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 Ravindra Mallapur
fred@cs.utah.edu - u0810018 - anil.konasale@utah.edu - u0939372
anilmr@cs.utah.edu - u0942509

Table of contents

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 Secretariat[1]. 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 Nations'[2] and from Google Public[3]. 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 Globo[4] and the official Brazilian Data website[5].

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.

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

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

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.

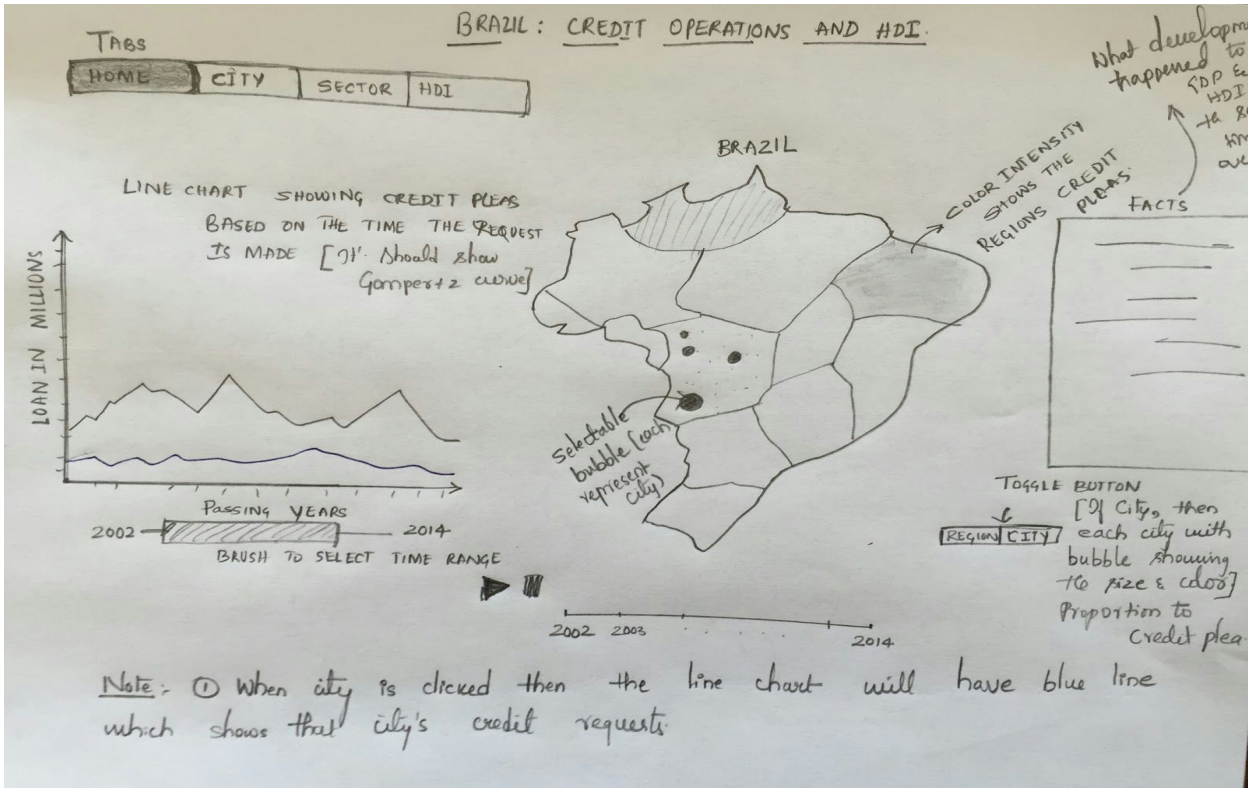
The data for HDI is very straightforward and won't need any cleanup nor derivation. This is also valid for the geographical data. Lastly, the data processing will be done on-the-fly using Javascript on the loading of the page. As all the data will be well-defined and offered as a JSON this will enable fast update on the page.

Exploratory Data Analysis

We have been consulting the research work done by Leno Rocha. Tableau's visualization on the number of credit operations in Brazil over the course of last decade indeed falls into Gompertz curve. The lender and credit requester data on node link diagram visualization shows that there are couple of banks which receives majority of the credit requests and the rest of the banks receive less than 20% of the credit requests.

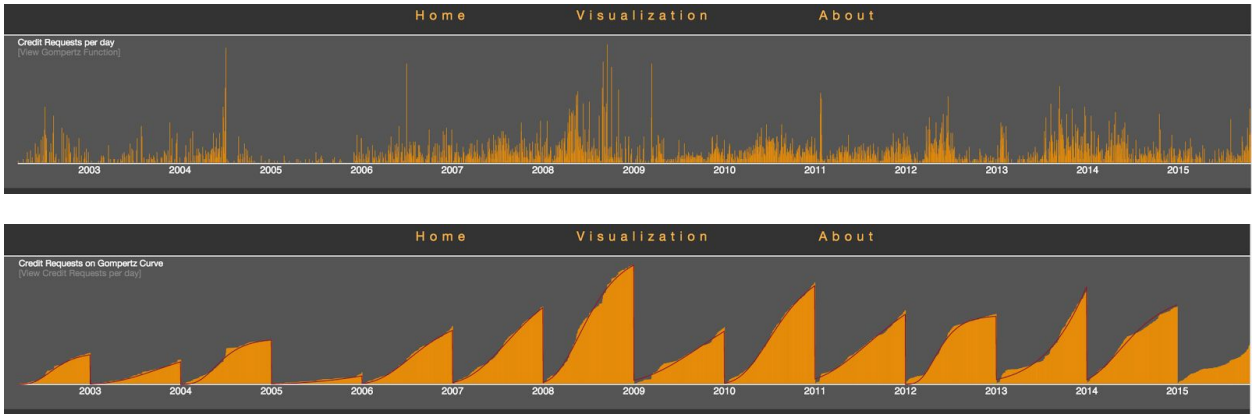
Design Evolution

Initially our design that we came up are shown below. Our main dataset is of credit operations performed in Brazil from year 2002 to 2014. Our initial design had a feature of including Brazil map to show geographical variation of credit operations using a color scale for visually appealing visualization that would help users to analyze the data. To select the timeline as our data is divided based on days granularity, we had time scale beneath map and line chart beside map. During the initial implementation of map visualization data we realized our visualization needed a different approach to blend our map and time selection data to fit perfectly in the available space of homepage. So we decided that we will have our time selection feature using brush in the top of the page and map towards the right side of page along with a information box to display the selected state and time line and a donut charts with bar graph showing human development index in the left part of page.



