

```
1 # encoding: utf-8
2 import sys
3 reload(sys)
4 sys.setdefaultencoding('utf8')
```

```
5
6 import csv
7 import datetime
8 from math import floor
```

```
9
10 # Ordering
11 output_keys = ['id', 'item', 'item_sub', 'item_name', 'item_name_english', 'material', 'designer', 'year']
```

```
12
13
14 # For creating value counts
15 multiple_value_fields = ['materials', 'material_categories', 'material_subcategories', 'material_subcategories_without_unspecified_values']
16 counts = {}
17 for k in output_keys:
18     counts[k] = { 'all_values': [], '(blank)': 0 }
19 for k in multiple_value_fields:
20     counts[k] = { 'all_values': [], '(blank)': 0 }
```

```
21
22 def increment(key, value):
23     if value == '':
24         value = '(blank)'
25     if value not in counts[key]['all_values']:
26         counts[key][value] = 1
27         counts[key]['all_values'].append(value)
28     else:
29         counts[key][value] += 1
```

```
30
31 # Read the data in
32 rawdata = []
33 for row in open('ESINE-tsv-csv.tsv', 'rU'):
34     rawdata.append(row.split('\t'))
35
36 keys = rawdata[0]
37 for i, key in enumerate(keys):
38     keys[i] = key.strip()
```

```
39
40 # Create dicts based on the tabular data
41 data = []
```

# Information design tools & data sources

Jonatan Hildén  
@jhilde  
@koponenhilden

KOPONEN + HILDÉN





“Save yourself some time: Sketch the rough visual shape to catch thinking errors. Don’t just go straight from idea to computer. If you can’t get it to work on a piece of paper, it definitely is not going to work on a computer.”

This is a perspective view of 3-layers of graphite

### "Weighing" atoms with electrons

All materials are made of atoms. However, fast electrons sometimes remove atoms from the material. It happens on average with fewer electrons for lighter atoms.

The "lead" of a pencil is actually called graphite, made of carbon atoms.

$^{12}\text{C}$  has 6 protons and 6 neutrons,  $^{13}\text{C}$  6 protons & 7 neutrons. Both have 6 electrons around the nucleus.

Carbon has two stable isotopes, for which only difference is one neutron more in the core of  $^{13}\text{C}$  as compared to  $^{12}\text{C}$ .

Although they can be seen by electron microscopy,  $^{12}\text{C}$  and  $^{13}\text{C}$  look identical.

Should be actual microscopy image

This is a cross section of a metal cylinder. Detector looks like a CD. The electron gun in this illustration is a sharp metal tip.

Until now, all C can be grey or black, here we could color  $^{12}\text{C}$  &  $^{13}\text{C}$  different. Electron beam could be shown looking like a laser beam or similar.

## "Weighing" atoms with electrons

universität wien  
Faculty of Physics

All materials are made up of atoms. The "lead" in a pencil is actually graphite, a material made of stacked sheets of carbon atoms. A single sheet is called graphene.

A carbon atom has six protons and six electrons, and comes in two stable variants called isotopes:  $^{12}\text{C}$  and  $^{13}\text{C}$ . Their only difference is one more neutron in the core of  $^{13}\text{C}$ .

Although atoms in graphene can be "seen" by shooting electrons through the material in what is called transmission electron microscopy, different isotopes such as  $^{12}\text{C}$  and  $^{13}\text{C}$  appear identical.

The lighter the atom, the fewer electrons are on average needed to create a vacancy.

However, the electrons may sometimes remove atoms from the material.

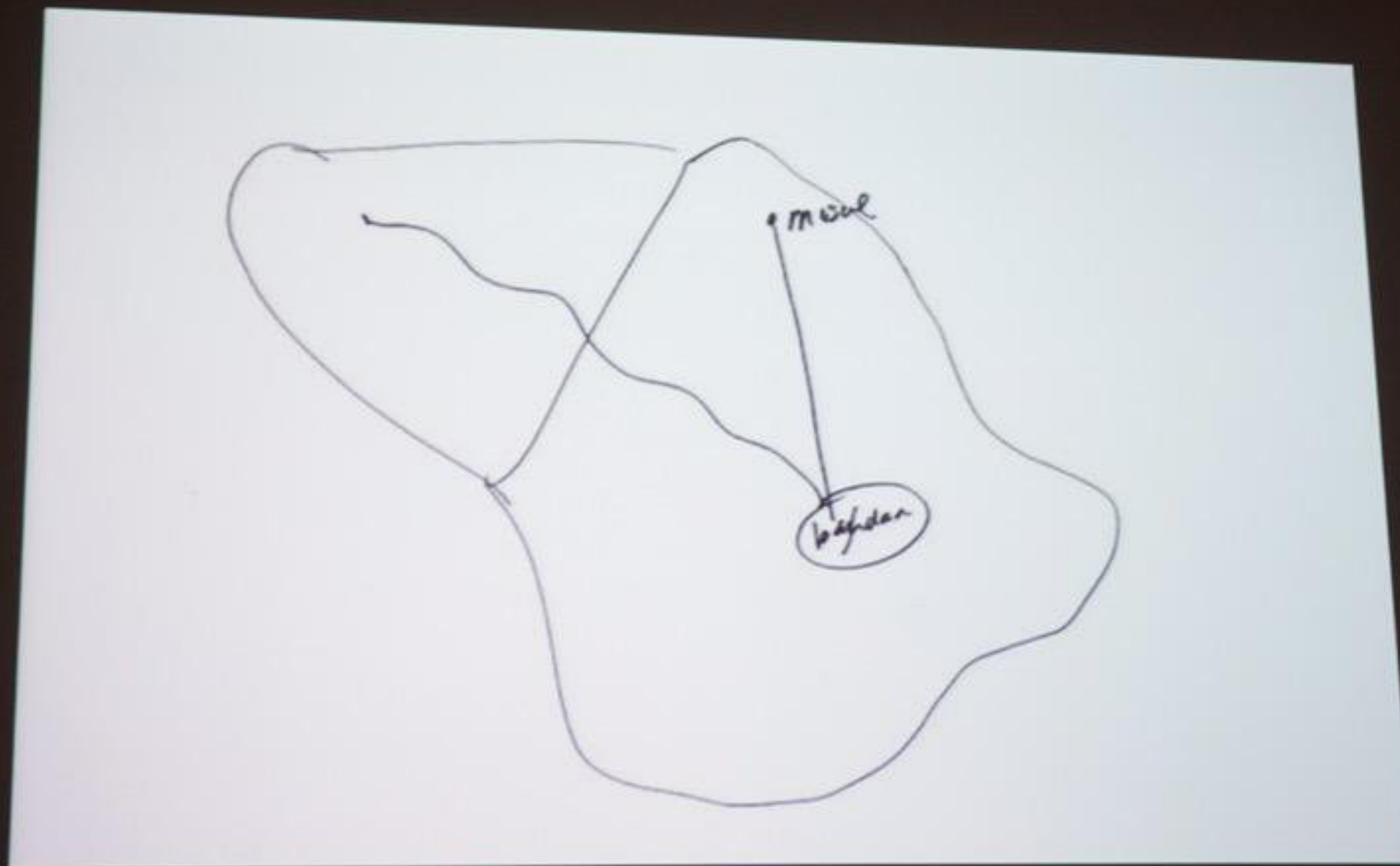
Measuring isotopes can help understand and improve the synthesis of materials.

DISORDER + MATERIALS  
T. Susi, C. Hofer, G. Argentero, G.T. Leuthner, T.J. Pennycook, C. Mangler, J.C. Meyer & J. Kotakoski. Isotope analysis in the transmission electron microscope. Nature Communications 7:13040 (2016) DOI: 10.1038/ncomms13040

Graphics: Koponen+Hildén  
Creative Commons BY 4.0

“Everybody is capable  
of sketching.”

— Michael Brenner



# A Rogue State Along Two Rivers

## How ISIS Came to Control Large Portions of Syria and Iraq

By JEREMY ASHKENAS, ARCHIE TSE, DEREK WATKINS and KAREN YOURISH July 3, 2014

The militant group called the Islamic State in Iraq and Syria, or ISIS, seemed to surprise many American and Iraqi officials with the recent gains it made in its violent campaign to create a new religious state. But the rapid-fire victories achieved over a few weeks in June were built on months of maneuvering along the Tigris and Euphrates Rivers.

### The Euphrates

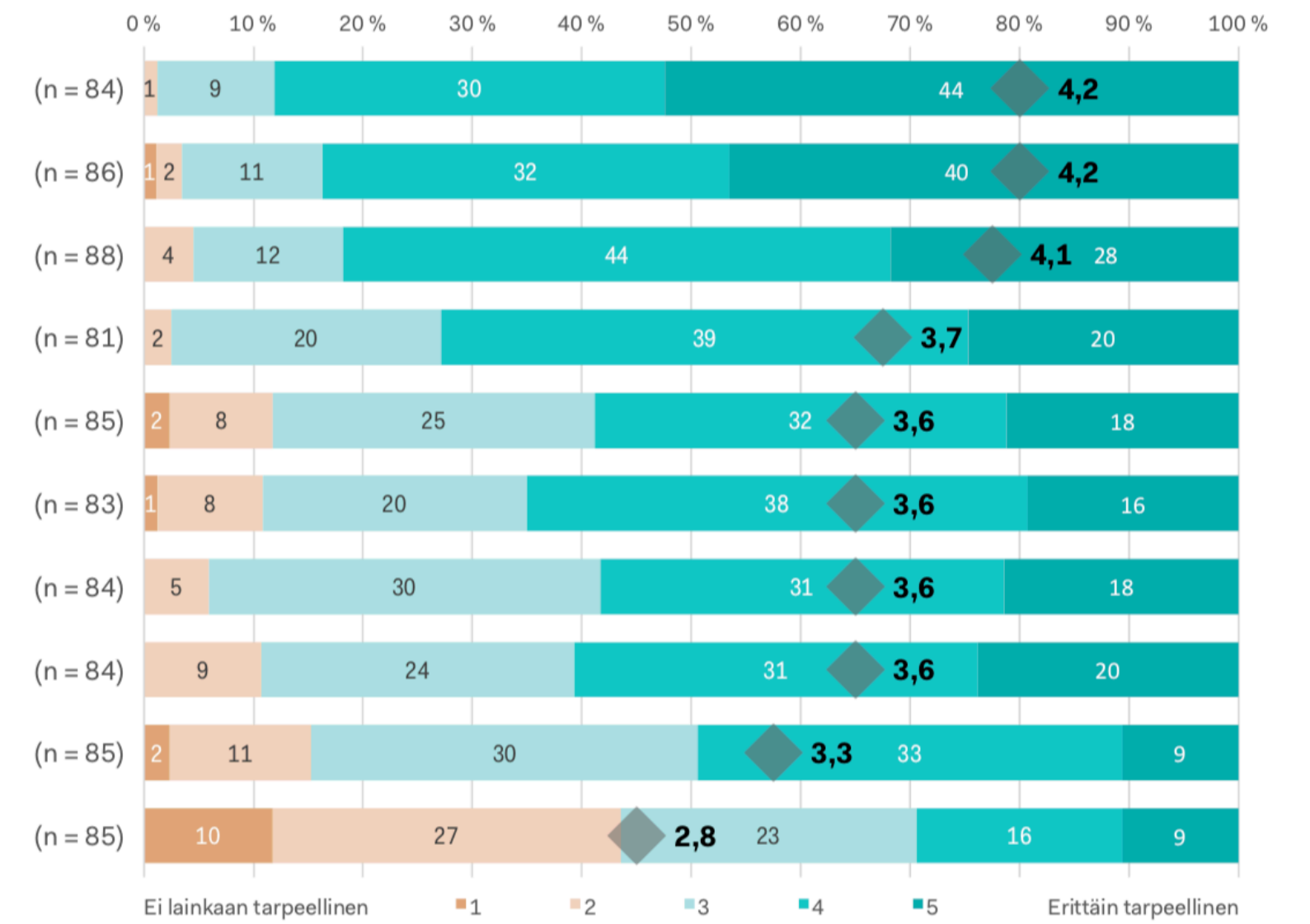


# Desktop tools



**Excel and Numbers:** Basic spreadsheet programs are surprisingly useful for creating basic visualizations. (Especially Excel.)

COMMERCIAL



**Kuntien terveydenhuollon menot (milj. €)**



**Kuntien sosiaaliturvan menot (milj. €)**



\*) Ahvenanmaalla maakunta vastaa terveydenhuollon järjestämisestä

Lähde: Tilastokeskus

**Adobe Illustrator** is the de facto standard tool for visual designers—useful for finishing charts made in Excel or other programs, and for more complex custom graphics.

**Affinity Designer** is more affordable

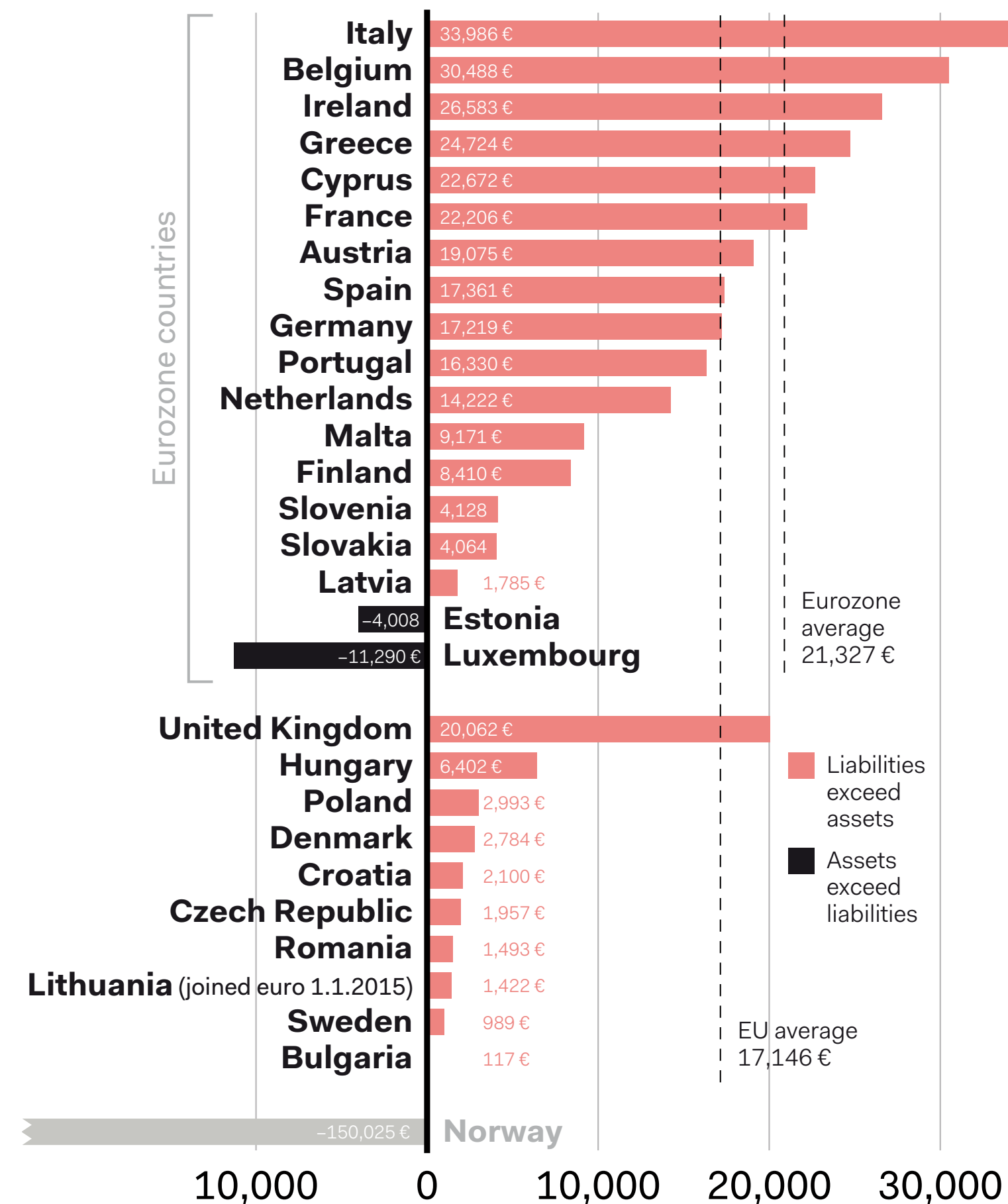
COMMERCIAL

**Inkscape** is the open source alternative.

KOPONEN + HILDÉN

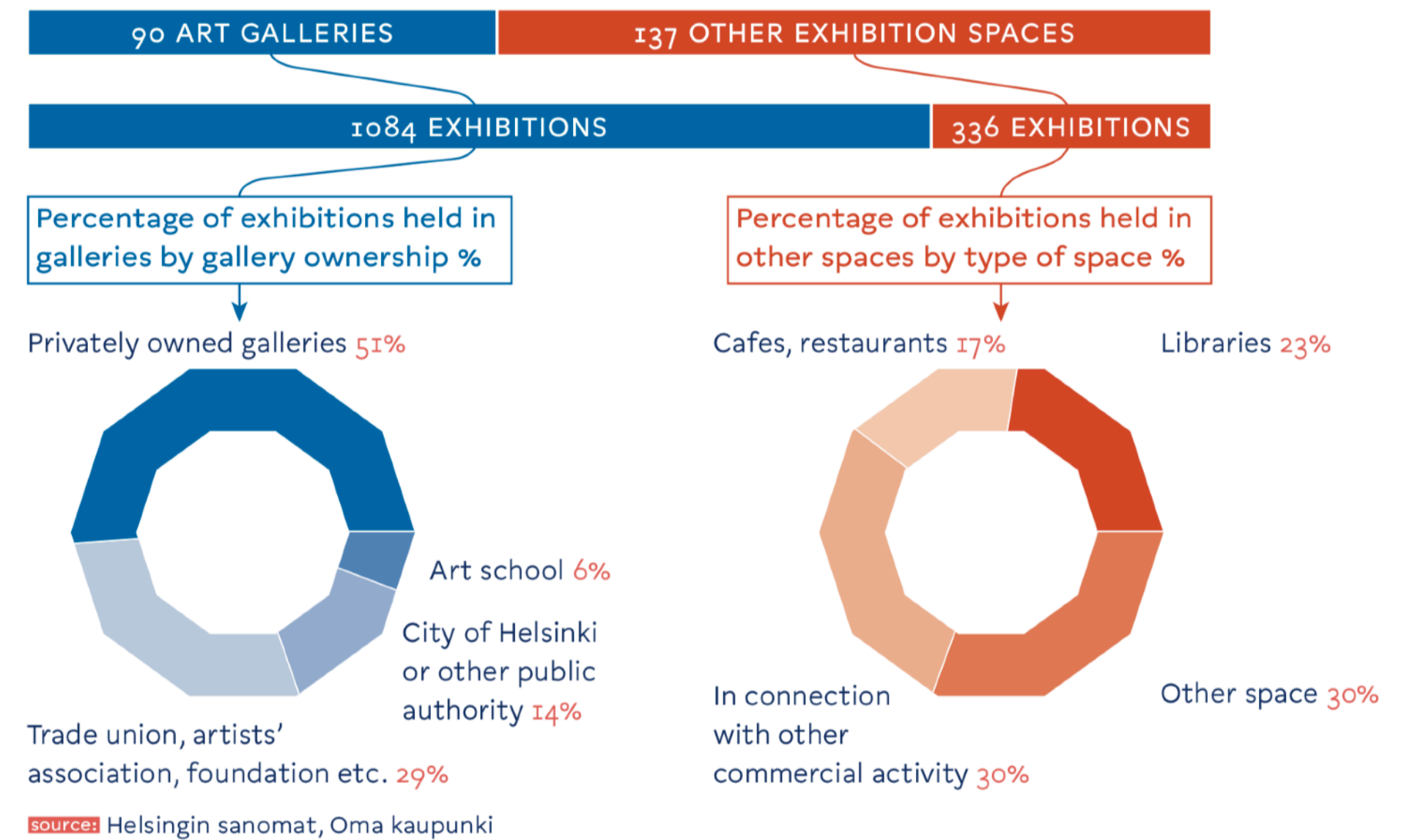
## Public sector net debt per capita in EU countries, Q1/2014

