

MAP-MAKING ON A BUDGET

Cut-price cartography tools are making light work of map-making.

ILLUSTRATION BY THE PROJECT TWINS



BY JEFFREY M. PERKEL

Glaciologist Kenichi Matsuoka was leading a team across an Antarctic ice sheet in 2005 when their crucial mapping software cut out.

Matsuoka relied on a commercial geographic information system (GIS) to review data and plan excursions on the remote ice. But amid his travel preparations, Matsuoka, then at the University of Washington, Seattle, had forgotten to renew the software license. “It was a disaster,” he says.

Or nearly so. The team also had a smattering of other tools on their laptops, which they used to cobble together a solution. “We managed,” says Matsuoka, now at the Norwegian Polar Institute in Tromsø, Norway. The team went on to develop a free, self-contained and open-source Antarctic-mapping resource called Quantarctica, which today has several hundred users, according to George Roth, who coordinates the project.

Maps are essential across a wide swathe of science, from ecology and anthropology

to sociology and climatology, and today’s researchers have a rich variety of inexpensive or free tools to choose from. They range from full-blown desktop GIS packages and cloud-based portals to libraries for scientists who code in the R and Python languages. Researchers can use them to chart their study locations, integrate multiple datasets and detect spatial relationships that otherwise would be hidden. But map-making is a subtle science, and the learning curve can be steep.

Mikel Maron, who leads community outreach at Mapbox, a mapping-services company in San Francisco, California, says that maps “can tell very good stories”. With such a rich and growing toolset, researchers are finding it easier than ever to tell these tales.

MECHANISMS FOR MAPPING

Oliver Gruebner, for instance, is a health geographer at Humboldt University, Berlin, who studies post-disaster mental health. He applied a ‘sentiment detection’ algorithm to Twitter posts that included location data (‘geotagged’ tweets) following such events as

the landfall of Hurricane Sandy in New York in 2012, and the 2015 terrorist attack in Paris. This let him identify regional clusters of emotional trauma — a finding that could help in deploying limited mental-health resources.

“It’s not surprising that we see something,” Gruebner says. “But the good thing is that we can actually measure these things.”

Gruebner identified those clusters using a free spatiotemporal statistical-analysis package called SaTScan, and mapped them using the open-source desktop tool, QGIS. “QGIS is free and it’s updated continuously, and it has great functionality,” he says. “And for most of the things you want to map, it’s pretty self-sufficient and efficient.”

But coming to grips with QGIS takes time, and easier alternatives exist. With nothing more than a smartphone, researchers can capture geotagged photos of their study sites — a feature that is enabled by default on many smartphones. They can then plot, style, and share maps of those data using cloud-based tools such as Google Maps, Mapbox Studio or ArcGIS Online (the latter from Esri, ▶

► a mapping-tools company in Redlands, California), as well as R and Python (see ‘Mapping in R’).

Mapbox Studio, for instance, provides exquisite control of a map’s appearance, whereas Google Maps is all about simplicity. Esri’s Story Maps tool focuses on the audience’s experience. The tool allows users to create and publish online documents that integrate spatial data with text, video and images, extracting location information from photos if available. Story Maps team member Owen Evans says that scientists could also create supplementary online resources using the tool. Users can even design their narrative to highlight different map features as the reader navigates the story. In one published example, researchers at the US Fish and Wildlife Service charted the locations of fish hatcheries across the Pacific Northwest before focusing the map, and the story, on one hatchery in particular.

Anita Graser, a spatial-data scientist at the Austrian Institute of Technology, in Vienna, and a member of the QGIS steering committee, says that the challenge comes when researchers attempt more detailed analyses than simply plotting points. Many mapping applications can calculate distance and area, for instance, but QGIS has plug-ins for tasks such as classifying land coverage (using categories such as forest or desert), measuring ground slope, calculating travel times, and modelling the path of flowing water. Programmers can perform a range of similar analyses in R and Python.

“The first step is creating the picture, but eventually you want to get your hands dirty with the analysis and get those numbers that you need for your papers,” Graser says.

One complication, says Michele Tobias, a GIS data curator at the library of the University of California, Davis (UCD), is that mapping data can be represented using any of several projections. A projection is “the math that translates from latitude and longitude into a flat thing like a map or a computer screen”, Tobias explains. If different data sets use different projections, overlaying multiple datasets will result in locations not matching up.