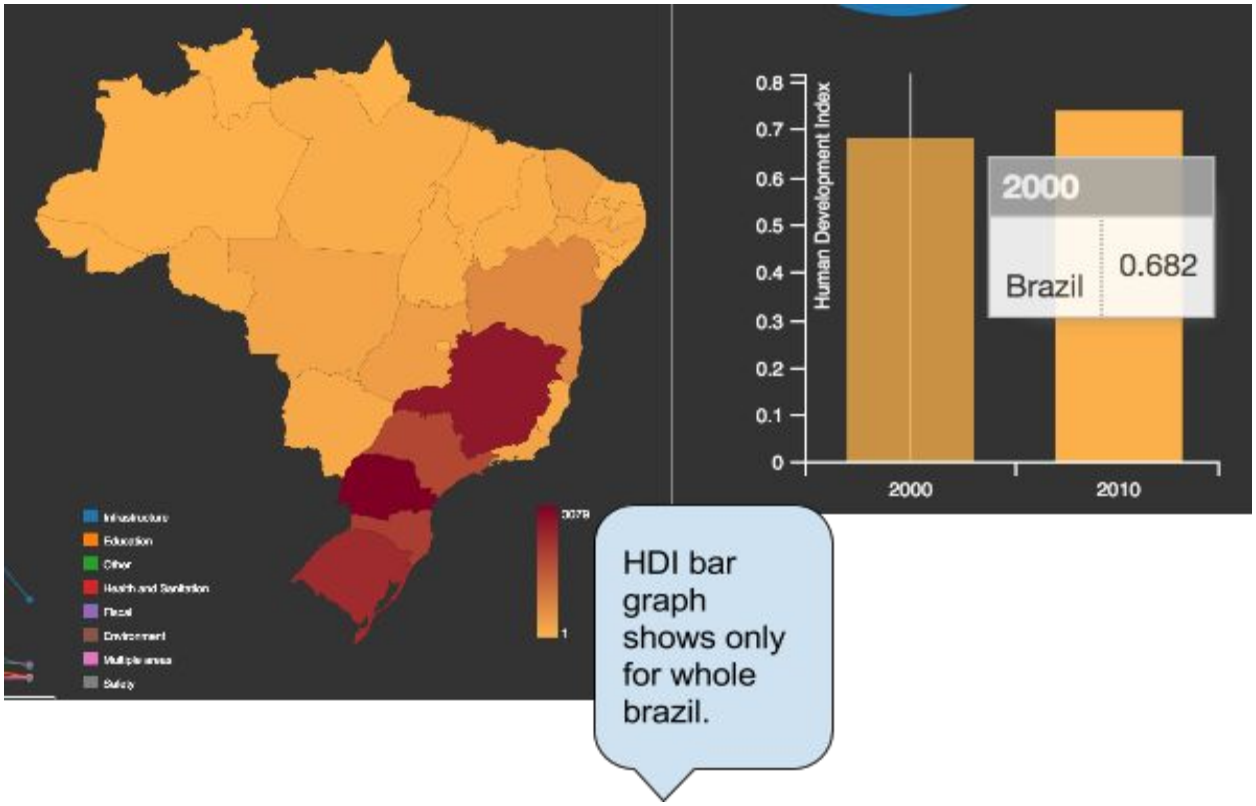
Our proposed design was as shown above. In this design, we wanted to show the geographical distribution of the credit operations in brazil, with that thought in our mind, we decided that Brazil Map with color intensity showing different intensity is the right choice, with option for user to choose a particular state and see the credit operations of that state instead of whole Brazil. As our data spans from 2002-2014, we needed a mechanism to let user visualize data for a particular year. Our first approach was a drop down button to let user select a particular year to visualize credit operations data for that year. But this function has a limitation that we won't be able to do a cumulative analysis of multiple years or from a particular time range. So we changed our design to accommodate selection of time range by implementing a brush feature. Instead of just letting a user select time range by using brush, we figured out that we can actually show how the credit operations varied across the whole dataset and embed Gompertz curve on that bar chart with time line selection. This design is visually appealing and accommodates more data in a shorter space.

