

# Economic analysis through alternative data and big data techniques: what do they tell about Brazil?

Matheus Pereira Libório (✉ [m4th32s@gmail.com](mailto:m4th32s@gmail.com))

Pontifícia Universidade Católica de Minas Gerais <https://orcid.org/0000-0003-1411-0553>

Petr Iakovlevitch Ekel

Pontifical Catholic University of Minas Gerais: Pontificia Universidade Catolica de Minas Gerais

Carlos Augusto Paiva da Silva Martins

Pontifical Catholic University of Minas Gerais: Pontificia Universidade Catolica de Minas Gerais

---

## Research Article

**Keywords:** alternative data, Google Trends, Google Mobility, big data, economic analysis

**Posted Date:** January 28th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1268799/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Alternative data are now widely used in economic analyses worldwide but still infrequent in studies on the Brazilian economy. This research demonstrates how alternative data extracted from Google Trends and Google Mobility contribute to innovative economic analysis. First, it demonstrates that the search for the future on the internet is correlated ( $R=0.62$ ) with the average household income in Brazilian states. The three Brazilian states with the most people looking for the future on the internet have an average household income 1.6 times higher than people from states that do not have this behavior. The search for the future represents 10.9% of the economic development potential of the states, while the proportion of people with university degrees, scientific publications, and researchers represents another 60.4%. The reduction in mobility in retail/recreation locations averaged 34.28% in Brazil, Ecuador, Paraguay, and Uruguay. This group of countries had COVID-19 infection and death rates 1.25 and 1.74 times higher than in countries that reduced their mobility in retail/recreation locations by 45.03%. The impact of reduced mobility in retail/recreation locations on the unemployment rate, gross domestic product degrowth, and inflation in countries such as Brazil was 1.1, 2.2, and 2.6 times lower than in countries that reduced mobility more of people. The research contributions are associated with identifying new indicators extracted from alternative data and their application to carry out innovative economic analyses.

**JEL Code:** A12; C00; E01; J1

## 1. Introduction

Big data machine learning techniques that enable the treatment and analysis of large volumes of data have favored diverse economic analyzes through the discovery and application of hitherto unknown sets of indicators (Cheng et al. 2021). This new set of indicators of great interest for economic analysis was called alternative data. Although many researchers use the terms big data and alternative data synonymously (Hassani and Silva 2015; Buono et al. 2017; Narita and Yin 2018), alternative data are conceptualized as data not yet fully incorporated into the economic mainstream (Jain 2019).

More and more researchers have taken advantage of big data advances in economic analysis based on alternative data (Garboden 2020; Blazquez and Domenech 2018). The books "Big data and AI strategies: Machine learning and alternative data approach to investing" (Kolanovic and Krishnamachari 2017) and "The Book of Alternative Data: A Guide for Investors, Traders, and Risk Managers" (Denev and Amen 2020) offer useful information for practitioners in this field. Among this information are concepts, types, and alternative data sources, techniques and methods of big data and econometrics, as well as the main risks and challenges faced by practitioners in the field.

Examples of applications of alternative data in economic analysis are easily found in the literature (Varian 2014; Blazquez and Domenech 2018; Charoenwong and Kwan 2021). Among these examples, the work of Choi and Varian (2012) occupies a prominent position. The research entitled "Predicting the

Present with Google Trends" developed by Google economists Hyunyoung Choi and Hal Varian exerts enormous influence among researchers in the field.

In particular, economic analysis based on alternative data extracted from Google Trends and Google Mobility are already very popular. Economists worldwide have been presenting several findings from these data, bringing relevant contributions to the field of economics. Although there is sufficient evidence in the literature on the usefulness of data extracted from Google Trends and Google Mobility in economic analyses, little is known about using this type of data in economic analysis in Brazil.

Therefore, the objective of this research is twofold. First, to explain the recent popularity of alternative data in economic analyses, indicating how much this popularity is reflected in studies on the Brazilian economy. Second, to use alternative data from Google Trends and Google Mobility, indicating how these data contribute to understanding the Brazilian economic reality.

The originality of this research is associated with the discovery of indicators extracted from alternative data for carrying out innovative economic analyses, such as the identification of the factors that most impact the potential for the economic development of Brazilian states and the analysis of the impact of COVID-19 on mobility in retail/recreation locations and the relationship of this impact with economic indicators from Brazil and other South American countries.

## **2. Alternative Data Applied To Economic Analysis**

### **2.1. Google trends data**

Until December 17, 2021, 385 articles were indexed in Scopus and Web of Science databases that explicitly mention the use of Google Trends data in the title, abstract, or keywords published in economic journals.

These articles bring quite varied contributions. Choi and Varian (2012) demonstrated how to use data extracted from Google Trends to predict short-term economic indicators such as car sales, unemployment claims, and consumer confidence. Vosen and Schmidt (2011) introduced a new indicator of private consumption from data extracted from Google Trends. The authors show that the new indicator performs better than research-based indicators, suggesting that adding Google Trends data helps significantly to predict private consumption (Vosen and Schmidt 2011). Carrière-Swallow and Labbé (2013) introduced a car interest index based on data extracted from Google Trends to forecast car sales in Chile. Kearney and Levine (2015) use Google Trends data to analyze the influence of television programs on interest in contraceptives and carrying out abortions.

While Google Trends data provide information for several economic analyses, a significant portion of publications that use this type of alternative data focus on oil and stock market analysis.

Yu et al. (2019) show that using data extracted from Google Trends significantly improves the prediction of the direction and level of oil consumption. Ma et al. (2019) introduce the event-triggered indicator through data extracted from Google Trends to predict oil price volatility. Qadan and Nama (2018) reveal that the greatest interest in the oil topic is related to oil price shocks and that the data extracted from Google Trends is a good predictor of oil price volatility.

Ding and Hou (2015) employed data extracted from Google Trends to capture the active attention that retail investors pay to stocks. The authors show that retail investor attention significantly broadens the shareholder base and improves stock liquidity. Vlastakis and Markellos (2012) study the demand and supply for information on companies with shares traded on stock exchanges. Based on Google Trends data, the authors reveal that risk aversion levels are positively related to the demand for information about stocks by investors.

Five studies that use data extracted from Google Trends in the economic analysis have been done by Brazilian authors. However, three of these studies have been dedicated to economic analyses for other countries and not for Brazil (Perlin et al. 2017; Corbi and Picchetti 2020; Piccoli and Castro 2021). Google Trends data is used by Neto and Candido (2021) to assess and predict loans to households and by Piccoli (2021) to predict currency prices in a speculative attack. This indicates that Google Trends data is still underused in Brazil despite its generally recognized usefulness, including in economic analysis.

