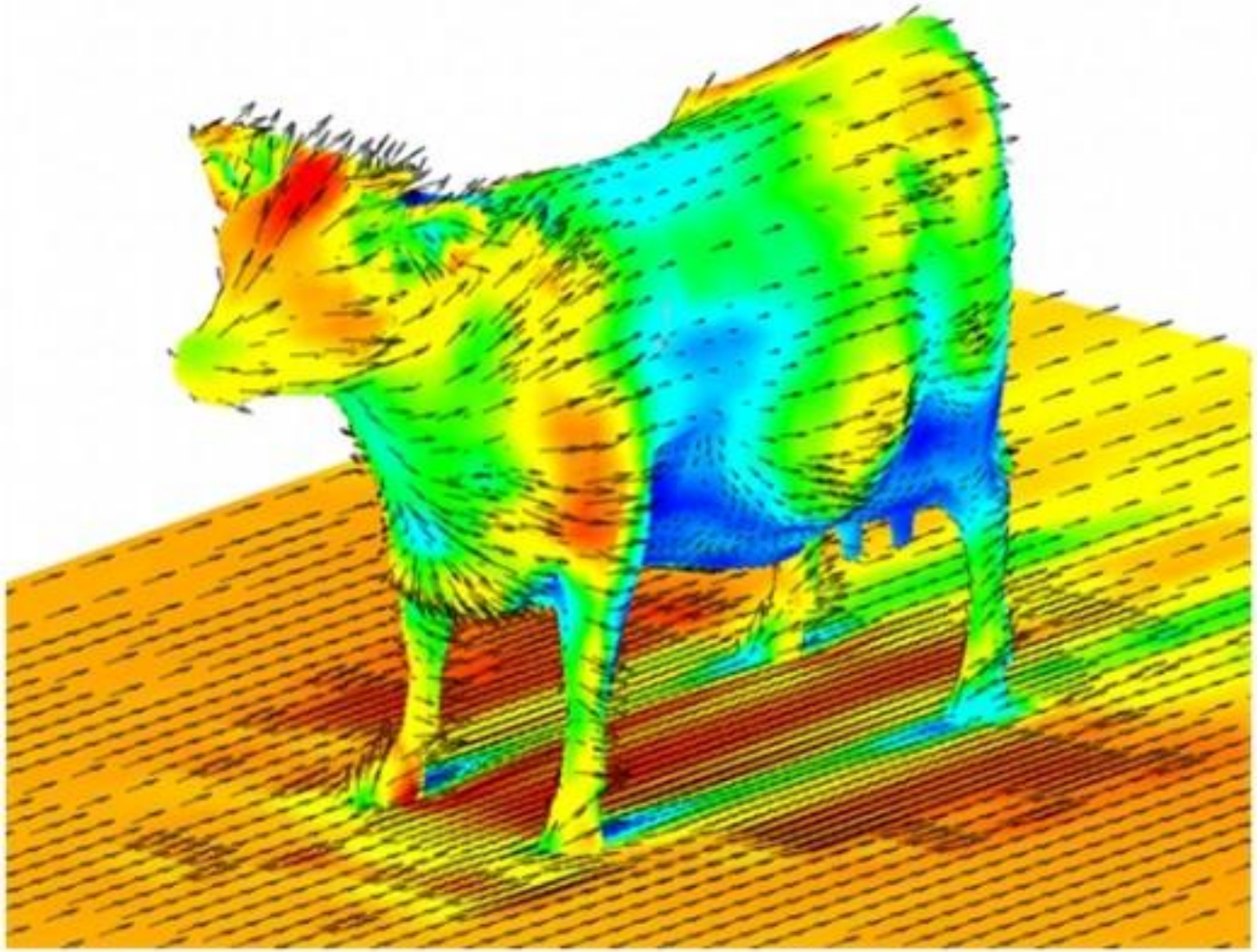
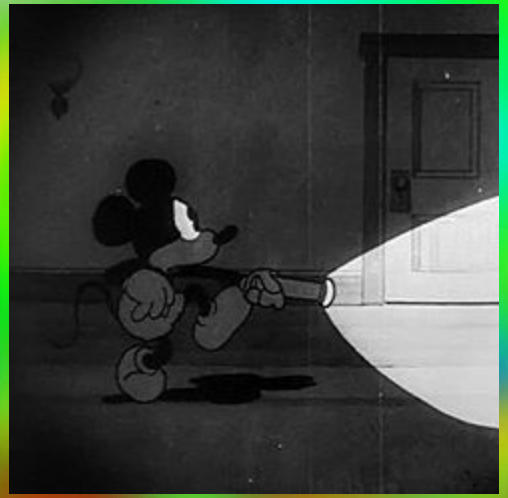
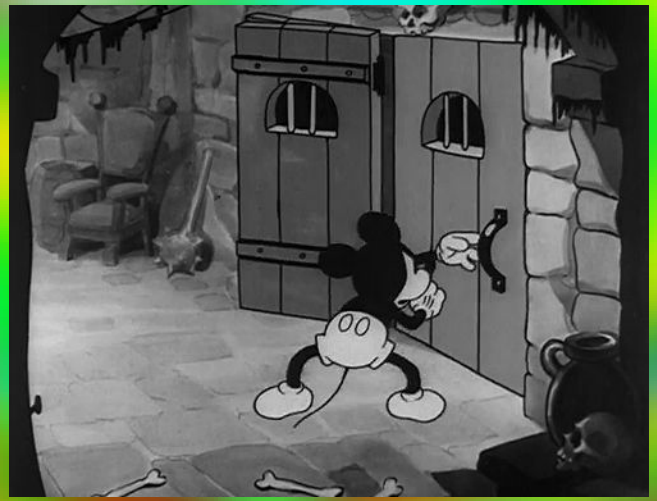
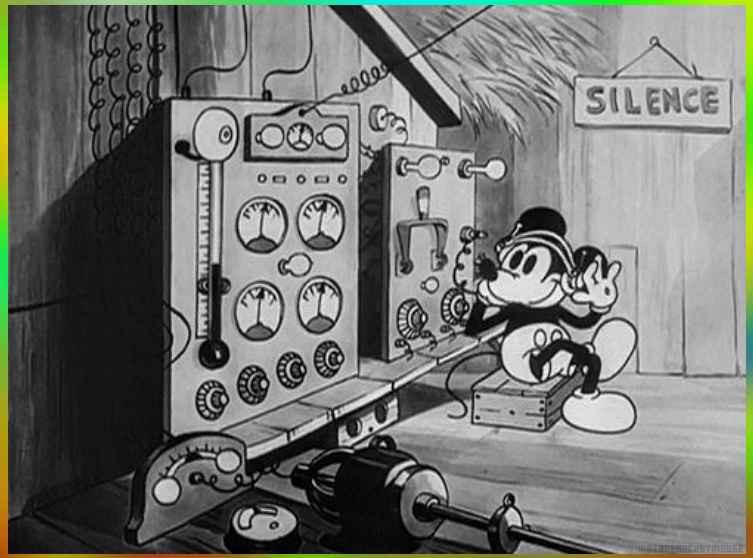
VIZU ALIZ ACE DAT



7. ZÍSKÁVÁNÍ A ANALÝZA DAT






SILENCE

Petr Koubský
38 mins · Prague ·

Kdo v těchto volbách prohrál opravdu katastrofálně, to jsou datoví analytici. Vůbec to nemyslím jako příspěvek do soutěže o nejsarkastičtější komentář, je to vážná věc sama o sobě. Bude se teď blbě vysvětlovat, že velká data atd., když chroustači dat do jednoho spáchali morální sebevraždu v přímém přenosu. (Nate Silver!!) Garbage in garbage out – pilovali detaily matematických modelů a ignorovali, že do nich cpou zcela nespolehlivé údaje. Život se ukázal zas jednou být složitější než matematika. Otřese to důvěrou v celý obor.

Hillary Clinton has an **85% chance** to win.

Last updated Tuesday, November 8 at 10:20 PM ET

CHANCE OF WINNING

85%
Hillary Clinton

15%
Donald J. Trump

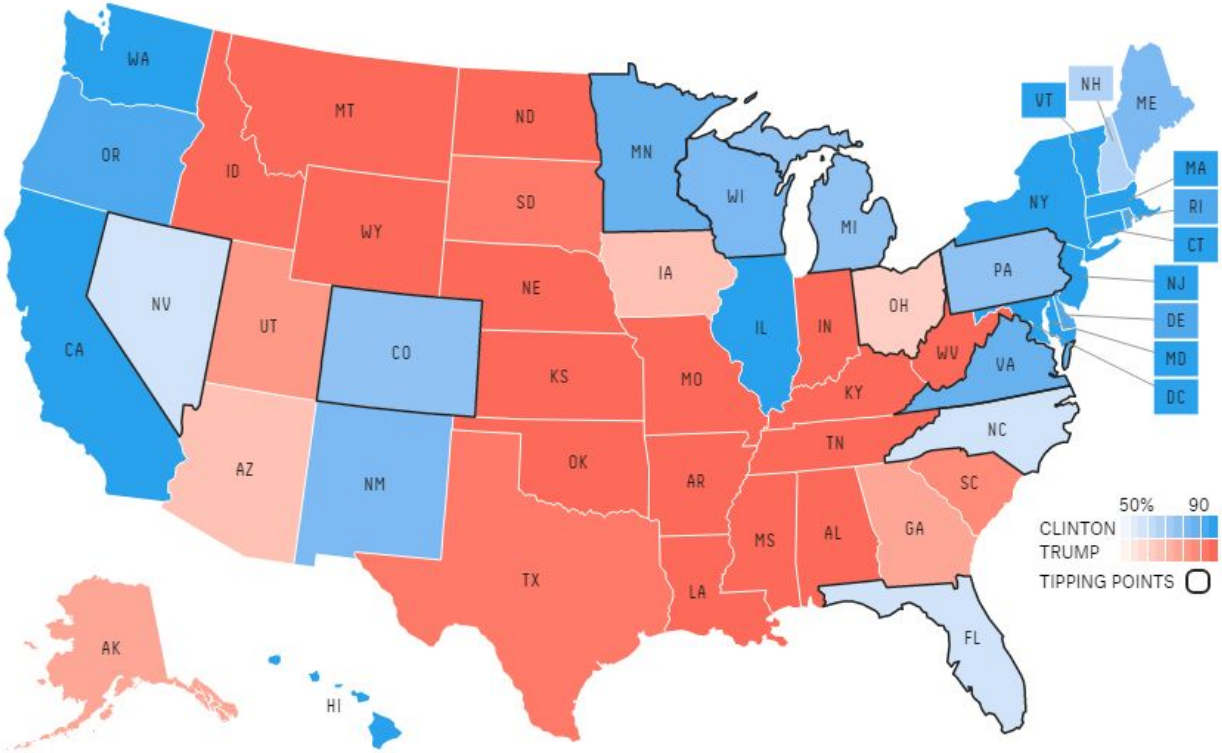
Chance of winning

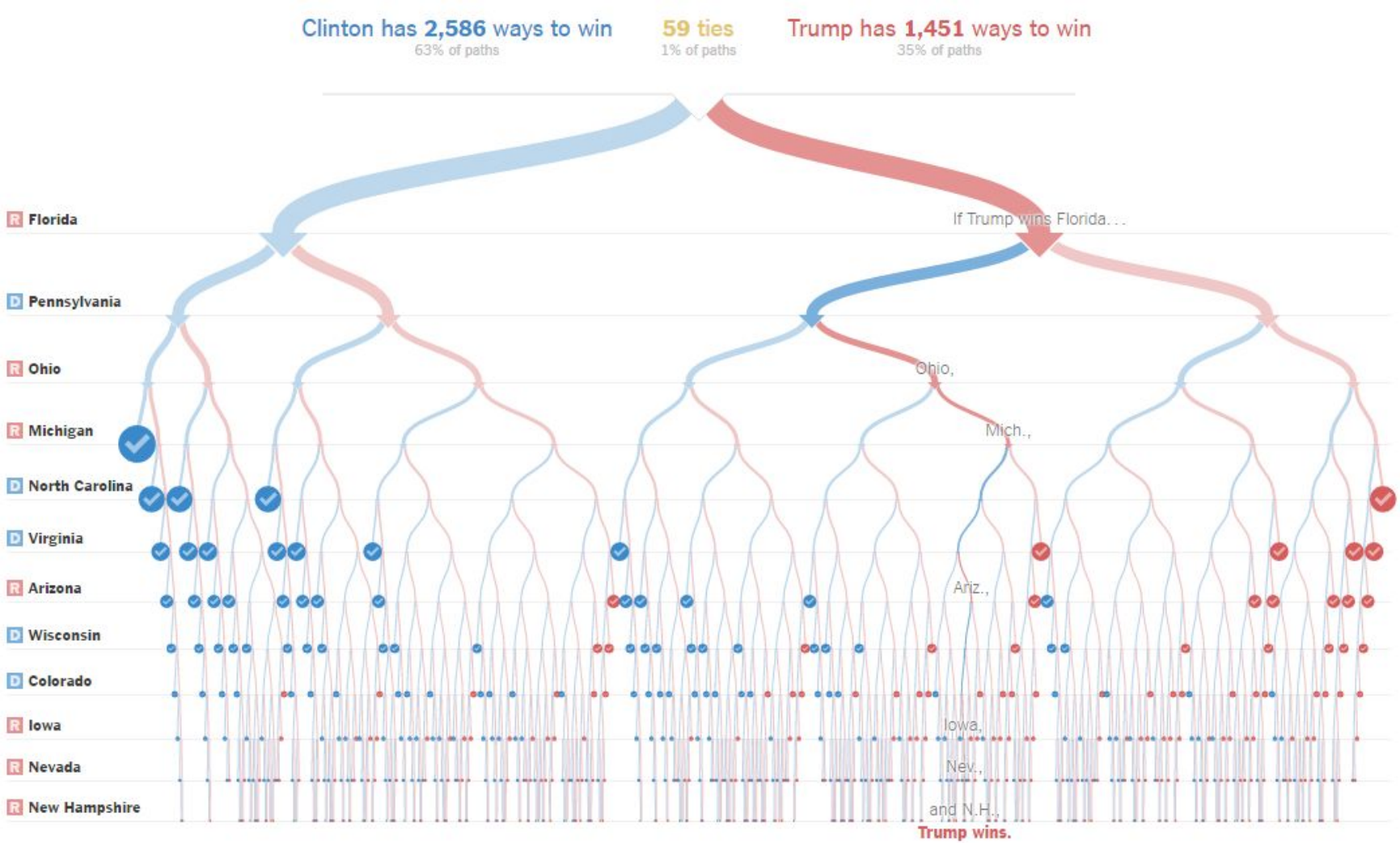
Hillary Clinton

71.4%

Donald Trump

28.6%

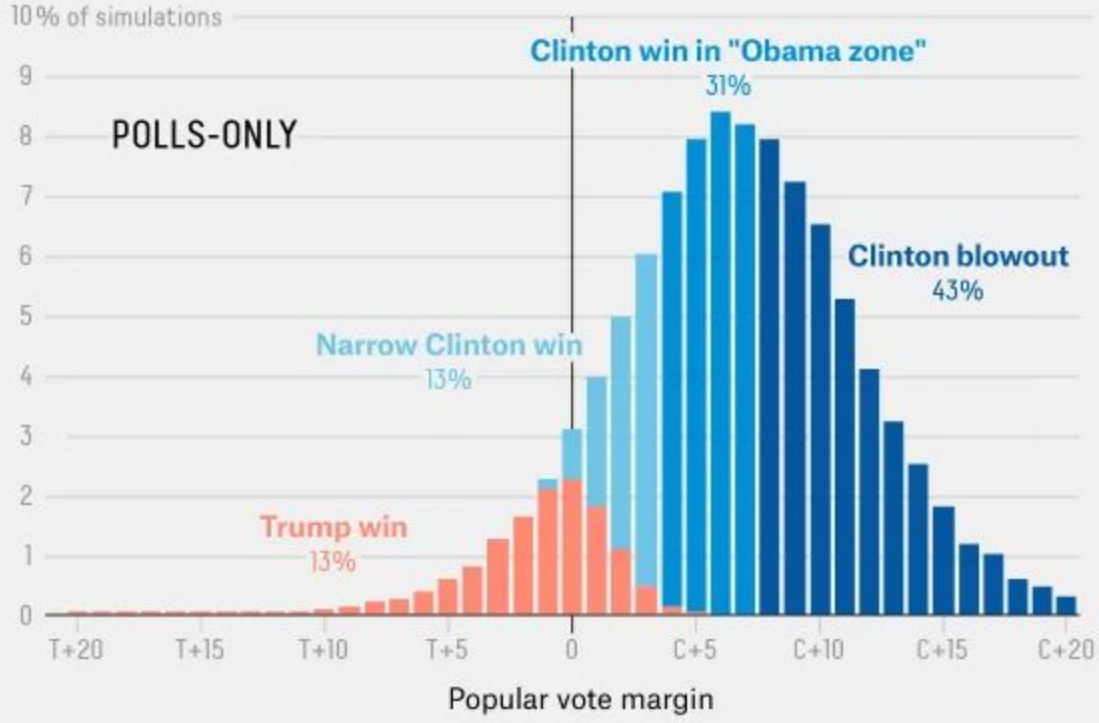



Clinton has 2,586 ways to win
63% of paths
59 ties
1% of paths
Trump has 1,451 ways to win
35% of paths
R Florida
D Pennsylvania
R Ohio
R Michigan
D North Carolina
D Virginia
R Arizona
D Wisconsin
D Colorado
R Iowa
R Nevada
R New Hampshire
If Trump wins Florida...
Ohio,
Mich.,
Ariz.,
Iowa,
Nev.,
and N.H.,
Trump wins.