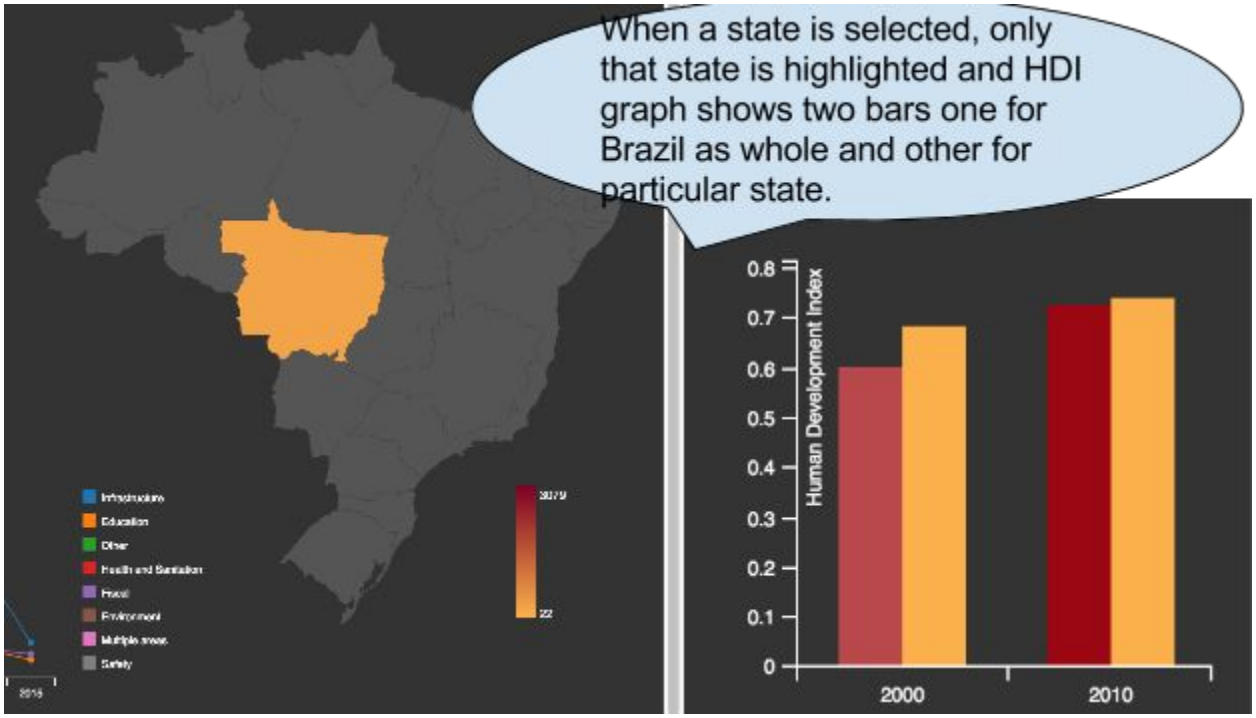
The below picture shows our brush displaying number of credit operations in a day for the available time range from 2002-2014. This component has a toggle feature which plots the gompertz curve of the data and successfully shows the claim that we made in our initial proposal that the Brazilian finance credit operations fits Gompertz curve.



Our Map initially didn't have an option to select a particular state, but in the final design we thought of an alternative design when we select a state, only that state displays in base color and rest of the map in background color with little border to differentiate a map. This helps us

in highlighting a particular state and our Human Development index bar graph shown in the left part of the page will show up two bars for two years, one for whole Brazil and other for the state that is being selected.



