VŠB TECHNICKÁ UNIVERZITA OSTRAVA | FAKULTA ELEKTROTECHNIKY A INFORMATIKY | KATEDRA APLIKOVANÉ MATEMATIKY

Počítačová cvičení
Škola matematického modelování
2020

Petr Beremlijski, Rajko Ćosić,
Marie Sadowská, Adéla Vrtková

Katedra aplikované matematiky
Fakulta elektrotechniky a informatiky
VŠB - Technická univerzita Ostrava
2020

Předmluva

Matematika je krásná. Co bylo pravda včera, je pravda i dnes.

Jaroslav Kurzweil

A výše uvedené (je řeč o druhé větě ...) platí i o mém loňském úvodníku. Nebudu ho proto měnit, ale jen (červeně) aktualizovat.

Milí studenti,
rád bych Vás jménem Katedry aplikované matematiky a Jednoty českých matematiků a fyziků přivítal na již 16. ročníku našeho semináře ŠKOMAM (ŠKOla MAtematického Modelování). Matematika je z mnoha důvodů krásná disciplína a přestože je vytvářená a vybrušovaná už několik tisíc +1 let (a to co vzniklo, je prostě úžasné), je pořád v obrovském rozmachu. Je totiž nesmírně užitečná.

Posledních pár desetiletí žijeme v době počítačů a dnes už i superpočítačů (jeden z největších ve střední Evropě je i u nás na Vysoké škole báňské – Technické univerzitě Ostrava – můžete se podívat na stránky Národního superpočítačového centra IT4Innovations: https://www.it4i.cz), což přináší dříve netušené možnosti rozvoje vlastně všech oblastí lidské činnosti. A ukazuje se, že chceme-li opravdu využít všech možností, které přináší intenzivní rozvoj výpočetní techniky, je třeba vyřešit spoustu nově vznikajících problémů, upravit stávající (a nebo vymyslet zcela nové) algoritmy, které efektivně využijí architekturu superpočítačů, a použijí nejnovější poznatky z matematiky nebo znalosti získané díky spojení různých oborů matematiky. A zde je obrovský prostor a spousta výzev pro matematiku a matematiky.

V reakci na tuto potřebu doby u nás na katedře vzniknul před více než 16 lety studijní obor „Počítačová matematika“ (o pár let později přejmenovaný na „Výpočetní matematika“) spojující studium aplikované matematiky se základy informatiky. Tento obor (ve všech stupních VŠ vzdělání, tj. bakalářském, magisterském a doktorském) již absolvovalo více než 250 studentů. Že se věc daří, dokládají nejen úspěchy našich studentů v celostátních soutěžích, ale především ta skutečnost, že je o naše absolventy velký zájem. Předminulý rok jsme náš obor podstatně upravili a přetavili ho do studijního programu „Výpočetní a aplikovaná matematika“. Budeme moc rádi, pokud si ho vyberete. Jsme přesvědčeni, že tím neuděláte chybu!

Následující dny trávené u nás na ŠKOMAMu můžete brát jako malou ochutnávku toho, co by Vás u nás čekalo. Poznáte některé z členů naší katedry, prohlédnete si posluchárny, ale především se budete bavit matematikou.

A já vám za nás všechny přeju, ať si to tu náramně užijete.

V Ostravě 10.1.2020 Jirka Bouchala

Úvod

Tento text je určen pro účastníky semináře Škola matematického modelování (http://skomam.vsb.cz) a slouží jako pomůcka k úlohám, které řeší studenti v průběhu tohoto semináře. Tento seminář, pro který používáme zkratku ŠKOMAM, organizuje Katedra aplikované matematiky (http://am.vsb.cz) Fakulty elektrotechniky a informatiky Vysoké školy báňské – Technické univerzity Ostrava jednou ročně již od roku 2005. V tomto roce probíhá již 16. ročník tohoto semináře.

Pro počítačové řešení úloh jsme v prvních 13 ročnících používali komerční systém Matlab. Ve čtrnáctém ročníku jsme se rozhodli udělat změnu a nově jsme zvolili programový balík R. Software R je zejména prostředí určené pro statistickou analýzu dat a jejich grafické zobrazení. Jazyk R ale umožňuje navíc i manipulaci s daty, numerické výpočty a grafické výstupy. Obsahuje také (podobně jako Matlab) řadu dalších knihoven s mnoha připravenými funkcemi. Jeho hlavní výhodou je jeho volná dostupnost pro výukové i vědecké účely. Podrobný popis tohoto jazyka je k dispozici v [2]. Programový balík R je volně k dispozici na webové adrese https://www.r-project.org/. Pro práci s balíkem R doporučujeme prostředí RStudio, které je volně k dispozici na webové adrese https://www.rstudio.com/products/rstudio/download/.

