Visualization of google mobility data for provinces in Indonesia using seaborn python programming package

R Aurachman¹

¹Telkom University, street Telekomunikasi Number 1, Bandung, Indonesia

E-mail: rioaurachman@telkomuniversity.ac.id

Abstract. One important data needed in handling the COVID-19 pandemic is human mobility data. Based on mobility data provided by Google, analysis can be done as a basis for decision making. This research produces several visualization methods that provide insight to decision makers. Visualization is done using Python Programming

1. Introduction

The COVID-19 pandemic influences the behavior of the people. One behavior that changes is mobility. Society tends to reduce mobility. Changes in behavior affect economic conditions in an area. The government needs a guide to understand the change in mobility. The knowledge gained can be a reference in making accurate and right decisions. These decisions are taken to mitigate the risks and impacts felt by the community.

Google mobility report releases mobility data. The data records changes in people's behavior, especially mobility during the COVID-19 pandemic. The data provided by Google includes data per province in Indonesia for each date since February. The data does not specifically display mobility figures. The data shown is a mobility value relative to the baseline. The data needs to be interpreted correctly and quickly. One of the tools that can be used is visualization using data processing software that automatically visualizes data according to user needs. It is hoped that with this data, it will provide quick insight to decision makers so that they can make.

Some researchers study visualizations based on data from Google [1] [2] [3]. The visualization was done outside the interests of handling the COVID-19 pandemic. Other research examines the importance of studies on human mobility for handling COVID-19 [4] [5] [6] [7]. Meanwhile research on visualization using Python Programming also continues to grow [8] [9] [10]. However, no specific study has yet been found that focuses on methods of visualization and analysis of human mobility in the midst of the COVID-19 pandemic. So this research is intended to contribute to the theme.

2. Method

The data used is COVID-19 Community Mobility Report. The data is in the form of a table that contains mobility data for each country. Some countries are also equipped with data for each region and sub-region. The data displayed includes certain dates that are quite complete. Some dates are not available. Google makes statements about when data can be included and when data cannot be listed.

The following table shows the locations where the mobility data are listed. The definition is obtained from Google:
Table 1. Definition of Data

| Classification          | Definition                                                                 |
|-------------------------|---------------------------------------------------------------------------|
| Retail and recreation   | Restaurants, cafes, shopping centers, theme parks, museums, libraries, and movie theaters. |
| Grocery and pharmacy    | grocery markets, food warehouses, farmers markets, specialty food shops, drug stores, and pharmacies |
| Parks                   | national parks, public beaches, marinas, dog parks, plazas, and public gardens |
| Transit stations        | Mobility trends for places like public transport hubs such as subway, bus, and train stations |
| Workplaces              | Mobility trends for places of work                                          |
| Residential             | Mobility trends for places of residence.                                    |

Processing data using python programing. The packages used mainly are Panda and Seaborn. In the initial stage, data preparation is carried out. The initial data has the CSV (Comma Separated Value) format. First, the process of importing data becomes a form of Dataframe. From the dataframe then selected so that a subset of data frames is made specifically for the State of Indonesia. Indonesian dataframes are then processed using a number of basic descriptive statistical techniques and visualized using the available tools.

3. Result and Discussion

Indonesian dataframes are called dfina. The dataframe is then analyzed initially using the dfina. describe () feature. The data needs to be transposed first in order to get a summary for each column in the mobility area.

rangkuman=dfina.describe().transpose()

Table 2. Data Summary

| Classification          | count  | mean     | std       | min  | 25%   | 50%   | 75%   | max  |
|-------------------------|--------|----------|-----------|------|-------|-------|-------|------|
| Retail and recreation   | 2660   | -20.63759| 19.281145| -89  | -37   | -24   | -1    | 22   |
| Grocery and pharmacy    | 2660   | -10.51015| 12.826330| -92  | -20   | -9    | 0     | 27   |
| Parks                   | 2660   | -18.99436| 19.048274| -78  | -32   | -18   | -4    | 53   |
| Transit stations        | 2637   | -27.95601| 26.079187| -95  | -51   | -32   | -2    | 55   |
| Workplaces              | 2660   | -13.98609| 18.055350| -89  | -29   | -13   | 4     | 24   |
| Residential             | 2660   | 8.2229323| 6.9624469| -2   | 1     | 10    | 14    | 37   |

It can be seen in Table 2 that residential has the largest increase with an average of 8.22 and a standard deviation of 6.96. While the lowest mobility is in Transit Stations which is at -27.9% compared to the baseline.

The next step is to analyze the correlation between mobility data. Correlation is done using the features on the panda package namely .corr (), round (1). Then the visualization is done using a feature in the seaborn package, namely sns.heatmap (). Correlation analysis is performed on World and Indonesian data.
correlation_matrix = dfina.corr().round(1)
sns.heatmap(data=correlation_matrix, annot=True)
correlation_matrix = df1.corr().round(1)
sns.heatmap(data=correlation_matrix, annot=True)

**Figure 1.** Correlation Between Variable of Mobility (World and Indonesia)

It can be seen that the correlation of residential value is negative with other mobility both in Indonesia or the world. What is quite different is the correlation between residential and parks. The correlation between the two variables for Indonesia is more negative than the world.