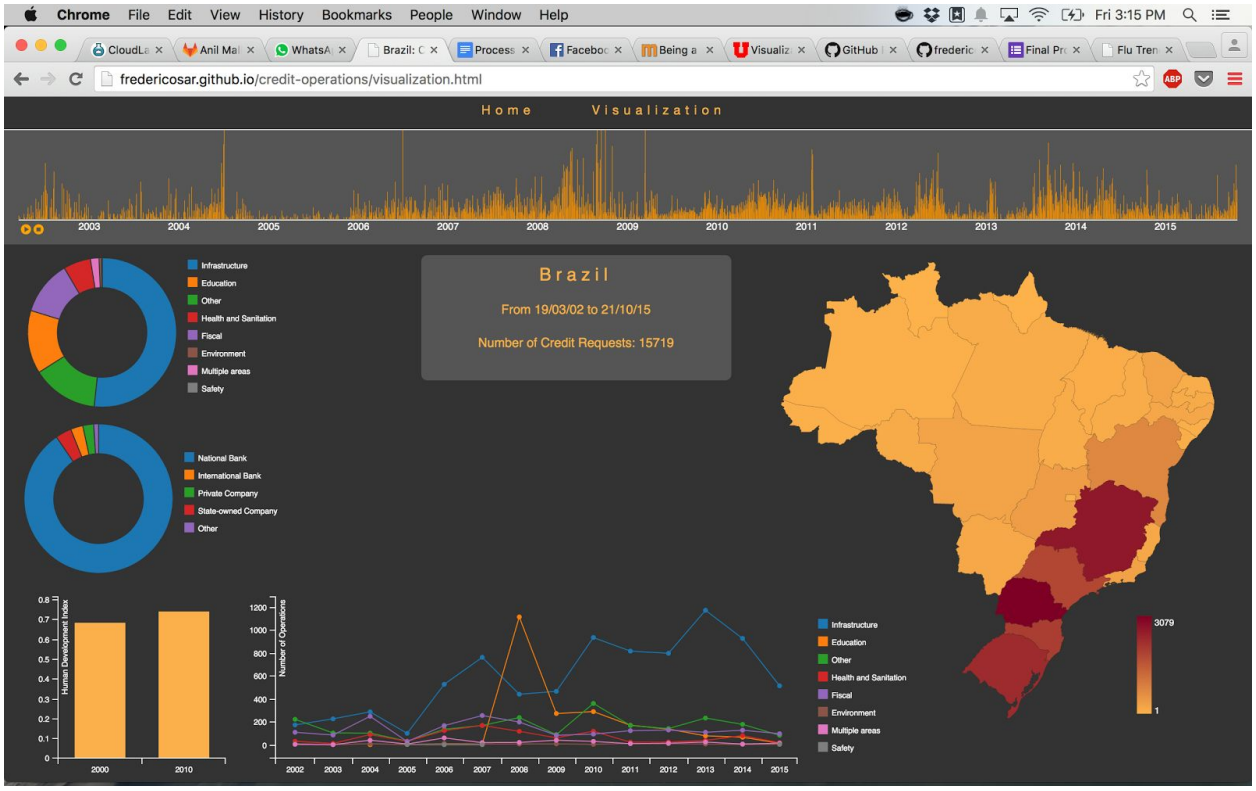


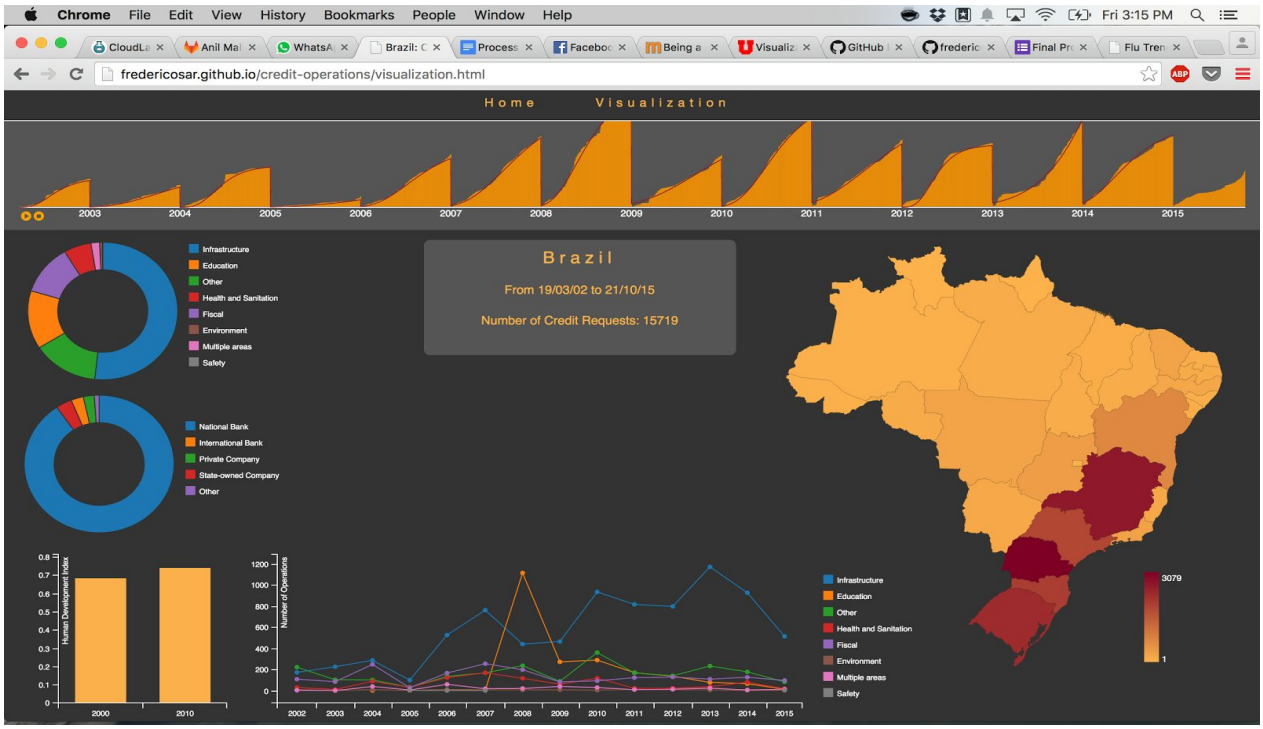
Finally, we decided to include an autoplay feature in our visualization to automatically visualize the data from start of initial date to the end date to show the user how the number of credit operations varied. Here an user can select the autoplay for all of Brazil or a particular state of in Brazil. The data keeps populating and all visualizations reflect the changes in data from start to time range. There is an option to pause the autoplay so that initial visualization will appear.

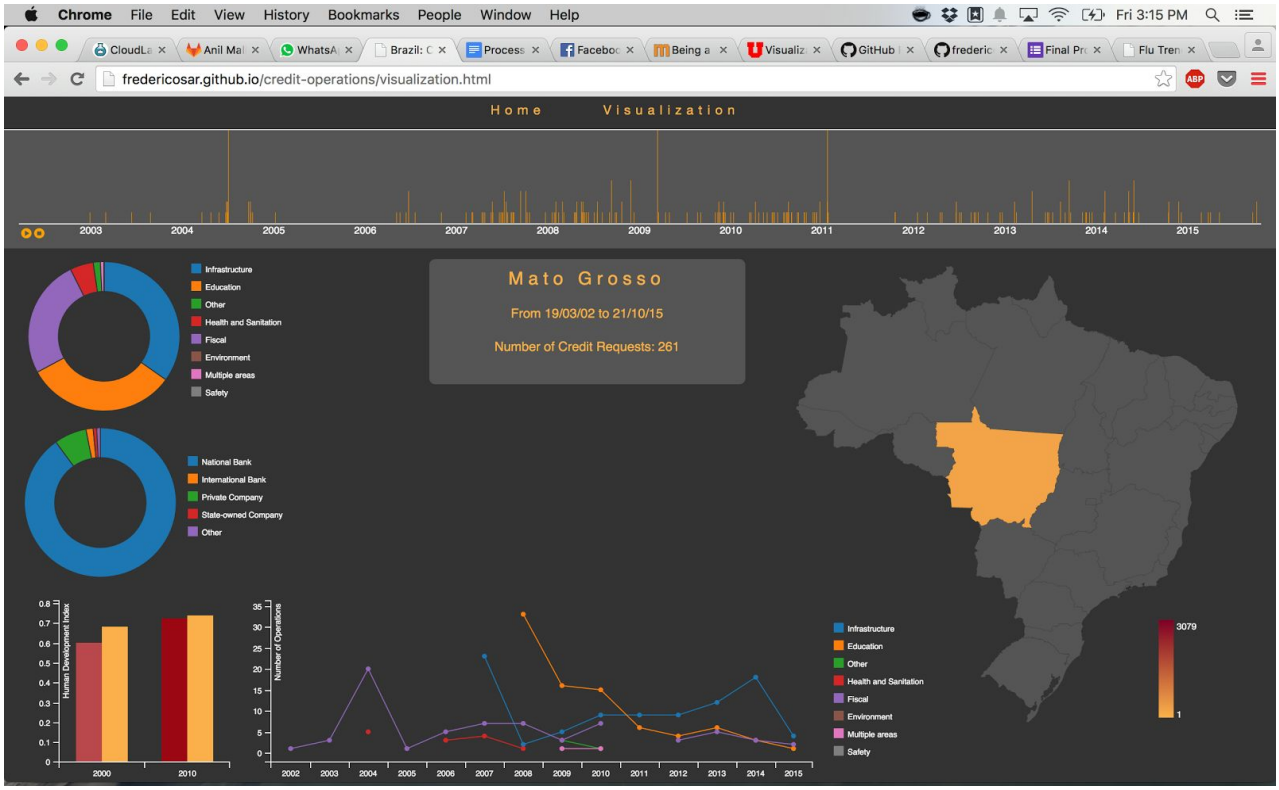
Implementation

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. The bar graph beneath the donut chart shows the human development index for a whole of brazil

