



ARISTOTLE  
UNIVERSITY OF  
THESSALONIKI



# Workbench data

Andreas Veglis – professor  
School of Journalism & Mass Media  
Aristotle University of Thessaloniki

# The two paths for conducting data journalism

- Based on individual tools (menu driven) – no code skills are necessary.
  - Microsoft Excel, Google Spreadsheet, Tableau, Infogram, Piktochart, DataWrapper, etc.
- Based exclusively on coding (Python, R)
  - Jupiter notebook, Google Colaboratory, etc.



And the  
intermediate  
solution

- Workbench Data.
- Combines features from both approaches.
- It can be employed by journalists with no coding skills and with coding skills.
- It is free.
- It supports group work (share files).
- It supports all data journalism stages (to some degree).

## My workflows

[Create Workflow](#)

Title	Updated <span>↕</span>	Privacy	
Δοκιμαστική συλλογή Tweets	6h ago	private	⋮
Δοκιμές γραφημάτων	8h ago	public	⋮
Παράδειγμα συλλογής Tweets	10d ago	private	⋮
Workbench_visualization_reactors	26d ago	private	⋮
word cloud	30d ago	public	⋮
Άσκηση workbench 2	32d ago	private	⋮
Άσκηση Workbench 1	33d ago	private	⋮
Copy of Άσκηση workbench 2	47d ago	private	⋮
Untitled Workflow	59d ago	private	⋮



# How we are working in Workbench Data



Creation of Workflows.



Each Workflow is comprised of a number of steps.



Every step is usually a certain action (for example apply a filter or delete/move a column).



Ability to see the data in every step of the workflow.



Understanding of the process (all steps are visible, not just the end result).

