

Viri podatkov

3. srečanje: Vizualizacija podatkov

Mojca Bavdaž (mojca.bavdaz@ef.uni-lj.si)

Petek, 6.3.2020



Prvi stolpčni grafikon in linijski grafikon za časovno vrsto

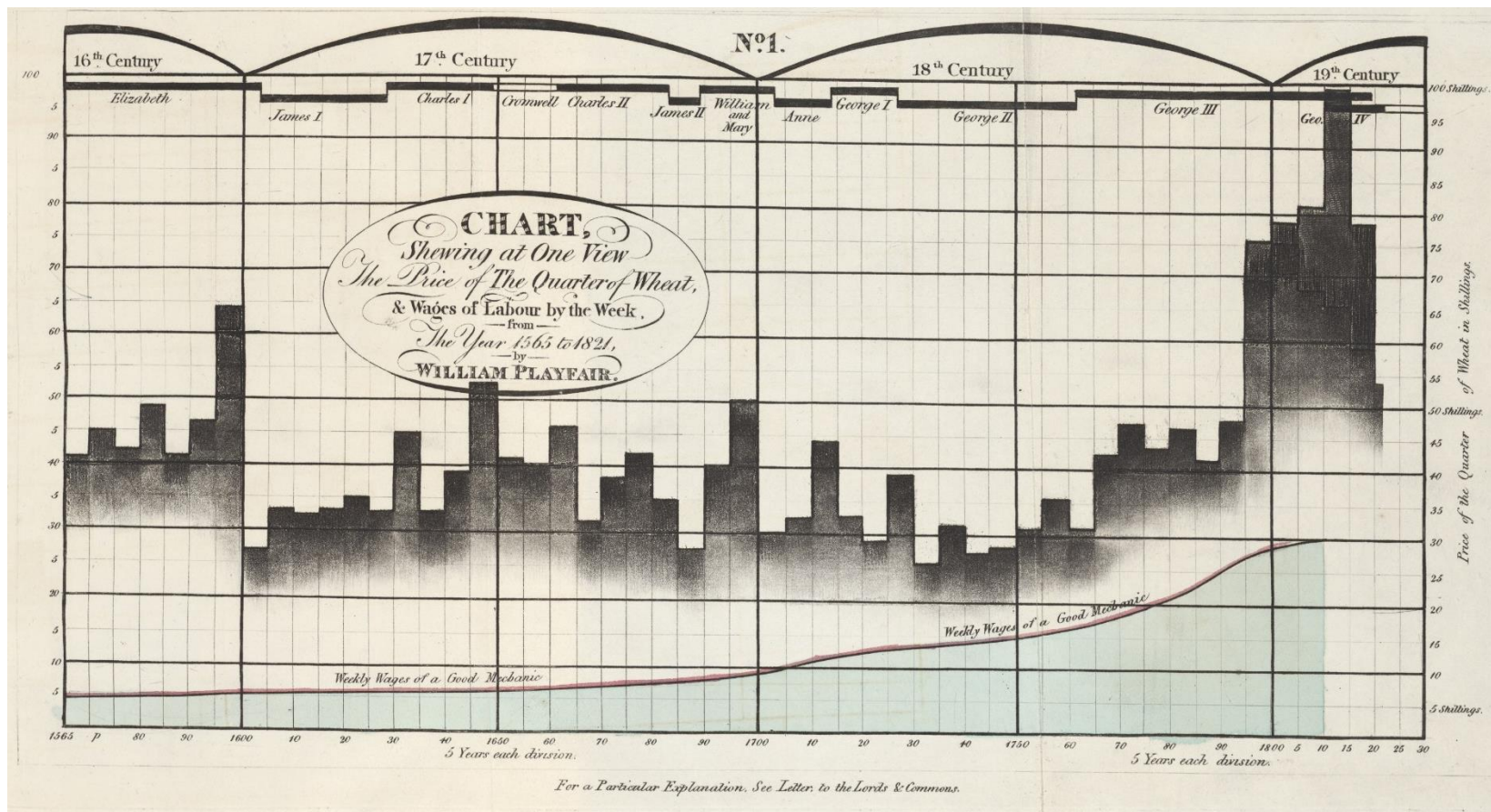
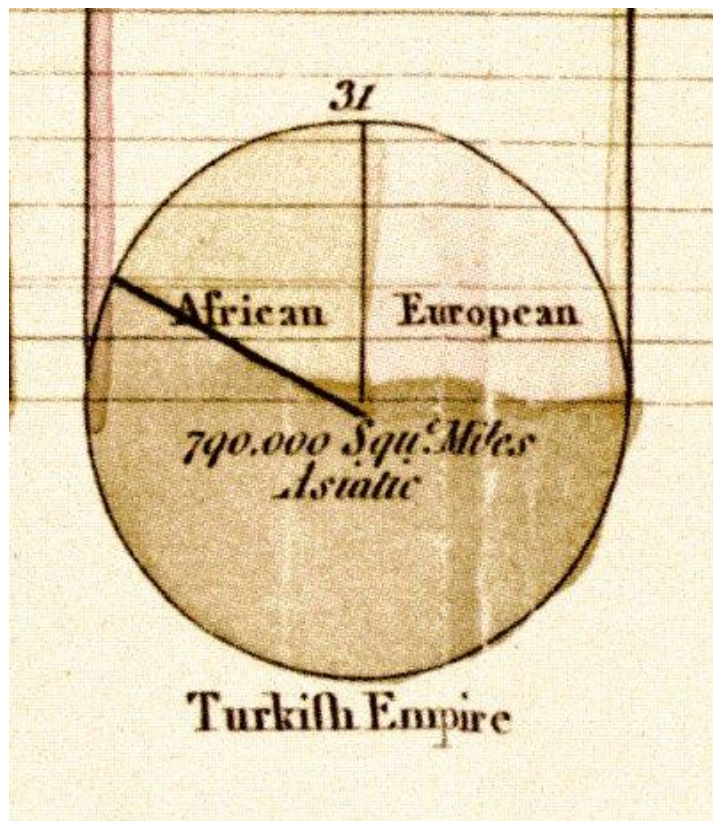


Chart shewing at one view the price of the quarter of wheat, & wages of labour by the week, from the year 1565 to 1821 by William Playfair (1822). (The second edition / with an additional chart.). Retrieved from <http://brbl-dl.library.yale.edu/vufind/Record/3566707>

Prvi strukturni krog



The pie chart Turkish Empire by William Playfair (1801). In *Statistical Breviary*. Retrieved from <https://commons.wikimedia.org/wiki/File:Playfair-piechart.jpg>

Poveži opis z ustreznim grafikonom

- | | |
|---|----------------------|
| 1. Prikazati morate velikost izvoza (v evrih) iz Slovenije v štiri največje trgovinske partnerice. | A. Linijski grafikon |
| 2. Prikazati morate število zaposlenih po plačnih razredih. | B. Stolpčni grafikon |
| 3. Prikazati morate letne stopnje natalitete za Slovenijo za zadnjih 10 let. | C. Tabela |
| 4. Prikazati morate strukturo slovenske odrasle populacije po štirih zaposlitvenih kategorijah (študenti, zaposleni, brezposelni, upokojeni). | D. Strukturni krog |
| 5. Prikazati morate velikost prebivalstva za vse države EU. | E. Histogram |

Mejniki v vizualizaciji

- Milestones in the History of Thematic Cartography, Statistical Graphics, and Data Visualization.

<http://www.datavis.ca/milestones>

- Hans Rosling: The best stats you've ever seen.
Ted Talks, 2006

http://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen

Hans Rosling (1948–2017)

So how do we let Hans Rosling rest in peace?

By forgetting that he sometimes swallowed swords in a heavy metal style tank top.

And by remembering that mothers in Bangladesh no longer give birth to five children on average, nor four, but TWO POINT TWO children.

How do we let Hans Rosling rest in peace?

By forgetting that he got more clicks than Lady Gaga online.

And by remembering that 80 percent of the children of the world now have access to the most important and most cost efficient of all vaccines: the one for measles.

How do we let Hans Rosling rest in peace?

By forgetting that Time Magazine put him on some list.

And by remembering that Hans Rosling was certain that the world, if it got it's act together, can reach the goal that the United Nations set for the year 2030: to exterminate extreme poverty for *everyone, everywhere*.

Izpeljanke

Scientific Proof that Americans are Truck Addicts

<http://www.bloomberg.com/graphics/2015-auto-sales/>

A Quick Guide to the Worst Corporate Hack Attacks

<http://www.bloomberg.com/graphics/2014-data-breaches/>

Brexit Could Disrupt Millions of Expats' Lives

<https://www.bloomberg.com/graphics/2017-brexit-lives>

Investors Are Piling Into Loans That Banks Have Avoided Since the Crash

<https://www.bloomberg.com/graphics/2018-private-credit-yields/>

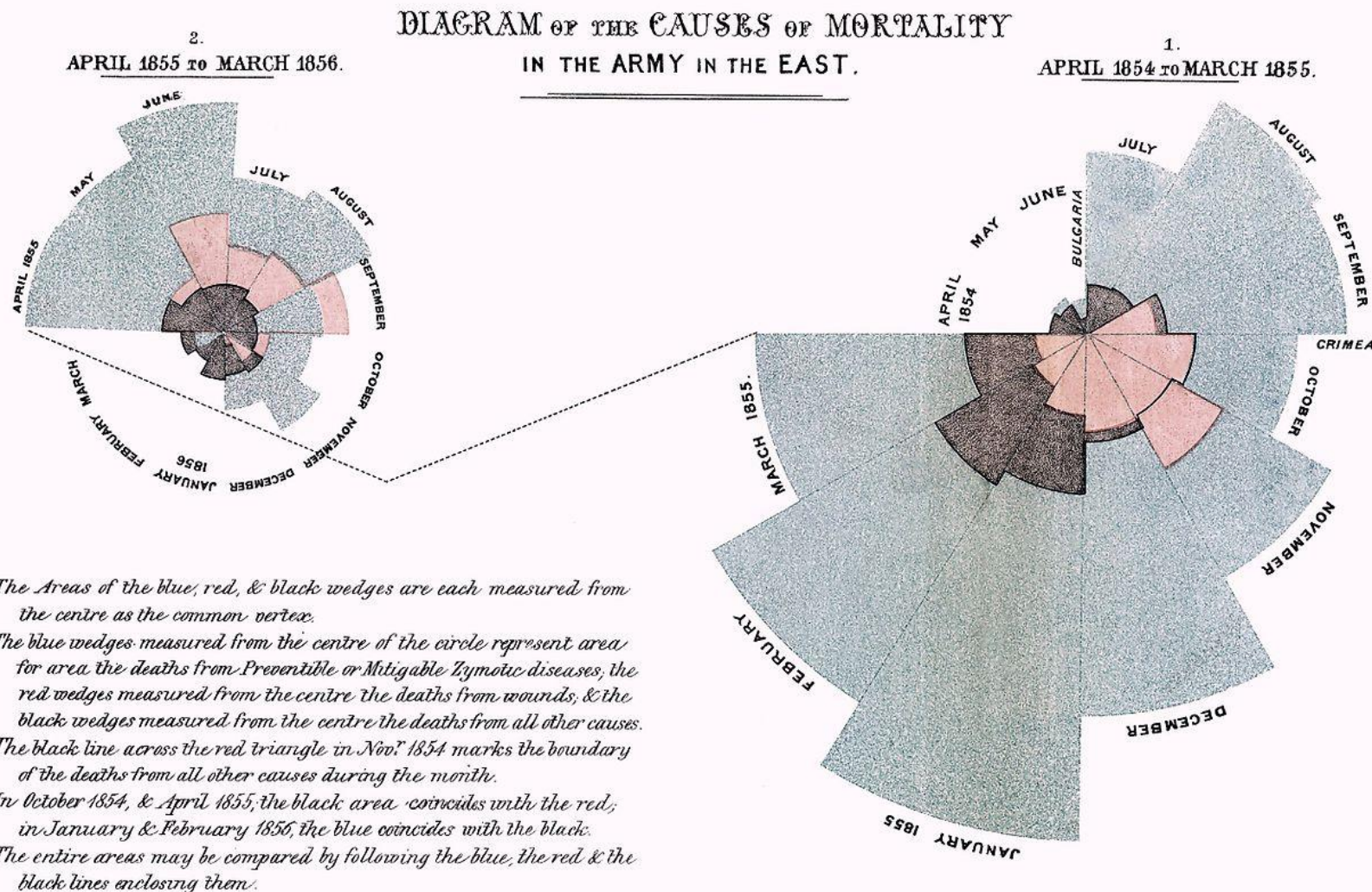
Zakaj je vizualizacija tako aktualna?

