

Data Science with Orange Toolbox:

Data Science for Everyone

Blaž Zupan
University of Ljubljana, Slovenia &
Baylor College of Medicine, Houston



Data Science

Linear algebra

Probability

Statistics

Mathematical optimization

Data & model representation

Machine learning

High performance computing

Data visualization

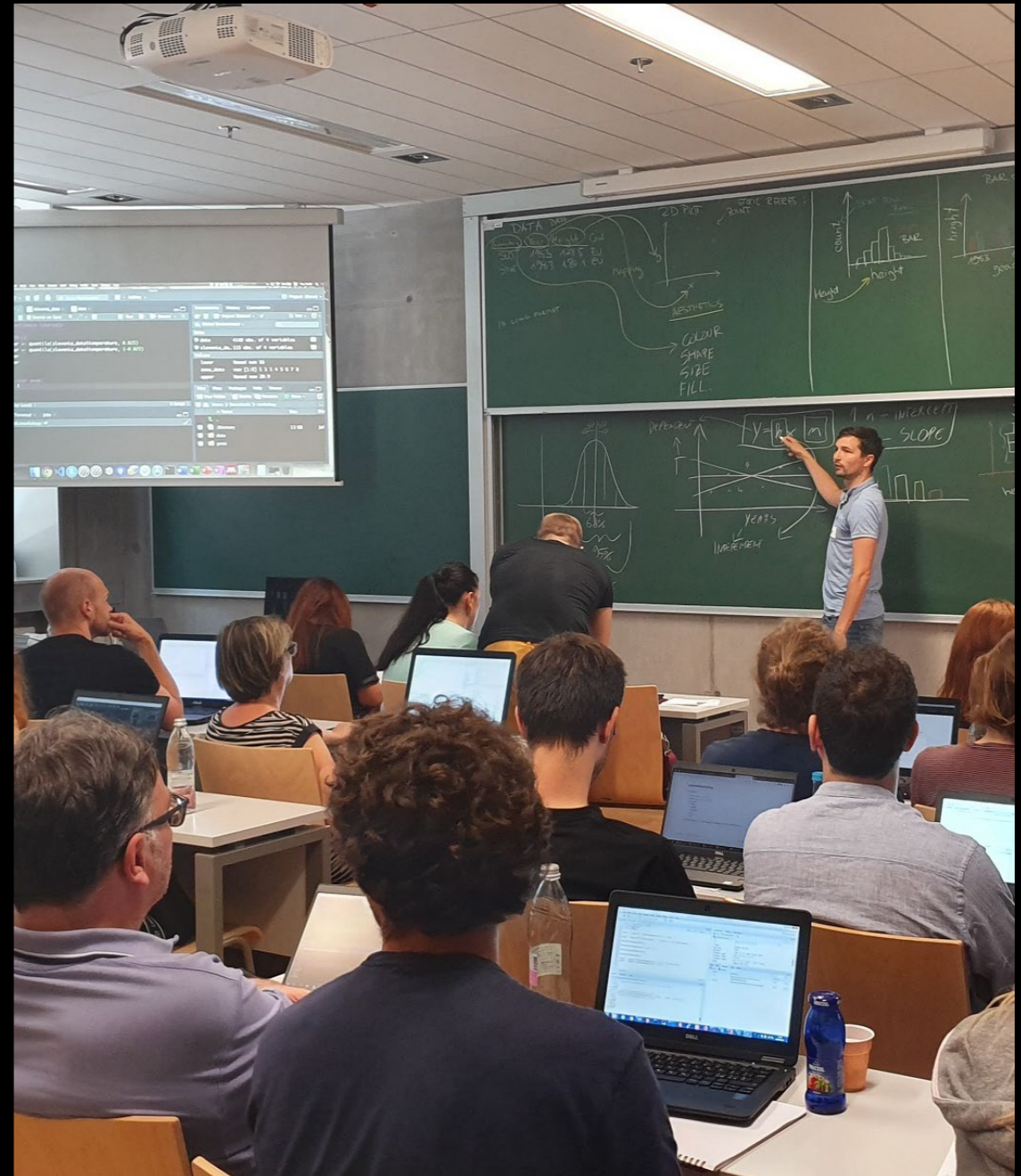
Scripting languages



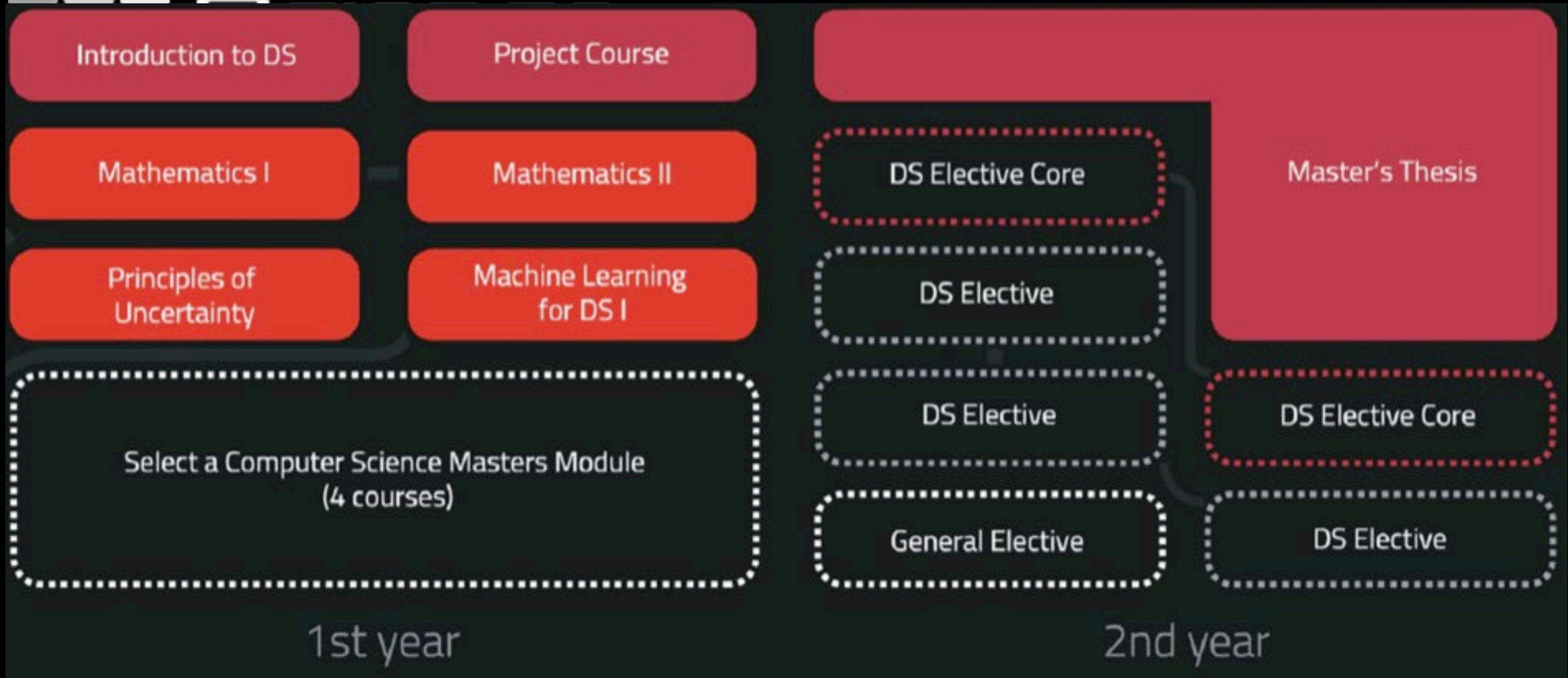
Data Science



2-5 hours to train concepts



Data Science



Data Science



2-5 hours to train concepts

data preparation, feature engineering, clustering, modeling, evaluation, data understanding