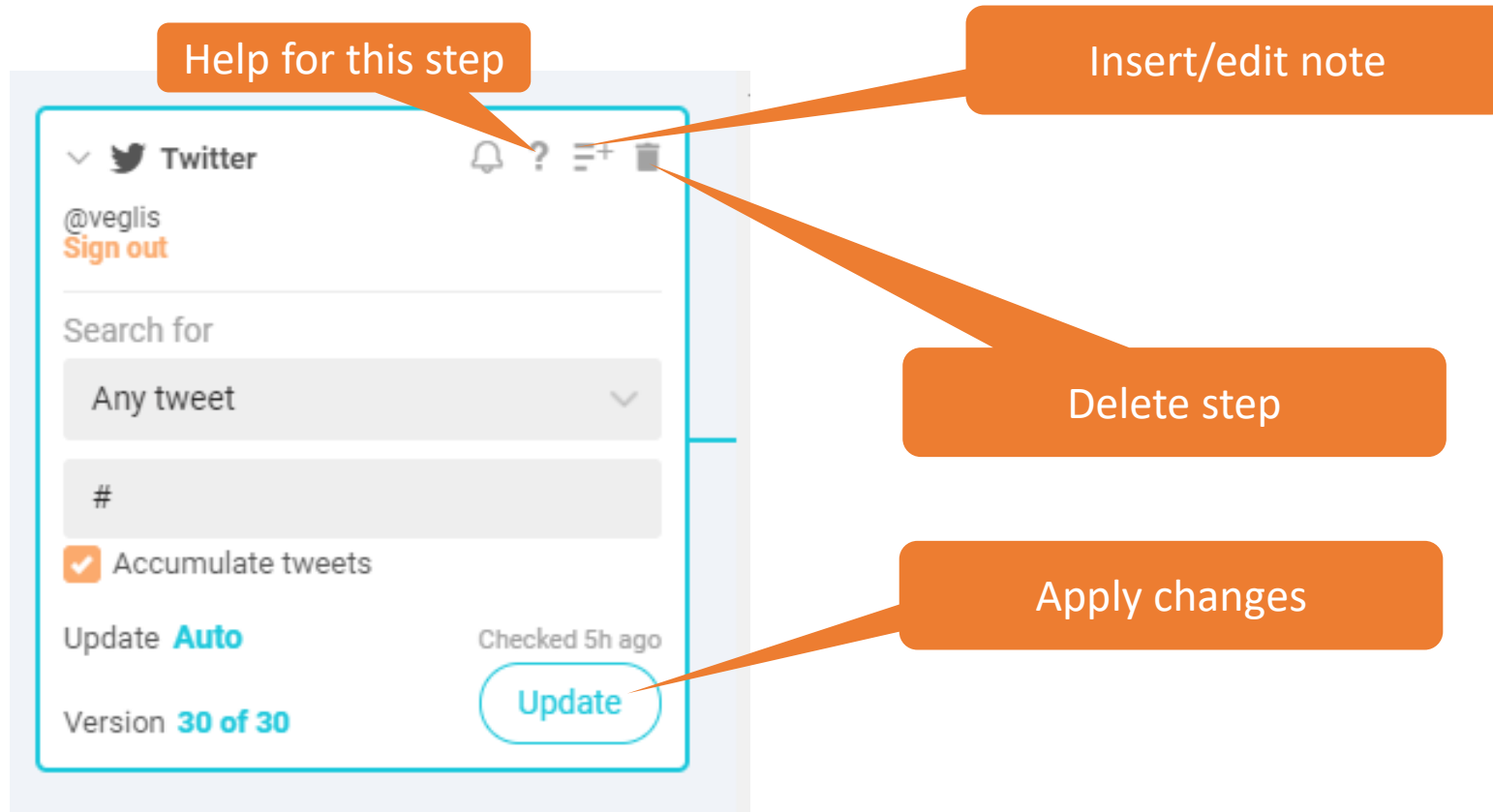


Ability to move/edit/delete steps.



Ability to directly replicate the workflow with different data set (which has the same organization).

# Step features



# Workbench Data characteristics

Large number of Available steps (45).

Accepts data from file (CSV, XLSX, JSON).

Able to perform Scraping.

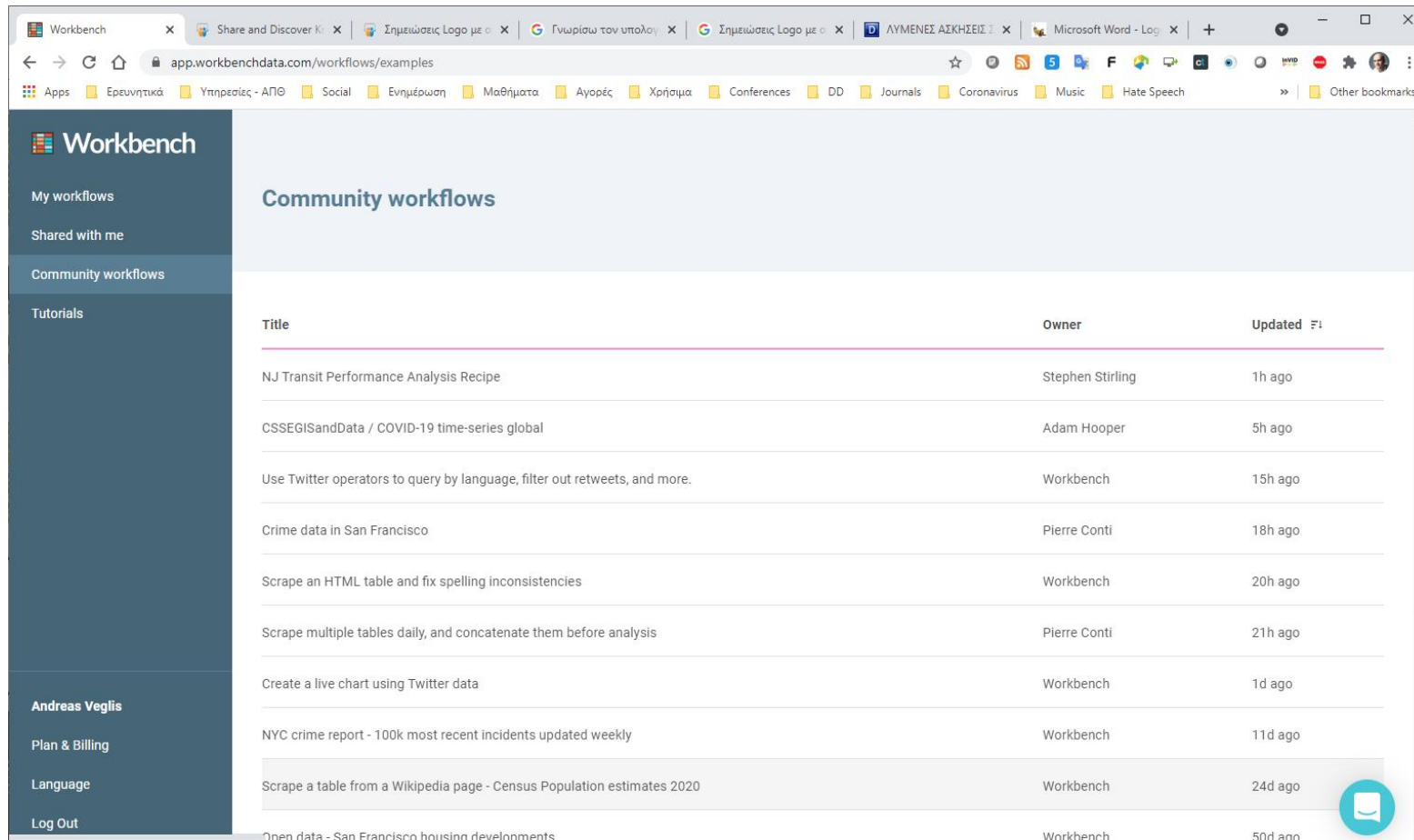
Able to link to Twitter and harvest tweets.