Win Presidency
TOSSUP
LEANING
LEANING
LIKELY
LIKELY
VERY LIKELY
VERY LIKELY
>95% Trump
FORECAST
Popular vote margin
OBAMA '12
OBAMA '08
D+1
D+2
D+3
D+4
D+5
D+6
D+7
D+8
D+9
D+10
D+11
R+1
R+2
R+3
R+4
R+5
R+6
R+7
R+8
R+9
R+10
R+11
Clinton +1.1
FORECAST, in pct. points
Electoral votes
OBAMA '12
OBAMA '08
260
280
300
320
340
360
380
400
420
440
460
480
500
520
306 Trump
FORECAST

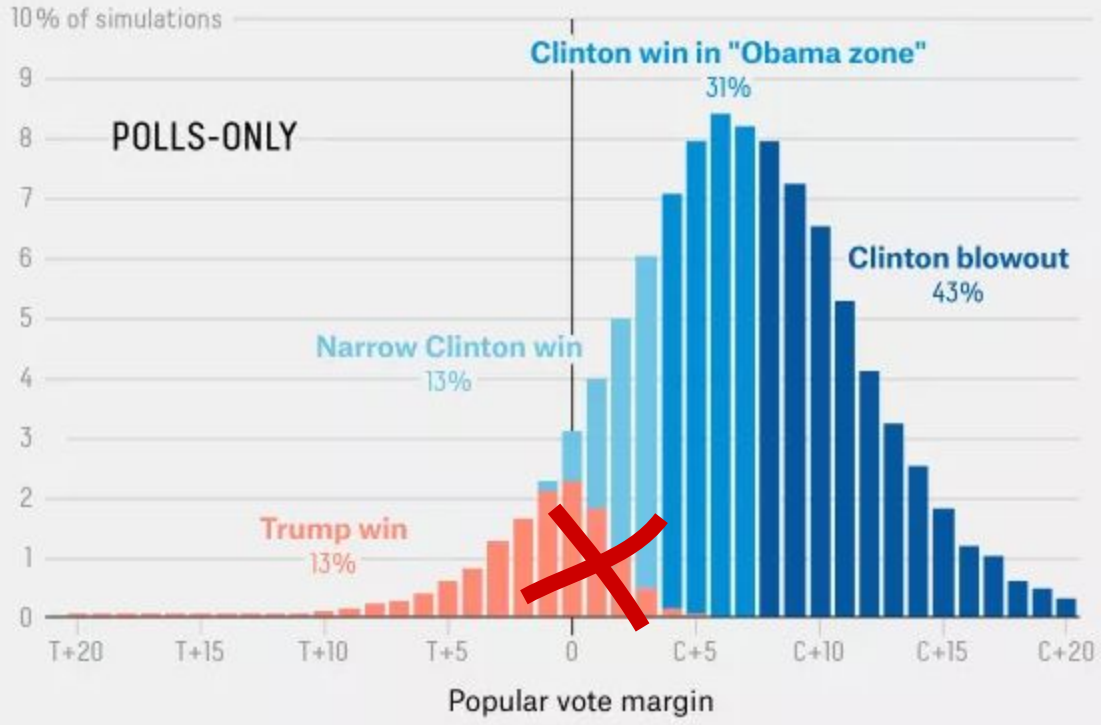
Potential outcomes of the presidential election

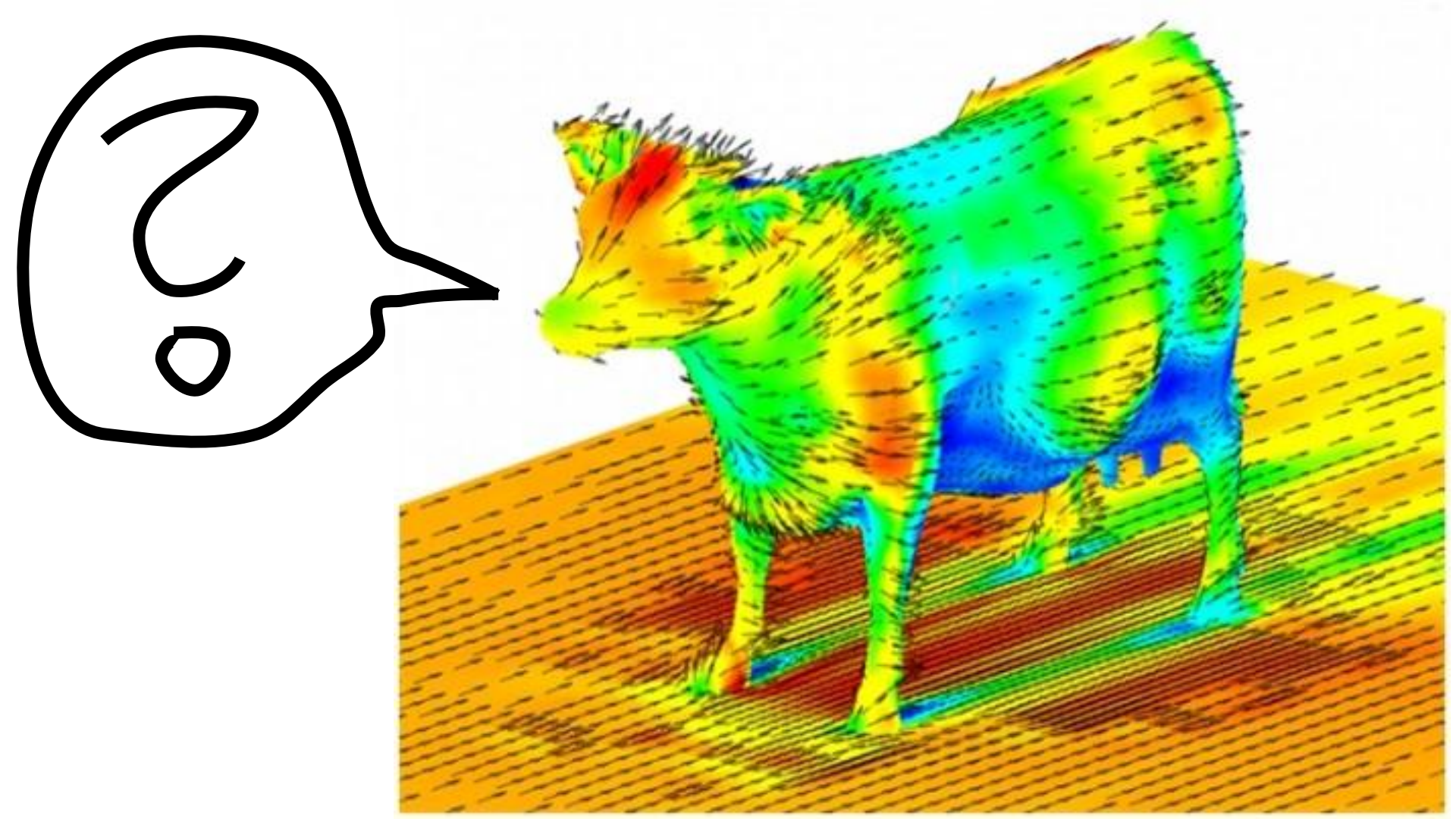
Likelihood of scenarios given by 20,000 simulations of the FiveThirtyEight polls-only and polls-plus election forecasts

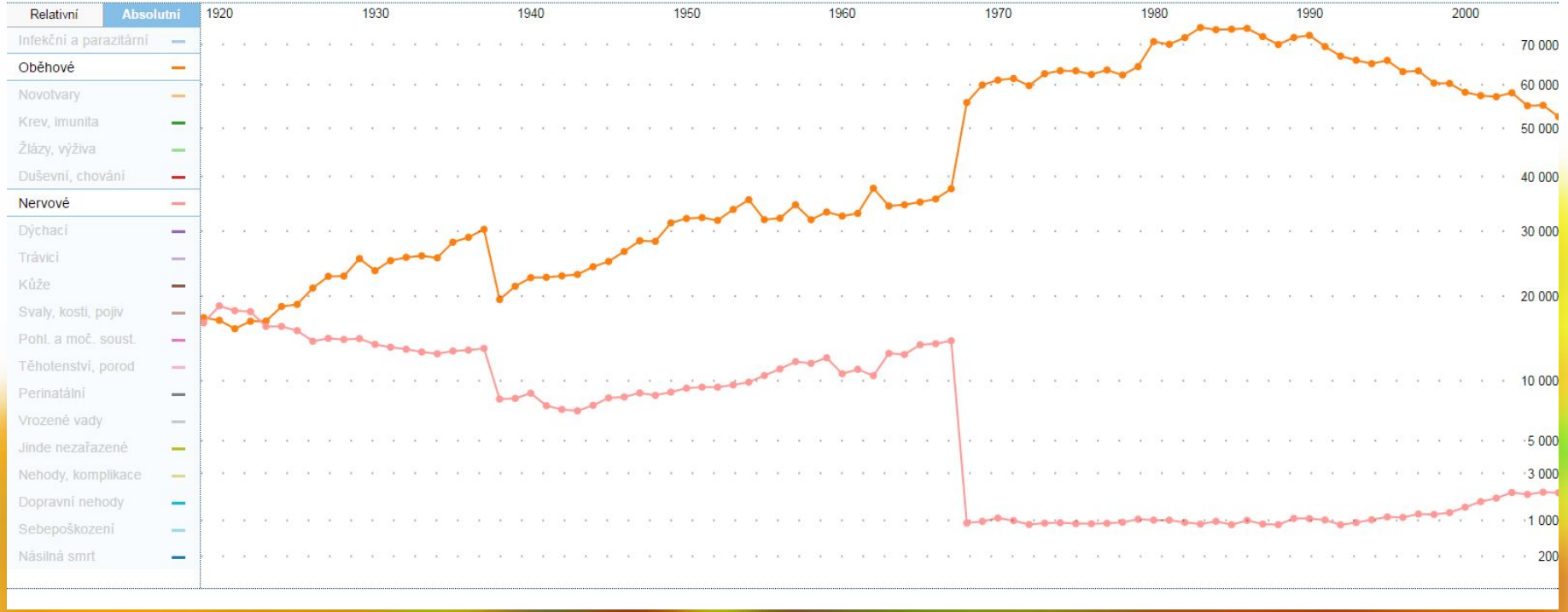


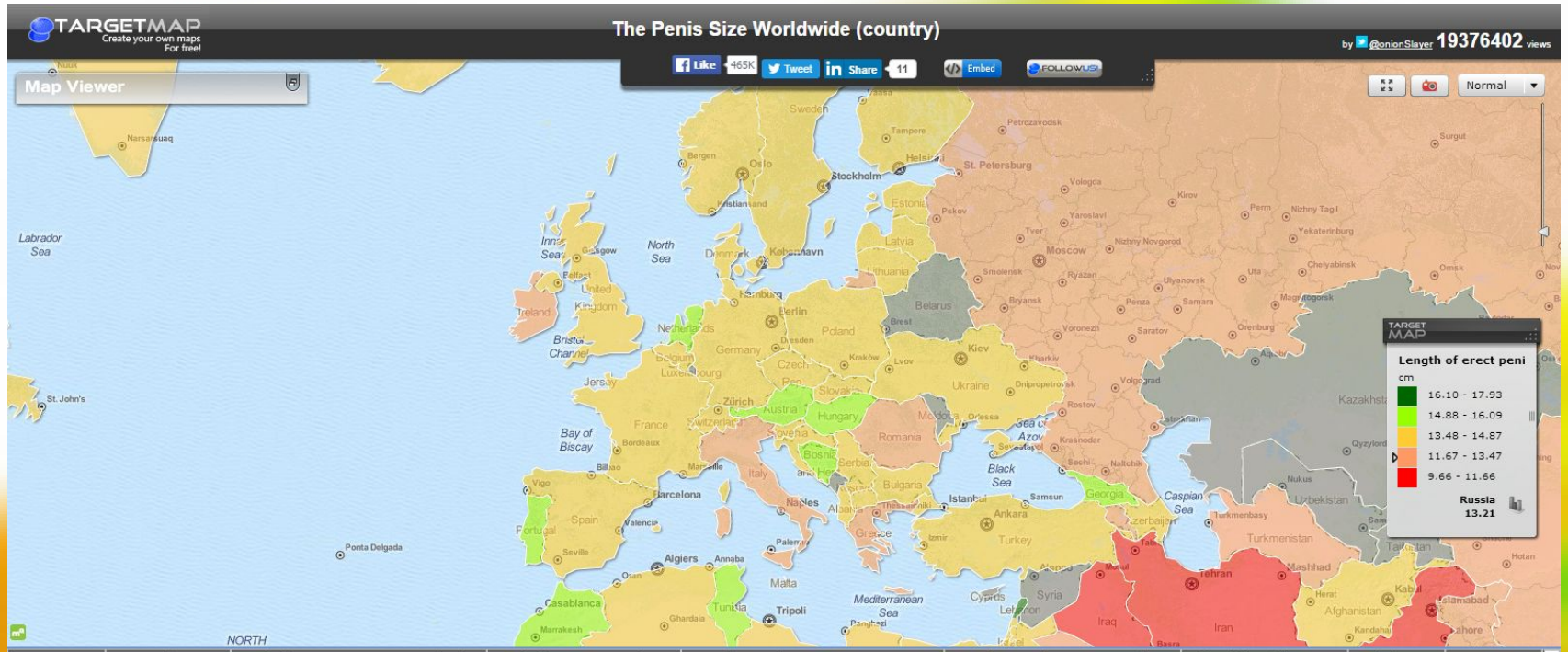
Potential outcomes of the presidential election

Likelihood of scenarios given by 20,000 simulations of the FiveThirtyEight polls-only and polls-plus election forecasts



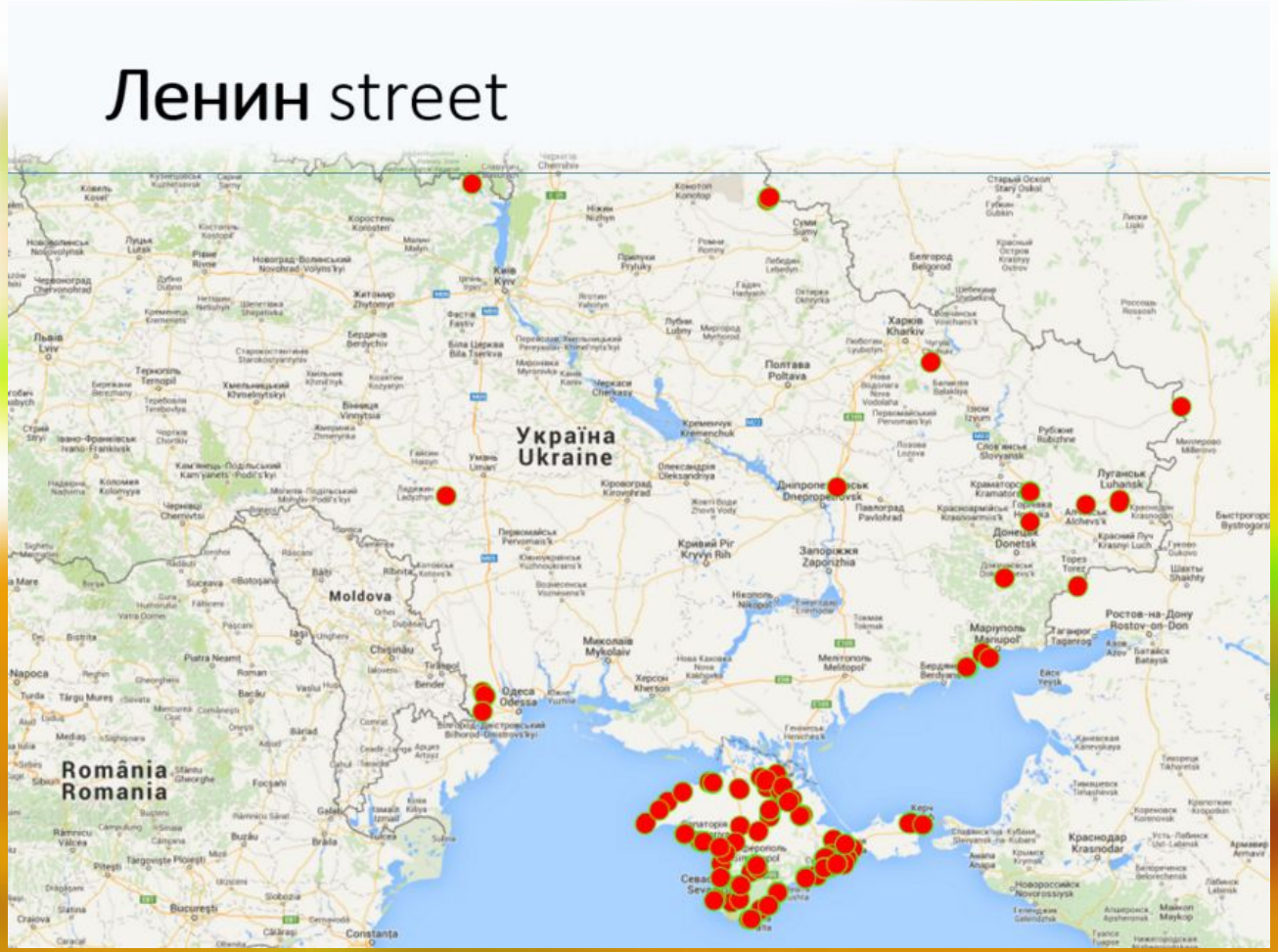



Relativní
Absolutní
Infekční a parazitární
Oběhové
Novotvary
Krev, imunita
Žlázy, výživa
Duševní, chování
Nervové
Dýchací
Trávicí
Kůže
Svaly, kosti, pojiv
Pohl. a moč. soust.
Těhotenství, porod
Perinatální
Vrozené vady
Jinde nezařazené
Nehody, komplikace
Dopravní nehody
Sebepoškození
Násilná smrt
1920
1930
1940
1950
1960
1970
1980
1990
2000
70 000
60 000
50 000
40 000
30 000
20 000
10 000
5 000
3 000
1 000
200