and that particular state for two time intervals. A line chart shows for a selected time interval in brush, number of credit operations that are being sanctioned for different sectors. Our final implementation is as shown below. We have used d3.js and c3.js library which helps in visualization development of donut charts. The below figures shows our final visualization on hosted page.



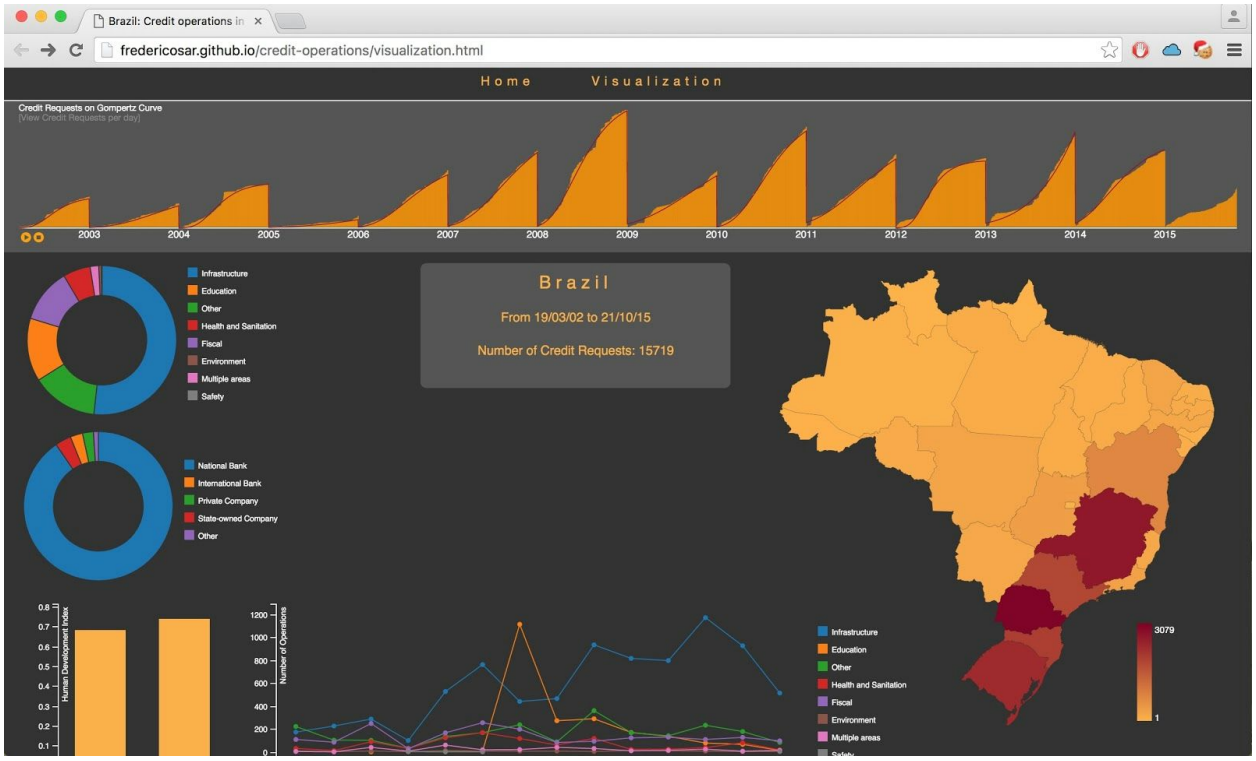

Chrome File Edit View History Bookmarks People Window Help
fredericosar.github.io/credit-operations/visualization.html
Home
Visualization
2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015
Infrastructure
Education
Other
Health and Sanitation
Fiscal
Environment
Multiple areas
Safety
National Bank
International Bank
Private Company
State-owned Company
Other
Brazil
From 19/03/02 to 21/10/15
Number of Credit Requests: 15719
Human Development Index
2000
2010
Number of Operations
2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015
3079
1


Chrome File Edit View History Bookmarks People Window Help
fredericosar.github.io/credit-operations/visualization.html
Home
Visualization
2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015
Infrastructure
Education
Other
Health and Sanitation
Fiscal
Environment
Multiple areas
Safety
National Bank
International Bank
Private Company
State-owned Company
Other
Mato Grosso
From 19/03/02 to 21/10/15
Number of Credit Requests: 261
Human Development Index
2000
2010
Number of Operations
2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015
3079
1

