

# PV251 Vizualizace

Jaro 2016

Výukový materiál

---

## 10. přednáška: Návrh efektivních vizualizací

Cílem této přednášky je poskytnout určité návody pro návrh efektivních vizualizací.

### Efektivní vizualizace

Vizualizaci označujeme za úspěšnou, jestliže efektivně a přesně sděluje zamýšlenou informaci cílovému publiku a přitom bere v úvahu účel vizualizace (prozkoumání, potvrzení hypotézy, prezentace, ...).

Pro danou vstupní množinu dat existuje nesčetně mnoho možných metod pro mapování datových komponent na grafické entity a atributy. Podobně existuje celá řada interaktivních nástrojů, které může uživatel využívat. Je zřejmé, že výběr vhodné kombinace technik mapování a interakce není jednoduchý a přímočarý proces.

Vizualizace může být **neefektivní** z mnoha důvodů. Například může být příliš matoucí či složitá pro interpretaci cílovým publikem. Nebo některá data mohou být nežádoucím způsobem deformována, ořezána nebo ztracena během procesu mapování. Další známky nekvalitní vizualizace jsou nedostatek podpory pro možnost změny pohledu na data nebo nemožnost ovlivnit použitou barevnou škálu. Dalším aspektem, který výrazně ovlivní úspěšnost dané vizualizace, je i estetická složka. Vizuelně ne příliš přitažlivá prezentace může významně ovlivnit ochotu publika tuto vizualizaci vůbec sledovat. V každém z uvedených případů dochází k interferenci některé z komponent vizualizace se sdělením informace uživateli.

Pokusíme se tedy o prezentování různých možných návrhů komponent, které jsou považovány za nezbytné pro kvalitní vizualizace. Zároveň se zaměříme na některé běžné problémy, se kterými se setkáme při procesu vizualizace a navrhujeme techniky, abychom se těmito problémům vyhnuli.

Říká se, že je mnohem jednodušší navrhnout špatnou vizualizaci než tu kvalitní. Proto je cílem této kapitoly zejména pomoci návrhářům vizualizací udělat dobrá návrhová rozhodnutí, která povedou k efektivním vizualizacím. Tato oblast je tak rozsáhlá, že samozřejmě nepokryjeme veškeré používané techniky. Nicméně na toto téma vyšla řada zajímavých knih, které by zájemcům o tuto oblast neměly uniknout.

## Hlavní kroky při návrhu vizualizace

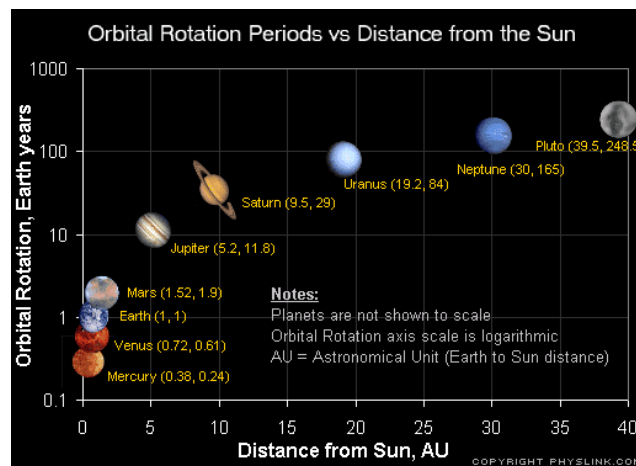
Vytváření vizualizací zahrnuje rozhodování, jakým způsobem mapovat datová pole na grafické atributy, dále výběr a implementaci metod pro změnu pohledu a rovněž výběr množství dat pro vizualizaci. Další nezbytnou součástí vizualizace je zobrazení doplňkových informací, jako například označení (labels) nebo použité barvy. Poslední součástí, která je hodně subjektivní, je celková estetika výsledného zobrazení.

Nyní se zaměříme na tyto jednotlivé fáze návrhu efektivní vizualizace a na problémy, které by měl návrhář v těchto fázích řešit.

## Intuitivní mapování dat na vizualizaci

Abychom mohli vytvořit co nejeftivnější vizualizaci pro danou aplikaci, je nezbytné vzít v úvahu sémantiku dat a kontext typického uživatele. Interpretaci výsledného obrázku posílíme rovněž tím způsobem, že vybereme vhodný způsob mapování dat na grafické entity, který zcela vyhovuje mentálnímu modelu uživatele v dané doméně. Dále, čím více je návrhář konzistentní při předpovídání očekávání uživatele, tím je menší riziko dezinterpretace.

Intuitivní mapování vedou k rychlejší interpretaci. Příkladem je obrázek, který znázorňuje planety a vztah jejich vzdálenosti ke Slunci a dobu jejich oběhu kolem Slunce.



Pravděpodobně nejběžnějším způsobem mapování, které je při vizualizaci používáno, je mapování prostorových atributů dat (jako například zeměpisná délka či šířka) na pozice na obrazovce. Některé z prvních vizualizací využívaly lidské schopnosti určit vztah mezi pozicí vykreslenou na výstupním zařízení a odpovídající pozicí ve 3D světě. Podobně s příchodem animace došlo k intuitivnímu a přirozenému způsobu zobrazení dat měnících se v čase, včetně možnosti ovládat rychlost animace.

Další mapování se stávají intuitivními při spojení s příslušným kontextem. Například velmi běžné je mapování teploty na barvu, protože většina kultur asociuje vysoké teploty s červenou či bílou barvou. Barva má specifickou interpretaci i v dalších oblastech, jako

například kartografie či geologie. Aplikační doména vizualizace tedy významně ovlivňuje použití barevného atributu.

Výška, nebo alternativně délka čáry, je jiný způsob reprezentace teploty v případě, že teplotu vyčítáme z teploměru. Délka je rovněž velmi přirozeně používána v medicíně pro zobrazení krevního tlaku či jiných skalárních hodnot.

### **Výběr způsobu mapování**

Při výběru způsobu mapování je jedním z důležitých aspektů kompatibilita mezi škálováním datových polí a škálováním odpovídajících grafických entit nebo atributů. Pro uspořádané datové atributy (jako například věk) není vhodné použít grafické atributy, které nejsou uspořádané, jako například tvar. Podobně neuspořádané datové atributy (jako například místo narození) by neměly být mapovány na uspořádané grafické atributy, jako je například délka.

Přesto je v některých případech zajímavé prozkoumat data za použití ne zcela intuitivního mapování pro daný případ, což může občas odhalit nečekané zajímavé vlastnosti dat. Například mapování času na barvu podél paprsků může odhalit rozdíly v rychlosti částic, které by jinak bylo obtížné detekovat.

Proto je dobrým pravidlem nastavení defaultního mapování co nejintuitivněji, aby odpovídalo požadavkům běžného uživatele. Zároveň však, zejména při prozkoumávání dat, je vhodné umožnit uživateli měnit různá nastavení a tím měnit výsledný vzhled dle jeho aktuálních požadavků.