(liabilities minus assets, excluding social security funds)



By: Juuso Koponen/Koponen+Hildén Creative Commons BY 4.0  
Source: Eurostat (gov\_q\_ggfa, demo\_gind)

## Art galleries and other exhibition spaces in Helsinki 2013



[https://www.adobe.com/Creative\\_Cloud](https://www.adobe.com/Creative_Cloud)

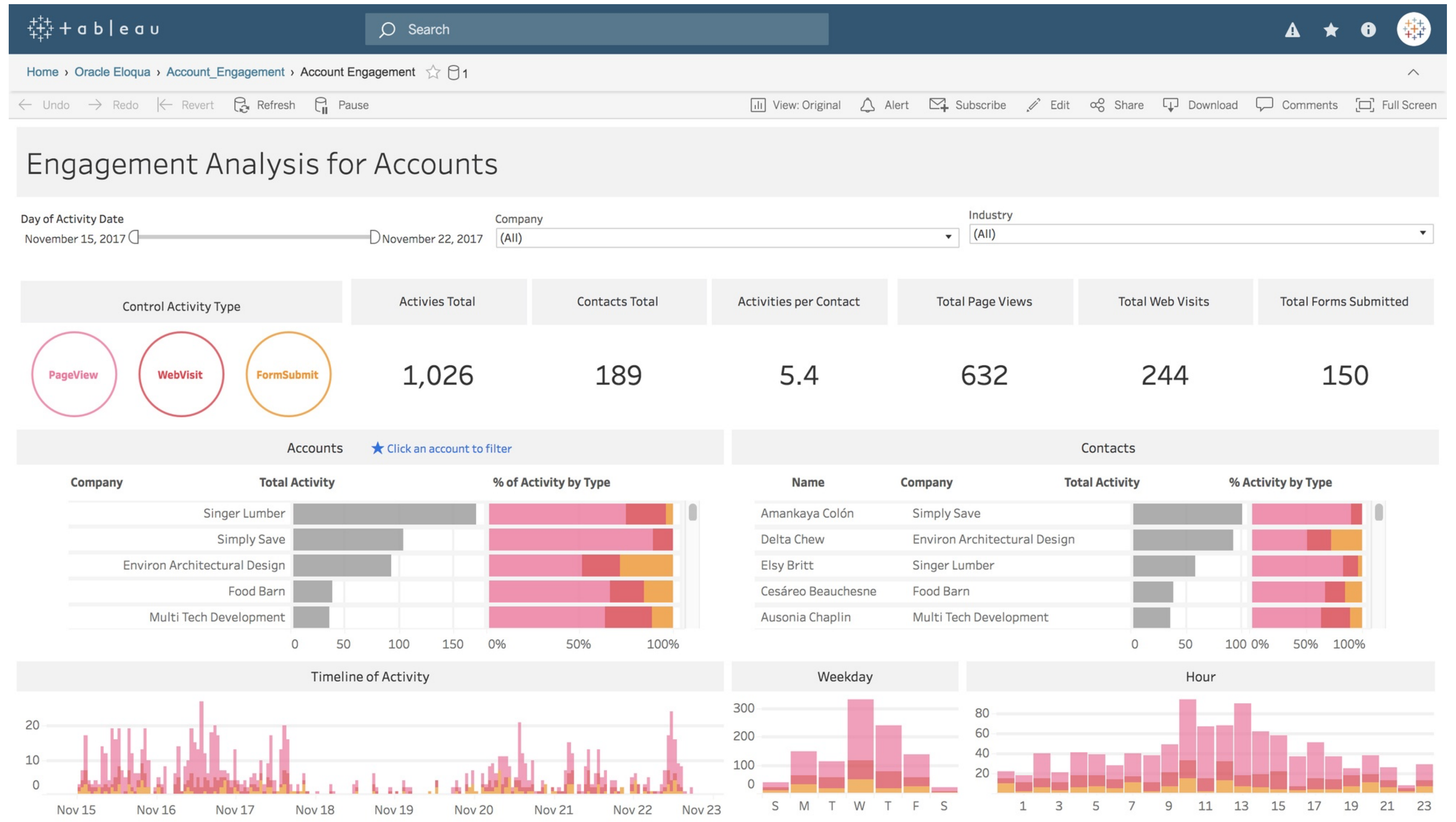
<https://affinity.serif.com/en-gb/> <https://inkscape.org>

# Tableau, Qlik, Power BI + iNZight: Business intelligence software that include good visualization tools.

**Tableau:** COMMERCIAL, FREE FOR ACADEMIC USERS

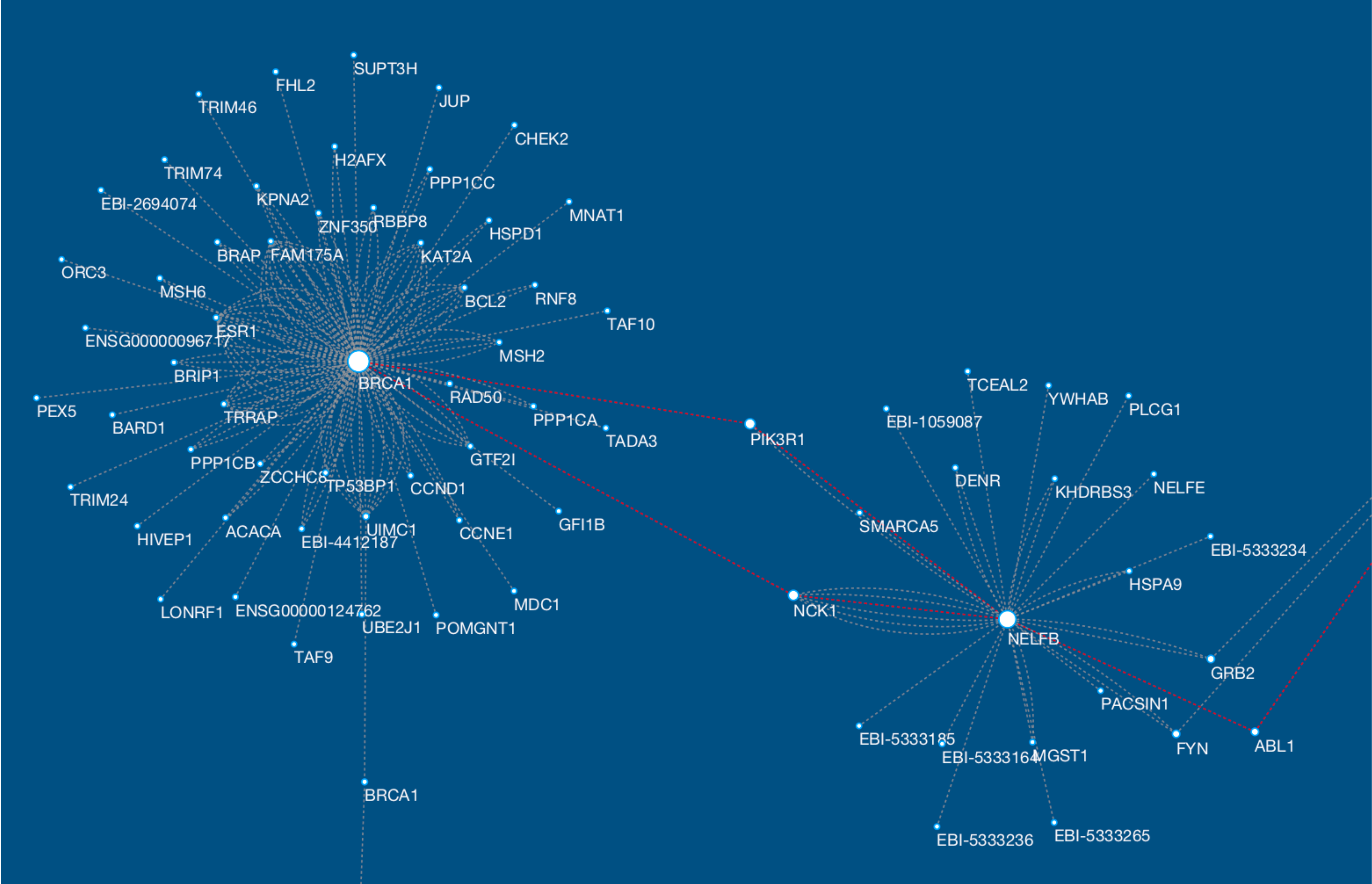
**Qlik & Power BI:** COMMERCIAL

**iNZight:** OPEN SOURCE



# Cytoscape and Gephi: For network visualization.

OPEN SOURCE



# THE CHARTMAKER DIRECTORY

ABOUT



☰ dot plot ×

Reference Type: ○ Example ● Solution | Chart Families: ● Categorical ● Hierarchical ● Relational ● Temporal ● Spatial

	Kibana	Mapbox	Matplotlib	Microsoft Excel	Microsoft Power BI	Microsoft PowerPoint	Microstrategy	Panorama
Connected dot plot								
Dot plot								

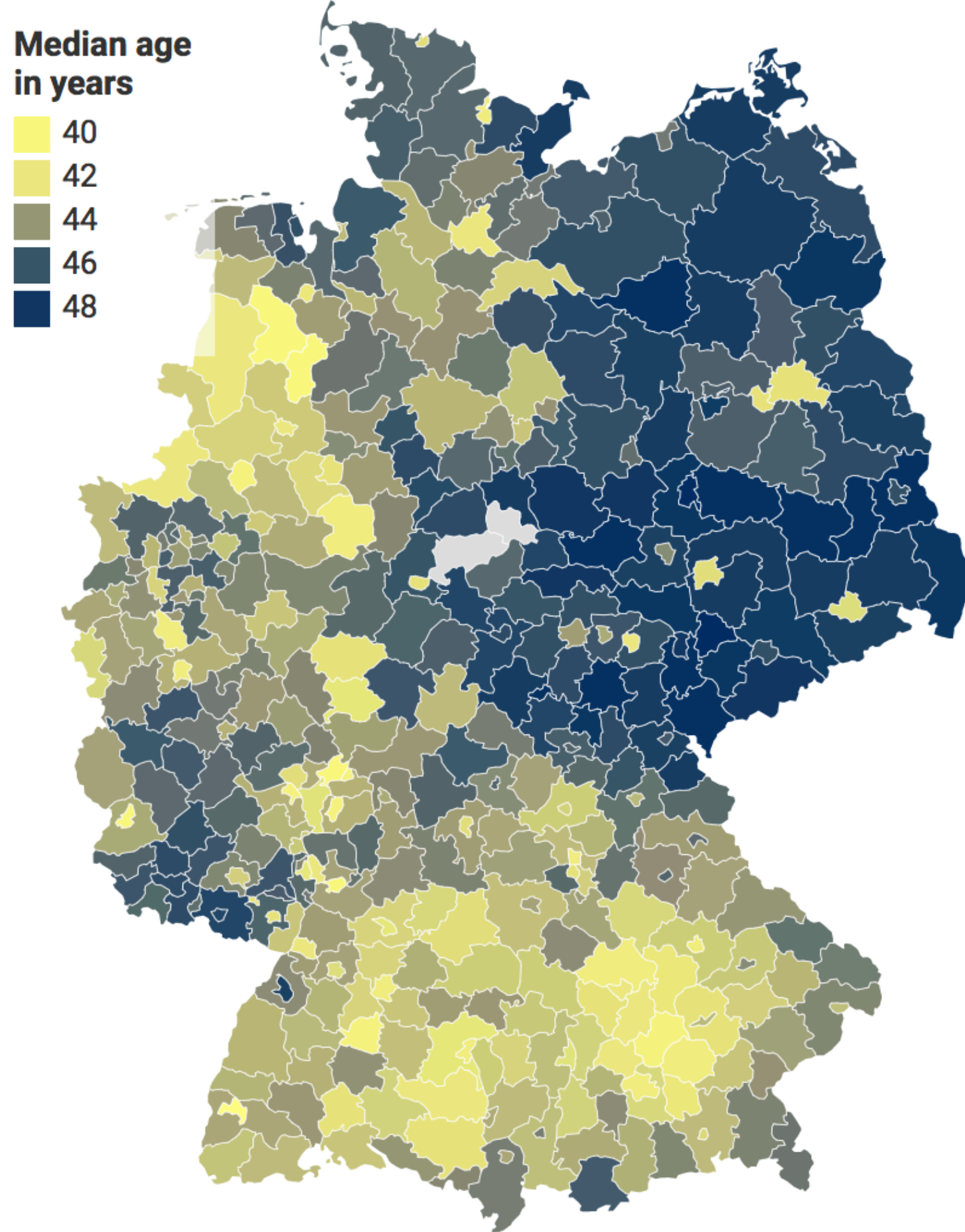
# Browser-based tools

# Datawrapper: An easy-to-use cloud-based tool for creating charts and maps.

COMMERCIAL, FREE FOR  
PUBLIC USE

## Young Germans live in the West and in cities

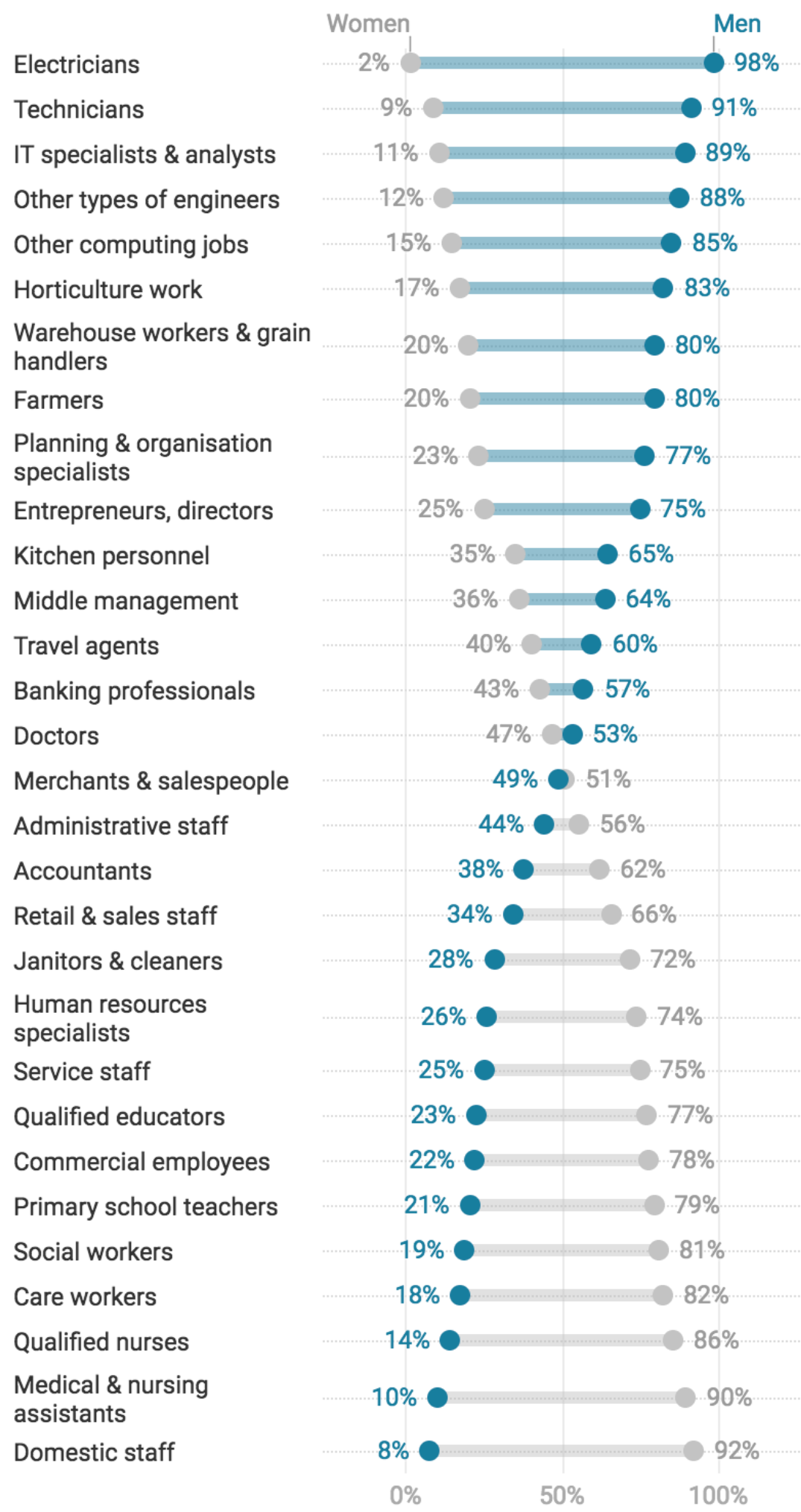
Median age in all German districts, 2015