Tento seminář je pořádán s finanční podporou Fakulty elektrotechniky a informatiky (http://www.fei.vsb.cz), statutárního města Ostravy a projektu Math Exercises for You 2 (http://math4u.vsb.cz) podpořeného programem Erasmus+. Nad akcí převzala záštitu Jednota českých matematiků a fyziků (http://jcmf.vsb.cz).

Cvičení 1: R – nejen nástroj pro statistickou analýzu dat

Abychom se mohli věnovat pokročilejším matematickým úlohám, potřebujeme vhodné prostředí, které nám jejich řešení umožní. Software R je zejména statistický software určený pro statistickou analýzu a grafickou reprezentaci dat, nicméně zvládne řešení i jiných úloh. Podstatnou výhodou softwaru R je jeho volná šiřitelnost. Podrobně se tomuto pracovnímu prostředí věnuje přiložený dokument „Jak pracovat s jazykem R“ [2]. My si zde uvedeme pouze stručný přehled Rkových proměnných a příkazů, které budeme potřebovat.

Prostředí

- *help, demo, methods, getAnywhere, library, ls, remove, class, dim, length, setwd*

Proměnné

- Skaláry
- Vektory
- Matice
- Datové rámce

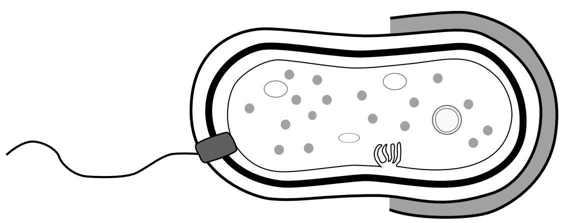
Příkazy

- Skalární funkce - *sin, cos, tan, asin, acos, exp, log, abs, sqrt, round*
- Vektorové funkce a generování vektorů - *max, min, mean, median, sort, seq, rep*
- Maticové funkce a generování matic - *matrix, det, ones, zeros, diag*
- Změna typu objektu - *as.numeric, as.character, as.factor, as.matrix, as.data.frame*
- Importování dat - *read.csv, read.csv2*
- Skalární operace - $+,\ -,\ *,\ /, \hat{}$
- Maticové a vektorové operace - $+,\ -,\ \%*\%,\ t$ (transponování), *solve* ($solve(A, v) = x \Leftrightarrow Ax = v$)
  Operace „po prvcích“ - $*, \hat{}, /$
- Základní grafika (vykreslení grafů) - *plot, lines, points, dev.set, dev.next, dev.off*
- Pokročilá grafika - *plot, ggplot*
- Řídící příkazy - *if*, *ifelse* (podmíněné příkazy), *for, while* (příkazy cyklu se známým počtem opakování a podmínkou na začátku)

- Relace a logické operace - $<, >, <=, >=, ==, !=, \&, |, !$
- Skripty a funkce - *function*

Úkol 1.1 Tričko je vyrobeno z látky o tloušťce 1 mm. Po jednom sušení v sušičce se látka ztenčí o 5 %. K proděravění dojde, má-li látka tloušťku menší než 0,2 mm. Tričko sušíme v sušičce jedenkrát týdně. Jaká bude tloušťka látky po 5 týdnech? Zaokrouhlete na 2 desetinná místa. Kolik týdnů nám tričko vydrží? ▲

Úkol 1.2 Bakterie Yersinia pestis, která způsobuje onemocnění morem, se v příznivých podmínkách dělí jednou za 100 minut. Jak dlouho by trvalo, pokud by nedocházelo k úhynu bakterií a mohly se bez omezení množit, než by jejich hmotnost překročila hmotnost Země? Předpokládejme, že v čase $t = 0$ žije jedna bakterie Yersinia pestis. Předpokládejme, že hmotnost jedné bakterie je $6 \cdot 10^{-15}$ kg a hmotnost Země je $6 \cdot 10^{24}$ kg. ▲



Obrázek 1: Bakterie

Úkol 1.3 Sestrojte grafy následujících funkcí:

a) $f(x) := x^2$

b) $f(x) := \sqrt{1-(|x|-1)^2}, x \in \langle -2, 2\rangle$

c) $f(x) := \arccos(1-|x|) - \pi, x \in \langle -2, 2\rangle$

d) Pomocí vhodné funkce vykreslete funkce z bodu b) a c) do jednoho obrázku. ▲

Úkol 1.4 Pomocí funkce *seq* si zadefinujte libovolný vektor (např. posloupnost čísel od 1 do 30 s krokem 3).

