

---

# **A Strategy for Temporal Visual Analysis of Labor Accident Data**

---

**Luciana Lima Brito**



FEDERAL UNIVERSITY OF UBERLÂNDIA  
FACULTY OF COMPUTING  
POST-GRADUATION PROGRAM IN COMPUTER SCIENCE

Uberlândia  
2019



**Luciana Lima Brito**

**A Strategy for Temporal Visual Analysis of  
Labor Accident Data**

Master Thesis presented to the Faculty of  
Computing Post-Graduation Program of  
the Federal University of Uberlândia as part  
of the requirements for obtaining the ti-  
tle of Master of Science in Computer Science.

Concentration Area: Computer Science

Advisor: José Gustavo de Souza Paiva

Uberlândia

2019


**UNIVERSIDADE FEDERAL DE UBERLÂNDIA**

Coordenação do Programa de Pós-Graduação em Ciência da Computação  
 Av. João Naves de Ávila, nº 2121, Bloco 1A, Sala 243 - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902  
 Telefone: (34) 3239-4470 - www.ppgco.facom.ufu.br - cpgfacom@ufu.br


**ATA DE DEFESA - PÓS-GRADUAÇÃO**

Programa de Pós-Graduação em:	Ciência da Computação				
Defesa de:	Dissertação de Mestrado Acadêmico, 19/2019, PPGCO				
Data:	26 de novembro de 2019	Hora de início:	<b>14h50min</b>	Hora de encerramento:	<b>16h45min</b>
Matrícula do Discente:	11722CCP007				
Nome do Discente:	Luciana Lima Brito				
Título do Trabalho:	A Strategy for Temporal Visual Analysis of Labor Accident Data				
Área de concentração:	Ciência da Computação				
Linha de pesquisa:	Ciência de Dados				
Projeto de Pesquisa de vinculação:	-				

Reuniu-se na sala 1B132, Bloco 1B, Campus Santa Mônica, da Universidade Federal de Uberlândia, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Ciência da Computação, assim composta: Professores Doutores: Jefferson Rodrigo de Souza - FACOM/UFU, Bianchi Serique Meiguins - ICEN/UFPA e José Gustavo de Souza Paiva - FACOM/UFU, orientador da candidata.

Iniciando os trabalhos o presidente da mesa, Prof. Dr. José Gustavo de Souza Paiva, apresentou a Comissão Examinadora e a candidata, agradeceu a presença do público, e concedeu à Discente a palavra para a exposição do seu trabalho. A duração da apresentação da Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir a candidata. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando a candidata:

**Aprovada**

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Jefferson Rodrigo de Souza, Professor(a) do Magistério Superior**, em 27/11/2019, às 10:12, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Bianchi Serique Meiguins, Usuário Externo**, em 27/11/2019, às 15:39, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).





Documento assinado eletronicamente por **José Gustavo de Souza Paiva, Professor(a) do Magistério Superior**, em 27/11/2019, às 23:13, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [https://www.sei.ufu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **1710987** e o código CRC **73D10826**.

Referência: Processo nº 23117.101278/2019-18

SEI nº 1710987



Ficha Catalográfica Online do Sistema de Bibliotecas da UFU  
com dados informados pelo(a) próprio(a) autor(a).

B862 Brito, Luciana Lima, 1993-  
2019 A strategy for temporal visual analysis of labor accident data  
[recurso eletrônico] / Luciana Lima Brito. - 2019.

Orientador: José Gustavo de Souza Paiva.  
Dissertação (Mestrado) - Universidade Federal de Uberlândia,  
Pós-graduação em Ciência da Computação.  
Modo de acesso: Internet.  
Disponível em: <http://doi.org/10.14393/ufu.di.2019.2579>  
Inclui bibliografia.  
Inclui ilustrações.

1. Computação. I. Paiva, José Gustavo de Souza, 1979-, (Orient.).  
II. Universidade Federal de Uberlândia. Pós-graduação em Ciência  
da Computação. III. Título.