član skupine Sandoz



```
def edges_from_osmdb(osmdb, vertex_namespace, slogs, profiledb=None):
    """generates (vertex1_label, vertex2_label, edgepayload) from osmdb"""

    street_id_counter = 0
    street_names = {}

    # for each edge in the osmdb
    for i, (id, parent_id, node1, node2, distance, geom, tags) in enumerate(

        # Find rise/fall of edge, if profiledb is given
        rise=0
        fall=0
        if profiledb:
            profile = profiledb.get( id )
            if profile:
                rise, fall = get_rise_and_fall( profile )

        # insert end vertices of edge to graph
        vertex1_label = "%s-%s"%(vertex_namespace,node1)
        vertex2_label = "%s-%s"%(vertex_namespace,node2)

        # create ID for the way's street
        street_name = tags.get("name")
        if street_name is None:
            street_id_counter += 1
            street_id = street_id_counter
        else:
            if street_name not in street_names:
                street_id_counter += 1
                street_names[street_name] = street_id_counter
            street_id = street_names[street_name]

        # Create edges to be inserted into graph
        s1 = Street( id, distance, rise, fall )
        s2 = Street( id, distance, fall, rise, reverse_of_source=True )
        s1.way = street_id
        s2.way = street_id

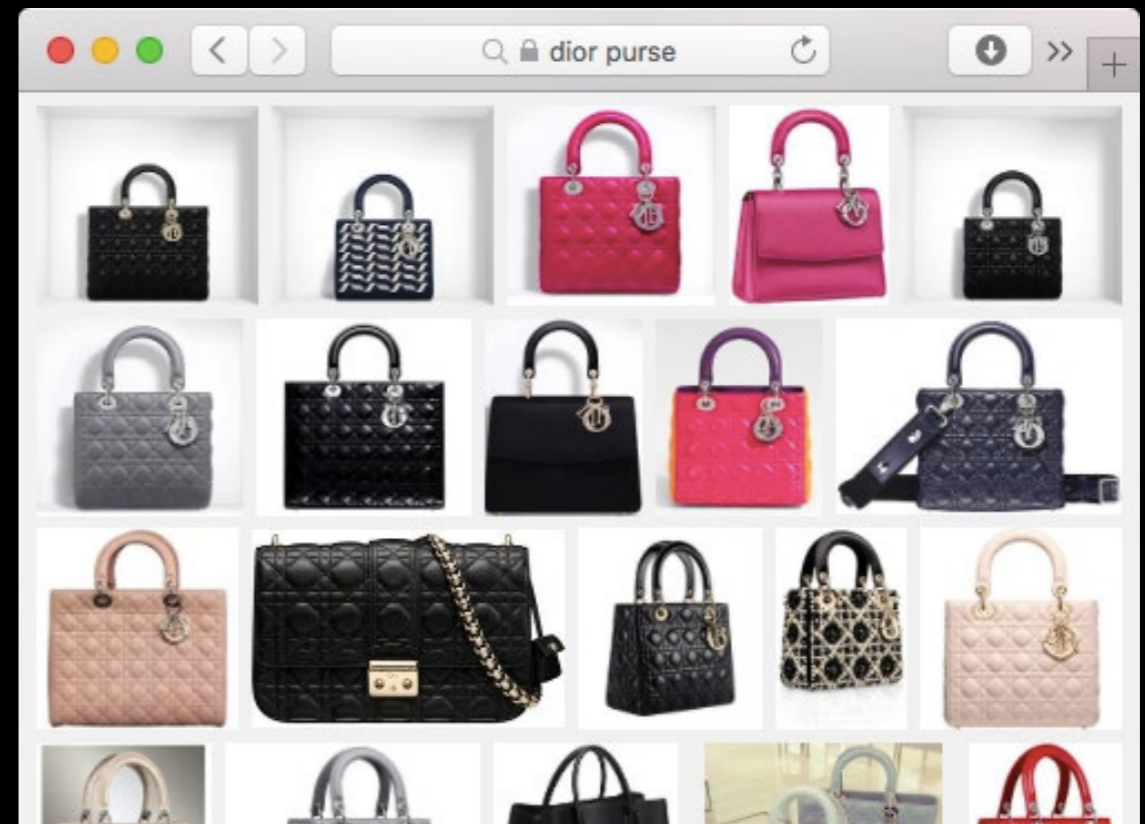
        # See if the way's highway tag is penalized with a 'slog' value; if s
        slog = slogs.get( tags.get("highway") )
        if slog:
            s1.slog = s2.slog = slog

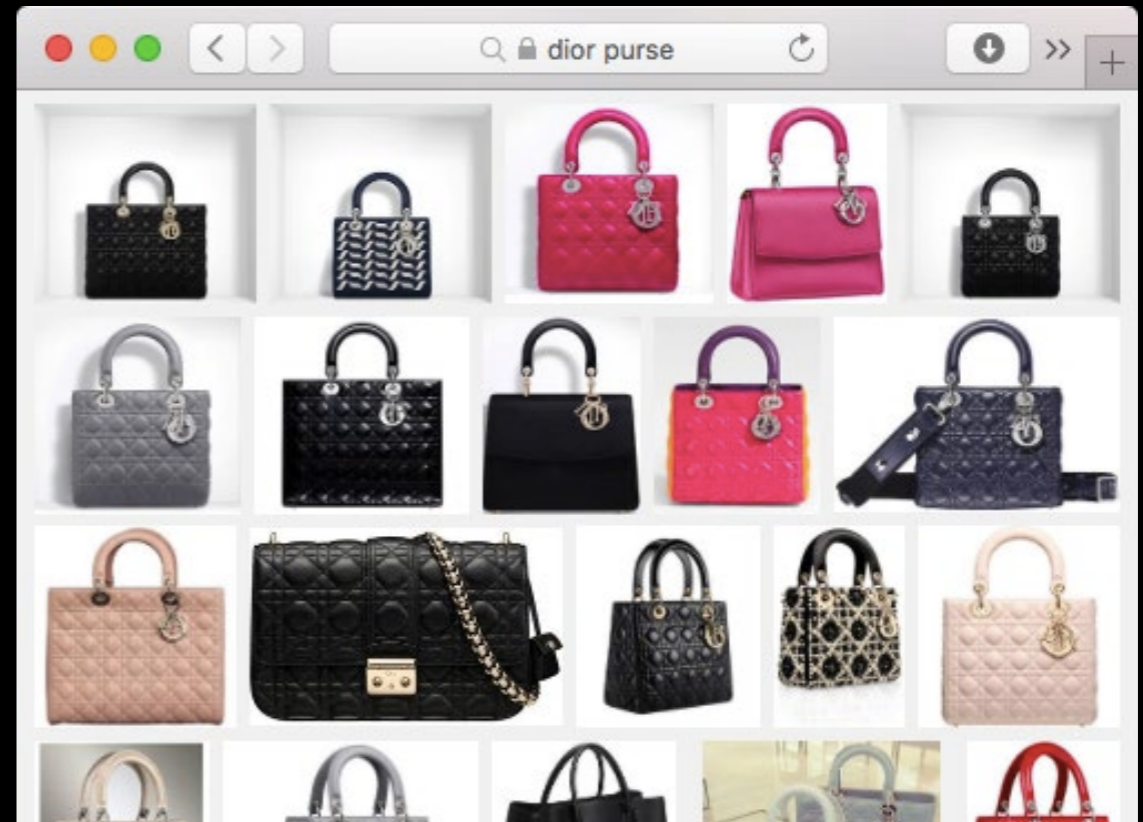
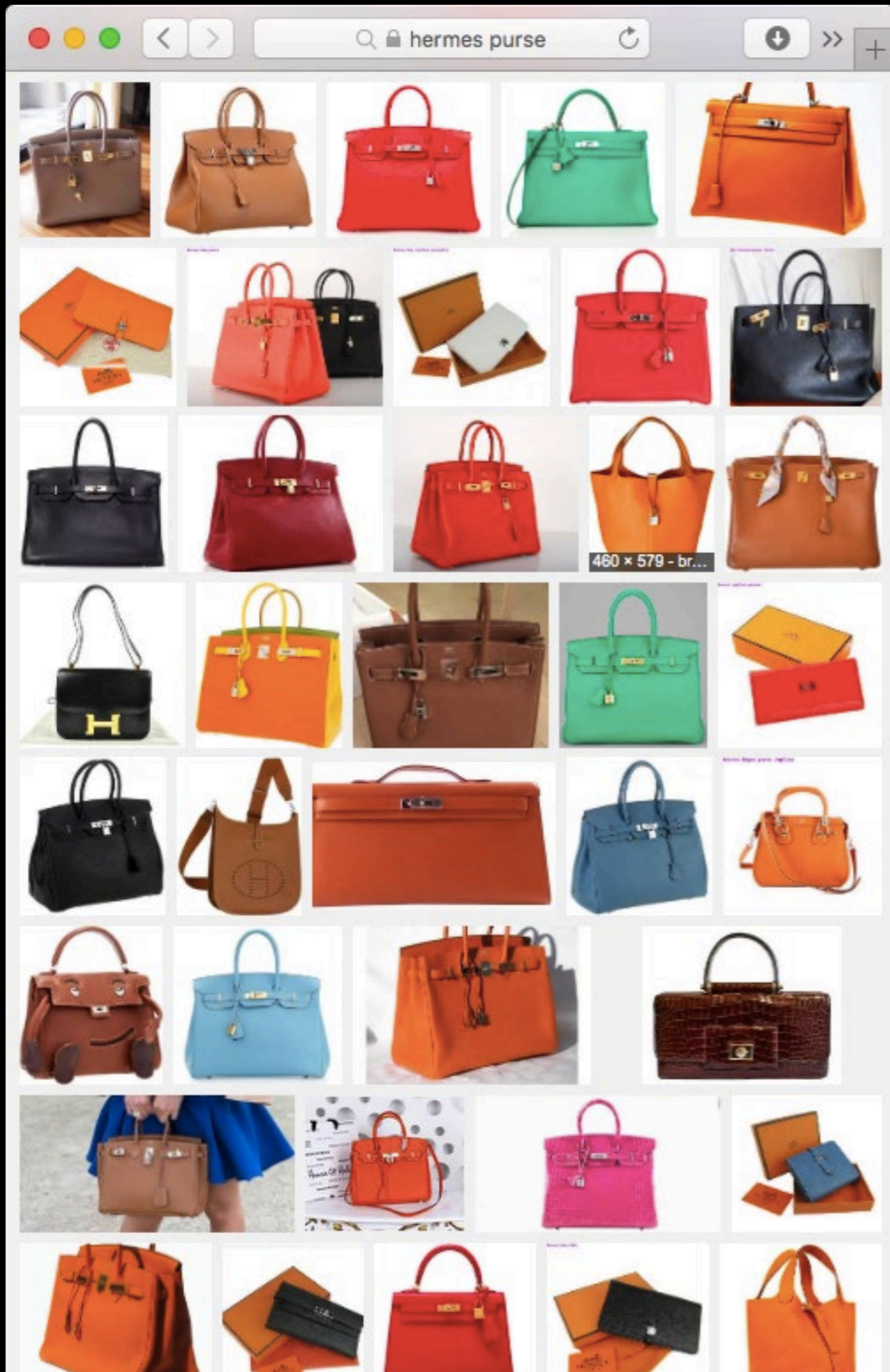
        # Add the forward edge and the return edge if the edge is not oneway
        yield vertex1_label, vertex2_label, s1

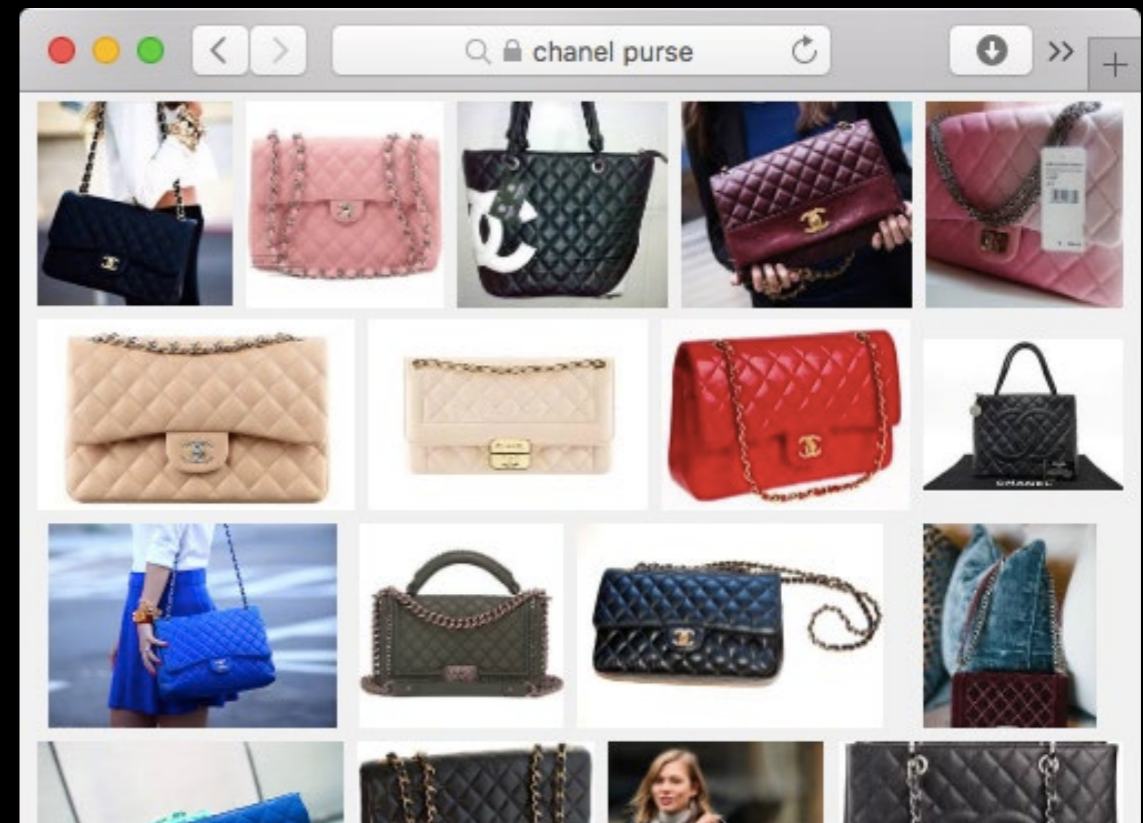
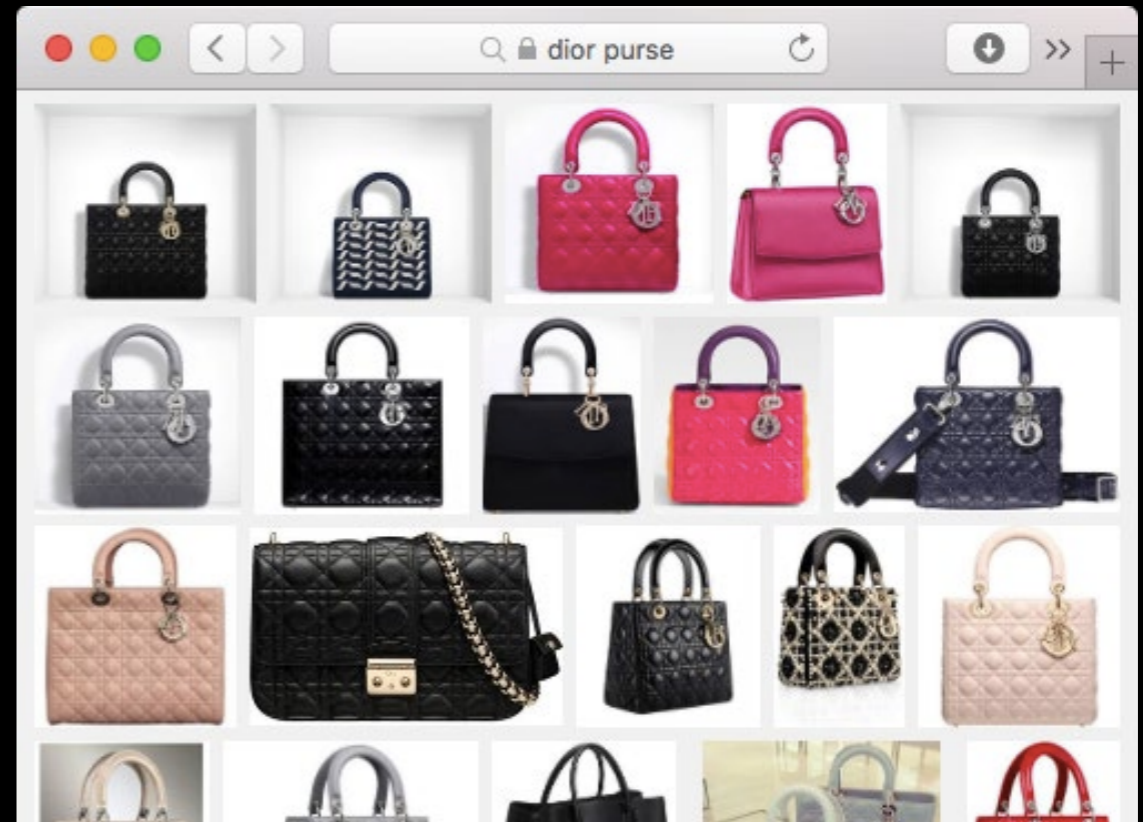
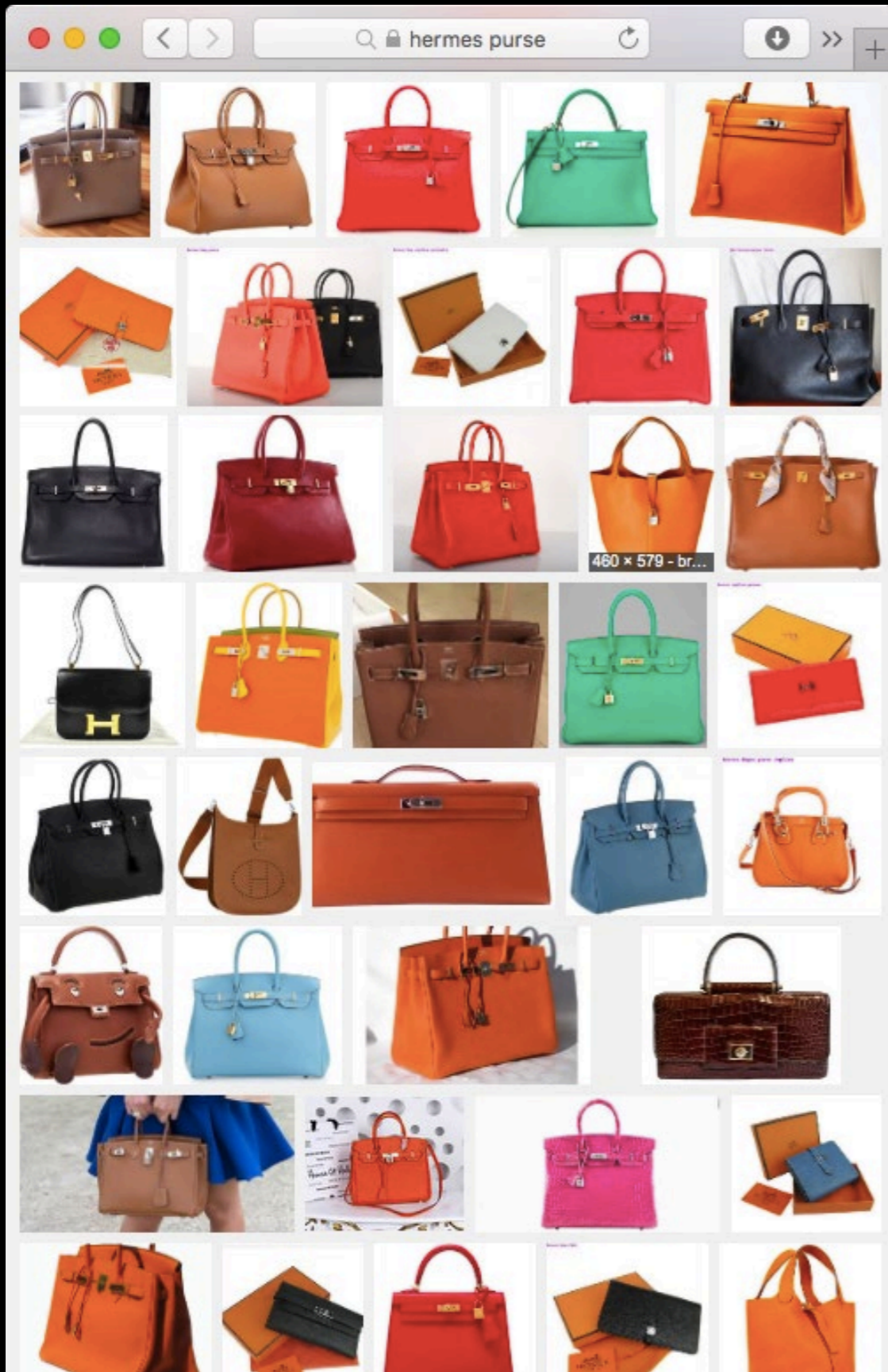
        oneway = tags.get("oneway")
        if oneway != "true" and oneway != "yes":
            yield vertex2_label, vertex1_label, s2
```