### **Výběr a modifikace pohledů**

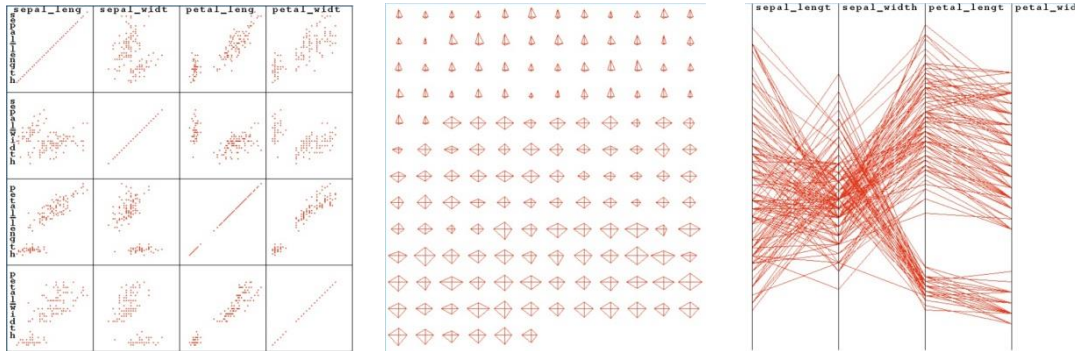
S výjimkou velmi jednoduchých datových množin je použití jednoho pohledu na data zřídka dostatečné pro prozkoumání všech vlastností a vztahů obsažených v těchto datech. Klíčem pro vytváření efektivních vizualizací je schopnost odhadnout způsoby pohledů a změny pohledů, které používá většina typických uživatelů, a poskytnout intuitivní ovládací nastavení a úpravu pohledů na základě požadavků uživatele.

Každý podporovaný pohled by měl být jasně označen a výběr nového pohledu by měl vyžadovat minimální akce ze strany uživatele.

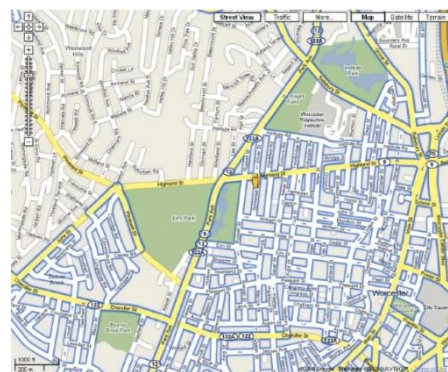
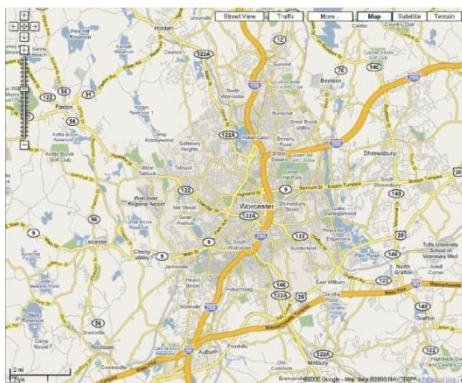
Změny pohledu spadají do několika kategorií a jejich zahrnutí do základní funkcionality by mělo odrážet priority uživatele. Tyto kategorie jsou:

- Operace **rolování (posouvání) a zoomování** jsou nutné v případech, kdy nelze celou datovou množinu zobrazit uživateli v požadovaném rozlišení.

- **Modifikace barevné palety** je požadována téměř vždy – minimálně je požadována možnost používat různé barevné palety. Výhodou je rovněž možnost ovlivňovat buď jednotlivé barvy, nebo celé palety.
- **Modifikace mapování** umožňuje uživateli přepínat mezi různými způsoby vizualizace stejných dat. Vlastnosti dat, které jsou v jednom mapování skryty, se objeví v jiných (viz obrázek – matice bodových grafů, star glyphs, paralelní souřadnice).



- **Řízení škálování** dovoluje uživateli modifikovat rozsah a rozložení hodnot příslušného datového pole ještě před jeho mapováním. Podobně ořezávání a různé formy filtrace umožňují uživateli soustředit se na požadované podmnožiny dat.
- **Řízení level-of-detail (LOD)** poskytuje schopnost eliminovat nebo naopak zvýraznit detaily a podporovat pohledy na různých úrovních abstrakce. Podle typu úkolu pak může uživatel přepínat mezi různými úrovněmi (viz obrázek).

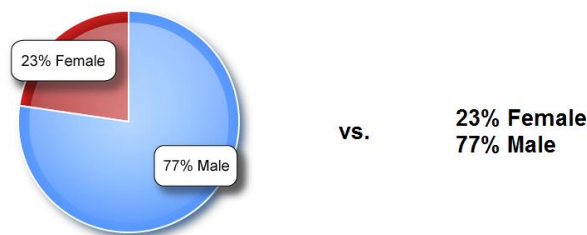


Ve všech případech je zásadní, aby manipulace s pohledy byly intuitivní a uživatelsky lehce zapamatovatelná a aby měly vhodně zvolenou přesnost. Pokud je to možné, je obecně preferována možnost přímé manipulace (změny jsou specifikovány přímo na obrázku).

### Hustota informace – kdy je to moc a kdy ještě málo?

Jedním z klíčových rozhodnutí při návrhu vizualizace je určení, kolik informace zobrazit. Rozlišujeme dvě extrémní situace.

První z nich se objevuje v případě, že máme pouze malé množství informace pro zobrazení. Příkladem jsou případy, kdy je třeba rozlišit mezi pouze dvěma nebo třemi odlišnými hodnotami, jako například procentuální poměr mužů a žen uvnitř daného vzorku. Dalším příkladem je odvození dalších veličin ze základních – například zobrazení dvou čísel, jejich součtu a rozdílu. V takových případech je mnohem efektivnější jednoduše zobrazit tyto kvantitativní hodnoty jako text. Vyžaduje to totiž mnohem méně místa na obrazovce, což je u většiny vizualizací poměrně zásadní. To odpovídá skutečnosti, že **pouze proto, že můžeme vizualizace vytvářet, ještě neznamená, že je musíme vytvářet.**



Druhým extrémem je případ, kdy se snažíme sdělit příliš mnoho informace, což je rovněž častým problémem. Přílišné množství informace může vést ke zmatení a obtížné interpretaci na straně uživatele. Důležitá informace obsažená v datech může být tímto rozptýlena a uživatel může mít problém s určením, kam zaměřit svoji pozornost.

Pro problém vysoké hustoty informace existuje mnoho efektivních řešení. Jednou z metod je umožnit uživateli zobrazovat a skrývat různé komponenty zobrazení. V tomto případě se uživatel často rozhodne, které části jsou pro něj nejdůležitější a ostatní části jsou zobrazeny na požádání.

Dalším možným řešením je využití násobných obrazovek – buď ve formě disjunktních panelů, nebo můžeme povolit částečné překryvy.

Další typickou vlastností zobrazení velkého množství dat je nerovnoměrné rozložení datových množin. Odstranění tohoto artefaktu můžeme dosáhnout pomocí škálování jedné nebo více dimenzí dat.