CDU: 681.3

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:  
Gizele Cristine Nunes do Couto - CRB6/2091  
Nelson Marcos Ferreira - CRB6/3074

*This work is dedicated to the Human Rights.*

---

## Acknowledgment

Agradeço a Deus pela vida e oportunidades a mim concedidas. Agradeço a meus pais Jonilda & Lucivaldo, que sempre lutaram para me dar a melhor educação possível. Agradeço a meu marido Mateus que sempre me apoia com tanta dedicação e que foi parte fundamental no desenvolvimento deste trabalho. Agradeço a minha gata persa Shizuka por sempre alegrar meus dias. Agradeço a minha irmã Luciene e meu cunhado Cyro por todo apoio e por terem colocado em minha vida um novo sentido, chamado João Brito, meu sobrinho que me inspira a querer ser melhor todos os dias. Agradeço a meus cunhados Camilo & Marília que apostaram em mim e me deram a oportunidade de iniciar este mestrado. Agradeço aos demais familiares que sempre estiveram em contato mesmo com a distância, pelo apoio de todas as formas.

Agradeço também aos amigos feitos em Uberlândia, Lucas, Juliana, Caio, Claudiney, Vinícius e Jean.

Agradeço ao meu orientador, Professor Dr. José Gustavo de Souza Paiva pela orientação, pelo trabalho realizado e oportunidade de trabalharmos juntos.

Por fim, agradeço à agência de fomento CAPES pela bolsa concedida.



*“Why had the problem not been previously identified, or, if it had been identified, why had it not been addressed?”*  
*(U.S Department of Labor)*





---

# Resumo

Acidentes de trabalho (ATs) são um problema sério para a sociedade que podem resultar em danos físicos e/ou psicológicos para os empregados, perda de mão de obra, despesas com indenizações e multas para o empregador e gastos de seguro social com indenizações e hospitalizações para o Estado. O Ministério Público do Trabalho(MPT) coleta dados sobre ATs e esses dados apresentam potencial para análise. Neste trabalho apresentamos uma estratégia visual empregando técnicas de Visualização da Informação para exploração e análise de dados de ATs, focando no aspecto temporal, para identificar padrões nas ocorrências de ATs e associá-los a informações estratégicas sobre como as ocorrências se comportam no tempo. Nós desenvolvemos um sistema interativo para validar nossa estratégia de análise de dados de ATs do Brasil e realizamos as análises utilizando os dados providos publicamente pelo MPT, usando informações geográficas e temporais associadas ao dado, para demonstrar o potencial de nossa estratégia. Como resultado de nossa análise visual nós pudemos destacar a evolução das ocorrências de ATs nas localidades brasileiras ao longo do tempo, identificar tendências, eventos sazonais e comportamentos anormais. Pudemos também compara o comportamento de localidades em mesmo/diferente nível hierárquico, entre outros. A estratégia proposta pode fornecer um ambiente eficaz para orientar governantes na identificação de localidades que carecem de mais atenção para criar políticas públicas que melhorem a fiscalização, ajudem na redução dos acidentes e garantam a segurança dos empregados. As contribuições deste trabalho são uma estratégia visual para análise temporal de dados de AT, e um sistema computacional para melhorar as capacidades analíticas de especialistas do governo. Nós esperamos incentivar a transparência nos governos, bem como a participação pública nas decisões dos governos.

**Palavras-chave:** Acidente de Trabalho. Dados Governamentais. Visualização da Informação.



---

# Abstract

Labor accidents (LAs) are a serious social problem that can result in physical or/and psychological damages to employees, loss of manpower, expenses with compensations and fines to employers, and social security expenditures with compensations and hospitalizations to the State. The Brazilian Federal Labor Prosecution Office (BFLPO) collects a massive volume of data regarding LAs and this data presents potential to be analyzed. In this work we present a visual strategy employing Information Visualization techniques to explore and analyze LA data, focusing on the temporal aspects to identify patterns on labor accident occurrences, and to associate it to strategical information about how they behave over time. We developed an interactive system to validate our strategy in analyzing LA data from Brazil, and performed the analysis on the data publicly provided by the BFLPO, using the geographical and temporal information associated with the data, to demonstrate the potential of our strategy. As the result of our visual analysis we were able to highlight the evolution of accidents occurrences in the localities of Brazil over time, to identify trends, seasonal events, and abnormal behavior. We could also compare localities' behavior in the same/distinct hierarchical levels, among other task. Our proposed strategy may provide an effective environment to guide the governors in identifying localities lacking attention in order to create public policies to improve inspection, to reduce accidents and to grant safety for employees. The contributions of this work are a visual strategy for temporal analysis of LA data and a computational system, that implements this strategy, to enhance the analysis capabilities of experts from government. We expect to encourage transparency in governments as well as public participation in the governments' decisions.