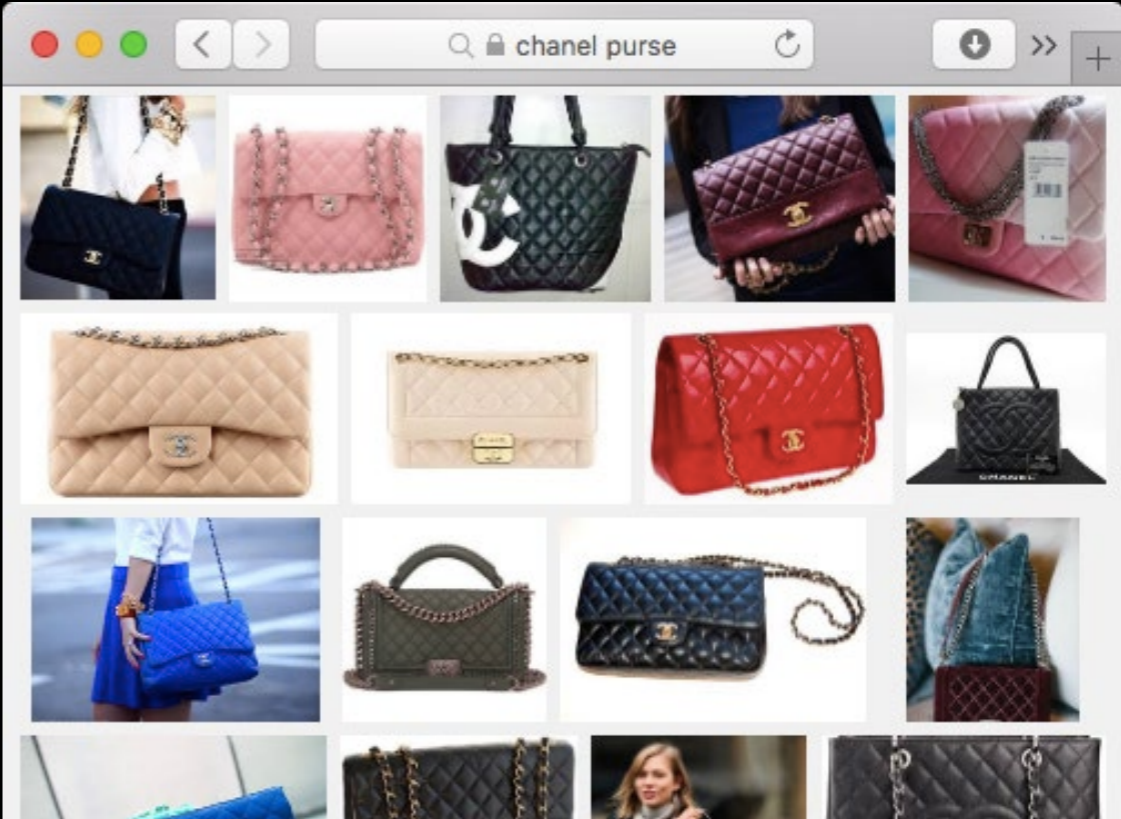


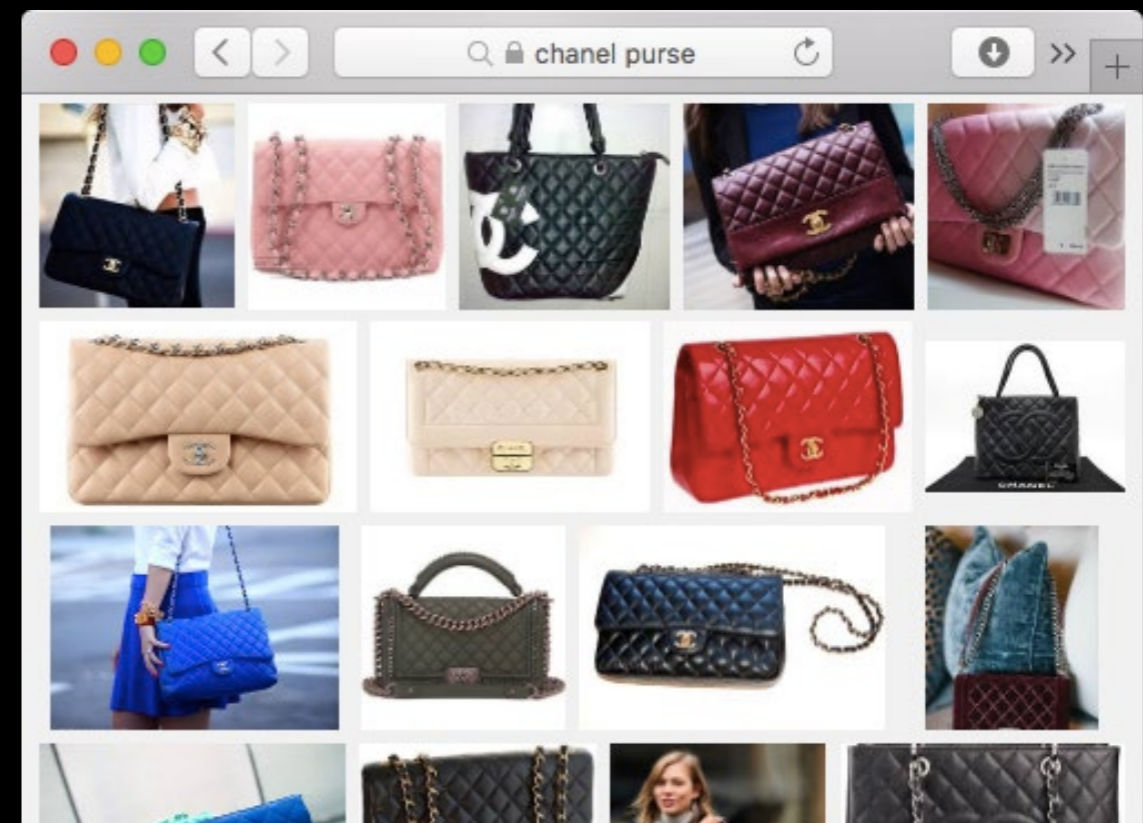
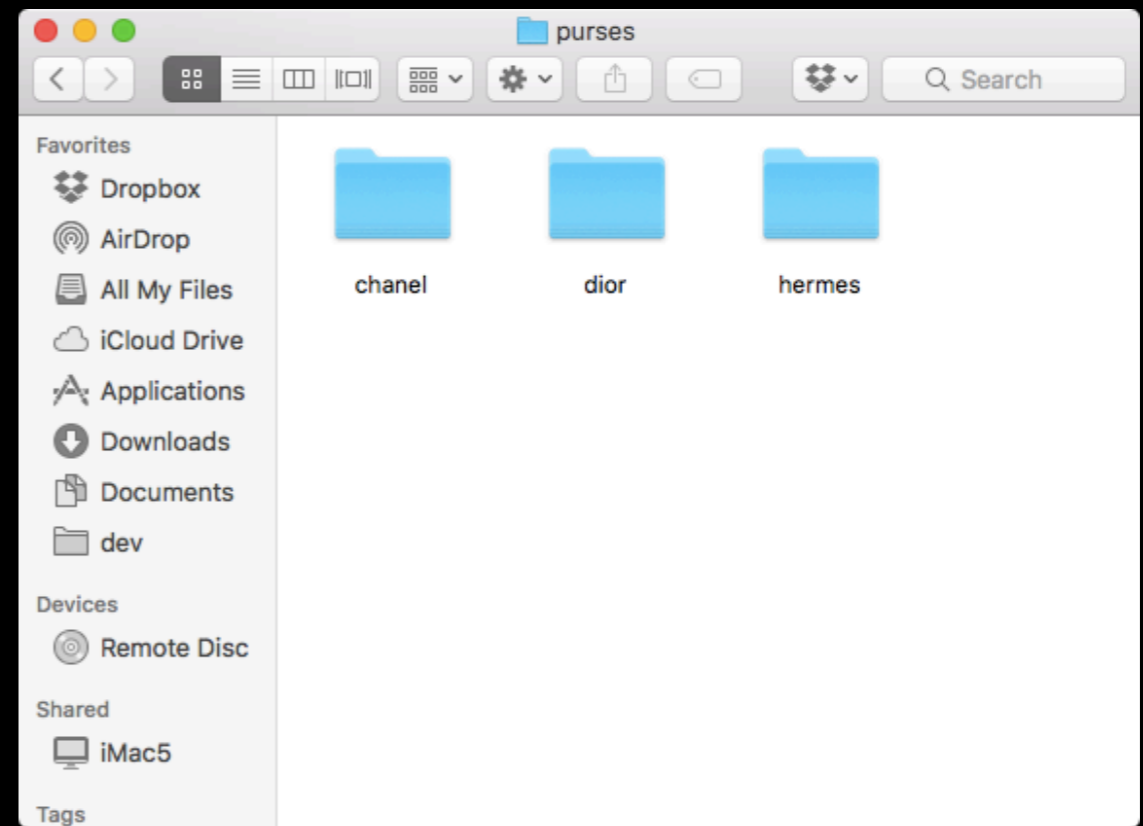












Mac OS Finder window titled "purses" showing a file list in table view.