## **2.2. Google mobility data**

Until December 17, 2021, 180 articles were indexed in the Scopus and Web of Science databases that explicitly mention the use of Google Mobility or Mobility Reports data in the title, abstract, or keywords.

Two reasons explain this lower number of studies compared to Google Trends. First, mobility data was only made available in 2020 to support studies on the COVID-19 pandemic. Second, the vast majority of articles are associated with the health area. Only twenty-four articles, or thirteen percent of the total, were published in economic journals.

Studies based on Google Mobility data focus on analyses that relate COVID-19 and the economy (Fernández-Villaverde and Jones 2020). These studies show that population vaccination rates are positively related to country wealth (Auld and Toxvaerd 2021). In the United States, COVID-19 cases would have been 17–78% higher in the states without containment measures such as business closures (Chernozhukov et al. 2021). Non-metropolitan indebtedness increased with reduced geographic mobility (Allen and Whitledge 2021). Mobility restriction measures are directly and negatively related to the economic value of green space for the citizens of England (Day 2020). Mobility in workplaces during the coronavirus pandemic is significantly lower in poorer regions than in other regions (Bargain and Aminjonov 2021).

The study by Silva et al. (2021) is the only one to analyze Brazil. The authors rely on Google Mobility data to show that the mobility in European and Canadian cities increased in the second wave compared to the first. In contrast, mobility in the first and second waves remained stable in the Brazilian city of Manaus.

## 3. Application Examples

Two case studies based on Google Trends and Mobility data were designed to demonstrate the applicability and usefulness of this type of alternative data in economic analysis in Brazil. The case studies bring a brief contextualization, followed by the materials and methods used, ending with the presentation and discussion of the results.

### 3.1. Unveiling the development potential of Brazilian states using Google trend data

Studies show that economic development is directly or indirectly associated with non-economic indicators. Several studies relate the gross domestic product to the level of education, number of Higher Education Institutions, number of scientific publications, number of scientific journals, number of researchers, and volume of expenditure on Research, Development, and Innovation (Meo et al. 2013; Tümer and Akkuş 2018; Pastor et al. 2018).

In turn, factors associated with Research, Development, and Innovation influence and are influenced by economic and technological factors (Libório et al. 2020). Broadband Internet access positively impacts labor productivity and is strongly associated with economic indicators such as gross domestic product and employment levels in the economy (Holt and Jamison 2009; Gruber et al. 2014; Li and Forzati 2018; Gallardo et al. 2021).

In an innovative approach, researchers show that the economic performance of countries can be explained and predicted through the behavior of people on the internet. How often people search for future years on the internet is correlated with the gross domestic product of countries (Preis et al. 2012). Time orientation refers to individual differences in the relative emphasis one places on the past, present, or future and is related to academic, financial, and health outcomes (Park et al. 2017). In short, these studies suggest that people from developed countries are more future-oriented than people from developing countries.

#### 3.1.1. Development Potential Composite Indicator (DP-CI): data and methods

Composite indicators are one-dimensional measures that facilitate the comprehension of multiple sub-indicators associated with multidimensional economic phenomena. Among these multidimensional economic phenomena, it is possible to mention the uncertainty (Charles et al. 2018), risk (Akin et al. 2016), costs of doing business (Bernardes et al. 2021; Ekel et al. 2022), sustainable development (Salvati and Carlucci 2014; Alaimo and Maggino 2020), competitiveness (Moirangthem and Nag 2022) among others (Mazziotta and Pareto 2016). Composite indicators are constructed by a method that aggregates normalized sub-indicators, weighted or not (El Gibari et al. 2019). Although there is no method exempt from limitations, researchers agree that composite indicators are tools that facilitate the

interpretation of complex realities and support decision-makers (Nardo et al. 2005; Kuc-Czarnecka et al. 2020).

This research employs Pearson's (1901) Principal Component Analysis to construct the Development Potential Composite Indicator. Principal Component Analysis is a very popular method in data science because it allows the reduction of a large volume of data in a few components with the least possible loss of the original information (Libório et al. 2018). The original data, that is, the sub-indicators, directly or indirectly related to economic development, which were used in the Principal Component Analysis, are presented in Table 1.

Table 1. Sub-indicators of the Development Potential Composite Indicator

ID	Subindicator	Description	Source
BA	Broadband access	Percentage of households with broadband internet	Brazilian National Telecommunications Agency <sup>[1]</sup>
FI	Future-oriented index	The search index for future years on the internet	Google Trends <sup>[2]</sup>
UD	University degree	Percentage of population with a university degree	Brazilian Institute of Geography and Statistics <sup>[3]</sup>
AI	Average income	Average monthly household income	Brazilian Institute of Geography and Statistics <sup>[4]</sup>
IC	Innovative companies	Percentage of companies classified as innovative	Ministry of Science, Technology, Innovations, and Communications <sup>[5]</sup>
RP	Research professionals	Percentage of the population employed in the research sector	Ministry of Labor and Social Security <sup>[6]</sup>
SP	Scientific papers	Articles published by a person	Web of Science <sup>[7]</sup>

**Note:** the future-oriented index was generated for each Brazilian state, considering the internet search for the word "2023" between January 2020 and December 2021.

Following the specialized literature, three parameters were considered to accept the Principal Component Analysis model. First, the variance explained in the first component is greater than 0.50 (Libório et al. 2021). Second, the sampling adequacy measured by the Kaiser-Meyer-Olkin criterion is greater than 0.70 (Kaiser 1974). Third, the degree of deviation between the correlation matrix and an identity matrix measured by Bartlett's (1937) sphericity test is lesser than 0.05.

### 3.1.2. Development potential of Brazilian states

In Brazil, future-oriented behavior is more common in the Federal District and states of Minas Gerais and Santa Catarina and less common in the Rio Grande do Norte, Amazonas, and Tocantins states.

Figure 1 shows that searching for the future on the internet positively correlates with the average income in Brazilian states.

These results suggest that the relationship between the search for the future with a higher gross domestic product per capita (Preis et al. 2012) can also occur on a regional scale. Therefore, it is possible to

consider the future-oriented index as a relevant sub-indicator for the Development Potential Composite Indicator. Figure 2 shows that the future-oriented index sub-indicator contributes 13% to the Principal Component Analysis model and 10.9% to the Principal Component, that is, to the Development Potential Composite Indicator.

These results indicate that sub-indicators related to research and teaching have greater weight in the Development Potential Composite Indicator. It also indicates that the sub-indicators representing technology and innovation have the lowest weights. In Figure 3, it is possible to conclude that these results are statistically consistent. The Principal Component Analysis model validity parameters Variance Explained and Kaiser-Meyer-Olkin exceeded the acceptance threshold of 0.50 and 0.70. The Bartlett test was below the maximum threshold of 0.05.