a) Vyberte ty prvky, které jsou větší než 5.

b) Vyberte ty prvky, které jsou větší než 5 a zároveň menší než 12.

c) Vyberte ty prvky, které jsou menší než 5 nebo větší než 12. ▲

Úkol 1.5 Pomocí funkce *matrix* vytvořte matice se 4 řádky a 3 sloupci, které budou obsahovat čísla od 1 do 12 řazeny následovně

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 5 & 9 \\ 2 & 6 & 10 \\ 3 & 7 & 11 \\ 4 & 8 & 12 \end{pmatrix}$$

Aplikujte na tyto matice funkce *as.numeric*, *as.character* a *as.data.frame* a prozkoumejte výstup. ▲

Cvičení 2: Dáte si kávu v ateliéru u Vermeera?

V tomto cvičení se pokusíme vytvořit matematický model, který popisuje rozpad atomů radioaktivních prvků. Abychom mohli takovýto model sestrojit, potřebujeme pracovat se speciálním typem rovnice – s tzv. diferenciální rovnicí. Podrobně se tomuto typu rovnic věnuje text [3]. Navíc musíme vědět, jak můžeme diferenciální rovnice numericky řešit. Tomu se budeme věnovat nyní.

`Obyčejnou diferenciální rovnicí 1. řádu` rozumíme rovnici tvaru

$$y'(t) = f(t, y(t)),$$

kde $f : \mathbb{R}^2 \to \mathbb{R}$ je zadaná funkce a $y : \mathbb{R} \to \mathbb{R}$ je „hledaná“ funkce. `Řešením této rovnice na otevřeném intervalu` (a, b), kde $a < b$, rozumíme každou funkci $\bar{y} : (a, b) \to \mathbb{R}$, takovou, že pro všechna $t \in (a, b)$ platí

$$\bar{y}'(t) = f(t, \bar{y}(t)).$$

Například funkce $\bar{y}(t) = t$, $t \in (0, +\infty)$, je řešením diferenciální rovnice $y'(t) = \frac{y(t)}{t}$ na intervalu $(0, +\infty)$. Jiným řešením této rovnice na intervalu $(0, +\infty)$ je například funkce $\bar{y}(t) = 2t$, $t \in (0, +\infty)$, či $\bar{y}(t) = 3t$, $t \in (0, +\infty)$. Není těžké ukázat, že každá funkce $\bar{y}_k(t) = kt$, $t \in (0, +\infty)$ $(k \in \mathbb{R})$, je řešením diferenciální rovnice $y'(t) = \frac{y(t)}{t}$ na intervalu $(0, +\infty)$. Naše úloha má nekonečně mnoho řešení. Zkusme navíc přidat k naší rovnici například podmínku $y(1) = 2$, tj. chceme nalézt funkci, která řeší naši rovnici a navíc její funkční hodnota v $t = 1$ je rovna 2. Lze ukázat, že taková úloha má na $(0, +\infty)$ pouze jediné řešení $\bar{y}(t) = 2t$.

Úlohu, která se skládá z hledání řešení obyčejné diferenciální rovnice 1. řádu, která má navíc splňovat tzv. počáteční podmínku $y(t_0) = y_0$ $(t_0 \in \mathbb{R}, y_0 \in \mathbb{R})$, nazýváme `Cauchyovou úlohou` a zapisujeme ji obecně takto:

$$\begin{cases} y'(t) = f(t, y(t)), \\ y(t_0) = y_0. \end{cases} \tag{1}$$

Pokud má funkce $f(t, y(t))$ speciální tvar[1], pak existují metody, jak analyticky řešit výše popsanou Cauchyovu úlohu. Často však analytické řešení nalézt nelze nebo by jeho nalezení bylo příliš náročné. V takovém případě se nabízí použití některé z numerických metod pro přibližné řešení diferenciálních rovnic 1. řádu s počáteční podmínkou. Podívejme se nyní na jednu z těchto metod – Eulerovu metodu[2]. O funkci f budeme dále předpokládat, že je v množině $D = \{(t, y) \in \mathbb{R}^2 : t_0 \leq t\}$ spojitá a také jsou v D spojité derivace této funkce podle proměnné y.

[1] Například f závisí lineárně na y, tj. $f(t, y) = a(t)y + b(t)$, kde a a b jsou reálné funkce.
[2] Publikoval ji významný švýcarský matematik a fyzik Leonhard Euler v roce 1768.

`Eulerova metoda`

Eulerova metoda je opravdu jednoduchý způsob numerického řešení Cauchyových úloh. Výstup Eulerovy metody nám aproximuje řešení Cauchyovy úlohy (1) na intervalu $\langle t_0, t_N \rangle$. Metoda využívá aproximace derivace[3] funkce y v bodě t pomocí tzv. dopředné diference y v tomto bodě

$$y'(t) \approx \frac{y(t+h) - y(t)}{h},$$

kde h je „malé" a kladné.

Po jednoduché úpravě dostaneme

$$y(t+h) \approx y(t) + hy'(t).$$

Použijeme-li (1), pak získáme vztah

$$y(t+h) \approx y(t) + hf(t, y(t)). \tag{2}$$

Dále zvolme „dostatečně malé" $h > 0$ a sestrojme posloupnost

$$t_0,\ t_1 \stackrel{\text{def.}}{=} t_0 + h,\ t_2 \stackrel{\text{def.}}{=} t_0 + 2h,\ \ldots,\ t_N \stackrel{\text{def.}}{=} t_0 + Nh$$

Označme pomocí y_n aproximaci hodnoty přesného řešení $y(t_n)$. Z (2) a po dodání počáteční podmínky z (1) dostaneme rekurzivní vztah

$$\boxed{\begin{array}{l} y_0 = y(t_0), \\ y_{n+1} = y_n + hf(t_n, y_n),\ n = 0, 1, \ldots, N, \end{array}}$$

který použijeme pro numerické řešení Cauchyovy úlohy (1). Dá se ukázat, že chyba aproximace řešení pomocí Eulerovy metody v bodě t_1 je přímo úměrná druhé mocnině velikosti kroku h. Chyba aproximace řešení v bodě t_N je přímo úměrná velikosti h.

Než se začneme zabývat tvorbou matematického modelu, který popisuje rozpad atomů radioaktivních prvků, vrátíme se v čase do doby krátce po konci druhé světové války. Těsně po válce zjistila nizozemská policie, že během války bylo prodáno několik Vermeerových[4] obrazů německému ministrovi letectví Hermannu Göringovi. Tuto transakci zprostředkoval nizozemský občan Han van Meegeren[5]. Na základě těchto zjištěných faktů byl 29.5.1945 van Meegeren zadržen a obviněn z kolaborace s nepřítelem. 12.7.1945 van Meegeren vydal prohlášení, že Göringovi nikdy žádný Vermeerův obraz neprodal. Naopak Göringa napálil, protože obrazy, které mu prodal, jsou podvrhy Vermeerových obrazů a sám je vytvořil.

A aby dokázal své tvrzení, začal jeden z „Vermeerových" obrazů[6] napodobovat. Van Meegeren přizvaným znalcům předvedl způsob, jakým vytváří barvy, jak připravuje plátno,

[3] Pro jistotu zde připomínáme její definici $y'(t) \stackrel{\text{def.}}{=} \lim\limits_{h \to 0} \frac{y(t+h)-y(t)}{h}$.

[4] Jan Vermeer van Delft (31. 10. 1632 - 15. 12. 1675) byl nizozemský barokní malíř.

[5] Han van Meegeren (10. 10. 1889 - 30. 12. 1947) byl nizozemský malíř.

[6] „Ježíš mezi znalci Písma"

či jak zařídí, aby povrch malby vypadal jako u několik set let starého obrazu. Těsně před dokončením podvrhu Vermeerova obrazu se van Meegeren dozvěděl, že obvinění z kolaborace, bude nahrazeno obviněním z padělatelství, a tak odmítl tuto kopii dokončit. I tak ale většina přizvaných odborníků uznala, že obrazy prodané Göringovi jsou pravděpodobně falzum a van Meegeren byl 12.10.1947 odsouzen za padělatelství na rok do vězení, ve kterém 30.12.1947 na infarkt zemřel. I přesto, že komise, která posuzovala pravost „Vermeerových“

Obrázek 2: Han van Meegeren

obrazů uznala, že to jsou pravděpodobně podvrhy vytvořené van Meegerenem, zůstávali odborníci u některých obrazů, k jejichž autorství se také van Meegeren přihlásil, na pochybách. Zejména zpochybňování pravosti obrazu Emauzští učedníci, který zakoupilo muzeum v Rotterdamu za 170 000 dolarů, vyvolával velké spory. Proto se přistoupilo u tohoto obrazu v roce 1967 k metodě radioaktivního datování, která měla tyto pochyby rozhodnout. Tato pravost byla zkoumána na Carnegie Mellon University (Pittsburgh, USA).