- Skoraj 50% možganov je vključenih v vizualno procesiranje
- 70% vseh čutilnih receptorjev je v očeh
- Vizualni prizor dojamemo v desetinki sekunde
- Pomnjenje je (bistveno) boljše, če je prenos informacij podprt z vizualnimi pripomočki

Pomen vizualizacije

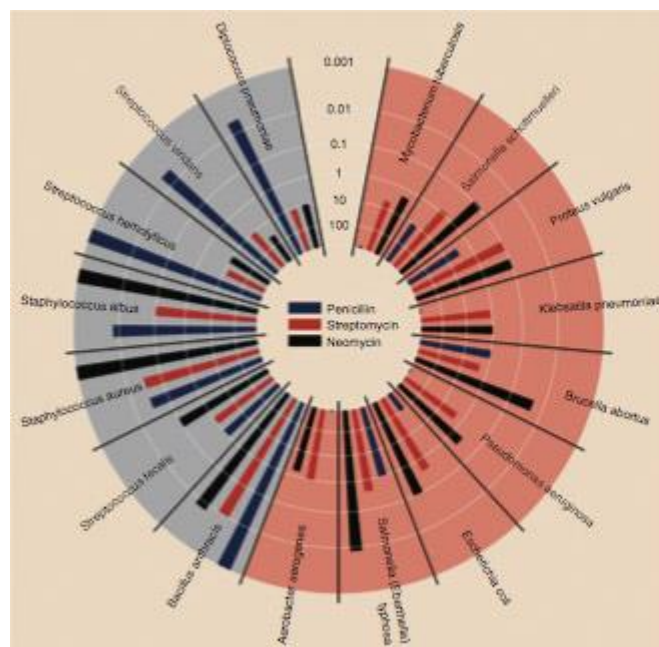
- Učinkovite predstavitve so danes samoumevne (čeprav še vedno relativno redke!), obstajajo pa že nekaj časa:
 - The Joy Of Stats: The Lady With A Data Visualisation.
 - Florence Nightingale
<http://www.open.ac.uk/openlearn/body-mind/health/health-sciences/the-joy-stats-the-lady-data-visualisation>
- John Tukey (1915–2000): “the greatest value of a picture is when it forces us to notice what we never expected to see.”

Polar area diagram: Causes of mortality in Crimean War



Pomen vprašanja, na katerega vizualizacija odgovarja

Will Burtin (1951)



Bacteria	Penicillin	Antibiotic Streptomycin	Neomycin	Gram stain
<i>Aerobacter aerogenes</i>	870	1	1.6	–
<i>Brucella abortus</i>	1	2	0.02	–
<i>Bacillus anthracis</i>	0.001	0.01	0.007	+
<i>Diplococcus pneumoniae</i>	0.005	11	10	+
<i>Escherichia coli</i>	100	0.4	0.1	–
<i>Klebsiella pneumoniae</i>	850	1.2	1	–
<i>Mycobacterium tuberculosis</i>	800	5	2	–
<i>Proteus vulgaris</i>	3	0.1	0.1	–
<i>Pseudomonas aeruginosa</i>	850	2	0.4	–
<i>Salmonella (Eberthella) typhosa</i>	1	0.4	0.008	–
<i>Salmonella schottmuelleri</i>	10	0.8	0.09	–
<i>Staphylococcus albus</i>	0.007	0.1	0.001	+
<i>Staphylococcus aureus</i>	0.03	0.03	0.001	+
<i>Streptococcus fecalis</i>	1	1	0.1	+
<i>Streptococcus hemolyticus</i>	0.001	14	10	+
<i>Streptococcus viridans</i>	0.005	10	40	+

Wainer, H, & Lysen, S. (2009). That's Funny... A window on data can be a window on discovery, *American Scientist*, 97(4), p. 272, DOI: 10.1511/2009.79.272

Revija *Chance* Ameriškega statističnega društva je ob stoti obletnici rojstva leta 2008 pripravila razpis za izboljšanje originalne Burtinove predstavitve.

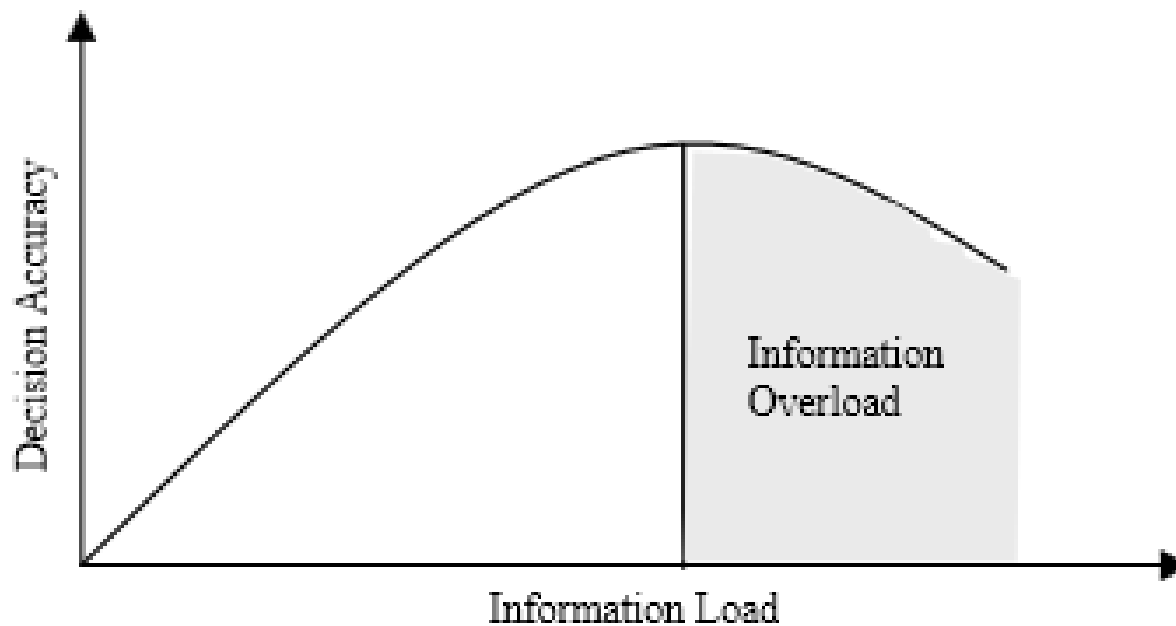
Nekaj primerov prijav:

- <http://peltiertech.com/antibiotic-effectiveness-a-study-of-chart-types/>
- <http://www.justintalbot.com/2009/01/my-submission-to-the-chance-contest/>

Razpis, o Burtinu in zmagovalci razpisa:

- <http://chance.amstat.org/> (4. št. 2008, 1. in 2. št. 2009)

(Pre)velika količina informacij

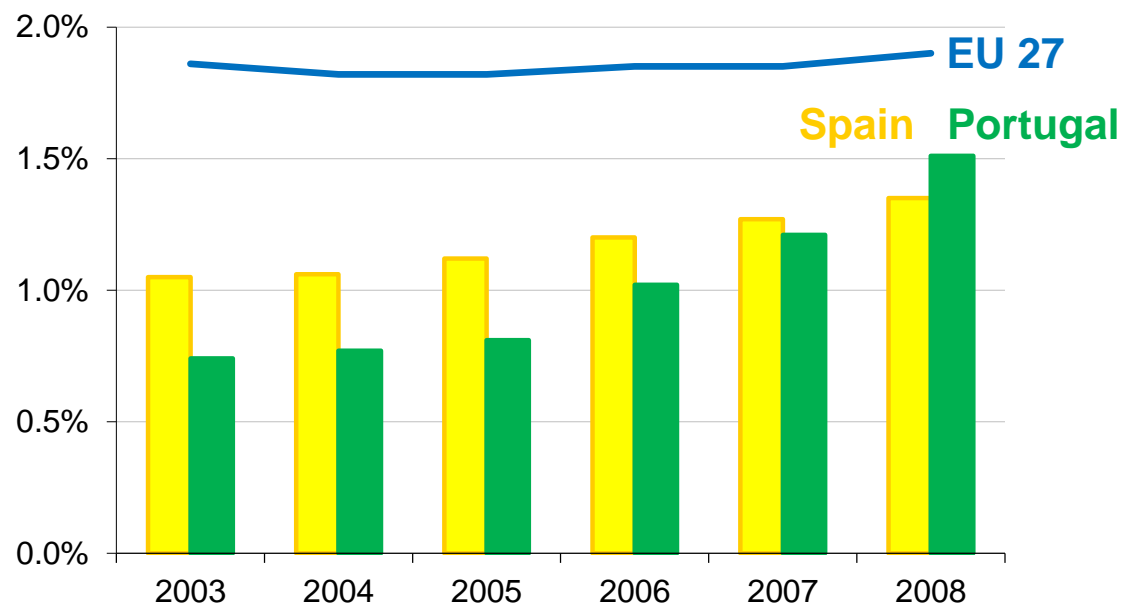
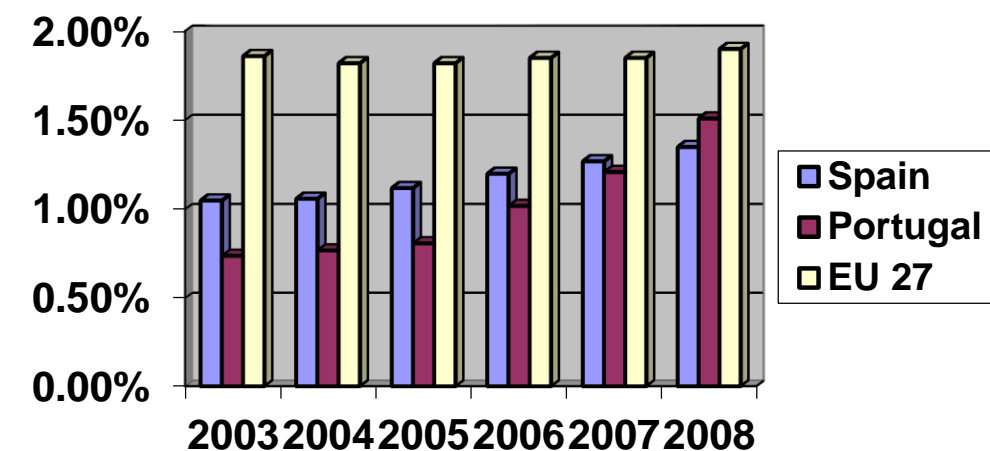


Tuftejeva načela za grafično odličnost

1. Above all else show the data.
2. Maximize the data-ink ratio.
3. Erase non-data ink.
4. Erase redundant data ink.
5. Revise and edit.



