# SAS<sup>®</sup> GLOBAL FORUM 2021

# Paper 1046-2021

# Spatial Analytics With SAS®: Examining Contributions to OpenStreetMap for the Covid-19 Response

Michael Matthews, Innotwist Pty Ltd

# **ABSTRACT**

Base SAS software includes powerful tools for spatial analytics that can be used in a variety of circumstances. This case study examines the contributions made by the Humanitarian OpenStreetMap Team (HOT) volunteers to support diverse global projects assisting impacted communities. SAS ODS Graphics procedures were used to analyze the OpenStreetMap data. As a result, work undertaken by global volunteers and mapping communities to identify and update infrastructure and housing can be analyzed to show the impact both spatially and over time. This analysis demonstrates techniques that are applicable for any organization using the open data from OpenStreetMap or other mapping data sources.

### INTRODUCTION

According to the United Nations, the number of people affected by a humanitarian crisis had doubled in the past decade prior to the significant impacts of the Covid-19 crisis. Open source technology is increasingly being used in a number of applications to create new solutions to allow aid organizations to respond to disasters. Particularly in the developing world, many communities have not been included in traditional mapping products. This presents a challenge for humanitarian organizations responding to disasters, whose volunteers may at best be given a hand-drawn, basic map to reach isolated and vulnerable communities.

The Humanitarian OpenStreetMap Team (HOT) is changing this access and inclusion by bringing together a community of volunteers from around the world (over 100,000 people from more than 60 countries) who are placing previously untracked communities on the map through collaborative geospatial data collection. Organizations such as the Red Cross, Médecins Sans Frontières, USAID and local government organizations post projects through the HOT web interface for mappers to work on.

Using satellite imagery, volunteers are allocated specific grids within a geographic region to trace infrastructure such as buildings, roads, rail or rivers. These map edits are validated and pushed live into OpenStreetMap so that on-the-ground responders can access the maps on their mobile devices as they move into the crisis areas. The mapping data is also being used for communities' evidence-based decisions, predictive analysis and preparation for both natural and medical disasters.

Global contributors to OpenStreetMap result in a rich source of data for analysis that can be used for many purposes. This paper analyses OpenStreetMap data and some of the Humanitarian OpenStreetMap Team mapping tasks in Peru to support local communities impacted by the Covid-19 crisis. Analysis is performed on the data using the SAS® ODS Graphics procedure PROC SGMAP to visualize the contributions to OpenStreetMap both spatially and over time. A review of additional OpenStreetMap tags that were created to globally support communities with Covid-19 businesses opening hours and delivery during lockdown periods is also performed for Italy.

### **OPENSTREETMAP**

OpenStreetMap is a map of the world built by a community of mappers who donate their time and skills to create mapping data available for everyone.

Consumers of OpenStreetMap, and other commercial mapping solutions, often consider the visual display of maps to be the final product, however with OpenStreetMap the data itself is the primary asset, with the rendering of the maps only one of the many possible uses.

The OpenStreetMap data is open; licensed under the Open Data Commons Open Database License (ODbL) by the OpenStreetMap Foundation (OSMF).

The default OpenStreetMap presentation at <a href="https://www.openstreetmap.org">https://www.openstreetmap.org</a> is shown in Figure 1:

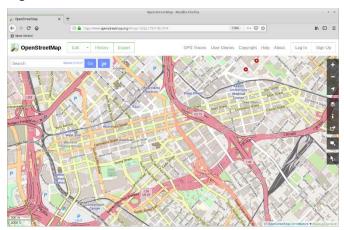


Figure 1. Standard OpenStreetMap layer

The underlying OpenStreetMap data supports producing many displays with a different emphasis; for example, Figure 2 shows a transport layer including the train and other public transport routes as the primary focus:

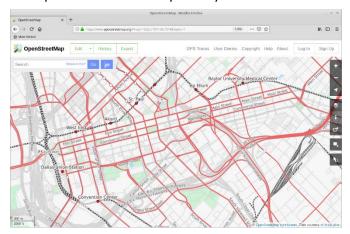


Figure 2. Transport Map OpenStreetMap layer

The generation of the maps shown on the OpenStreetMap website are rendered using a Tile Server which generates a group of image tiles that are then seamlessly joined in the browser to create the complete map display. Different maps can be rendered based on the specific requirements; along with the standard view, there are transport maps (that highlight train, bus and ferry routes), humanitarian views (that highlight medical facilities, water sources, and other important infrastructure), cycling maps and sea maps for navigation as a few examples.

