Building Contextual Data from Online Sources

Nathan R. Jones, PhD
University of Wisconsin Survey Center

Abstract

The purpose of this presentation is to demonstrate ways to extend the utility of data you already have, build rich new sources of data from publicly available sources, and interact efficiently with web application programming interfaces (API). Although the discussion will focus on general approaches, I use R with some popular helper packages (tidyverse, rvest, and ggmap). The presentation will feature an example of mapping medical facilities and travel times using an API.

Extended description

Location data are available from many online sources. The R package *ggmap* provides access to Google Maps, OpenStreetMap, and Stamen Maps APIs and facilitates geocoding, distance and travel time calculations, and sophisticated map rendering and data visualization options. I will use the *mapdist()* function to obtain location information and calculate distance from a list of addresses to a list of urgent care clinics, make a data file of the results, identify the closest clinic to each address, and plot the results on a map.

A Google Places API search for "Urgent care near Dane County" provides longitude, latitude, and address for each clinic.

name	lon	lat	address
UW Health Urgent Care Clinic	-89.506	43.061	7102 Mineral Point Rd, Madison, WI 53717, USA
SSM Health Urgent Care	-89.308	43.072	1821 S Stoughton Rd, Madison, WI 53716, USA
UW Health Urgent Care Clinic	-89.350	43.098	2402 Winnebago St, Madison, WI 53704, USA
GHC-SCW Urgent Care Clinic	-89.394	43.067	675 W Washington Ave, Madison, WI 53703, USA
GHC-SCW East Clinic	-89.289	43.145	8339, 5249 E Terrace Dr, Madison, WI 53718, United States
SSM Health Dean Medical Group	-89.520	43.077	752 N High Point Rd, Madison, WI 53717, USA
Concentra Urgent Care	-89.525	43.068	358 Junction Rd, Madison, WI 53717, USA
Concentra Urgent Care	-89.324	43.120	1619 N Stoughton Rd, Madison, WI 53704, USA
Mental Health Center of Dane County	-89.483	43.049	25 Kessel Ct Ste. 105, Madison, WI 53711, USA
Sauk Prairie Healthcare	-89.753	43.287	260 26th St, Prairie Du Sac, WI 53578, USA
Journey Mental Health Center	-89.481	43.049	49 Kessel Ct, Madison, WI 53711, USA
Unitypoint Health - Meriter Hospital	-89.402	43.066	202 S Park St, Madison, WI 53715, USA
UnityPoint Health - Meriter - McKee Clinic	-89.528	43.017	3102 Meriter Way, Madison, WI 53719, USA

Then, the *mapdist()* function calculates travel distances, duration, other details, by mode:

```
from <- homes$address
to   <- clinics$address

dist <- mapdist(
     from,
     to,
     mode="bicycling",
     output="simple”
     )
```

Use *ggmap()* to create a map with addresses in blue and clinics in red

```
ggmap(mdata,
  geom_point(data=homes, aes(x = lon, y = lat), color="blue")+
  geom_point(data=clinics, aes(x = lon, y = lat), color="red")
```

Resulting map: