

TOOLBOX

SCIENCE ON THE MAP

Easy-to-use mapping tools give researchers the power to create beautiful visualizations of geographic data.

ILLUSTRATION BY THE PROJECT TWINS



BY MARK ZASTROW

When linguist Lauren Gawne roams the valleys of Nepal documenting endangered Tibetan languages, she takes pains to distinguish each dialect's geographical origin. But when it came to producing maps of her results, for many years her cartographic methods were somewhat crude.

"My old maps were [made] using MS Paint on top of some copyrighted map that I really shouldn't have been using," she says. Her next solution wasn't much better: "My mum

tracing a map off an atlas so that I had something a bit cleaner to work with." The one after that — "using Google Earth and dropping pins on it" — was generic, ugly and "looks horrible in a PowerPoint".

So in 2013, she jumped at the chance to join a workshop on mapping and visualization at the University of Melbourne in Australia, where she was working on her PhD. There she discovered the free, open-source program TileMill, created

➔ NATURE.COM
For more on scientific software, apps and online tools, visit: nature.com/toolbox

by the company Mapbox, which has offices in San Francisco, California, and in Washington DC. It lets users create maps from their data and pre-existing online cartographic databases.

TileMill is just one tool in the emerging field of customized mapping, where a bevy of open-source technologies and start-ups have given rise to an abundance of offerings for researchers and enthusiasts. These tools are more approachable for novices than the conventional geographic information systems (GISs) that geographers have long used for analysis of geospatial data sets. They allow non-specialists ▶

► to easily visualize, manipulate and share their data in formats that are as slickly browsable as Google Maps but with greater power and flexibility. “TileMill allows you to be a complete control freak,” says Gawne, now at Nanyang Technological University in Singapore. From line styles to font spacing and kerning, “I can really manipulate all the variables quite easily.”

Until recently, Google, which is based in Mountain View, California, itself had staked the biggest claim in this space, providing various ways to access and decorate its maps through application programming interfaces (APIs). But as demand grew, the tech giant began limiting public access to its APIs in 2011 — and this allowed slightly more sophisticated open-source tools to flourish, says Oliver O’Brien, a geographer at University College London. Today, a fully fledged ecosystem of start-ups with open-source technology at their core offer platforms that many say have surpassed Google’s offerings.

“Google really nailed down having maps on the web,” says Javier de la Torre, a founder and current chief executive of one of Google’s emerging rivals, CartoDB of New York City. “What I think they didn’t see coming was that there was going to be this explosion of new map-makers.”

THE NEW MAPPING LANDSCAPE

In 2011, de la Torre was part of a team researching biodiversity informatics. The group was seeking an online platform to make a map of all known species on the planet. “There wasn’t technology for doing that,” he says — no tool could handle the amount of data, nor visualize how they changed over time.

The researchers decided to develop the tool themselves and created what became the open-source platform CartoDB. The company offers free and paid plans for hosting and visualizing data through its website. Unlike TileMill, which is primarily intended for drawing and designing static maps, CartoDB specializes in visualizing dynamic layers of data on top of basemaps. Users can import their geo-located data into CartoDB’s web-based interface and then filter or cluster data points, change the colour or size of symbols, and animate data changes over time. “CartoDB wants to be a place where your data lives,” says Steve Bennett, a research-oriented technologist at the University of Melbourne who takes workshops on mapping, including the one that Gawne attended.

Peter Desmet, who collaborates with a bird-tracking research team at the Research Institute for Nature and Forest in Brussels, was a colleague of de la Torre and became an early adopter of CartoDB. “I was never a desktop GIS person,” he says. But in CartoDB, “you can create and share a visualization in literally minutes.” Being able to simply send a link to the map online also makes it much faster to point out data-quality issues to colleagues, he says.