Except for the Federal District, the map in Figure 3 reveals that the south and southeast region concentrates the states with the greatest development potential. Maranhão, Alagoas, and Bahia, all in the country's northeast region, have the lowest development potential. These results show important regional inequalities between Brazilian states and regions. The development potential of the top five states in the ranking is, on average, 10.9 times greater than the development potential of the five lowest-ranked states. Furthermore, these results suggest that inequality tends to increase due to the states' development potential.

Therefore, the introduction of public policies aimed at promoting research and education can be a more efficient strategy to reduce regional inequalities with innovative investments or access to the internet. In particular, these policies should seek to increase the proportion of people at the university level engaged in research activities and the production of scientific knowledge.

## **3.2. Measuring the impact of retail and recreation mobility on economic indicators in South American countries through Google mobility data**

Studies that use Google mobility data reveal how people's mobility during the COVID-19 pandemic relates to economic indicators. These studies reveal that social distancing leads to declines in the growth rate of coronavirus cases (Milani 2021) and that self-protection, social distancing, and mask use is a behavior directly related to income (Papageorge et al. 2021). Studies also show that the reduction in population mobility was greater in cities with the highest socioeconomic index, in countries with higher sociodemographic indices and universal health coverage (Liu et al. 2021), and in areas where people trust medicine more. and science (Brodeur et al. 2021).

From an economic point of view, the correlation of COVID-19 cases with mobility in retail/recreation, grocery/pharmacy, and public transportation locations is greater than mobility in parks/workplaces or homes (Casa Nova et al. 2021). Mobility reduction policies impact the economy, particularly Italian,

German, French, and Spanish (Spelta and Pagnottoni 2021). Milani (2021) also shows that the impact of COVID-19 on unemployment is very heterogeneous, being higher in Spain and the United States and lower in countries that have adopted subsidy programs for employers and employees.

### **3.2.1. Mobility and COVID-19 in South America: data and multivariate analysis**

This case study aims to answer two questions. First, what was the impact of mobility in retail/recreation locations on the number of COVID-19 cases and deaths in South American countries? Second, how does this impact relate to economic indicators? Data concerning mobility, COVID-19, economic indicators, and the methods of Cooks distance, k-means cluster analysis, and Two-factor Analysis of Variance (ANOVA) with replication were employed to answer these two questions.

Mobility data at retail/recreation locations were taken from Google COVID-19 Community Mobility Reports[8]. Data from the pre-vaccination period, between February 15 and December 31 2020, were retrieved to ensure comparability across countries. Demographic and COVID-19 data were obtained from Our World in Data[9]. The economic indicators were obtained from the World Bank from DataBank, the World Development Indicators database[10].

*Cook's (1977) distance:* is a measure that allows identifying atypical elements in a multivariate set of data, which can distort the result and precision of multivariate analyses. As a general rule, observations with a Cook's distance greater than three times the mean are possible outliers (Bernardes et al. 2021).

*MacQueen's (1967) k-means clustering:* is an algorithm that uses the total variation of Euclidean distances from elements to cluster centroids to separate elements into k-clusters to maximize differences between clusters and minimize differences within each cluster. The average silhouette width of Rousseeuw's (1987) measures the groups' cohesion and resolution and guarantees their internal homogeneity and external heterogeneity. Average silhouette width above 0.5 indicates that the clusters have cohesion and resolution. Positive values indicate that the elements are well grouped. Negative coefficients indicate that the elements are poorly placed between the groups (Libório et al. 2021).

*Two factor ANOVA with replication:* is an analysis that tests the hypothesis of significant differences between the indicators of countries of different groups (Ekel et al. 2021). In other words, this analysis verifies that the differences between the means of the data sets are significant, allowing several groups to be compared simultaneously (Quirk 2012).

This second case study was developed in four stages. First, the monthly mobility variation in retail/recreation locations in South American countries was calculated. Second, the Cooks Distance multivariate outlier detection method was used to identify and exclude countries that showed atypical variations in monthly mobility. Third, k-means cluster analysis was used to group countries according to their similarity in monthly mobility in retail/recreation locations. Fourth, Two-factor ANOVA with replication was used to verify whether the impact of monthly variations in mobility in retail/recreation

locations on the number of COVID-19 cases and deaths and economic indicators is significantly different between country groups.

### **3.2.1. The impact of retail and recreation mobility on economic indicators in South American**

Figure 4 shows that Bolivia and Venezuela show atypical variations in monthly mobility in retail/recreation locations. These two countries were excluded from the cluster analysis that separated the countries into two groups.

Average silhouette width of 0.56, above the acceptance threshold of 0.50, indicates that monthly variations in mobility in retail/recreation locations are similar between countries within the same group and dissimilar between groups. All observations show silhouette width positive, indicating that countries were grouped correctly.

The Cluster 1 countries, Argentina, Chile, Colombia, and Peru, showed monthly mobility variations in similar retail/recreation locations. In ten of the eleven months analyzed, the reduction in mobility in retail/recreation locations in Cluster 1 countries was greater than in Cluster 2 countries, Brazil, Ecuador, Paraguay, and Uruguay.

From the Two Factor ANOVA with Replication, it is possible to state that the indicators of the two groups are statistically different at a confidence level of 0.05. Figure 5 shows that the reduction in mobility in retail/recreation locations in Cluster 1 countries was 10.75% greater than in Cluster 2 countries. The number of COVID-10 cases and deaths per thousand inhabitants was 1.25 and 1.74 times higher in Cluster 2 countries than Cluster 1 countries. It is noteworthy that countries with populations with higher average age and a higher proportion of older people do not seem to have adopted effective measures to reduce mobility in retail/recreation locations. The absence of these measures significantly increased the number of cases and deaths from COVID-19 in Cluester 2 countries.

Naturally, reduced mobility in retail/recreation locations has consequences for economic activity. The South American countries that most reduced mobility in retail/recreation locations had higher unemployment and inflation rates and a greater decline in gross domestic product. The most impacted economic indicators were gross domestic product degrowth and inflation. Cluster 1 countries had gross domestic product degrowth and inflation 2.2 and 2.6 times lower than Cluster 2 countries. In turn, the unemployment rate was only 1.1 times higher in the countries that most reduced mobility in retail/recreation locations.

These results are in line with the literature that associates social distancing and mobility in retail/recreation locations with a lower number of coronavirus cases (Milani 2021; Casa Nova et al. 2021). From an economic point of view, the results reinforce the evidence that associates reduced mobility with a worsening of the real economy (Spelta and Pagnottoni 2021). This research also suggests that the impact of COVID-19 on unemployment rates in South American countries is not very

different regardless of reduced mobility in retail/recreation locations. Similar to what was observed in Spain and the United (Milani 2021), these results suggest that populations in South American countries suffer from the absence of subsidy programs for employers and permanent employees.

