Received: 18 June 2020, **Accepted:** 29 Sept. 2020
Published online: 31 Oct. 2020
DOI: https://doi.org/10.22201/igg.25940694e.2020.2.74 **Volume:** 4, **Issue:** 2, **Pages:** 1-5

terra digitalis
international e-journal of maps

Mapping COVID-19 in Los Angeles with Power BI☆

Mapeo de la COVID-19 en Los Angeles utilizando Power BI

Steven M. Graves*[a], Lauren He[b]

[a]*Department of Geography and Environmental Studies, California State University, Northridge. 18111 Nordhoff Street, Northridge, California, 91330. United States of America.*
[b]*Westlake High School. 4100 Westbank Drive, Austin, Texas, 78746. United States of America.*

Abstract

The purpose of the map and dashboard is twofold. First, Microsoft's Power BI software presents a reasonably easy means by which many can make reasonably robust data visualizations with minimal expertise in mapping and programming languages. Unlike some competing data visualization and online mapping software packages that require specialized software or skills, ordinary citizens and students, new to data collection and management can create an appealing dashboard-style presentation within a few minutes with minimal guidance. Familiarity with the function and logic of spreadsheet software seems to be the only prerequisite. Many businesses, colleges, and universities already provide access to the Microsoft Office suite of software, further extending the appeal of using Power BI. The COVID-19 dashboard featuring case data by date, city, county, and/or neighborhood in Southern California was created in less than one-half hour using a connection to publicly available data shared by the Los Angeles Times via GitHub. The map allows visitors to select individual communities from a map, which prompts dynamically linked graphs and tables to display data for only that community. The second purpose of this dashboard is to allow users to rapidly analyze local data trends, perhaps following school openings/closings, changes in the status of businesses, or social protests movements. Epidemiologists have expressed concern over the effects of the protest marches and untimely changes in rules regarding businesses and schools on COVID-19 rates. Compounding the need for such a map are concerns about differential COVID-19 infection rates in communities of color, whose citizens were already at an elevated risk of death from COVID-19. This map and dashboard demonstrates that these concerns were not unfounded.

Keywords: COVID19; map; Los Angeles; visualization; dashboard

Resumen

El propósito del mapa y el tablero es doble. En primer lugar, el software Power BI de Microsoft es un medio razonablemente sencillo para realizar visualizaciones de datos razonablemente sólidas con una experiencia mínima en lenguajes de programación y cartografía. A diferencia de algunos paquetes de software de visualización de datos y mapas en línea de la competencia que requieren software o habilidades especializadas, los ciudadanos y estudiantes comunes y corrientes, nuevos en la recopilación y gestión de datos, pueden crear una presentación atractiva al estilo de un tablero en unos pocos minutos con una orientación mínima. La familiaridad con la función y la lógica del software de hoja de cálculo parece ser el único requisito previo. Muchas empresas, facultades y universidades ya brindan acceso al paquete de software Microsoft Office, lo que amplía aún más el atractivo del uso de Power BI. El panel de COVID-19 con datos de casos por fecha, ciudad, condado y/o vecindario en el sur de California se creó en menos de media hora utilizando una conexión a datos disponibles públicamente compartidos por Los Angeles Times a través de GitHub. El mapa permite a los visitantes seleccionar individualmente comunidades de un mapa, y muestra gráficos y tablas vinculados dinámicamente a los datos de la comunidad seleccionada. El segundo propósito de este tablero es permitir a los usuarios analizar rápidamente las tendencias de los datos locales, tal vez después de las aperturas/cierres de escuelas, cambios en el estado de los negocios o movimientos de protesta social. Los epidemiólogos han expresado su preocupación por los efectos, sobre las tasas de COVID-19, de las marchas de protesta y los cambios inoportunos en las reglas aplicadas a negocios y escuelas. A la necesidad de un mapa de este tipo, se suma la preocupación por las tasas diferenciales de infección por COVID-19 en las comunidades de color, cuyos ciudadanos ya tenían un riesgo elevado de muerte por COVID-19. Este mapa y tablero demuestran que estas preocupaciones no eran infundadas.