Another strength of CartoDB is its selection

MORE ONLINE



For more on scientific mapping, including a gallery and links to tools, see go.nature.com/u76knj

of global basemaps — ranging from familiar geopolitical and satellite-image formats to more stylish black-and-white and even pencil-and-watercolour-themed renditions. Some are produced by TileMill’s maker Mapbox, which boasts a growing list of corporate and media clients — in many cases supplanting Google in a growing ‘battle of the basemaps’.

Mapbox first released TileMill in 2011. The team took a powerful but complex open-source cartographic renderer called Mapnik, built an easy-to-use interface around it and created a simple styling language, CartoCSS, to customize the maps’ appearance.

“TileMill was a game-changer, absolutely,” says Bennett. It allowed non-experts to produce professional-looking maps — either for publication as static figures or for use as basemaps in other visualization tools — without the need for more-complicated GIS programs.

The landscape continues to shift rapidly. In January, Google announced that it would shut down some premium and paid forms of Google Maps and focus on its basic Maps API. In response, CartoDB introduced tools to help users migrate their data to CartoDB, while still allowing them to integrate the Google Maps APIs. Mapbox, for its part, has shifted development from TileMill to its intended replacement, Mapbox Studio.

Cost of data storage is a potential stumbling block for scientists with large data sets — although CartoDB is open source, its convenience comes in large part from using it on the company’s hosted web service. The firm offers 75 megabytes of storage for free, but to store more than 1 gigabyte of data, the price rises quickly to hundreds of US dollars per month. CartoDB also charges to keep data and maps private on the site. “We’ve had real problems,” says Bennett. “If you’re a PhD student with no funding, it just doesn’t work.” Mapbox works with a similar pricing model for hosting maps on its servers, although TileMill itself is a free, downloadable program. However, CartoDB does work with academic users to try to find a solution, says de la Torre, and awards grants of up to US\$3,500 to researchers studying the impacts of climate change, in recognition of the company’s environmental roots.

Power users can daisy-chain these tools together: for example, one could create a

basemap in TileMill and data layers in CartoDB, then wrap them in an online interface using Leaflet, a mobile-friendly visualization package that runs in the program JavaScript and meshes with other JavaScript visualization packages such as D3. Duncan Smith, a geographer at University College London, has made one such combination: an online map of UK census data called LuminoCity that uses Leaflet to display the map data over basemaps produced in TileMill, and a variant of D3 called Dimple to show graphs of the data onscreen.

STORAGE HUBS

Researchers can also store their data sets in a CartoDB account, then access them (using the ubiquitous SQL database language) for other online applications, notes Desmet. For one project, he used D3 to build a map depicting radar observations of bird migration as wind-like flowing curves. The source code is stored in the repository GitHub, but the map pulls the scientific data from his CartoDB account.

Despite the visual sophistication of these tools, the level of computational analysis they provide is limited. But after using these programs to get to grips with the basic principles, researchers can progress to more-powerful GIS platforms. Many scientists — including those involved in public policy, such as urban planning and crisis mapping — use arcGIS, a suite of products maintained by Esri, based in Redlands, California. But there is also an open-source alternative: QGIS, a project of the Open Source Geospatial Foundation.

Researchers who already write code as part of their work can use programming languages such as Python and R, which already have capable mapping packages that users may not even be aware of, points out astronomer James Davenport of the University of Washington in Seattle. He says that astronomers often “end up bastardizing scientific visualization software to make maps”. He now uses the Python package matplotlib in tandem with the rest of his Python-based analysis to project his infrared observations onto maps of the sky.

Even researchers who would rather not touch a line of code can accomplish a lot with the help of CartoDB and TileMill. “You don’t have to be particularly technically competent,” says Gawne, who produced the Tibetan-language maps for her thesis in TileMill and now teaches mapping workshops herself. “You have to be not afraid to try it.” ■

Mark Zastrow is a science writer in Seoul. He reported this article from Washington DC.

CLARIFICATION

The Toolbox story ‘Adventures with R’ (*Nature* **517**, 109–110; 2015) did not make clear what stopped Rabih Murr from practising R — he was preparing a paper for publication.