Name	Size	Date Mod
▼ chanel	--	01/12/2
c1.jpeg	6 KB	30/11/2
c2.jpeg	7 KB	30/11/2
c3.jpeg	10 KB	30/11/2
c4.jpeg	8 KB	30/11/2
c5.jpeg	14 KB	30/11/2
c6.jpeg	6 KB	30/11/2
c7.jpeg	7 KB	01/12/2
c8.jpeg	9 KB	01/12/2
c9.jpeg	6 KB	01/12/2
c10.jpeg	6 KB	01/12/2
▼ dior	--	01/12/2
d1.jpeg	7 KB	30/11/2
d3.jpeg	6 KB	30/11/2
d4.jpeg	8 KB	30/11/2
d5.jpeg	7 KB	30/11/2
d6.jpeg	7 KB	30/11/2
d7.jpeg	10 KB	30/11/2
d8.jpeg	9 KB	01/12/2
d9.jpeg	9 KB	01/12/2
d10.jpeg	7 KB	01/12/2
d11.jpeg	9 KB	01/12/2
d12.jpeg	5 KB	01/12/2
▼ hermes	--	01/12/2
a.jpeg	7 KB	30/11/2
b.jpeg	6 KB	01/12/2
c.jpeg	7 KB	30/11/2
d.jpeg	5 KB	30/11/2
e.jpeg	4 KB	30/11/2
f.jpeg	4 KB	30/11/2
g.jpeg	5 KB	30/11/2
i.jpeg	8 KB	01/12/2
j.jpeg	5 KB	01/12/2

Left sidebar includes Favorites (Dropbox, AirDrop, All My Files, iCloud Drive, Applications, Downloads, Documents, dev), Devices (Remote Disc), Shared (iMac5), and Tags (Red, Orange, Yellow, Green, Blue, Purple, Gray, All Tags...).

Mac OS Finder window titled "purses" showing a folder view.

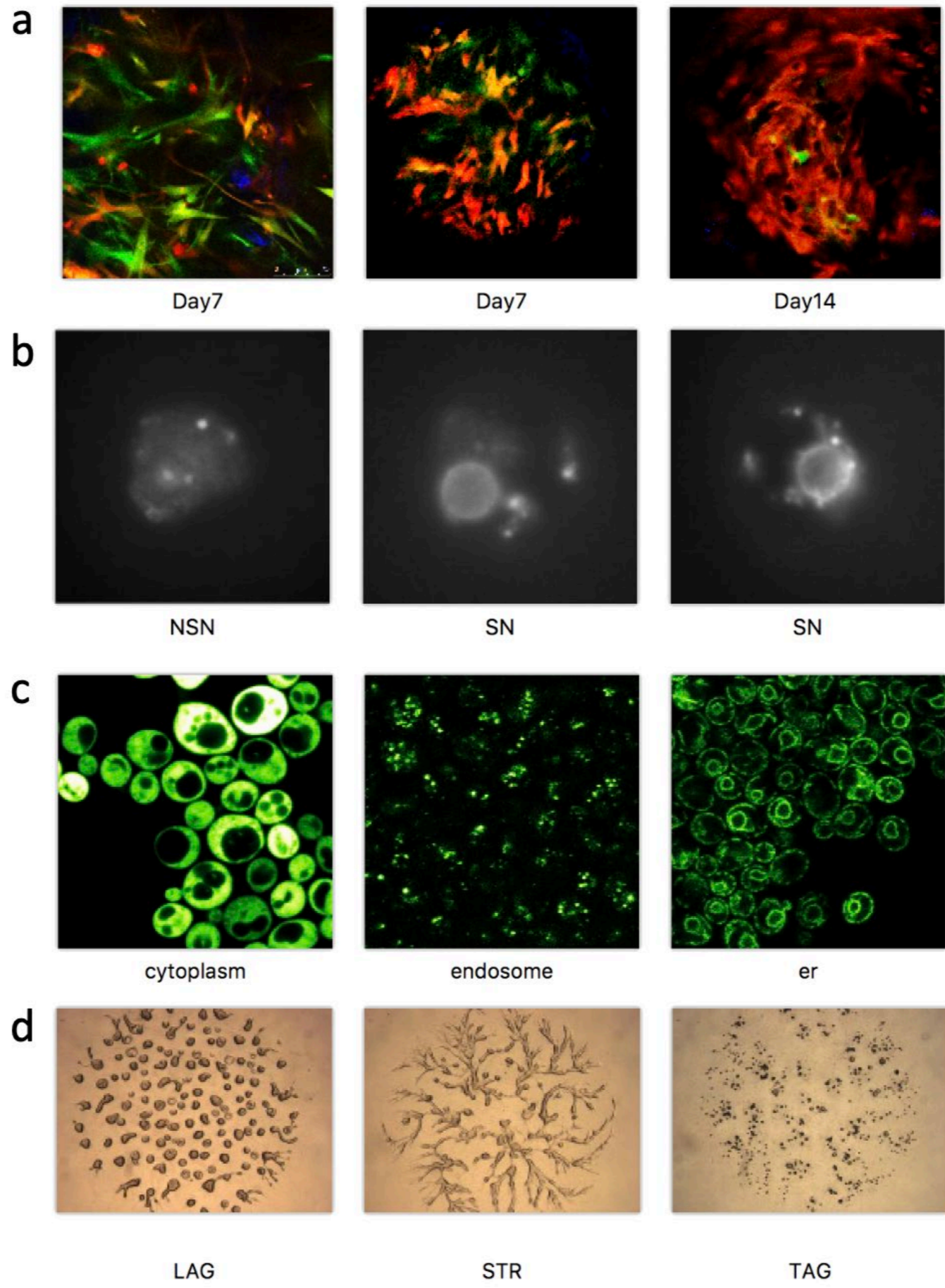
Three folders are visible: chanel, dior, and hermes.

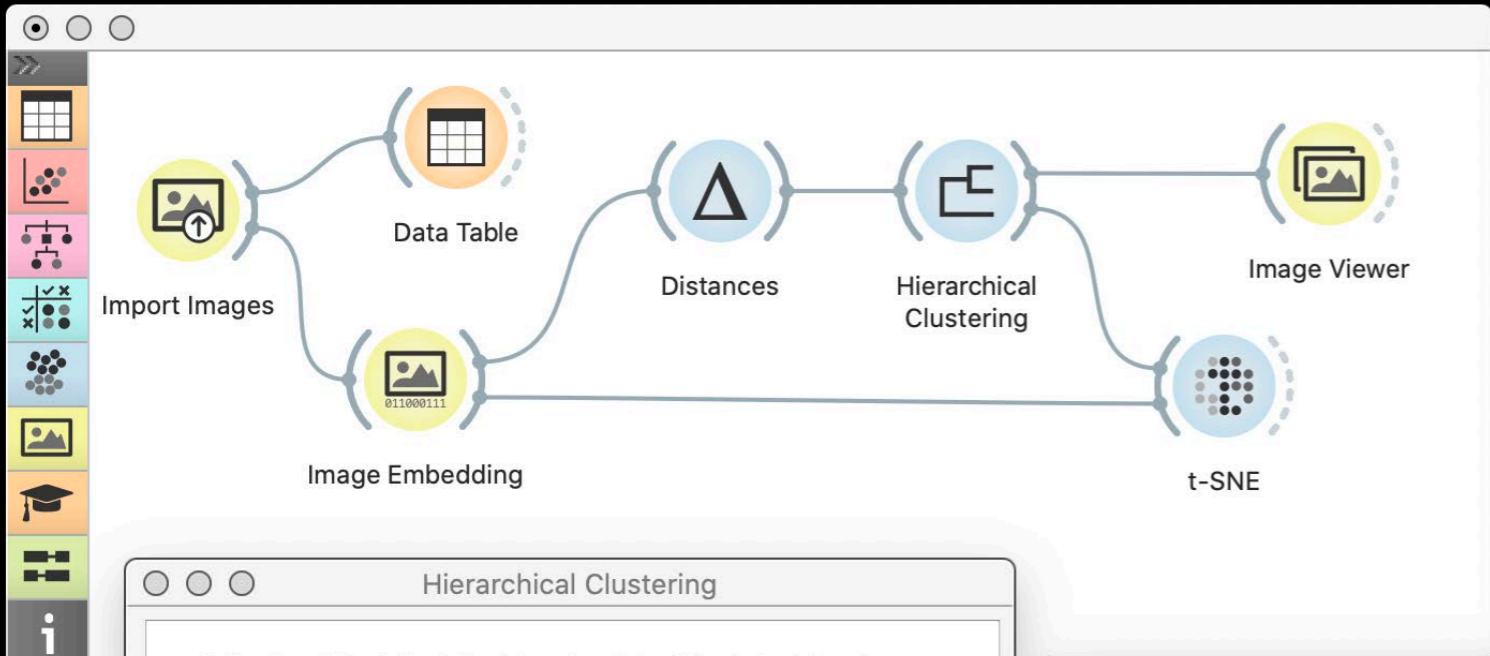
Left sidebar includes Favorites (Dropbox, AirDrop, All My Files, iCloud Drive, Applications, Downloads, Documents, dev), Devices (Remote Disc), and Shared (iMac5).

Mac OS Photos window titled "chanel purse" showing a grid of images.

The grid contains 16 images of various Chanel handbags in different colors and styles, including quilted and chain-link designs.

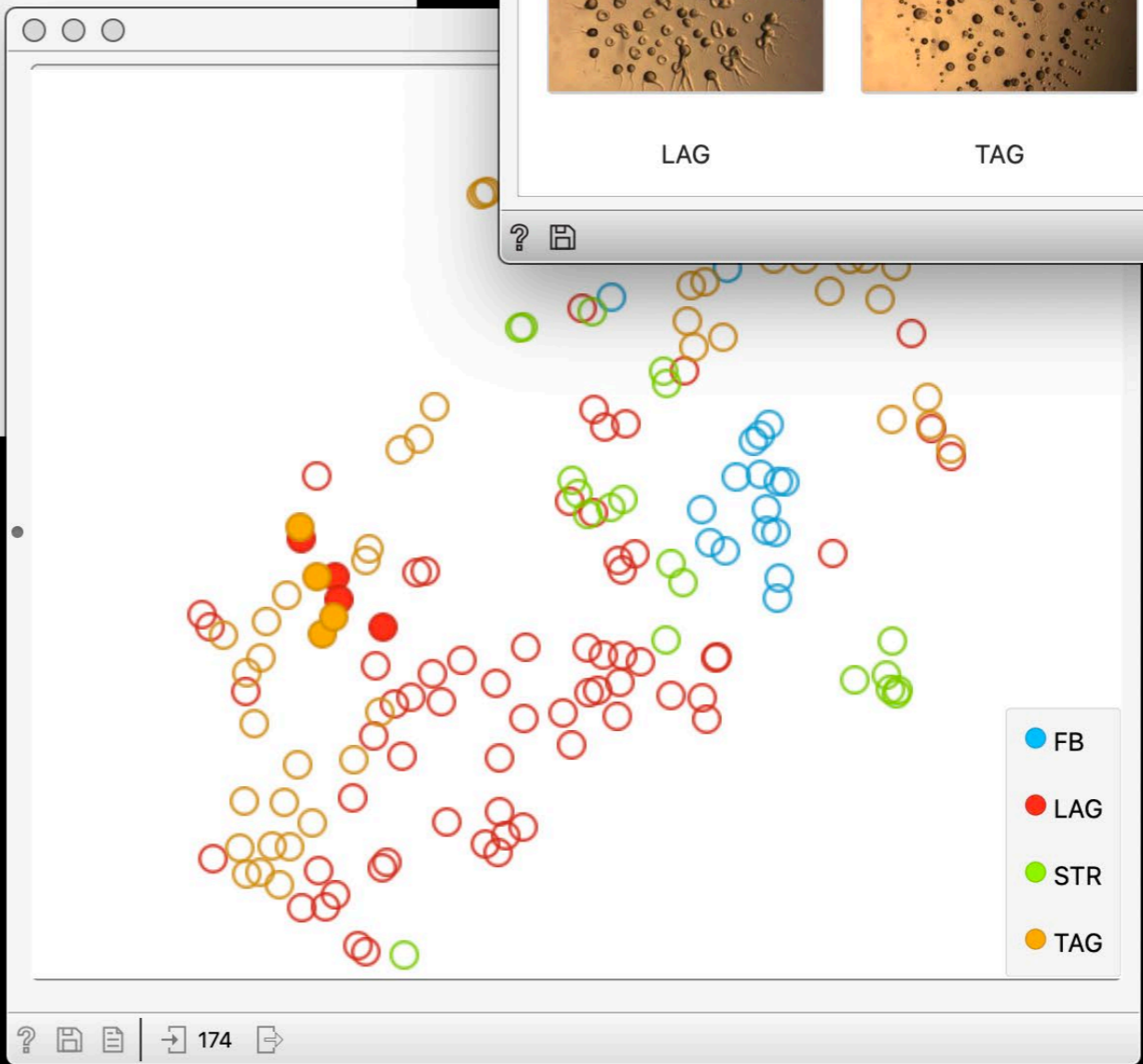
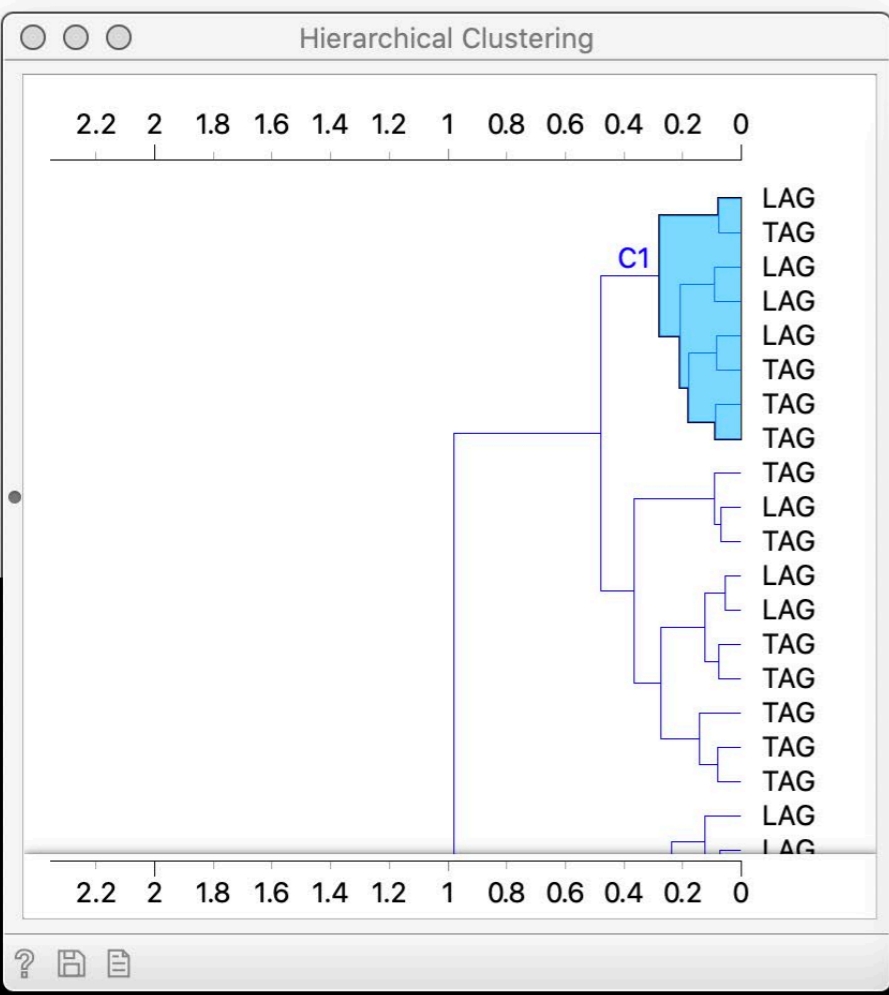
Purse Analytics (live demo)