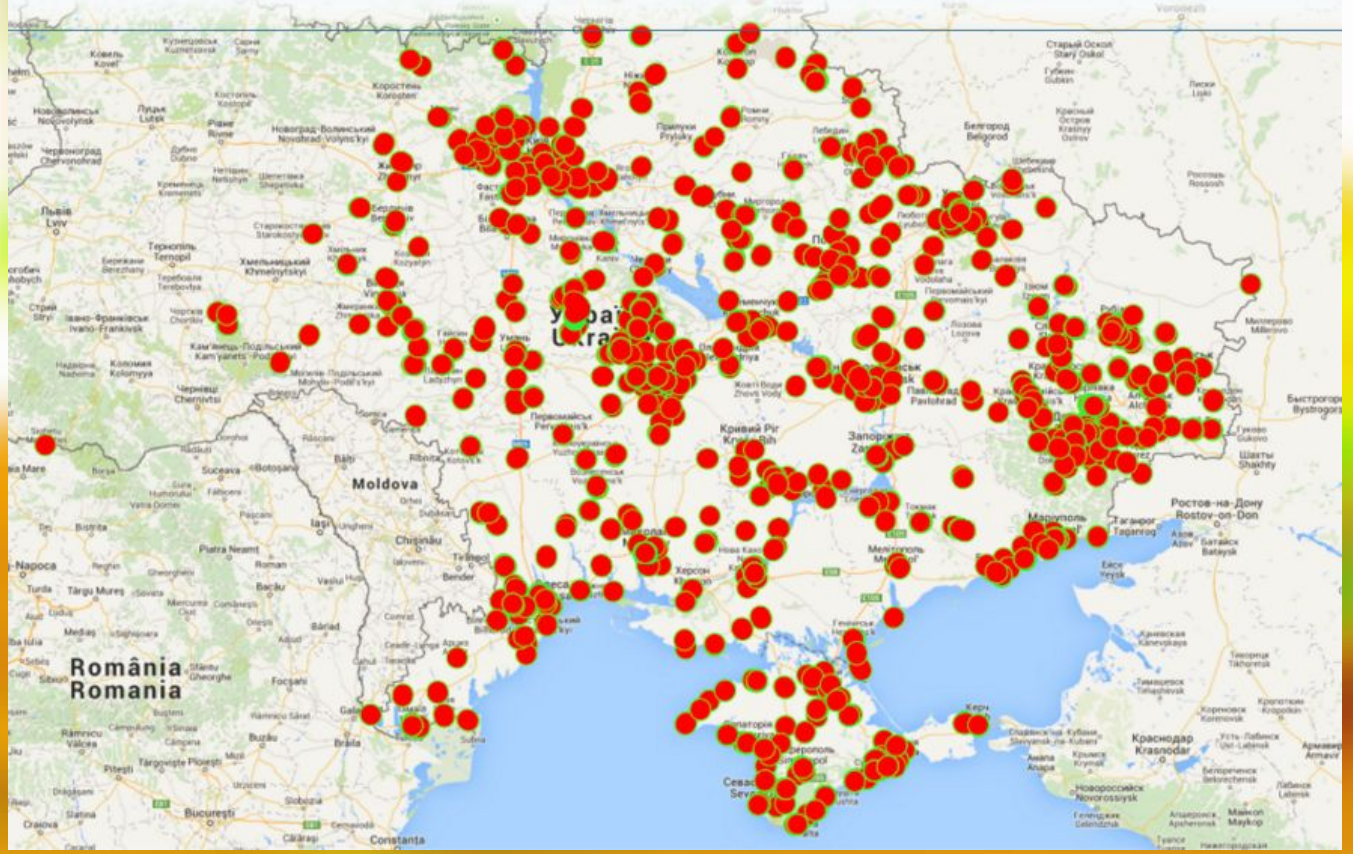


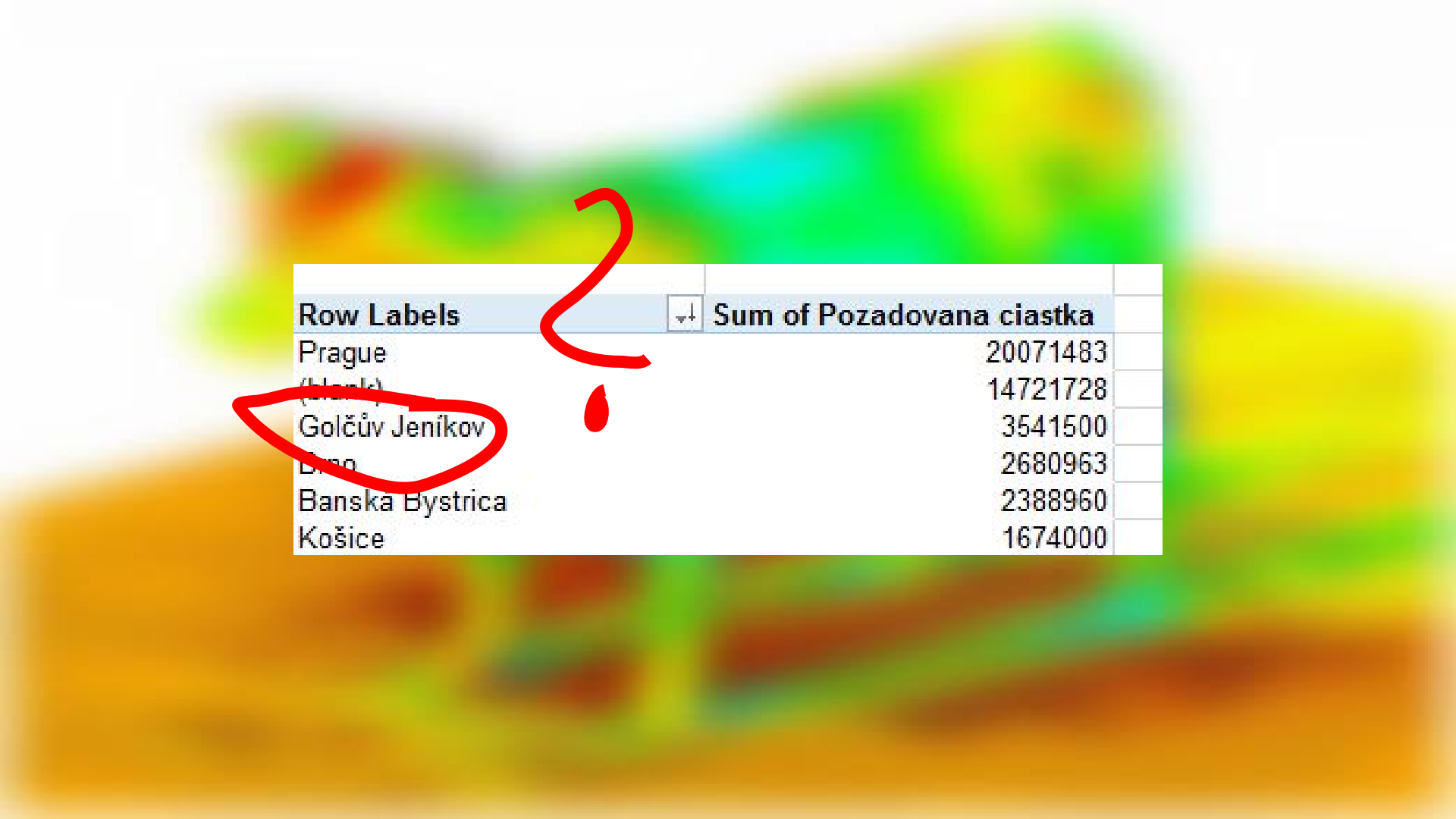
Country	Continent/Region	Length of erect penis in centimeter (SUM)	Length of erect penis in inches	Circumference of erect penis in centimeter	Circumference of erect penis in inches	Volume (n*r^2*L) in cm3	Method	N
Afghanistan	South Asia	13.69	5.4	11.42	4.5	142	Measured	100
Albania	Europe	14.19	5.6	12.16	4.79	167	Self reported	95
Algeria	Africa	14.83	5.8	10.97	4.32	142	Self reported	738
Angola	Africa	15.73	6.2	13.08	5.15	214.2	Measured	978

Ленин street



Ленин + Ленін street

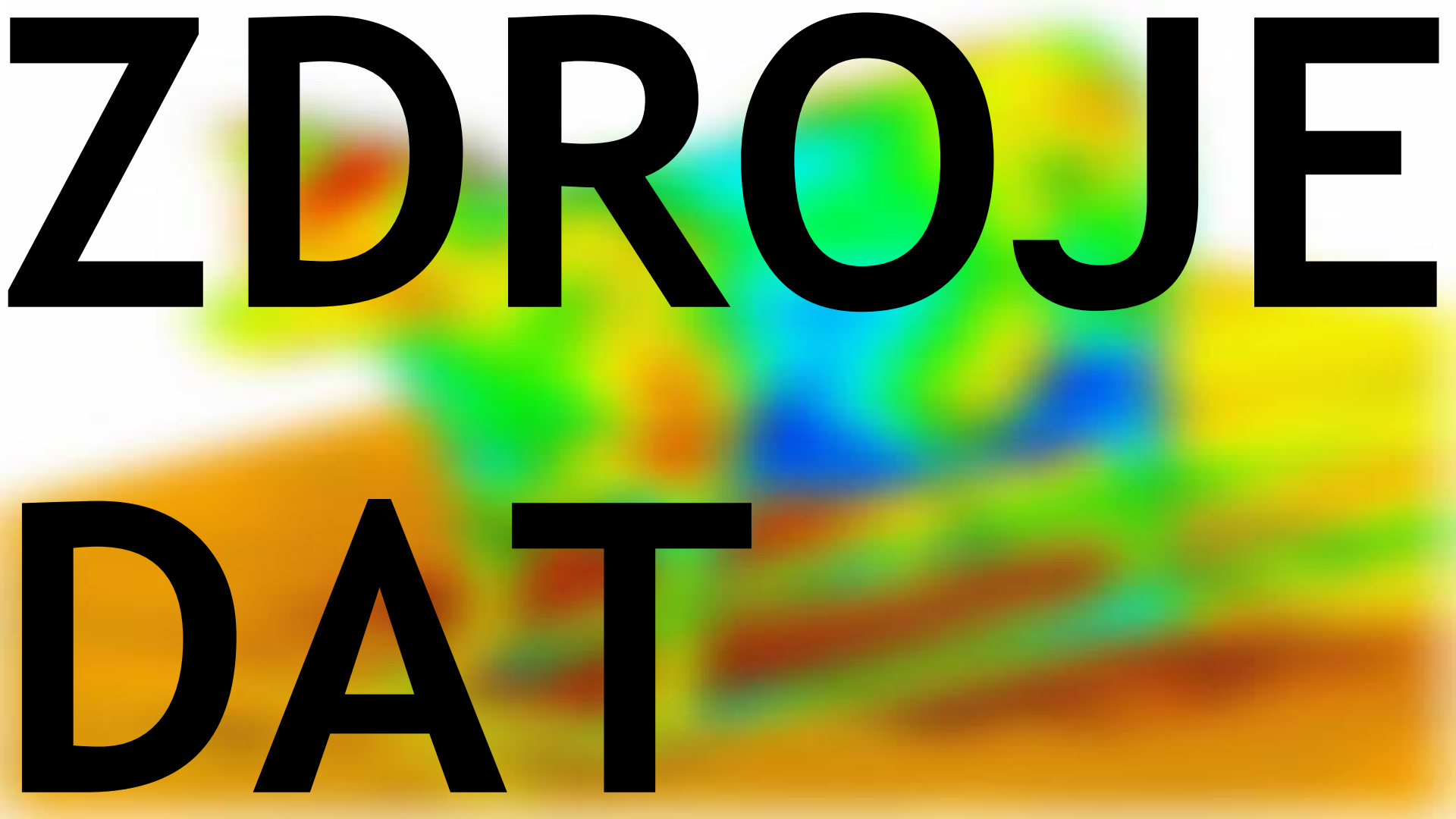




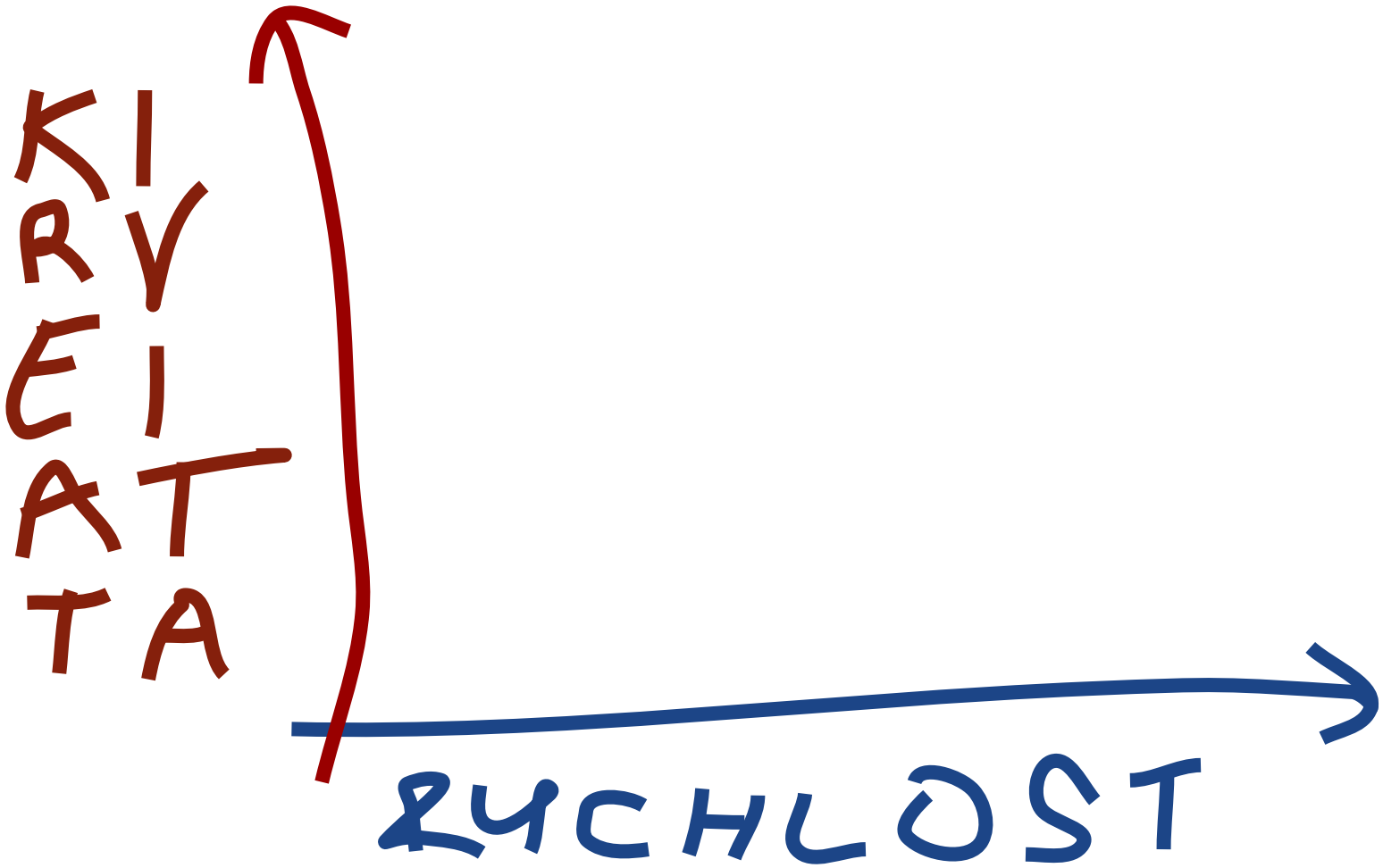
Row Labels	Sum of Pozadovana ciastka
Prague	20071483
(blank)	14721728
Golčův Jeníkov	3541500
Brno	2680963
Banska Bystrica	2388960
Košice	1674000