## Klíče, označení, legendy

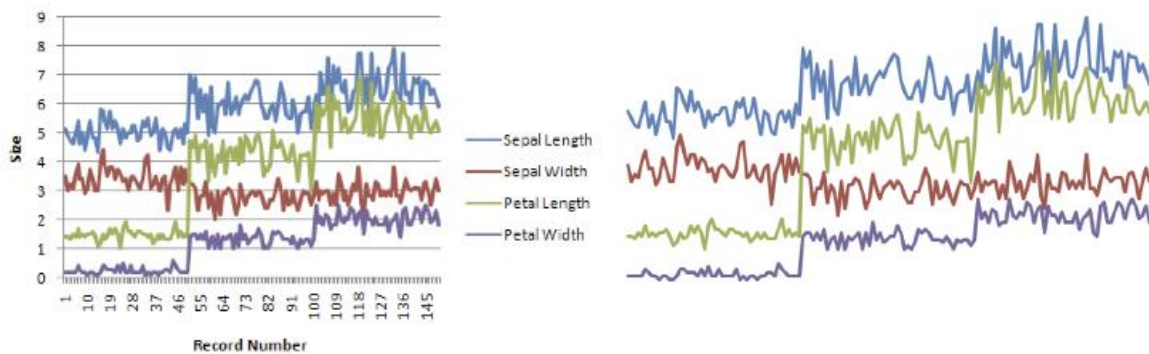
Dalším běžným problémem mnoha vizualizací je neposkytnutí dostatečné doplňkové informace, jejíž přítomnost by zabránila nejednoznačné a nepřesné interpretaci. Tato dodatečná informace by měla obsahovat detailní popis příslušných zobrazených datových polí a použitého mapování.

Dále je potřeba zahrnout značky, které znázorní rozsahy a hodnoty pro číselná pole. Všechny osy by měly být označeny a doplněny odpovídajícími jednotkami.

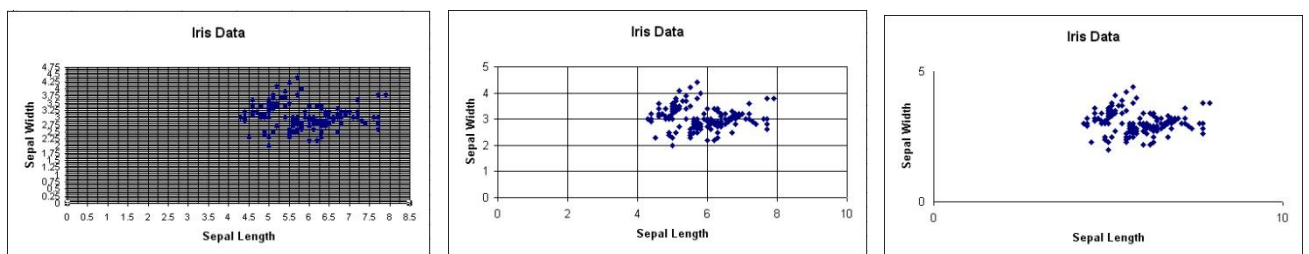
Pokud jsou použity symboly, musí být poskytnut klíč pro určení jejich významu – často podél hranice zobrazovacího okna nebo v samostatné grafické komponentě.

Pokud má i barva svůj význam, musí být přítomna dostatečná informace pro její jednoduchou interpretaci (např. označená barevná stupnice).

Obrázek ukazuje důležitost zobrazení těchto doplňkových informací – vlevo jsou zahrnuty popisky, zobrazení os i legenda, vpravo chybí.

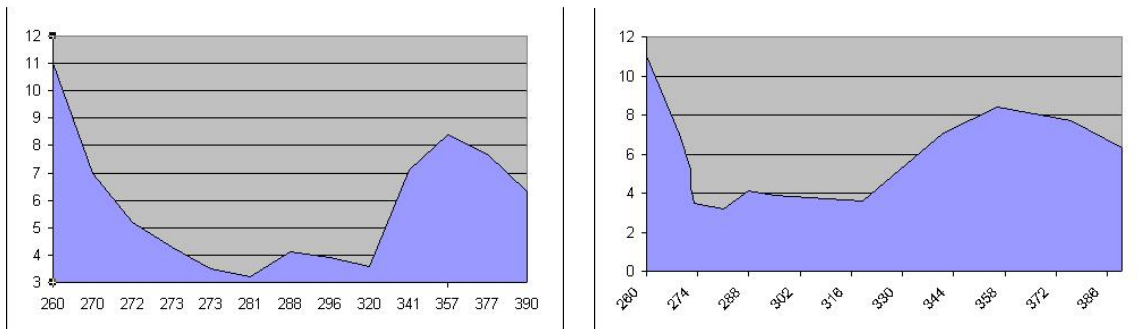


Využití značek má ovšem svá pozitiva i negativa. Špatný výběr značek a jejich hustota mohou zastínit samotné zobrazené datové hodnoty. Příklad je uveden na obrázku – vlevo přílišná hustota značek, uprostřed vhodná hustota, vpravo malá hustota.



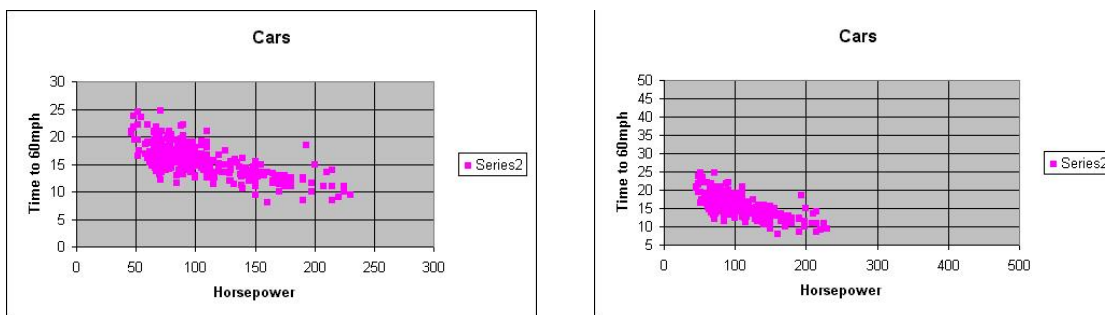
Samotné pozice označení mají vztah k tomu, jak čitelná bude daná interpretace dat. S ohledem na sémantiku dat mohou mít jisté mezery mezi označeními pro uživatele větší smysl než jiné. Defaultní nastavení některých vizualizačních nástrojů nemusí dokonce uživateli dávat žádný smysl (viz obrázek – vlevo nelogické rozmístění hodnot, vpravo logické).





Návrhář se musí rovněž rozhodnout, jaký rozsah hodnot bude zobrazen (toto rozhodnutí je prováděno v dřívějších fázích procesu vizualizace). Pokud nezobrazíme samotný rozsah, riskujeme špatnou interpretaci dat. Například pokud pracujeme s procentuálním vyjádřením, očekáváme, že zobrazená data jsou v rozsahu 0 až 100. Avšak v mnoha případech to může vést k plýtvání prostorem obrazovky (například pokud všechny hodnoty spadají do rozsahu 10 procent). Proto je nutné rozsah pečlivě vybrat a jasně označit.