Based on [Lisa Charlotte Rost, Datawrapper](#) • Source: [BBSR study, 2015](#) • Map data: © GeoBasis-DE / BKG 2013 • [Get the data](#) • Created with Datawrapper

## Occupations by gender

The 30 most common jobs in Switzerland in 2016 and proportion of men and women in each

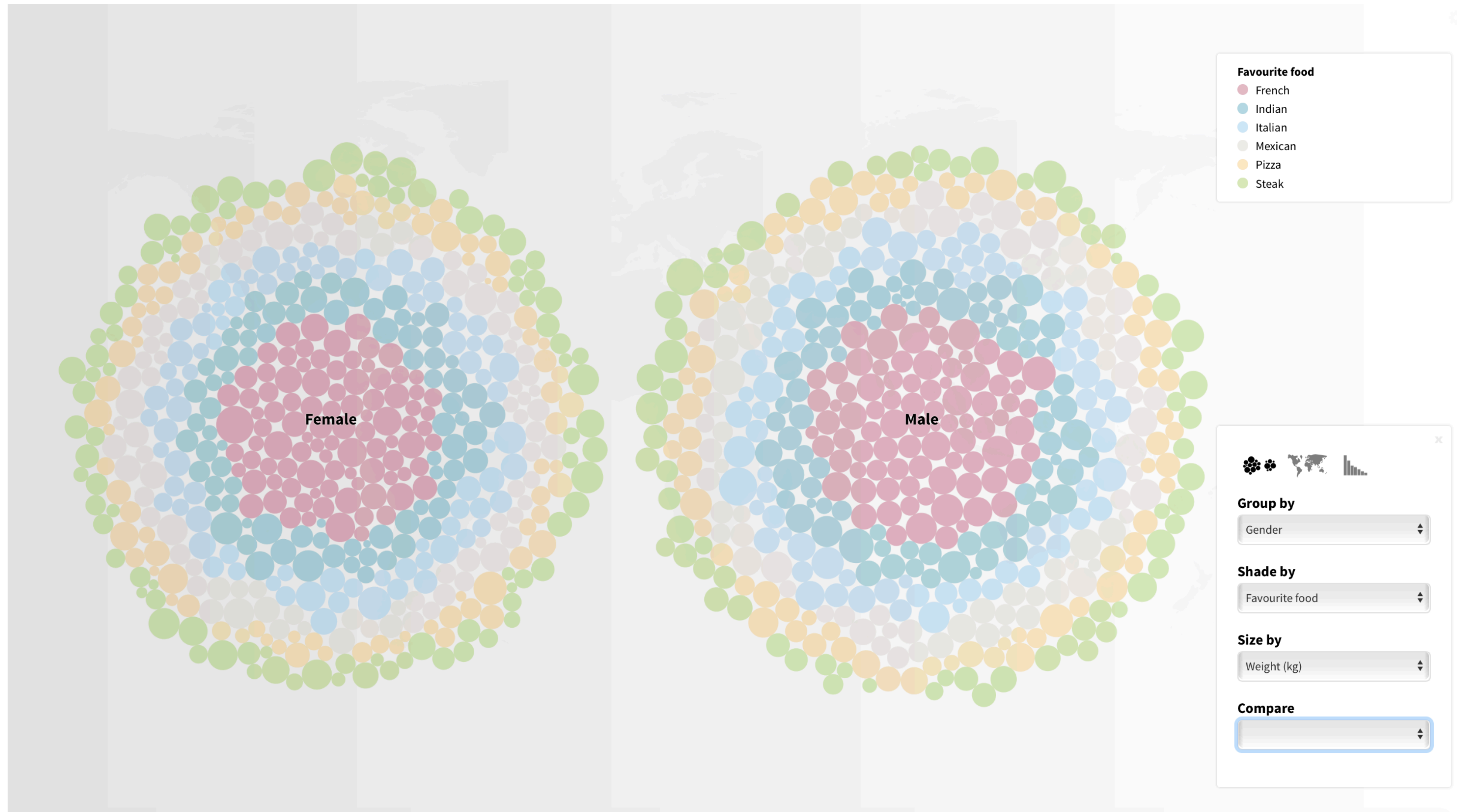


Based on [on](#) • Source: [Federal Statistical Office](#) • [Get the data](#) • Created with Datawrapper



# Flourish: A promising WYSIWYG tool for creating visualizations.

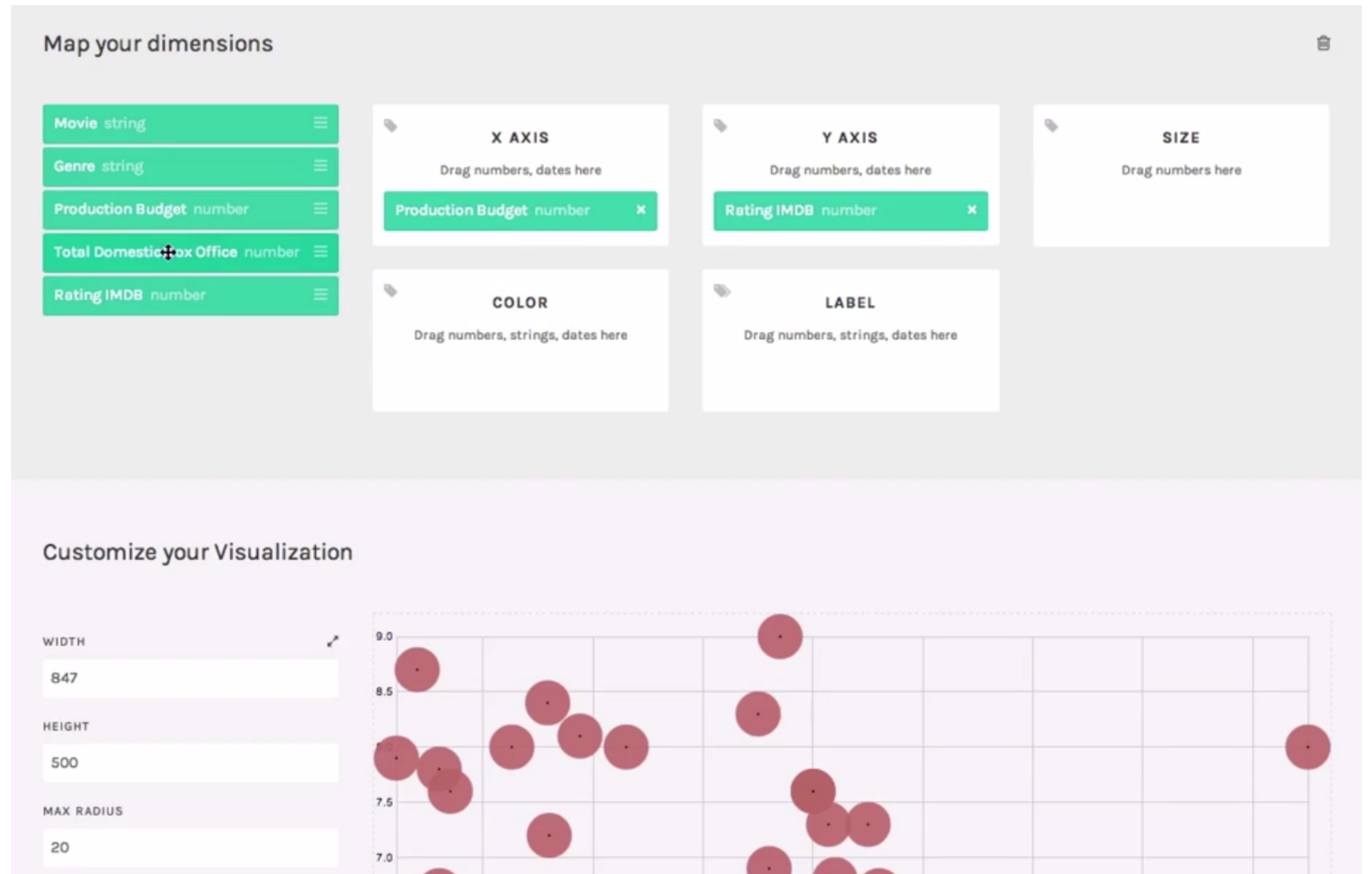
COMMERCIAL, FREE FOR PUBLIC USE, STUDENTS AND JOURNALISTS





**Raw:** A browser based tool for quick prototyping and data exploration.

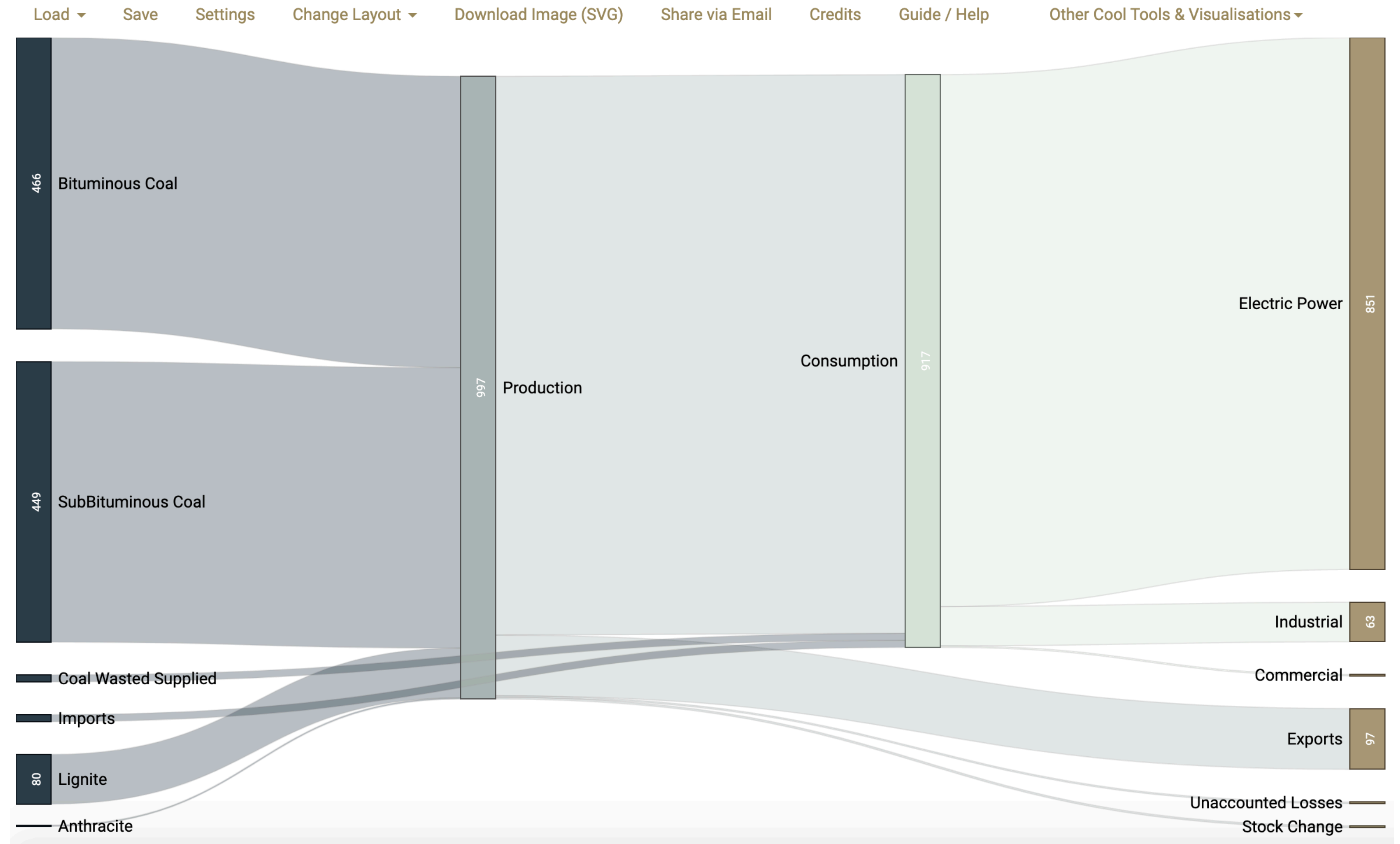
FREE



# The Sankey Diagram Generator, SankeyMATIC

Simple browser-based tools for creating sankey diagrams.