Primerjava levega in desnega prikaza

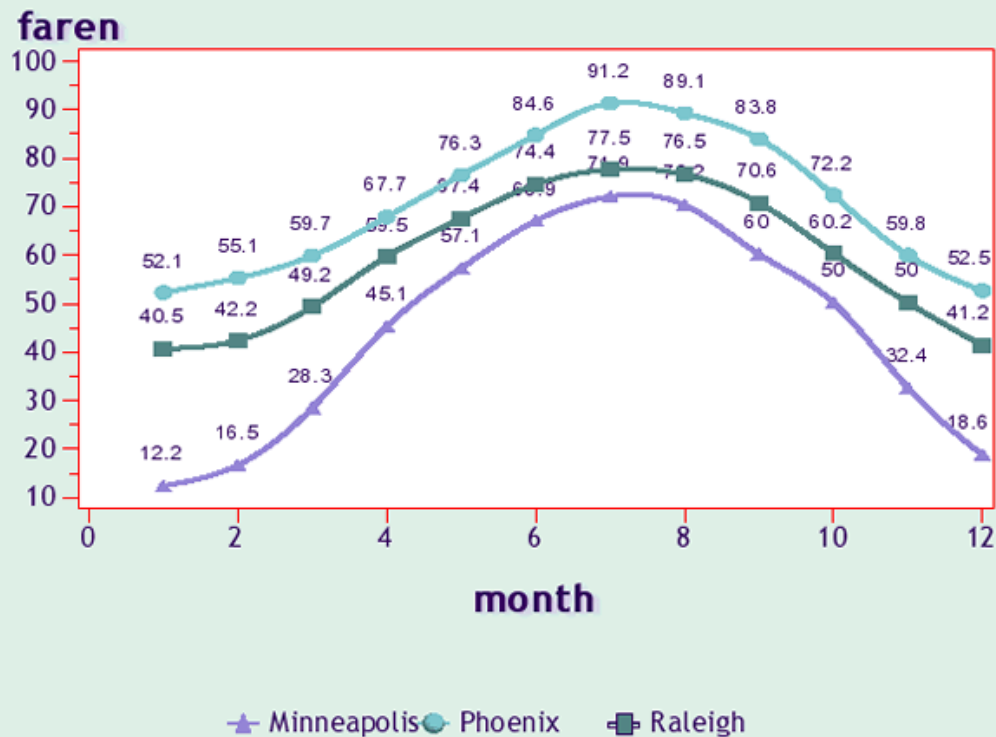


Več primerov...

Stephen Few, Perceptual Edge

<http://www.perceptualedge.com/examples.php>

Average Monthly Temperature



Vodila za pripravo tabel

- Informativni naslov (kaj, kje, kdaj, kako)
- Vir podatkov
- Oblikovanje besedila:
 - naslovna/zbirna vrstica/stolpec
 - decimalni tabulator
 - značilne decimalke
 - velikost
 - črkovanje
- Samozadostnost
 - vse kratice in okrajšave morajo biti razložene
 - vsi simboli morajo biti razloženi
 - opombe
- Barve uporabljajmo premišljeno

Primerjava leve in desne tabele

	Groupa			
	A		B	
Faktor	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Podoba	14,23	3,52	14,54	3,19
Breme	12,97	1,51	11,89	1,36
Interes	6,06	2,39	6,15	2,00
Veščine	10,71	3,02	11,49	3,15
Trud	11,26	2,70	11,18	2,43

Dejavnik	Eksperimentalna skupina		Kontrolna skupina	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Podoba	14,2	3,5	14,5	3,2
Breme	13,0	1,5	11,9	1,4
Interes	6,1	2,4	6,1	2,0
Veščine	10,7	3,0	11,5	3,1
Trud	11,3	2,7	11,2	2,4

M = aritmetična sredina; SD = standardni odklon

Heavy borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z.ZZZ
Group 2	\$X.X	Y%	Z.ZZZ
Group 3	\$X.X	Y%	Z.ZZZ
Group 4	\$X.X	Y%	Z.ZZZ
Group 5	\$X.X	Y%	Z.ZZZ

Light borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z.ZZZ
Group 2	\$X.X	Y%	Z.ZZZ
Group 3	\$X.X	Y%	Z.ZZZ
Group 4	\$X.X	Y%	Z.ZZZ
Group 5	\$X.X	Y%	Z.ZZZ

Minimal borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z.ZZZ
Group 2	\$X.X	Y%	Z.ZZZ
Group 3	\$X.X	Y%	Z.ZZZ
Group 4	\$X.X	Y%	Z.ZZZ
Group 5	\$X.X	Y%	Z.ZZZ

FIGURE 2.4 Table borders

Table

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Heatmap

LOW HIGH

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

FIGURE 2.5 Two views of the same data

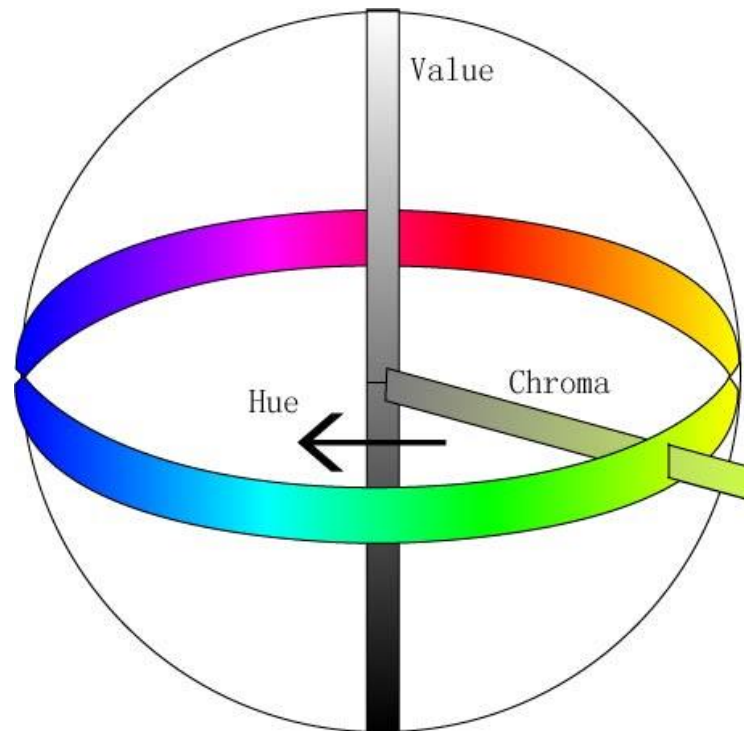
Kako najbolje izkoristiti senzorni spomin?

Predpozornostne lastnosti:


- Barva
- Oblika
- Postavitve/umestitev v prostor


Gestalt načela vidnega zaznavanja

Predpozornostne lastnosti: Barva

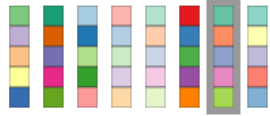



Barvne palette


Number of data classes: 6  how to use updates downloads credits

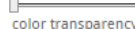
Nature of your data:  sequential ☐ diverging ☒ qualitative





Pick a color scheme:









Only show:  ☐ colorblind safe ☐ print friendly ☐ photocopy safe

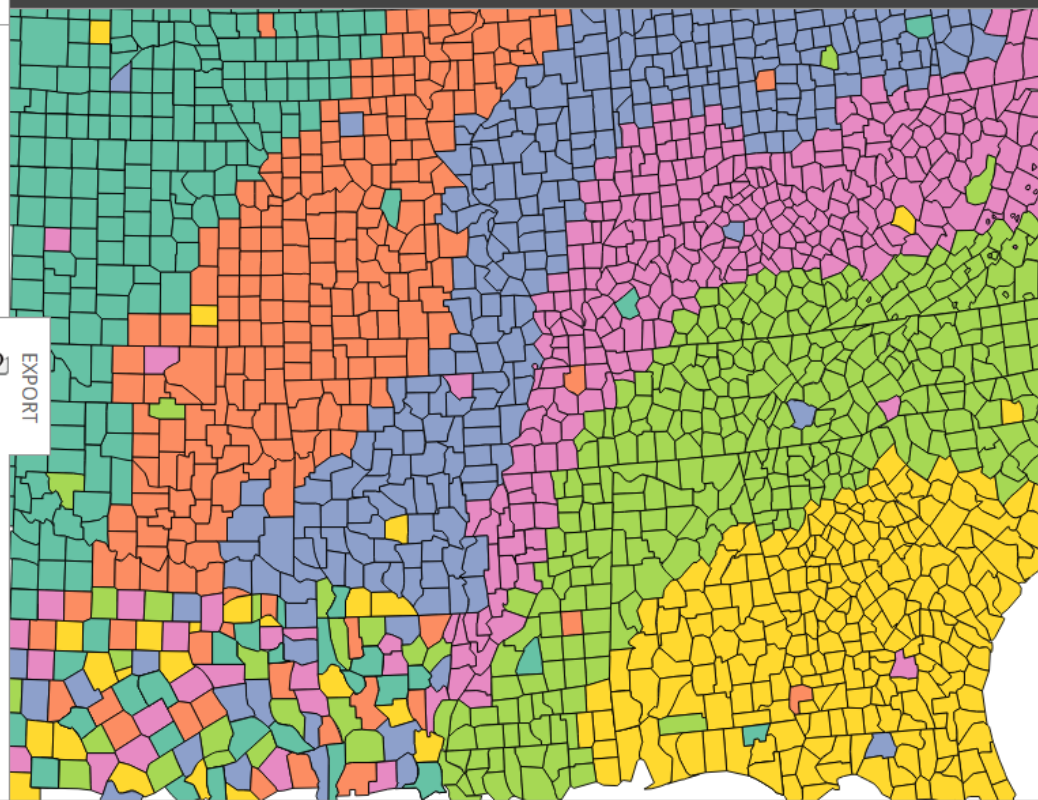
Context:  ☐ roads ☐ cities ☒ borders


Background: ☒ solid color ☐ terrain 


6-class Set2     EXPORT

RGB

	102,194,165
	252,141,98
	141,160,203
	231,138,195
	166,216,84
	255,217,47



© Cynthia Brewer, Mark Harrower and The Pennsylvania State University
 [Source code and feedback](#)
[Back to Flash version](#)
[Back to ColorBrewer 1.0](#)