Metoda radioaktivního datování využívá toho, že některé tzv. radioaktivní prvky jsou nestabilní a část jejich atomů se samovolně rozpadá na atomy jiných prvků. Experimenty bylo zjištěno, že rychlost rozpadu atomů radioaktivních prvků je přímo úměrná počtu těchto atomů. Pokud funkci udávající počet atomů radioaktivního prvku v čase t v gramu látky označíme jako $N(t)$, pak výše zmíněnou závislost můžeme popsat diferenciální rovnicí

$$N'(t) = -\lambda N(t), \tag{3}$$

kde λ je konstanta, která popisuje rychlost rozpadu atomů daného radioaktivního prvku. Tato konstanta je dána pro každý radioaktivní prvek tímto vztahem

$$\lambda = \frac{\ln 2}{\text{poločas rozpadu prvku}}.$$

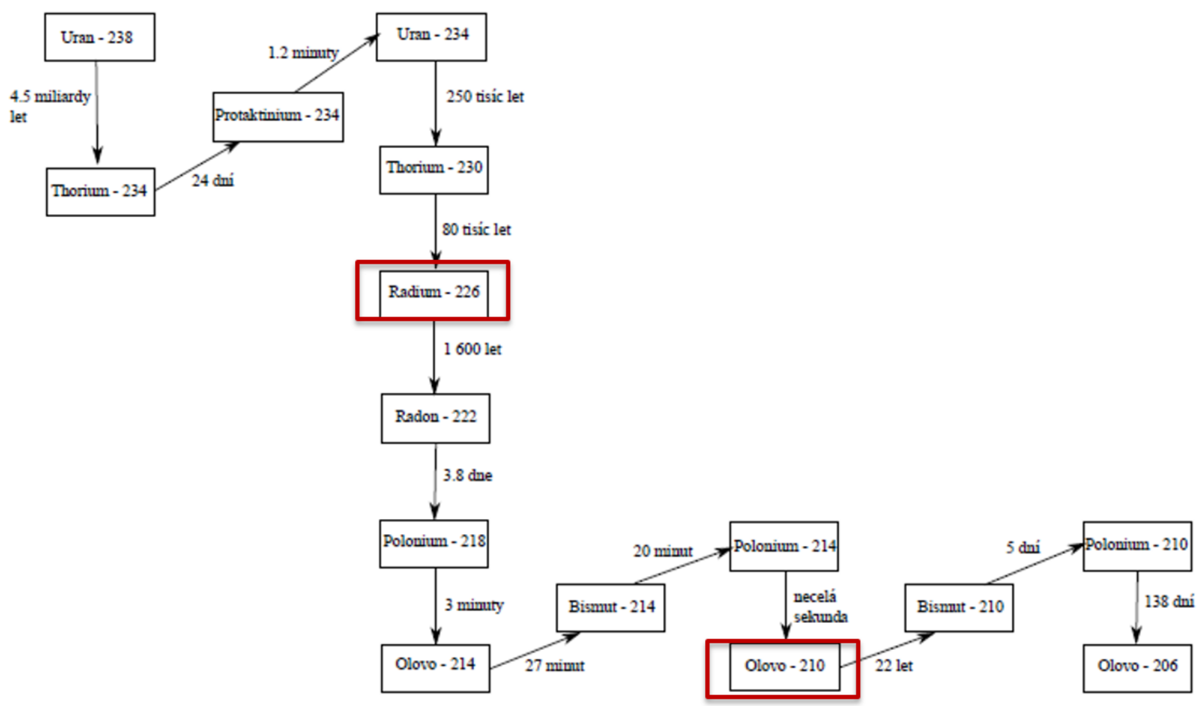
Čas t v našem modelu budeme měřit v rocích a jednotka konstanty λ je v rok^{-1}.

Metoda radioaktivního datování je založena na jednoduchém pozorování. Pokud bychom věděli, kolik atomů radioaktivního prvku měla látka v jednom svém gramu při svém vzniku (tzn. známe hodnotu N_0, pro kterou platí $N(0) = N_0$), a znali bychom také aktuální počet těchto atomů v gramu látky, mohli bychom řešením úlohy

$$\begin{cases} N'(t) = -\lambda N(t), \\ N(0) = N_0. \end{cases} \tag{4}$$

zjistit, jak je tato látka stará.

Než se začneme zabývat datováním „Vermeerových“ obrazů, uvědomme si, že všechny horniny na Zemi obsahují malé množství radioaktivního uranu, který se rozpadá na atomy dalšího prvku. Tyto atomy se opět samovolně mění na další atomy atd. (viz obrázek 3).



