Exploring Canadian Geospatial Data with GeoPandas in Colab Notebooks

Software & Data Reviews

Section Editor: Nicole Stradiotto, Brock University

Author: Christopher Macdonald Hewitt,
Assistant Professor (PT), Western University, Research Associate, Wilfrid Laurier University

Keywords: Google Colab, Python, GeoPandas, Open Source, Statistics Canada, Cartography

Abstract

Google Colab with the GeoPandas Python package is an effective platform for GIS analysis. Following a tutorial prepared by the author, this review discusses how users can load, project, query and visualize GIS data through graphs and maps. Future directions of where this platform could be taken are also discussed. Lastly, the benefits and drawbacks of GIS analysis with Google Colab are presented.

Introduction

Google Colab is a Jupyter Notebook that you can operate in your web browser, made available by Google (n. d.). Users upload their data or can access it through a link. The content that a user might add to a Colab file is the same as a regular Jupyter Notebook: a mix of narrative explanatory text, code, and output including visualizations (Google, n. d.). Multiple Python packages are already installed, including GeoPandas, which allows users to manage geographic data with a traditional Pandas dataframe (Van den Bossche et al., 2022; Van den Bossche et al., 2023). In this review, Google Colab and the GeoPandas package are evaluated in accordance with a 1-hour workshop developed by the author using both. First, the tutorial itself will be described in detail. This will be followed by a discussion of the merits of the GeoPandas package in a Google Colab environment for GIS analysis.

Tutorial

The tutorial can be accessed here and it is divided into four sections. These include an introduction, projections, information on working with data, and map production. It was presented virtually at the fall 2023 GIS Day conference. It was also incorporated into an introductory GIS course at Western University. Participants were instructed to have a Google account beforehand and were provided with the link. There were no other requirements.
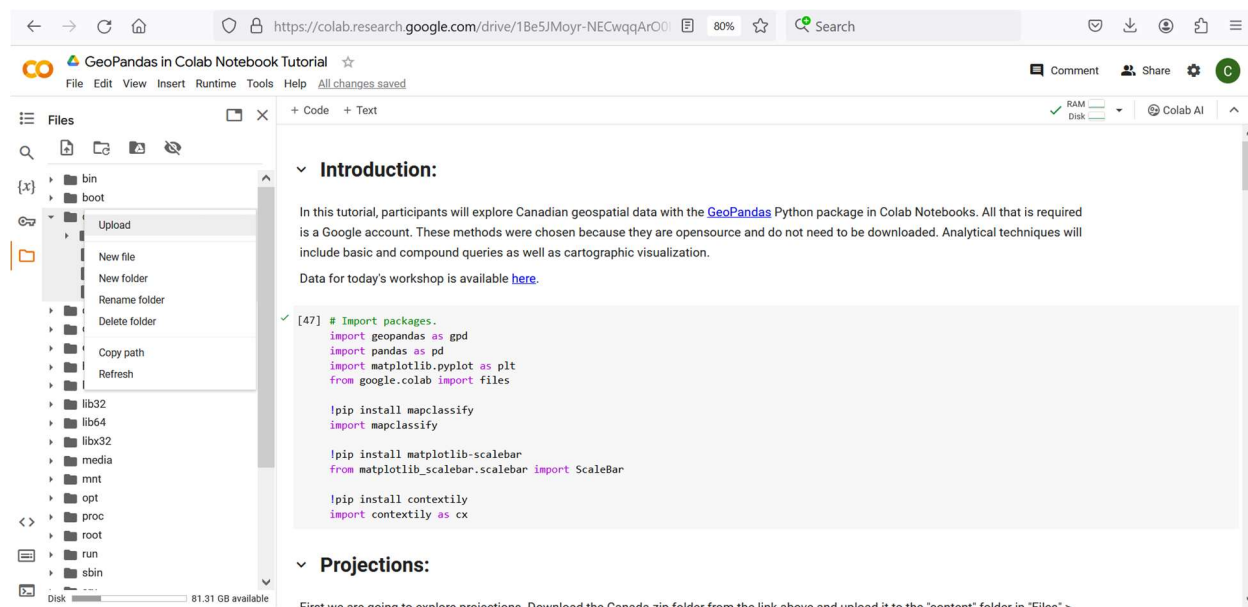

Figure 1: Uploading data

Readers are encouraged to experiment first-hand with the tutorial at the link above, but what follows is a narrative explanation summarizing the experience. Users are first instructed to download the data which were provided by the author but originated from *Scholars Geoportal* and the *Computing in the Humanities and Social Sciences* (*CHASS*) website (Ontario Council of University Libraries, n. d.; University of Toronto, 2010). They are then required to upload it to Google Colab and place it in the 'content' folder (Figure 1). Next, the required GeoPandas, Pandas and matplotlib.pyplot packages are imported and the mapclassify, matplotlib-scalebar and contextily packages are installed and imported. Projections are discussed with a map of