Palabras clave: COVID19; mapa; Los Angeles; visualización; tablero

1. Introduction

The COVID-19 pandemic of 2020 vastly increased the public awareness of online mapping tools, particularly the "dashboard" style tools used by many public health agencies across the United States and to some extent, Mexico and elsewhere. Many universities and other research concerns produced similar dashboards with maps and creative visualizations. Several of these efforts functioned as valuable additions to the public and scientific discourse regarding the pandemic. Public officials regularly referenced these data tools in both policymaking efforts, as well as nightly briefings on the television news.

The academic community worldwide responded quickly to the crisis and geographers figured prominently among those publishing analyses, even in the earliest days of the pandemic. Kamel Boulos and Geraghty (2020) produced an excellent primer on health mapping and some of the challenges of mapping Covid-19, noting the difficulty of mapping at local scales, and data sourcing during the early stages of the pandemic. Pardo et. al. (2020) published an excellent review of some 60-plus articles addressing the pandemic during the early months of the outbreak, categorizing the research efforts into five different research foci: Spatio-temporal analyses, Data Mining, Web-based mapping, Health and Social geography, and Environmental variables.

As the Kamel Boulos and Geraghty (2020) article notes, many dashboards, maps, and data visualizations shared several unfortunate drawbacks. We noted quite a number as well. The first problem was that the construction of many of these visualization tools required moderately advanced skills in database management and/or programming skills to create dynamic connections between data servers and the mapping and/or visualization tools, especially if data calculations were involved. Second, some maps appeared to have been constructed in the absence of regard for basic cartographic principles and occasionally were bereft of common-sense approaches to statistical communication or epidemiological standards. Third, many of the impressive and/or widely circulated efforts were focused on macro-level datasets. Statewide counts and analyses were very popular. Microsoft's Bing Google's Covid-19 Dashboards focused on case number and rates at the state level and provided only the most basic of statistical analysis. Even the well-known dashboard from Johns Hopkins failed to include case rates for the first month or two of the pandemic. County-level analyses and visualizations were less popular and therefore were less commonly referenced in the news media. Sub-county analyses were almost non-existent, except in very large cities with especially well-funded news media companies with data analysis and visualization teams. This last problem seems especially troubling since the transmission of COVID-19 appears to occur largely at the local level. This effect was certainly heightened during the so-called 'lockdown' period, where travel was, supposedly, curtailed. "Community Transmission" was the watchword, but mapping and analysis were rarely occurring at the community or neighborhood level.

We suspect that the reasons behind the shortcoming in the visualization and mapping of COVID-19 in the United States may be interrelated. The sophistication required of the software user interested in analyses and visualization may have restricted the number of researchers and media outlets from attempting to create more local maps and analysis dashboards, or perhaps inhibited the public distribution of the maps and analysis. We are aware of at least one researcher who produced a series of compelling 'hot-spot' and relative risk maps that were of significant public interest, and potentially useful for policymakers, but this researchers' lack of programming skills necessary to both automate and serve his maps and analyses severely restricted his audience (Lucas, 2020). In another instance, a very high-profile dashboard with a national audience lacked a simple infection rate statistic for a month or more, although this statistic is the most common statistic used by epidemiologists and health geographers to express the prevalence of nearly every disease. One presumes this surprising oversight was due to a lack of epidemiological expertise among the programmers and a lack of programming expertise among the epidemiologists. As a result, only research organizations with a robust team of programmers, statisticians, and/or epidemiologists with advanced web capabilities were capable of producing high quality, dynamic, publicly accessible dashboards and maps that were informative and useful.

The lack of visualizations and analysis regarding specific populations and at specific scales concerned us and roused our interest in creating maps to satiate our curiosity and to provide data for audiences that shared our specific interests. Our interests center around the visualization of data in Southern California at the local level, particularly the effect of the pandemic on residential care facilities, and in turn the effect of residential care facilities on the overall data picture. According to the Los Angeles Times, at one point in the pandemic, patients living in nursing homes accounted for 43% of all fatalities in California (Dolan & Mejia, 2020). Despite the high percentage of COVID-19 cases associated with nursing homes, no mapping or visualization of infections in nursing homes or assisted living facilities in California could be identified by the researchers.