Obrázek 3: Uranová rozpadová řada (časy u šipek udávají poločasy rozpadu jednotlivých radioaktivních prvků)

Dále je známo, že olovnatá běloba používaná na Vermeerových malbách obsahuje oxid olovnatý, který obsahuje malé množství olova-210 a ještě menší množství radia-226. V okamžiku, kdy je barva obsahující oxid olovnatý vyrobena, začnou se atomy olova-210 velmi rychle rozpadat s poločasem rozpadu 22 let a množství olova-210 v této barvě klesá. Na druhé straně vzniká malé množství olova-210 rozpadem radia-226 s poločasem rozpadu 1600 let (a prvků, které následují v rozpadové řadě za ním). Tento proces můžeme popsat následující Cauchyovou úlohou

$$\begin{cases} N'(t) = -\lambda N(t) + r(t), \\ N(0) = N_0, \end{cases} \tag{5}$$

kde $N(t)$ je funkce udávající počet atomů olova-210 v čase t v gramu látky, $r(t)$ je funkce udávající počet atomů olova-210, které vzniknou v čase t v gramu oxidu olovnatého za rok.

Protože poločas rozpadu radia-226 je 1600 let a metodu radioaktivního datování chceme použít pro rozpoznání stáří obrazů, které měly v roce 1967 buď přibližně 300 let nebo 20 let, můžeme funkci $r(t)$ považovat za konstantní. Pak $r(t) = r =$ konst. a úlohu (5) můžeme nahradit úlohou

$$\begin{cases} N'(t) = -\lambda N(t) + r, \\ N(0) = N_0. \end{cases} \tag{6}$$

Mnohem více podrobností o metodě radioaktivního datování může čtenář nalézt v [1].

Také v případě úlohy (6) jsme schopni, pokud známe počet atomů olova-210 v gramu oxidu olovnatého v době výroby olovnaté běloby určit stáří obrazu, na kterém je tato barva použita. K řešení této úlohy můžeme použít Eulerovu metodu, se kterou jste se seznámili dříve. V naší úloze počet atomů olova-210 v gramu oxidu olovnatého v době výroby barvy bohužel neznáme. I přesto jsme schopni rozlišit obraz jehož stáří je 300 let od obrazu, který má 20 let. Je totiž známo, jaké bývají koncentrace radioaktivního olova-210 v rudách, ze kterých se vyrábí oxid olovnatý. Na Zemi není známo ložisko, které by obsahovalo rudu s koncentrací větší než $5 \cdot 10^{11}$ atomů olova-210 v gramu rudy. Proto můžeme zjistit, pokud známe potřebné parametry, zda je možné, aby bylo stáří obrazu 300 let.

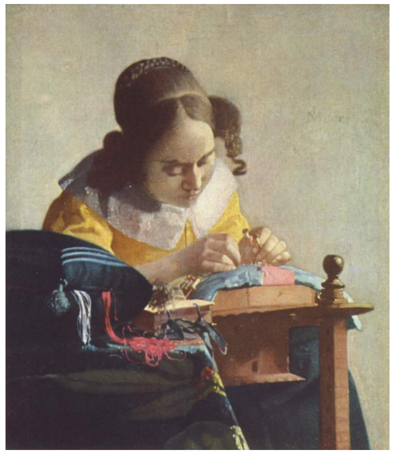
Úkol 2.1 Je možné, aby obraz Emauzští učedníci byl v roce 1967 starý cca 300 let? V roce 1967 bylo změřeno, že v čase $t = 1967$ platí

$$N(t) = 1{,}42 \cdot 10^8, \;\; r = 420480.$$

Obrázek 4: Emauzští učedníci - Museum Boijmans Van Beuningen, Rotterdam

Úkol 2.2 Je možné, aby obraz Krajkářka byl v roce 1967 starý cca 300 let? V roce 1967 bylo změřeno, že v čase $t = 1967$ platí

$$N(t) = 0{,}25 \cdot 10^8, \; r = 735840.$$



Obrázek 5: Krajkářka - Louvre, Paříž

Úkol 2.3 Je možné, aby obraz Voják a smějící se děvče byl v roce 1967 starý cca 300 let? V roce 1967 bylo změřeno, že v čase $t = 1967$ platí

$$N(t) = 1 \cdot 10^8, \; r = 2733120.$$

Obrázek 6: Voják a smějící se děvče - Frick Collection, New York

Cvičení 3: Koluje Ti Google krví? Statistika napoví!