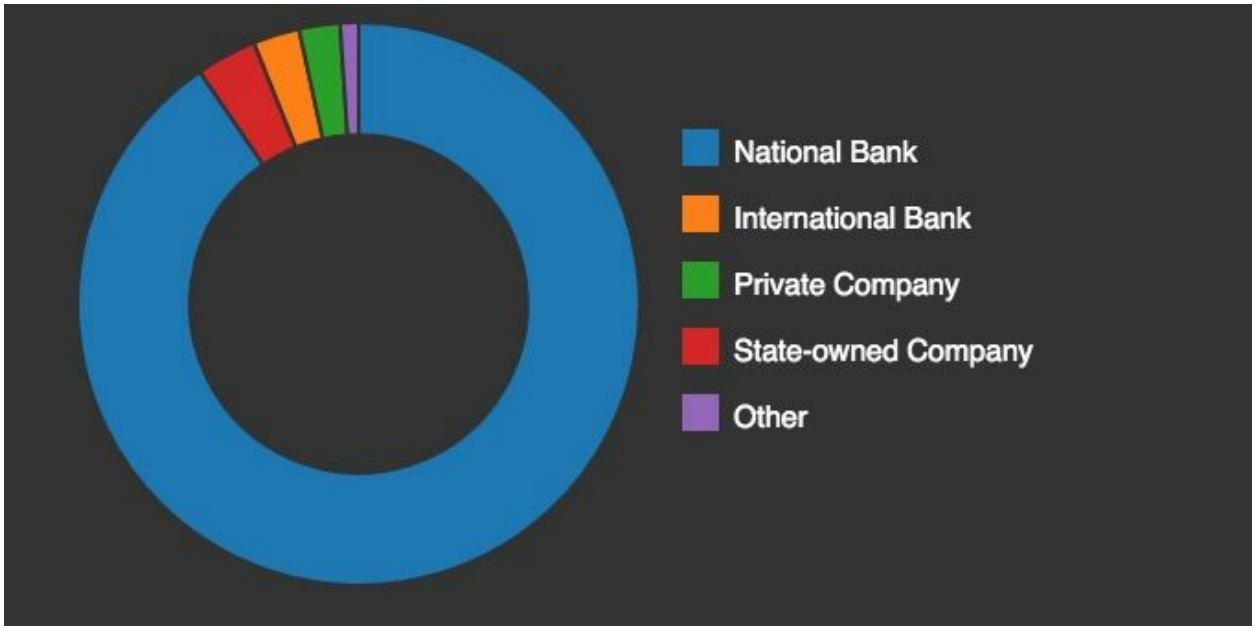
Evaluation

Our visualization helps answering the questions discussed in the section **Project Objectives.**

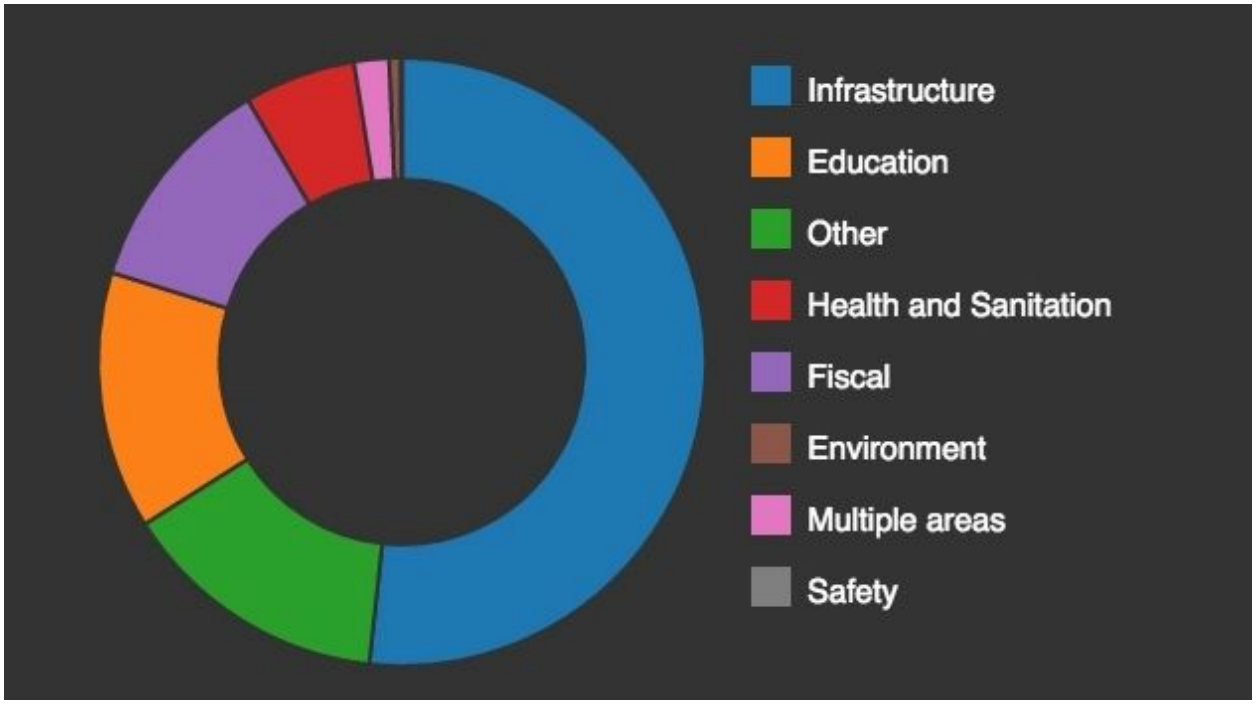
- **Gompertz curve** As shown in the below Figure, number of credit operations made by federative entities fits the Gompertz curve