FREE



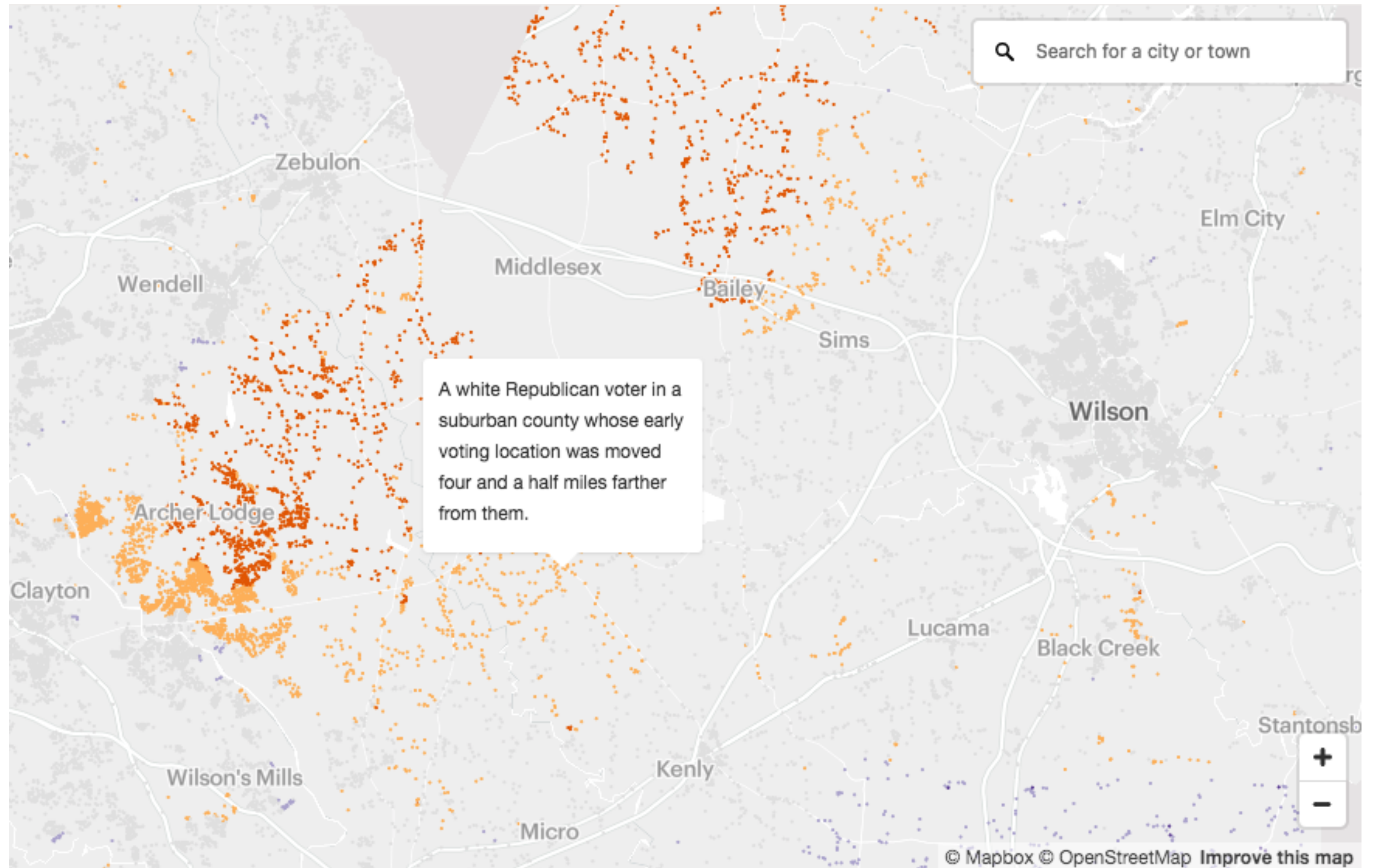
# Mapping

# Leaflet and Mapbox: For online mapping.

**Leaflet:** OPEN SOURCE

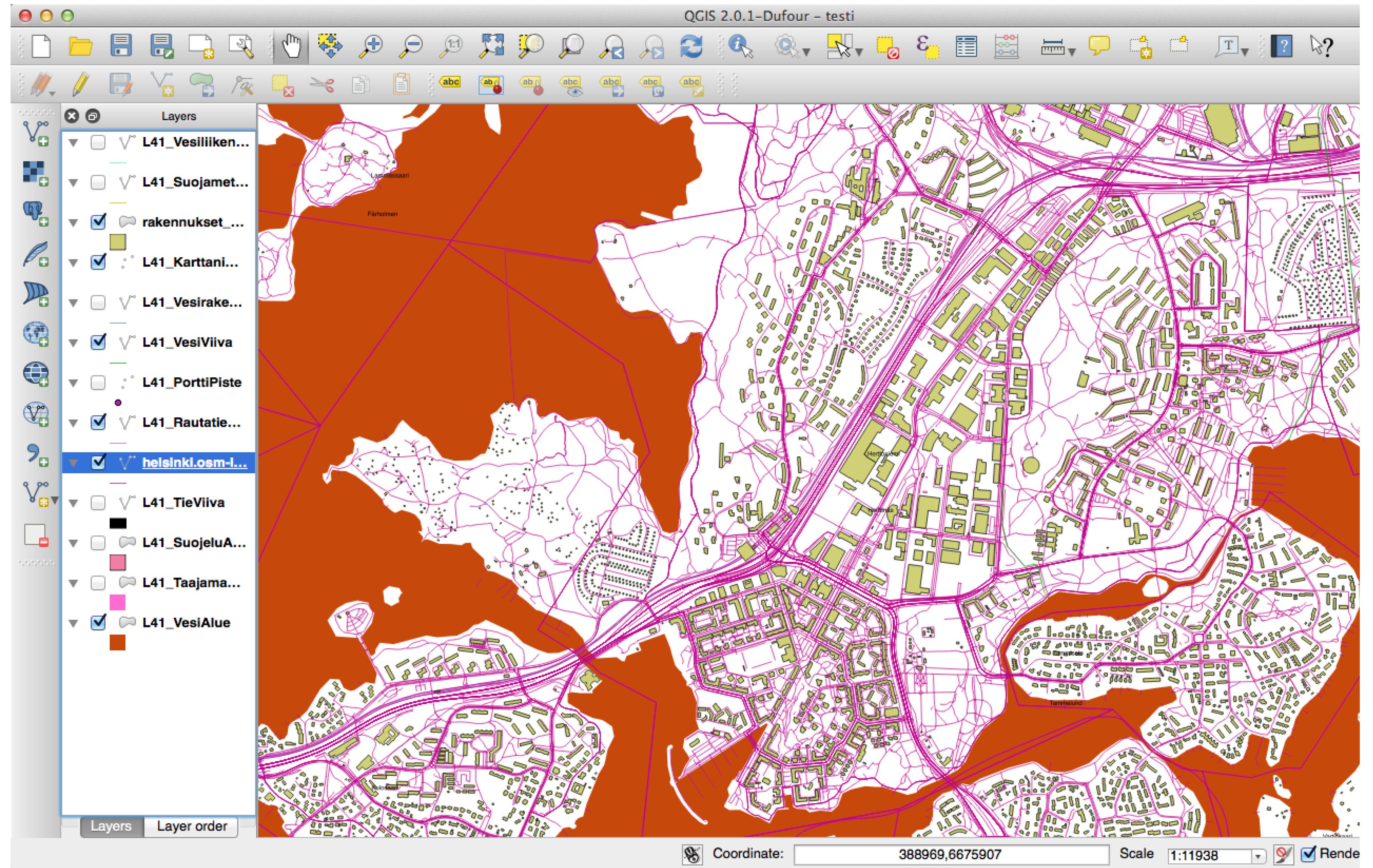
**Mapbox:** COMMERCIAL, FREE VERSION AVAILABLE

**Google Maps** also has extensive customization options nowadays:  
[mapstyle.withgoogle.com](https://www.google.com/maps/heatmap/data/)



**QGIS:** Open-source GIS and spatial analysis/ visualization software. Steep(ish) learning curve, but many features and plug-ins.

OPEN SOURCE



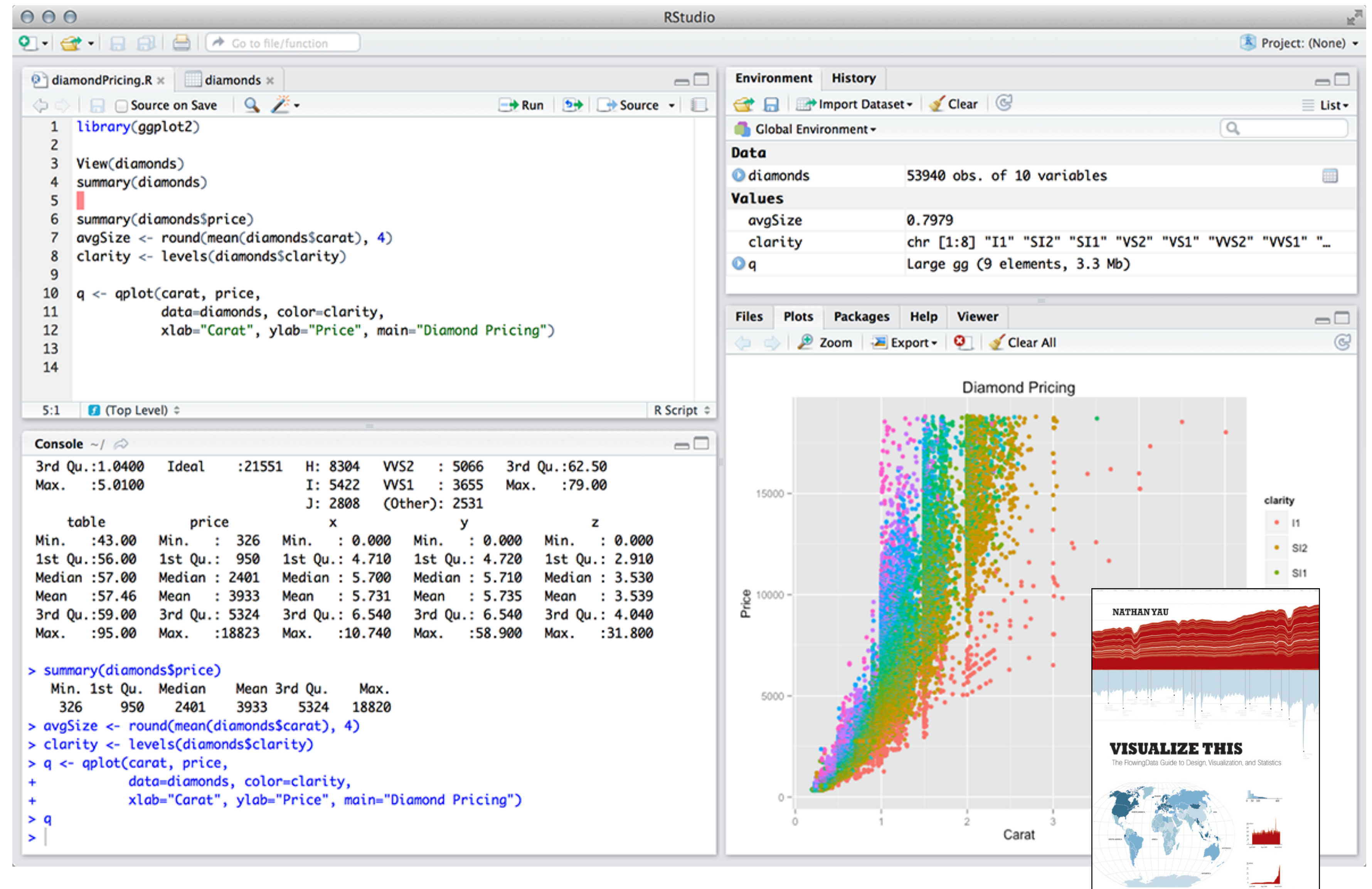
# Code-based tools

**R:** A statistical analysis & visualization environment.

**RStudio:** an integrated developing environment for R.

**R:** OPEN SOURCE

**RStudio:** COMMERCIAL, FREE VERSION AVAILABLE



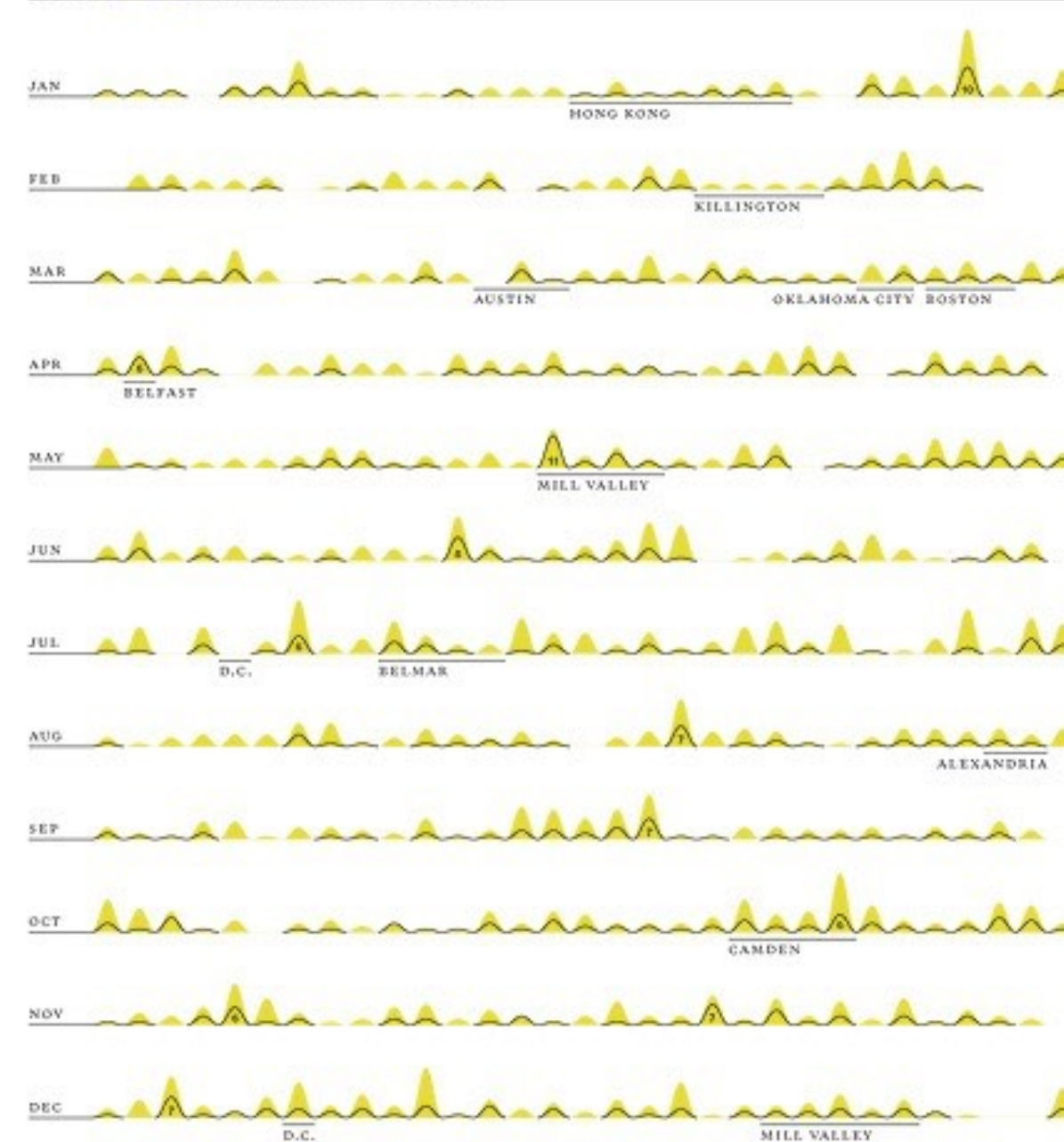
**Processing and P5.js:**  
 Environments for artistic/design programming.  
 Large community and lots of learning resources. P5.js is the web version of Processing, based on Javascript.

OPEN SOURCE

# Distribution

Date and location of encounters.

FIGURE 1. ENCOUNTERS / RESPONSES



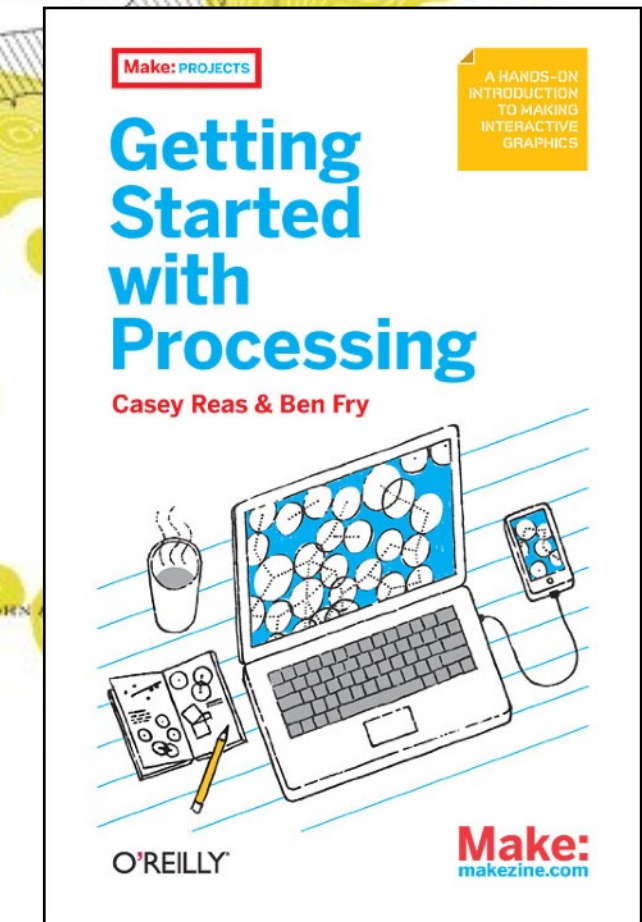
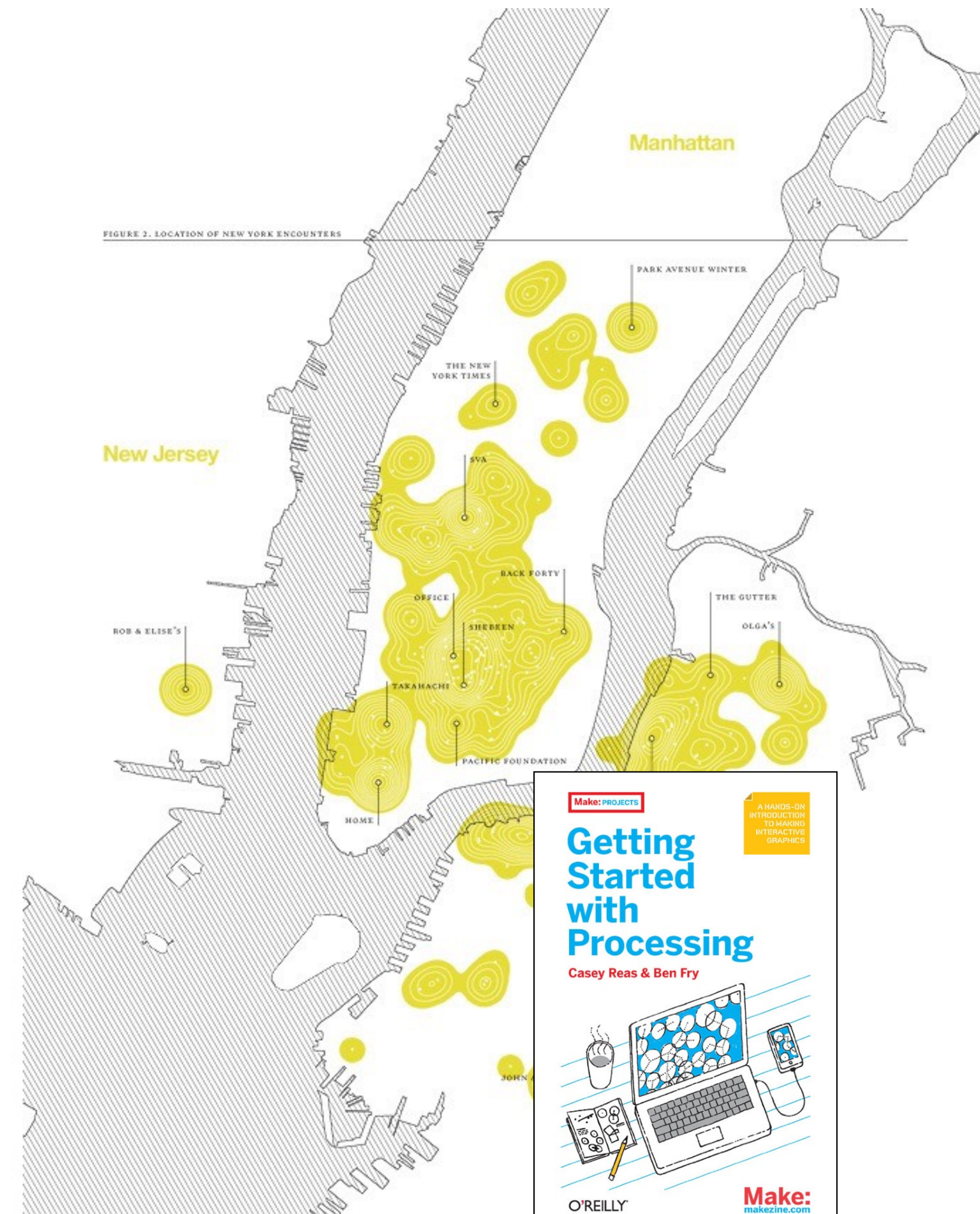
TOTAL ENCOUNTERS	AVERAGE ENCOUNTERS PER DAY	SURVEYS COMPLETED	CUMULATIVE RESPONSE RATE
<b>1,761</b>	<b>4.8</b>	<b>560</b>	<b>32%</b>
COUNTRIES INCLUDED	STATES INCLUDED	DAYS WITH REPORTS	CONTRIBUTORS
<b>Three</b>	<b>Nine</b>	<b>254</b>	<b>210</b>
<small>U.S.A., HONG KONG AND NORTHERN IRELAND</small>	<small>CALIFORNIA, MAINE, MASSACHUSETTS, NEW JERSEY, NEW YORK, OKLAHOMA, TEXAS, VERMONT, VIRGINIA, PLUS WASHINGTON D.C.</small>	<small>70% OF THE YEAR</small>	<small>AVERAGE 2.66 REPORTS PER PERSON</small>

METHODOLOGY

Throughout 2009, friends, family, co-workers and acquaintances of Nicholas Felton were asked to report on his activities whenever they met.