Vizualne spremenljivke

Vir: Halik, Ł. (2012). The analysis of visual variables for use in the cartographic design of point symbols for mobile Augmented Reality applications, *Geodesy and Cartography*, 61(1), str. 21, najdeno na http://www.iag-aig.org/attach/30dee1f85f7bd479367f1f933d48b701/V61N1_2FT.pdf



Visual Variable	Author	Example
Size	Bertin (1967/83), Morrison (1974), MacEachren (1995), Kraak & Ormeling (2003), Krygier & Wood (2005), Dent et al. (2009), Sloum et al. (2010), Tyner (2010).	
Shape	Bertin (1967/83), Morrison (1974), MacEachren (1995), Kraak & Ormeling (2003), Krygier & Wood (2005), Dent et al. (2009), Sloum et al. (2010), Tyner (2010).	
Lightness/ value	Bertin (1967/83), Morrison (1974), MacEachren (1995), Kraak & Ormeling (2003), Krygier & Wood (2005), Dent et al. (2009), Sloum et al. (2010), Tyner (2010).	
Color (hue+saturation)	Bertin (1967/83).	
Orientation	Bertin (1967/83), Morrison (1974), MacEachren (1995), Kraak & Ormeling (2003), Dent et al. (2009), Sloum et al. (2010), Tyner (2010).	
Texture	Bertin (1967/83), Morrison (1974), MacEachren (1995), Kraak & Ormeling (2003), Krygier & Wood (2005), Dent et al. (2009), Tyner (2010).	
Location	Bertin (1967/83), MacEachren (1995), Kraak & Ormeling (2003), Krygier & Wood (2005), Dent et al. (2009), Sloum et al. (2010), Tyner (2010).	
Hue	Morrison (1974), MacEachren (1995), Kraak & Ormeling (2003), Krygier & Wood (2005), Dent et al. (2009), Sloum et al. (2010), Tyner (2010).	
Saturation/ intensity	Morrison (1974), MacEachren (1995), Krygier & Wood (2005), Dent et al. (2009), Sloum et al. (2010), Tyner (2010).	
Arrangement	Morrison (1974), MacEachren (1995), Dent et al. (2009), Sloum et al. (2010), Tyner (2010).	
Focus/ crispness	MacEachren (1995).	
Resolution	MacEachren (1995).	
Transparency	MacEachren (1995).	
Spacing	Sloum et al. (2010).	
Perspective Height	Sloum et al. (2010).	

Kombinacije več vizualnih spremenljivk

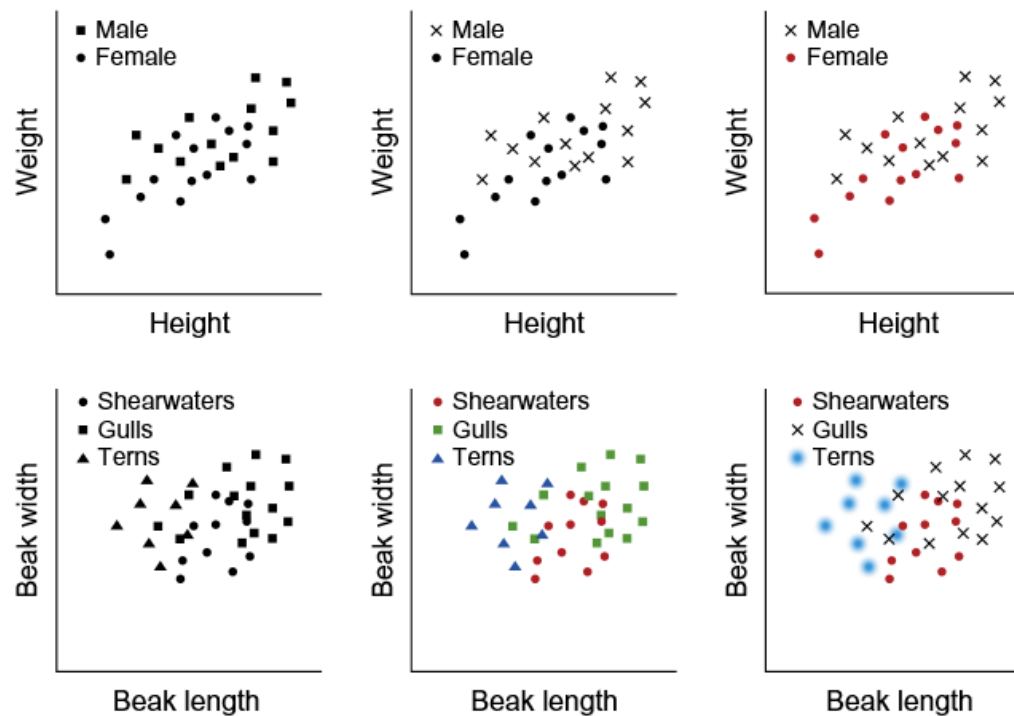


Figure 5.8 Feature channels can be used to make symbols more distinct from one another. The graphs on the right use redundant color coding in addition to more distinctive shapes.

Zaznavanje grafičnih elementov

Weber. Če je razlika med dvema črtama relativno

- velika \Rightarrow enostavno zaznamo
- majhna \Rightarrow težko zaznamo

Stevens. Pristranskost v presoji
dolžine < ploščine < prostornine

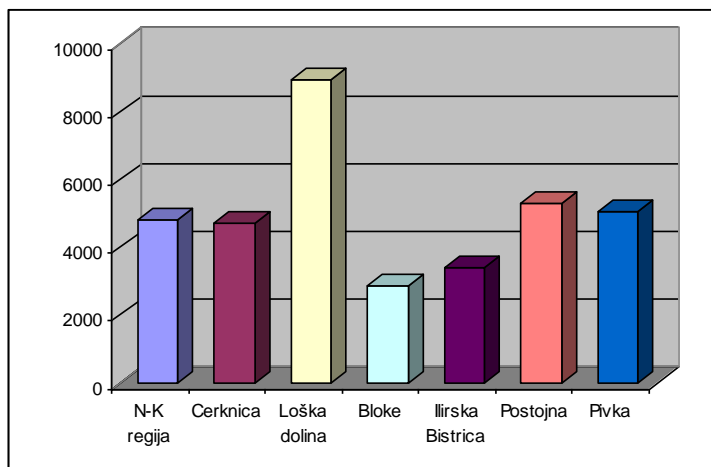
Cleveland. Razvrstitev zaznavanja kvantitativnih vrednosti od najbolj do najmanj točnega:

Položaj/Lega
Dolžina
Kot, naklon
Ploščina
Prostornina
Barva

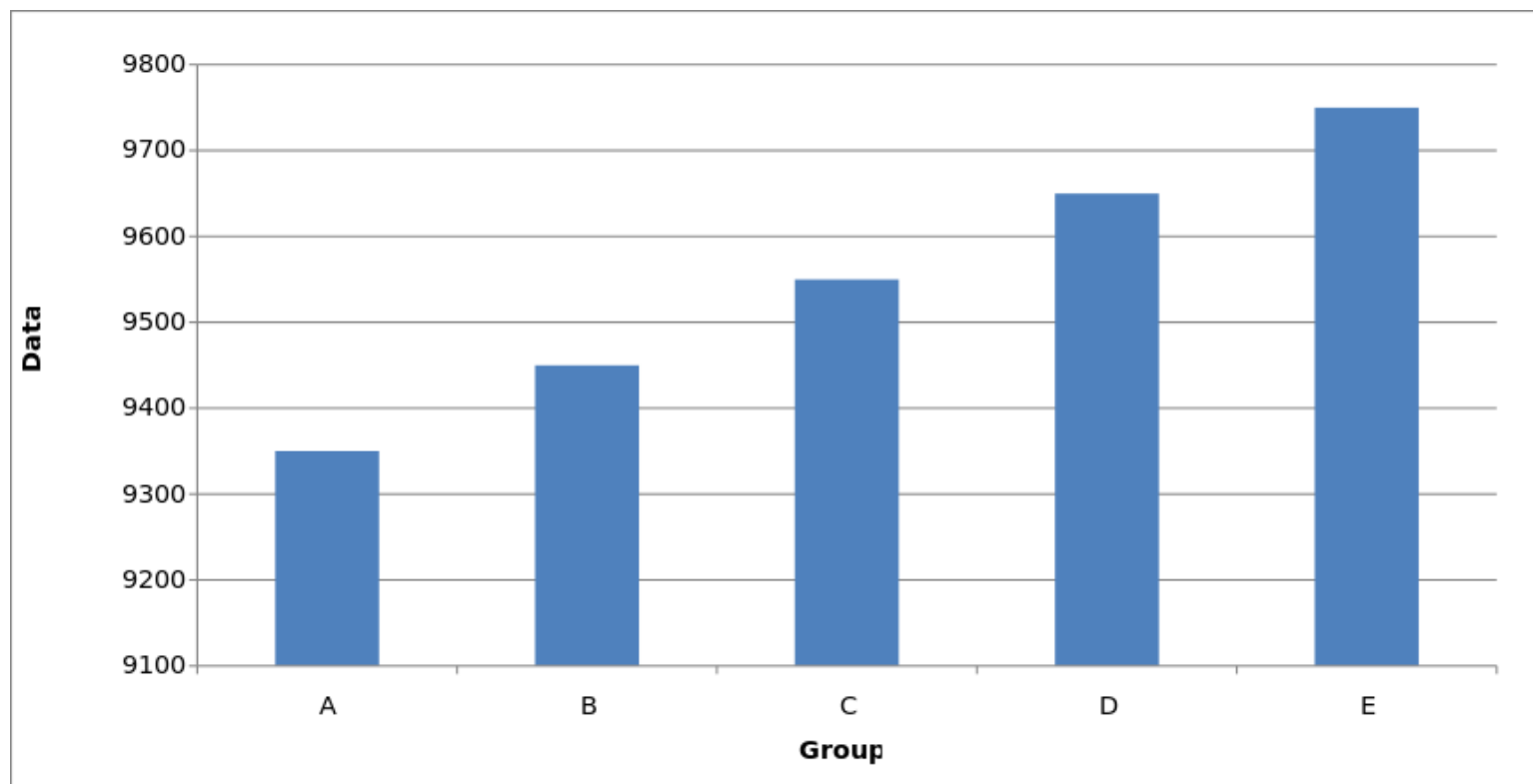
Problematicni prikazi (1)

How to make bad charts in 7 simple rules?