The next step is to conduct a comparative analysis of mobility for all regions in Indonesia for each mobility variable. The visualizer uses the features of the panda package; pd.pivot_table(). Visualization uses the sns.heatmap() feature.

heatmap1_data = pd.pivot_table(dfina, values='retail_and_recreation_percent_change_from_baseline', index=['sub_region_1'], columns='date')
sns_plot=sns.heatmap(heatmap1_data, cmap="Blues")

**Figure 2.** Retail and recreation percent change from baseline

**Figure 3.** Grocery and pharmacy percent change from baseline
Figure 4. Parks percent change from baseline

Figure 5. Transit Station percent change from baseline

Figure 6. Workplace percent change from baseline

Figure 7. Residential percent change from baseline

Each picture contains a trend of changes in community mobility for each date for each region in Indonesia. Bright colors indicate small numbers. While the dark colors indicate large numbers. For residential heatmaps in all provinces are dark. This shows the occurrence of higher mobility in the house. Whereas for other heatmaps change to bright colors. This shows a reduction in mobility for other regions. And that happened fairly evenly in each province. It needs to be further analyzed how the application of lockdown and its effect on mobility.

Then the correlation calculation can be done more specifically for certain variables. Visualization of the correlation between two variables can use sns.lmplot().

    sns.lmplot('workplaces_percent_change_from_baseline', 'residential_percent_change_from_baseline', data=dfina, fit_reg=False)
For two indicators that have a positive correlation, the gradient formed has a positive value. When one indicator increases in value, the other indicator will increase. As for the two indicators that have a negative correlation, it will give a negative gradient value. When one indicator increases in value, the other indicator will decrease.

In addition to analysis on a national scale, analysis and visualization on a regional scale can also be carried out. As an example in Figure 10, there is a change in the mobility of Jakartans from mid-February to the end of April. It can be seen that since the beginning of mid-March there has been a change in the mobility behavior of the people. This continued until the end of April. An analysis of the application of lockdown to this mobility can also be carried out.

4. Conclusion

Analysis and visualization can be done using Python programming. This was done with the help of the Pandas and Seaborn libraries. Several programming lines have been obtained which can later be used to analyze and visualize the same or different data. By analyzing and visualizing it is gained new
knowledge and insight in understanding conditions and problems. Knowledge and insight are difficult to obtain if you only see rows of numbers in the data table.

Some development of this research can be done. One of them is to test the correlation of mobility variables with other variables. Can also be compared between several mobility variables between several regions. In addition, evaluation research can also be conducted on a policy that is applied to an area based on changes in mobility that occur.

Acknowledgments

Authors wishing to acknowledge assistance or encouragement from google for releasing this data.

5. References

[1] F. Wu, M. Zhu, Q. Wang, X. Zhao, W. Chen and R. Maciejewski, “Spatial--temporal visualization of city-wide crowd movement,” Journal of Visualization, vol. 20, pp. 183-194, 2017.

[2] T. Sobral, T. Galvão and J. Borges, “Visualization of urban mobility data from intelligent transportation systems,” Sensors, vol. 19, p. 332, 2019.

[3] E. Mešković and D. Osmanović, “A system for monitoring and visualization of big mobility data,” in 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2018.

[4] M. U. G. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D. M. Pigott, L. Plessis, N. R. Faria, R. Li, W. P. Hanage and others, “The effect of human mobility and control measures on the COVID-19 epidemic in China,” Science, vol. 368, pp. 493-497, 2020.

[5] N. Oliver, E. Letouzé, H. Sterly, S. Delataille, M. De Nadai, B. Lepri, R. Lambiotte, R. Benjamins, C. Cattuto, V. Colizza and others, “Mobile phone data and COVID-19: Missing an opportunity?,” arXiv preprint arXiv:2003.12347, 2020.

[6] P. Bonato, P. Cintia, F. Fabbri, D. Fadda, F. Giannotti, P. L. Lopalco, S. Mazzilli, M. Nanni, L. Pappalardo, D. Pedreschi and others, “Mobile phone data analytics against the COVID-19 epidemics in Italy: flow diversity and local job markets during the national lockdown,” arXiv preprint arXiv:2004.11278, 2020.

[7] C. O. Buckee, S. Balsari, J. Chan, M. Crosas, F. Dominici, U. Gasser, Y. H. Grad, B. Grenfell, M. E. Halloran, M. U. G. Kraemer and others, “Aggregated mobility data could help fight COVID-19.,” Science (New York, NY), vol. 368, p. 145, 2020.

[8] J. VanderPlas, B. Granger, J. Heer, D. Moritz, K. Wongsuphasawat, A. Satyanarayan, E. Lees, I. Timofeev, B. Welsh and S. Sievert, “Altair: Interactive statistical visualizations for python,” Journal of open source software, vol. 3, p. 1057, 2018.

[9] K. Dale, Data visualization with python and javascript: scrape, clean, explore & transform your data, ” O'Reilly Media, Inc.”, 2016.

[10] D. A. R. Eaton, “Tooytree: A minimalist tree visualization and manipulation library for Python,” Methods in Ecology and Evolution, vol. 11, pp. 187-191, 2020.