All data on the following pages was compiled from the responses of these participants to a variety of questions concerning their encounter.

FIGURE 2. LOCATION OF NEW YORK ENCOUNTERS



<https://processing.org>    <http://p5js.org/>  
 Tools for data visualization: <http://www.gicentre.net/utills>



# Paper.js: Besides P5.js, another good starting point for learning Javascript is Paper.js.

OPEN SOURCE

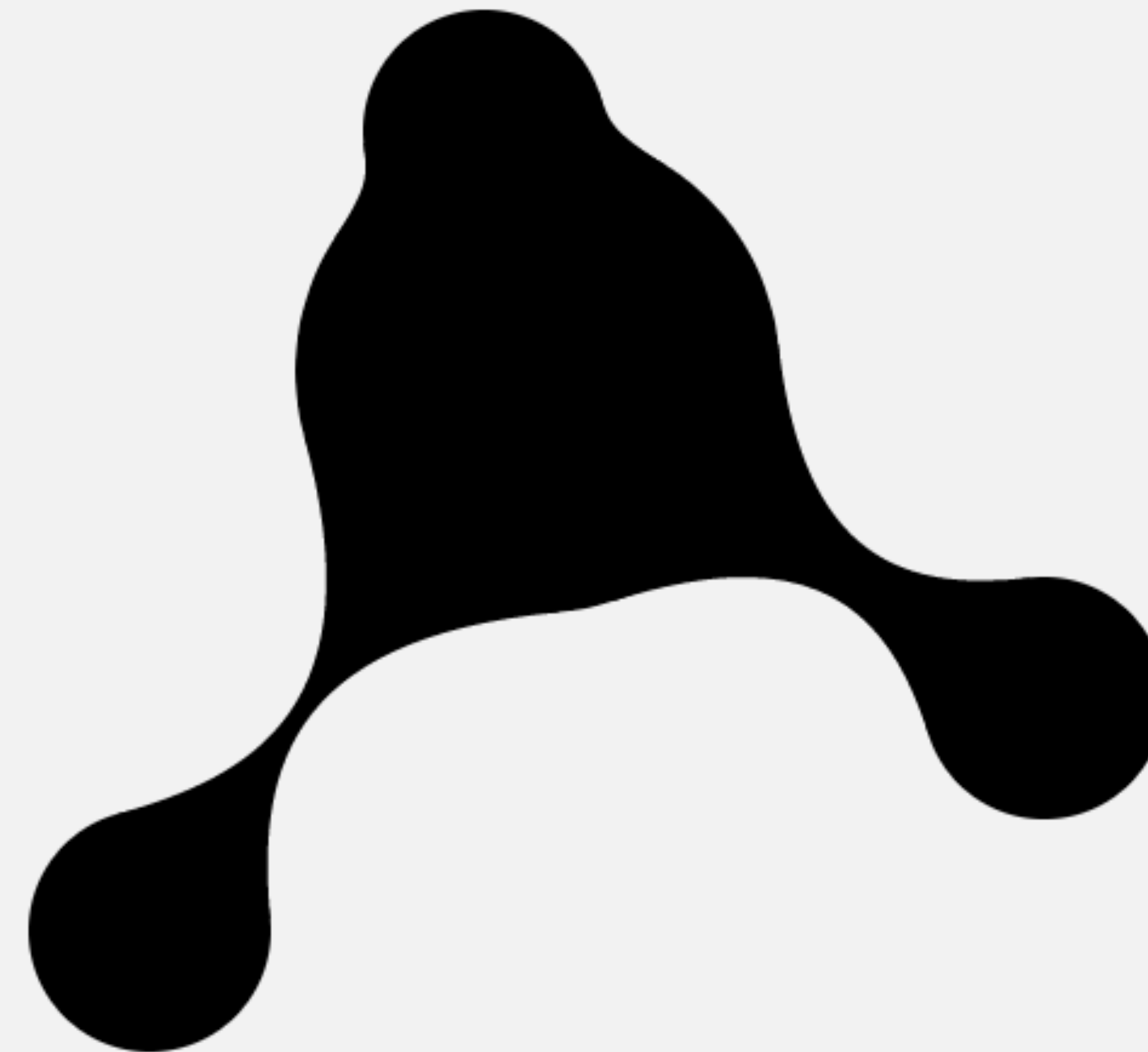
## Paper.js

About  
Features

Examples

[Satie Liked To Draw Chain](#)  
[Tadpoles](#)  
[Nyan Rainbow](#)  
[Rounded Rectangles](#)  
[Radial Rainbows](#)  
[Meta Balls](#)  
[Voronoi](#)  
[Future Splash](#)  
[Smoothing](#)  
[Spiral Raster](#)  
[Division Raster](#)  
[Q-bertify](#)  
[Path Intersections](#)  
[Path Simplification](#)  
[Hit Testing](#)  
[Bouncing Balls](#)  
[Showcase](#)

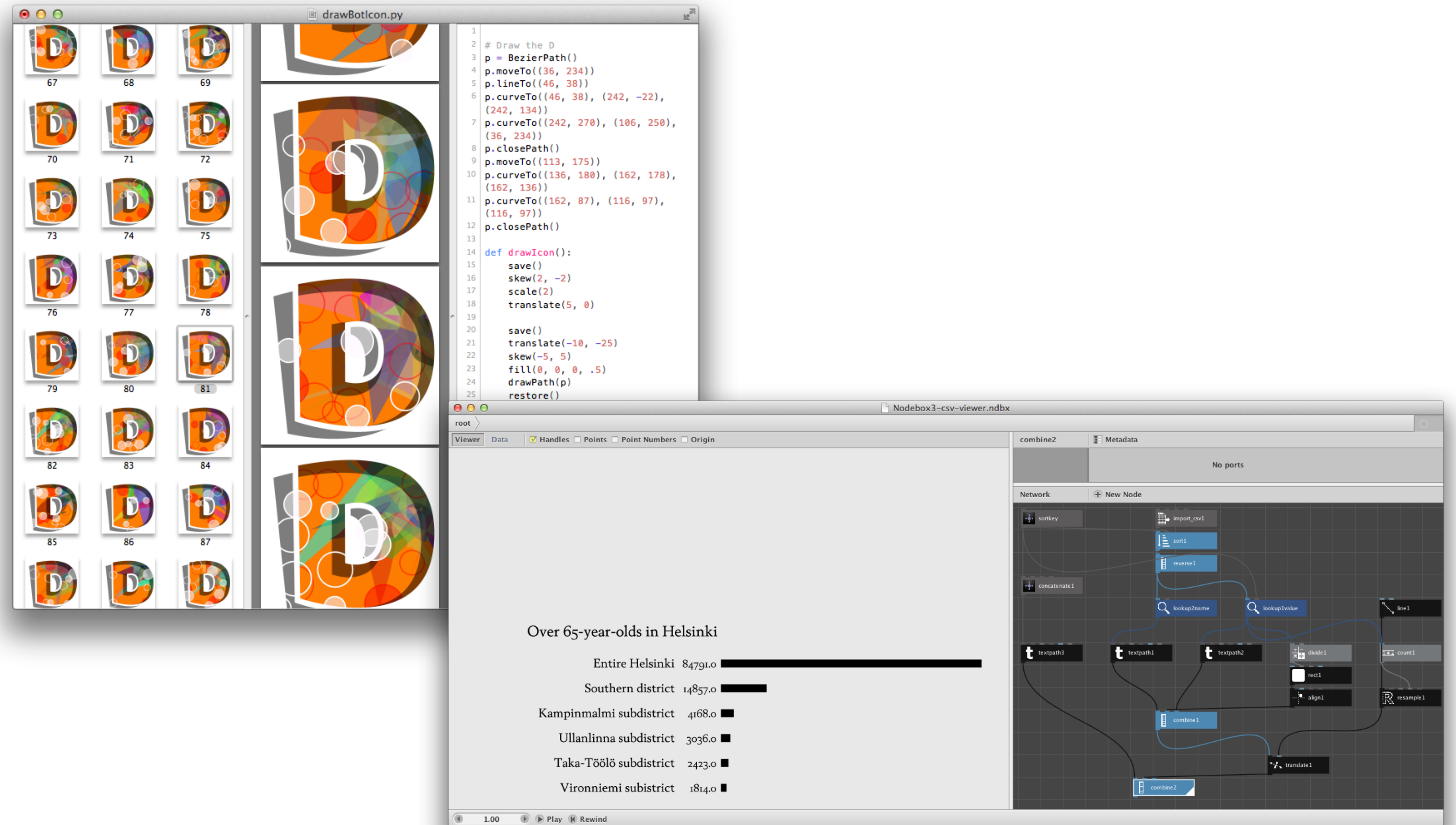
Note: You can view and even edit the source right here in the browser [Source](#)



**Drawbot and Nodebox 1/ Nodebox OpenGL:** Similar to Processing, but built on Python.

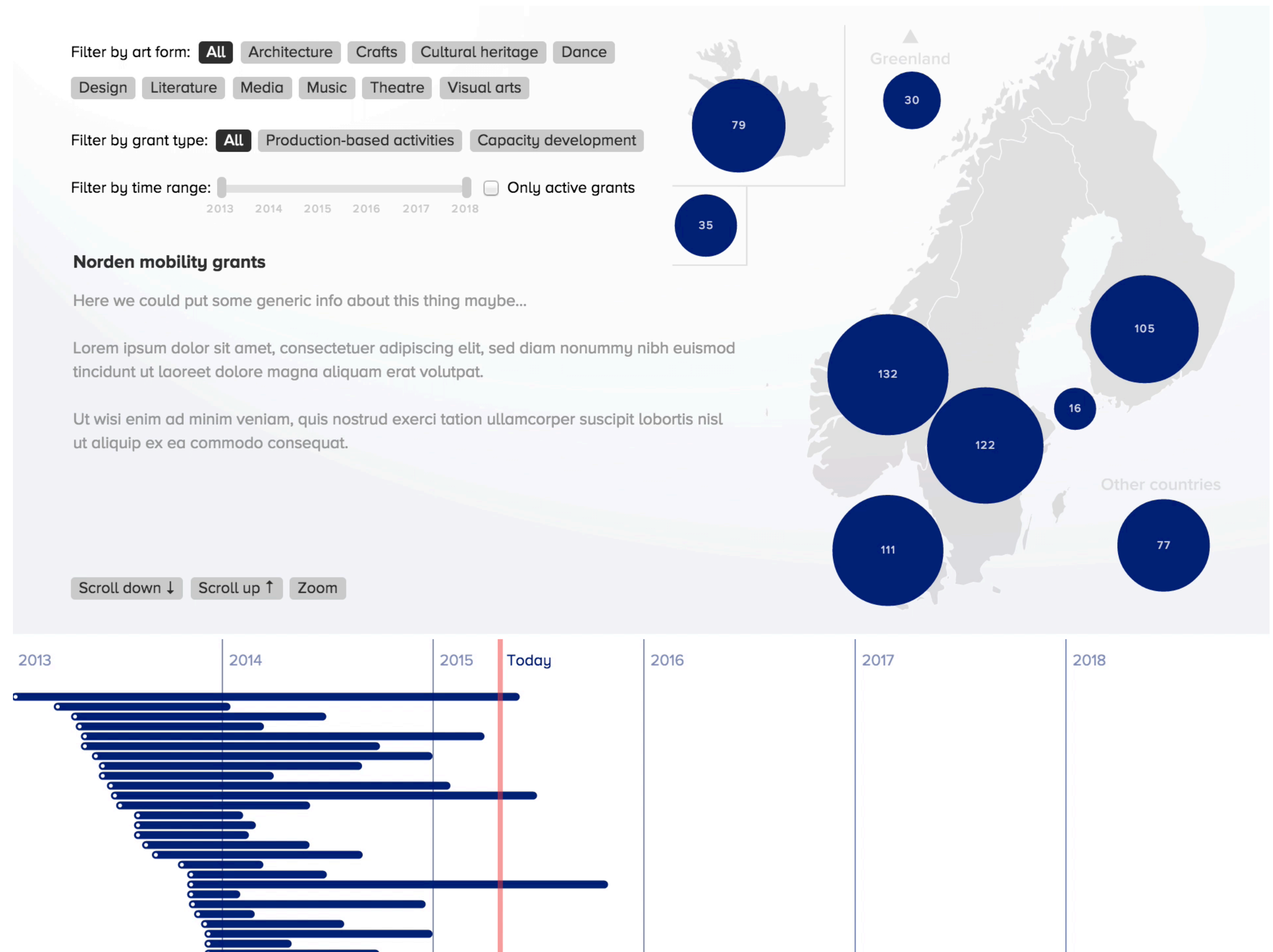
**Nodebox 3:** Uses node-based visual programming instead of directly manipulating code.

OPEN SOURCE



**D3.js:** Extremely versatile and powerful SVG based Javascript library for online visualization. Steep learning curve!  
**Observable:** online notebook interface using D3.

OPEN SOURCE



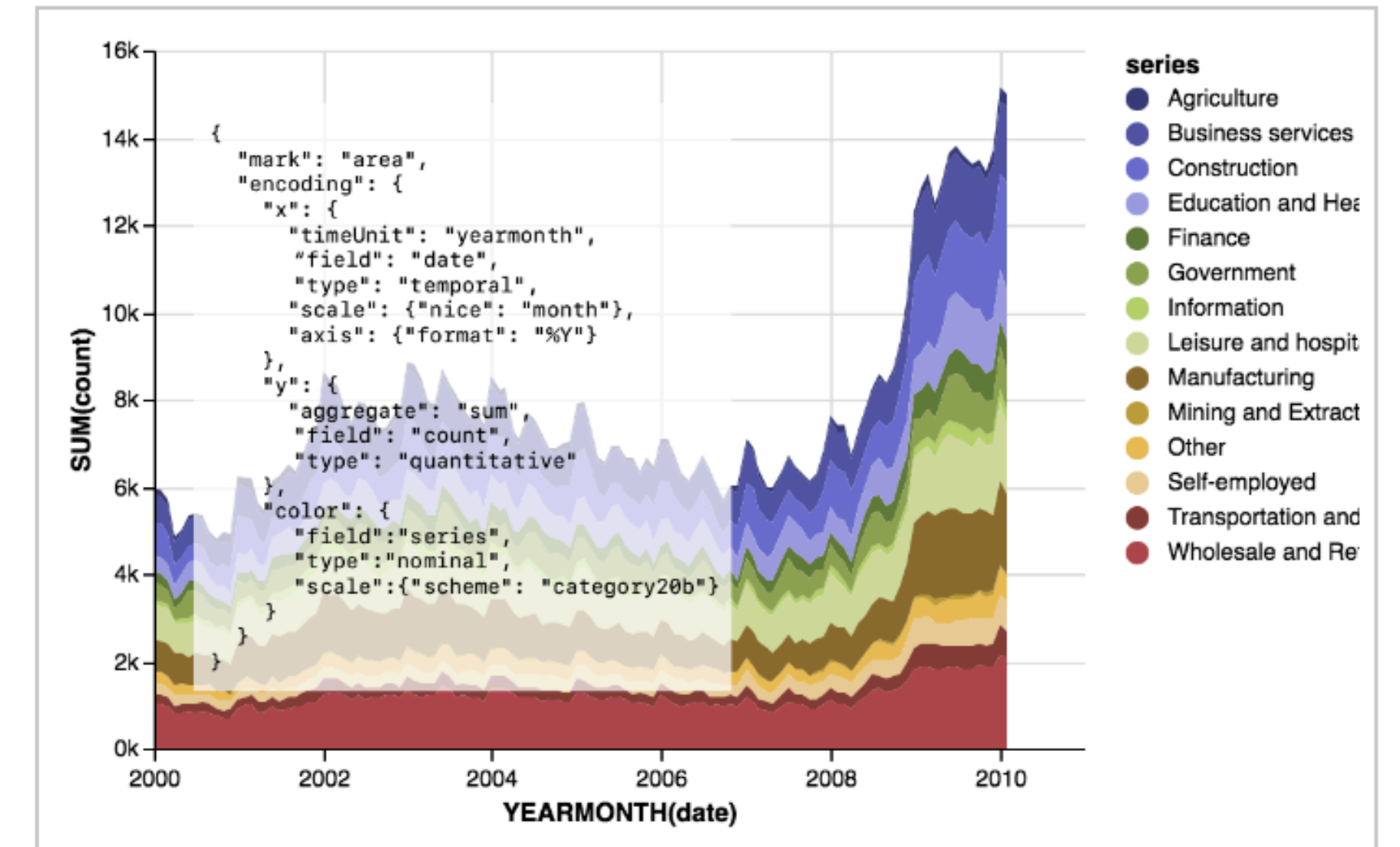
**Vega and Vega-Lite:**  
JSON based visualization grammars for D3. Make writing D3 code much simpler. The Vega suite also includes browser based tools such as Polestar.

OPEN SOURCE



**NEW** VEGA 3.0 is a visualization grammar, a declarative language for creating, saving, and sharing interactive visualization designs. With Vega, you can describe the visual appearance and interactive behavior of a visualization in a JSON format, and generate web-based views using Canvas or SVG.