- [1] [https://dados.gov.br/dataset/densidade\\_banda\\_larga](https://dados.gov.br/dataset/densidade_banda_larga)
- [2] <https://trends.google.com.br/trends/?geo=BR>
- [3] <https://sidra.ibge.gov.br/tabela/5919>
- [4] <https://sidra.ibge.gov.br/tabela/5442>
- [5] [https://antigo.mctic.gov.br/mctic/opencms/tecnologia/Lei\\_do\\_bem/Noticia/Resltados.html](https://antigo.mctic.gov.br/mctic/opencms/tecnologia/Lei_do_bem/Noticia/Resltados.html)
- [6] <http://pdet.mte.gov.br/>
- [7] <https://www.webofscience.com/wos/woscc/basic-search>
- [8] <https://www.google.com/covid19/mobility/>
- [9] <https://github.com/owid/covid-19-data/tree/master/public/data>
- [10] <https://databank.worldbank.org/source/world-development-indicators#>

## 4. Conclusions

The use of alternative data for economic analysis is becoming a common practice among researchers worldwide. Data extracted from internet searches and mobility can be easily obtained through Google Trend and Google Mobility platforms. Although these data allow for innovative analyses, this research shows that alternative data are still rarely used in Brazil. Only 0.52% of the publications found in the Scopus and Web of Science databases that explicitly mention Google Trends analyze the Brazilian economy. The case of Google Mobility is no different. Only 0.55% of publications analyze Brazil. This research demonstrates through two case studies how alternative data and big data techniques can contribute to the realization of innovative economic analyses in Brazil.

First, we show that searching for the future on the internet is associated with the higher average household income in the Brazilian states. The average household income in the three states where people are most looking for their future on the internet is 1.56 times higher than in the three states where people are least looking for their future. People in the three more forward-looking states have incomes 1.56 times higher than people in the three least forward-looking states. By aggregating this future-oriented index with six other economic and non-economic indicators through Principal Component Analysis, we revealed that the Federal District and the southern and southeastern states of Brazil have greater potential for economic development than the other Brazilian states.

The second case study shows the impact of COVID-19 on mobility in retail/recreation locations in South American countries and the relationship of this impact with economic indicators. The results show that countries like Brazil reduced mobility in retail/recreation locations by 34.28%. This reduction negatively impacted employment levels by 10%, economic growth by -4%, and inflation by 4%. The rates of infection and deaths by COVID-19 per thousand inhabitants in Brazil were 1.25 and 1.74 higher than in countries that reduced mobility in retail/recreation locations by 45.03%.

The alternative data and the big data techniques used in this research bring two innovative contributions to the Brazilian economic literature. First, the economic development potential of Brazilian states is more associated with knowledge factors than with technology, innovation, or income. Second, the smaller reduction in mobility in retail/recreation locations in Brazil compared to four other South American countries minimized the effects of the coronavirus pandemic on unemployment rates, economic growth, and inflation. However, this lower impact on economic indicators is associated with more infections and deaths from COVID-19 per thousand inhabitants.

## Declarations

### Compliance with ethical standards

**Funding:** This work was carried out with the support of the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Financing Code 001 and the National Council for Scientific and Technological Development of Brazil (CNPq) Grant 311032/2016-8.

**Disclosure of potential conflicts of interest:** The authors declare they have no potential conflicts of interest.

**Research involving Human Participants and/or Animals:** This article does not contain any studies with human participants or animals performed by any of the authors.

**Data availability:** <https://doi.org/10.17632/m3y4jncvch.4>

## References

Akin T, Iqbal Z, Mirakhor A (2016) The composite risk-sharing finance index: Implications for Islamic finance. *Review of Financial Economics* 31:18-25.

Alaimo LS, Maggino F (2020) Sustainable development goals indicators at territorial level: Conceptual and methodological issues—The Italian perspective. *Social Indicators Research* 147(2):383-419.

Allen KD, Whitley MD (2021) Further evidence on the effectiveness of community banks in the Paycheck Protection Program. *Finance Research Letters* 19:102583.

- Auld C, Toxvaerd F (2021) The Great COVID-19 Vaccine Rollout: Behavioral and Policy Responses. *National Institute Economic Review* 257:14-35. <https://doi.org/10.1017/nie.2021.23>
- Bargain O, Aminjonov U (2021) Poverty and COVID-19 in Africa and Latin America. *World development* 142:105422. <https://doi.org/10.1016/j.worlddev.2021.105422>
- Bartlett MS (1937) Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences* 160(901):268-82.
- Bernardes P, Ekel PI, Rezende SF, Pereira Júnior JG, dos Santos AC, da Costa MA, Carvalhais RL, Libório MP (2021) Cost of doing business index in Latin America. *Quality & Quantity*. <https://doi.org/10.1007/s11135-021-01221-8>
- Blazquez D, Domenech J (2018) Big Data sources and methods for social and economic analyses. *Technological Forecasting and Social Change* 130:99-113.
- Brodeur A, Grigoryeva I, Kattan L (2021) Stay-at-home orders, social distancing, and trust. *Journal of Population Economics* 19:1-34.
- Buono D, Mazzi GL, Kapetanios G, Marcellino M, Papailias F (2017) Big data types for macroeconomic nowcasting. *Eurostat Review on National Accounts and Macroeconomic Indicators* 1:93-145.
- Carrière-Swallow Y, Labbé F (2013) Nowcasting with Google Trends in an emerging market. *Journal of Forecasting* 32(4):289-98.
- Casa Nova A, Ferreira P, Almeida D, Dionísio A, Quintino D (2021) Are Mobility and COVID-19 Related? A Dynamic Analysis for Portuguese Districts. *Entropy* 23(6):786.
- Charles A, Darné O, Tripier F (2018) Uncertainty and the Macroeconomy: Evidence from an uncertainty composite indicator. *Applied Economics* 50(10):1093-107.
- Charoenwong B, Kwan A (2021) Alternative Data, Big Data, and Applications to Finance. In: Choi P.M.S., Huang S.H. (eds) *Fintech with Artificial Intelligence, Big Data, and Blockchain*. Blockchain Technologies. Springer, Singapore. [https://doi.org/10.1007/978-981-33-6137-9\\_2](https://doi.org/10.1007/978-981-33-6137-9_2).
- Cheng P, Ferrara L, Froidevaux A, Huynh TL (2021) Massive Data Analytics for Macroeconomic Nowcasting. In: Consoli S., Reforgiato Recupero D., Saisana M. (eds) *Data Science for Economics and Finance*. Springer, Cham. [https://doi.org/10.1007/978-3-030-66891-4\\_7](https://doi.org/10.1007/978-3-030-66891-4_7).
- Chernozhukov V, Kasahara H, Schrimpf P (2021) Causal impact of masks, policies, behavior on early covid-19 pandemic in the US. *Journal of econometrics* 220(1):23-62.
- Choi H, Varian H (2012) Predicting the present with Google Trends. *Economic record* 88:2-9.