Mapping data can also come in two basic forms, says Sergio Rey, a geographic-information scientist at the University of California, Riverside. In vector files, the world is populated with discrete objects — polygons representing roads, buildings and political boundaries, for instance. Raster files are used to model continuous data, such as maps of rainfall or elevation. Those two forms use different file types, and the GDAL (Geospatial Data Abstraction Library) project has developed free tools that can read and write popular formats. An online tool called GeoJSON.io allows users to interactively create, manipulate and export GeoJSON files.

Shanan Peters, a geologist at the University

of Wisconsin–Madison, is the principal investigator on the Macrostrat project, which is an online encyclopaedic atlas for geological data. Although most of the Macrostrat mapping data are publicly available, importing them required “a fair bit of time”, Peters says. The files needed to be converted to a single vector format, modified to use a common vocabulary and checked for accuracy. “A lot of these maps actually come with small geometry errors from the publisher,” Peters says.

The Macrostrat team relied mostly on two tools. QGIS converted the different input files to a single format, and PostGIS enabled storage and analysis of the data set. Peters says that PostGIS, an extension that adds geospatial capabilities to the open-source database system PostgreSQL, “basically turns a relational database into a full-blown GIS”. And, Tobias notes, it does so while avoiding the computational overhead required to actually draw a map — a process that can be computationally intensive.

Nistara Randhawa, a veterinarian-turned-PhD-candidate also at UCD, used ArcGIS to build a network of population centres and roads, which she then exported into R to model an influenza outbreak in Rwanda. Buoyed by her model’s fidelity to real-world observations, Randhawa scaled up her analysis to encompass western Africa. But as the network ballooned from 1,300 locations to 17,000, the graphical interface froze up. So, working with Tobias and Alex Mandel, a geospatial scientist at UCD, Randhawa tested a range of tools and opted for GrassGIS, an open-source GIS that she could control by issuing text-based instructions without the bother of drawing the map. “It’s enabled me to programmatically create my network,” she says. (ArcGIS and QGIS have Python interfaces that enable similar functionality.)

Indeed, for an ever-larger number of researchers, programming tools provide an attractive alternative to desktop tools, says Rey. “They’re just a lot more flexible.”

They also foster reproducibility, because researchers can repeat runs while specifying exactly which version of the code to use. Coding also allows you to use the most up-to-date algorithms, which typically are developed in languages such as R or Python. Such algorithms cannot just be plugged into a desktop GIS without extra work from the developer.

Robert Hijmans, a computational geographer at UCD, recalls the frustration of switching repeatedly between ArcGIS and R in order to apply a new algorithm to study income distribution in Asia. “The whole process was very cumbersome,” he says. But by transitioning fully to R, “all of a sudden I had this freedom of data analysis that is so much more powerful”.

Indeed, whatever tool you choose, it has never been easier to tell those map-based stories. ■

Jeffrey M. Perkel is technology editor for Nature.

Mapping in R

The Leaflet library for R is a set of easy-to-use tools for visualizing geospatial data. Researchers can create and publish interactive maps featuring panning, zooming and informational markers. To test the library, which is also available in Python, I compiled points of interest in southeast England in a comma-separated values table, and plotted them on a topographical map. Then, with help from John Czaplewski, lead developer on the Macrostrat project, I pulled in Macrostrat map tiles for the same region, colouring the base map to reflect the underlying geology of the region. The whole script required just seven lines of code (see go.nature.com/2iztpta).

Next, I downloaded the projected path for Hurricane Irma that the National Hurricane Center calculated on 4 September 2017. The US organization supplies those predictions as ZIP files each containing 20 files: 5 for the predicted position of the storm at each point in the forecast, a line connecting those points, the ‘cone of uncertainty’ of that predicted track, and the

locations of storm watches and warnings. Collectively, these 5 files comprise a ‘shape file’, and for some tools, they must be repackaged into individual ZIP archives. The files can then be imported as individual layers, each of which can be formatted for colour, opacity, and so on. But using `rgdal`, which provides an R interface to the Geospatial Data Abstraction Library’s collection of data-access tools, I could read the unzipped files and plot them alongside a selection of potential coastal targets, which I stored as a table of comma-separated values.

Finally, I overlaid a plot of public sea surface temperature data. Those data are not available in a readily mappable form, so I asked marine biologist Luke Miller for help. Miller, at San José State University, California, has developed software to extract and plot these data to inform his studies on coastal molluscs and crustaceans. He kindly worked out how to get those data into a Leaflet-compatible form, yielding the final figure (see go.nature.com/2j4nx9o). **J.M.P.**