**Keywords:** Labor Accident. Governmental Data. Information Visualization.



---

## List of Figures

Figure 1 – Pipeline of the visual analysis process using INFOVIS. . . . .	34
Figure 2 – Representation of the time series of the Lin-spiration. . . . .	35
Figure 3 – Layout proposed by Li et al. (2013) displays a tree-like diagram presenting five pollutant indexes and the air pollution index(API). The root node represents the province of Shandong, the next layer of nodes represents the cities being monitored, the third layer represents the monitor stations of air pollution and the last layer represents the pollutants measured in the station. Each bar chart represents one pollutant index measured during the year 2010 and the API. . . . .	36
Figure 4 – Layout proposed by Burch, Beck e Diehl (2008) presenting a tree-like diagram in which each node is a region of the world. A thumbnail structure and a bar chart is associated to each node. . . . .	37
Figure 5 – Disposal of the nodes of the graph over the timeline. . . . .	37
Figure 6 – (a) Interaction to zoom in or out on a specific period. (b) Interaction of transitioning among different time scales in specific periods.(c) Interaction to hide some specific period to allow comparison of non adjacent periods. . . . .	38
Figure 7 – Small multiples of Time Matrices depicting observations of ionospheric scintillation variables. Each Time Matrix represents the period 1-31 January, 2014, and each line of the Time Matrix represent the observations of one day. . . . .	39
Figure 8 – Context visualization in (SHARMIN et al., 2015) displaying the stress level of an individual according to her activities during the time span of one day. . . . .	40
Figure 9 – Context visualization about stock exchange prediction in (BADAM et al., 2016). . . . .	40

Figure 10 – Visualization of people flow departing from the station of Xujiahui during two hours of observation. The arches indicate the destination of the passengers, the colors indicate the volume of passengers and the sector chart beside indicates which subway lines were more used and the proportion of use among them. . . . .	42
Figure 11 – Visualization of German power plants data. . . . .	43
Figure 12 – Representation of the mean compensation payed by the loss of one arm in the states of the USA. . . . .	44
Figure 13 – Bubble map of Brazil provided by ODSST. The bubbles indicate the localization of a LA occurrence and the size of the bubbles indicates the amount of missing work days due to the the accident. . . . .	45
Figure 14 – Basic charts provided by ODSST. (a) The sector chart indicates the 10 more frequent injuries and the 10 economic sectors with more reports of LA between 2012 to 2018, in São Paulo state. (b) The bar chart above indicates the accumulated amount of missing work days by year due to injured employee removal. The chart below indicates the spent volume with expenses arising from removals. . . . .	46
Figure 15 – Dendrogram showing Brazil’s geographical territorial division, in which the nodes represent (1) region, (2) state, (3) mesoregion, (4) microregion, and (5) city, all aligned. . . . .	52
Figure 16 – Bar chart showing the daily amount of LACs in the period of one year. . . . .	52
Figure 17 – Three different localities’ Bar charts showing the distributions of LACs in a single year. The alignment of bar charts favors behavior comparison among them. . . . .	53
Figure 18 – Clockwise arrangement of the glyph’s edges considering the years from 2012 to 2017. The lines create a geometric shape. . . . .	53
Figure 19 – Comparison strategy using radial glyphs. Glyph 2 quickly catches the attention due to its non-homogeneous shape. Glyphs 1, 3 and 4 are more homogeneous. . . . .	54
Figure 20 – Dendrogram composition visualization: A dendrogram showing Brazil’s hierarchy based on the territorial division, the bar charts of LACs in a single year and glyphs with an overview of the amount of LACs in the six years. . . . .	54
Figure 21 – Choropleth map showing the proportion of LACs to the population size for each Brazil’s state in the year 2012. . . . .	55
Figure 22 – Small multiples showing the evolution of the percentage of out-of-wedlock births in France in a range of 14 years. . . . .	56