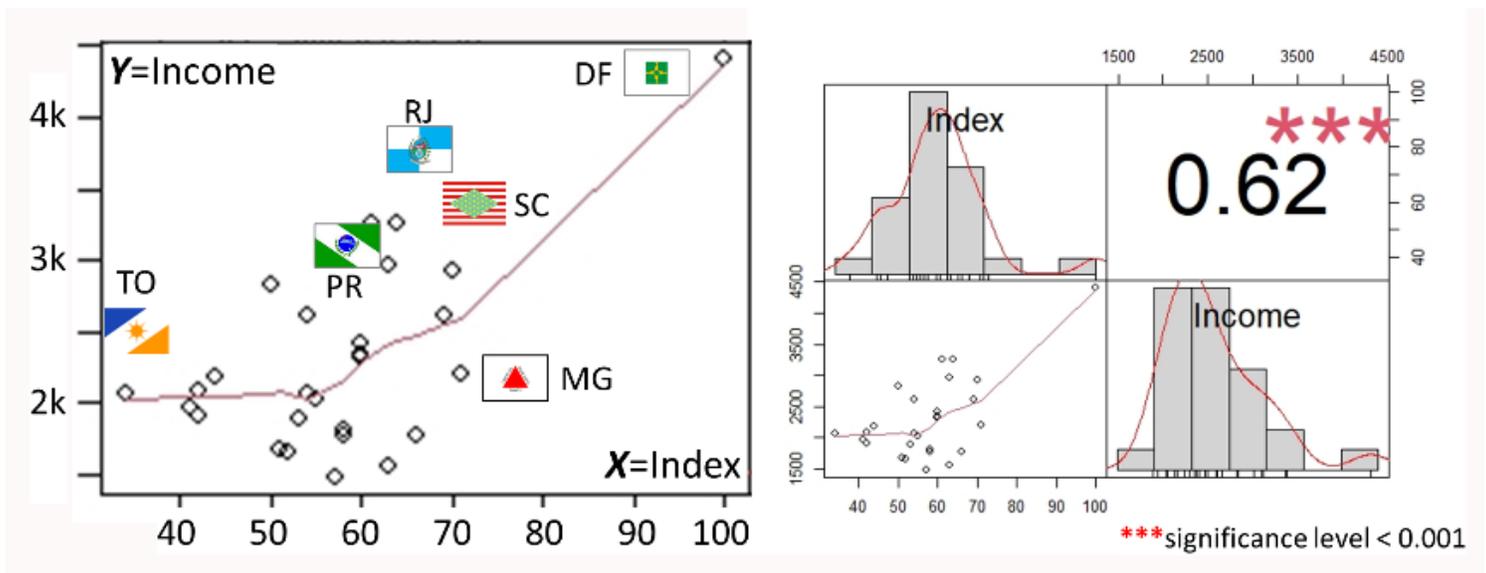
- Cook RD (1977) Detection of influential observation in linear regression. *Technometrics* 19(1):15-8. <https://doi.org/10.1080/00401706.1977.10489493>
- Corbi R, Picchetti P (2020) The cost of gendered attitudes on a female candidate: Evidence from Google Trends. *Economics Letters* 196:109495.
- da Silva GC, de Almeida FM, Oliveira S, Wanner EF, Bezerra LC, Takahashi RH, Lima L (2021) Comparing community mobility reduction between first and second COVID-19 waves. *Transport Policy* 112:114-24.
- Day BH (2020) The value of greenspace under pandemic lockdown. *Environmental and Resource Economics* 76(4):1161-85.
- Denev A, Amen S (2020) *The Book of Alternative Data: A Guide for Investors, Traders and Risk Managers*. John Wiley & Sons.
- Ding R, Hou W (2015) Retail investor attention and stock liquidity. *Journal of International Financial Markets, Institutions and Money* 1;37:12-26.
- Ekel PI, Bernardes P, Vale GMV, Libório MP (2022) South American business environment cost index: reforms for Brazil. *International Journal of Business Environment*.  
<http://dx.doi.org/10.1504/IJBE.2021.10043256>
- El Gibari S, Gómez T, Ruiz F (2019) Building composite indicators using multicriteria methods: a review. *Journal of Business Economics* 89(1):1-24.
- Fernández-Villaverde J, Jones CI (2020) Macroeconomic outcomes and COVID-19: a progress report. National Bureau of Economic Research. <http://dx.doi.org/10.3386/w28004>
- Gallardo R, Whitacre B, Kumar I, Upendram S (2021) Broadband metrics and job productivity: a look at county-level data. *The Annals of Regional Science* 66(1):161-84.
- Garboden PM (2020) Sources and types of big data for macroeconomic forecasting. In: Fuleky P. (eds) *Macroeconomic Forecasting in the Era of Big Data*. *Advanced Studies in Theoretical and Applied Econometrics*, vol 52. Springer, Cham. [https://doi.org/10.1007/978-3-030-31150-6\\_1](https://doi.org/10.1007/978-3-030-31150-6_1)
- Gruber H, Hätönen J, Koutroumpis P (2014) Broadband access in the EU: An assessment of future economic benefits. *Telecommunications Policy* 38(11):1046-58.
- Hassani H, Silva ES (2015) Forecasting with big data: A review. *Annals of Data Science* 2(1):5-19.
- Holt L, Jamison M (2009) Broadband and contributions to economic growth: Lessons from the US experience. *Telecommunications Policy* 33(10-11):575-81.
- Jain A (2019) Macro forecasting using alternative data. In *Handbook of US Consumer Economics* Jan 1 (pp. 273-327). Academic Press. <https://doi.org/10.1016/B978-0-12-813524-2.00011-1>

