

Datu dzīves cikls mūsu acīs

Kā labāk datus pasniegt

2021. gada 6. janvāris

Jānis Lazovskis

Latvijas Universitāte + Rīgas Biznesa Skola

Pārskats

1. Abstraktais
2. Iespējamais
3. Praktiskais

1. Krāsošana
2. Savienošana

*caur piemēriem

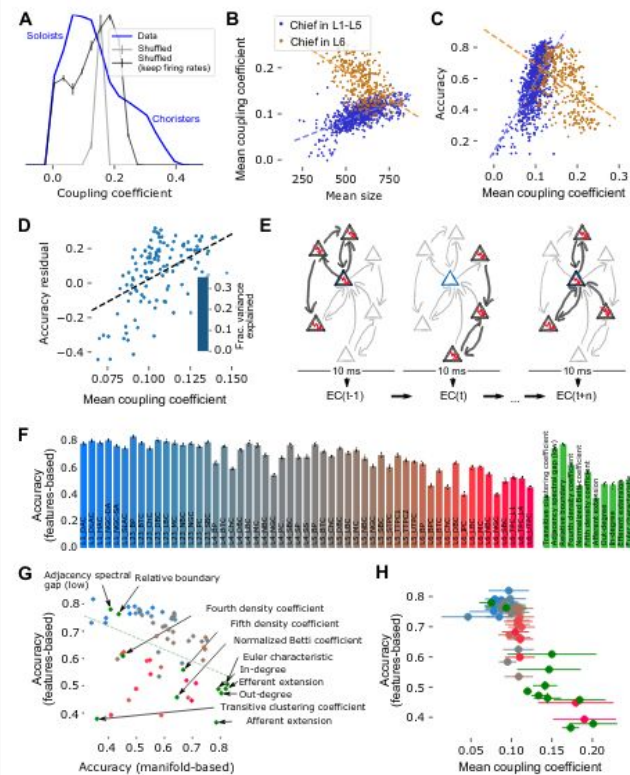
Abstraktais

Redzot datus:

- Kas te notiek?
- Kas man būtu jāredz?

Pasniedzot datus:

- Ko es vēlos pateikt?
- Kas ir lieks?
- Ko domā cilvēks pirms tam?



	A	B	C	D	E	F
1	song_name	danceability	energy	key	loudness	mode
2	Mercury: Retrograde	0.831	0.814	2	-7.364	1
3	Pathology	0.719	0.493	8	-7.23	1
4	Symbiote	0.85	0.893	5	-4.783	1
5	ProductOfDrugs (Prod. The Virus and Antidote)	0.476	0.781	0	-4.71	1
6	Venom	0.798	0.624	2	-7.668	1
7	Gatteka	0.721	0.568	0	-11.295	1
8	kamikaze (+ pulse)	0.718	0.668	8	-4.162	1
9	T.R.U. (Totally Rotten Underground)	0.694	0.711	8	-5.525	1
10	I Put My Dick In Your Mental	0.774	0.751	1	-2.445	1
11	Andromeda	0.893	0.907	11	-10.406	1
12	BRAINFOOD	0.864	0.365	8	-10.219	1
13	Troll Under the Bridge	0.736	0.932	1	-3.726	1
14	1000 Rounds	0.825	0.761	8	-5.389	1
15	Sacrifice	0.767	0.576	10	-9.683	0
16	Backpack	0.765	0.726	5	-5.58	1
17	D(R)Own	0.617	0.541	6	-4.113	1
18	Okay, But This Is The Last Time	0.755	0.298	1	-15.032	1
		0.814	0.575	11	-9.635	1