Obrázek ukazuje rozdíl mezi logickou (vlevo) a nelogickou (vpravo) volbou rozsahu.



Při použití **násobných oken** je nutné zachovat konzistentní označení. Změna pozice označení a klíčů nebo rozsahu zobrazených hodnot pro stejné pole dat může způsobit zmatení a zvyšuje riziko špatné interpretace dat. Pokud jsou například změny v rozsahu nezbytné, měla by označení tyto změny rovněž reflektovat.

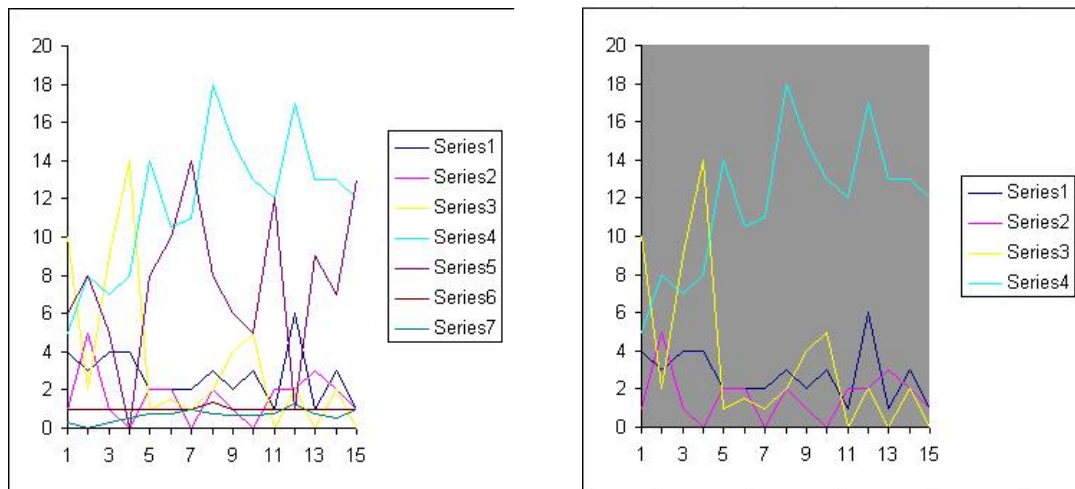
## Použití barvy

Jeden z nejčastěji špatně používaných parametrů při návrhu vizualizace je **barva**. Výběr špatné barevné škály nebo pokus o sdělení přílišného množství kvantitativní informace za použití barvy vede k neefektivní a zavádějící vizualizaci. Vnímání barvy je navíc závislé na použitém kontextu – příslušná barva je vnímána jinak v závislosti na použitých sousedních barvách.

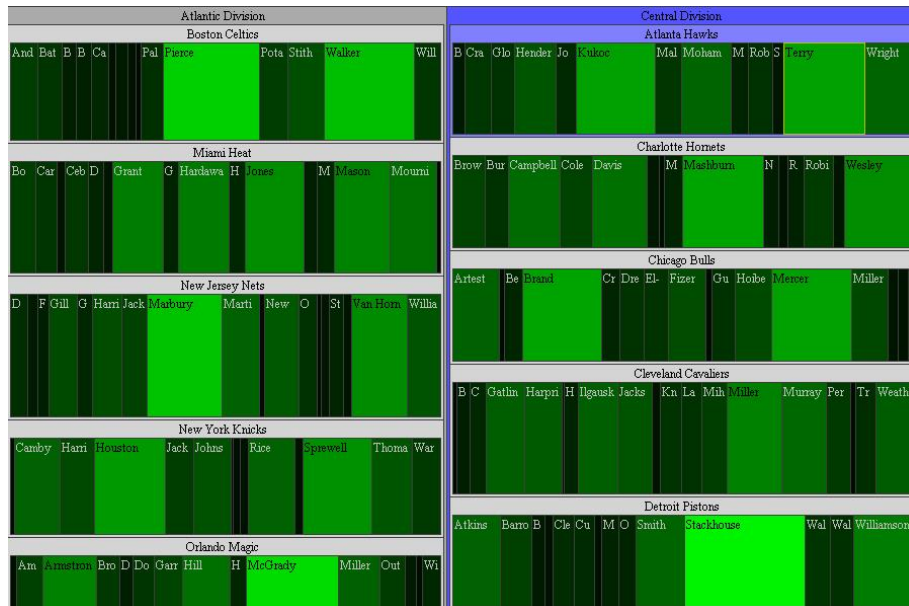
Dále musíme mít na paměti, že řada lidí je barvoslepých nebo trpí jinou poruchou vnímání barvy – průzkumy ukázaly, že až 10% mužů trpí nějakou formou této poruchy.

Následující pravidla mohou pomoci při návrhu efektivního využití barvy ve vizualizaci.

1. Pokud daný vizualizační úkol vyžaduje absolutní rozhodnutí, měli bychom použít pouze omezený počet různých numerických úrovní (viz obrázek – ukazující využití přílišného vs. přiměřeného množství barev).

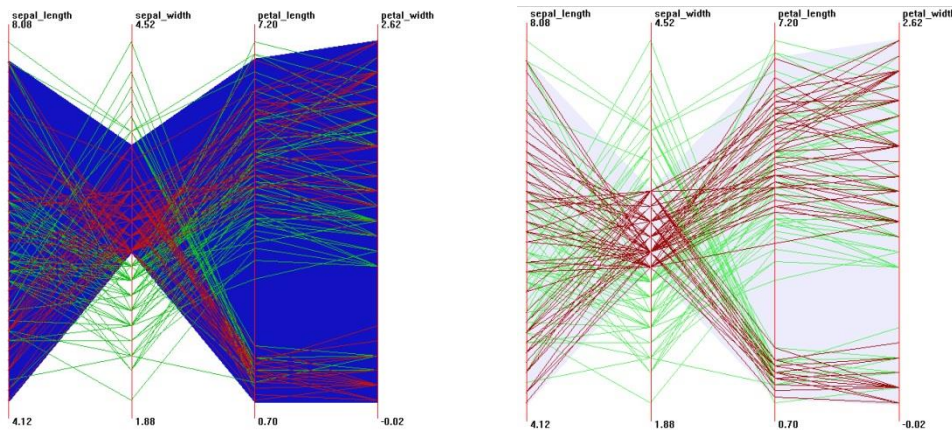


2. Pokud je to možné, využívat redundantní mapování, například mapování určitého pole dat na barvu a velikost zároveň (viz obrázek – treemap basketbalových statistik, kde jsou body získané za hru redundantně mapovány na barvu a velikost). Zvýšíme tak pravděpodobnost, že data budou interpretována tak, jak mají.



3. Při vytváření barevných map pro reprezentaci numerické informace se musíme ujistit, že pro každý záznam je změněna barva i její sytost (viz obrázek – vlevo změna pouze barvy, vpravo barvy a sytosti).