[Examples](#) | [Docs & Tutorials](#) | [Porting Guide](#) | [GitHub](#)

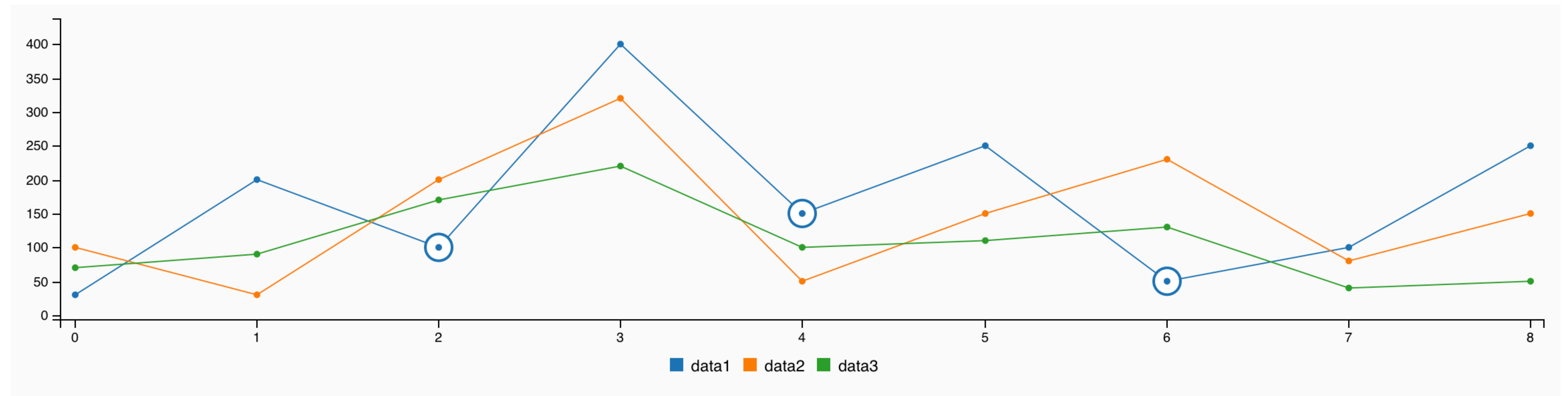
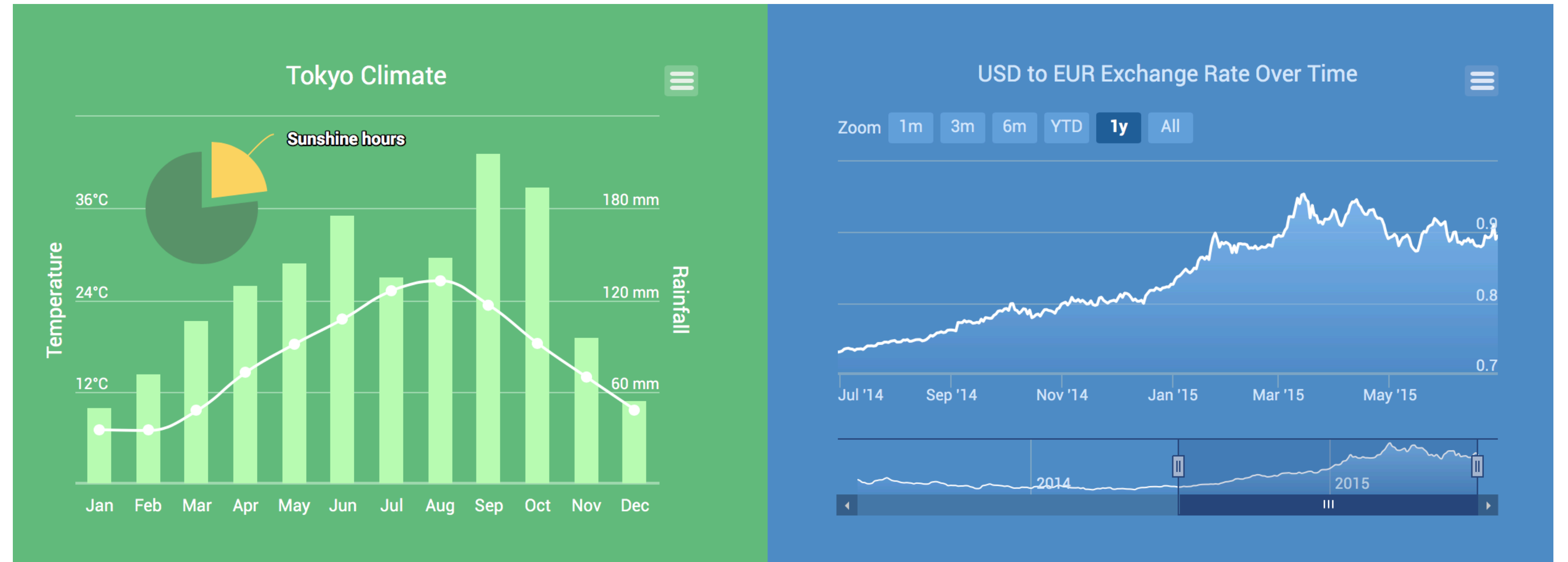


**NEW** VEGA-LITE 2.0 is a high-level visualization grammar. It provides a concise JSON syntax for supporting rapid generation of visualizations to support analysis. Vega-Lite 2 support interactive multi-view graphics. Specifications can be compiled to Vega.

[Examples](#) | [Docs & Tutorials](#) | [Online Editor](#) | [GitHub](#)

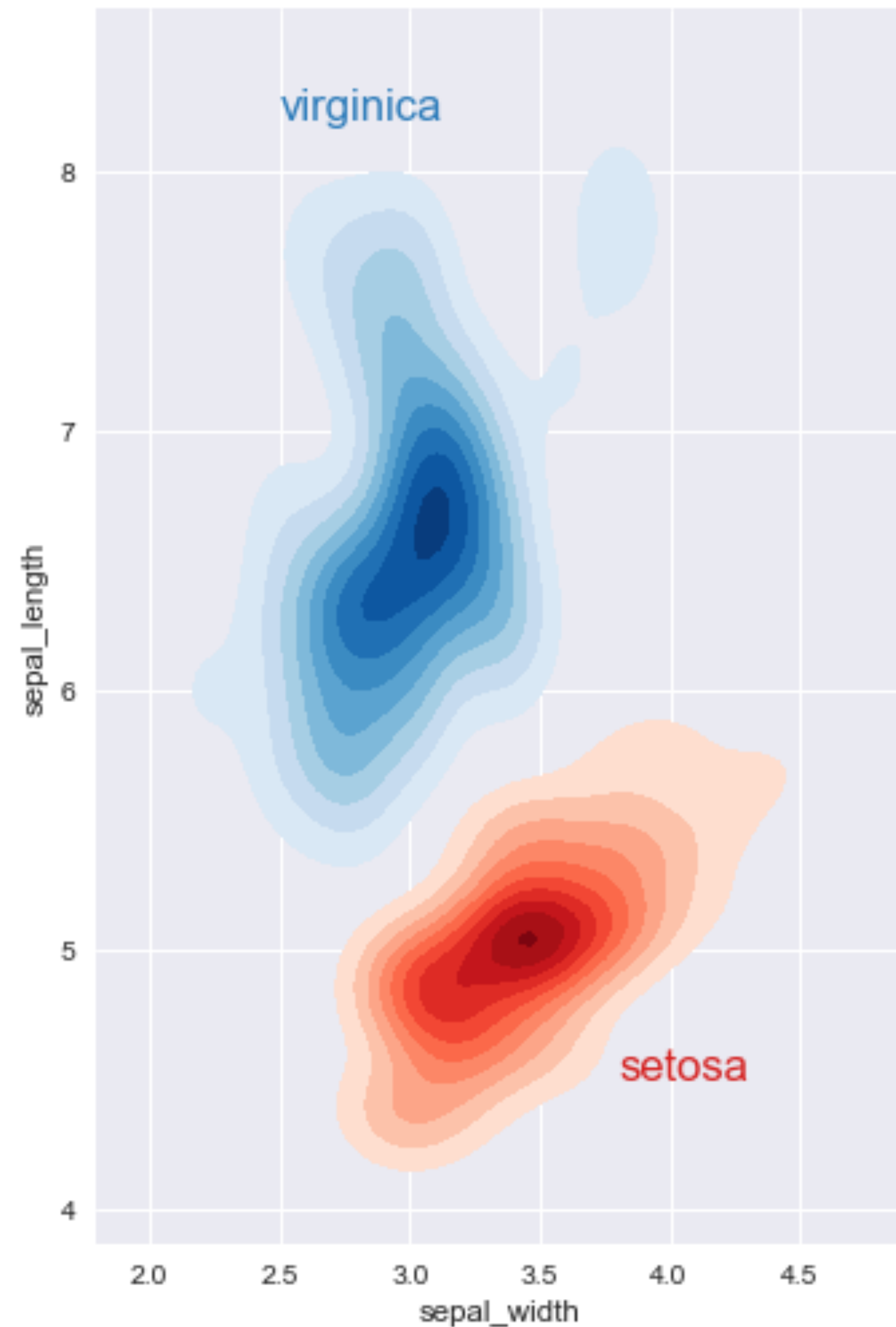
**Highcharts,  
Anychart,  
Polychart,  
C3.js...:** There are many charting libraries for Javascript make enable getting results quicker than D3 or P5 but with less customization options.

VARIES

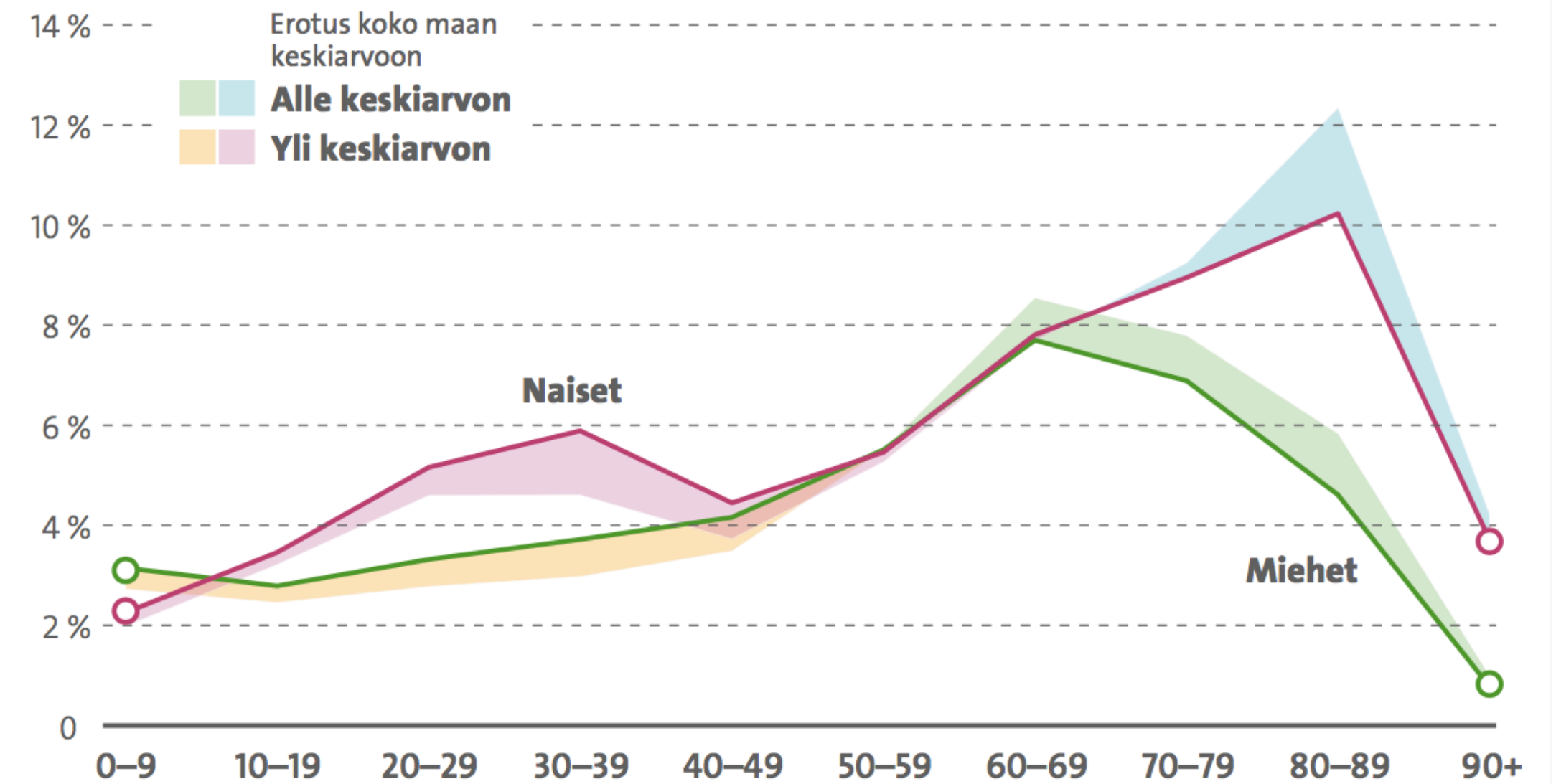


# Altair, Matplotlib, Seaborn, Bokeh, and other Python visualization libraries

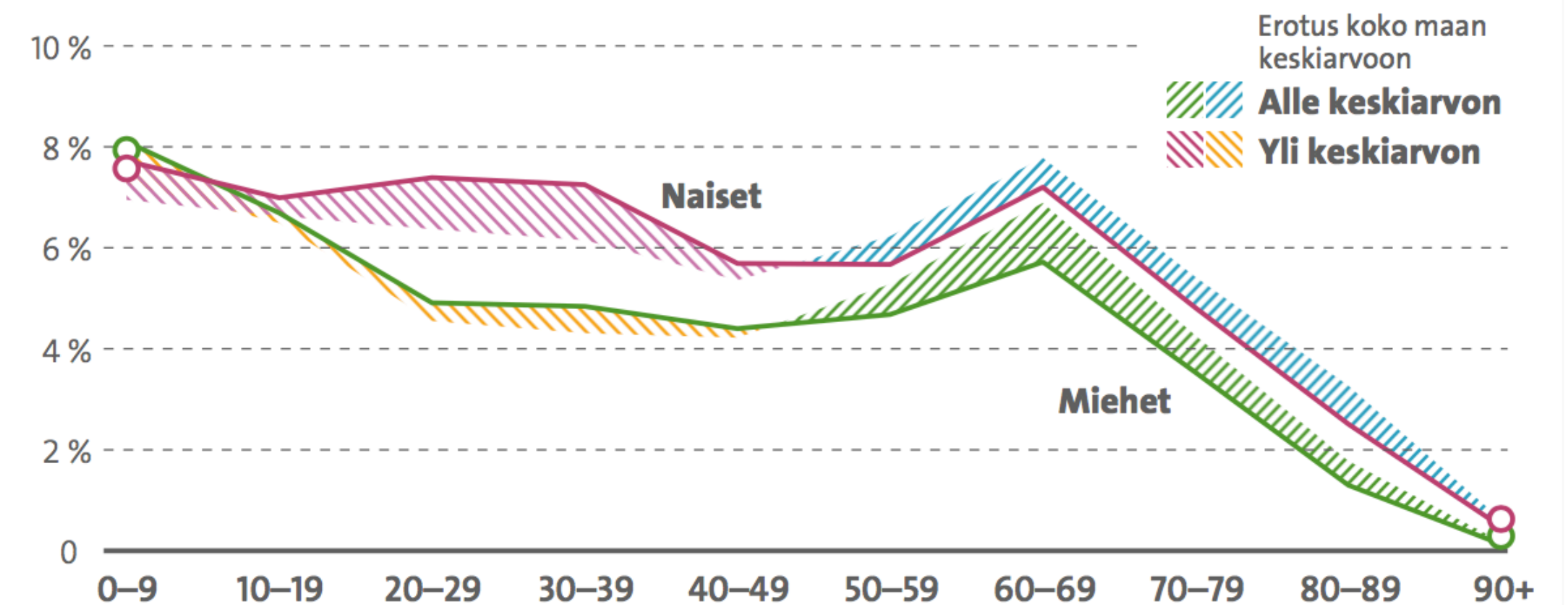
OPEN SOURCE



## Eri ikäryhmien osuus terveydenhuollon **menoista**



## Eri ikäryhmien osuus terveydenhuollon **asiakkaista**



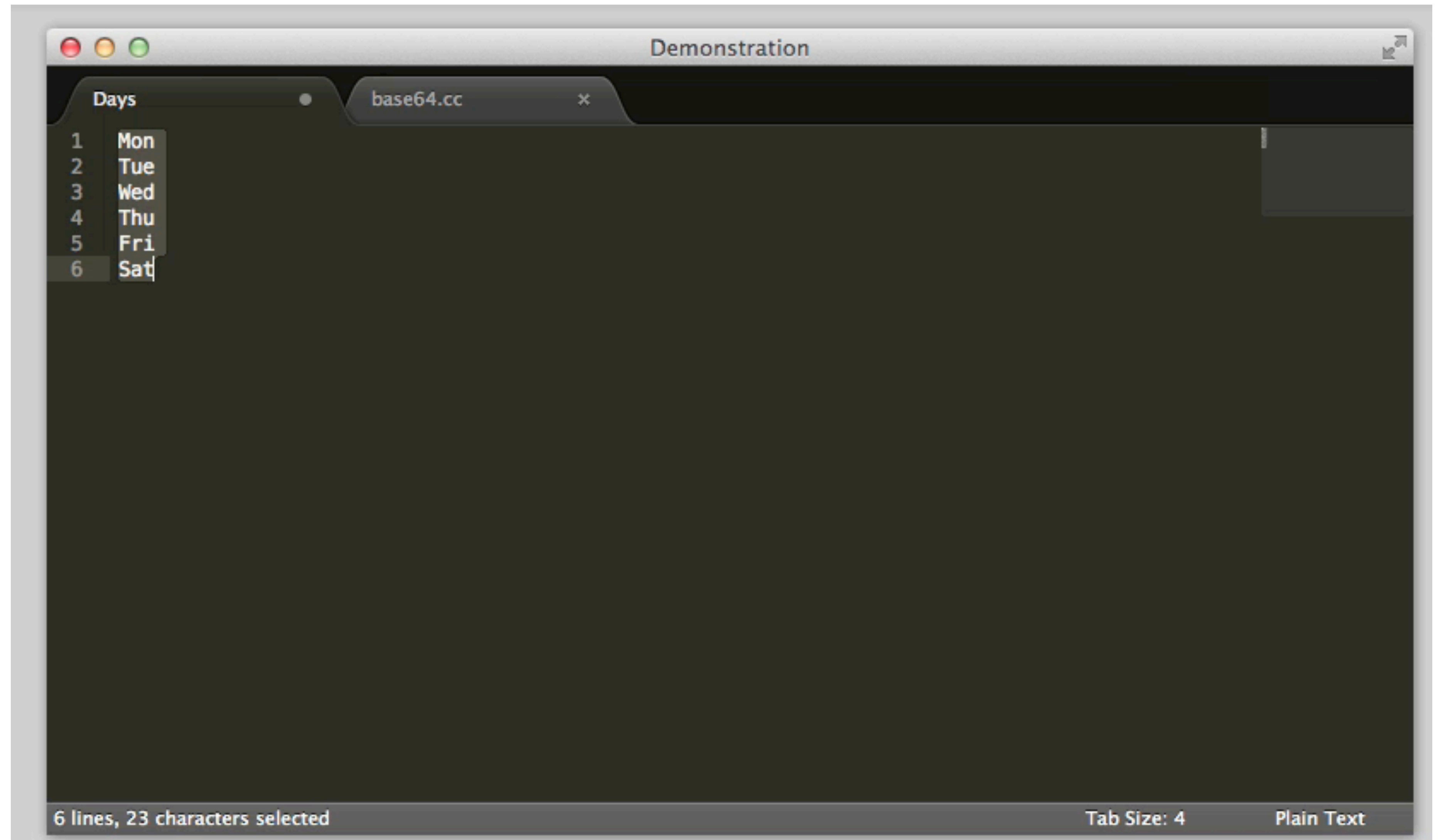
# Utilities

# Visual Studio Code, Sublime Text:

If you do information design, you need a good plaintext editor. These are the two mosty versatile options for programming and wrangling text data.