Individually, we turned to ESRI mapping tools to create the visualization and data analysis that we envisioned. One of us had been using ESRI products since the 1990s, and the other was a novice with the software. Neither of us had the programming expertise, nor the troubleshooting expertise to build a functional, dynamic dashboard using the publicly available data. The ESRI product also presented some additional challenges in terms of making public dynamically updating maps as

**E-mail address:* steve.graves@csun.edu

well. So, we turned to Microsoft's Power BI (Business Intelligence) software as an alternative platform for data analysis and visualization because it had fewer entry barriers and a friendly learning curve.

2. Data

To create our visualizations and analyses, we identified several sources of public data. Address data for nursing homes and residential care facilities were downloaded from the California Department of Health and the California Department of Social Services respectively (California Department of Public Health, 2020; California Department of Social Services, 2020). The Los Angeles Times data team collated data from across the state of California, and from County of Los Angeles Department of Public Health (Los Angeles County, 2020; Los Angeles County Department of Public Health, 2020) and graciously provided access to the data via a GitHub webpage (Los Angeles Times, 2020). The GitHub Service greatly facilitated the ease of data acquisition and renewal or 'refresh'.

The first task was to geocode the addresses of all 391 skilled nursing facilities (SNFs) and 2,053 residential care facilities (RCFs) in our study area. The addresses of the facilities were geocoded using ArcMap 10.7 and the geocoding add-on tool Awesome Table in Google Sheets. Once the addresses were geocoded and plotted as points in ArcMap 10.7, they were joined to a polygon layer of "communities" in Los Angeles County, so that each facility could be identified by its community. This was an important step because the Los Angeles County Public Health Department reports case data by "community" rather than ZIP code, census tract, or city. This peculiar data reporting is necessary because Los Angeles County is home to 88 cities and at least 140 unincorporated communities. The City of Los Angeles itself has an additional 140+ neighborhoods. In total, depending on the accounting, there were between 320 and 350 "communities" where a RCF or SNF could operate.

Once the SNF and RCF's attribute tables included a "community" column (PLACE), each table was exported as a Microsoft Excel File, stored on a university-based Microsoft One Drive folder, and uploaded into the PowerBI software, along with several other Excel files, including a table of dates, a table of communities containing columns of place names, coordinates and population of each "community" in Los Angeles County; and the three online COVID-19 data tables (all cases, SNF cases, and RCF cases – all by data and 'place') from the Los Angeles Time's GitHub.

The organization of the databases is key to allowing the software to aggregate data. So for example, if the neighborhood of "Hollywood" has five SNFs, two of which have seven COVID cases, and two of which have five COVID cases, and one has six COVID cases, the software can aggregate this data (many to one relationship) and point on the map with representing Hollywood will have a value of 30 for total number of COVID cases. For those accustomed to working with database functions, this would seem a simple matter, but for those new to the functionality and logic of databases, this capability is revelatory. To help matters, the software recognized many of the common column headers and automatically made several of the connections among the multiple sources stored as database tables. An intuitive database connectivity tool allowed us to quickly establish and modify connections between the tables in case the software didn't make those connections automatically, or if the nature of the connection was incorrect.

3. Design and Methodology

Once the tables were connected in the "Model" tool, it was only a matter of selecting the appropriate visualizations from the list in the Report window. We chose to make three separate pages or tabs of visualizations, each with a graduated point map depicting the two most basic statistics: total cases (by point size) and case rate (by point color) (See dashboard). We calculated the case rate with the standard formula for disease rates (cases per 100,000 persons) using a simple division equation in the software that pulled the cumulative cases per place per date and divided that variable by the population data from the 'place' table. Interestingly, this operation did not require the calculation of a new column in any table but is instead calculated 'on the fly', therefore saving processing power and increasing processing speed. It was not necessary with this data collection, but it is a helpful software feature for large datasets. In a slight breach of standard cartographic principle, we used a divergent color ramp (blue to red) to indicate the severity of the case rate in each community.

Accompanying the map on the first 'page' of the overall report, "Active Case Curves" we added a dual-line graph showing the cumulative number of cases "curve" on the primary vertical axis, and a rolling average of new cases during the previous five-days. We were able to find a formula to calculate a rolling multiday average within minutes via an internet search engine. The formula was copied and pasted from a helpful software blogger's website and modified to use the table names and column headers in our dataset.