4. Zahrnout klíč určující, jakým způsobem jsou barvy namapovány na data – pomáhá interpretaci barvy.

Barva může vizualizaci přidat významnou vizuální působivost, ale může naopak i významně snížit efektivitu celého komunikačního procesu. Proto někteří návrháři preferují při iniciálním návrhu využití pouze šedotónních odstínů. Přidání barvy až po otestování prvotního návrhu může být provedeno mnohem efektivněji.

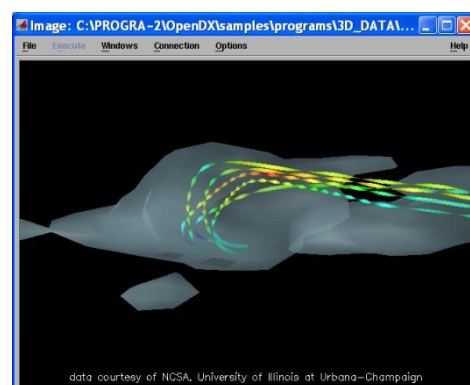
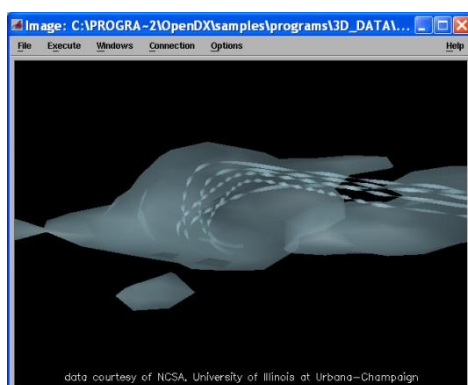
### Důležitost estetiky

Poté, co máme zajištěno, že navržená vizualizace umožňuje předat požadovanou informaci uživateli, musíme zhodnotit estetickou stránku návrhu. Samozřejmě nejlepší vizualizace jsou jak informativní, tak vizuálně přitažlivé. Na opačném pólu stojí vizualizace, které jsou tak vizuálně nehezké, že přímo odrazují od komunikačního procesu. Vizuálně přitažlivé vizualizace naopak lákají uživatele prozkoumat data více do hloubky.

Existuje řada pravidel pro návrh atraktivních vizualizací, která pochází zejména z komunity umělců a grafických designérů. Tato pravidla jsou:

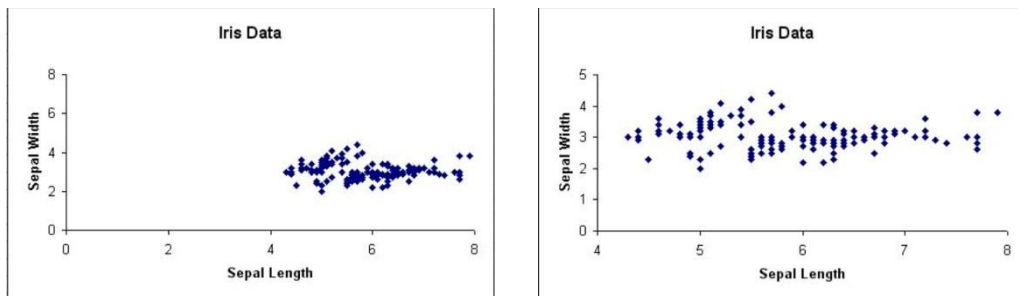
#### 1. Focus

Uživatel by měl být směřován na tu část vizualizace, která je nejdůležitější. Pokud nejsou důležité komponenty řádně zvýrazněny, uživatel nemá dostatečný návod pro jejich prozkoumání (viz obrázek – vlevo zobrazení proudových částic bez jejich zvýraznění, vpravo se zvýrazněním).



## 2. Balance

Prostor obrazovky by měl být využit efektivně, nejdůležitější komponenty by měly být umístěny do středu. Neměla by být zvýrazněna žádná hranice (viz obrázek – vlevo veškerá data na jedné straně, vpravo data vyvážená nalevo i napravo).

