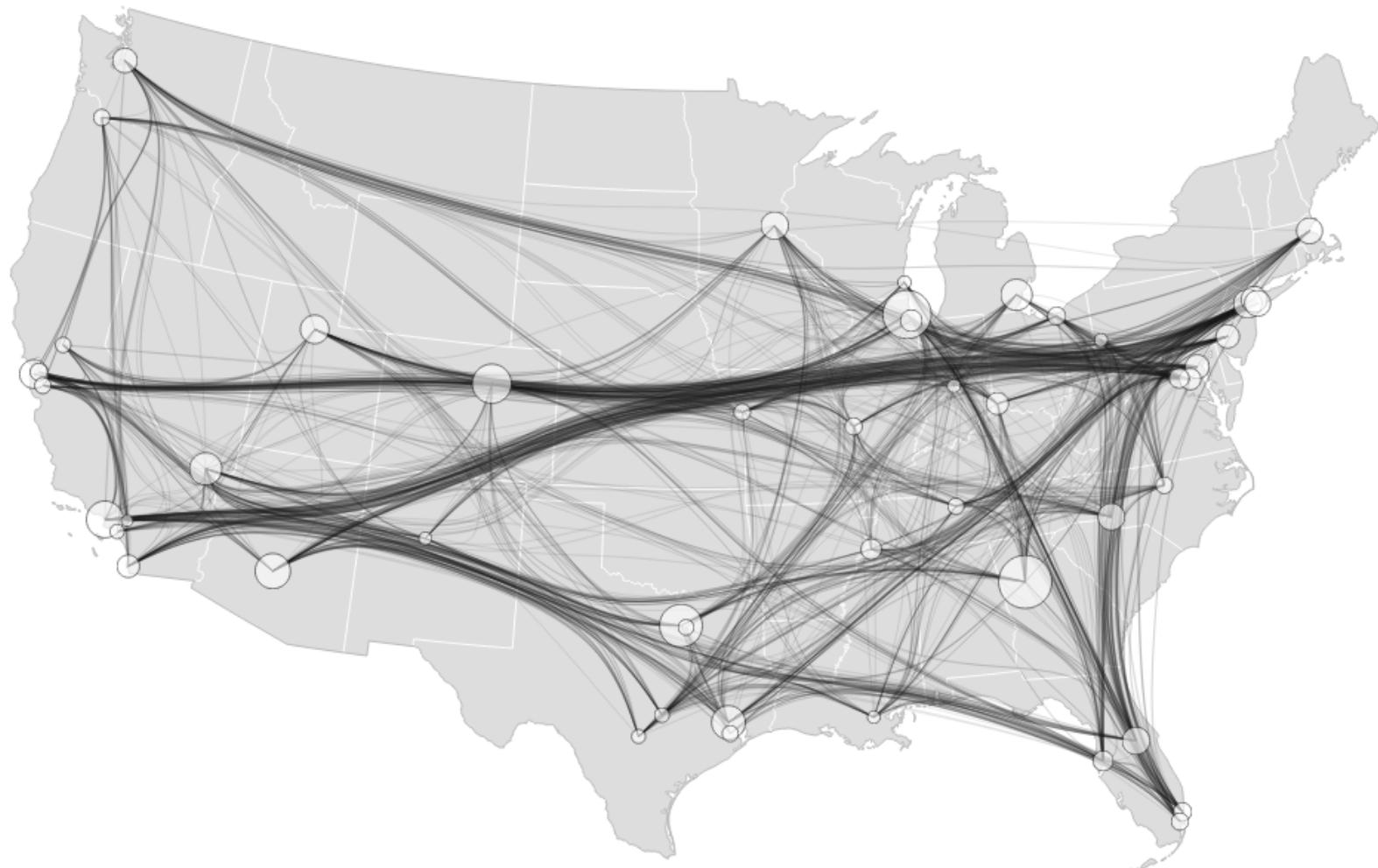
- vizualizuje vztahy mezi uzly na mapě



- nutné řešit velké množství průsečíků (přehlednost)

(Edge bundling map)