The Image Viewer window displays a 2x3 grid of images:

- Top row: Three images labeled **LAG**.
- Bottom row: Three images labeled **LAG**, **TAG**, and **TAG**.



A vertical toolbar on the left side of the interface contains various icons for data manipulation and visualization, including a grid, scatter plot, tree, matrix, image, and other symbols.



Search



E-alert



Submit



Login

Article | [Open Access](#) | Published: 07 October 2019

Democratized image analytics by visual programming through integration of deep models and small-scale machine learning

Primož Godec, Matjaž Pančur, Nejc Ilenič, Andrej Čopar, Martin Stražar, Aleš Erjavec, Ajda Pretnar, Janez Demšar, Anže Starič, Marko Toplak, Lan Žagar, Jan Hartman, Hamilton Wang, Riccardo Bellazzi, Uroš Petrovič, Silvia Garagna, Maurizio Zuccotti, Dongsu Park, Gad Shaulsky & Blaž Zupan

Nature Communications **10**, Article number: 4551 (2019) | [Cite this article](#)

4505 Accesses | **1** Citations | **135** Altmetric | [Metrics](#)

Abstract

Analysis of biomedical images requires computational expertise that are uncommon among biomedical scientists. Deep learning approaches for image analysis provide an opportunity to develop user-friendly tools for exploratory data analysis. Here, we use the visual programming toolbox Orange (<http://orange.biolab.si>) to

Download PDF



Sections

Figures

References

[Abstract](#)

[Introduction](#)

[Results](#)

[Discussion](#)

[Methods](#)

[Data availability](#)

[Code availability](#)

[References](#)

[Acknowledgements](#)

[Author information](#)

[Ethics declarations](#)

[Additional information](#)