A filter on the graphic restricts the plotted data to the previous 90 days, a period at the time of this writing showed the rise and apparent decline of new cases from June through mid-August of 2020. We added a date "slicer" option, allowing users to change the date range of the visualizations, which overrides the 90-day filter we placed on the default line graph.

Map users can select any point, or multiple points on the map simultaneously, and the graphic (tables and line graphs) will automatically filter the data, showing only the table data and line graph curves for the selected locations, which is very helpful for those wishing to compare the differences in several communities at once while eliminating hundreds of others from the visualization and calculations.

On the second page, entitled "Cases and Rates", we recycled the cases vs rates point map, and added a graphic that shows the differential caseload curves for each community in the study area after it had reached its 100th case of COVID-19. This graphic, which required again the adaptation of a formula we found on the internet, permits easy visualization of the relative rates at which each community in the study area is "flattening the curve" (or not). We added a different 'slicer' to this page, choosing instead to allow users to select specific communities from a drop-down menu of communities (alphabetized).

The third page in our overall report, "Facilities" included once again the point map of cumulative cases and infection rates, but instead included a table and a horizontal bar graph indicating the number of COVID-19 cases by place for SNFs and RCFs. This was a useful graphic (and table) because it allows users to quickly analyze the relative contribution of these facilities to the COVID-19 caseload in each community. A similar graphic could be included that showed deaths, rather than simple cases of the disease. Repeated inquiries to the Los Angeles County Department of Health regarding the nature of the inclusive or exclusive relationship between the overall caseload and the number of cases associated with institutional settings went unanswered. Therefore, we assumed, possibly incorrectly, that the facility data was not included in the sum of cases per community.

4. Study Limitations

Although we consider our experience with Power BI successful, there are some limitations to the software and to the maps we constructed. The main problem with PowerBI is the limited number of mapping options, especially for those who want to construct and share polygon/choropleth maps. Power BI does have an integration with ESRI, which is a promising development, but as of this writing, their interface is restricted to the desktop application, so sharing dynamic maps using custom, or ESRI shapefiles is limited. Microsoft does provide several common polygonal basemaps for making choropleth maps, including the US states and counties. From the basemap of cities pre-packaged with the software, the matching algorithm was unable to distinguish communities within Los Angeles County from those beyond Los Angeles County when more than one location shared a single place name. To overcome this limitation, we used latitude and longitude coordinates to display our data. The lack of attention to mapping tools extends to a sparse consideration for traditional cartographic principles as well. For example, Power BI includes no provision for adding a scale bar, a north arrow, or legend (in the traditional sense). However, we felt comfortable that for our primarily local audience, a north arrow was unnecessary, and the legend is replaced by a dynamically connected table and timeline-curve that instantly displays the data associated with each point on the map.

Another irritation with the PowerBI software is the difficulty we had in understanding the process for automating the data refresh task. Because the data from the Health Department is updated daily, the maps and graphics should be updated daily as well. One can manually refresh the data easily, and for the data served by the GitHub, the process was also reasonably simple. However, for the static data tables stored as Excel files on the Microsoft cloud drive (One Drive) there were some additional steps, including the installation of a "Gateway" software that allowed PowerBI to access the files from a cloud drive.

The other limitations or problems we encountered with the entire process were the inconsistencies and/or peculiarities with the data provided by the state and the Los Angeles Times. Again, the presence of distinct locations in their database that shared a common name identifier presented problems. For example, "Brentwood" is a neighborhood in Los Angeles and a city in Contra Costa County. The software combined the data from both cities in some of the data visualizations, but the mapping tool treated them separately. We eventually solved the problem by filtering the excluded counties outside our study area data prior to importing, rather than afterward as we had initially thought sufficient. Troubles of this sort are common for research leveraging public datasets produced by persons unaccustomed to, or unconcerned with the secondary uses of the data, like mapping and/or statistical analysis, and so this type of limitation is not exclusive to this study.