Kdo by neznal Google – amerického giganta se sídlem v Mountain View v Silicon Valley v USA, jehož název byl odvozen od krásného čísla Googol, které má hodnotu 10^{100}. Chtěli byste v něm pracovat? Pokud ano, máte na to? Díky serveru Kaggle (online komunita uživatelů zajímajících se o data science) máme k dispozici data o zhruba 1200 inzerátech na pracovní pozice právě do Googlu [6]. Jaké musíte mít vzdělání a kolik let praxe? Které programovací jazyky jsou nezbytné? Na tyto otázky se pokusíme odpovědět pomocí statistické analýzy dat a třeba zjistíte, že jste stvořeni pro Google...

Data se zpravidla zapisují do tzv. `datové matice`, což není nic jiného než tabulka, jejíž počet řádků je určen počtem pozorování a počet sloupců je určen počtem proměnných. Podívejte se na obrázek 7, kde je zobrazena část datové matice, se kterou budeme dále pracovat.

ID	Nazev_pozice	Kategorie	Misto_vykonu_mesto	Misto_vykonu_s
1	Google Cloud Program Manager	Program Management	NA	Singapore
2	Supplier Development Engineer (SDE), Ca...	Manufacturing and Supply Chain	Shanghai	China
3	Data Analyst, Product and Tools Operatio...	Technical Solutions	New York	United States
4	Developer Advocate, Partner Engineering	Developer Relations	Mountain View	United States
5	Program Manager, Audio Visual (AV) Dep...	Program Management	Sunnyvale	United States
6	Associate Account Strategist (Czech/Slov...	Technical Solutions	Dublin	Ireland
7	Supplier Development Engineer, Camera, ...	Hardware Engineering	Mountain View	United States
8	Strategic Technology Partner Manager, H...	Partnerships	Sunnyvale	United States
9	Manufacturing Business Manager, Googl...	Manufacturing and Supply Chain	Xinyi District	Taiwan

Obrázek 7: Část datové matice

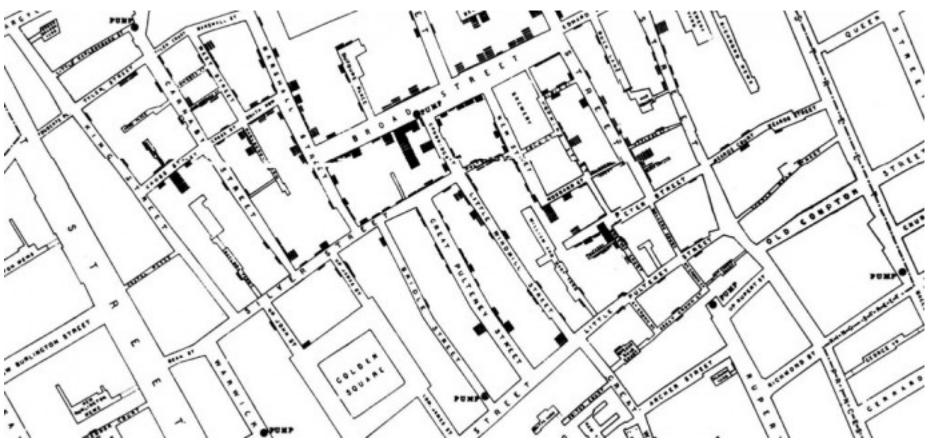
Každý řádek obsahuje informace o jednom pracovním inzerátu. Níže je uveden seznam všech proměnných se stručným popisem:

- `ID` – identifikační číslo záznamu,
- `Nazev_pozice` – název pracovní pozice,
- `Kategorie` – zařazení pracovní pozice do kategorie,
- `Misto_vykonu_mesto` – místo výkonu práce – město,
- `Misto_vykonu_stat` – místo výkonu práce – stát,
- `Min_kvalif` – minimální kvalifikace a znalosti potřebné pro danou pracovní pozici,
- `Pref_kvalif` – další preferované znalosti a kvalifikace.

Při prvním náhledu na celá data asi zjistíte, že nebude tak jednoduché odpovědět na výše položené otázky. Budeme se totiž muset vypořádat s tím, že požadavky na pracovní pozice jsou sepsány do vět. Naučíme se tedy pracovat s takto zapsanými informacemi a ukážeme si funkce `grep` a `grepl`, které nám umožní vyhledávat požadovaná klíčová slova v delším textu.

Se získanými informacemi pak budeme pracovat dále a pokusíme se je i `vizualizovat` pomocí knihovny `ggplot2`. Velmi důležitou úlohou statistiky je totiž nejen efektní, ale hlavně efektivní vizualizace dat, jejímž účelem je převedení dat a informací do grafické podoby, což umožní proniknout do zkoumaného problému, pochopit pozorované jevy nebo identifikovat hlavní trendy. Právě volba vhodného grafického zobrazení, vhodného grafu či diagramu, je největším úskalím kvalitně provedené vizualizace. Jaké jsou různé možnosti vizualizace dat, můžete najít v katalogu grafických nástrojů [4].

