

CS-6630 - Data Visualization

Process Book

Brazil: Credit operations in public sector and the Human Development Index

<https://github.com/fredericosar/credit-operations>

By,

Frederico Rocha - Anil Kumar Konasale Krishna - Anil Kumar Ravindra Mallapur
fred@cs.utah.edu - u0810018 - anil.konasale@utah.edu - u0939372 -
anilmr@cs.utah.edu - u0942509

Background and Motivation :-

Brazil has entered in a fiscal crisis, not as intense as the one Greek is going through, but equally worrying. In his latest research, Leno Rocha, an analyst for the Brazilian National Treasury, showed how the number of loan pleas made by federative entities evolves during a fiscal year. He was able to prove that the diffusion phenomenon is determinant on the dynamics of the pleas. Furthermore, his research demonstrated how the number of credit operations fits in the Gompertz curve. The theory shows excellent results when compared to the practical data from 2002 to 2014. His work demonstrates that large financial volume of credit is concentrated in a small number of operations, enhancing the predictability and the performance capacity of debt managers in the Brazilian National Treasury.

We believe this work is especially interesting because it allows us to make predictions on Brazilian financing pleas. It also allows researchers to understand the important consequences of the federal debt, such as the moment of the credit impact in the social and Brazilian domestic macroeconomic scenario. Moreover, it is possible to categorize the finality of the credit operations (e.g. Health, Infrastructure, *etc*) and use the Human Development Index to analyse if the HDI determines who is more likely to take the loans.

Thus, the reason for us to implement a visualization for this research and data, is to analyse the credit operation environment that surrounds Brazil's development and show how it might be possible to connect it to an humanized development index, and provide a better understanding of how the credit operations affect quality of people's life. We can also use visualization to show the predictions of these pleas using the Gompertz curve.

Finally, the visualization can be used in the management of human resources. Understanding the demand of pleas creates the possibility to foresee the amount of work that will be necessary and efficiently allocate the workers that analyze the pleas for subnational loans. This efficient allocation improves the quality of spending on personnel expenses in the administration of Brazil's public finances.

Related Work :-

When we were exploring for some ideas for designing a visually appealing design, we came across the following visualizations like <http://teamdatahub.github.io/> and <http://laurenwood.github.io/> . In both of these visualizations, Map has been used to convey useful information to users. As our project is for analyzing the credit operations of public sector for Brazil, we gained useful insights on how map can be used to convey the credit operations information to user.

Project Objectives :-

In this project, we are trying to answer the following questions that stem from the data showing credit requests made by different public sector organizations in Brazil.

1. Demonstrate that the number of credit operations made by brazilian federative entities fits the Gompertz curve. According to Leno Rona's studies, the number of credit please fits the Gompertz curve. Thus we would like to plot the graph and see if it fits the Gompertz curve.
2. The data that we have collected for this project provides information about who is sanctioning the loan, this data is crucial in analyzing which financial organizations across the world are investing in Brazil and how much money each organization is investing and how frequently they are investing. This data might be helpful if someone is trying to put a new plea, it would be helpful in understanding which organization is more likely to sanction the credit request and for what purpose an organization is more likely to sanction the loan.
3. How the loan money being sanctioned is distributed for different kind of purpose ? It's always true that most of credit requests are made for operations or infrastructure, so we would like to crunch the data and visually show that most of the loan being sanctioned is a part of one or two major purpose.

4. Next important question to answer is why a city or a state is requesting a credit ? We would like to categorize the credit finality and understand how the money is sanctioned on these finalities.
5. Compare the HDI evolution during the years of 2002 to 2014 and analyse if the number of credit operations (loans) has any role in the index changes.
6. Try to describe facts about the requested loans. A main reason to answer this question is to understand if there is any strong reason the federative entities are asking for money at a particular moment?
7. Show how the HDI changed during the 12 years that are being analyzed. We would like to answer if the credit is anywhere related to HDI changes.

Data :-

The credit data comes from the the Brazilian Treasury Secretariat¹. It provides the information about the credit operations in the Brazilian public sector from 2002 to 2014. The data is offered in a CSV (comma separated value) file.

The data contains the Human Development Index (HDI) for brazilian cities and states. It comes from the United Nations² and from Google Public³. The data in both sites are available in a table format.

The brazilian latitudes, longitudes and areas were located in multiple places. We decided to use the ones offered by Globo⁴ and the official Brazilian Data website⁵.

1. <http://sadipem.tesouro.gov.br>
2. <http://www.pnud.org.br/atlas/ranking/Ranking-IDHM-Municipios-2000.aspx>
3. <http://www.google.com/publicdata>
4. <http://app.globoesporte.globo.com/futebol/mapa-das-torcidas-no-facebook/data/fbz.json>
5. <http://dados.gov.br/>

Data Example

The credit data contains 14 fields. However, only 7 fields are useful for this work. They are presented with sample data below.

City	State	Credit Finality	Type of Creditor	Creditor Name	Amount in US Dollars	Date
Goiânia	Goiás	Infrastructure	Public National	Caixa	\$80,922.21	09-10-2013
Rio de Janeiro	Rio de Janeiro	Health	Public National	Banco do Brasil	\$3,140,016.00	06-07-2011

The HDI data contains 6 fields. Again, we are only interested in the fields shown below.

City	Human Development Index
Goiânia	0,715
Rio de Janeiro	0,716

8. Data Processing

The credit operations data is provided in Portuguese and will be translated to English. As it already has well-defined fields, it will not be necessary to do a lot of cleanup, just the exclusion of some fields. However, the data file needs to be parsed to JSON to simplify the manipulation using D3. For that, we will use free online converting tools such as CSV to JSON web-site⁶. Using this credit data, we generate the Gompertz curve, we categorize the type of creditor in three categories (Public National, Private National, International) and the credit finality in four categories (Infrastructure, Health, Education, and Safety). Also, as the amount of the credit loan is based on the currency of the country in which the creditor is located, we will convert the amount to U.S. dollars.

The data for HDI is very straightforward and won't need any cleanup nor derivation. This is also valid for the geographical data.

Lastly, the data processing will be done on-the-fly using Javascript on the loading of the page. As all the data will be well-defined and offered as a JSON this will enable fast update on the page.

6. <http://www.csvjson.com/csv2json>

Exploratory Data Analysis:

We still haven't performed any exploratory data analysis on this data, we have been consulting the research work done by Leno Rocha. In the coming week, we would like to use Tableau's visualization software to see and find some important patterns in the data.

Design Evolution:

We are currently implementing the design that we proposed, our current design places brush feature to select a time range at the top of the page, but after using it for ourselves, we decided putting the brush feature below the map helps in choosing the time range and observe the map changing at the same time without much of an effort. We are still working on our placements, our current implementation shows the brush being implemented at the top.

Implementation:

Our current implementation looks like this. To clean up the data, we converted data from CSV to JSON format. We started with our implementation which is shown in the below picture. Main parts of visualizations that we have implemented is interactive Brazilian Map showing the total credit requests from a state for a particular range of time selected in the top from the brush scale. Once a user clicks on a particular state, the Pie Charts beside describes which financial institutions have contributed how much of credit for that state in that time range, and second pie chart shows the distribution of credit between different finality for a particular state.