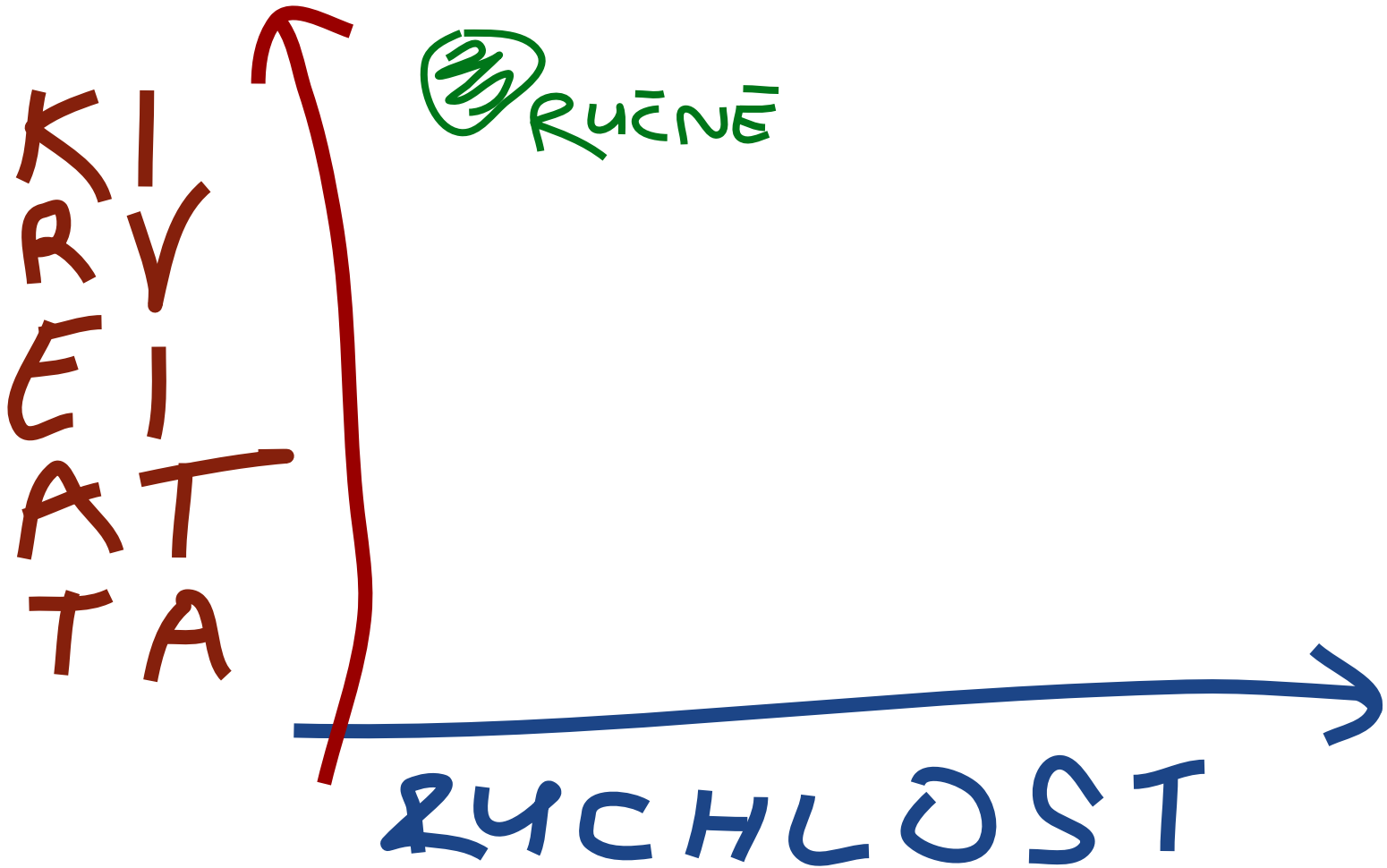
Nebodované cvičení: tabulka **Brno-střed.xls** v úschovně mapuje neobydlené byty v majetku městské části. Pomocí kontingenčních tabulek odpovězte na otázky:

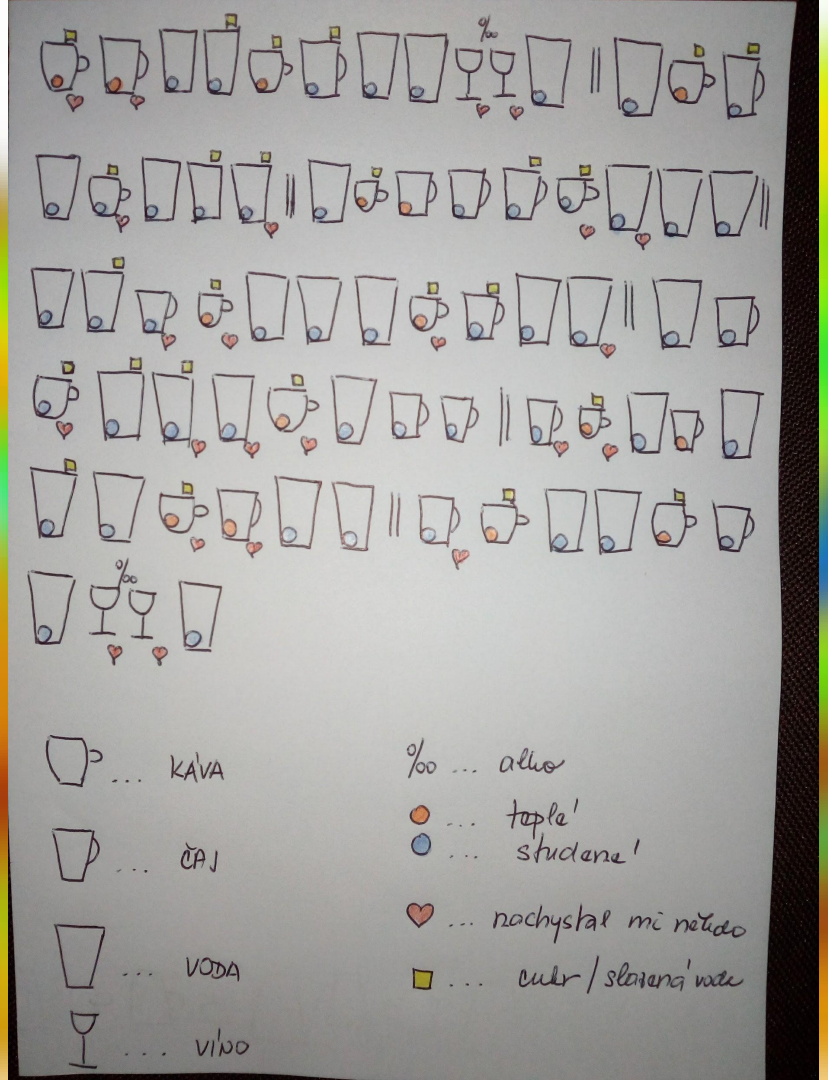
1. O kolik měsíčně Brno-střed na nepronajatých bytech přichází? (Cena nájmu se spočítá jako výměra krát cena za m2.)
2. O kolik přichází jen na Bratislavské?
3. Která ulice má průměrně nejdražší nájem?
4. Která ulice má průměrně nejdražší nájem, když nás zajímají nás pouze ulice, kde je aspoň pět nepronajatých bytů?
5. Jaký je mezi volnými byty podíl bytů 2+1? Jaká je průměrná cena nájmu v takovém bytě? (Bonusová otázka: ne průměr, ale medián ceny.)


ZDROJE
DAT

RYCHLOST

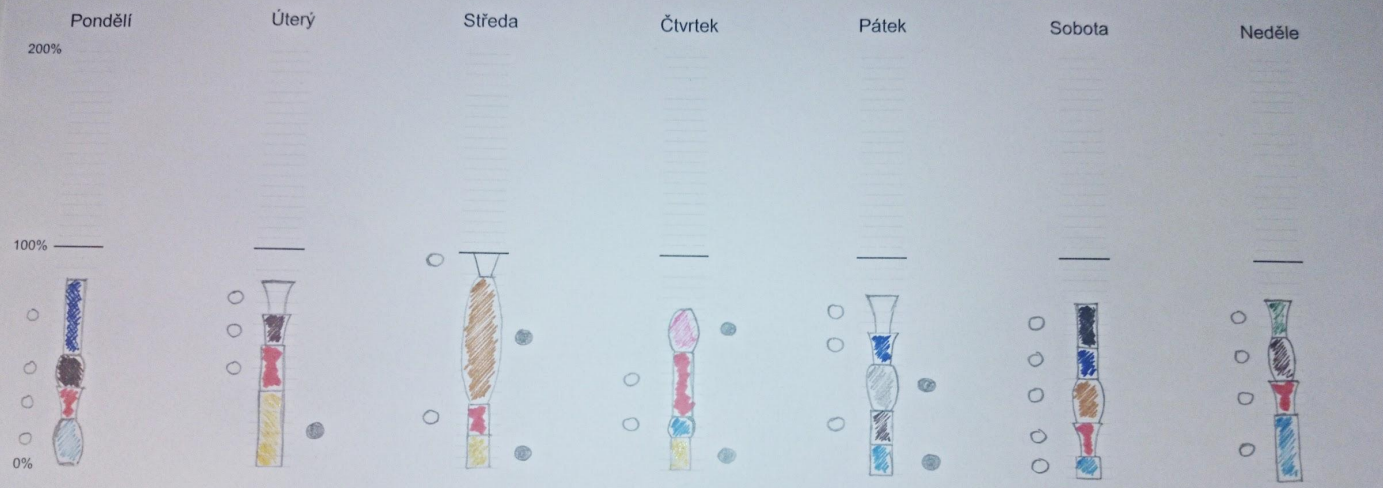

KVALITA
RYCHLOST


KVALITA
RUČNĚ
RYCHLOST


... KÁVA
... ČAJ
... VODA
... VÍNO
‰ ... alko
... teplé
... studené
... nachystal mi někdo
... cukr / sladená voda

1

DENNÍ KONZUMACE TEKUTIN VS. DOPORUČENÁ DENNÍ DÁVKA TEKUTIN



LEGENDA

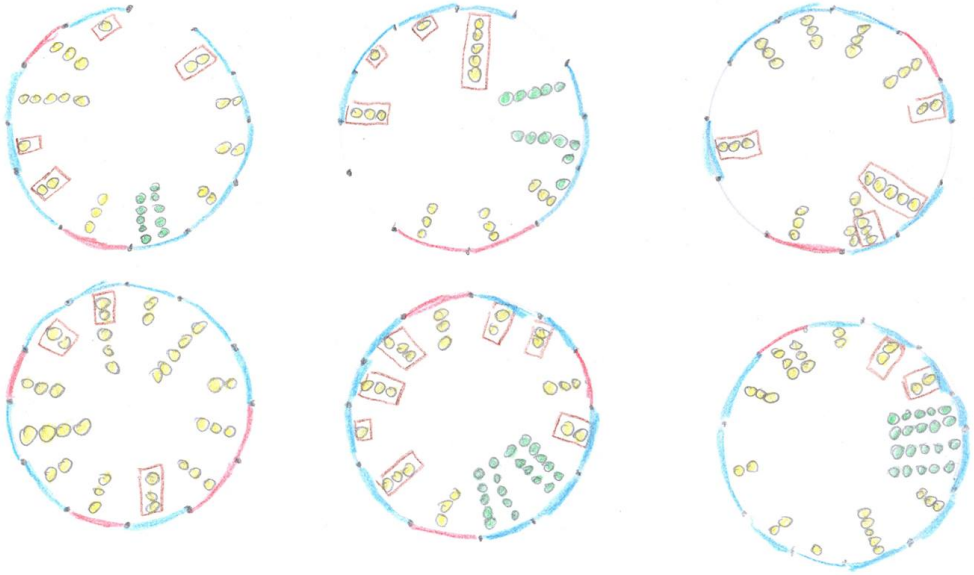
- VODA
- VODA + SIRUP
- ČAJ
- MINERÁLKA
- CLUB MATE
- MLÉKO
- PIVO
- MALINOVKA
- BEZOVKA
- KOFOLA
- DŽÚS

- OBLÁ NÁDOBA
- KONVEXNÍ
- ROVNÁ

100 % = 2,5 l

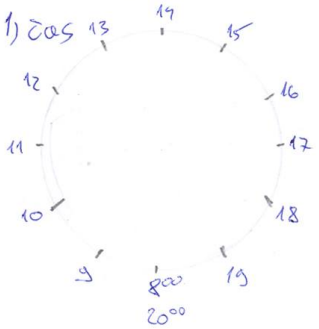
MÍSTO KONZUMACE

- DOMOV
- VENEK



Legenda

1) čas



2) teplota
- teplý
- studený

3) množství
- 1 dcl

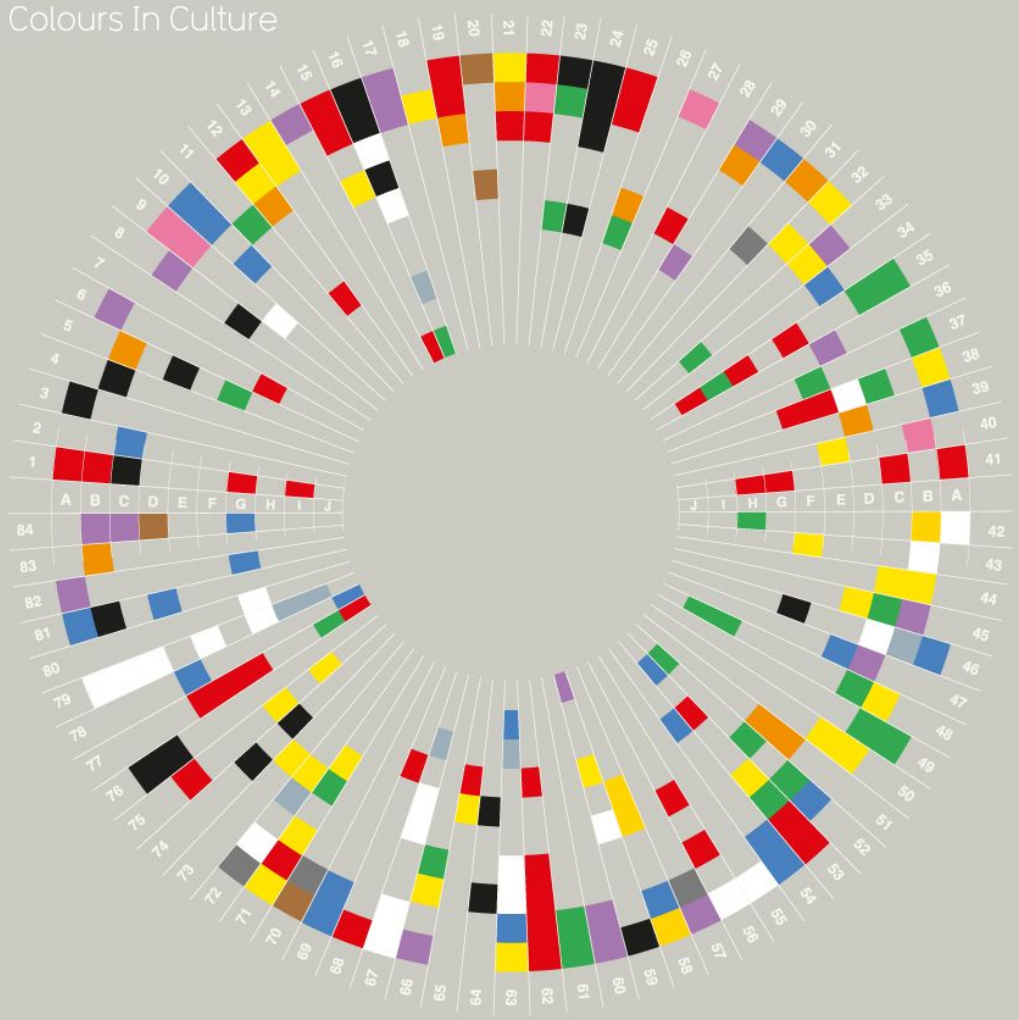
4) alkohol / nealko
- nealko
- alko

5) Plast
- balený nápoj

Každý kruh znázorňuje jeden den od pondělí. Kruhy končí pátkem a sobotou, což dokládá množství vypitého alkoholu i) Nedělí si bohužel nepamatuji, resp. data nešla možné sbírat (od rána)

Lukáš Elias, 425697

Colours In Culture



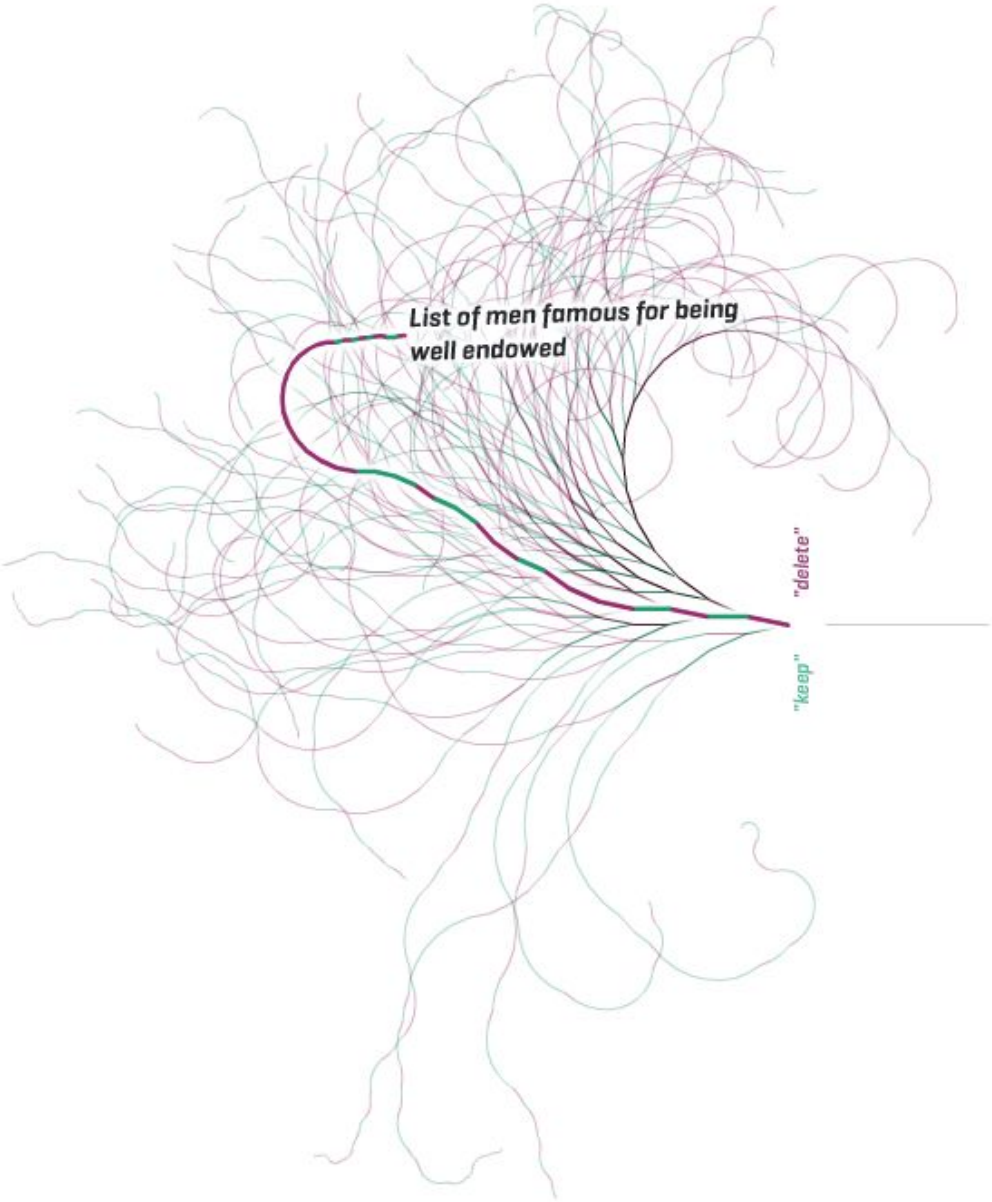
A Western / American	1 Anger	19 Desire
B Japanese	2 Art / Creativity	20 Earthy
C Hindu	3 Authority	21 Energy
D Native American	4 Bad Luck	22 Erotic
E Chinese	5 Balance	23 Eternity
F Asian	6 Beauty	24 Evil
G Eastern European	7 Calm	25 Excitement
H Muslim	8 Celebration	26 Family
I African	9 Children	27 Femininity
J South American	10 Cold	28 Fertility
	11 Compassion	29 Flamboyance
	12 Courage	30 Freedom
	13 Cowardice	31 Friendly
	14 Cruelty	32 Fun
	15 Danger	33 God
	16 Death	34 Gods
	17 Decadence	35 Good Luck
	18 Deceit	36 Gratitude