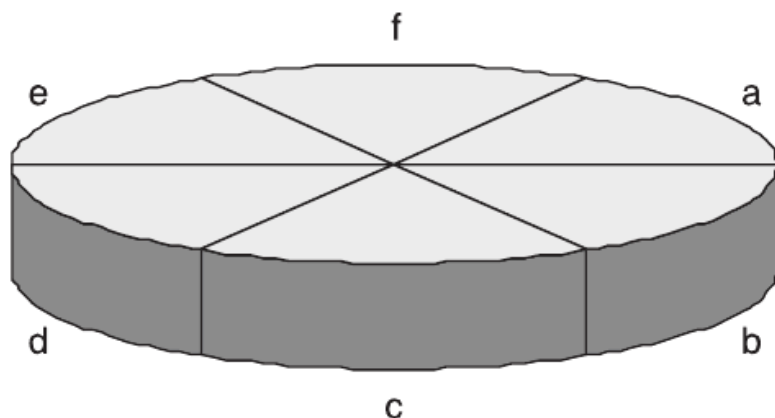
https://fullfact.org/factchecks/top_bad_infographics_charts-29075



Problematicni prikazi (2)



Problematicni prikazi (3)



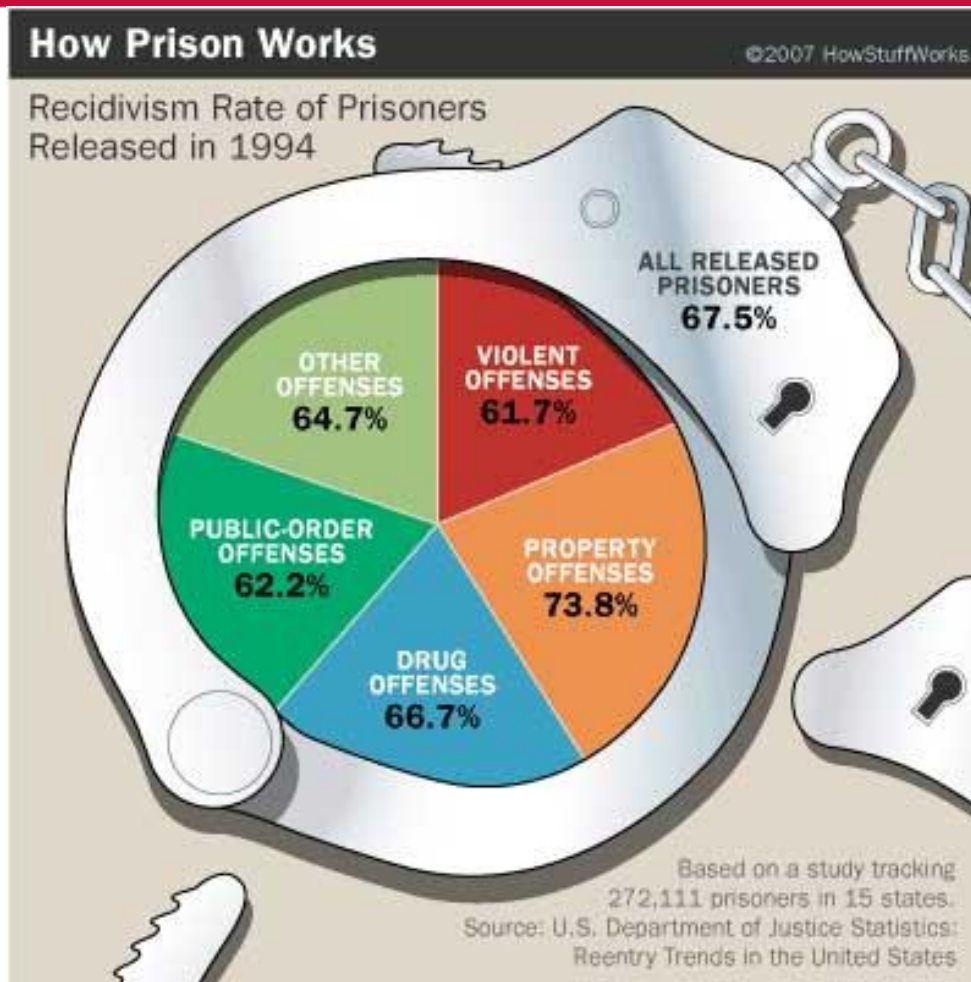
Vir: Hunt, N., & Mashhoudy, H. (2008). The Humble Pie – Half-Baked or Well Done? *Teaching Statistics*, 30(1), 6-12.



Resolution 5000 x3500 px - free download - www.psdgraphics.com

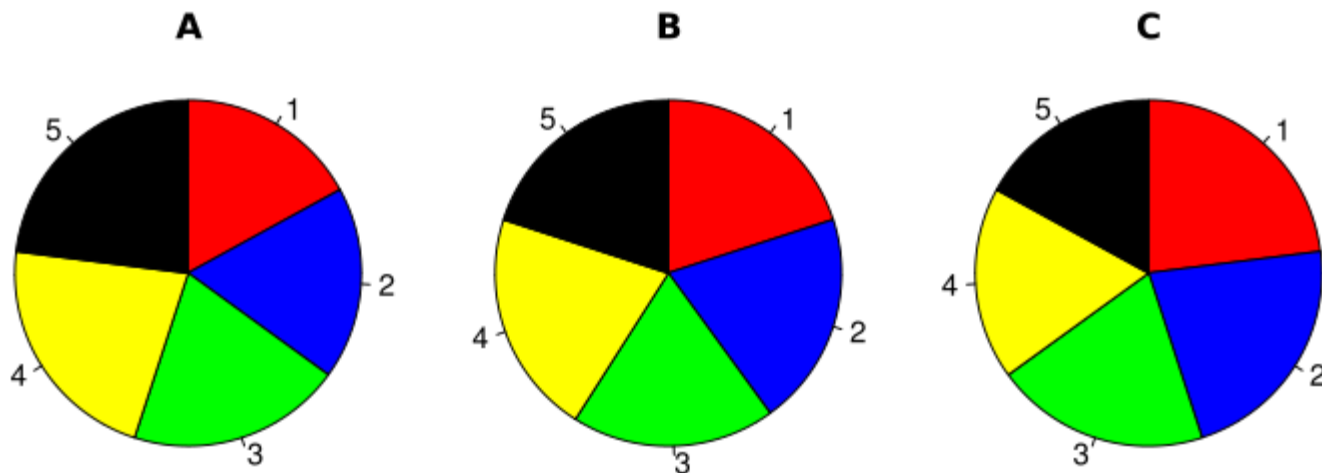
Vir: psdGraphics (2009). *3D pie chart*. Najdeno na <http://www.psdgraphics.com/icons/3d-pie-chart>

Problematicni prikazi (5)

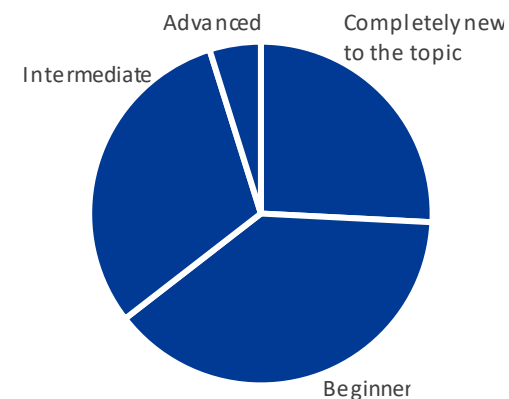
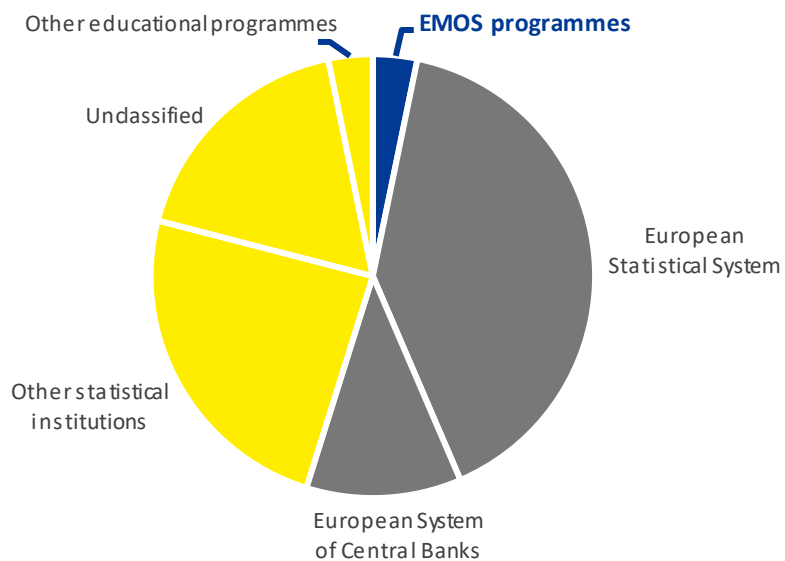


Problematicni prikazi (6)

Pri vsakem strukturnem krogu razvrsti krožne izseke po velikosti:



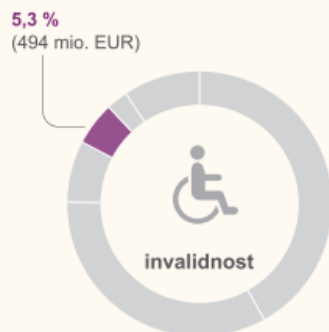
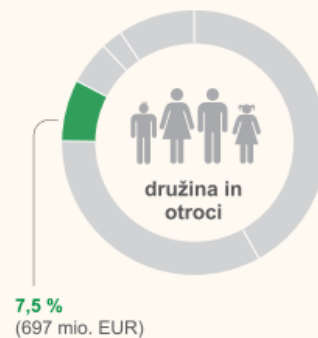
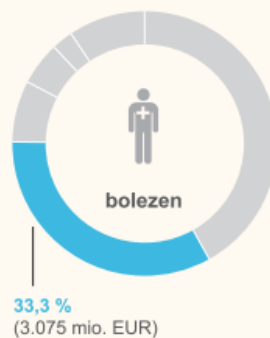
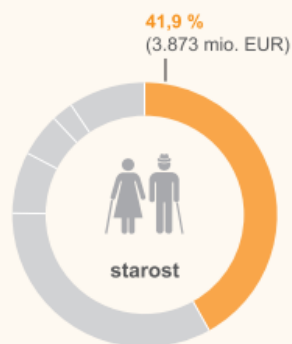
Alternativa (1)



Alternativa (2)

Izdatki za socialne prejemke po področjih socialne zaščite, 2016

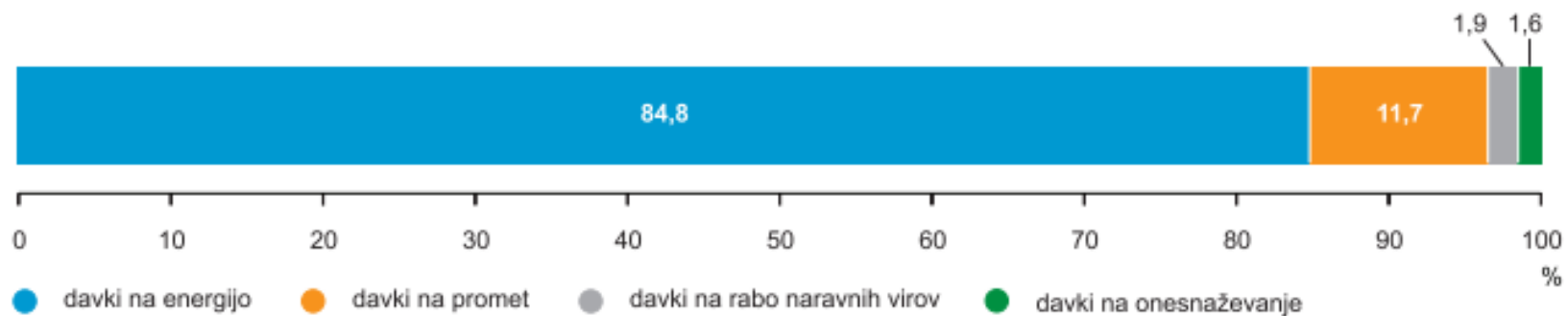
SOCIALNA ZAŠČITA



Alternativa (3)

Plačanih več okoljskih davkov

V Sloveniji je bilo v 2017 plačanih v proračun Republike Slovenije za 1.602 milijona EUR okoljskih davkov ali za skoraj 3 % več kot v 2016. Največji delež teh davkov je bil pobran z davki na energijo (85 %).