Podíváme-li se kousek do historie, pravděpodobně o jednu z prvních vizualizací dat se pokusil již v 19. století anglický lékař, který zmapoval šíření cholery v Londýně a identifikoval tak zdroj nákazy. Tento lékař se jmenoval `John Snow` a na jeho znázornění šíření cholery se můžete podívat na obrázku 8. Zájemci se poté mohou podívat na celý článek o tomto vědci vydaný v britském deníku The Guardian [5].



Obrázek 8: Vizualizace šíření cholery z 19. století

Cílem cvičení je seznámit se s funkcemi `grep` a `grepl` pro práci s tzv. řetězci. Dále si vyzkoušet práci s knihovnou `dplyr`, která poskytuje funkce pro práci s datovou maticí, např. funkce `filter` nebo `select`. V průběhu analýzy totiž výzkumníka můžou zajímat pouze pozorování, která splňují určitou podmínku, nebo jen některé proměnné – v našem případě se třeba můžeme zajímat jen o ty pracovní pozice nabízené v České republice nebo na Slovensku. Na závěr si pak ukážeme základní principy vizualizace dat s knihovnou `ggplot2`.

Reference

[1] M. Braun: Differential Equations and Their Applications. Springer Verlag (1993).

[2] K. Konečná, J. Koláček: Jak pracovat s jazykem R. [cit. 2.1.2020]. Dostupné z: `https://www.math.muni.cz/~kolacek/vyuka/vypsyst/navod_R.pdf`

[3] B. Krajc, P. Beremlijski: Obyčejné diferenciální rovnice. [cit. 2.1.2020]. Text vytvořený při realizaci projektu „Matematika pro inženýry 21. století“, Vysoká škola báňská - Technická univerzita Ostrava (2012). Dostupné z: `http://mi21.vsb.cz/modul/obycejne-diferencialni-rovnice`

[4] *The Data Visualisation Catalogue* [online]. [cit. 2.1.2020]. Dostupné z: `https://datavizcatalogue.com/`

[5] *The Guardian* [online]. John Snow's data journalism: the cholera map that changed the world. [cit. 2.1.2020]. Dostupné z: `https://www.theguardian.com/news/datablog/2013/mar/15/john-snow-cholera-map`

[6] *Kaggle* [online]. Google Job Skills. [cit. 2.1.2020]. Dostupné z: `https://www.kaggle.com/niyamatalmass/google-job-skills`

Definice A.1 Buď $f: \mathbb{R} \to \mathbb{R}$ a $x \in \mathbb{R}$. Existuje-li

$$\lim_{h\to 0} \frac{f(x+h) - f(x)}{h},$$

značíme ji $f'(x)$ a nazýváme `derivací funkce` f `v bodě` x.

Poznámka A.1 Většinou – a nejinak je to v tomto textu – se pod pojmem derivace rozumí konečná (tzv. `vlastní`) derivace.

Věta A.1 Buď $f, g: \mathbb{R} \to \mathbb{R}$ a $x \in \mathbb{R}$. Pak platí

- $(f \pm g)'(x) = f'(x) \pm g'(x)$, má-li pravá strana rovnosti smysl,
- $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$, existují-li (vlastní) derivace $f'(x)$ a $g'(x)$,
- $\left(\dfrac{f}{g}\right)'(x) = \dfrac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}$, existují-li (vlastní) derivace $f'(x)$ a $g'(x)$ a je-li $g(x) \neq 0$.

Pozorování A.1

- $(c)' = 0,\ c \in \mathbb{R}$ (konst.), $x \in \mathbb{R}$,
- $(x^r)' = rx^{r-1},\ r \in \mathbb{R},\ x \in (0, +\infty)$,
- $(\sin x)' = \cos x,\ x \in \mathbb{R}$,
- $(\cos x)' = -\sin x,\ x \in \mathbb{R}$,
- $(\operatorname{tg} x)' = \dfrac{1}{\cos^2 x},\ x \in \mathbb{R} \setminus \left\{\dfrac{\pi}{2} + k\pi :\ k \in \mathbb{Z}\right\}$,
- $(\operatorname{cotg} x)' = -\dfrac{1}{\sin^2 x},\ x \in \mathbb{R} \setminus \{k\pi :\ k \in \mathbb{Z}\}$,
- $(\mathrm{e}^x)' = \mathrm{e}^x,\ x \in \mathbb{R}$,
- $(\ln x)' = \dfrac{1}{x},\ x \in (0, +\infty)$.