Canada from the 2001 census. The first projection is the North American Datum 1983, which is the default projection of this dataset. The Canada map is then reprojected in the WGS 84 / World Mercator (EPSG 3395) and the NAD83(CSRS) / Statistics Canada Lambert (EPSG 3348) projections, and visualized. This task is to demonstrate: 1) how to easily reproject data and 2) how the shape of large land bodies such as Canada's changes on a map with different projections. Code is provided to export the map at 300 DPI (Figure 2).
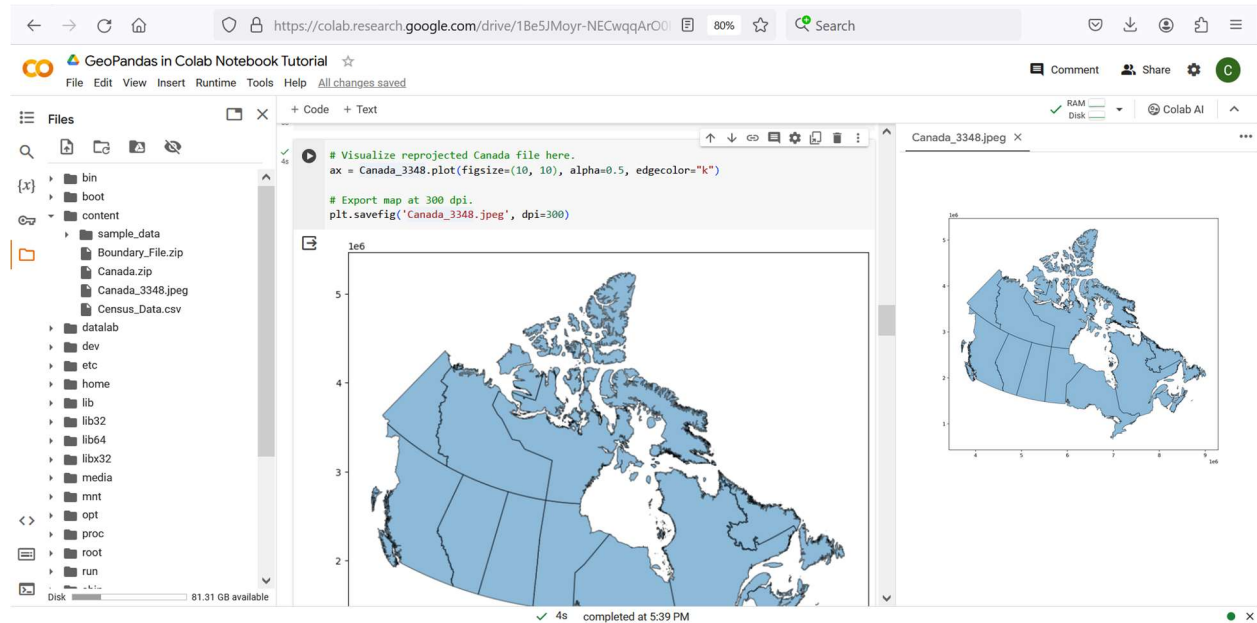

Figure 2: Exporting Map of Canada

Next, the notebook section "Working with data" is presented, and users are prompted to access 2016 Census data for the St. Catharines Census Metropolitan Area (CMA). Data columns were renamed, and data tables were joined together. Then, graphs including boxplots, histograms and scatter plots were visualized to showcase the value to statistics education. Multiple queries were presented. For instruction purposes, the Python code for some queries is not complete, allowing participants to fill in the code themselves. Lastly, map production is presented with different choropleth classification schemes and colour ramps. The final cell has participants add in a scale bar. Code is provided to export the map at 300 DPI.
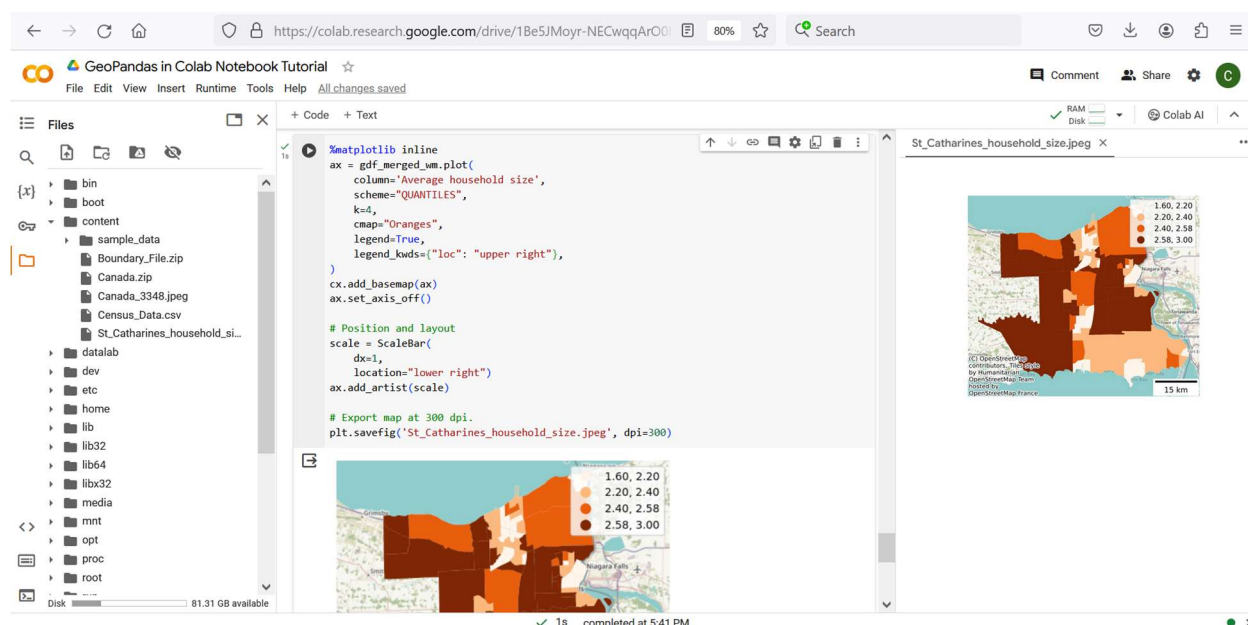
Figure 3: Exporting Household Size Map of St. Catharines CMA

