

GIS-Technologie-News

SOGI-Informationsblatt 4/2019

Apache Superset und andere Datenvisualisierungs- und -Publikations-Werkzeuge

In letzter Zeit machen neuartige Anwendungen von sich reden, die sich als *Datenvisualisierungs- und -Publikations-Werkzeuge* bezeichnen lassen. Diese werden oft auch *Business Intelligence-Werkzeuge* genannt.

Eine gute Visualisierung ist wichtig, damit Daten, wie beispielsweise Umfragen, Sensordaten oder räumliche Daten, veranschaulicht werden. Das verbessert deren Verständlichkeit und Akzeptanz. Häufig erkennt man zudem Zusammenhänge erst durch geschickte Diagramme und Karten. Das heisst, dass Visualisierungen auch zur Datenanalyse eingesetzt werden können. Darüber hinaus können die zu Listen, Diagrammen oder Karten aufbereiteten Daten organisations-intern publizieren oder übers Web mit externen Kunden geteilt werden. Diese meist webbasierten Anwendungen haben das Potential, bestimmte Funktionen von Geoinformationssystemen (GIS) zu übernehmen.

Die beiden wohl verbreitetsten Produkte sind die kommerziellen Anwendungen Tableau und Microsoft PowerBI. Im GIS-Umfeld gibt es Insights for ArcGIS und CARTO. Bei genauerem Hinsehen findet man viele weitere Produkte – besonders wenn man noch Produkte dazu nimmt, die aus dem System-Monitoring kommen, wie beispielsweise Kibana. Es scheint nicht ganz einfach, Definitionsgrenzen zu ziehen.

Folgende Funktionen sind gemeinsam an diesen Werkzeugen:

- Analysieren: Daten selektieren, filtern und aufbereiten.
- Visualisieren: Listen, interaktive Diagramme und Karten erstellen.
- Präsentieren: Listen, Diagramme und Karten zu einem „Cockpit“ zusammenfassen (in Superset „Dashboard“ genannt).
- Publizieren: Die Cockpits/Dashboards über das Web mit anderen teilen.

Alle Werkzeuge erlauben eine einfache Erstellung von Diagrammen. Dazu kommen Funktionen zur Benutzerverwaltung wie das Anlegen von Konten und Kontengruppen, inklusive Vergabe von Rechten, die bestimmen, auf welche Datenquellen und Funktionen ein Benutzer Zugriff bekommt.

Mit Apache Superset ist seit 2015 auch ein Open Source-Projekt verfügbar. Der Hauptautor dieses Beitrags hat sich darum diese Anwendung etwas näher angeschaut und nachfolgend seine Eindrücke kurz zusammengefasst, beginnend mit folgenden drei Punkten:

- Die Analyse mit Apache Superset umfasst verschiedenste Datenquellen, von Dateien im CSV- oder GeoJSON-Format über Datenbanksysteme bis zu Onlineservices.
- Die Erstellung von Visualisierungen (Diagramme und Karten) ist mit wenigen Klicks möglich.
- Die Handhabung und die Publikation der Resultate gehen allgemein recht leicht von der Hand.

Es gibt einige Möglichkeiten, die Datensätze zu analysieren. Und mit 30 Diagramm-Typen und zwölf Karten-Typen steht eine ansehnliche Auswahl an Visualisierungs-Varianten zur Auswahl.

Es kann nicht genug betont werden, dass bei Werkzeugen dieser Art weder Programmier- noch besondere GIS-Kenntnisse nötig sind. Und falls die bestehenden Funktionen nicht ausreichen, kann man SQL-Anfragen einbinden (lassen), die dann als neue Datenquellen zur Verfügung stehen.

Das Open Source-Projekt Apache Superset ist zurzeit noch nicht so mächtig wie die oben genannten kommerziellen Werkzeuge und die webbasierte Anwendung muss noch umständlich installiert werden – ausser man findet jemanden, der sich schon auskennt. Diese Unzulänglichkeiten erklären sich vor allem dadurch, dass das Projekt zwölf Jahre jünger ist als beispielsweise die Firma Tableau, die 2003 gegründet wurde.

Die Funktionalität von Superset ist jedenfalls bereits gut genug für interessante In-House-Anwendungen und wird u.a. von Unternehmen wie Airbnb oder Twitter eingesetzt.

Stefan Keller und SOGI Fachgruppe 4 GIS Technologie

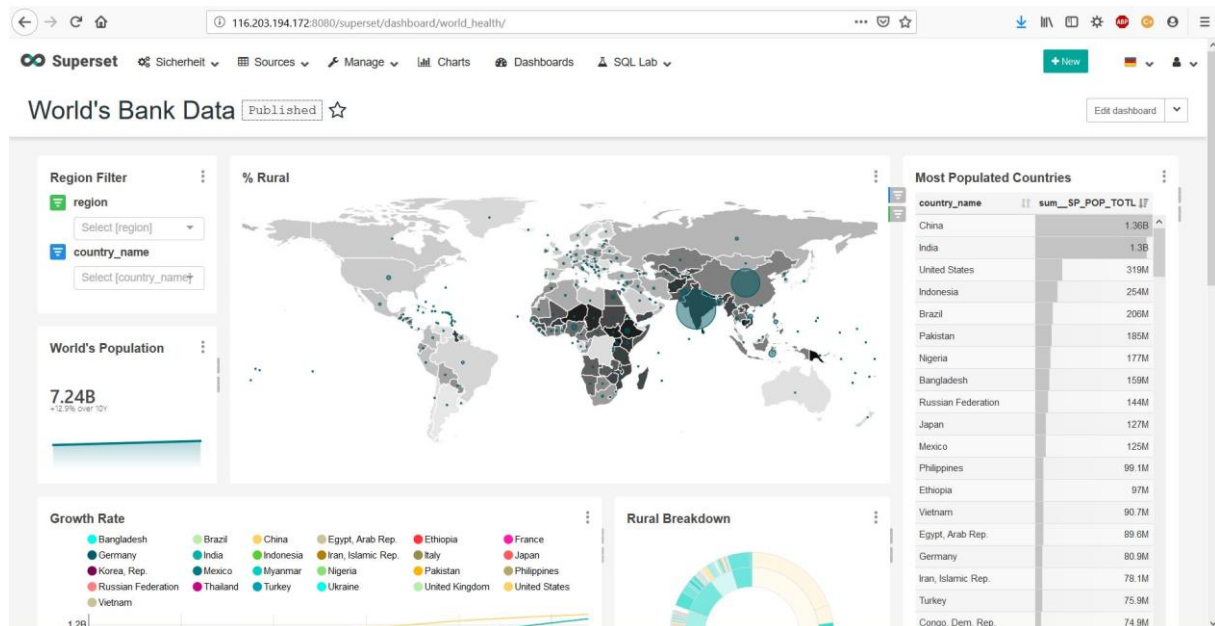


Abbildung: Bildschirmfoto von Diagrammen über Länder-Daten der Weltbank visualisiert als „Dashboard“ mittels Apache Superset.