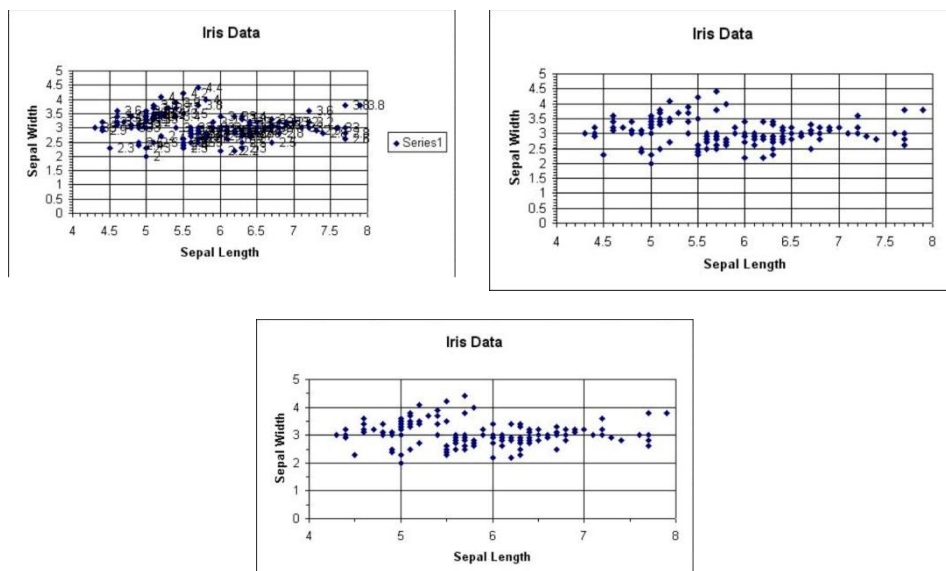


## 3. Simplicity

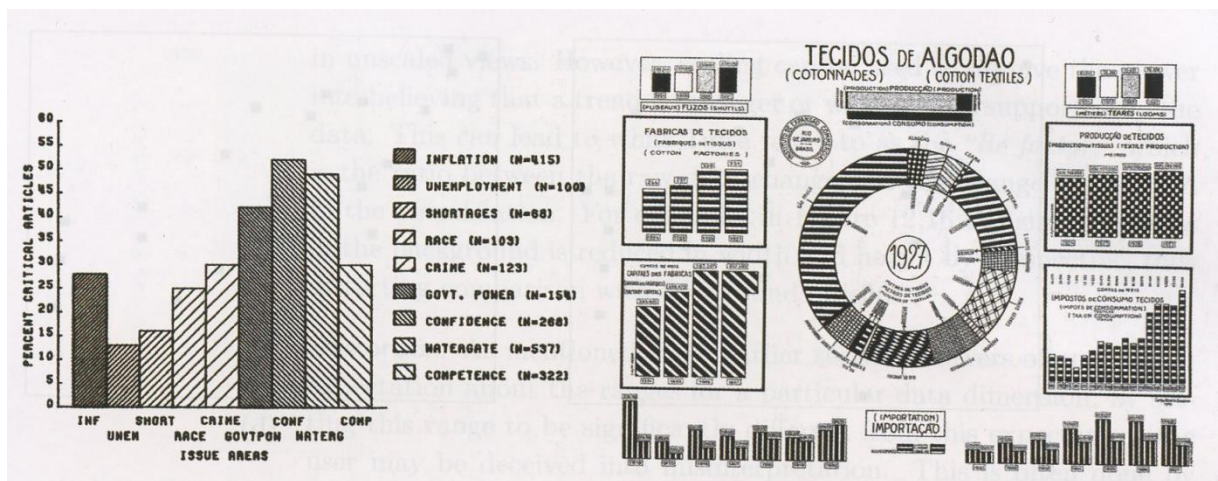
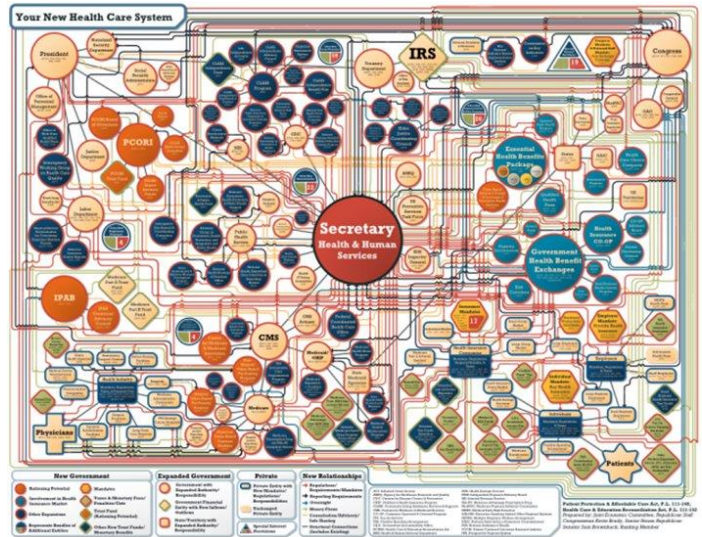
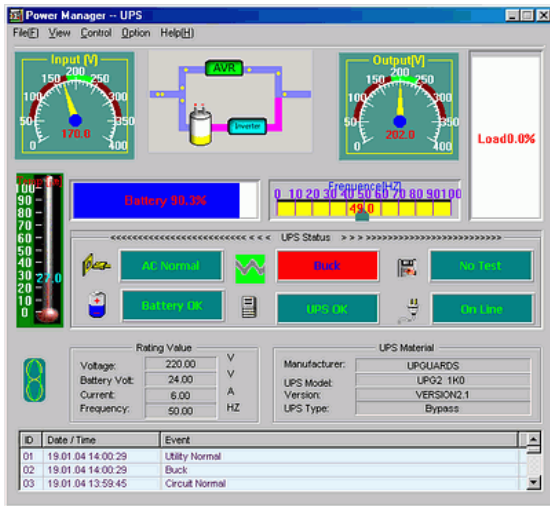
Hlavním pravidlem je nesnažit se nahustit co nejvíce informace na jednu obrazovku a nepoužívat různé grafické triky jenom proto, že jsou dostupné (například nepoužívat histogramy s 3D Phongovým stínováním, když tu stejnou informaci jednoduše sdělíme pomocí bodového nebo čárového grafu).

Běžná procedura pro eliminaci těchto „chyb“ u již navržených vizualizací je založena na iterativním odstraňování vlastností a poté měření velikosti ztráty sdělované informace.

Vlastnosti, po jejichž odstranění vizualizace vykazuje minimální ztráty ve sdělování informace, mohou být pravděpodobně zahozeny (viz obrázek ukazující postupný přechod z přeplněného grafu do zjednodušeného).



V literatuře nalezneme mnoho příkladů „ošklivých“ vizualizací. Některé z nich ukazují obrázky. Po provedení návrhu vizualizace je vřele doporučováno nechat si výsledek esteticky zhodnotit, než bude prezentován uživatelům. Dalším doporučením projít si některou z knih o grafickém designu.



## Problémy při návrhu efektivních vizualizací

V následující části se zaměříme na prozkoumání některých běžných problémů vyskytujících se při vizualizacích, které se mohou objevit, i když dodržujeme předchozí uvedená pravidla. Tyto problémy mají hlubší kořeny a souvisí s rozhodnutím, co vlastně vizualizovat a jaká je pro to nejvhodnější metoda. Některé z problémů zahrnují úmyslné nebo naopak nechtěné deformace dat, které mohou vést k dezinterpretaci, atd.

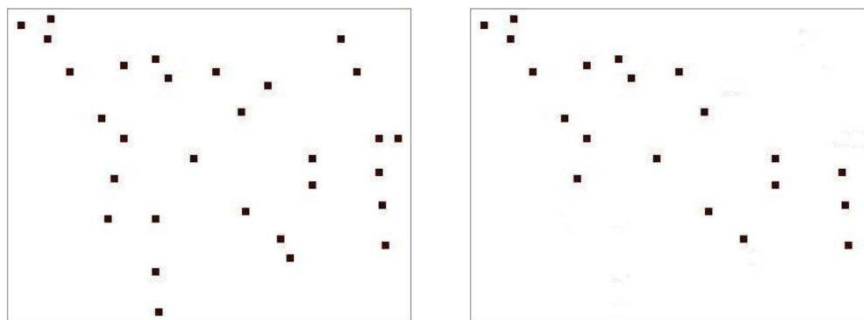
## Zavádějící vizualizace

Jedním z hlavních pravidel při vizualizaci je skutečnost, že obrázek by měl přesně znázorňovat požadovaná data. Avšak v historii je mnoho příkladů, kdy deformovaná vizualizace dat dokázala změnit názor uživatelů a v podstatě jim lhát. Tyto takzvané „viz lies“ můžeme najít všude – od nejprestižnějších časopisů po portfolia firem.

Nyní se podíváme na některé běžné strategie používané pro vytváření zavádějících vizualizací – ne pro to, abychom se je naučili vytvářet a používali je, ale proto, abychom se jim mohli v budoucnu vyhnout.

- **Data scrubbing**

Surová data mohou být často velmi hrubá, a proto jsme při návrhu vizualizace často v pokušení tuto hrubost odstranit. Naneštěstí je tento výběr, která data odstranit, proveden tak, že po eliminaci dat zobrazíme určité vztahy, které v původních datech vůbec nejsou (viz obrázek – vlevo surová data, která nevykazují korelaci, vpravo ořezaná data odhalující špatné korelace).

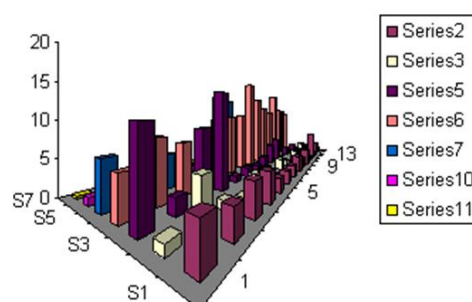


Běžnou taktikou používanou v této situaci je odstranění významně se odlišujících hodnot (outliers). I když je oprávněný důvod věřit, že tyto odlišné hodnoty vznikly díky chybám při procesu získávání dat, neměly by být odstraněny bez toho, aniž bychom o tom uživatele informovali. Rovněž bychom měli uživateli dát možnost si tyto hodnoty zobrazit.