Future Directions

There are a number of future directions that students and researchers could take using the skills introduced in this tutorial. For one, the package's different uses and functionality–including clips, overlays and spatial joins, as well as links to other Python packages such as PySAL– can be viewed through the link to the GeoPandas documentation (Van den Bossche et al., 2022; Arribas-Bel et al., 2020).   Additionally, from a statistics perspective, it would be straightforward to incorporate aspatial and spatial statistical tests into the scripts (e.g. Rogerson, 2021).   Lastly, to build user confidence, more sections of code could be left incomplete for users to fill out.

Discussion

There are multiple benefits to running the code and analysis on an online platform such as Google Colab. For one, no software, packages or updates need to be downloaded as they are all accessed on the Google cloud. This is contrary to RStudio, for example, where new software and packages need to be downloaded and updated from time to time (RStudio Team, 2020). Another important benefit is the ability to easily collaborate on projects. In Google Colab, notebooks can be shared like any other document with viewing or editing mode enabled. Viewing mode works for course content to students as was demonstrated here. Alternatively, editing mode allows multiple authors to edit and revise a document at the same time from different locations. Thus, the notebook is an excellent tool for collaborative projects.

However, there are a selection of drawbacks to this framework. For one, researchers have to upload the data or check the link each time they run the code. This could take time, especially in a class setting if the input files are large. Also, the packages not part of Google Colab have to be installed each time you run the code. Additionally, while the output maps can be printed as images, the entire notebook cannot be downloaded as a PDF. This functionality would be extremely helpful for those who wish to present the notebook as a document.

Conclusion

In conclusion, this tool will make an important contribution to collaborative research in the GIS field as well as teaching the foundational elements in introductory GIS and geography-based statistics. As was demonstrated in the tutorial, geospatial data can be easily viewed, reprojected and particular variables visualized. The links to other packages would enable further spatial analyses to be performed. Incorporated graphs and statistical tests highlight the value to geography-focused statistics education as well. Thus, there are many benefits to using a Google Colab notebook for GIS or statistical teaching and analysis.

References:

Arribas-Bel, D., Knaap, E., Barcelos, G., Shao, H., Gaboardi, J., Sauer, J., et al. (2020). *PySAL*. https://pysal.org/

Google. (n. d.). *Welcome to Colaboratory - Colaboratory*. https://colab.research.google.com/notebooks/intro.ipynb

Ontario Council of University Libraries (n. d.). *Scholars GeoPortal*. geo.scholarsportal.info

Rogerson, P. A. (2021). *Spatial Statistical Methods for Geography*. London: Sage.

RStudio Team. (2020). *RStudio: Integrated development for R*. Boston: RStudio, Inc. http://www.rstudio.com/.

University of Toronto (2010). *Computing in the Humanities and Social Sciences*. http://chass.toronto.edu/facilities/

Van den Bossche, J., Fleischmann, M., McBride, J., Ward, B., Wolf, L., & Richards, M. (2022). *GeoPandas 0.dev+untagged*. https://geopandas.org/en/stable/index.html

Van den Bossche, J., Jordahl, K., Fleischmann, M., McBride, J., Wasserman, J., Richards, M., et al. (2023, June 6). *geopandas/geopandas*: v0.13.2 (Version v0.13.2). Zenodo. http://doi.org/10.5281/zenodo.3946761