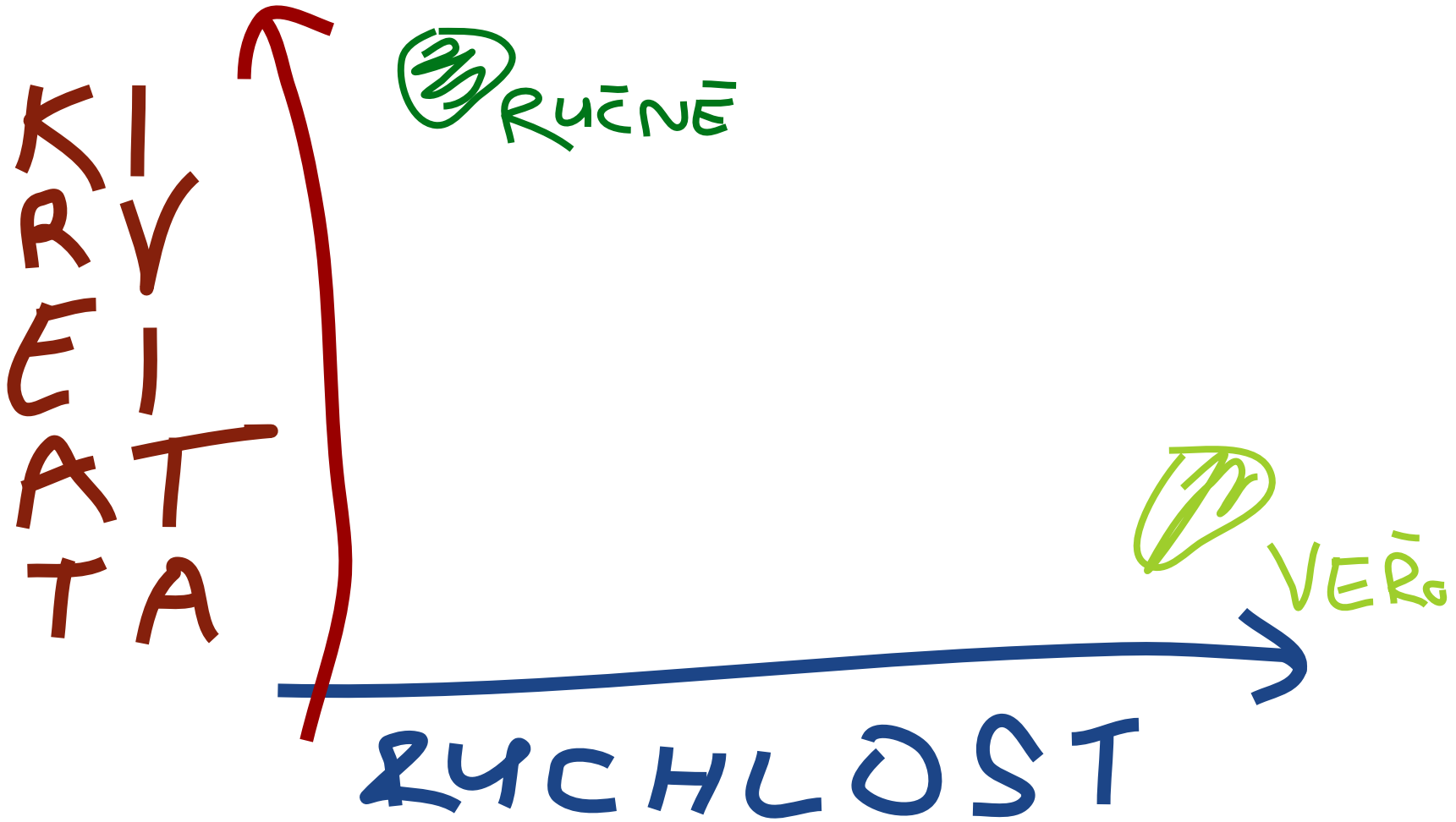
37 Growth	55 Luxury	73 Royalty
38 Happiness	56 Marriage	74 Self-cultivation
39 Healing	57 Modesty	75 Strength
40 Healthy	58 Money	76 Style
41 Heat	59 Mourning	77 Success
42 Heaven	60 Mystery	78 Trouble
43 Holiness	61 Nature	79 Truce
44 Illness	62 Passion	80 Trust
45 Insight	63 Peace	81 Unhappiness
46 Intelligence	64 Penance	82 Virtue
47 Intuition	65 Power	83 Warmth
48 Religion	66 Personal power	84 Wisdom
49 Jealousy	67 Purity	
50 Joy	68 Radicalism	
51 Learning	69 Rational	
52 Life	70 Reliable	
53 Love	71 Repels Evil	
54 Loyalty	72 Respect	

Yellow
Gold
Grey
Silver

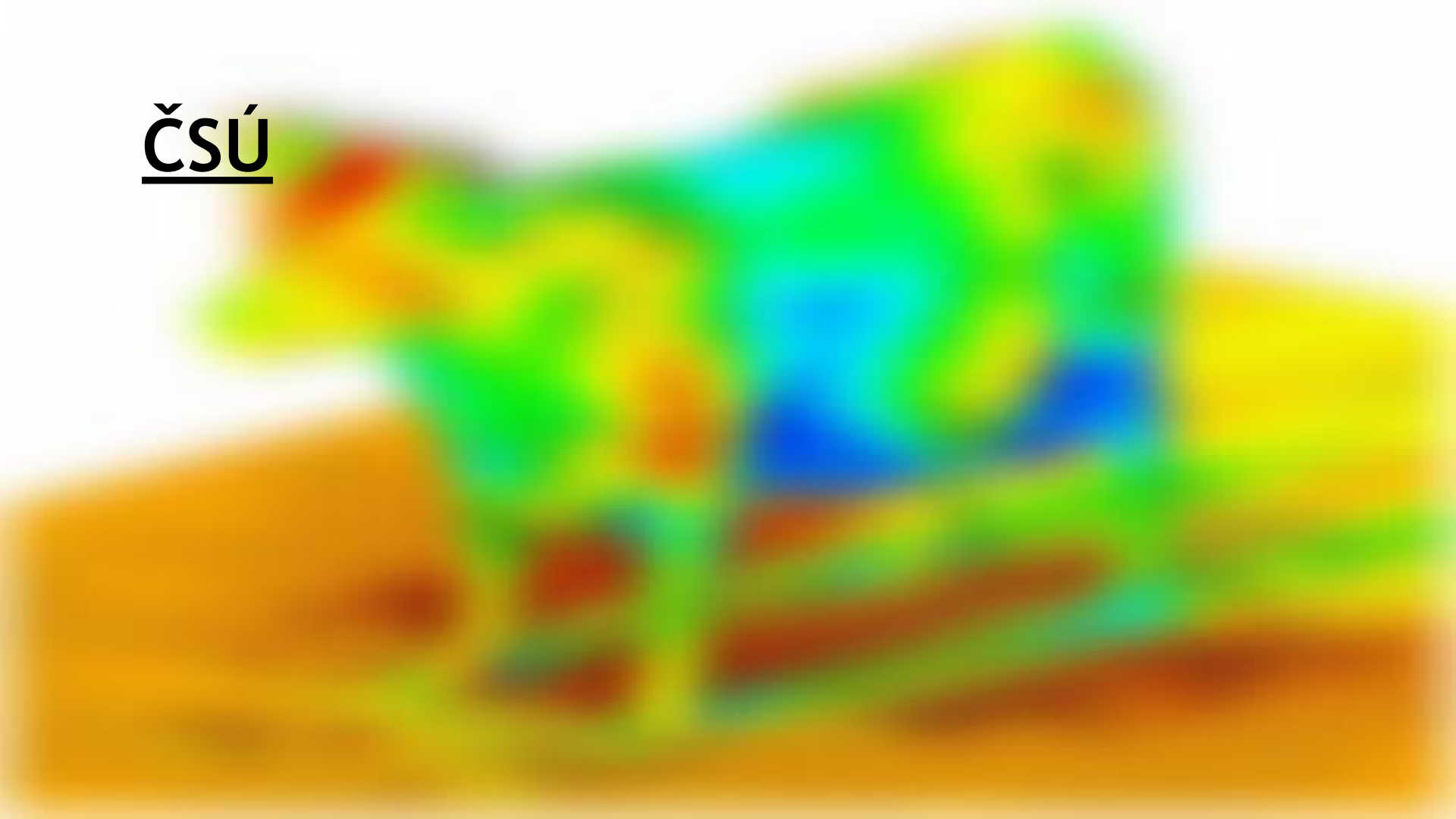
David McCandless & AlwaysWithHonor.com // v1.0 // Apr 09 // InformationIsBeautiful.net

source: Pantone, ColorMatters & web sources


List of men famous for being well endowed
"delete"
"keep"

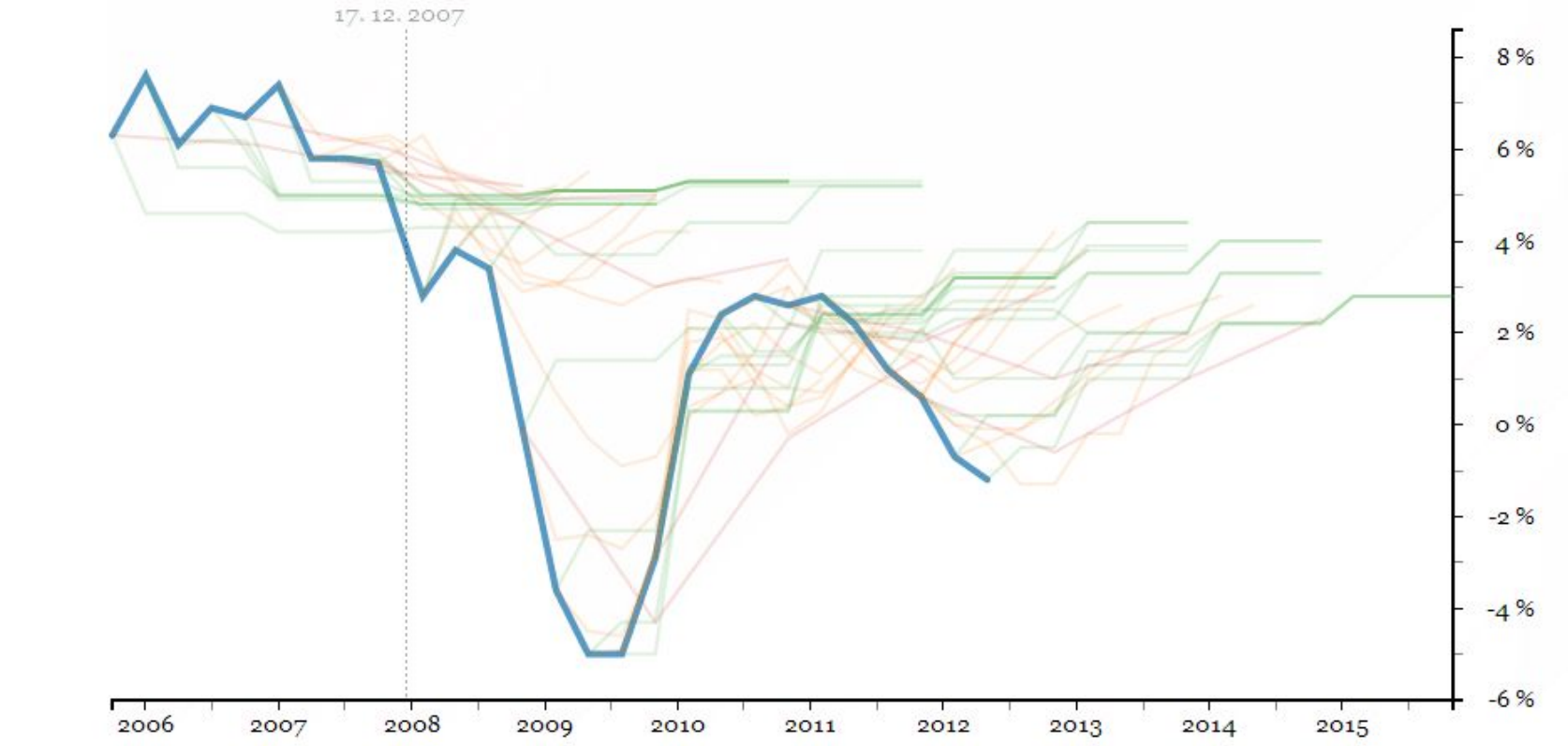

KVALITA
RUČNĚ
VERO
RYCHLOST

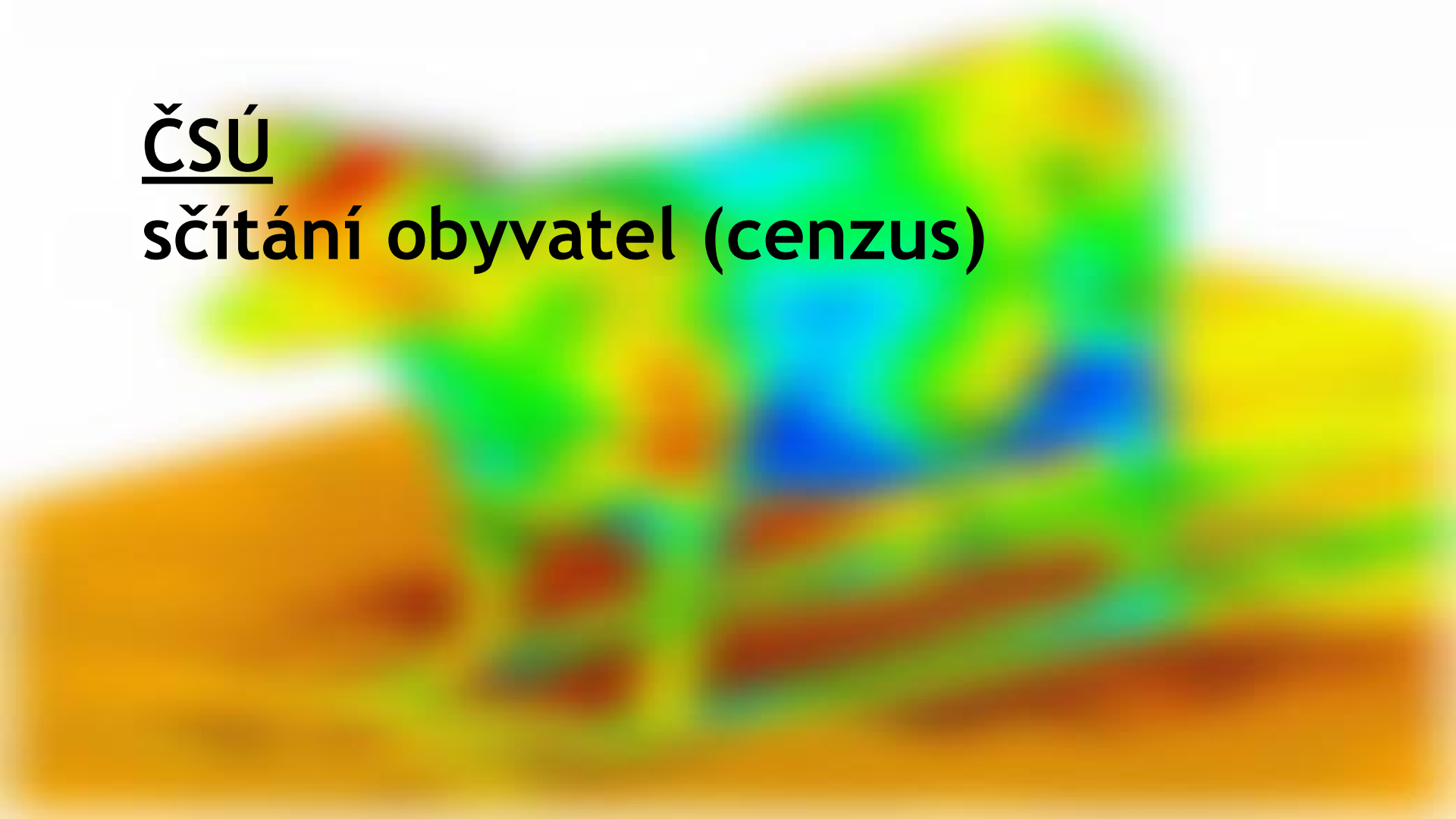
ČSÚ



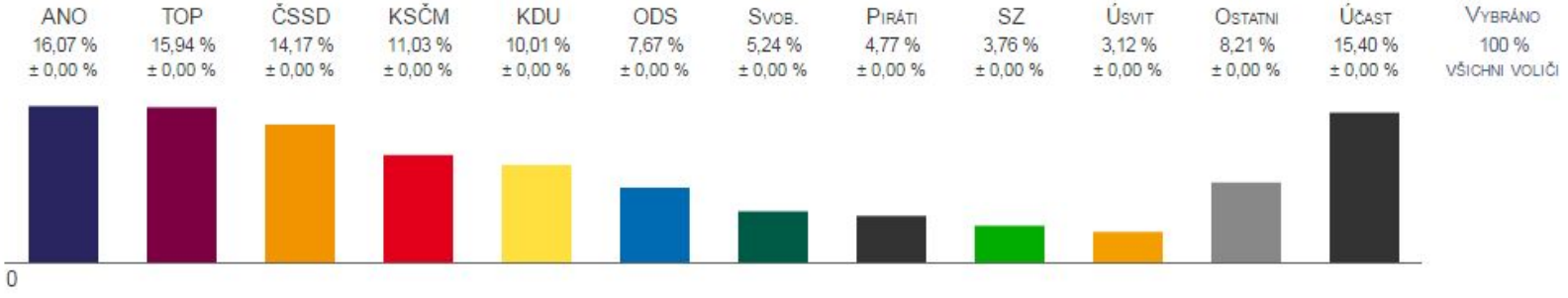
Instituce:

Prognóza ČNB Prognóza MF Prognóza MPSV



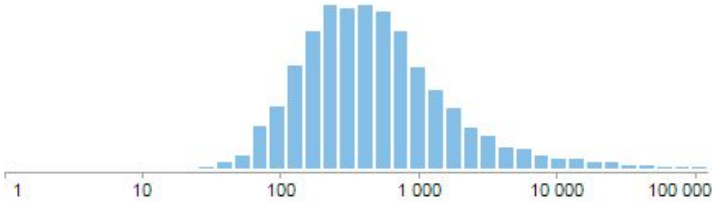


ČSÚ
sčítání obyvatel (cenzus)

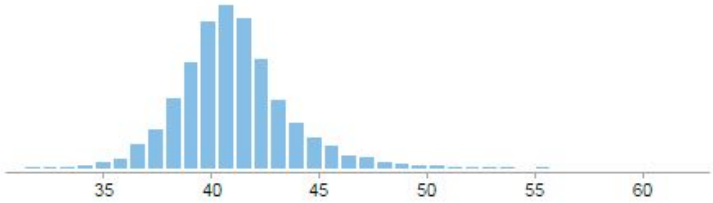


VELIKOST OBCE

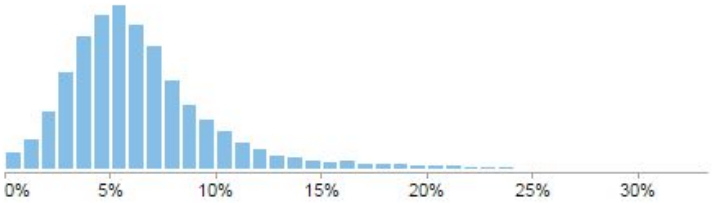
TAŽENÍM MYŠI VYBERETE ČÁST GRAFU, KTERÁ VÁS ZAJÍMÁ



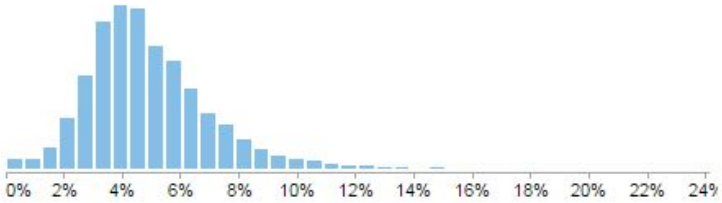
VĚKOVÝ PRŮMĚR



LIDÉ S VYSOKOU ŠKOLOU



PODÍL NEZAMĚSTNANÝCH



ČSÚ

sčítání obyvatel (cenzus)

ministerstva, kraje, obce, …

ČSÚ

sčítání obyvatel (cenzus)

ministerstva, kraje, obce, …

otevřená data: Praha

Stíny Jižního Města
2014-06-15 17:00:00

ČSÚ

sčítání obyvatel (cenzus)

ministerstva, kraje, obce, …

otevřená data: Praha

apps: VOR, VZ, VášMajetek, mapy.cz

ČSÚ

sčítání obyvatel (cenzus)

ministerstva, kraje, obce, …

otevřená data: Praha

apps: VOR, VZ, VášMajetek, mapy.cz

tematické databáze: GTD

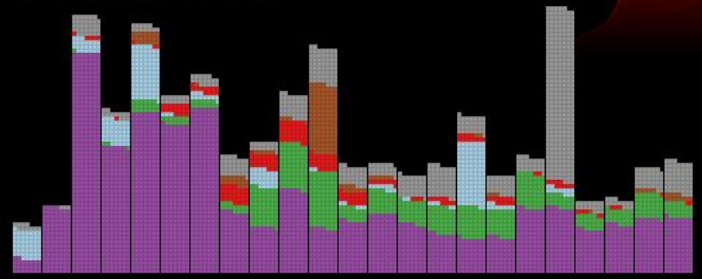
Česko se bojí terorismu. Evropa přitom prožívá relativně klidné období

Přesně před čtrnácti lety odstartoval útok na New York a Washington mezinárodní „válku proti terorismu“. Zmapovali jsme moderní historii teroristických útoků v Evropě a ukázalo se, že jejich obětí dlouhodobě ubývá.

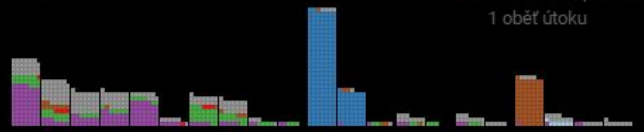
Češi přesto považují terorismus, obvykle ve spojení s imigrací, za největší aktuální hrozbu - ať už pro Českou republiku, nebo pro Evropskou unii. Ukazuje to výzkum české vlády z ledna i výzkum agentury STEM z května.

Graf ve spodní části obrazovky znázorňuje, kolik lidí teroristé v jednotlivých letech zabili – jeden čtvereček představuje jednu oběť. Vybarvené jsou podle toho, která teroristická skupina má útok na svědomí. Každý útok je také zakreslený v mapě.

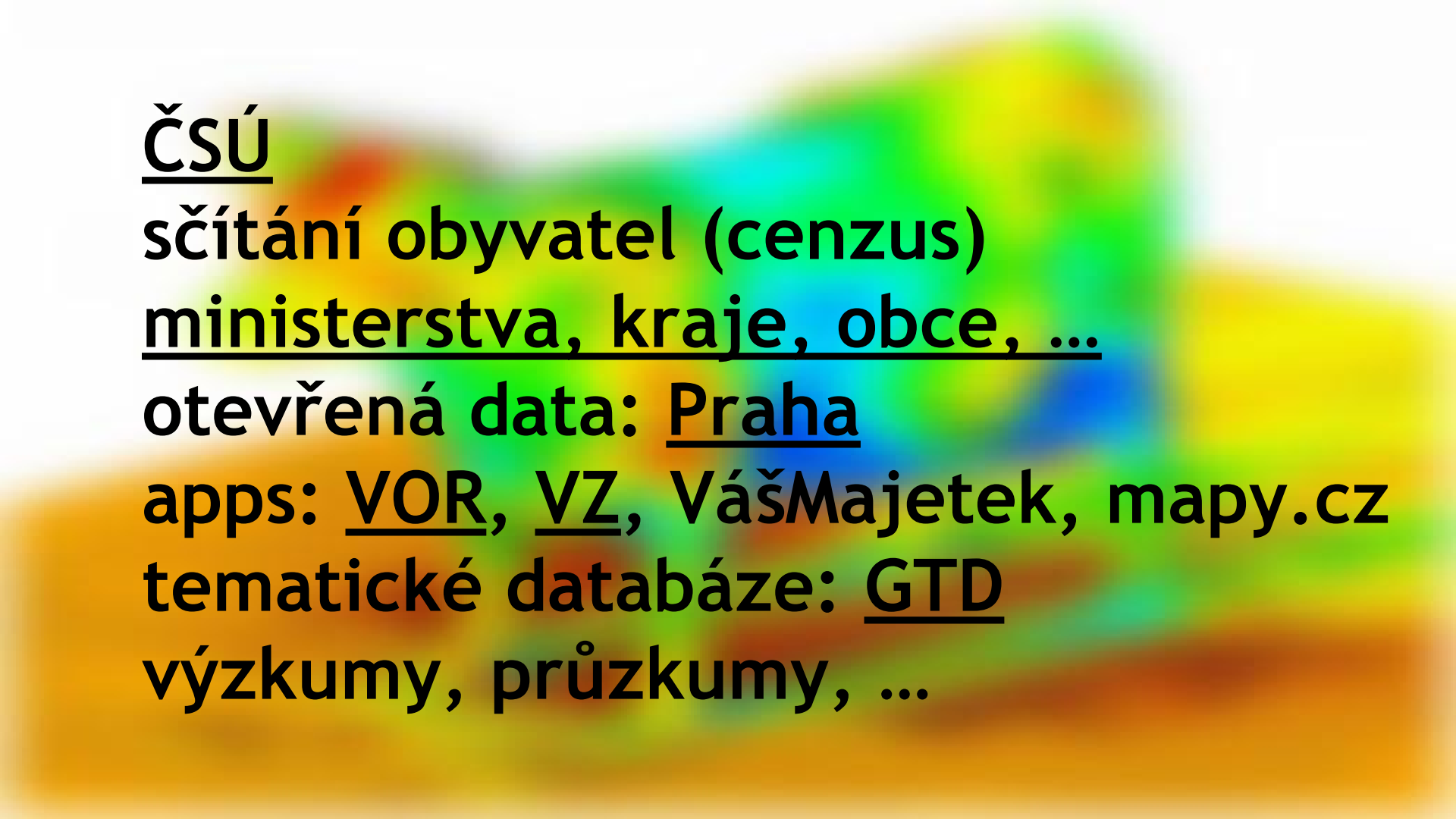
Číst dál ›

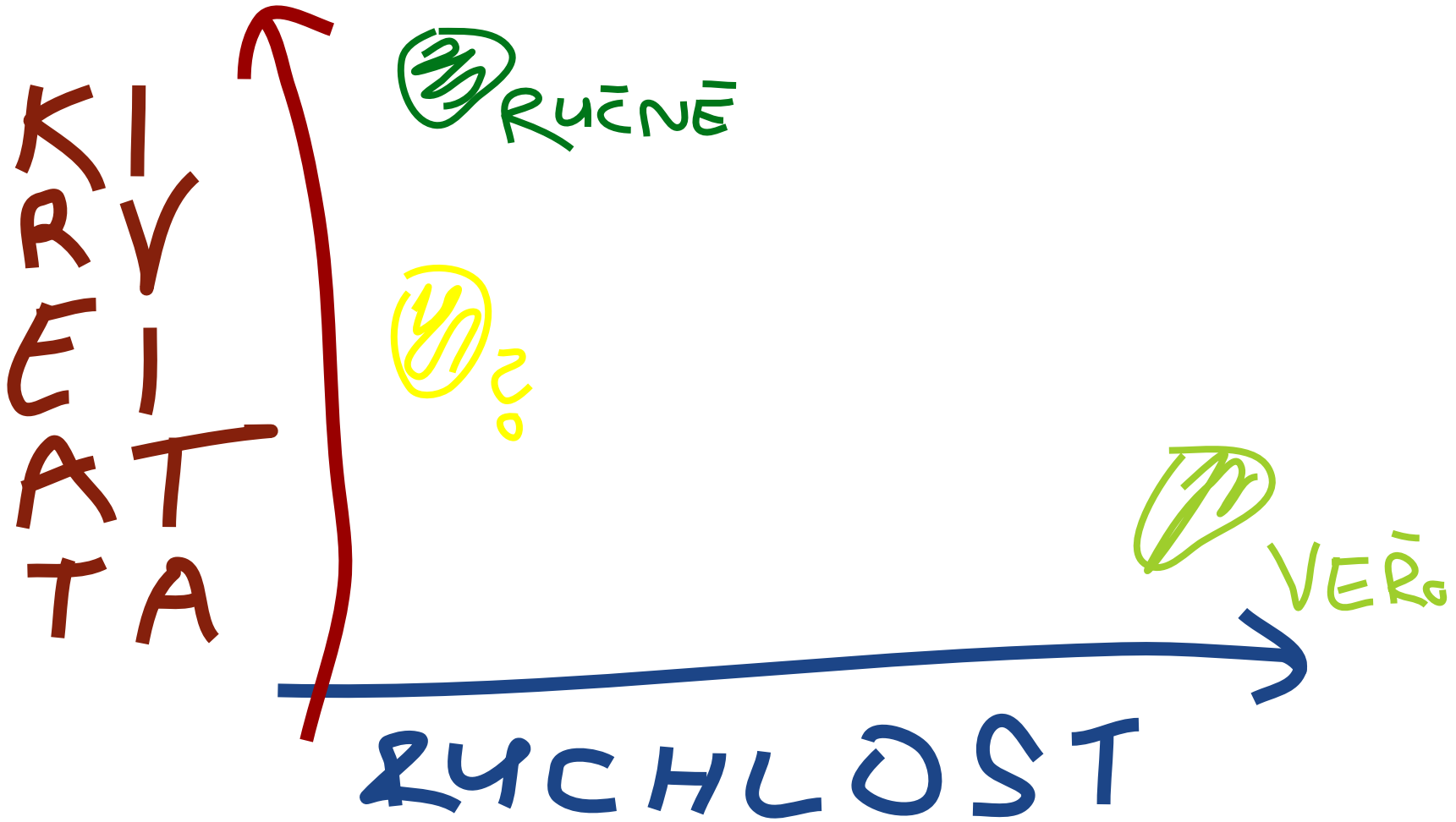


V datech bohužel chybí rok 1993, kdy ale na evropském území podle dostupných zdrojů nedošlo k žádnému útoku s velkým počtem obětí



1 čtvereček představuje 1 oběť útoku


ČSÚ
sčítání obyvatel (cenzus)
ministerstva, kraje, obce, …
otevřená data: Praha
apps: VOR, VZ, VášMajetek, mapy.cz
tematické databáze: GTD
výzkumy, průzkumy, …


KVALITA
RUČNĚ
VE
VERO
RYCHLOST

KDYŽ NEVÍTE, PTEJTE SE

kolik je agentů StB ve vedení policie?

KDYŽ NEVÍTE, PTEJTE SE

kolik je agentů StB ve vedení policie?

kdo létá na podnikatelské mise s prezidentem?

KDYŽ NEVÍTE, PTEJTE SE

kolik je agentů StB ve vedení policie?

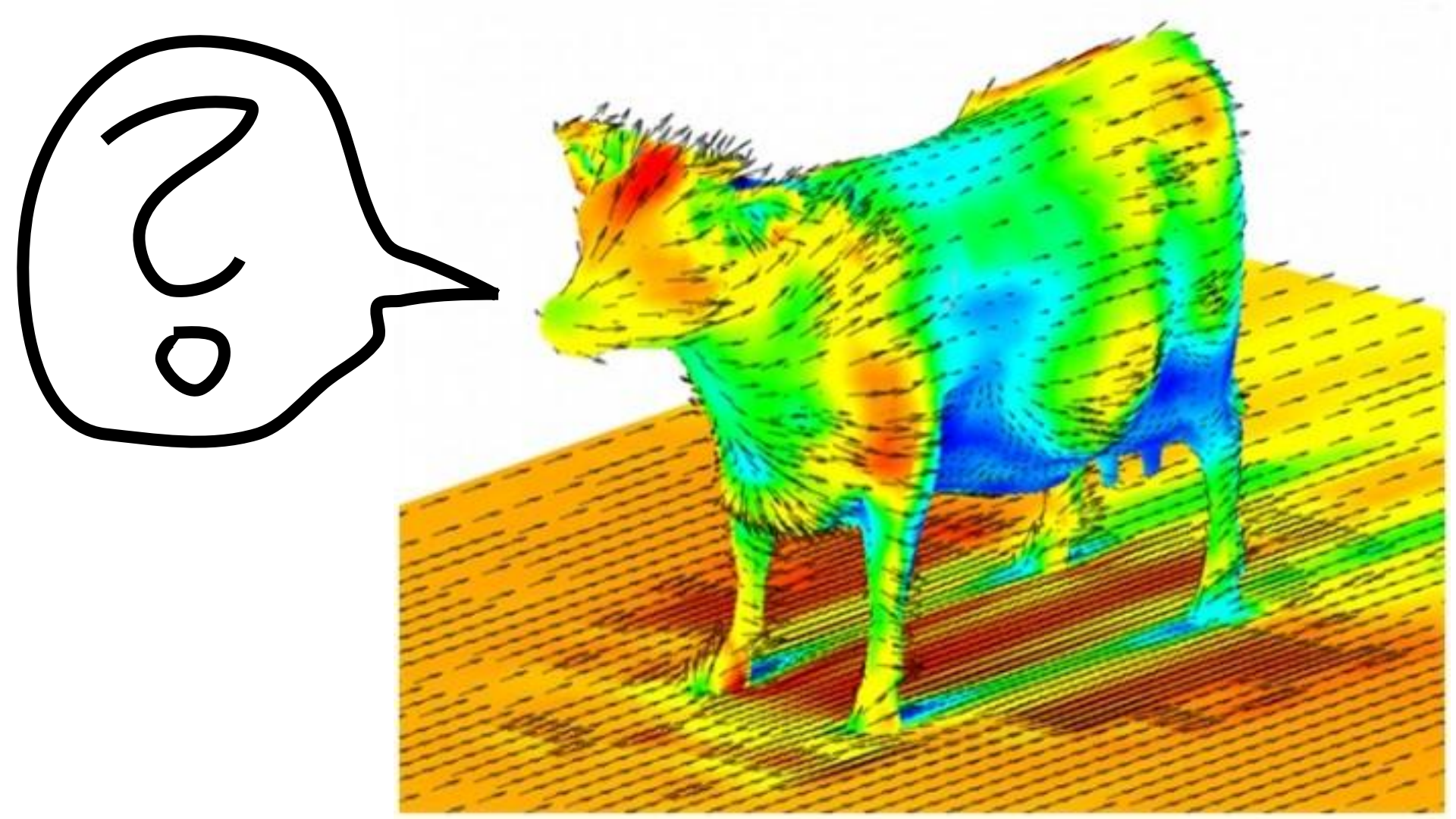
kdo létá na podnikatelské mise s prezidentem?

kde myslivci zastřelí nejvíc srnečků?