- **Unbalanced scaling**

Škálování je velmi silným nástrojem používaným při vizualizaci, protože pečlivý výběr škálovacích faktorů může odhalit vzory a struktury, které nejsou v neškálovaných datech viditelné. Avšak škálování může být použito i k oklamání pozorovatele, který pak věří, že daný trend v datech je silnější nebo slabší, než je ve skutečnosti. To vede k tomu, co Tufte označuje jako „**lie factor**“, což je poměr mezi změnou v surových datech a změnou znázorněnou pomocí vizualizace.

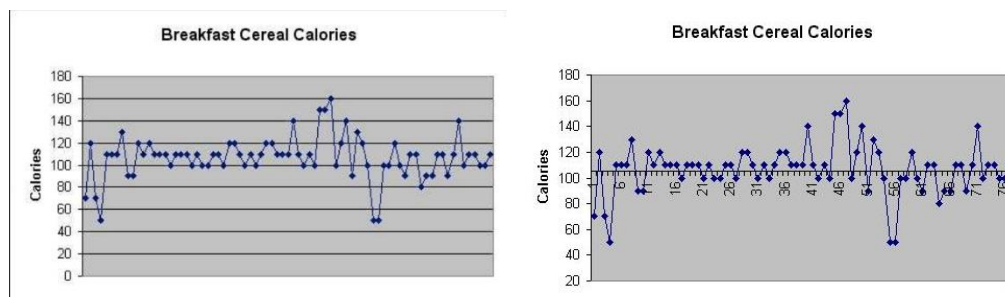
Příklad je uveden na obrázku, kde velikost (šířka i výška) objektů v pozadí je redukována s perspektivou, čímž narušujeme možnost jejich porovnání s objekty vepředu.





- **Range distortion**

Jak jsme již zmínili dříve, uživatelé často mají jistá očekávání týkající se rozsahu hodnot v příslušné dimenzi. Proto nastavení rozsahu zcela odlišného od tohoto očekávání může opět vést k dezinterpretaci. Častým příkladem je posun osy takovým způsobem, že už dále nekoresponduje s očekávanou nulovou hodnotou (viz obrázek), což může mít významný vliv na interpretaci obrázku. Návrhář samozřejmě může uživateli nabídnout možnost posunu počátku souřadné soustavy, aby se vyhnul plýtvání prostorem obrazovky. Vše ale musí být provedeno jasně, zejména pokud to vybočuje ze zavedených norem.

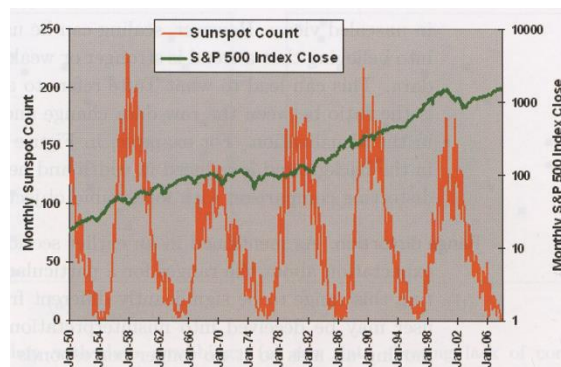


- **Abusing dimensionality**

Počet chyb v interpretaci je úměrný mocnině počtu dimenzí zobrazovaných dat. Proto naše chybovost při posuzování objemu je mnohem vyšší než při posuzování plochy a ta je vyšší než při posuzování délky. Mapování skalární hodnoty na grafický atribut objemu tedy významně zvyšuje pravděpodobnost chybné interpretace. Jak již bylo zmíněno dříve, i zde platí pravidlo: **čím jednodušší, tím lepší**.

### Vizuální nesmysly – porovnávání hrušek a jablek

Vizualizace jsou vytvářeny pro sdělení informace a je důležité, aby tato informace byla smysluplná. Vizualizace jsou často vytvářeny kombinováním datových sad pocházejících z různých zdrojů. Je jednoduché kombinovat nesouvisející komponenty do jedné vizualizace a identifikovat v nich strukturu – například vykreslení hodnot na burze proti výskytu skvrn na Slunci (viz obrázek).

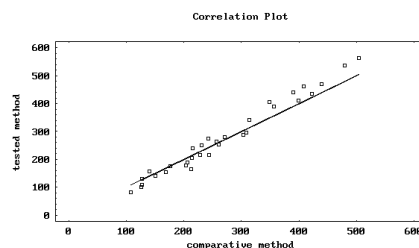


Při rozhodování, která data kombinovat, je důležité se nejdříve ujistit, že je v jejich kombinaci určitá logika.

Další faktor, který musíme vzít v úvahu, je kompatibilita mezi časovými a prostorovými rozsahy porovnávaných dat. Například bychom tedy neporovnávali prodejnost určitého produktu v jednom roce v daném regionu země s prodejností téhož produktu v jiném roce a v jiném regionu – podobnou analýzu bychom mohli provádět třeba pouze v případě, že bychom měli podezření na migraci zájmu o daný produkt.

Při vytváření datových sad pro vizualizaci musíme vzít v úvahu i kompatibilitu jednotek. Například produkty, které jsou měřeny v jednotkách „cena za objem“ jsou často kombinovány s produkty měřenými v jednotce „cena za kilogram“. Efektivní vizualizace takovýchto dat by jednotky normalizovala na „cenu za porci“.

Nakonec jsme často v pokušení aplikovat operace, které jsou vhodné pro uspořádaná nebo spojitá data, na kategorická, neuspořádaná data – jednoduše protože výsledkem mapování byla uspořádaná grafická reprezentace. Příkladem je pokus napasovat přímkou nebo křivkou na sekvenci datových bodů, které mapují jméno společnosti na pozici na obrazovce. Očividně toto mapování nemá žádný sémantický význam, ale protože vizuálně se mapování pohybuje kolem nějaké křivky, uživatel může mít dojem, že toto napasování je vhodné.



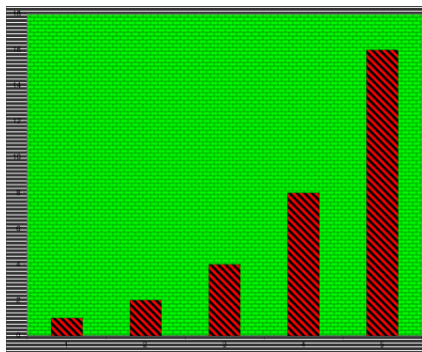
### Ztráta dat díky Chart Junk

**Chart Junk** = veškeré vizuální prvky v grafech a diagramech, které nejsou nezbytné pro pochopení informace prezentované v grafu a spíše rozptylují pozornost uživatele.

Příklady jsou uvedeny na obrázcích. Obrázek vlevo zobrazuje pouze pět číselných hodnot, které jsou velmi těžko čitelné. Jedním z důvodů je právě využití velkého množství zbytečných symbolů, čar a barev.

Druhý obrázek obsahuje naprosto zbytečný gradient v každém regionu.