tabula (table)
nekārtoti objekti ar īpašībām

tabula (weighted adjacency matrix)
objektu pāru savstarpējās īpašības

0	-4	-2	-2	-1	-3	-2	1
-4	0	1	0	0	-2	1	-1
-2	1	0	-1	-3	1	1	0
-2	0	-1	0	-2	-3	-1	-2
-1	0	-3	-2	0	-3	-1	-2
-3	-2	1	-3	-3	0	-3	-1
-2	1	1	-1	-1	-3	0	-3
1	-1	0	-2	-2	-1	-3	0

```

2
3 {
4   "particles": {
5     "number": {
6       "value": 80,
7     "density": {
8       "enable": true,
9       "value_area": 800
10    }
11  },
12  "color": {
13    "value": "#01b6ed"
14  },
15  "shape": {
16    "type": "circle",
17    "stroke": {
18      "width": 0,
19      "color": "#01b6ed"
20    },
21    "polygon": {
22      "nb_sides": 10
23    },
24    "image": {
25      "src": "img/github.svg",
26      "width": 100,
27      "height": 100
28    }
29  },
30  "opacity": {
31    "value": 0.5,
32    "random": false,
33    "anim": {
34      "enable": false,
35      "speed": 1
36    }
37  }
38 }
39
40 0.1,

```

vārdnīca (dictionary)
hierarhiskas īpašības

date	2017-01-01	2017-01-02	2017-01-03	2017-01-04	2017-01-05	2017-01-06	2017-01-07	2017-01-08	2017-01-09	2017-01-10	2017-01-11	2017-01-12	2017-01-13
meantemp	15.91	18.5	17.11	18.7	18.39	19.32	14.71	15.68	14.57	12.11	11	11.79	13.24
humidity	85.87	77.22	81.89	70.05	74.94	79.32	95.83	83.53	80.81	71.94	72.11	74.58	67.06
wind_speed	2.74	2.89	4.02	4.55	3.3	8.68	10.04	1.95	6.54	9.36	9.77	6.63	6.44
meanpressure	59	1018.28	1018.33	1015.7	1014.33	1011.77	1011.38	1015.55	1015.95	1016.89	1016.78	1016.37	1017.53

tabula (time series)
nekārtoti objekti ar kārtotām īpašībām

Iespējamais

Datu paveidi

Objekti un to īpašības

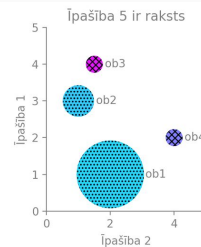
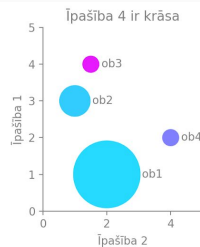
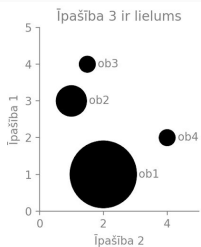
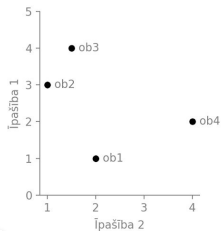
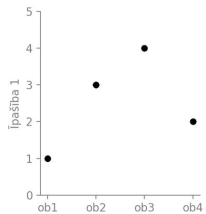
- Mērījumi, aptaujas
- Parametri

Objekti vai īpašības mainās

- Atkārtoti mērījumi
- Parametru maiņa

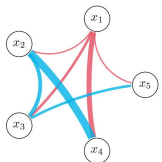
Objektu attiecības

- Savstarpējie mērījumi



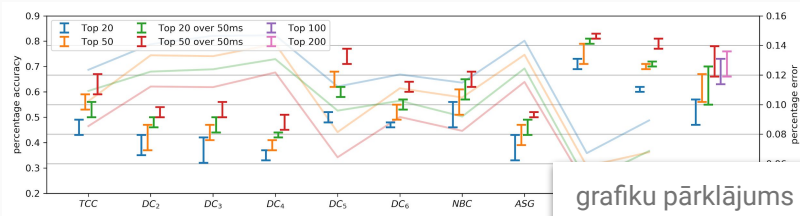
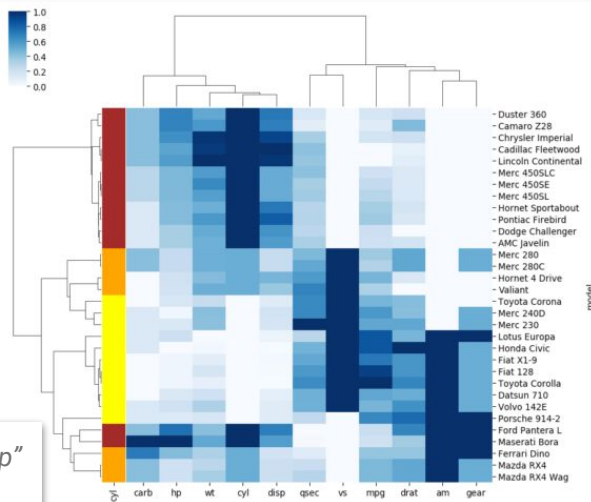
Īpašību vizualizācija