FREE/NAGWARE

KOPONEN + HILDÉN

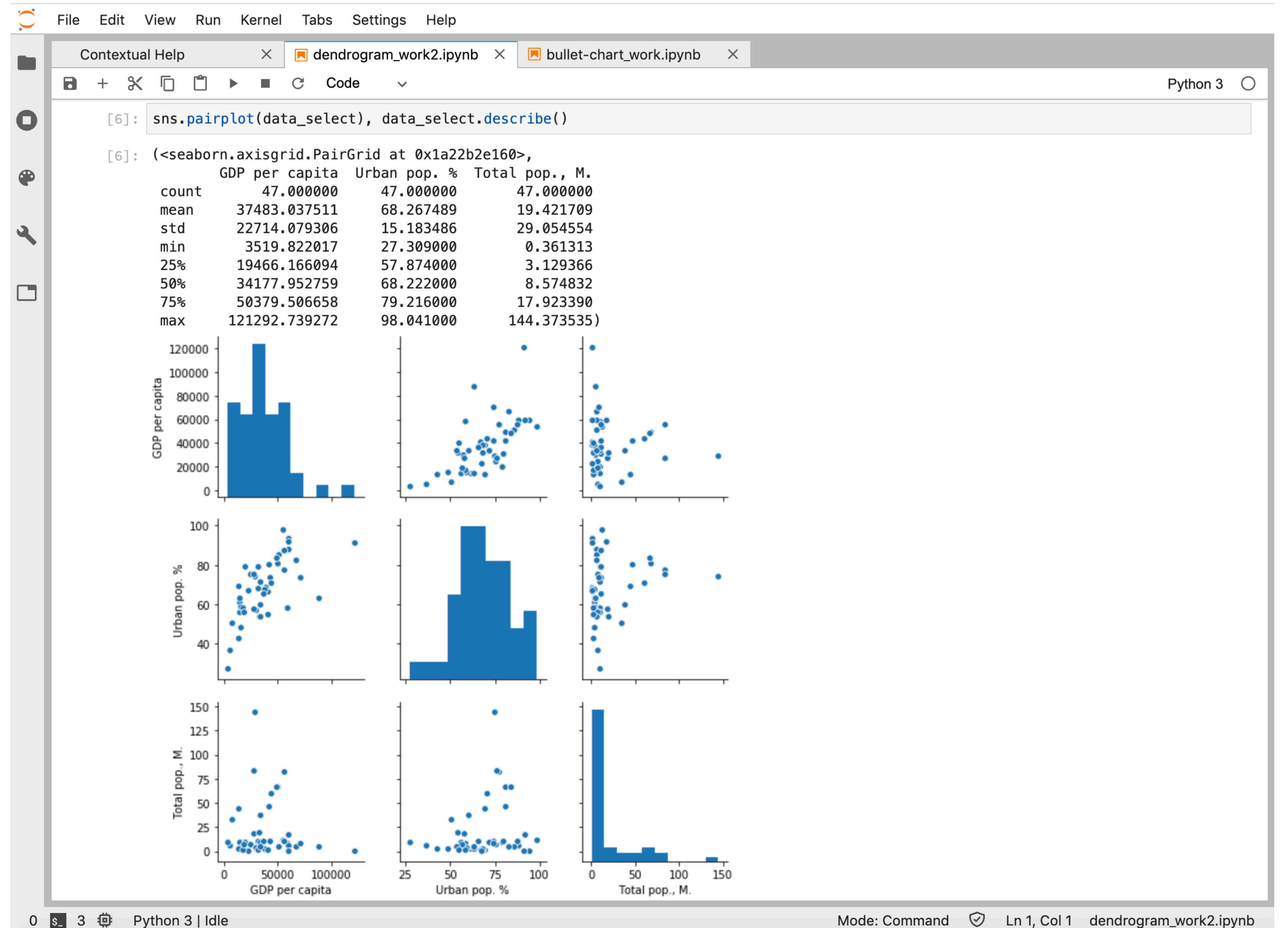


<https://code.visualstudio.com/>  
<http://www.sublimetext.com/>



**JupyterLab:**  
Interactive web  
interface for working  
with data and  
visualization  
libraries in *Jupyter  
Notebooks*:  
documents that  
combine live code  
with text, equations,  
images, interactive  
visualizations...

OPEN SOURCE



# Open Refine:

## A powerful tool for cleaning up messy data.

OPEN SOURCE

ESINE tsv Permalink

Facet / Filter Undo / Redo 51

34006 rows

Show as: rows records Show: 5 10 25 50 rows « first < previous 1 - 10 next > last »

All	original ID	original title	clustered title	original seconds	clustered seconds	original model	clustered model	original designe	clustered designe	original manufa	
☆	1.	0:0	tuoli	tuoli	pikkutuoli	pikkutuoli	Tempo puutuoli	Tempo puutuoli	Saarnio Timo	Saarnio Timo	Arktis Oy
☆	2.	0:0	muistikirja	muistikirja	kierrevihko	kierrevihko	Unikko	Unikko	Isola Kristina	Isola Kristina	Marimekko
☆	3.	0:0	muistikirja	muistikirja			Unikko	Unikko	Isola Kristina	Isola Kristina	Marimekko
☆	4.	0:0	kuppi	kuppi			Oiva / Unikko	Oiva / Unikko	Isola Kristina	Isola Kristina	
☆	5.	0:0	hillipurkki	hillipurkki					Still-McKinney Nanny	Still-McKinney Nanny	Riihimäen Lasi, Noormarkun Käsityöt
☆	6.	0:0	purkki	purkki					Still-McKinney Nanny	Still-McKinney Nanny	Noormarkun Käsityöt
☆	7.	0:0	maljakko	maljakko	taide-esine	taide-esine	3235	3235	Wirkkala Tapio	Wirkkala Tapio	littala
☆	8.	0:0					Paratiisi	Paratiisi			Arabia
☆	9.	0:0	vati	vati	tarjoiluvati	tarjoiluvati	AH, Aino	AH, Aino			Arabia
☆	10.	0:0	painokangas	painokangas	pöytäliina	pöytäliina	Mandariini	Mandariini	Kellomäki Kaarina	Kellomäki Kaarina	Marimekko

# Tabula: A tool for extracting tabular data from PDFs.

OPEN SOURCE

The screenshot shows the Tabula web application interface. At the top, there is a navigation bar with links for 'My Files', 'My Templates', 'About', 'Help', and 'Source Code'. On the right, there is a link to 'Support Tabula on OpenCollective!'. Below the navigation bar, the file name 'Vaalien\_tulostiedostojen\_kuvaus\_EKV-E...' is displayed. There are several buttons: 'Templates', 'Clear All Selections', 'Autodetect Tables', and 'Preview & Export Extracted Data'. The main content area shows a PDF document with a table of election results. The table is highlighted with a red dashed border, and a 'Repeat this Selection' button is visible at the bottom right of the table.

Koska äänioikeus eri vaaleissa muodostuu eri tavoilla, jää osa äänioikeutettujen lukumäärä- ja äänestysaktiivisuuskerrosta aina nolaksi. Seuraavassa taulukossa kerrotaan missä vaaleissa aluetiedostossa asuinpaikan ja kansalaisuuden mukaan eritellyt ryhmät voivat olla äänioikeutettuja:

	Suomessa asuvat Suomen kansalaiset	Ulkomailla asuvat Suomen kansalaiset	Suomessa asuvat muiden EU-valtioiden kansalaiset	Suomessa asuvat Norjan ja Islannin kansalaiset	Suomessa asuvat muiden valtioiden kansalaiset
Maakuntavaalit	x	-	x	x	x
Eduskuntavaalit	x	x	-	-	-
Europarlamenttivaalit	x	x	x	-	-
Kuntavaalit	x	-	x	x	x
Presidentinvaali	x	x	-	-	-

# SVG Crowbar: A Chrome extension for extracting SVGs (e.g. visualizations made with D3.js) from web pages.

OPEN SOURCE

# ai2html: A tool for creating responsive web pages with Illustrator.

OPEN SOURCE

KOPONEN + HILDÉN

## SVG Crowbar

A Chrome-specific bookmarklet that extracts SVG nodes and accompanying styles from an HTML document and downloads them as an SVG file—A file which you could open and edit in Adobe Illustrator, for instance. Because SVGs are resolution independent, it's great for when you want to use web technologies to create documents that are meant to be printed (like, maybe on newsprint). It was created with [d3.js](#) in mind, but it should work fine no matter how you choose to generate your SVG.

### The Bookmarklet

**SVG Crowbar** ← Drag this to your bookmarks bar.

After you've installed the bookmarklet, you can execute it on any page. Go ahead and try it out on this [crazy map](#).

(You can click on the link instead to test it on this page immediately.)

### Update

Some users reported that styles were not stored with the SVG files, so we added a new version that should work everywhere. The new method is slower, so loading can take a while on pages with many SVG elements. Still in beta.

**SVG Crowbar 2** ← Drag this to your bookmarks bar.

## ai2html

ai2html is an open-source script for Adobe Illustrator that converts your Illustrator documents into html and css.

Here are [examples of how we've used the script](#) at The New York Times and [examples of how others](#) have used it. Share your ai2html projects on Twitter, Delicious, etc. using [#ai2html](#).

## Table of contents

- [How to install ai2html](#)
- [How to use ai2html](#)
- [Frequently asked questions](#)
- [Settings](#)
- [Point text vs. area text](#)
- [Which attributes are converted to html and css](#)
- [How does ai2html work](#)
- [Limitations](#)
- [What works well and what does not](#)
- [Using fonts other than Arial and Georgia](#)

[nytimes.github.io/svg-crowbar/](https://nytimes.github.io/svg-crowbar/)  
<http://ai2html.org>

# Color

# Color Brewer:

A simple and useful color tool based on predesigned sets, especially suitable for maps.

OPEN SOURCE

The screenshot displays the Color Brewer 2.0 interface. At the top right, it says "COLOR BREWER 2.0 color advice for cartography". The main interface is divided into several sections:

- Number of data classes:** Set to 3.
- Nature of your data:** Radio buttons for sequential (selected), diverging, and qualitative.
- Pick a color scheme:** Two columns of color swatches labeled "Multi-hue" and "Single hue".
- Only show:** Checkboxes for "colorblind safe", "print friendly", and "photocopy safe".
- Context:** Checkboxes for "roads", "cities", and "borders" (checked).
- Background:** Radio buttons for "solid color" (selected) and "terrain". A "color transparency" slider is also present.
- 3-class BuGn:** A legend showing three color swatches with their corresponding hex codes: #e5f5f9, #99d8c9, and #2ca25f.
- EXPORT:** A section with icons for eye, print, and download, and a "HEX" dropdown menu.

The main map area shows a map of the United States with a 3-class sequential color scheme applied to the data classes. The colors range from light cyan to dark green.

# Chroma.js palette helper: Simple online tool for custom building *equidistant* quantitative palettes.

OPEN SOURCE

What's so wrong with HSV? <https://vis4.net/blog/posts/avoid-equidistant-hsv-colors/>

KOPONEN + HILDÉN

1 What kind of palette do you want to create?  
Palette type: **sequential** diverging  
Number of colors:

2 Select and arrange input colors

3 Check and configure the resulting palette  
 correct lightness  bezier interpolation  
simulate: **normal** deut. prot. trit.  
✓ This palette is colorblind-safe.

lightness saturation hue

100 90 80 70 60 50 40 30

80 60 40 20 0

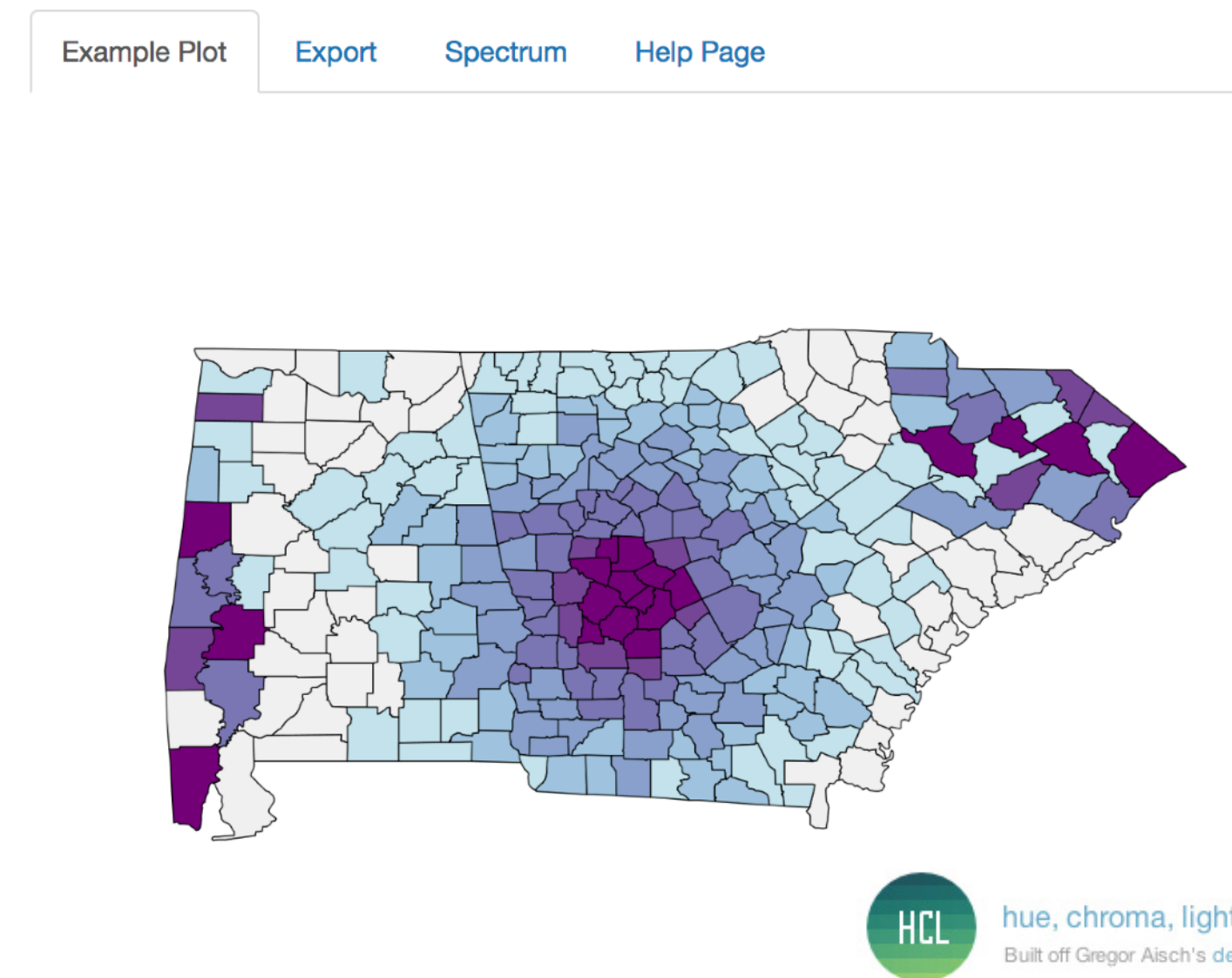
300 250 200 150 100

<https://gka.github.io/palettes/>  
JavaScript library: <https://github.com/gka/chroma.js>

# HCL Color Wizard and HCL Colorpicker: More tools for building equidistant quantitative color palettes.

OPEN SOURCE

The screenshot shows the HCL Color Wizard interface. It is divided into three main sections: Base Options, Color Settings, and Control Options. Base Options includes 'Nature of your data' (Sequential (multiple hues)), 'Base color scheme' (BuPu), and 'Example' (Map). Control Options includes checkboxes for 'Reverse', 'Correct colors' (checked), and 'Desaturated', as well as 'Vision' options (Normal, Deutan, Protan, Tritan). Color Settings includes sliders for Hue 1 (300), Hue 2 (200), Chron (60), Chrom (0), Lumin (25), Lumin (95), Power (0.7), Power (1.3), and Number (7). A color palette preview is shown at the bottom, and a 'Return to R' button is at the bottom right.



The screenshot shows the HCL Colorpicker interface. It features a color wheel with a white circle indicating the selected color. A list of color codes is shown on the right: #21313E, #20575F, #268073, #53A976, #98CF6F, and #EFEE69. The interface also includes a '1C' label, a '+' button, and a '-' button.