Tato doplňková informace může vést nejen k přehnaně složitým vizualizacím, ale rovněž k okluzi a zeslabování vizuální reprezentace původních dat.

Rozhodnutí, jaké množství doplňkové grafiky použít, je často obtížný proces. Hlavně proto, že návrhář nemusí znát potřeby všech potenciálních uživatelů. Avšak protože pracujeme s dynamickým a uživatelsky měnitelným médiem, můžeme uživateli rovněž umožnit ovlivňovat míru zobrazení těchto doplňkových informací.

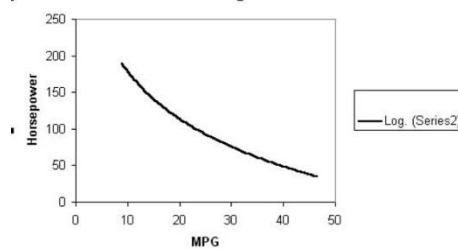
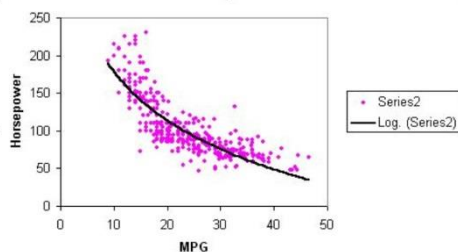
V některých vizualizačních úkolech může uživatel přepínat mezi kvalitativním náhledem a kvantitativní analýzou. V prvním případě je obvykle důležitější poskytnout uživateli čistý pohled na data, zatímco ve druhém případě jsou žádoucí doplňkové nástroje, které pomáhají kvantifikovat zobrazené prvky.

Vhodným **pravidlem** je tedy poskytnout dostatečné nástroje pro podporu kvantitativních náhledů a zároveň umožnit jejich odstranění nebo změnu stupně jejich zapojení do vizualizace.

### Surová vs. odvozená data

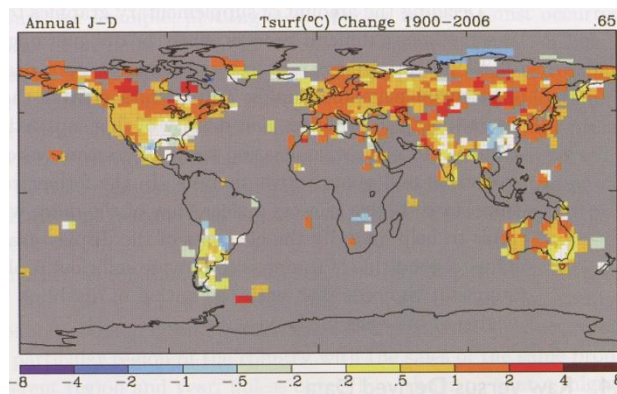
Běžnou praktikou při snaze o dosažení co nejlepšího vizuálního výsledku je spočtení analytického modelu dat reprezentovaného křivkami či povrchem. Tato technika může opět vést ke zkreslení a ve výsledku ke špatným předpokladům a odvozeným závěrům.

V některých vizualizacích je běžnou praktikou odstranit všechna surová data a nahradit je hladkou aproximací odvozenou z těchto dat. To nutí uživatele věřit, že daná aproximace je přesným zobrazením dat, což často není pravda. Proto je nejvhodnější v tomto případě zobrazit jak surová data, tak jejich aproximaci, a umožnit uživateli si tyto typy filtrovat na požádání (viz obrázek – vlevo surová data s aproximační křivkou, vpravo pouze aproximační křivka).

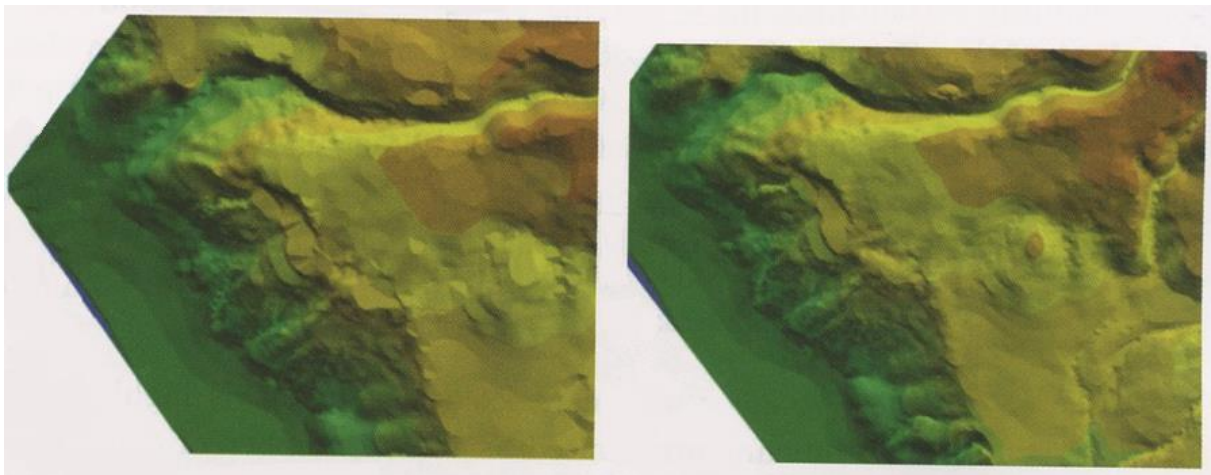


Další formou „čištění“ dat je proces **převzorkování**, kdy jsou surová data umístěná náhodně nebo v řídké mřížce použita pro vytvoření mnohem hustší mřížky. To vede k mnohem bohatším vizualizacím, které se blíží spojitému vzorkování. Avšak opět klameme uživatele, který věří, že zobrazená data jsou mnohem větší, než jsou ve skutečnosti. Čím hustší převzorkování, tím větší pravděpodobnost špatné interpretace dat.

Například obrázek ukazuje umístění stanic pro globální monitorování teploty. Je jasné, že jsou místa, kde není žádná stanice, proto převzorkování vede k odvození špatných závěrů, jako například ten, že celá severní část Jižní Ameriky by byla odvozena z dat vyčtených ze čtyř nebo pěti stanic a výsledek by byl takový, že v posledním století došlo k poklesu průměrné teploty v této oblasti.



Dalším problémem je nedostatečné vzorkování. Jak ukazuje obrázek, vzorkování, které nebere v potaz charakteristiky dat, může opomenout řadu důležitých vlastností. Levý obrázek je vzorkován a interpolován uniformně, zatímco pravý obrázek využívá informaci o kontuře, pomocí které přidáváme body tam, kde se objevuje výrazná změna.



Je nutné, aby měl uživatel vždy přístup k surovým datům a aby byl informován o jakékoliv operaci vyhlazení či převzorkování, která byla na data aplikována. V některých oborech, jako

například v radiologii, jsou analytici zásadně proti použití jakéhokoliv vyhlazení či filtrování, protože je nebezpečí, že důležitý signál by mohl být vyhodnocen jako šum. Proto je vhodné poskytnout uživateli pohled na surová data i na odvozená a nechat jej rozhodnout, zda dané odvození dostatečně reprezentuje původní data.