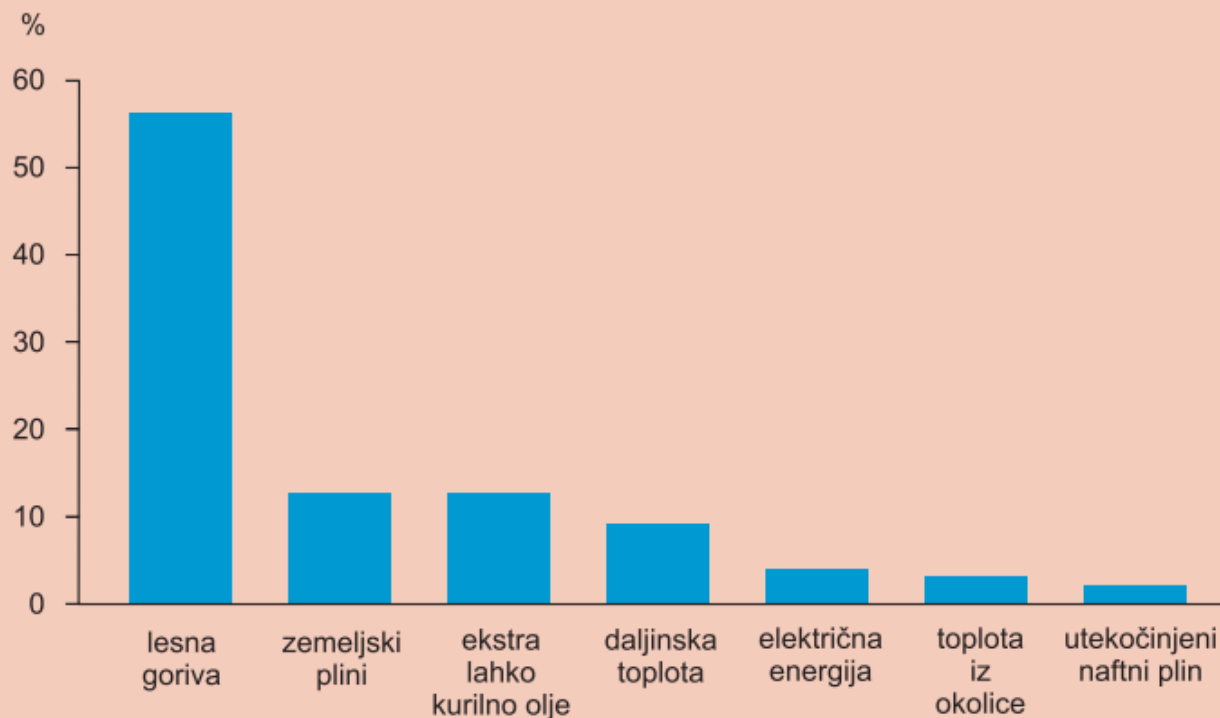
Vir: SURS

© SURS

Alternativa (4)

Poraba energije in goriv v gospodinjstvih za ogrevanje prostorov, 2017

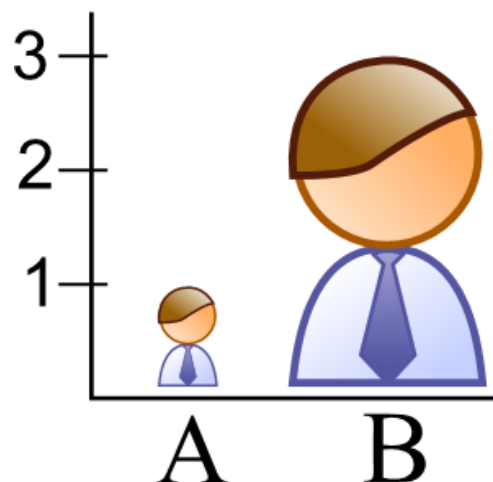
V gospodinjstvih se je v letu 2017 največ energije, namenjene končni rabi, porabilo za ogrevanje prostorov, in sicer 29.832 TJ ali 64 %. Največ gospodinjstev v Sloveniji je za ogrevanje prostorov uporabljalo lesna goriva (56,3 %); sledila so gospodinjstva, ki so prostore ogrevala z zemeljskim plinom (12,7 %) in ekstra lahkim kurilnim oljem (12,6 %).



Vir: SURS

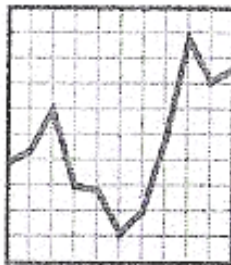
© SURS

Problematicni prikazi (7)



Problematicni prikazi (8)