- Kaiser HF (1974) An index of factorial simplicity. *Psychometrika* 39(1):31-6.
- Kearney MS, Levine PB (2015) Media influences on social outcomes: The impact of MTV's 16 and pregnant on teen childbearing. *American Economic Review* 105(12):3597-632.
- Kolanovic M, Krishnamachari RT (2017) Big data and AI strategies: Machine learning and alternative data approach to investing. JP Morgan Global Quantitative & Derivatives Strategy Report.
- Kuc-Czarnecka M, Piano SL, Saltelli A (2020) Quantitative storytelling in the making of a composite indicator. *Social Indicators Research* 23:1-28. <https://doi.org/10.1007/s11205-020-02276-0>
- Li J, Forzati M (2018) The Social-Economic Impact of Fiber Broadband: A Hype or a Reality?. In 2018 20th International Conference on Transparent Optical Networks (ICTON) (pp. 1-7). IEEE.
- Libório MP, da Silva Martinuci O, Machado AM, Machado-Coelho TM, Laudares S, Bernardes P (2020) Principal component analysis applied to multidimensional social indicators longitudinal studies: limitations and possibilities. *GeoJournal* 21:1-6. <https://doi.org/10.1007/s10708-020-10322-0>
- Libório MP, Ekel PY, Lyrio RD, Bernardes P, Soares GL, Machado-Coelho TM (2020) Structural Equation Modeling Applied to Internet Consumption Forecast in Brazil. *IEEE Access* 8:161816-24.
- Libório MP, Machado-Coelho TM, Bernardes P, Machado AM, Ekel PY, Soares GL (2018) Forecasting Internet demand using public data: a case study in Brazil. *IEEE Access* 6:65974-80.
- Libório MP, Martinuci OD, Machado AM, Lyrio RD, Bernardes P (2021) Time–Space Analysis of Multidimensional Phenomena: A Composite Indicator of Social Exclusion Through k-Means. *Social Indicators Research* 25:1-23. <https://doi.org/10.1007/s11205-021-02763-y>
- Liu Y, Wang Z, Rader B, Li B, Wu CH, Whittington JD, Zheng P, Stenseth NC, Bjornstad ON, Brownstein JS, Tian H (2021) Associations between changes in population mobility in response to the COVID-19 pandemic and socioeconomic factors at the city level in China and country level worldwide: a retrospective, observational study. *The Lancet Digital Health* 3(6):e349-59.
- Ma YR, Ji Q, Pan J (2019) Oil financialization and volatility forecast: Evidence from multidimensional predictors. *Journal of Forecasting* 38(6):564-81.
- MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* 1967 Jun 21 (Vol. 1, No. 14, pp. 281-297).
- Mazziotta M, Pareto A (2016) On a generalized non-compensatory composite index for measuring socio-economic phenomena. *Social indicators research* 127(3):983-1003.

- Meo SA, Al Masri AA, Usmani AM, Memon AN, Zaidi SZ (2013) Impact of GDP, spending on R&D, number of universities and scientific journals on research publications among Asian countries. *PloS one* 20;8(6):e66449.
- Milani F (2021) COVID-19 outbreak, social response, and early economic effects: a global VAR analysis of cross-country interdependencies. *Journal of population economics* 34(1):223-52.
- Moirangthem NS, Nag B (2022) Developing a competitiveness index at sub-national level for India: an economic growth measure. *International Journal of Emerging Markets*. <https://doi.org/10.1108/IJOEM-02-2021-0209>
- Nardo M, Saisana M, Saltelli A, Tarantola S (2005) Tools for composite indicators building. *European Commission, Ispra*. 15(1):19-20.
- Narita MF, Yin R (2018) In search of information: use of google trends' data to narrow information gaps for low-income developing countries. *International Monetary Fund*.
- Neto AR, Candido O (2021) What does Google say about credit developments in Brazil?. *Studies in Nonlinear Dynamics & Econometrics*. <https://doi.org/10.1515/snde-2019-0122>
- Papageorge NW, Zahn MV, Belot M, Van den Broek-Altenburg E, Choi S, Jamison JC, Tripodi E (2021) Socio-demographic factors associated with self-protecting behavior during the Covid-19 pandemic. *Journal of Population Economics* 34(2):691-738.
- Park G, Schwartz HA, Sap M, Kern ML, Weingarten E, Eichstaedt JC, Berger J, Stillwell DJ, Kosinski M, Ungar LH, Seligman ME (2017) Living in the past, present, and future: Measuring temporal orientation with language. *Journal of personality* 85(2):270-80.
- Pastor JM, Peraita C, Serrano L, Soler Á (2018) Higher education institutions, economic growth and GDP per capita in European Union countries. *European Planning Studies* 26(8):1616-37.
- Pearson K. LIII (1901) On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science* 2(11):559-72.
- Perlin MS, Caldeira JF, Santos AA, Pontuschka M (2017) Can we predict the financial markets based on Google's search queries?. *Journal of Forecasting* 36(4):454-67.
- Piccoli P (2019) Speculative attacks and investor attention. Available at SSRN 3503654. <https://doi.org/10.1111/twec.13207>.
- Piccoli P, de Castro J (2021) Attention-return relation in the gold market and market states. *Resources Policy* 74:102333.

- Preis T, Moat HS, Stanley HE, Bishop SR (2012) Quantifying the advantage of looking forward. *Scientific reports* 2(1):1-2.
- Qadan M, Nama H (2018) Investor sentiment and the price of oil. *Energy Economics* 69:42-58.
- Quirk TJ (2012) One-way analysis of variance (ANOVA). In: *Excel 2007 for Educational and Psychological Statistics*. Springer, New York, NY. [https://doi.org/10.1007/978-1-4614-3725-3\\_8](https://doi.org/10.1007/978-1-4614-3725-3_8)
- Salvati L, Carlucci M (2014) A composite index of sustainable development at the local scale: Italy as a case study. *Ecological Indicators* 43:162-71.
- Spelta A, Pagnottoni P (2021) Mobility-based real-time economic monitoring amid the COVID-19 pandemic. *Scientific Reports* 11(1):1-5.
- Tümer AE, Akkuş A (2018) Forecasting gross domestic product per capita using artificial neural networks with non-economical parameters. *Physica A: Statistical Mechanics and its Applications* 512:468-73.
- Varian HR (2014) Big data: New tricks for econometrics. *Journal of Economic Perspectives* 28(2):3-28.
- Vlastakis N, Markellos RN (2012) Information demand and stock market volatility. *Journal of Banking & Finance* 36(6):1808-21.
- Vosen S, Schmidt T (2011) Forecasting private consumption: survey-based indicators vs. Google trends. *Journal of forecasting* 30(6):565-78.
- Yu L, Zhao Y, Tang L, Yang Z (2019) Online big data-driven oil consumption forecasting with Google trends. *International Journal of Forecasting* 35(1):213-23.  
<https://doi.org/10.1016/j.ijforecast.2017.11.005>