npj | Climate and Atmospheric Science

Call for Papers

Key Concepts

Interactive Visualizations

Workflows

“Lego Bricks” for Data Science



**Democratization of
Data Science**

Easy Access to Data

Reproducibility

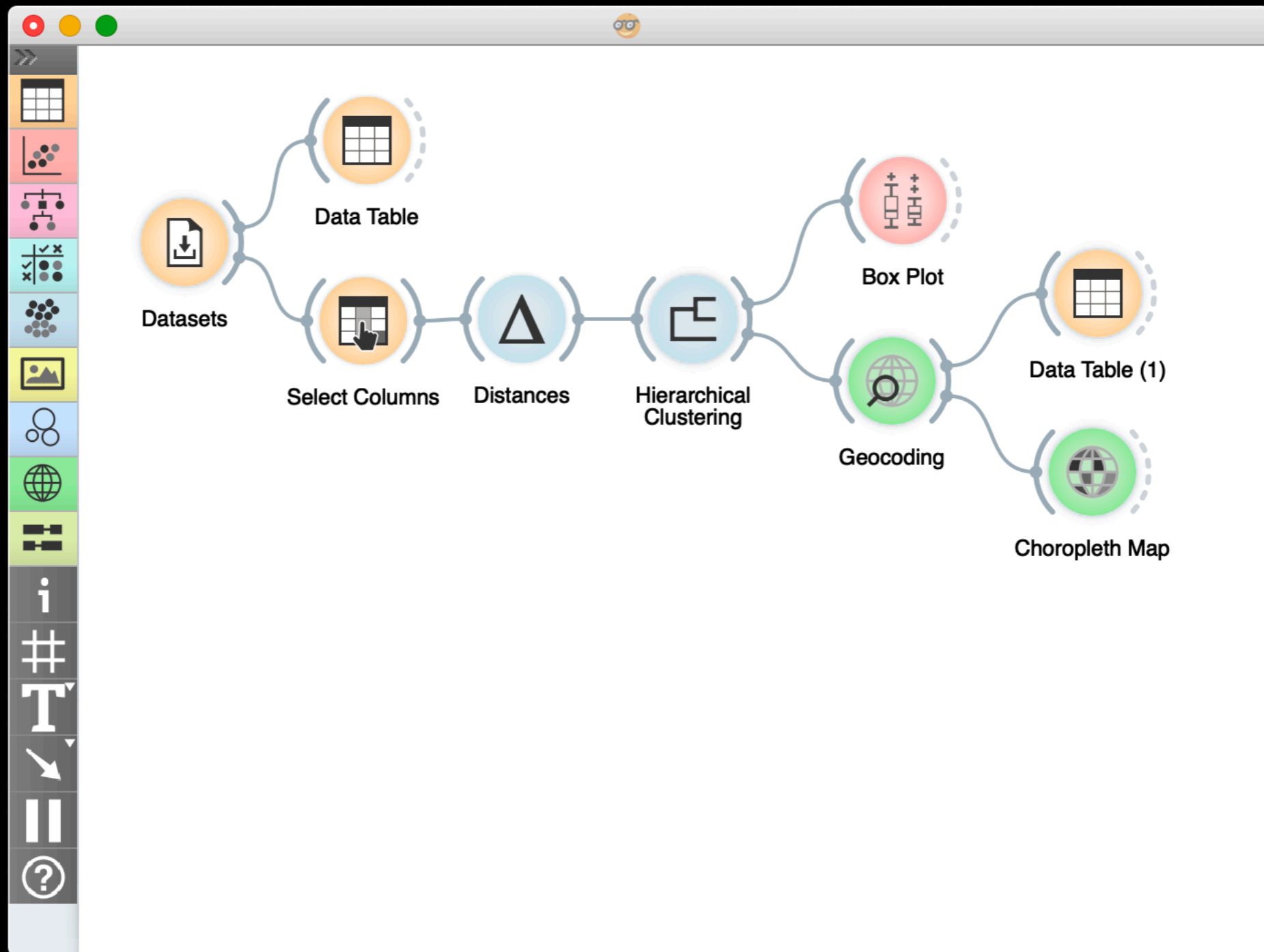
Customization

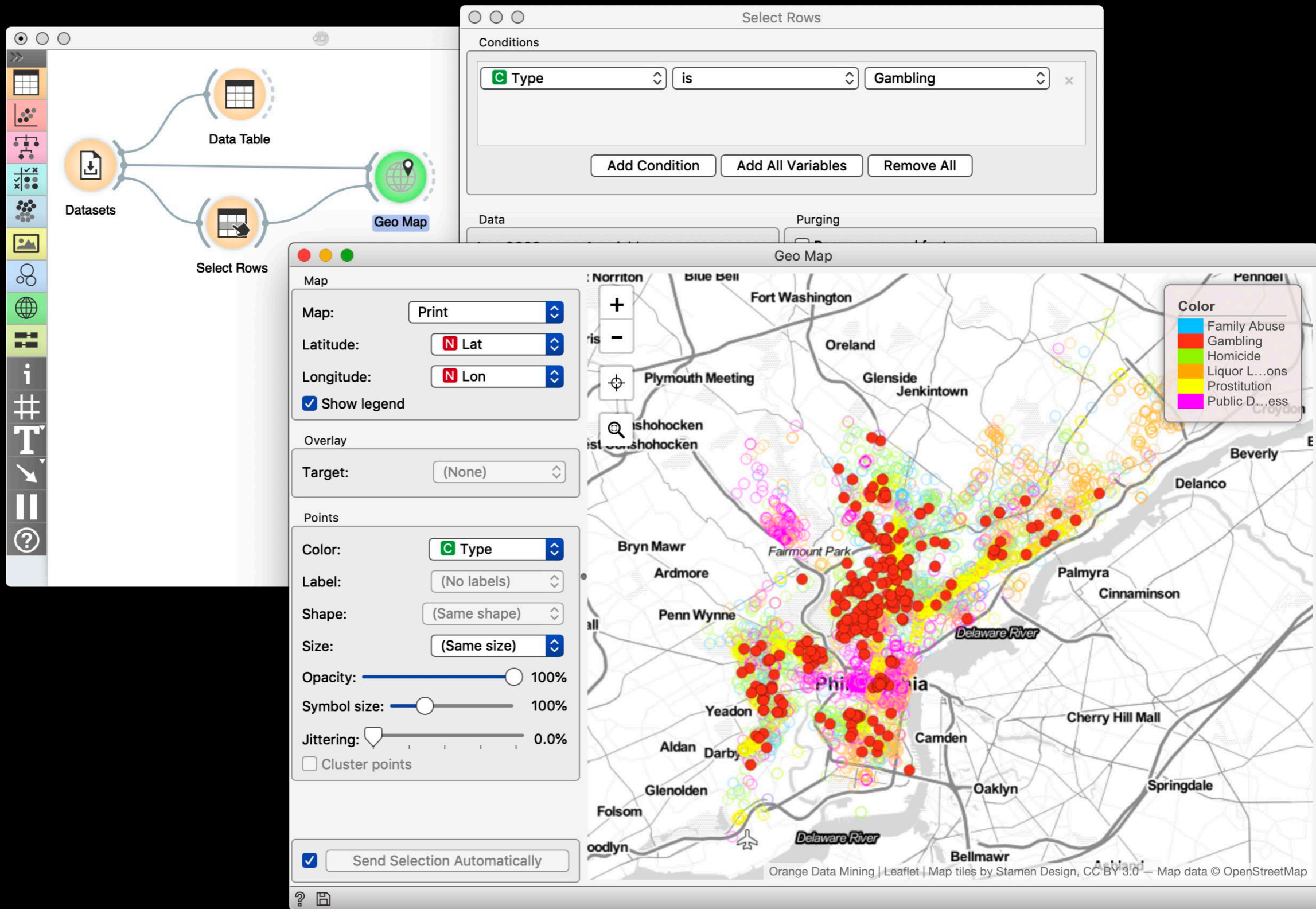
Story Telling

Experimentation

Tools for Training

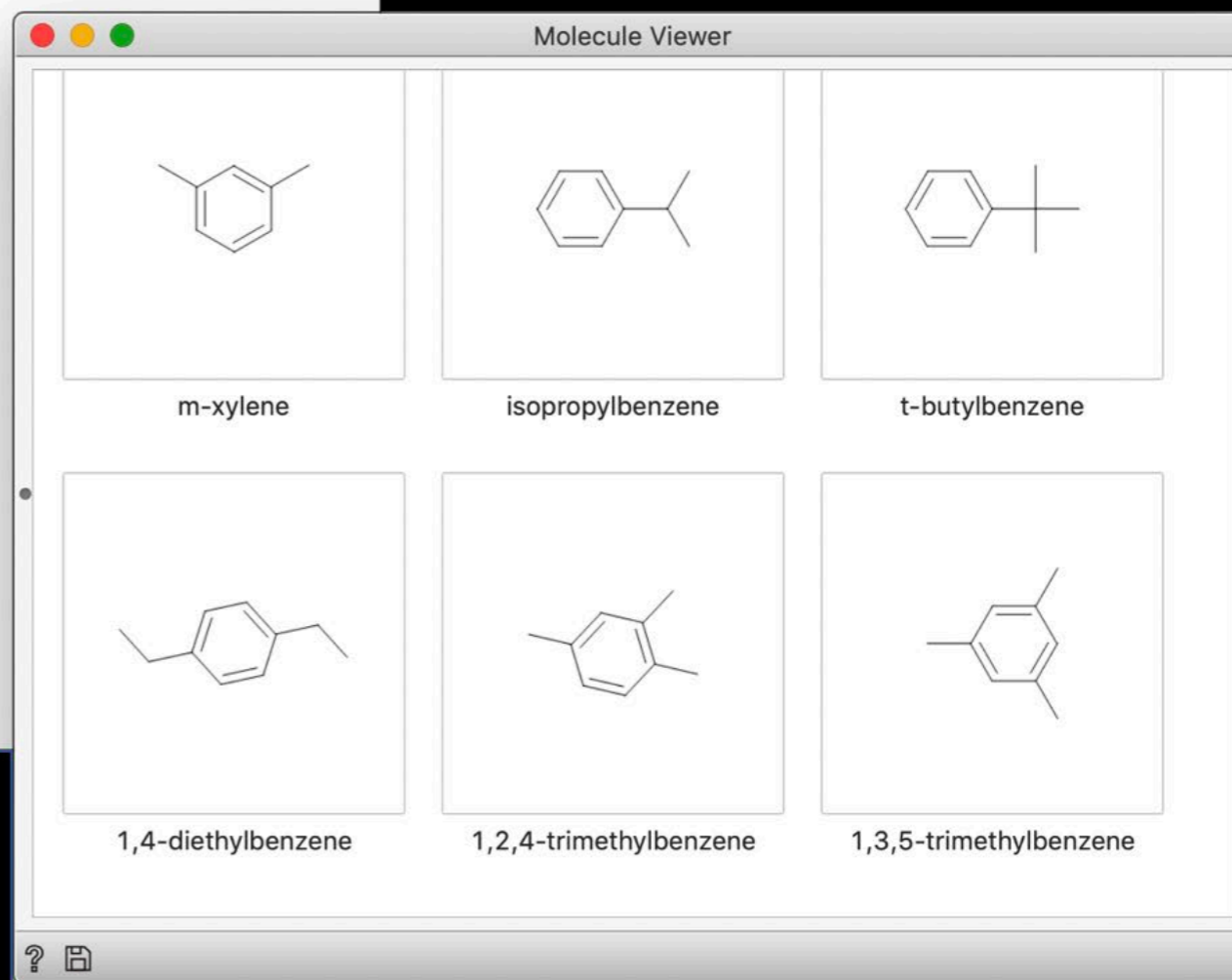
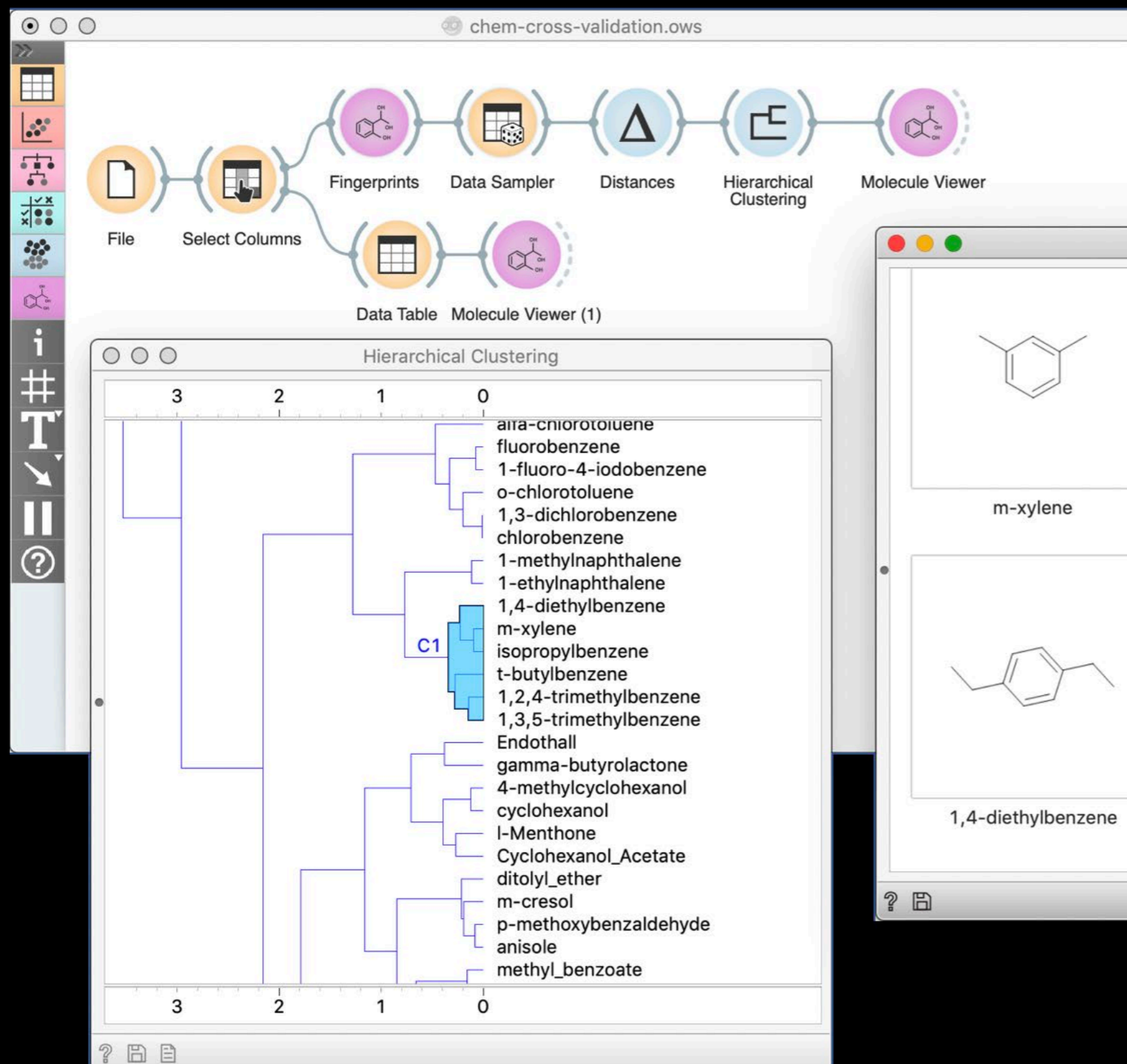
“Lego” Bricks for Data Science

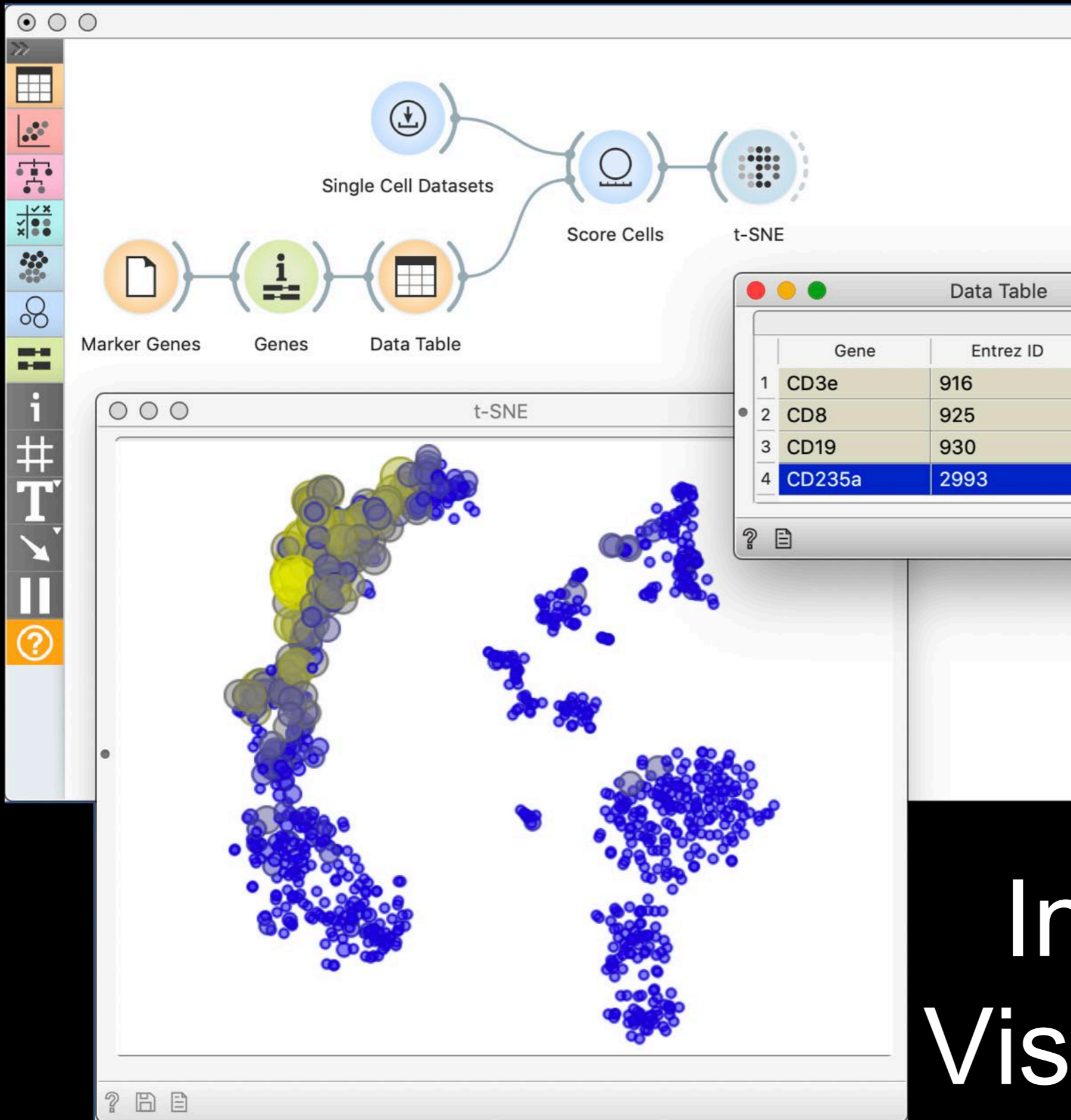




Responsive Workflows

Interactive Visualizations





	Gene	Entrez ID	Cell Type
1	CD3e	916	T Cell
2	CD8	925	T Cell
3	CD19	930	B Cell
4	CD235a	2993	Erythrocyte

Interactive
Visualizations

Data Integration

The interface displays a workflow for data integration:

- Single Cell Datasets**: Input data source.
- Louvain Clustering**: Clustering algorithm.
- t-SNE**: Dimensionality reduction visualization.
- Differential Expression**: Analysis of gene expression changes.
- GO Browser**: Gene ontology enrichment analysis.
- Genes**: List of identified genes.

The **t-SNE** plot shows several distinct clusters of cells, color-coded by cluster.