The OpenStreetMap data is available to copy and use for any purpose within the license agreement.

The data is available from numerous sources, including:

- 1. APIs: OpenStreetMap and Overpass API are two examples of many available. The licenses for these sources limit usage and should be reviewed before use.
- 2. Database copy: Planet OSM enables you to copy the entire global database (100+ GB). Other providers offer useful regional extracts at different country and/or state levels. It is also possible with a full copy to create a process where regular deltas are applied to keep a local copy of the OpenStreetMap database current without having to download the entire globe on a regular basis.
- 3. Tile Servers: The OpenStreetMap tile servers and others are available for personal use, however they are often not licensed for commercial use due to their limited funding and infrastructure. It is possible to create your own tile server, such as the way SAS Institute have done, to support the OpenStreetMap layers in their mapping solutions.

#### **DATA CONCEPTS**

The following fundamental concepts regarding the data used by OpenStreetMap are required to understand the analysis performed in this paper:

**Node**: A point which may be standalone as a marker to identify an item or place, or part of a way.

**Way**: An ordered collection of Nodes that make up a larger object. This may be a street, building, lake, park, or any other object represented as either a line or polygon.

**Relation**: A collection of related Nodes and Ways. An example might be multiple buildings that are related as a single hospital.

**Tag**: A descriptive attribute that can be added to Nodes, Ways or Relations. This might be a street name, building height, road surface etc. The tags provide rich data for both the tile rendering and for analysis.

# **BUILDINGS AT THE SAS CAMPUS**

The rich data in the OpenStreetMap database includes many tags describing the nodes, ways and relations. Many of these are not visible in the standard map displays, however the data is able to be used to perform analysis and reporting.

To introduce the use of OpenStreetMap data with SAS Software, this example uses the OverPass API to extract buildings with a name starting with the string "SAS Building" in the area "Cary". The data is then processed using BASE SAS and the SAS procedure SGMAP.

The following text file contains the Overpass API request format selecting the ways with a name starting with "SAS Building" in the "Cary" area:

```
[out:xml][timeout:25];
(area[name="Cary"];)->.searchArea;
(
   way["name"~"SAS Building"](area.searchArea);
);
out center;
```

The BASE SAS code below uses the SAS procedure HTTP to call the Overpass API, with the specifications detailed above in the file sas\_campus\_spec.txt. The response from the API call is written in XML format to the file sas campus.xml:

```
filename spec "sas_campus_spec.txt";
filename resp "sas campus.xml";
```

Note that it is worthwhile saving the results of any requests rather than calling the API repeatedly if the data is unlikely to change. This will improve your program performance and prevent unnecessary calls to the API servers.

The XMLV2 librame engine uses an XML Map to convert sas\_campus.xml into a SAS library with various tables including the associated ways and way\_tags:

```
filename osm_map "osm_map.xml";
libname osm data xmlv2 "%sysfunc(pathname(resp))" xmlmap=osm map;
```

The library osm\_data created above with the XML Map creates a collection of datasets as shown in Figure 3:

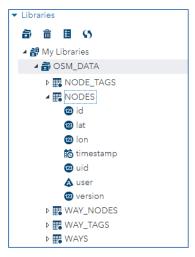


Figure 3. Library Contents Using XML Map

The XML Map used above can be used to read any OpenStreetMap XML data and is available at <a href="https://github.com/m-matthews/hot-sas/blob/master/data/osm\_map.xml">https://github.com/m-matthews/hot-sas/blob/master/data/osm\_map.xml</a>.

The XML data from the osm\_data libref can be joined using standard SAS code:

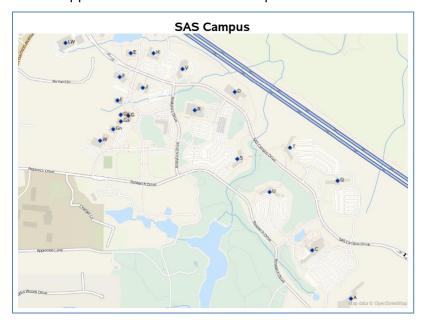
The resulting SAS dataset now contains the list of SAS buildings and coordinates as shown in Output 1:

Obs	id	lat	lon	name	building
1	32482276	35.8287	-78.7670	SAS Building LW	LW
2	32482304	35.8243	-78.7640	SAS Building Gn	Gn
3	32482305	35.8238	-78.7648	SAS Building W	W
4	32482345	35.8228	-78.7562	SAS Building S	S
5	47777592	35.8253	-78.7589	SAS Building R	R

**Output 1. List of SAS Buildings** 

The lat and lon variables are now in a form suitable for presenting visually on a map. The SAS procedure SGMAP is used to generate the map using BASE SAS:

The resulting image from the PROC SGMAP code is shown in Output 2, with the marker labels supplied with the DATALABEL options on the SCATTER statement:



# **Output 2. SAS Campus Buildings**

The OPENSTREETMAP statement supplied to PROC SGMAP ensures that the rendered background for the map is using the SAS supplied OpenStreetMap tile server.

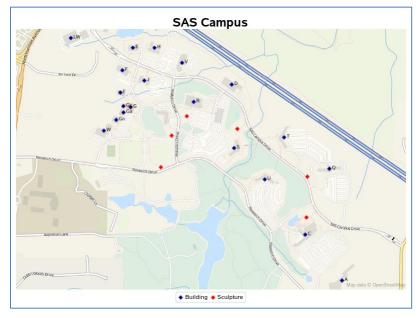
The rich data available in OpenStreetMap can be seen by also adding the sculptures on the SAS Campus through using an updated call to the Overpass API. Note that the Nodes with tags of "tourism" = "artwork" have been selected, along with a spatially limited bounding box (rather than just "Cary") to select the required area of interest:

```
[out:xml][timeout:25];
(
   node["tourism"="artwork"](35.815, -78.767, 35.829, -78.749);
);
out body;
(
   way["name"~"SAS Building"](35.815, -78.767, 35.829, -78.749);
);
out center;
;;;;
```

The following code combines the two sets of data, where the buildings are ways, and the sculptures are nodes (point data):

The following code includes the GROUP option on the SCATTER statement to provide the different appearance for the buildings and sculptures at the SAS Campus and the legend:

The resulting image from the PROC SGMAP code is shown in Output 3:



**Output 3. SAS Campus Buildings and Sculptures** 

This same technique for extracting OpenStreetMap data can be used for other items of interest, such as hospitals, golf courses, mines, parks, train stations etc., which can be useful for other analytics and visualizations.

An example is showing the OpenStreetMap 'way' surrounding Central Park in New York, displayed with a SERIES statement rather than the SCATTER statements from previous examples, is shown in Output 4:



**Output 4. Central Park** 

Detailed code for the content of this paper, including the additional example shown in Output 4, is available in the repository <a href="https://github.com/m-matthews/hot-sas">https://github.com/m-matthews/hot-sas</a>.

# **HUMANITARIAN OPENSTREETMAP TEAM**

The Humanitarian OpenStreetMap Team (HOT) is dedicated to humanitarian action and community development through open mapping. The HOT community is made up of volunteers, local community leaders, and professionals who work together to provide map data which revolutionizes disaster management and reduces risks for vulnerable communities. Many different organizations request projects through the HOTOSM website, including the Red Cross, Médecins Sans Frontières, USAID and local governments.

Output 5 shows the location of HOTOSM projects, with highlights for the Ebola crisis discussed in a 2019 SAS Global Forum paper, and tasks relating to the Covid-19 response:



Output 5. Humanitarian OpenStreetMap Team Tasks

The display in Output 5 demonstrates the ongoing varied requirement for Humanitarian mapping across the globe, and where projects specific to the Covid-19 response have been performed.

# **HOTOSM PROJECTS IN CUSCO, PERU**

Due to the Covid-19 situation in Peru, and the lack of tourism providing income to the local population, the Cusco regional government made requests to the Humanitarian OpenStreetMap Team to complete the maps in this area. This was to help to identify and provide cash transfers to families impacted by the quarantine and state of emergency declared in the region, and for longer term regional planning.

The following analysis is to examine the number of new buildings that were added to the map during the term of these projects.

The OpenStreetMap data is extracted from a regional subset of the Planet OSM data available at the Geofabrik website <a href="https://download.geofabrik.de/south-america/peru.html">https://download.geofabrik.de/south-america/peru.html</a>.