	x_1	x_2	x_3	x_4	x_5
x_1	0	1	2	3	1
x_2	1	0	-2	-6	0
x_3	2	-2	0	0	-1
x_4	3	-6	0	0	0
x_5	1	0	-1	0	0



attiecības kā grafs

Īpašības kā "heatmap" un "dendrogram"



grafiku pārklājums ar "error bars"

Iespējamais

Grafiku paveidi

Kas lec acīs?

- Krāsa
- Kontrasts
- Teksts / tukšums

** Svarīgais = pirmais, ko redz

matplotlib
Version 3.3.3

Installation Documentation Examples Tutorials Contributing

home | contents | Matplotlib: Python plotting

Matplotlib: Visualization with Python

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.

Matplotlib makes easy things easy and hard things possible.

Create Customize

- Develop publication quality plots with
- Take full control of line styles, font
- Explore the

matplotlib (matplotlib.org)

Plotly (dash-gallery.plotly.host)

Image Processing Support Vector Machine Drug Discovery

DISTRIBUTION

Violin Density Boxplot Histogram

CORRELATION

Scatterplot Connected Scatterplot Bubble plot Heatmap 2D density plot Correlation

RANKING

Barplot Stacked barplot Gantt plot Lollipop plot Word cloud Spider

PART OF A WHOLE

Stacked barplot Treemap Venn diagram Doughnut plot Pie plot Treemap

Python Graph Gallery
(python-graph-gallery.com)

PGFPlots gallery (pgfplots.sourceforge.net)

Iespējams

Piemērs: Grafiku piemēru kolekcijas

Viss gatavs

- Plotly
- Chart.js, D3.js

Pats visu kontrolē

- The Python Graph Gallery
- matplotlib

Pats visu kontrolē (ekstrēmi)

- TeXample.net
- PGFPlots Gallery

The screenshot shows the Kaggle website with a search bar at the top. Below it, there are several dataset listings with titles like 'World Health 2020', 'Pfizer Vaccine Tweets', and 'Covid-19 Case Surveillance Public Use Dataset'. Each listing includes a small image, a title, a description, and a 'Hotness' score.

Kaggle ([kaggle.com](https://www.kaggle.com))

The screenshot shows the Latvian Open Data Portal (data.gov.lv) with a search bar and a list of datasets. The main heading reads 'Laipni lūdzam Latvijas Atvērto datu portālā!' (Welcome to the Latvian Open Data Portal!). It features statistics for 'DATU KOPAS' (433) and 'PUBLICĒTĀJI' (83). Below, there are category filters for 'Ārlietas', 'Ekonomika un uzņēmējdarbība', 'Enerģija', and 'Izcīņotāji un sabiedrība'.

Atvērto Datu Portāls (data.gov.lv)

The screenshot shows the EU Open Data Portal with a search bar and a list of datasets. The main heading reads 'EU Open Data Portal'. It features a search bar and a list of datasets with titles like 'COVID-19 Coronavirus data' and 'Research Electronic Data (RED) - public procurement notices from the EU and beyond'.

EU Open Data Portal (data.europa.eu)

The screenshot shows the Janelia neuPrint website with a neuron morphology visualization and a connectivity matrix. The main heading reads 'Janelia neuPrint (neuprint.janelia.org)'. It features a search bar and a list of datasets with titles like 'Example Neuron Morphology' and 'Synapse Connectivity (broken down by cell type and region)'.

Janelia neuPrint (neuprint.janelia.org)

Iespējamais

Piemērs: datu krājumi

Vispārīgi

- Kaggle
- Atvērto Datu Portāls
- EU Open Data Portal

Specifiski

- Allen Brain Map
portal.brain-map.org
- Blue Brain Nexus
bluebrainnexus.io
- Janelia
neuprint.janelia.org

Praktiskais

Praktiska datu lietošana un pasniegšana

Sagatavoti

- MS Excel, Google Sheets
- MS Word
- Canva
- Plotly, Chart.js

Neierobežoti

- Python: CSV, JSON, pandas
- LaTeX: TikZ, PGFPlots

```

import pandas as pd
>>> df = pd.read_pickle('mc2_newparams2.pkl')
>>> df

```

	tpsg_strong_low	tpsg_strong_high	tpsg_low	tpsg_high	reciprocal_pairs	zero_outdeg	zero_indeg
0	0.456522	0.456522	0.266667	0.266667	1	15	3
1	0.110756	0.110756	0.094926	0.094926	44	22	11
2	0.119700	0.119700	0.111111	0.111111	25	20	7
3	0.000000	0.000000	0.333333	0.333333	0	9	1
4	0.133989	0.133989	0.114300	0.114300	20	23	3
...
31341	0.068787	0.068787	0.068950	0.068950	162	3	16
31342	0.140519	0.140519	0.139227	0.139227	40	4	3
31343	0.030780	0.030780	0.030934	0.030934	803	3	16
31344	0.174490	0.174490	0.168207	0.168207	17	9	10
		0.037836	0.038262	0.038262	590	2	20

Python + pandas

```

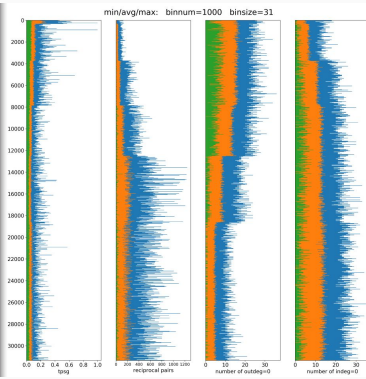
def plot_surfs():
    # load dataframe and initiate figure
    print('loading dataframe, flush=True')
    df = pd.read_pickle(database_dir+'mc2_newparams.pkl')
    df_strong = pd.read_pickle(database_dir+'mc2_tpsgstrong.pkl')
    fig = plt.figure(figsize=(10,10)) # default is (8,6)

    # tpsg
    print('computing and plotting: tpsg, flush=True')
    ax1 = fig.add_subplot(4,1,1)
    data_tpsg = df[['tpsg_high']].values
    data_tpsg_surf = data_tpsg
    ax1.plot(data_tpsg_surf, label='tpsg', alpha=0.8)
    data_tpsg_strong = df_strong[['tpsg_strong_high']].values
    data_tpsg_strong_surf = data_tpsg_strong
    ax1.plot(data_tpsg_strong_surf, label='tpsg_strong', alpha=0.8)
    data_st = sum(df[['tpsg_high']], axis=1)
    data_st_surf = data_st
    ax1.plot(data_st_surf, label='difference', alpha=0.8)
    ax1.set_ylabel('tpsg, fontsize=10'); plt.yticks(fontsize=10)
    ax1.set_xlabel('number of tpsg')

    # reciprocal edges
    print('computing and plotting: reciprocal edges, flush=True')
    ax2 = fig.add_subplot(4,1,2)
    data_rec = df[['reciprocal_pairs']].values
    data_rec_surf = data_rec
    ax2.plot(data_rec_surf, label='reciprocal pairs', alpha=0.8)
    ax2.set_ylabel('reciprocal pairs, fontsize=10'); plt.yticks(fontsize=10)
    ax2.set_xlabel('number of reciprocal pairs')

    # outdegree
    print('computing and plotting: out degree, flush=True')
    ax3 = fig.add_subplot(4,1,3)
    data_out = df[['zero_outdeg']].values
    data_out_surf = data_out
    ax3.plot(data_out_surf, label='zero outdeg', alpha=0.8)
    ax3.set_ylabel('number of outdeg=0, fontsize=10'); plt.yticks(fontsize=10)
    ax3.set_xlabel('number of outdeg=0')