## Figures

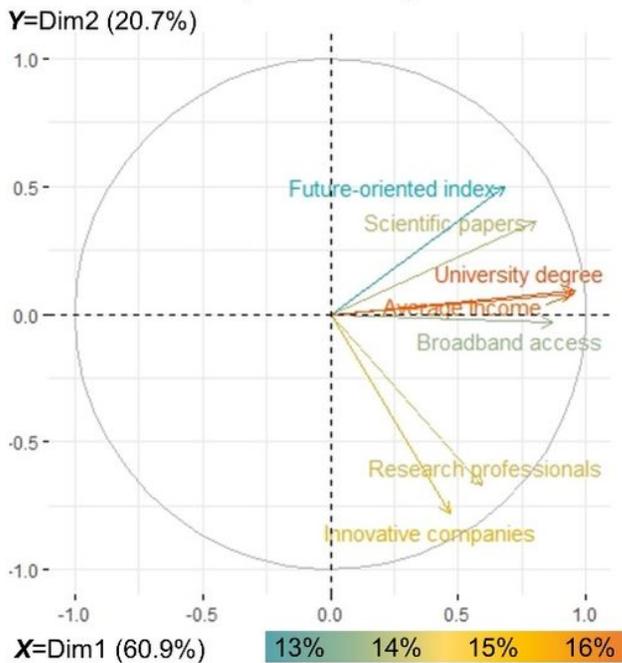


**Figure 1**

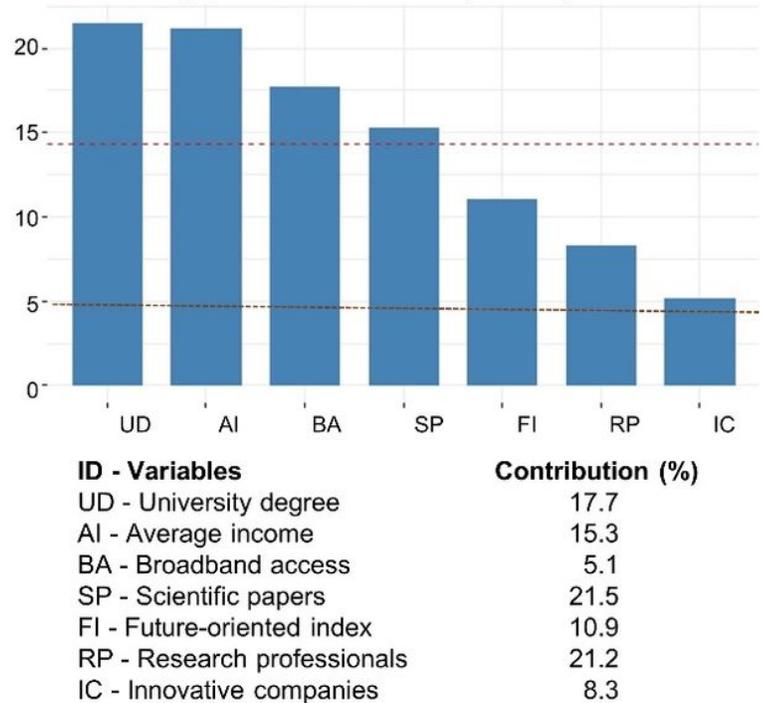
Correlation between the future-oriented index and average income in Brazilian states.

**Note:** the average household income was due to the absence of gross domestic product per capita data for Brazilian states.

**Contribution of sub-indicators in the Principal Component Analysis**



**Y=contribution (%) of variables to Principal Component**



**Figure 2**

Contribution of the sub-indicators in the Principal Component Analysis model and in the Development Potential Composite Indicator.

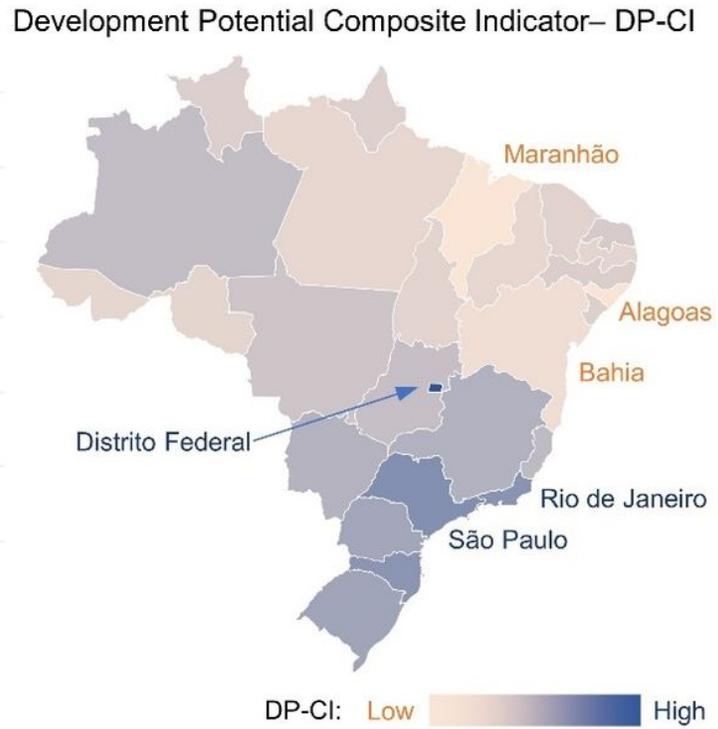
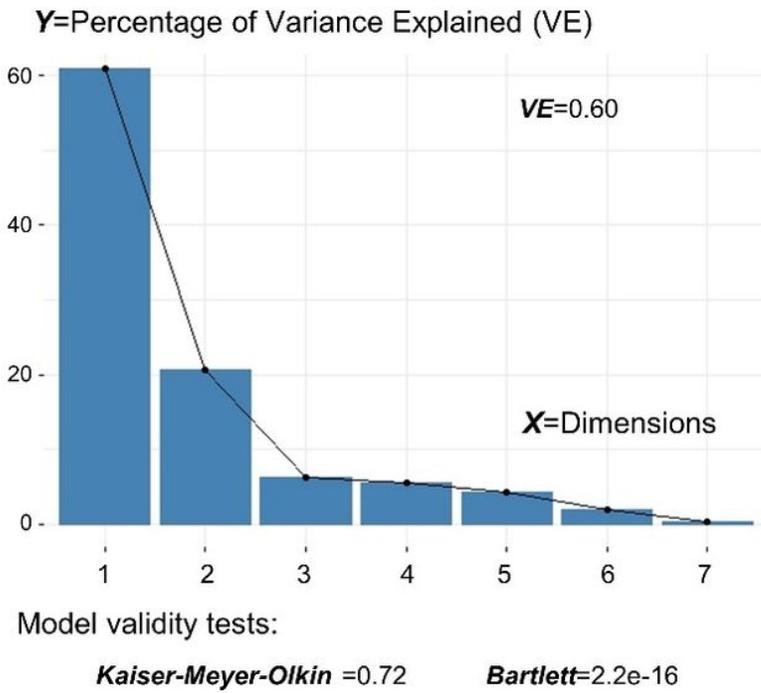


Figure 3

Acceptance parameters of the Principal Component Analysis and DP-CI model of Brazilian states.