The data is converted from the supplied OSM PBF format into an XML file to be easily consumed using SAS Software by using one of the many OpenStreetMap tools available. The following use of OSMOSIS extracts all 'building' Ways along with the geographical coordinates:

```
osmosis --read-pbf ./peru-latest.osm.pbf
--bounding-box top=-13.153 left=-72.551
bottom=-13.868 right=-71.374
--tf accept-ways --tf reject-relations
--way-key keyList="building" --used-node
--write-xml cusco buildings.osm
```

Using the XML Map described in previous sections, the XML data can be converted into a SAS dataset using the SAS procedure SQL:

```
filename osm map "osm map.xml";
libname osm data xmlv2 "cusco buildings.osm"
                 xmlmap=osm map;
proc sql;
  create table work.cusco buildings as
    select w.id, w.timestamp, w.version,
          wt.value as type,
          mean(n.lat) as lat,
          mean(n.lon) as lon
      from osm data.ways as w,
          osm data.way tags as wt,
           osm data.way nodes as wn,
          osm data.nodes as n
         where wt.key="building" and
               w.id = wt.id and w.id = wn.id and
               wt.id = wn.id and wn.ref = n.id
           group by w.id, w.timestamp, w.version, wt.value
             order by id;
```

Note that the above SQL code uses the mean (average) of the node coordinates for each building (way) to create an approximate centroid. While this is not the most accurate method, it is suitable for the resolution of this analysis.

The resulting data can be seen in Output 6, showing the latest update timestamp, the version of the object and the building type. Note that the default building type is 'Yes' as

remote mappers using satellite imagery are unable to determine the details, which may be updated later by mappers local to the area:

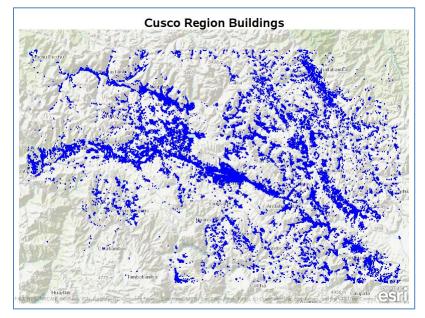
Obs	id	timestamp	version	type	lat	lon
1	4818980	23APR20:10:55:30	8	yes	-13.5380	-71.9435
2	23799145	02NOV17:15:16:19	7	church	-13.5149	-71.9743
3	23799161	26JAN20:13:27:20	11	yes	-13.5176	-71.9779
4	23949833	01FEB20:22:48:13	5	train_station	-13.5211	-71.9833
5	36115454	08APR20:18:06:03	9	yes	-13.1661	-72.5429

# **Output 6. Cusco Buildings**

There are more than 334,000 buildings included in this area, many of which have been manually added by the global volunteers through the Humanitarian OpenStreetMap Team projects.

The following SAS code uses PROC SGMAP to generate a view of the individual buildings using ESRIMAP imagery for the background layer rather than using the OPENSTREETMAP statement:

The PROC SGMAP SCATTER statement results are shown in Output 7, although due to the number of points it is difficult to determine the building density due to the overlapping pixels for individual buildings:



**Output 7. Cusco Buildings (Scatter)** 

One solution to the issue of overlapping points in a scatter plot is to use a binning technique, where the area is divided (often into a square based grid) and then counts within those areas are used to generate colored regions of intensity.

The default 'square' grid technique is known to have potential bias and there are many discussions on the benefits of using the alternative hexagonal based binning technique.

The SAS/STAT® software procedure SURVEYREG can produce hexagonal regions and statistics to generate the required style of map display:

A subset of the PROC SURVEYREG output is displayed in Output 8:

Obs	hID	XVar	YVar	WVar
1	1	-71.3919	-13.8618	22
2	1	-71.3795	-13.8664	22
3	1	-71.3795	-13.8757	22
4	1	-71.3919	-13.8803	22
5	1	-71.4043	-13.8757	22
6	1	-71.4043	-13.8664	22
7	2	-71.3795	-13.8479	50
8	2	-71.3671	-13.8525	50
9	2	-71.3671	-13.8618	50
10	2	-71.3795	-13.8664	50
11	2	-71.3919	-13.8618	50
12	2	-71.3919	-13.8525	50

### **Output 8. PROC SURVEYREG Hexagonal Binning**

There are 6 observations per value of hID; the XVar and YVar are the boundary coordinates for the hexagon, while the WVar is the number of observations (buildings) present within this boundary.