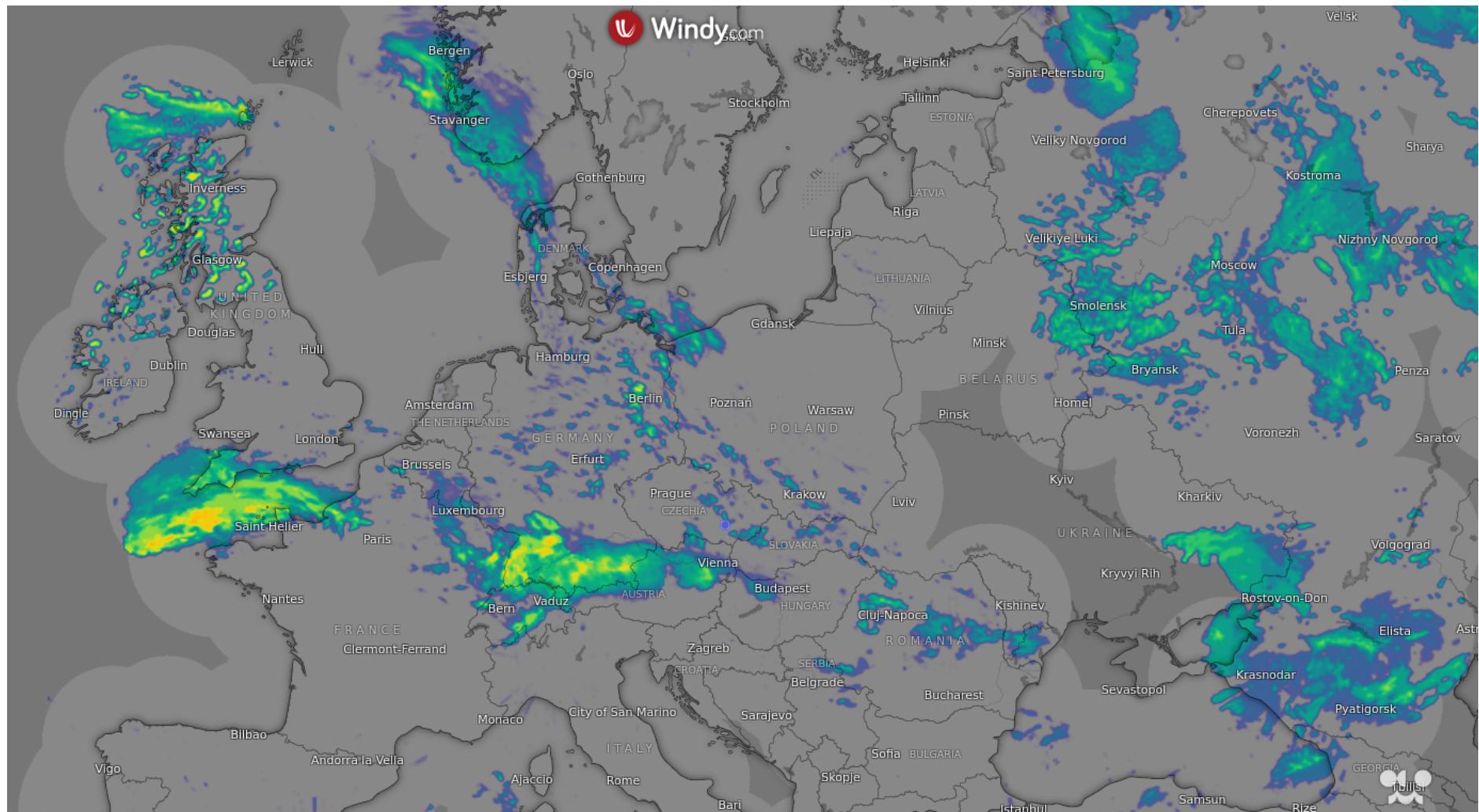
- aplikace edge bundling v mapě



- výpočetně náročné

Geografická teplotní mapa

- hodnoty teplotní mapy s geografickými souřadnicemi



Děkuji za pozornost!