See also: <http://www.husl-colors.org/>  
<https://github.com/d3/d3-color>

<http://hclwizard.org:64230/hclwizard/>  
<http://tristen.ca/hcl-picker/>



# I Want Hue and Adobe CC Color: Tools for building harmonious qualitative color palettes.

FREE



Colors for data scientists. Generate and refine palettes of optimally distinct colors.

The interface for 'i want hue' is divided into two main sections: 'Color space' and 'Palette'. The 'Color space' section includes sliders for H (78), C (0), and L (0), with corresponding color gradients. A 'Dark background' checkbox is also present. The 'Palette' section shows a grid of generated colors and a 'Reroll palette' button. A central visualization shows a cluster of points in a 2D space, colored according to the current settings.

## Colors

A preview of the generated color palette, showing a horizontal bar with a gradient from light green to dark green. Below it, the hex code #63DB5C and RGB values 99,219,92 are displayed. A second bar shows a darker shade of green.

## JSON

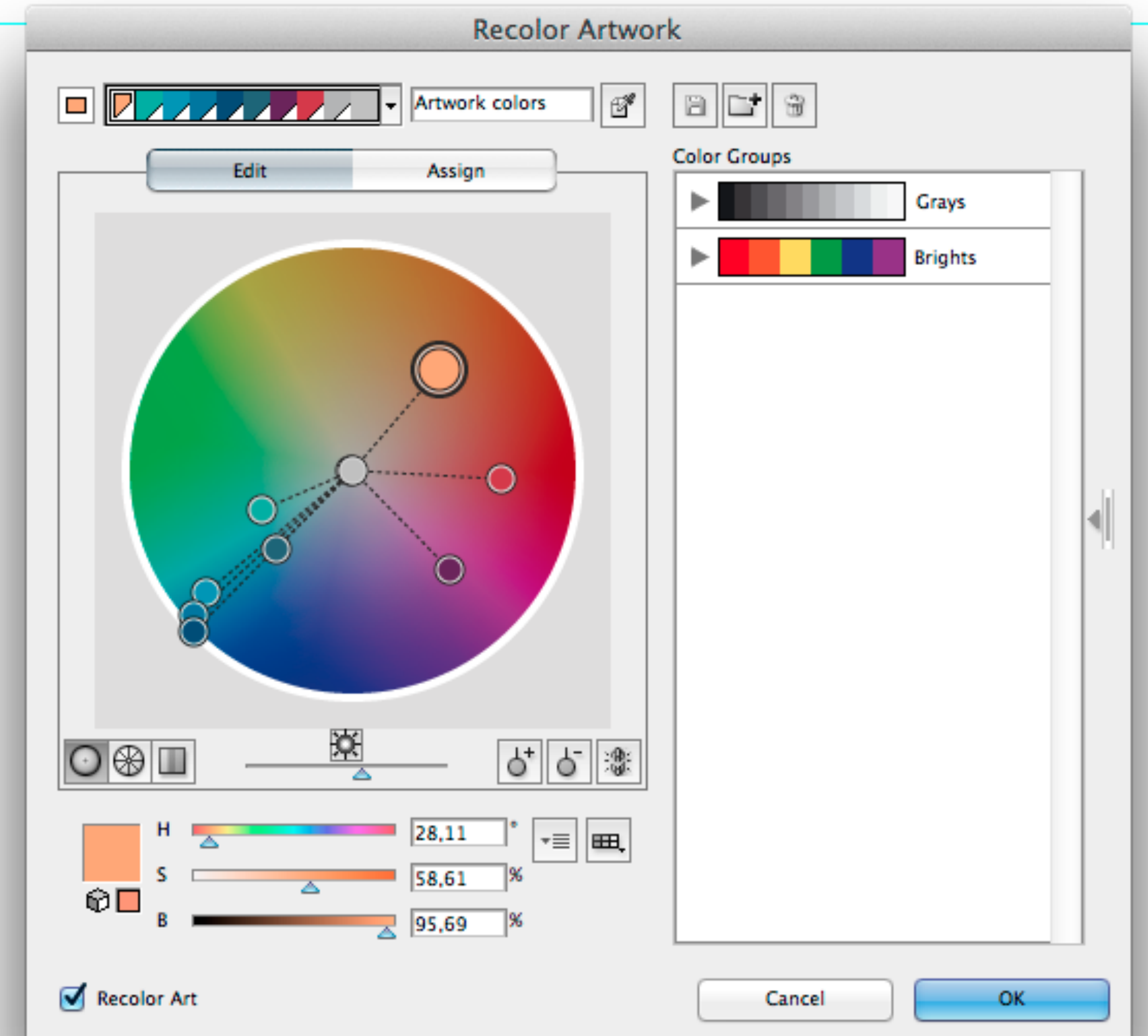
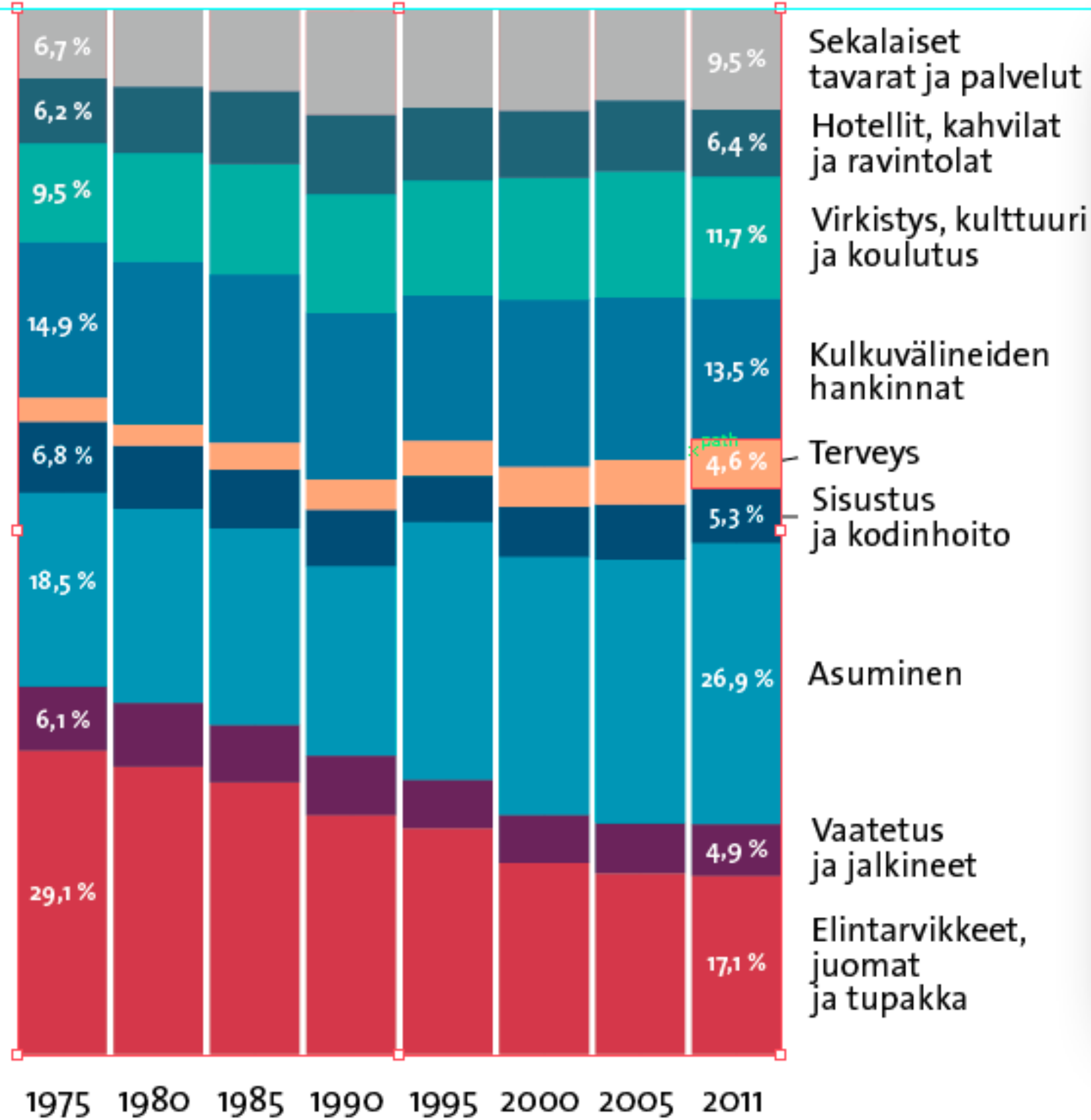
HEX json  
[ "#63DB5C",  
"#364B2C",  
"#7BDBC8",  
"#5E903E",  
"#B6D392",  
"#B8D848",  
"#5F9080" ]

The Adobe CC Color interface features a 'Color Rule' dropdown menu with options like Analogous, Monochromatic, Triad, Complementary, Compound, Shades, and Custom. A central color wheel allows for selecting colors. Below the wheel, a horizontal bar displays five color swatches: light green, blue, teal, yellow, and orange. At the bottom, there are sliders for adjusting the colors and a table of color data.

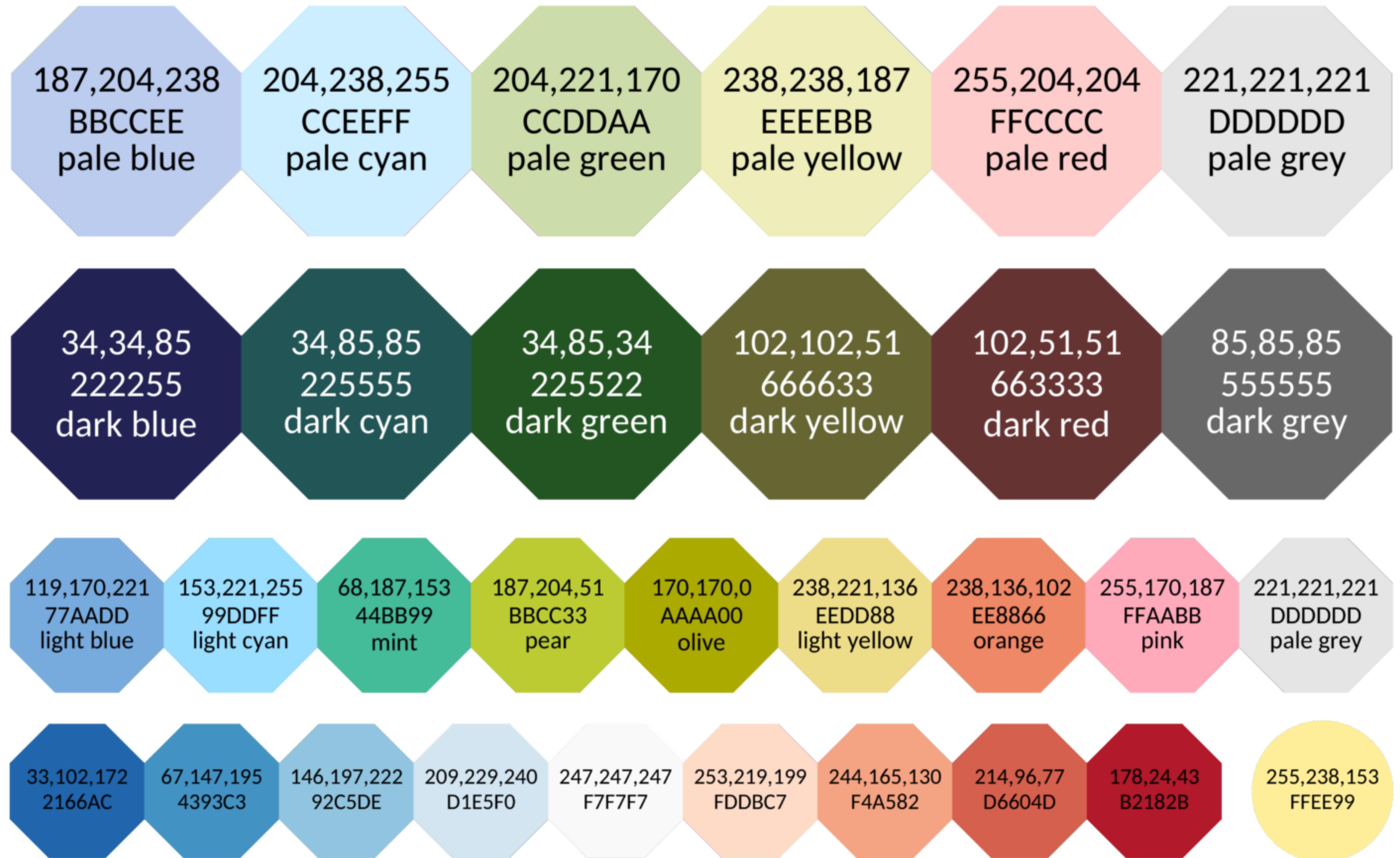
# The Recolor Artwork tool in Illustrator: Very handy for manual color tweaks.

COMMERCIAL

**Suomalaisten kulutusmenot käyttötarkoituksen mukaan**  
Prosenttiosuus kokonaiskulutuksesta 1975–2011



**Paul Tol's  
notes: A  
comprehensive  
guide to color  
in scientific  
visualizations.**



- **Excel, Numbers**
- **Open Refine, Tabula**
- **Powerpoint, Keynote**
- **Illustrator, Sketch**
- **QGIS (+ Mapbox)**
- **Python + Pandas + Matplotlib & Seaborn & Altair**  
(+ some **Nodebox 1.0**)
- a little **R** and **Tableau**
- **D3.js** (+ other javascript libraries)
- **Visual Code Studio, Sublime Text**
- **Slack**
- Currently in process of adding to our setup: **React, PostGIS**
- Try to use more: **GitHub, Trello**

- **Excel, Numbers**
- **Open Refine, Tabula**
- **Powerpoint, Keynote**
- **Illustrator, Sketch**
- **QGIS (+ Mapbox)**
- **Python + Pandas + Matplotlib & Seaborn & Altair**  
(+ some **Nodebox 1.0**)
- a little **R** and **Tableau**
- **D3.js** (+ other javascript libraries)
- **Visual Code Studio, Sublime Text**
- **Slack**
- Currently in process of adding to our setup: **React, PostGIS**
- Try to use more: **GitHub, Trello**

# Data sources

# Examples of types of data

- statistics
- official registries
- Big Data
- MyData
- scientific research data
- surveys
- GIS data

# Ways of accessing data

- statistical databases
- APIs
- scraping
- FOIA requests
- commercial data products
- data leaks
- crowdsourcing
- self-collection

# Data sources: Finland

- **Statistics Finland:** <http://stat.fi/tup/tilastotietokannat/>
  - StatFin: <http://pxnet2.stat.fi/PXWeb/pxweb/fi/StatFin/>
  - Paavo – open data by postcode: [http://pxnet2.stat.fi/PXWeb/pxweb/fi/Postinumeroalueittainen\\_avoin\\_tieto/](http://pxnet2.stat.fi/PXWeb/pxweb/fi/Postinumeroalueittainen_avoin_tieto/)
  - Eurostat key tables: <http://pxnet2.stat.fi/PXWeb/pxweb/fi/Eurostat/>
  - Statistics Finland GIS data: <https://www.stat.fi/org/avoindata/paikkatietoaineistot.html>
- **THL:** <https://thl.fi/fi/tilastot-ja-data/tilastot-aiheittain>
  - SotkaNet: <https://sotkanet.fi/>
  - Terveystemme.fi: <http://www.terveytemme.fi>



# Data sources: Finland

- **Aluesarjat:** <http://aluesarjat.fi>
- **Natural Resources Institute Finland:** <https://stat.luke.fi>
- **Helsinki Region Infoshare:** <https://hri.fi>
- **National Land Survey:** [www.maanmittauslaitos.fi/asioi-verkossa/avoimien-aineistojen-tiedostopalvelu](http://www.maanmittauslaitos.fi/asioi-verkossa/avoimien-aineistojen-tiedostopalvelu)
- **Paikkatietoikkuna:** <https://kartta.paikkatietoikkuna.fi/>
- **Environmental administration's GIS service Liiteri:** <https://liiteri.ymparisto.fi>

# Data sources: Finland

- **Road Safety Council:** [www.liikenneturva.fi/fi/tutkittua](http://www.liikenneturva.fi/fi/tutkittua)
- **Passenger transport survey :** [www.traficom.fi/fi/hlt](http://www.traficom.fi/fi/hlt)
- **Alcohol statistics:** [www.valvira.fi/alkoholi/tilastot](http://www.valvira.fi/alkoholi/tilastot)
- **Theatre statistics:** [www.tinfo.fi/fi/teatteritilastot](http://www.tinfo.fi/fi/teatteritilastot)
- **Publishers' statistics:** [kustantajat.fi/tilastot](http://kustantajat.fi/tilastot)

**Other sources:** ministries, municipalities, industry groups, trade unions, regulators, sectoral research institutes; companies (e.g. mobile operators, AirBnB, Instagram...)

# Data sources: International

- **World Bank:** <https://data.worldbank.org>
- DataBank: <https://databank.worldbank.org>
- **Gapminder:** <https://www.gapminder.org/data/>
- **OECD:** <http://data.oecd.org>
- **Eurostat:** <https://ec.europa.eu/eurostat>
- Statistics Explained: <https://ec.europa.eu/eurostat/statistics-explained/>
- **Natural Earth:** <https://www.naturalearthdata.com>
- **OpenStreetMap:** <https://www.openstreetmap.org>

# Data sources: International

- **FAO:** [www.fao.org/faostat/en/](http://www.fao.org/faostat/en/)
- **WHO:** [www.who.int/gho/en/](http://www.who.int/gho/en/)
- **UNHCR:** [popstats.unhcr.org/en/overview](http://popstats.unhcr.org/en/overview)
- **Nordic Statistics:** <http://www.nordicstatistics.org/pxweb/en/Nordic%20Statistics/>

# Thank you.

**Jonatan Hildén**

+ 358 50 3410 0461

[jonatan@koponen-hilden.fi](mailto:jonatan@koponen-hilden.fi)

@jhilden

@koponenhilden