Using the SAS procedure UNIVARIATE on the variable WVar produces useful percentiles to create visual banding ranges. Based on the current data the following format produces a suitable banding for display:

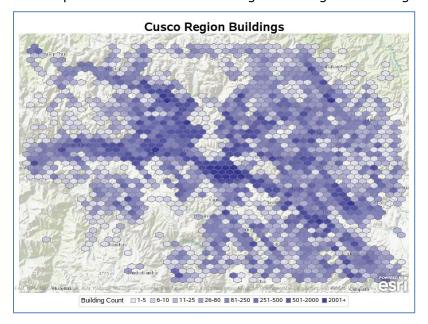
```
proc format;
  value hexgrp 1-5='A'
                6-10='B'
                11-25='C'
                26-80='D'
                81-250='E'
                251-500='F'
                501-2000='G'
                2001-high='H';
  value $hexgrp 'A'='1-5'
                 'B'='6-10'
                 'C'='11-25'
                 'D'='26-80'
                 'E'='81-250'
                 'F'='251-500'
                 'G'='501-2000'
                 'H'='2001+';
```

The dataset created by PROC SURVEYREG is then converted into two datasets using these formats; one describing the boundaries (work.hexmap\_map) and the other the response (work.hexmap\_resp):

```
data work.hexmap_map(keep=id x y)
     work.hexmap_resp(keep=id wvar gvar);
set work.hexmap(rename=(hid=id xvar=x yvar=y));
where id ne .;
by id;
output work.hexmap_map;
if last.id;
gvar=put(wvar,hexgrp.);
format gvar $hexgrp.;
output work.hexmap_resp;
run:
```

These two datasets can then be used with PROC SGMAP to create the hexagons and display the density of buildings present in each defined area by using a CHOROMAP statement:

The output from PROC SGMAP using the hexagonal binning is shown in Output 9:



**Output 9. Cusco Buildings (Hexagonal Binning)** 

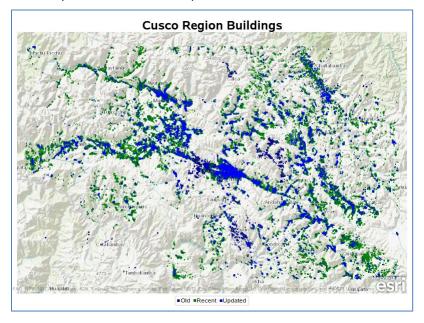
This method can be used to provide a clearer image of point density rather than overlaying many pixels together. The regional center in the middle of the map is clearly visible as an area with many buildings, however it is interesting to note the wide extent of mapped buildings in the remote and mountainous parts of the region.

The next analysis is to determine when the building was added or amended in OpenStreetMap. This can be used to approximate if the building was added as part of the HOTOSM projects of interest.

The following SAS code determines if the building existed prior to the project commencement dates and labels them "Old", if they are recent with version=1 then they are "Recent", otherwise they are "Updated" during the period of interest:

```
data work.cusco updates;
```

The output is visible in Output 10:



**Output 10. Cusco Buildings Analysis** 

The older buildings are present in the central, more populated area, however there have been many updates to the buildings in the broader region due to the HOTOSM Projects, including in many remote areas that were previously not mapped.

# **OPENSTREETMAP COVID-19 TAGS USAGE IN ITALY**

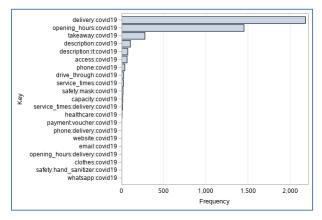
OpenStreetMap includes many tags to identify features, such as buildings, road, rivers, and also attributes related to those features such as names, addresses and opening hours. To support communities during the global Covid-19 crisis, OpenStreetMap have created additional standard tags specific for this crisis. This section will demonstrate the usage of those tags in Italy.

The data is extracted from a regional subset of the Planet OSM data available at the Geofabrik website https://download.geofabrik.de/europe/italy.html.

The data can be converted from the supplied OSM PBF format into an XML file to be easily consumed using SAS Software by using one of the many OpenStreetMap tools available. The following use of OSMOSIS extracts the covid19 tags of interest for ways and nodes along with the latitude and longitude:

Note that both ways and nodes are used in this analysis, as a point of interest may be either a building (way) or a tagged point (node) on a building.