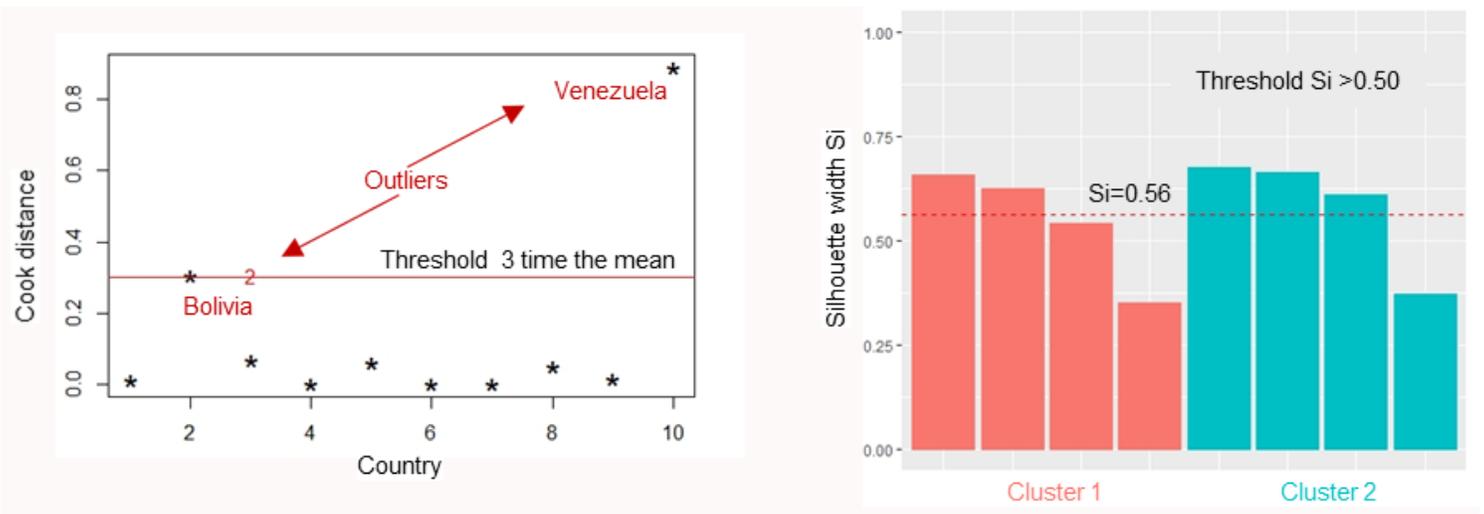


Figure 4

Exclusion of outliers and definition of groups of countries with similar behavior.

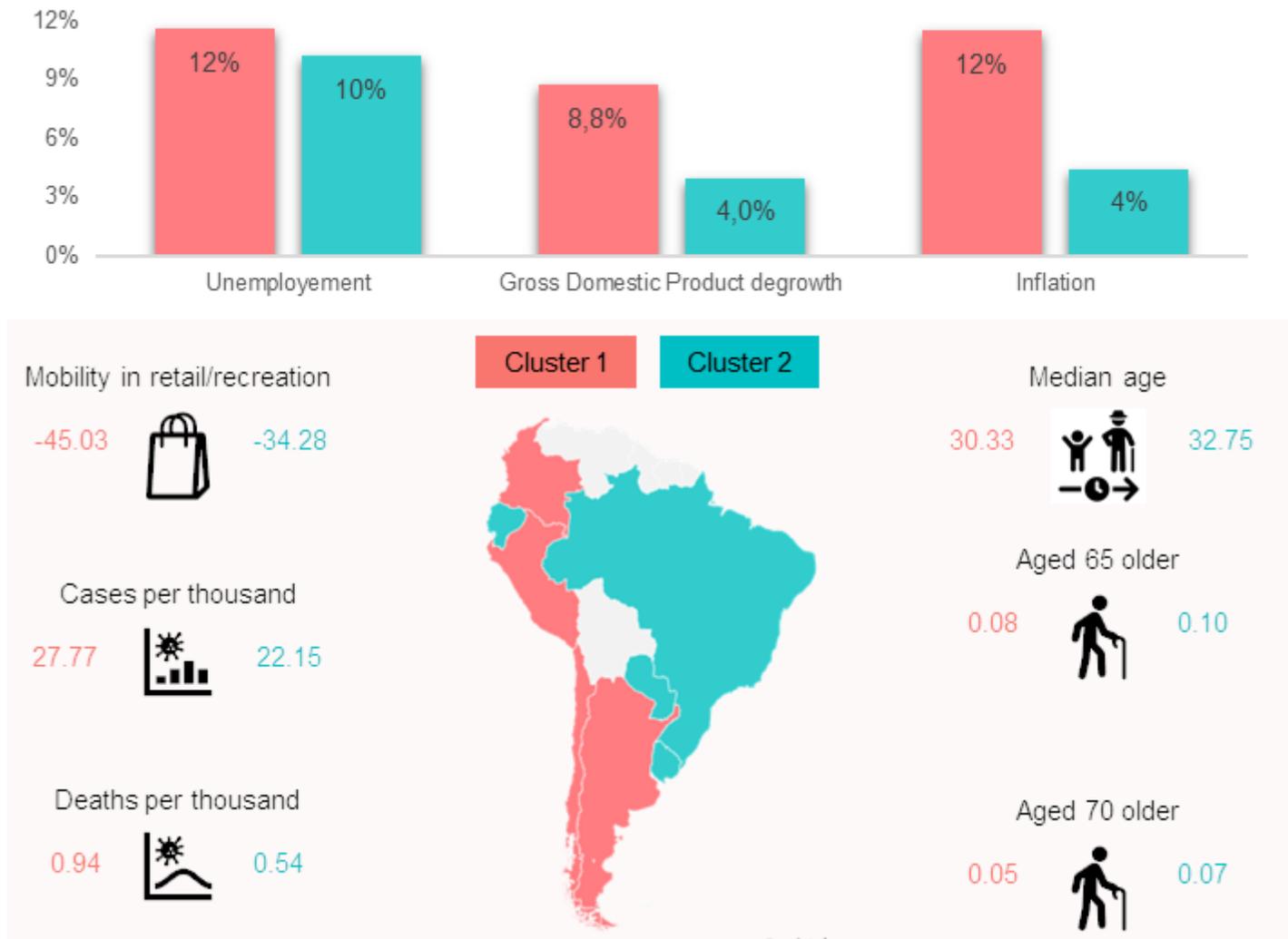


Figure 5

Impact of mobility in retail/recreation locations on the number of cases and deaths from COVID-19 and the relationship of this impact with economic indicators.