Specialized steps for cleaning data.

Able to create basic visualizations.

Custom steps (accepts Python code).

# Workflows created by other users



The screenshot shows the Workbench web application interface. The browser address bar displays `app.workbenchdata.com/workflows/examples`. The left sidebar contains navigation links: "My workflows", "Shared with me", "Community workflows" (selected), "Tutorials", "Andreas Veglis", "Plan & Billing", "Language", and "Log Out". The main content area is titled "Community workflows" and displays a table of workflows created by other users.

Title	Owner	Updated
NJ Transit Performance Analysis Recipe	Stephen Stirling	1h ago
CSSEGISandData / COVID-19 time-series global	Adam Hooper	5h ago
Use Twitter operators to query by language, filter out retweets, and more.	Workbench	15h ago
Crime data in San Francisco	Pierre Conti	18h ago
Scrape an HTML table and fix spelling inconsistencies	Workbench	20h ago
Scrape multiple tables daily, and concatenate them before analysis	Pierre Conti	21h ago
Create a live chart using Twitter data	Workbench	1d ago
NYC crime report - 100k most recent incidents updated weekly	Workbench	11d ago
Scrape a table from a Wikipedia page - Census Population estimates 2020	Workbench	24d ago
Open data - San Francisco housing developments	Workbench	50d ago



# Tutorials – Workbench basics

The screenshot shows a web browser window with the URL `app.workbenchdata.com/lessons/en`. The page is titled "Workbench basics" and features a sidebar on the left with navigation options: "My workflows", "Shared with me", "Community workflows", "Tutorials" (selected), "Workbench basics" (sub-selected), and "Intro to data journalism". The main content area displays nine tutorial cards arranged in a 3x3 grid. Each card includes a title, a brief description, and a set of icons representing the tools or data sources used in the tutorial.