KDYŽ NEODPOVÍ, ZKUSTE ZÁKON 106

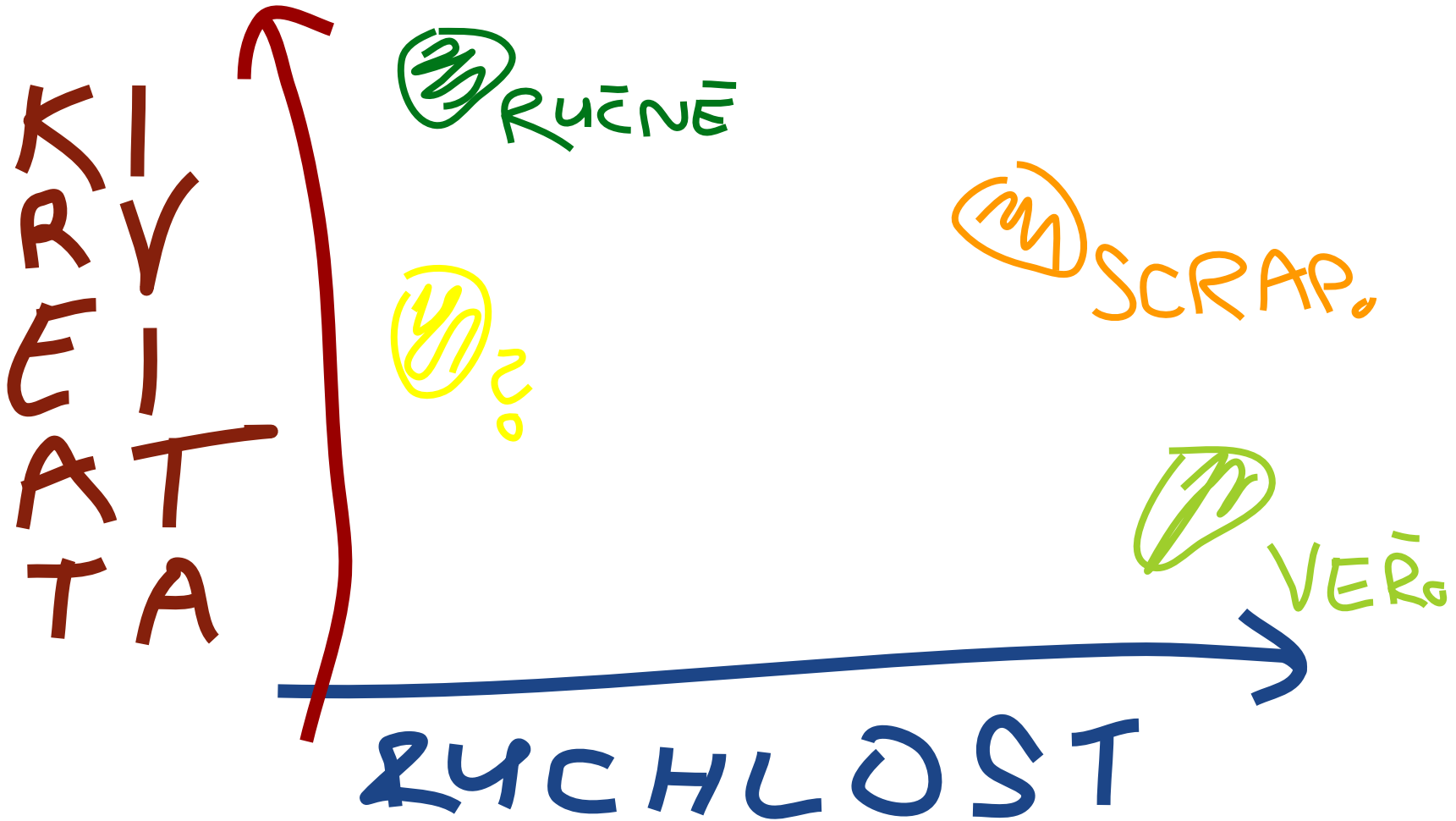
...o svobodném přístupu k informacím

JAK NA TO:
OTEVŘETE.CZ



zveřejnění platů manažerů státních firem:

povede ke zvýšení nebo snížení?


KVALITA
RUČNĚ
SCRAP.
VEŘ.
RYCHLOST

SCRAPOVÁNÍ

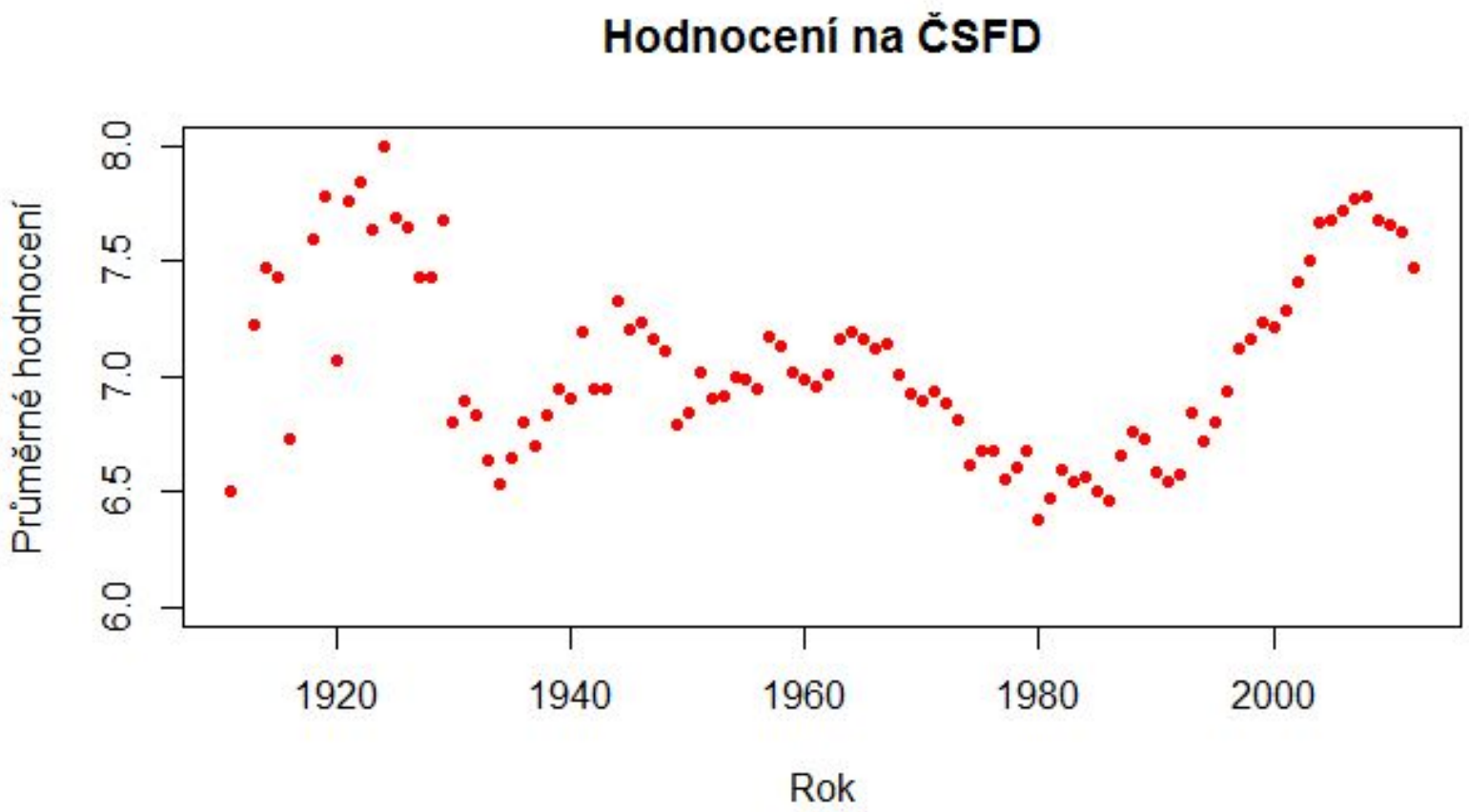
SCRAPOVÁNÍ

využití toho, že HTML je strukturované (tagy)

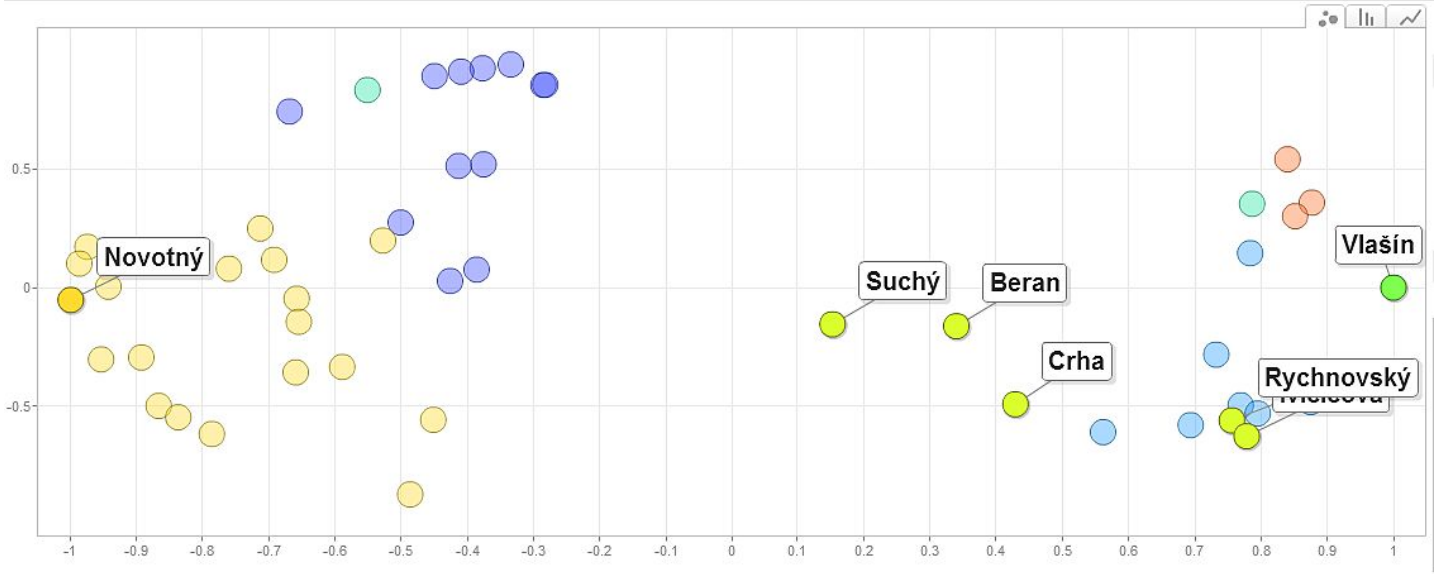
RO BO T!

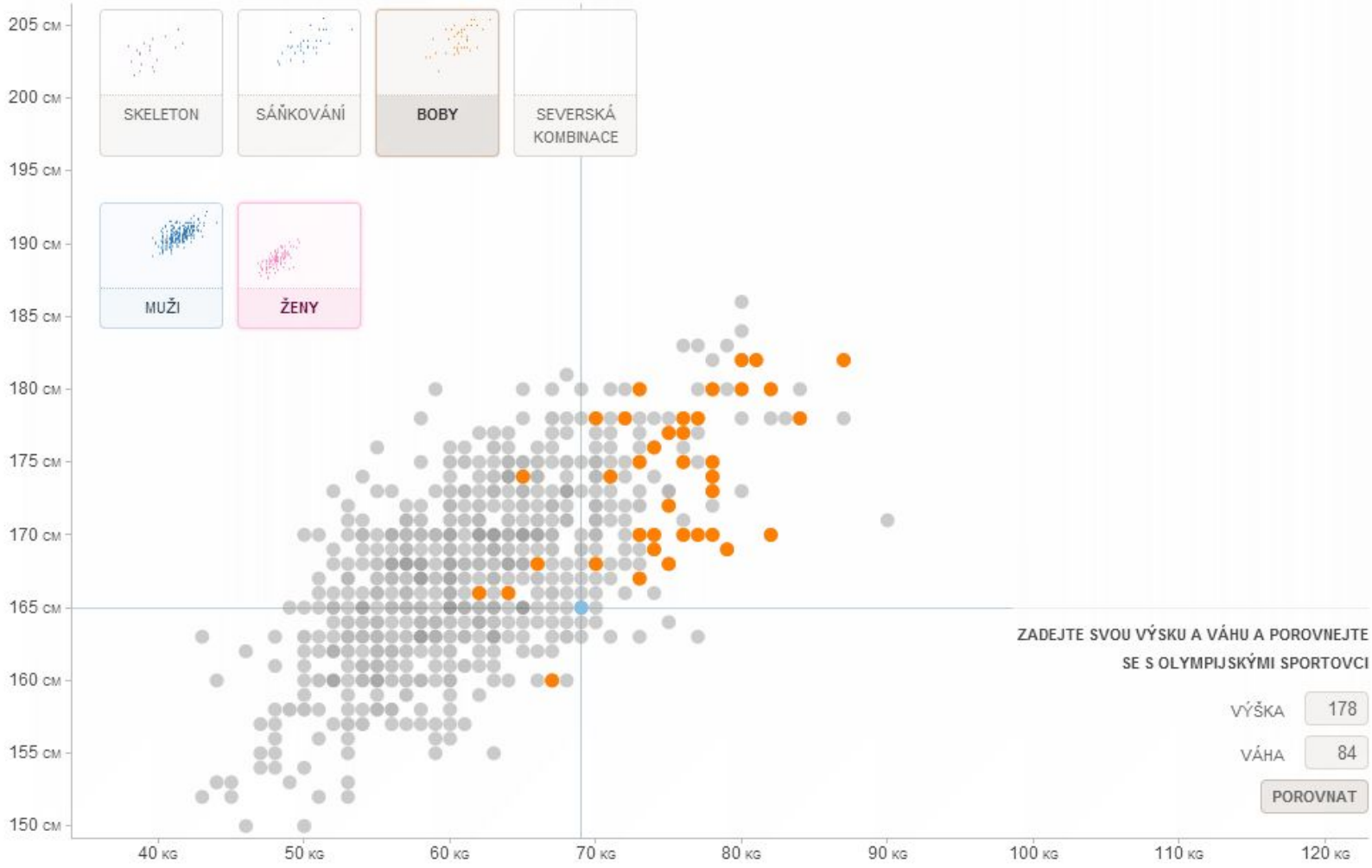
ROBOT!

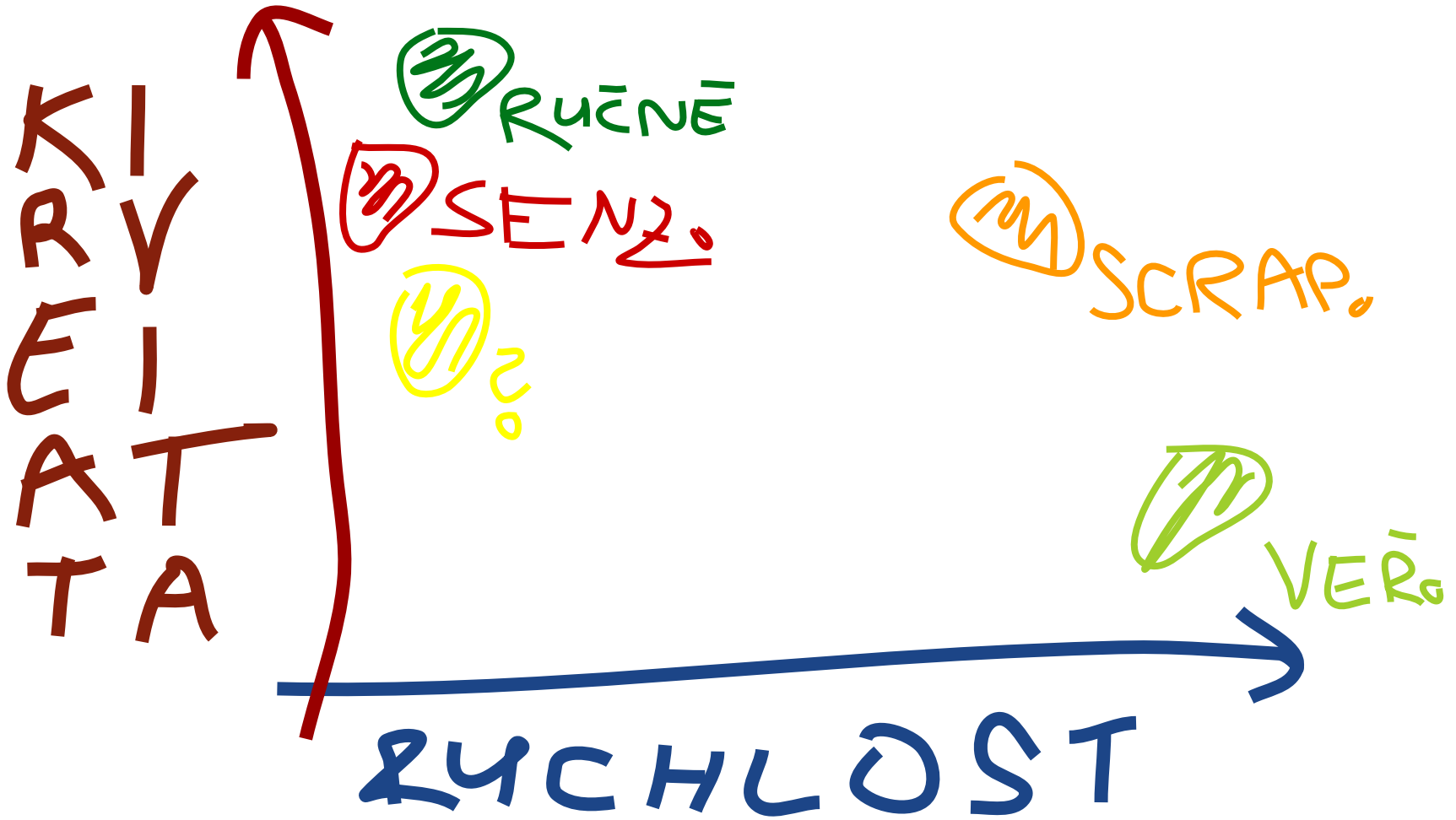
```r
library(RCurl)
library(XML)
library(stringr)

linky <- paste("http://www.sdruzeni-sova.cz/seznam-aut-s-podezrenim-na-zasah-do-stavu-tachometru?vir

df <- read.csv(text="datum-webu, VIN, znacka, model, datum-zaznamu, km-pred, zeme, datum-kontroly, k

for(i in 1:198) {

  web <- getURL(linky[i], .opts = list(ssl.verifypeer = FALSE), .encoding = "UTF-8")
  web <- readLines(tc <- textConnection(web)); close(tc)

  datum <- grep("Vozy nalezeny", web)
  start <- datum
  datum <- web[datum]
  datum <- gsub(".*dne ", "", datum)
  datum <- gsub("</h4>", "", datum)

  for(j in start) {           # pro každé datum

    startaut <- j+16

    while(grepl("<td>", web[startaut]))  {      # pro každé auto

      datumwebu <- web[j]
      datumwebu <- gsub(".*dne ", "", datumwebu)
      datumwebu <- gsub("</h4>", "", datumwebu)
      datumwebu <- as.character(datumwebu)

      VIN <- web[startaut]
      VIN <- gsub(".*<td>", "", VIN)
      VIN <- gsub("</td>", "", VIN)
      VIN <- as.character(VIN)

      znacka <- web[startaut+1]
      znacka <- gsub(".*<td>", "", znacka)
      znacka <- gsub("</td>", "", znacka)
      znacka <- as.character(znacka)

      model <- web[startaut+2]
      model <- gsub(".*<td>", "", model)
      model <- gsub("</td>", "", model)
      model <- as.character(model)