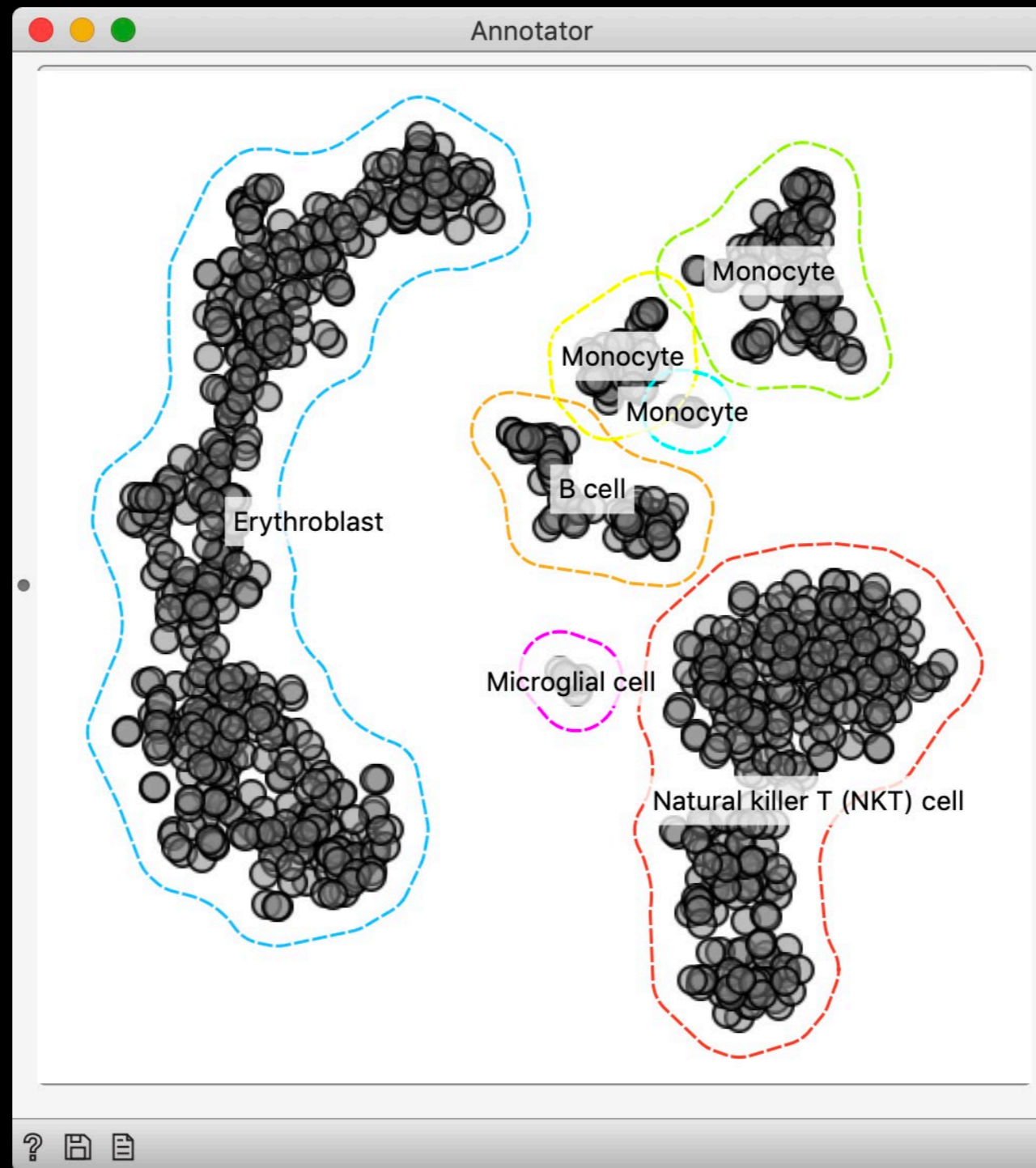
The **Genes** window shows a list of genes with their Entrez IDs and descriptions:

Input ID	Entrez ID	Name	Description	Synonyms
IL1B	3553	IL1B	interleukin 1 beta	IL-1, IL1-BET...
CD74	972	CD74	CD74 molecule	DHLAG, ...
FCER1G	2207	FCER1G	Fc fragment of ...	FCRG
CD14	929	CD14	CD14 molecule	
S100A11	6282	S100A11	S100 calcium ...	HEL-S-43, ...
CFD	1675	CFD	complement ...	ADIPSIN, AD...
MNDA	4332	MNDA	myeloid cell ...	PYHIN3
PETN	56720	PETN	resistin	ADSE, FI773

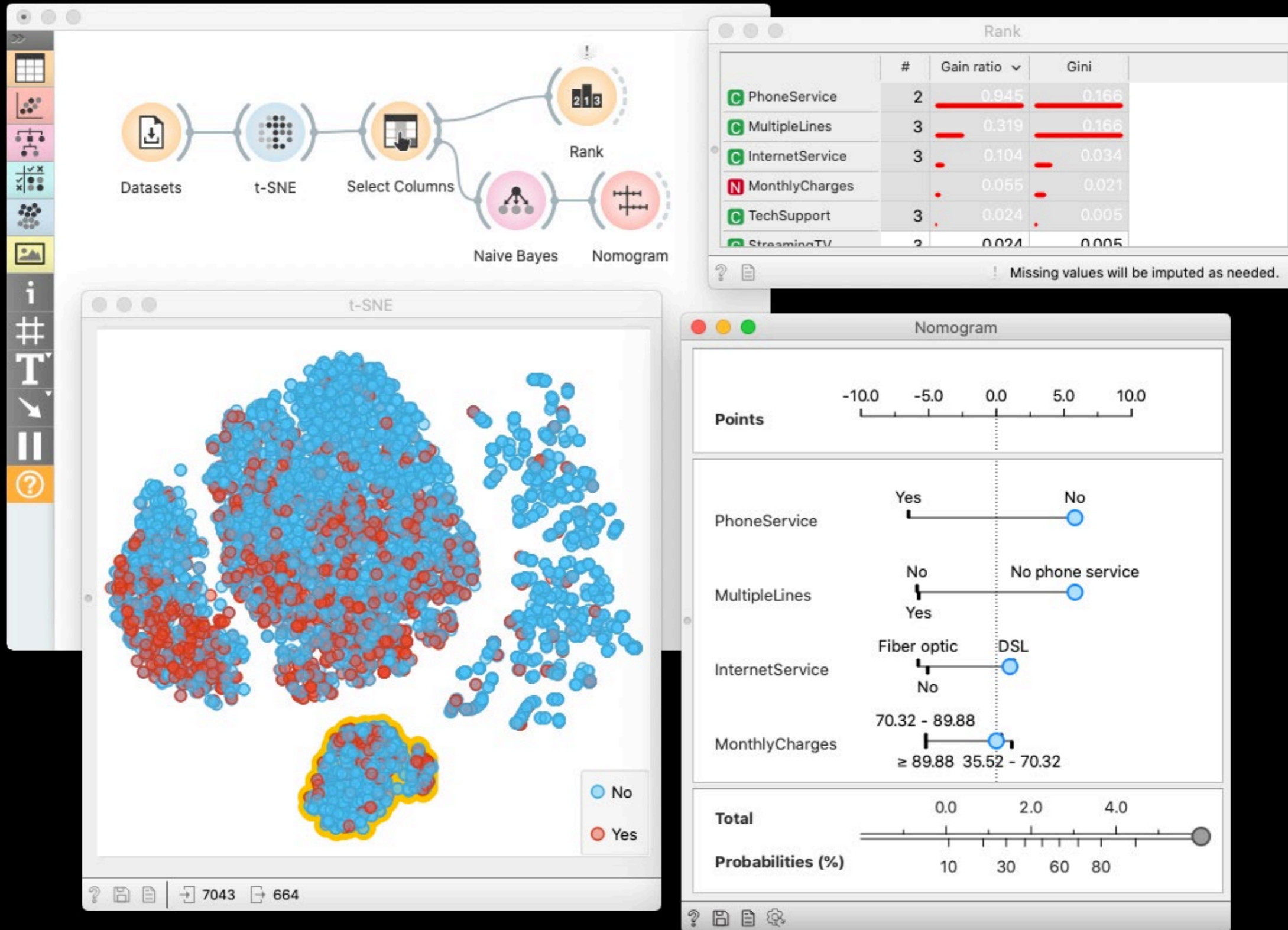
The **GO Browser** window shows the following GO terms and their enrichment statistics:

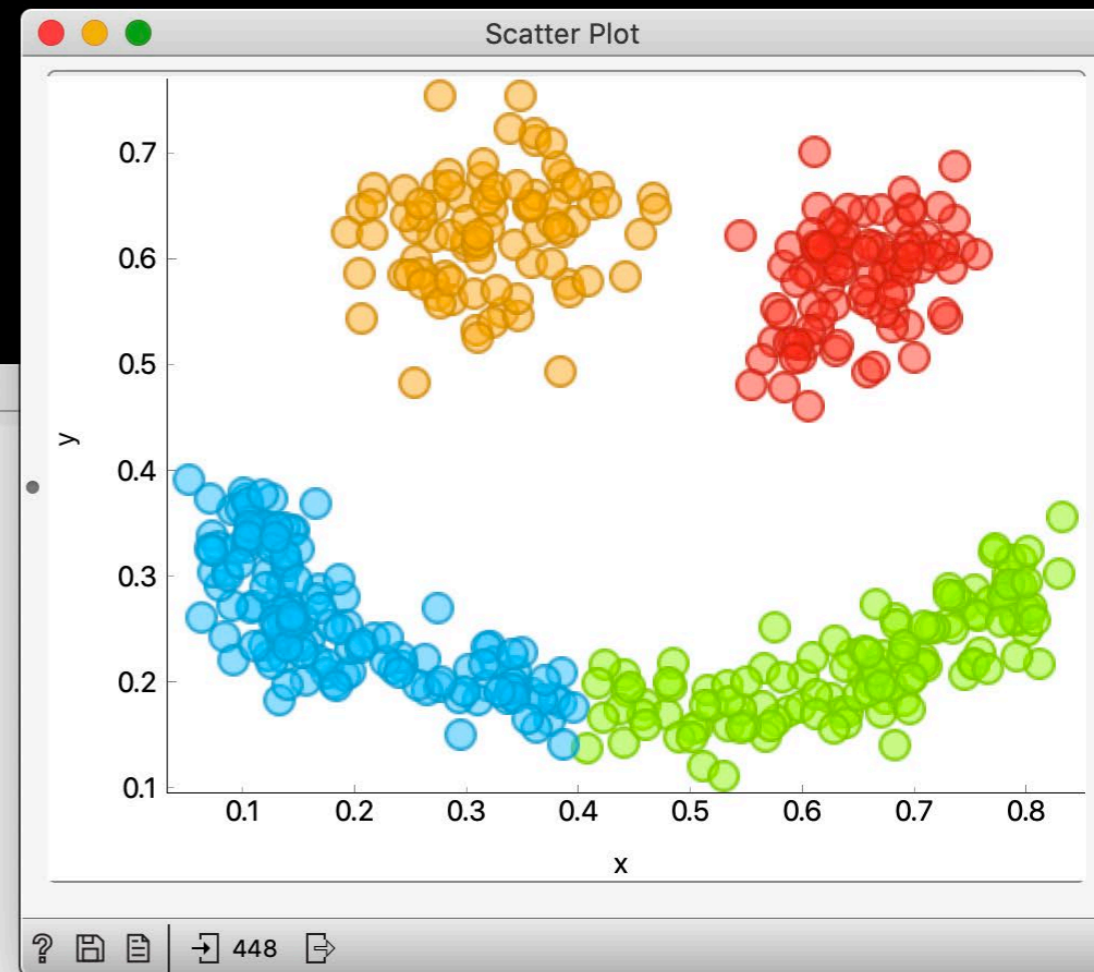
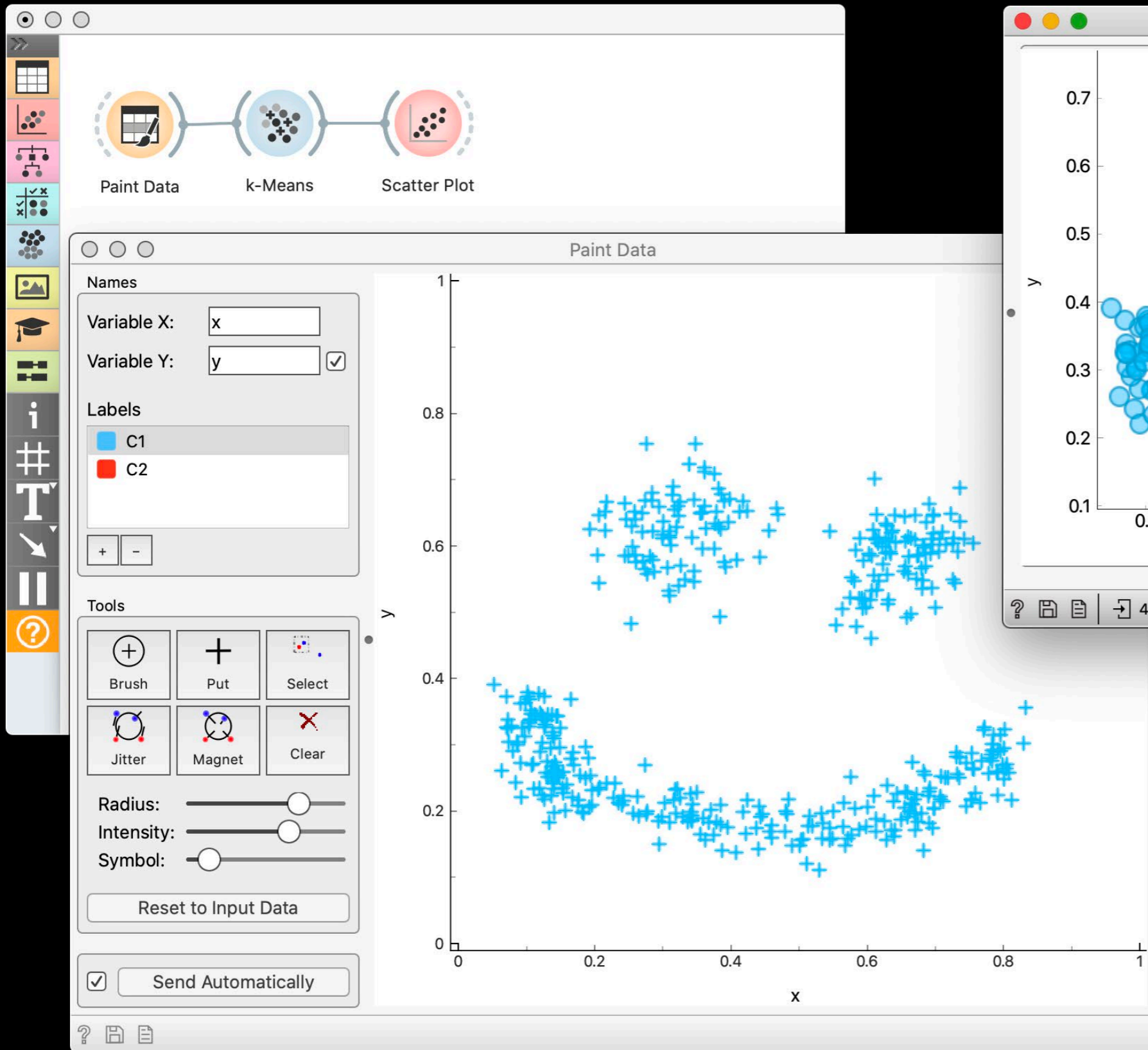
GO term	Cluster	Reference	p-value	Fl
biological_process	85 (94.44%)	17899 (87.36%)	0.02283	0
immune system process	65 (72.22%)	3068 (14.97%)	5.4e-34	1
immune response	55 (61.11%)	2160 (10.54%)	4.5e-31	7
leukocyte activation	41 (45.56%)	1180 (5.76%)	6.4e-27	5
immune effector process	37 (41.11%)	1196 (5.84%)	2.6e-22	6
cell activation involved in immune respon...	30 (33.33%)	690 (3.37%)	6.1e-22	1
leukocyte degranulation	27 (30.00%)	533 (2.60%)	2.3e-21	3
leukocyte mediated immunity	30 (33.33%)	815 (3.98%)	6.3e-20	7
myeloid leukocyte mediated immunity	27 (30.00%)	550 (2.68%)	5.0e-21	6
neutrophil mediated immunity	26 (28.89%)	499 (2.44%)	7.2e-21	8
neutrophil degranulation	26 (28.89%)	485 (2.37%)	3.6e-21	5