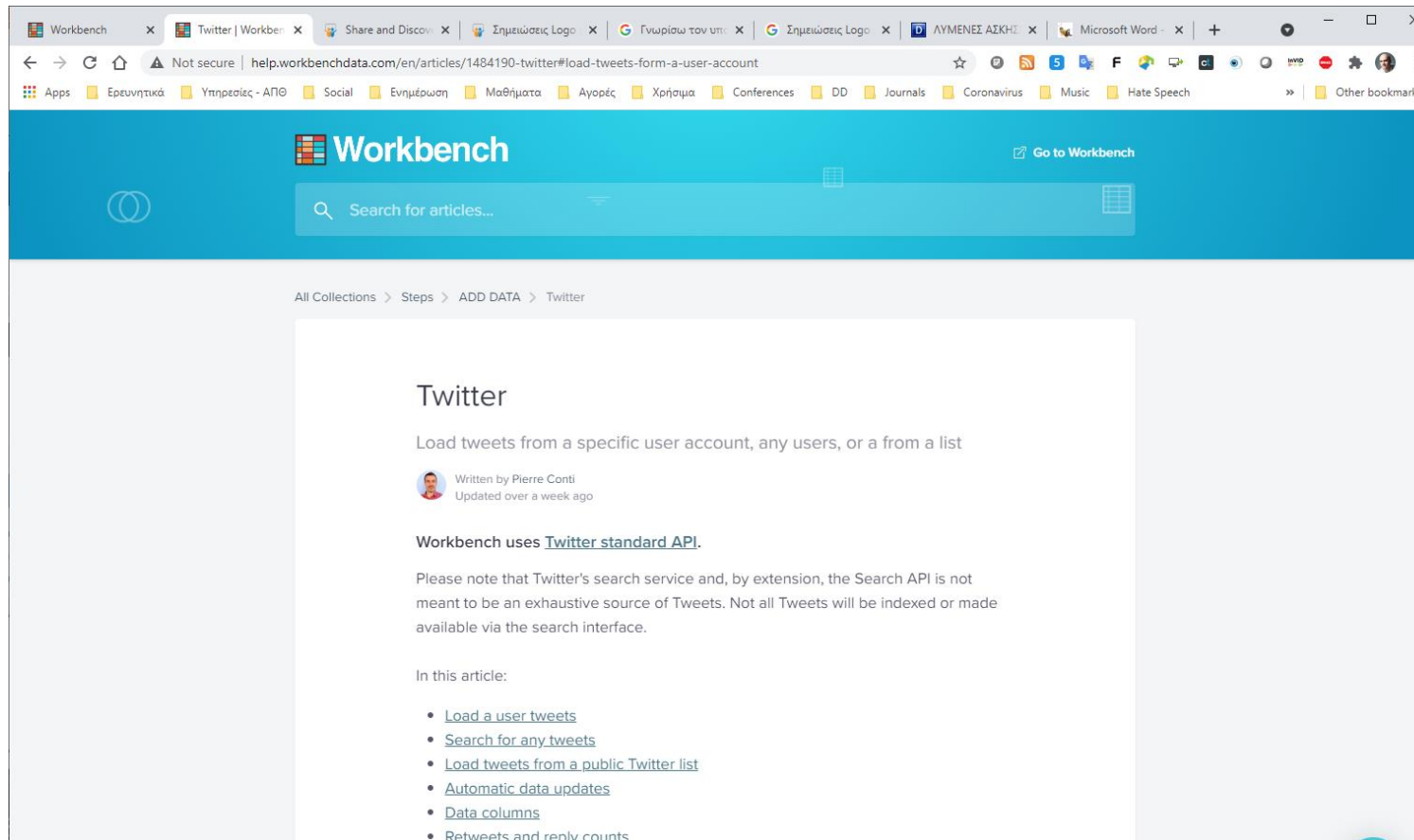
Tutorial Title	Description	Icons
1. Load public data and make a chart	Load data from a public URL and create an embeddable chart	Table, Bar Chart, Line Chart
2. Clean dirty data	Find and fix misspellings and typos in seconds. Format your table to visualize it.	Text Editor, Sparkles, Checkmark
3. Scrape a table from a web page	Load and store HTML tables using the page URL and position	Globe, Magnifying Glass
4. Analyze live data from Twitter	Build a live chart to visualize how often a specific word is tweeted from a specific account	Twitter Bird, Bar Chart, Bell
5. Monitor data and set custom alerts	Monitor a data feed and receive an email notification if a condition is true.	Table, Bell
6. Scrape a webpage to extract or monitor content	Scrape a section of a website on a schedule and receive alerts when its content changes	Globe, Magnifying Glass, Bell
7. Scrape multiple pages on one website	Scrape a list or table that spans multiple pages.	Globe, Magnifying Glass, Document
8. Make a chart of police stops by race.	Analyze data from 300k traffic stops in Oklahoma City, and make a chart by race.	Table, Bar Chart, Line Chart
9. Cities with the highest PM pollution	Examine particle pollution in U.S. cities by analyzing EPA air quality data.	Table, Sparkles, Bar Chart

# Intro to Data Journalism

The screenshot displays the Workbench application interface. At the top, a navigation bar includes a search bar and a list of categories: Apps, Ερευνητικά, Υπηρεσίες - ΑΠΘ, Social, Ενημέρωση, Μαθήματα, Αγορές, Χρήσιμα, Conferences, DD, Journals, Coronavirus, Music, Hate Speech, and Other bookmarks. The left sidebar features the Workbench logo and a menu with options: My workflows, Shared with me, Community workflows, Tutorials (with sub-items: Workbench basics and Intro to data journalism), and a user profile section for Andreas Veglis (Plan & Billing, Language, Log Out). The main content area is a grid of 12 tutorial cards, each with a title, description, and a set of icons representing the data operations involved.