5. Conclusions

The implications of our mapping project are several. First, we were able to produce several high-quality, dynamic, data visualizations of COVID-19 trends in Southern California, none of which, we presume, were being created by any other public or private research interest, or media outlet during 2020. Among the unique elements of our mapping and data visualization efforts were a community-level map of COVID-19 cases displayed as a clickable point map, that was dynamically linked on the 'dashboard' to a data table and a trend-line graphic that showed both overall cases and new daily cases, and a five-day rolling average of new cases. We also produced a series of maps showing how residential care facilities, including nursing homes and assisted living communities were factoring into the overall prevalence of the disease in Los Angeles County. This was important because of the outsized effect residential care facilities and nursing homes were having, at least initially, on the overall trends countywide, and upon the disease count in several communities. These maps and visualizations were widely shared on social media and hundreds of thousands of regular site visits suggest that it was a useful tool in the dissemination of important public health information.

The other important implication of our mapping project is that we demonstrated that two persons with limited database and programming knowledge could produce a useful, dynamic, and web-enabled series of maps using a free, or at least very inexpensive software. We were able to find public data, scrub the data, insert it into a dynamically linked database structure and make multiple visualizations of the data within hours of installing the software on ordinary laptop computers. The senior researcher, has several decades of experience with ESRI products and is an advanced user of Microsoft Excel, which was valuable in terms of conducting some of the ancillary steps in the process, but overall, the PowerBI software was intuitive and exceptionally easy to use for basic visualizations. Although the formulas necessary for completing calculations between two different datasets echoes what one might do in Microsoft Excel, the two more sophisticated visualizations were accomplished only after we watched a YouTube video or two to learn how to write a formula in the DAX programming language.

The low barriers to adoption and the gentle learning curve of this software will permit many novice data analysts and geographers to attain mastery of the main functions of this software within hours. Certainly, there are other plug-and-play mapping applications with 'smart' functionality allowing novices to make simple maps, but we are unaware of any other application that combines that functionality with facile database management and an array of complex data visualization tools – beyond the mapping.

We suspect that within a few years, many small businesses, government offices, schools, researchers and even ordinary citizens previously thwarted in their data sharing, analysis and transparency efforts by the complexity of the software and the logic of database management will adopt this, or similar software platforms. This platform may also prove useful as a pedagogic tool, especially in introductory programming, geography and/or data visualization courses. Empowered by the simplicity of analyzing, displaying and sharing data, it is plausible that we shall see far more efforts like our own in the near future.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

California Department of Public Health, 2020. Skilled Nursing Facilities. [Data set]. https://www.cdph.ca.gov/Programs/CID/DCDC/Pages/COVID-19/SNFsCOVID_19.aspx.

California Department of Social Services, 2020. Nursing homes addresses. [Data set]. https://www.cdss.ca.gov/#covid19, https://www.ccld.dss.ca.gov/carefacilitysearch/.

Dolan, J., Mejia, B., 2020. As coronavirus raged through nursing homes, inspectors found nothing wrong. [Online]. https://www.latimes.com/california/story/2020-06-28/coronavirus-nursing-homes-state-inspector-covid-19 (Accessed 30 July 2020).

Franch-Pardo, I., Napoletano, B. M., Rosete-Verges, F., Billa, L., 2020. Spatial analysis and GIS in the study of COVID-19. A review. The Science of the Total Environment 739, https://doi.org/10.1016/j.scitotenv.2020.140033.

Kamel Boulos, M. N., Geraghty, E. M., 2020. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. International Journal of Health Geographics 19, 8, https://doi.org/10.1186/s12942-020-00202-8.

Los Angeles County, 2020. Enterprise Geographic Information Systems. [Data set]. https://egis-lacounty.hub.arcgis.com/.

Los Angeles County Department of Public Health, 2020. COVID-19 Data. [Data set]. http://publichealth.lacounty.gov/media/coronavirus/data/index.htm.

Los Angeles Times, 2020. COVID-19 Data (GitHub). [Data set]. https://github.com/datadesk/california-coronavirus-data.

Lucas, B., Jul. 2020. Covid-19 Maps as of July 18th. https://facebook.com/brett.lucas.5/posts/10224143839087562 #CBSNews #NBCNews #ABCNews #FoxNews #CNN #Maptitude [Twitter]. https://twitter.com/BLucas71/status/1284976316278689792 (Accessed 1 October 2020).

This article accompanies the following material:
HTML: https://doi.org/10.22201/igg.25940694e.2020.2.74
Dashboard: https://doi.org/10.22201/igg.25940694e.2020.2.74.160