```



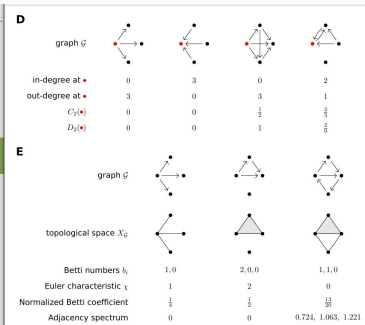
Python + matplotlib

```

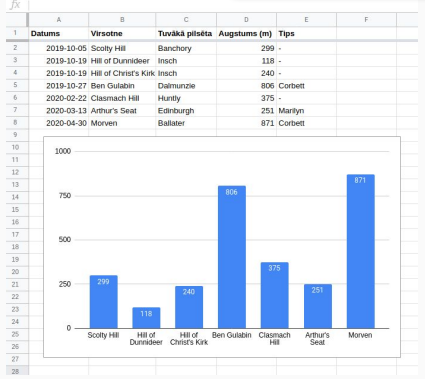
%matplotlib inline
132 | vertices = G.v() to [end right=30] (v);
133 |
134 | vertices(c) to [end right=45] (c);
135 | vertices(labeled) at (1,4,-5) (strongly connected);
136 |
137 | %v
138 | %v d
139 | %v
140 |
141 |
142 | %v[boxcol] | 2,5, 1 | rectangle (24,-8);
143 | vertices(only) at (1,2,1) (1) (largest box(D));
144 | %v[scope] | shift(1,4,-4.8));
145 |
146 |
147 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
148 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
149 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
150 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
151 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
152 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
153 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
154 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
155 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
156 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
157 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
158 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
159 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
160 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
161 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
162 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
163 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
164 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
165 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
166 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
167 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
168 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
169 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
170 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
171 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
172 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
173 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
174 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
175 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
176 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
177 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
178 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
179 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
180 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
181 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
182 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
183 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
184 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
185 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
186 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
187 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
188 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
189 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
190 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
191 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
192 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
193 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
194 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
195 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
196 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
197 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
198 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
199 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));
200 | vertices(labeled, anchor=west) at (0,2,4) (graph structure (G));

```

LaTeX + TikZ

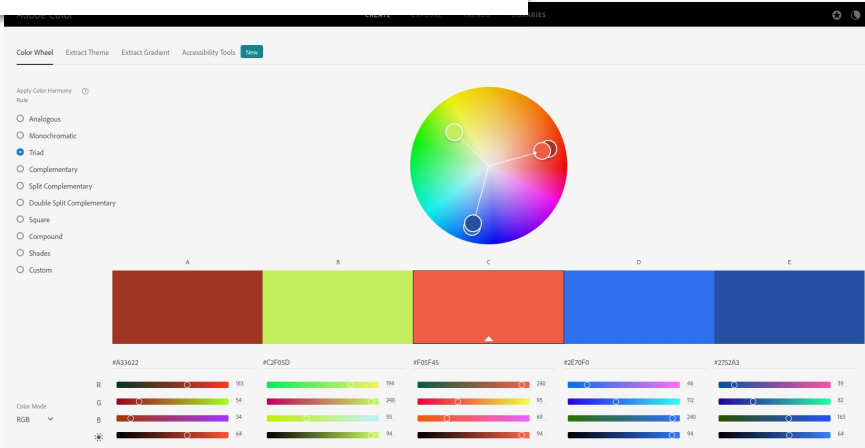


Virsnotes Google Sheets



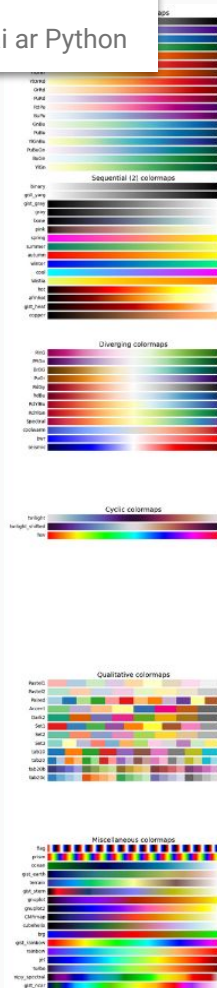
Adobe Color Wheel (color.adobe.com)