Explanation



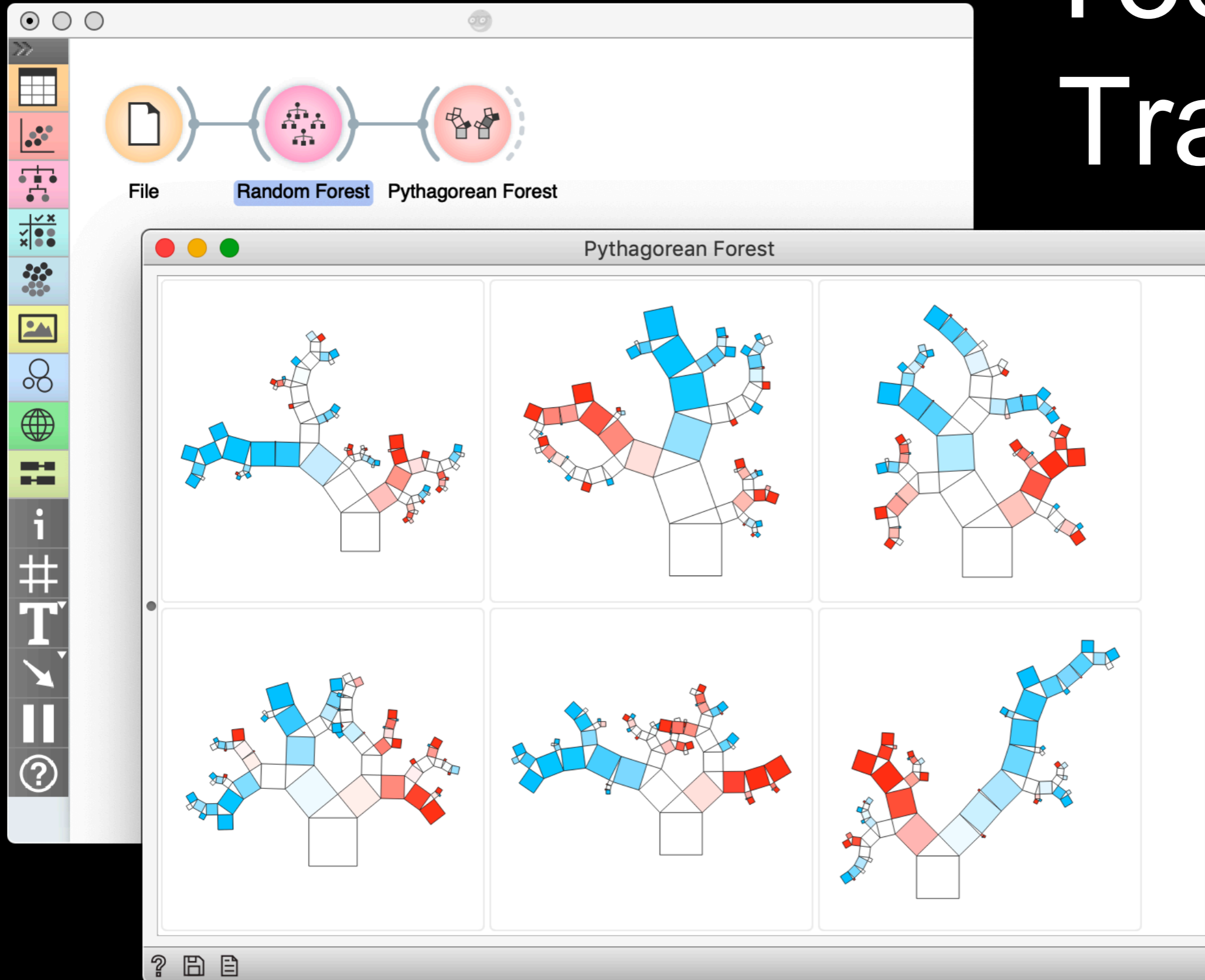
Interactive Explanation





Tools for Training

Tools for Training



Dashboards

**Interactive
Workflows**

Python, R



Simplicity



Flexibility

Training



Data Science Workshop for Public Administration of Slovenia, 2018 & 2019.

Houston



Kolkata



Pavia



YouTube

Search



Getting Started with Orange 01: Welcome to Orange

271,544 views • Dec 21, 2015

790 11 SHARE SAVE ...



Orange Data Mining
12K subscribers

SUBSCRIBED



Open Problems

Tech

Response Time & Interactivity

Servers-Side Compute & Visual Analytics

Data Fusion

Human

Simplicity vs. Flexibility

Gamification

Interpretation vs. Explanation

biolab

Andrej Čopar, GPUs, data fusion

Niko Colnerič, text mining

Tomaž Curk, bioinformatics

Janez Demšar, core development, visualization

Aleš Erjavec, lead developer

Tomaž Hočevar, algorithms, speed-up

Primož Godec, image analytics, text mining

Pavlin Poličar, clustering, embedding, speed-up

Ajda Pretnar, communication, digital humanities

Martin Stražar, bioinformatics

Vesna Tanko, development, testing

Marko Toplak, spectral analysis, bioinformatics

Robert Cvitkovič, bug fixes

Blaž Zupan, enjoys

Lan Žagar, code review, testing

Marinka Žitnik, method expert

external

Nejc Ilenič, deep learning, cloud

Nejc Debevc, web

Jaka Kokošar, bioinformatics

Matjaž Pančur, cloud services

Fabio Ricciato, telecommunications

Veljko Pejović, telecommunications

Anže Starič, software development

support

ARRS Slovenia, NIH USA, MVZT Slovenia

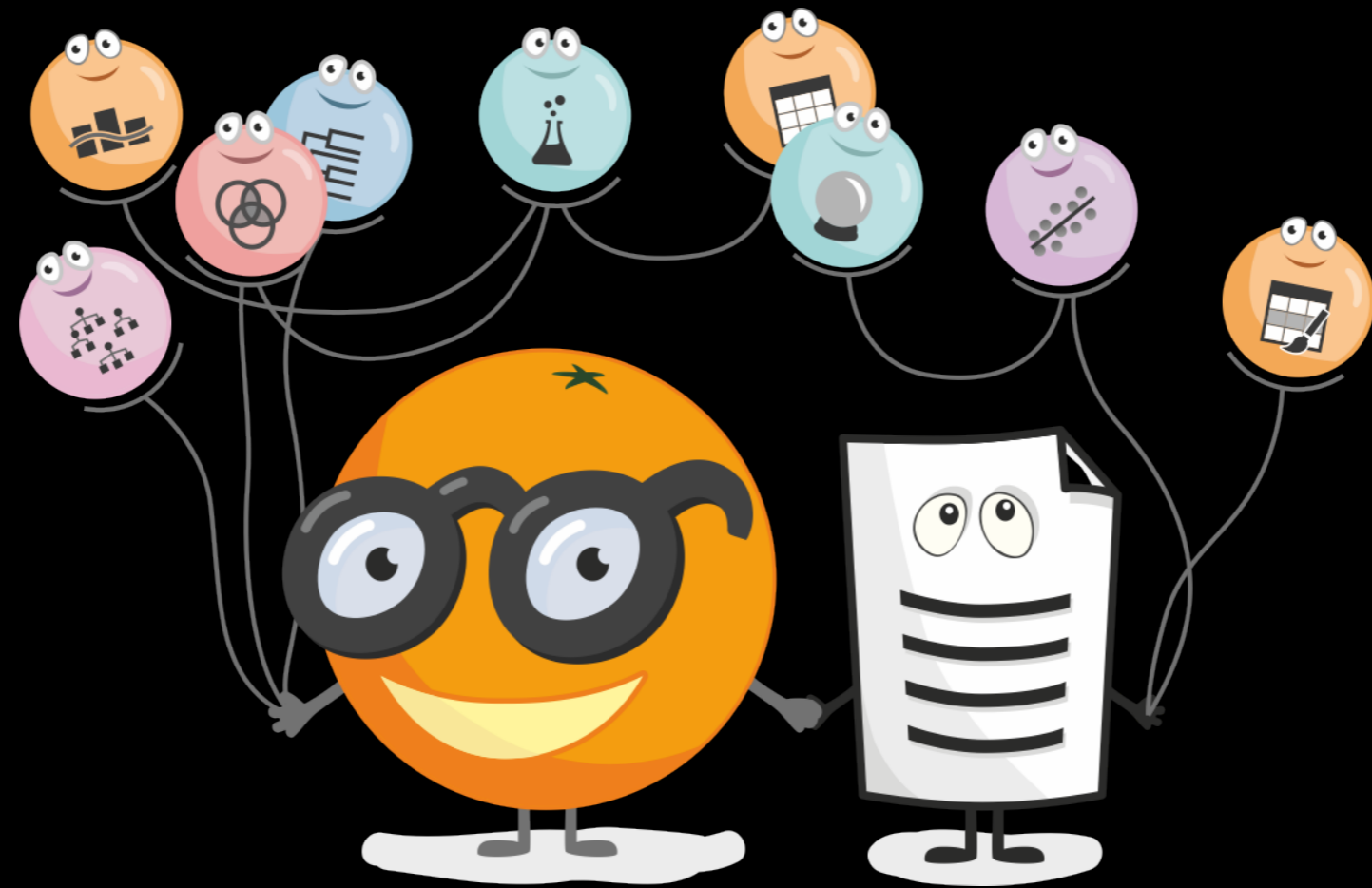
EU FP7 & Horizon 2020

Astra Zeneca, Lek, Elletra, Soleil

Fulbright

Google Summer of Code

NumFOCUS



orange.biolab.si