```

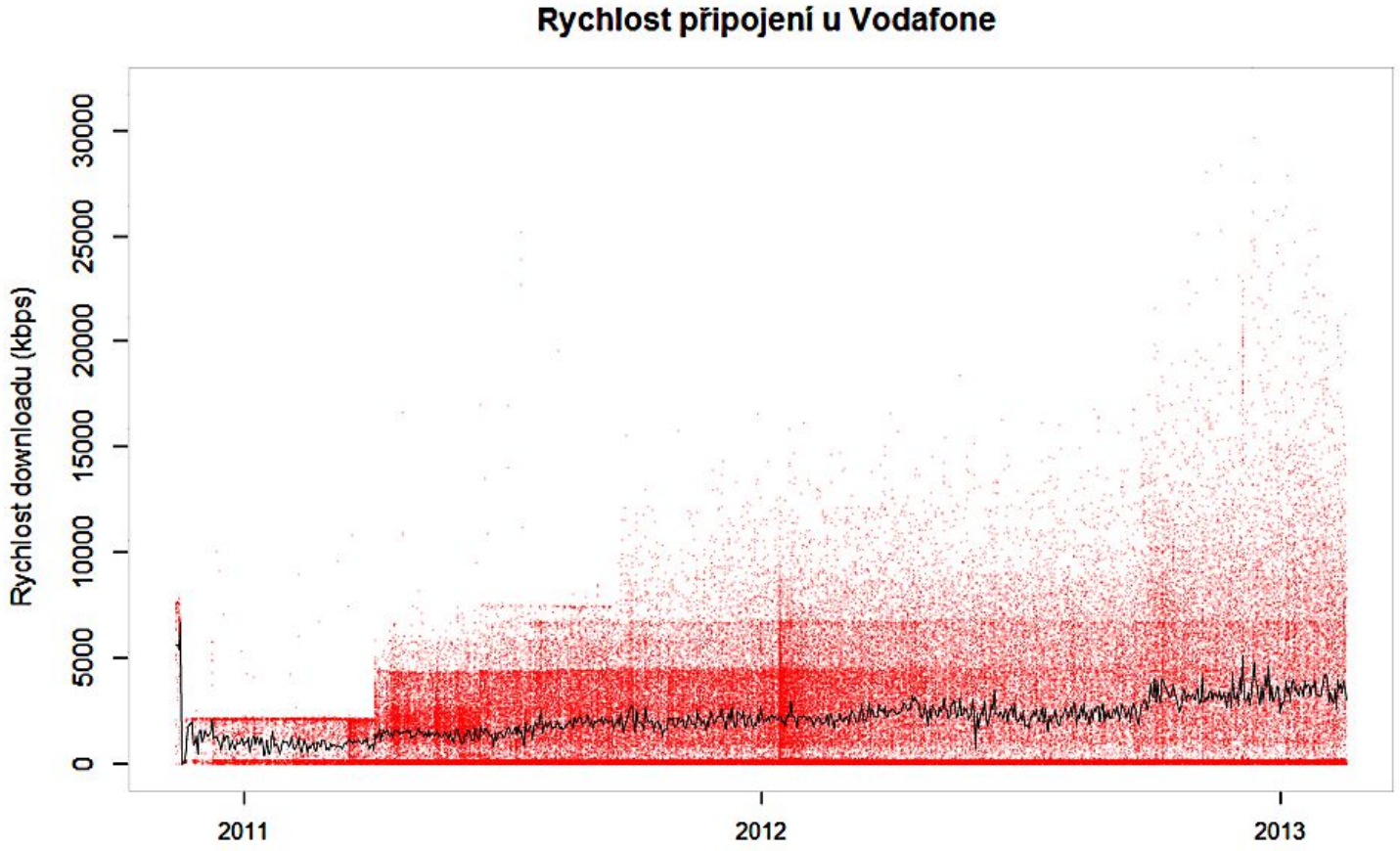

Hodnocení na ČSFD
Průměrné hodnocení
6.0
6.5
7.0
7.5
8.0
1920
1940
1960
1980
2000
Rok

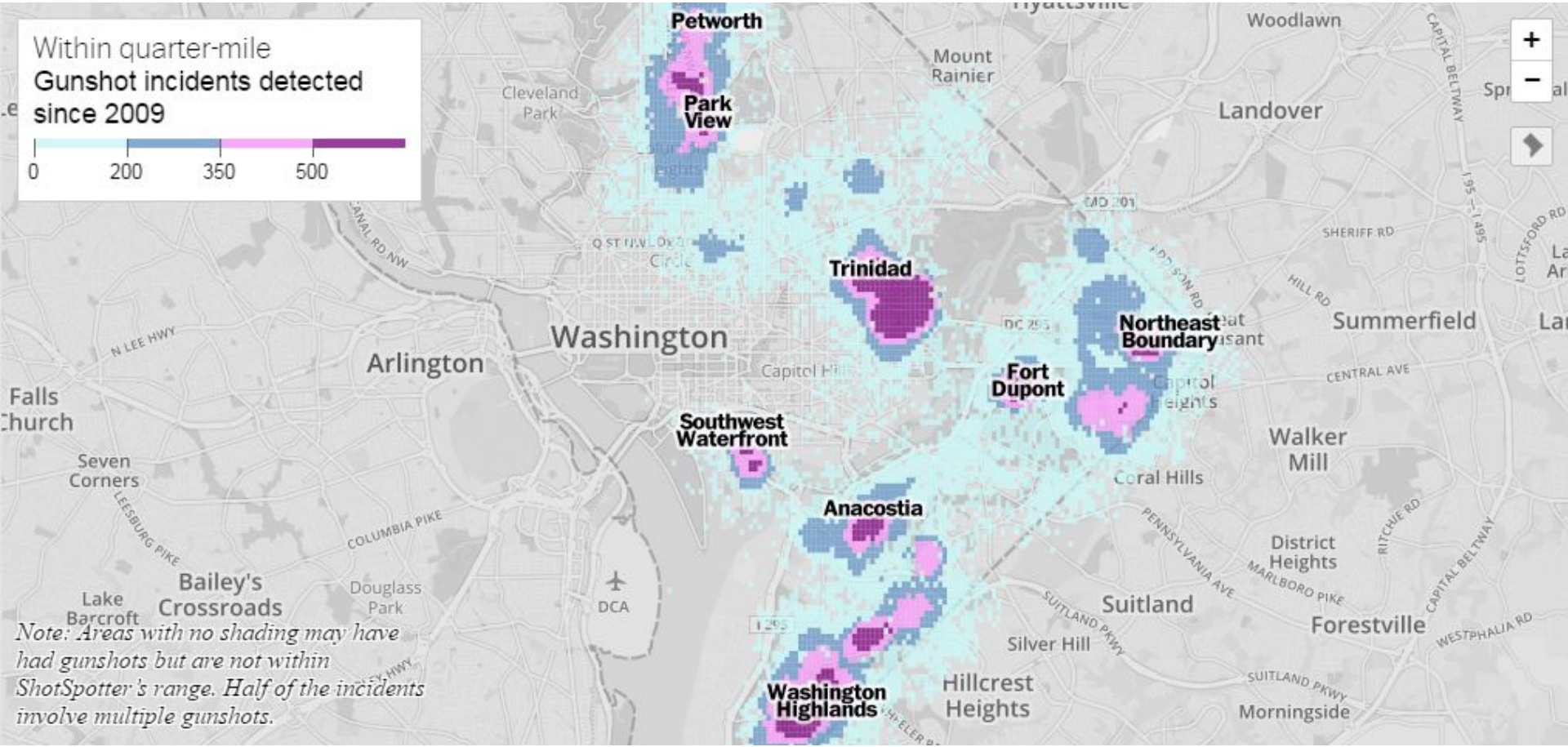
Zastupitelstvo 11/2012 až 01/2013

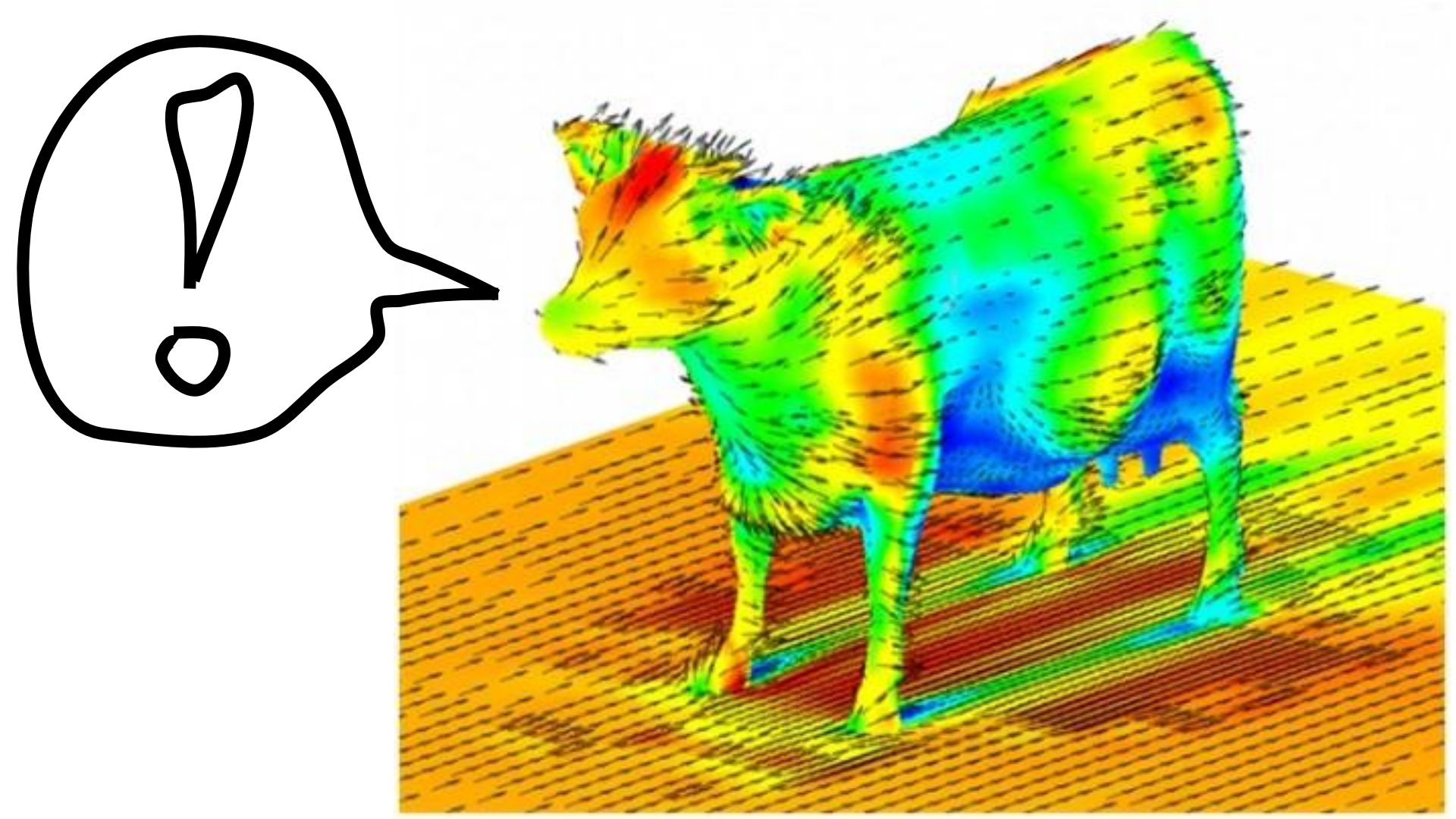



SKELETON
SÁŇKOVÁNÍ
BOBY
SEVERSKÁ KOMBINACE
MUŽI
ŽENY
205 CM
200 CM
195 CM
190 CM
185 CM
180 CM
175 CM
170 CM
165 CM
160 CM
155 CM
150 CM
40 KG
50 KG
60 KG
70 KG
80 KG
90 KG
100 KG
110 KG
120 KG
ZADEJTE SVOU VÝSKU A VÁHU A POROVNEJTE SE S OLYMPIJSKÝMI SPORTOVCI
VÝŠKA
178
VÁHA
84
POROVNAT


KVALITA
RUČNĚ
SENZ.
?
SCRAP.
VĚR.
RYCHLOST


Rychlost připojení u Vodafone
Rychlost downloadu (kbps)
0
5000
10000
15000
20000
25000
30000
2011
2012
2013


Within quarter-mile
Gunshot incidents detected since 2009
0
200
350
500
Petworth
Park View
Trinidad
Northeast Boundary
Fort Dupont
Southwest Waterfront
Anacostia
Washington Highlands
Washington
Arlington
Falls Church
Seven Corners
Bailey's Crossroads
Lake Barcroft
Douglass Park
DCA
Mount Rainier
Cleveland Park
Landover
Woodlawn
Summerfield
Walker Mill
Coral Hills
District Heights
Suitland
Forestville
Silver Hill
Hillcrest Heights
Morningside
Capitol Heights
Note: Areas with no shading may have had gunshots but are not within ShotSpotter's range. Half of the incidents involve multiple gunshots.



DÚ

čtyři témata, čtyři tabulky

postup:

1. stáhněte si ze studijních materiálů *(přednáška 7, data k DÚ)* a prozkoumejte jednu z tabulek
2. vymyslete a poznamenejte si otázky, které vás k tématu zajímají a v datech na ně je odpověď
3. na svoje otázky najděte a sepište odpovědi
4. objevte kontext: proč to tak je

to celé na 2-3 normostrany (1800 znaků)

do 24.11. (čtvrtek, zadávám to později, tak je později i odevzdání)

data jsou skutečná, v podstatě tak, jak jsem je dostal od zdroje, jen jsem je lépe pojmenoval a zformátoval

dělejte s nimi cokoliv, co vám pomůže odpovědět na otázky (pokud něco neumíte, vygůglete si to)

pokud k interpretaci potřebujete další data, dohledejte si je

jakékoliv problémy: honza.bocek@gmail.com

TÉMA 1: PŘESTUPKY V BRNĚ

- všechny přestupky, které eviduje městská policie v Brně od června 2013 do poloviny září 2014
- typ přestupku je v původním znění, pokud s ním chcete pracovat, zkuste www.zakonyprolidi.cz
- nechal jsem tam gps souřadnice, pokud byste je někdo chtěli a uměli dát na mapu; není to nutné

TÉMA 2: STARTOVAČE

- úspěšné i neúspěšné ukočené projekty dvou největších českých kopií Kickstarteru od spuštění doteď
- zatím jsou neveřejné, nechte si je pro sebe
- v každé záložce jeden startovač, každý má data v jiné struktuře
- vše, včetně lokality, zadávají autoři projektů
- pokud je u Hithitu projekt ve více kategoriích, každá má svůj řádek; poraďte si s tím!

TÉMA 3: ÚMRTNOST V EU

- úmrtnost podle diagnózy v evropských regionech
- standardizovaná (na 100 000 obyv.), průměr za roky 2011 až 2013
- regiony odpovídají evropským státům (NUTS1) a krajům (NUTS2)
- dělení podle pohlaví a podle věku

TÉMA 4: TERORISMUS

- data z webu Global Terrorism Database
- původní, bez jakýchkoliv změn, od roku 1970 do 2014
- spíš pro drsňáky ( )