- līdžīgu krāsu izvēle
- redzes traucējumu palīglīdzekļi



matplotlib colormap

- krāsu paletes lietošanai ar Python



Krāsošana

Krāsu izvēle

Pievilkt uzmanību

- Sarkans, zaļš
- “Complementary”

Mazināt uzmanību

- Gaišs
- Līdžīgs fonam

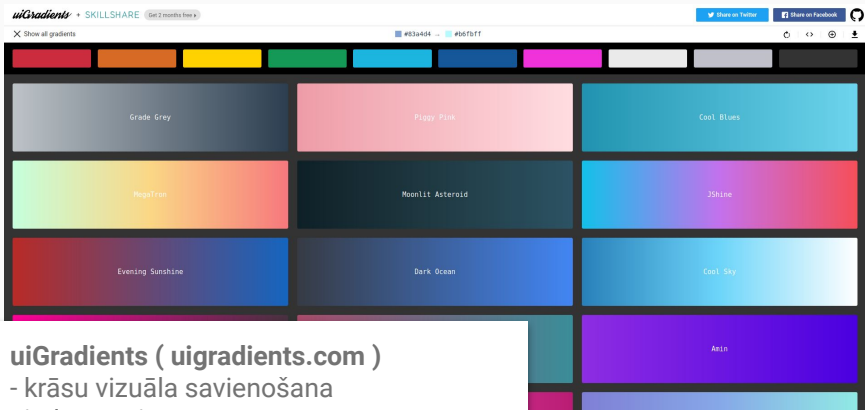
Papildināt

- Raksti (“patterns”)

** Organizācijām bieži ir “style guide”

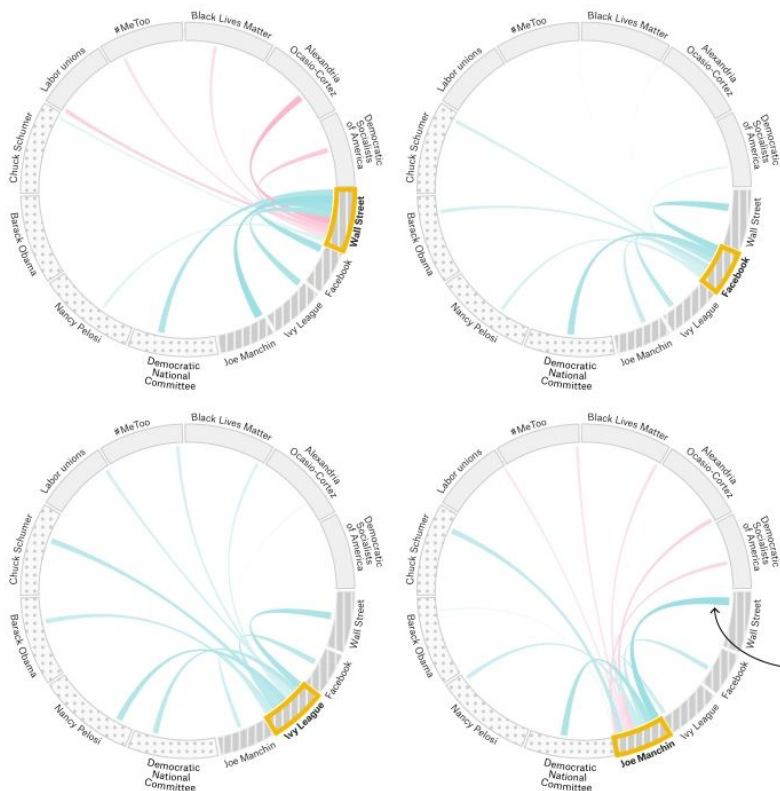
uiGradients (uigradients.com)

- krāsu vizuāla savienošana
- iedvesmai



Neoliberals seem at odds with the progressive left

How favorability ratings of 13 politicians and organizations are related, focusing on figures associated with “neoliberals,” from a poll of 2,900 likely Democratic primary voters



The strongest connection within the neoliberal group is between Joe Manchin and Wall Street

Krāsošana

Piemērs: Ryan Best, Shom Mazumder, FiveThirtyEight

- Galvenā doma virsrakstā
- Krāsu diapazons



- Mazāka vērtība = gaišāka krāsa
- Papildus informācija = “patterns”

Savienošana

Piemērs: Graphic Detail blog, The Economist

Vairāk informācijas

- Vai vērts rādīt?
- Ko tas papildus dod?
- Kā vizuāli labāk savienot?