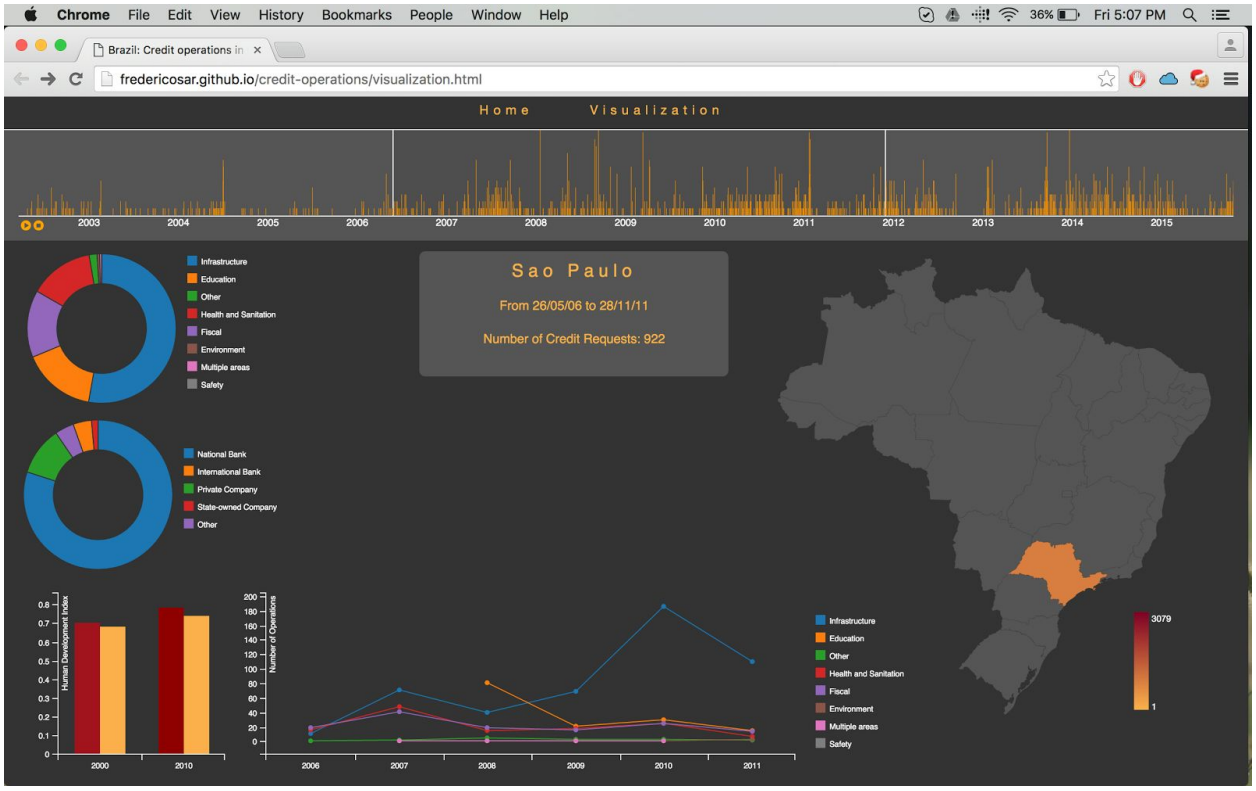
- **Categorize the type of creditors** As shown in the below snapshot, Creditors are categorized in the Donut Graph. This helps understand from where the loaned money is coming from. As it is evident from the Donut Graph that the vast majority of loaned money came from National Bank for the loans between 2003 to 2015.
    - Also users are provided with a feature to filter the creditors. This will help visualize 2nd, 3rd ... largest loan providers.
    - Creditors and their share during a particular period can be visualized by selecting corresponding time intervals



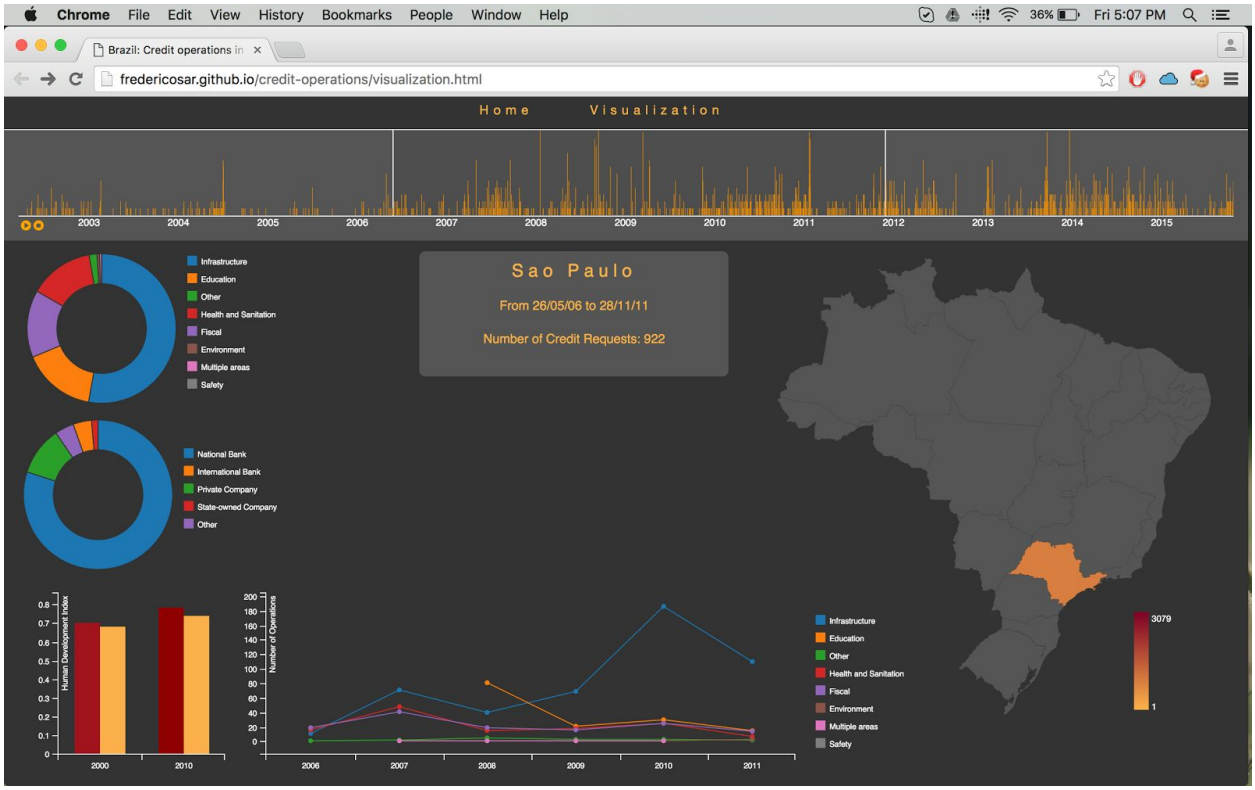
- **Loaned Money Distribution** As shown in the below Donut Graph, the most part of loaned money is concentrated in the few operations such as Infrastructure, Education and Fiscal.