Tutorial Title	Description	Icons
First story	Learn about the first questions you should ask your data before starting your analysis, and create your first chart.	Grid, Plus, Formula, Bar Chart
Finding the largest CO2 producers	Work with international CO <sub>2</sub> emission data and learn how to find the largest and smallest values in your data.	Grid, Plus, Double Arrow, Filter
Column and line charts	Learn when to use different types of charts, and how to transpose to prepare data for charting.	Bar Chart, Plus, Double Arrow, Line Chart
Filtering data about public housing	Filtering is a powerful analysis technique. Learn how to filter by condition and by value to explore your data.	Grid, Plus, Filter, Checkmark
Grouping rows (pivot tables)	Group rows together in order to summarize large amounts of data quickly.	Grid, Plus, Pivot Table, Bar Chart
Cleaning and analyzing campaign finance data	How much money came from out of state in New York governor 2014 re-election campaign?	Grid, Plus, Double Arrow, Sparkle
Introduction to data types	Understanding data type is essential. Some operations only work on specific types.	Grid, Plus, Text
Grouping by dates	Dates can be used to group events in time frames: Per hour, day, month, quarter..	Double Arrow, Plus, Sparkle, Line Chart
Long and wide format	Learn the difference between "long" and "wide" table formats and how to convert between them.	Grid, Plus, Clock, Line Chart
Calculating per capita crime rates in the U.S.	Many numbers are best interpreted by comparing them to the overall	
Comparing numbers	Learn three common types of comparisons and how to calculate them: difference, ratio, and	
Joining tables to adjust teacher salaries for inflation	Use joins and the Consumer Price Index to compare dollar amounts	

# Help available for every step



The screenshot shows a web browser window with multiple tabs. The active tab is titled "Twitter | Workbench" and shows the URL `help.workbenchdata.com/en/articles/1484190-twitter#load-tweets-form-a-user-account`. The page has a blue header with the "Workbench" logo and a search bar. Below the header, a breadcrumb trail reads "All Collections > Steps > ADD DATA > Twitter". The main content area is titled "Twitter" and describes how to load tweets from a specific user account. It includes a byline "Written by Pierre Conti" and "Updated over a week ago". A note states "Workbench uses [Twitter standard API](#)". A disclaimer follows: "Please note that Twitter's search service and, by extension, the Search API is not meant to be an exhaustive source of Tweets. Not all Tweets will be indexed or made available via the search interface." The article concludes with "In this article:" followed by a bulleted list of links: "Load a user tweets", "Search for any tweets", "Load tweets from a public Twitter list", "Automatic data updates", "Data columns", and "Retweets and reply counts".

Workbench

Go to Workbench

Search for articles...

All Collections > Steps > ADD DATA > Twitter

## Twitter

Load tweets from a specific user account, any users, or a from a list

Written by Pierre Conti  
Updated over a week ago

Workbench uses [Twitter standard API](#).

Please note that Twitter's search service and, by extension, the Search API is not meant to be an exhaustive source of Tweets. Not all Tweets will be indexed or made available via the search interface.

In this article:

- [Load a user tweets](#)
- [Search for any tweets](#)
- [Load tweets from a public Twitter list](#)
- [Automatic data updates](#)
- [Data columns](#)
- [Retweets and reply counts](#)



# Resources

---

- <https://app.workbenchdata.com/>
- <https://source.opennews.org/articles/workbench-data-journalism-open-source-platform/>



Thank you for  
your attention

- E-mail: [veglis@jour.auth.gr](mailto:veglis@jour.auth.gr)
- Webpage: <http://veglis.webpages.auth.gr>
- <https://auth.academia.edu/AndreasVeglis>
- <https://www.linkedin.com/in/veglis/>
- <https://scholar.google.gr/citations?user=Q2MCvhoAAAAJ>
- <https://publons.com/researcher/1654694/andreas-veglis/>
- Twitter: @veglis



ARISTOTLE  
UNIVERSITY OF  
THESSALONIKI