Interaktīva informācija

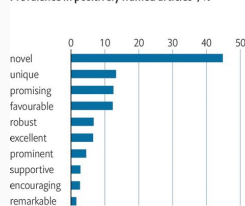
- Plotly
- D3.js

atsevišķi palīdz

Showvinism

Use of positive words in abstracts and titles of clinical-research articles, 2002-17

Prevalence in positively framed articles*, %



Prevalence in all sampled articles, %



Source: "Gender differences in how scientists present the importance of their research: observational study", M. Lechenmüller, O. Sorenson and A. Jena, 2019

*Containing at least one of 25 selected words

The Economist

atsevišķi nepalīdz

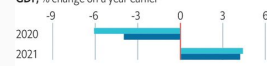
Taken by surprise

United States, economists' forecasts

Average forecast

Forecast made in: Jun 2020 Sep 2020

GDP, % change on a year earlier



Unemployment rate, %

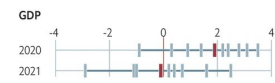


Sources: Bank of America; Barclays; Citigroup; Deutsche Bank; Federal Reserve; Goldman Sachs; IMF; JP Morgan Chase; Morgan Stanley; OECD; UBS; Wells Fargo; The Economist

The Economist

Revisions to forecasts, where available

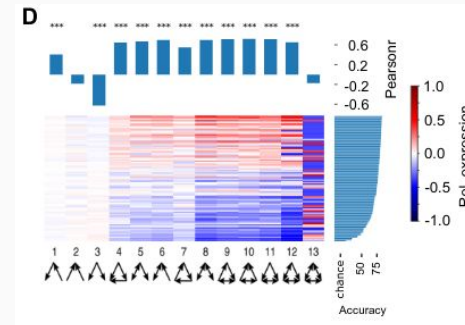
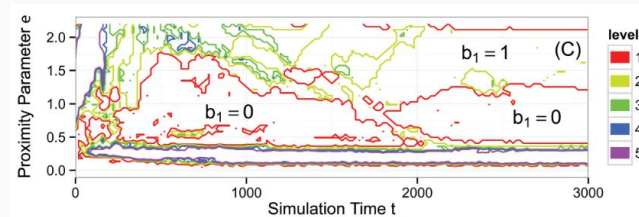
Jun-Sep 2020, percentage points



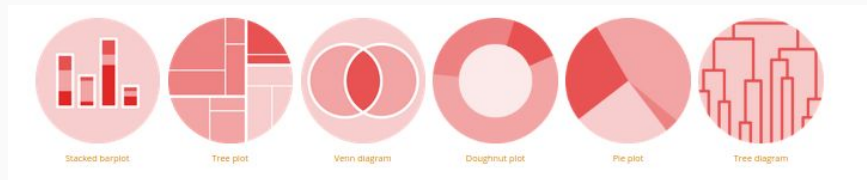
Unemployment rate



akademiski piemēri: jābūt priekšzināšanām, lai novērtētu



1. Abstraktais / Iespējamais / Praktiskais
2. Krāsošana / Savienošana



Avoti:

- The Economist Graphic Detail blog economist.com/graphic-detail
- FiveThirtyEight fivethirtyeight.com
- Adobe Color Wheel color.adobe.com/create/color-wheel
- The Python Graph Gallery python-graph-gallery.com
- PGFPlots Gallery pgfplots.sourceforge.net/gallery.html
- TikZ and PGF examples texample.net/tikz/examples
- Lisa Charlotte Rost github.com/lisacharlotterost/talk-slides
- James Gleick “The Information: A History, A Theory, A Flood”
- Topaz et al “Topological Data Analysis of Biological Aggregation Models”
- Reimann et al “Topology of synaptic connectivity constrains neuronal stimulus representation”