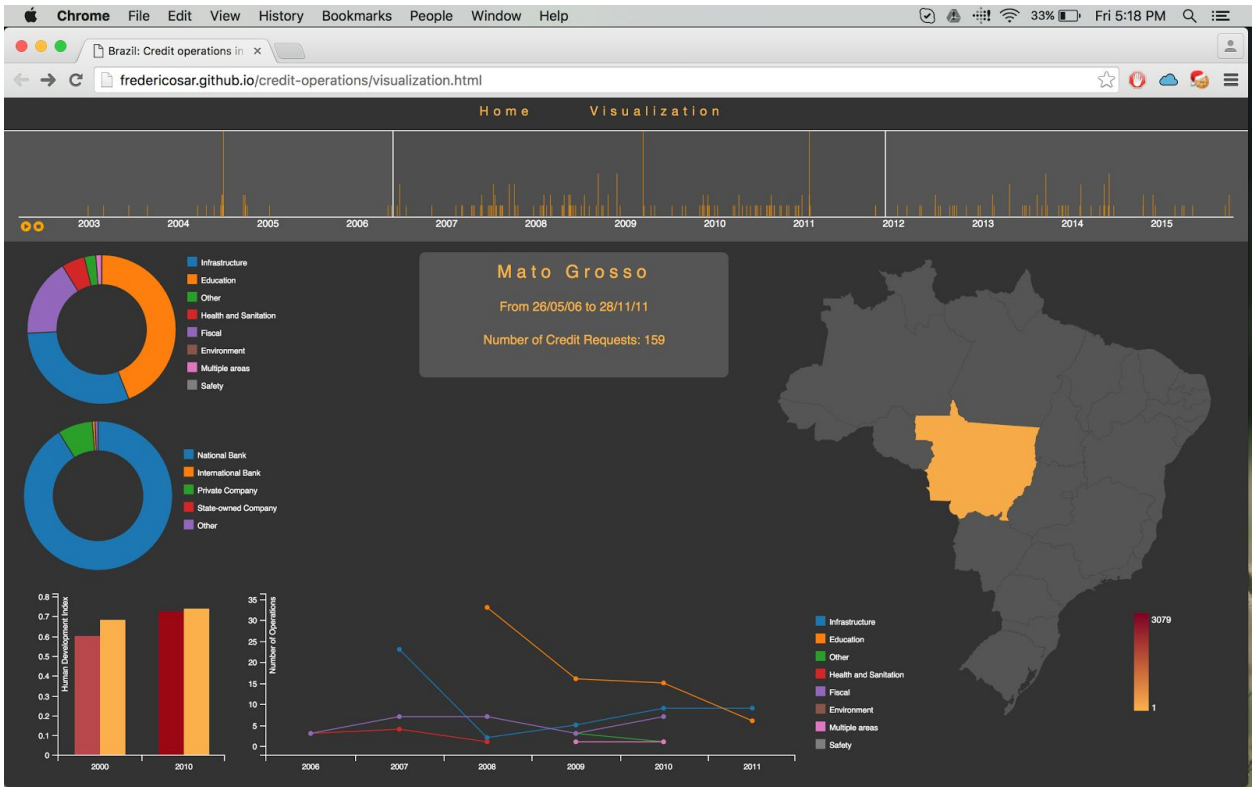


- **Purpose of Loan** The same Donut used to categorize federate operations also indicates the purpose of the loan. Following are some of the interesting patterns:
  - During all the times, purpose of the most of the applications is for Infrastructure
  - Between 2003 and mid of 2005, Fiscal and Health & Sanitation takes 2nd and 3rd highest number of loan requests. But number of loan application for Education increases gradually and it is the second highest one between 2005 and 2014.
- **HDI Evolution** It is interesting to know that states with the most loan applications have seen increased HDI index in both 2002 and 2010 when compared with the Brazil HDI.



- **Facts about requested loan** It is evident from the below Graphs that different federative entities have different purposes for loans. In the below example, applicants from Mato Grosso requests loans between 2006 and 2012 mostly for Education and Infrastructure whereas Sao Paulo applicants request loans for Infrastructure and Fiscal! So our visualization helps understand the purpose of loans of different federative entities.


Chrome File Edit View History Bookmarks People Window Help
36% Fri 5:07 PM
Brazil: Credit operations in
fredericosar.github.io/credit-operations/visualization.html
Home
Visualization
2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015
Infrastructure
Education
Other
Health and Sanitation
Fiscal
Environment
Multiple areas
Safety
National Bank
International Bank
Private Company
State-owned Company
Other
Sao Paulo
From 26/05/06 to 28/11/11
Number of Credit Requests: 922
Human Development Index
2000 2010
Number of Operations
2006 2007 2008 2009 2010 2011
3079
1


Chrome File Edit View History Bookmarks People Window Help
33% Fri 5:18 PM
Brazil: Credit operations in
fredericosar.github.io/credit-operations/visualization.html
Home
Visualization
2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015
Infrastructure
Education
Other
Health and Sanitation
Fiscal
Environment
Multiple areas
Safety
National Bank
International Bank
Private Company
State-owned Company
Other
Mato Grosso
From 26/05/06 to 28/11/11
Number of Credit Requests: 159
Human Development Index
2000 2010
Number of Operations
2006 2007 2008 2009 2010 2011
3079
1

- **Credit operations and Brazilian HDI** Yes, credit operations and number of applications have implication of the HDI of Brazilian HDI. As shown in the above snapshot, states with highest loan application have greater HDI than the Brazilian HDI in both the years : 2002 and 2010.