A frequency count of the covid19 tag usage in Italy (without the limiting keyList shown above) is displayed in Output 11:



Output 11. OpenStreetMap "covid19" Tags

Note that the usage of some tags is limited in the Italian extract, however these tags have been used globally with varying levels of use.

The data can be combined to produce geographical coordinates for the tags using the following code:

```
proc sql;
  create table work.combined as
    select id, lat, lon, "Node" as type
      from osm itan.nodes
        union corresponding
    select w.id,
           mean(n.lat) as lat,
           mean(n.lon) as lon,
           "Way" as type
      from osm itaw.ways as w,
           osm itaw.way nodes as wn,
           osm itaw.nodes as n
         where w.id = wn.id and
               wn.ref = n.id
           group by w.id
             order by id;
quit;
```

Note that the above SQL code uses the mean (average) of the way's node coordinates to create an approximate way centroid. While this is not the most accurate method, it is suitable for the resolution of this analysis.

# The data can be analyzed with PROC SGMAP:

The regions using the Covid-19 tags are visible in Output 12:



Output 12. OpenStreetMap Italian "covid19" Tags

The output is shown in Output 13:

The usage of the tags is concentrated around individual cities within Italy. As the data entered in OpenStreetMap can be used for multiple purposes, local governments can use this information to publish the details to their communities. This could be using a variety of distribution platforms, including customized mobile applications.

A closer analysis of Bologna is performed below, where the centroid is created using SAS Macro variables to allow simple code changes to analyze different cities or regions:



Output 13. OpenStreetMap Bologna "covid19" Tags

Details on Covid-19 business availability within Bologna has been entered for many points within the local area, with the information available for anyone to use. This method of tagging has been used across the globe to assist communities in communicating the availability of resources during the pandemic.

### CONCLUSION

In 1854 John Snow made medical history and changed the way health data is represented by mapping cholera cases across London to visualize the source of the problem. In addition to deaths at residences, he included the source of the problem - contaminated water pumps - providing data points for more effective outbreak control. Since then, maps have been helping aid organizations and medical professionals better respond to emergencies, track the spread of diseases, and improve access to healthcare facilities.

Today, collaborative mapping and geospatial data collection is changing the inclusion outcomes for many communities. From disaster risk reduction to the elimination of diseases such as malaria, global volunteers are literally putting communities on the map to change their access to support and services.

From a technical perspective, global contributors to OpenStreetMap result in a rich source of data for analysis that can be used for many purposes. Just-in-time mapping provides on-the-ground volunteers with accurate, detailed maps of hard-to-reach communities. Proactive mapping of infrastructure and risk assessment is assisting with early-warning and disaster management preparation.

SAS Software includes the functionality to extract, analyze and visualize data from OpenStreetMap. The data can be sourced from APIs (using PROC HTTP) or OpenStreetMap downloads, and then converted into SAS dataset format from the XML data format. SAS Software includes many tools for analysis and visualization of the data, including the ability to use OpenStreetMap tile servers as the background for map displays with PROC SGMAP and SAS® Visual Analytics.

The repository <a href="https://github.com/m-matthews/hot-sas">https://github.com/m-matthews/hot-sas</a> contains the complete source code and files used to perform the analyses in this paper.

# REFERENCES

OpenStreetMap. "Copyright and License." Accessed March 31, 2021.

https://www.openstreetmap.org/copyright

OpenStreetMap Wiki. "Planet OSM." Accessed March 31, 2021.

https://wiki.openstreetmap.org/wiki/Planet.osm

OpenStreetMap Wiki. "Overpass API." Accessed March 31, 2021.

https://wiki.openstreetmap.org/wiki/Overpass API

OpenStreetMap Wiki. "Osmosis." Accessed March 31, 2021.

https://wiki.openstreetmap.org/wiki/Osmosis

Humanitarian OpenStreetMapTeam. "What We Do." Accessed March 31, 2021. https://www.hotosm.org/what-we-do

OpenStreetMap Wiki. "COVID-19 - How to Map." Accessed March 31, 2021.

https://wiki.openstreetmap.org/wiki/COVID-19 - How to Map

Geofabrik. "Downloads." Accessed March 31, 2021.

https://www.geofabrik.de/data/download.html

# **CONTACT INFORMATION**

Your comments and questions are valued and encouraged. Contact the author at:

Michael Matthews
Innotwist Pty Ltd
michaelm@innotwist.com
https://github.com/m-matthews/hot-sas

Data used in this paper is © OpenStreetMap contributors.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.