Original Scale
Arrangement



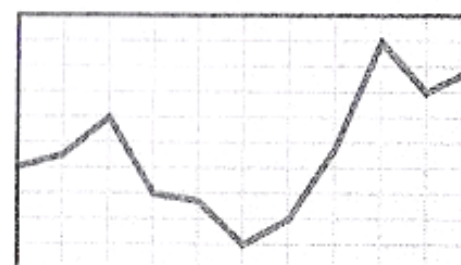
Expanding Vertical



Contracting
Horizontal



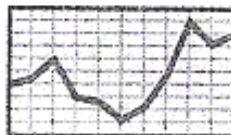
Expanding Horizontal



Expanding Vertical
and Contracting
Horizontal



Contracting Vertical

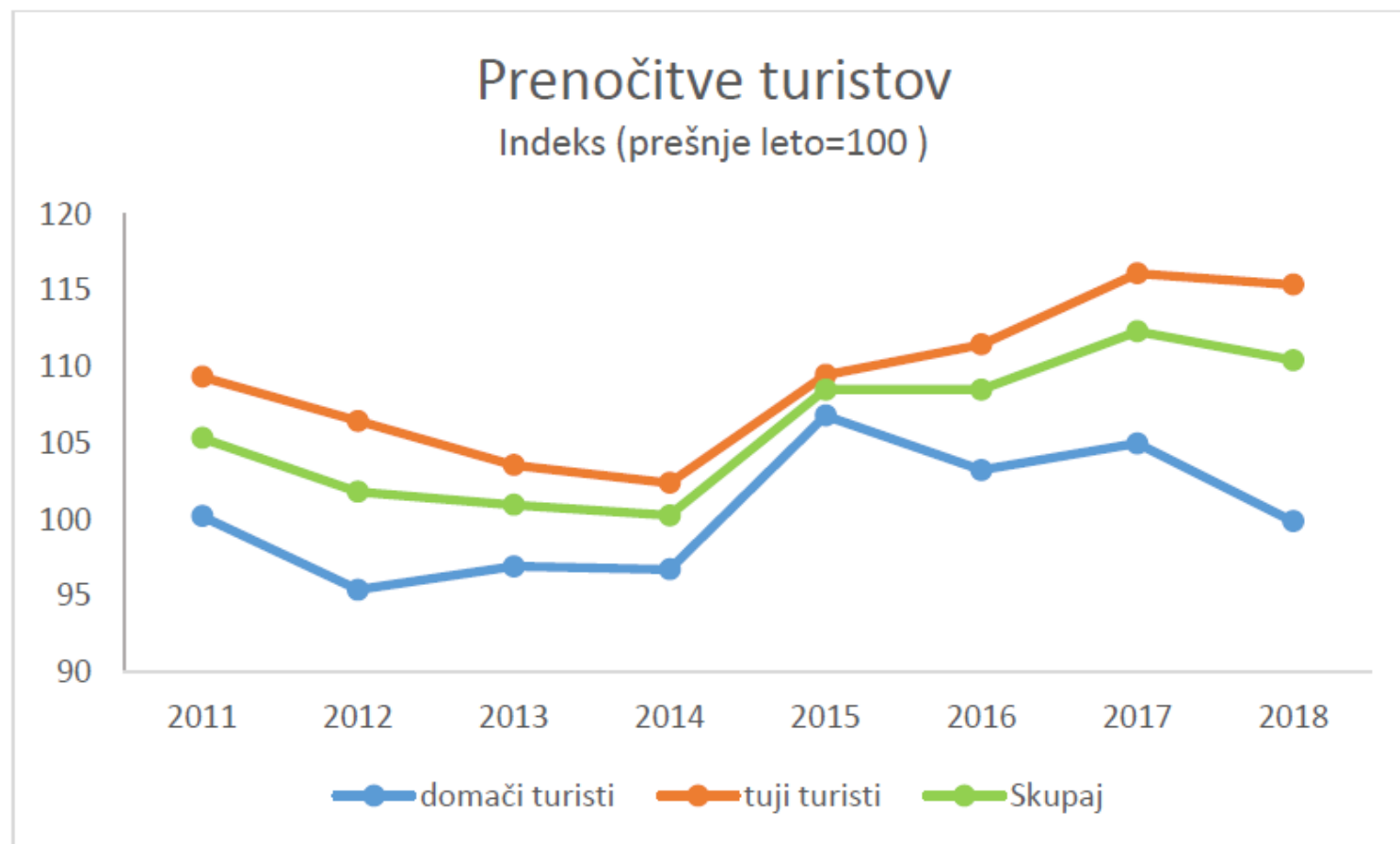


Contracting Vertical and Expanding Horizontal



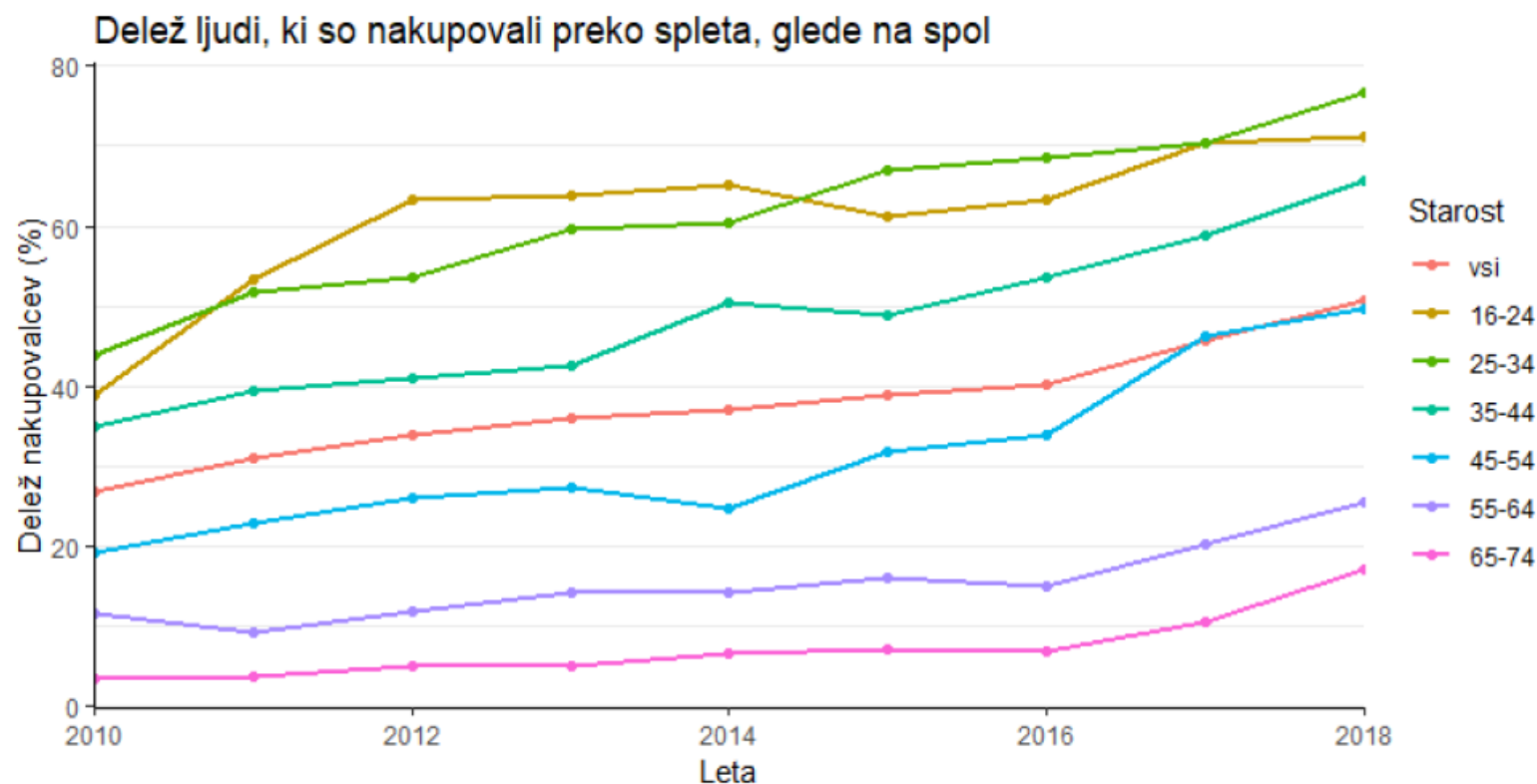
Indeks ali stopnje rasti?

Črta za primerjavo?



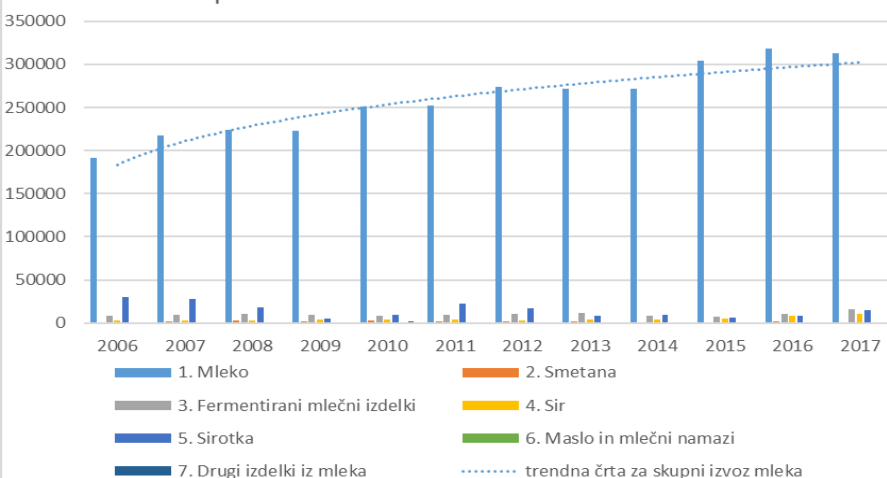
Naslov? Osi?

Barva? Vsi? Legenda?

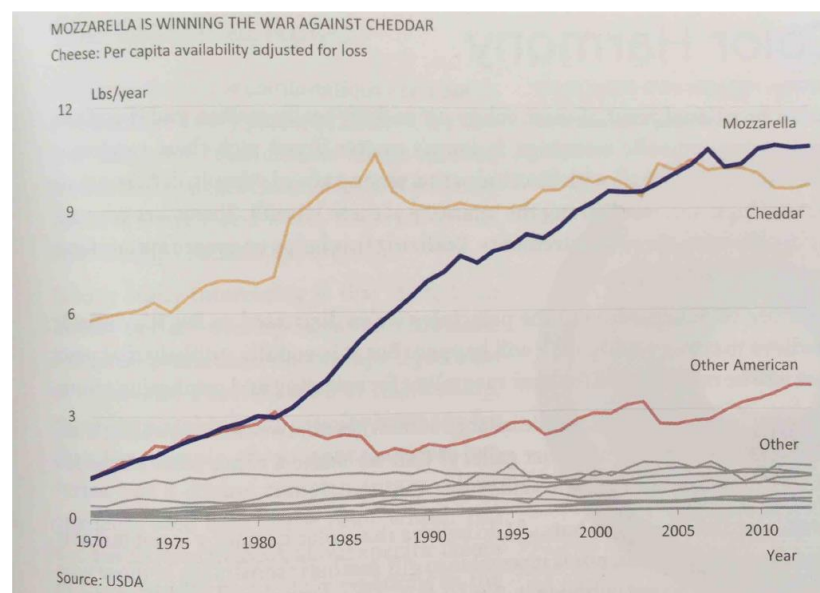
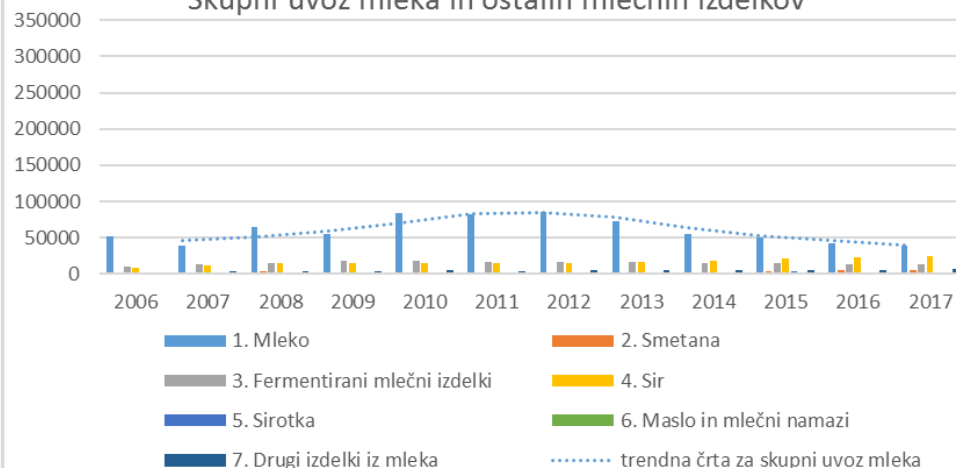


Preglednost? Vsebinsko sporočilo?

Skupni izvoz mleka in ostalih mlečnih izdelkov

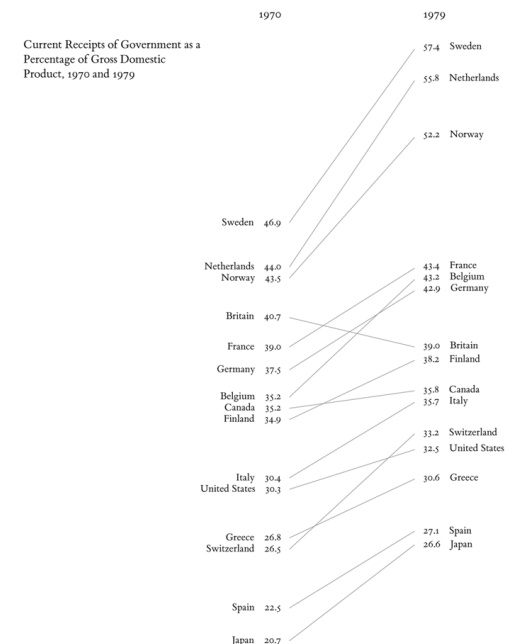
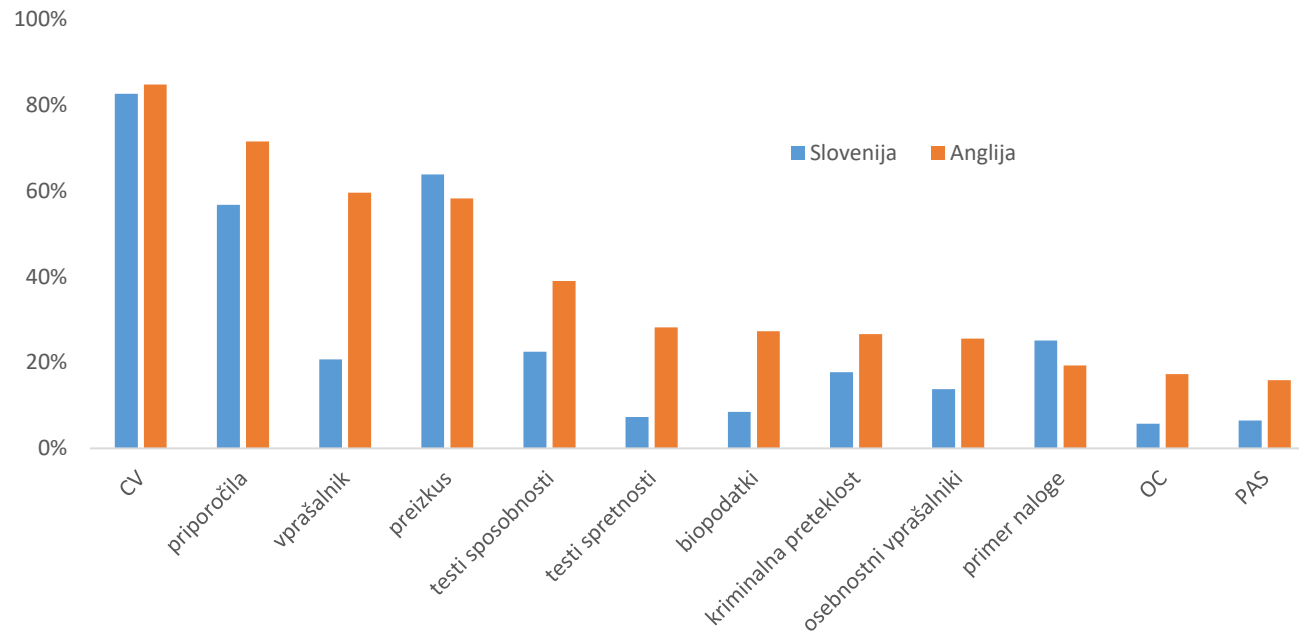


Skupni uvoz mleka in ostalih mlečnih izdelkov

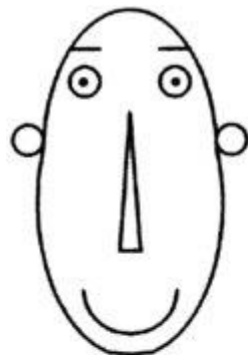


Slika 15.21. The leading actors (Mozzarella), the supporting actors (Cheddar and Other American cheeses), and the extras (the grey lines) (str. 391). Vir: Camoes, J. (2016). Data at work. New riders. 40

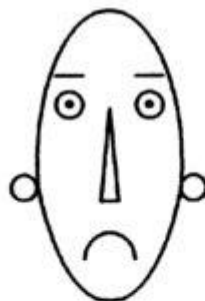
Idealen primer za slope graph?



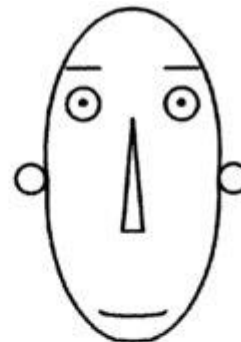
Problematicni prikazi (9)



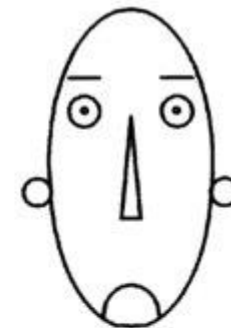
Control



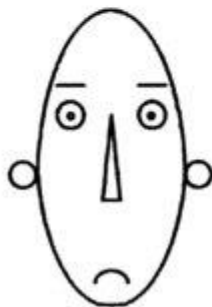
Cd-40



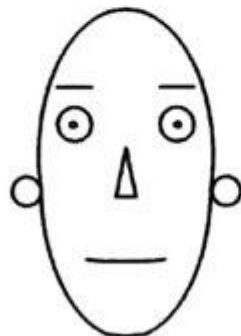
Cd-160



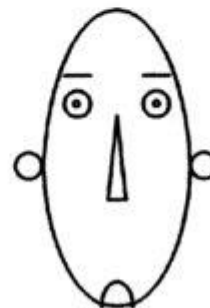
Cd-640



Cd-800



Zn-400



Zn-1600



Zn-6400

Pomembni vidiki vizualizacije

Narava podatkov

Vrsta uporabnikov & Namen

Medij



Vodila za pripravo grafikonov

Informativni naslov (kaj, kje, kdaj, kako)

Vir podatkov

- Grafikon:
 - najustreznejši grafikon
 - izbor fokusa
 - brez 3D ali drugih učinkov
 - pametna uporaba prostora
 - enostavno berljivo besedilo
- Osi:
 - ustrezna lestvica in intervali
 - značilne decimalke
- Prazno ozadje
- Samozadosten
- Premišljena uporaba barv (tiskanje, barvna slepota)

Kompleksnejši prikazi (1)

Small multiples

<https://archive.nytimes.com/www.nytimes.com/interactive/2013/12/20/health/a-racial-gap-in-breast-cancer-deaths.html>

Kompleksnejši prikazi (2)

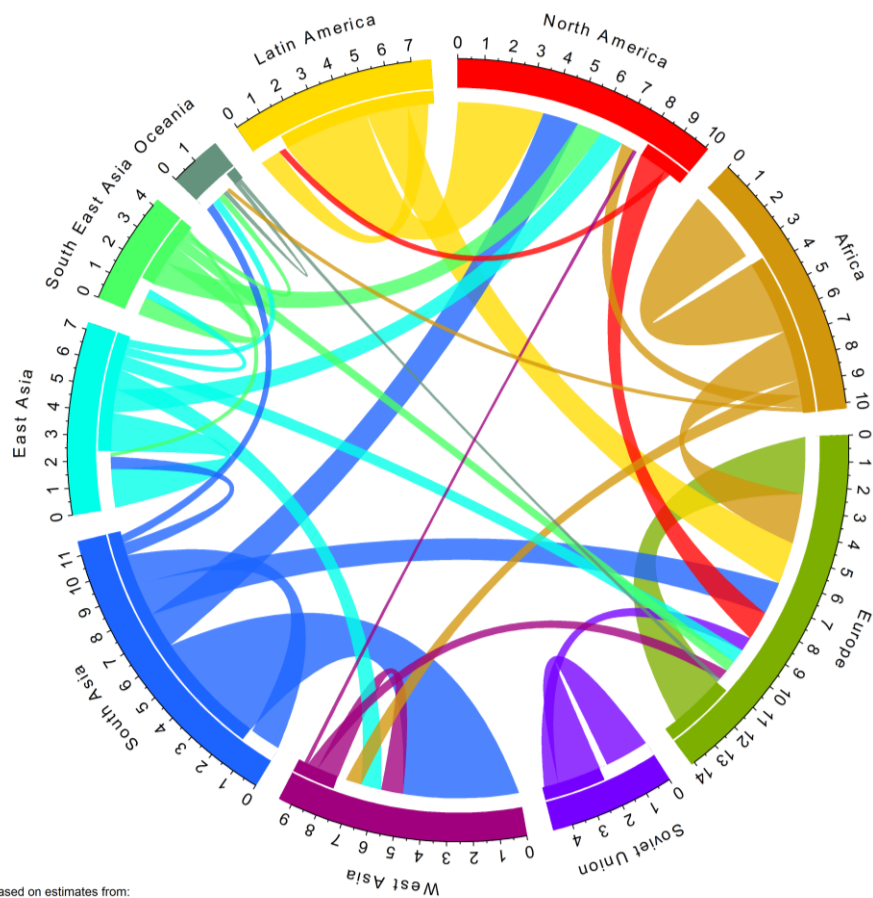
Prikaz kompleksnejših hierarhičnih struktur



⇒ Simoes, AJG, Hidalgo, CA (2011). The Economic Complexity Observatory: An Analytical Tool for Understanding the Dynamics of Economic Development. Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence.

https://oec.world/en/visualize/tree_map/hs92/export/svn/show/all/2017/

Kompleksnejši prikazi (3)



Based on estimates from:
Abel and Sander (2014) *Science* Vol. 343 no. 6178 pp. 1520 – 1522

Vir: Abel, G. (2014, 28. marec). Circular Migration Flow Plots in R. Najdeno na
<https://gjabel.wordpress.com/2014/03/28/circular-migration-flow-plots-in-r/>

Interaktivni prikaz na: http://download.gsb.bund.de/BIB/global_flow/

Kompleksnejši prikazi (4)

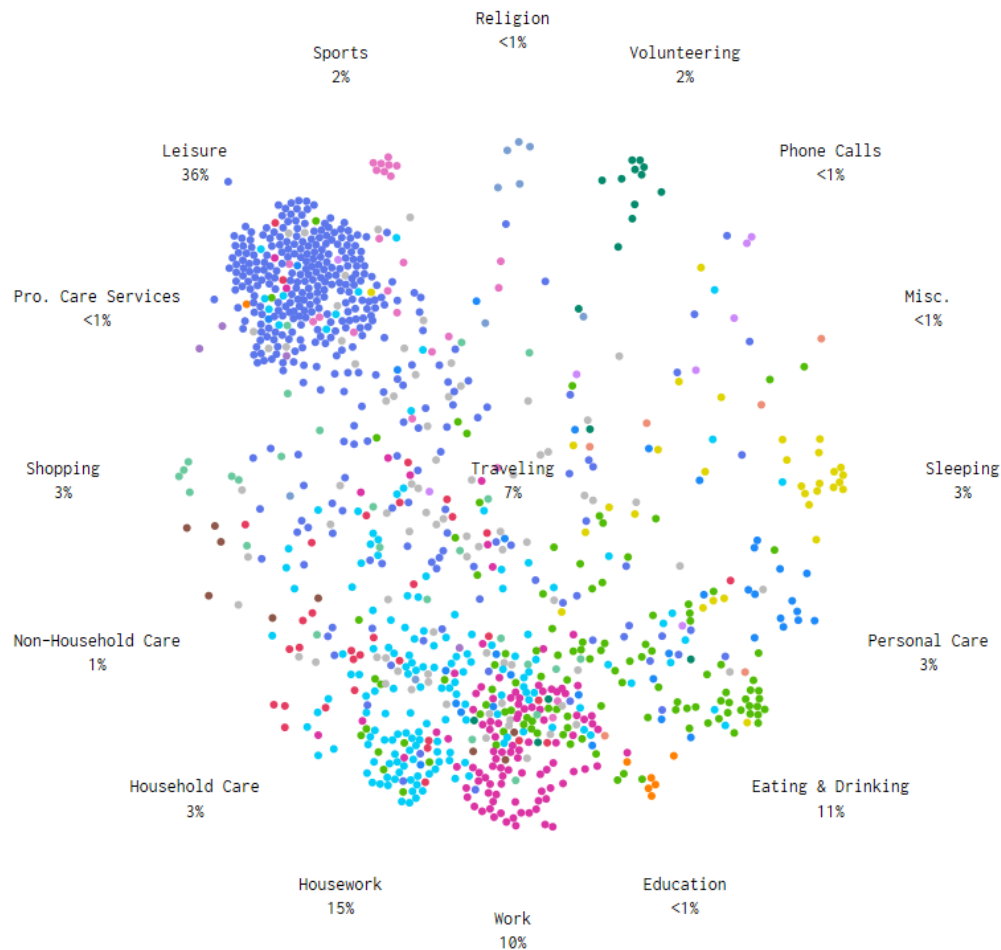
5:56pm

SLOW

MEDIUM

FAST

With the work day done, it's time to commute home and fix dinner or go out for a while.



This is a simulation of 1,000 people's average day. It's based on 2014 data from the [American Time Use Survey](#), made way more accessible by the [ATUS Extract Builder](#).



Vir: Yau, N. (2015). Time use simulation. Najdeno na <http://www.flowingdata.com/projects/2015/timeuse-simulation/>

Vizualizacije glede na ustvarjalca zgodbe

Table 1. Properties of Author-Driven and Reader-Driven Stories. Most visualizations lie along a spectrum between these two extremes.

Author-Driven	Reader-Driven
Linear ordering of scenes	No prescribed ordering
Heavy messaging	No messaging
No interactivity	Free interactivity