Figure 23 – A heat map table showing the period of one year. The months are presented as columns and the days as rows. Each cell present a different blue shade according to the number of LACs in the corresponding day.	57
Figure 24 – A heat map displayed as stripes, with days presented side-by-side and years, from 2012 to 2017, vertically arranged. The amount of LACs are presented in different blue shades, with low values associated to light shades.	57
Figure 25 – Choropleth composition visualization: A choropleth map presenting the states of Brazil, small multiples of choropleth maps presenting a specific a state of the country and two different representations of heat maps.	58
Figure 26 – Diagram of the Dendrogram Composition visualization. The structure in this strategy is shown in hard line boxes, comprised of multi-level perspective structure and comparison structures for evolution analysis in a single chosen year. The elements in the dotted line boxes shows the INFOVIS techniques chosen to represent each structure. The user first interacts with the multi-level structure(dendrogram), and through it she is able to interact with the attached layouts.	59
Figure 27 – Diagram of the Choropleth Composition visualization. The structure in this strategy is shown in hard line boxes, comprised of geographical context structure and temporal comparison structures for evolution analysis. The user first interacts with the geographical context structure then she is able to interact with the attached layouts. The elements in the dotted line boxes shows the INFOVIS techniques chosen to represent the each structure.	59
Figure 28 – The resulting Dendrogram Composition visualization.(A) shows the dendrogram, showing the regions of Brazil in its first hierarchical level, the five regions. (B) shows the bar charts aligned representing data from a single year. (C) shows the glyphs aligned beside the bar charts. (D) shows a tooltip, which is the result of the interaction with a glyph that shows the raw values encoded in each axes.	60
Figure 29 – Scale changes over a node's bar chart to highlight patterns. The two nodes refer to the same locality, (1) presents the normal scale and (2) presents the scale change.	60

Figure 30 – The resulting Choropleth Composition visualization. (A) a dropdown menu where a year, out of six, is selected to color the big choropleth map. (B) Brazil’s choropleth map divided in states and colored according to the specified year proportion of LACs to population size. (C) small multiples of six choropleth maps of one selected state divided in its mesoregions also according to the proportion. (D) six heat map tables each forming individual mosaics. (E) six heat map stripes that together form a mosaic in for a different perspective. . . . .	61
Figure 31 – Hover over the mesoregion Metropolitana de Belo Horizonte in the state of Minas Gerais. . . . .	62
Figure 32 – System overview. . . . .	63
Figure 33 – Bar charts of the five Brazilian geographic regions. . . . .	66
Figure 34 – Dendrogram regions’ glyphs. Higher hierarchical levels present glyphs with similar shapes, close to a perfect hexagon, since they summarize the amount of LACs of all its sub-levels. This homogeneity makes it harder to get meaningful analysis about their LACs’ distribution. . . .	67
Figure 35 – Nordeste’s bar chart distribution in 2012, showing two high peaks. The first relates to May 15th, and the second to May 25th. . . . .	67
Figure 36 – Bahia state’s bar chart distribution, with associated glyph, showing that the two high peaks observed in Nordeste, in Figure 35, are associated to this state. . . . .	67
Figure 37 – Expansion of Bahia state’s node showing its mesoregions’ bar charts. The two high peaks observed in Bahia come from Metropolitana de Salvador mesoregion and Sul Baiano mesoregion. . . . .	68
Figure 38 – The glyph of Itamarajú revealing a high peak in 2012. . . . .	68
Figure 39 – Distribution of LACs in Alagoas state in 2012 showing a seasonal pattern with higher bars in the beginning and end of the chart. . . . .	69
Figure 40 – Bar chart of Leste Alagoano. This mesoregion presents the pattern observed in Alagoas state more pronounced than the other mesoregions in the state. Its glyph also indicates reduction of LACs over the years. . . . .	69
Figure 41 – Alagoas state’s bar charts from 2012 to 2017. An increasing in the amount of LACs in the begin/end of each year can be noticed. . . . .	70
Figure 42 – Leste Alagoano’s microregions analysis, with bar charts and associated glyphs. At this level, it represents a regional event, since the behavior is observed in all these microregions. . . . .	70



Figure 43 – Comparative Analysis of bar charts from Alagoas’ mesoregions and Mata Pernambucana’s microregions. Similar patterns could be observed in two different hierarchical levels: (A) in a mesoregion of Alagoas named Leste Alagoano, and (B) in a microregion of Pernambuco named Mata Meridional Pernambucana. . . . .	71
Figure 44 – Analysis of São Miguel do Araguaia’s associated glyph (A), and the respective absolute values (B). This microregion presented a growth in the number of registered LACs. . . . .	71
Figure 45 – (A) shows the glyph associated with the city of Vila Rica and (B) shows its associated values. The number of LACs registered in 2013 is significantly bigger than in the other years. . . . .	72
Figure 46 – Comparative analysis of Ananindeua (A) and Paranaguá (B) glyphs. Ananindeua is three times bigger in population size than Paranaguá, but in Paranaguá the amount of registered LACs are 35% bigger. . . .	72
Figure 47 – Brazil’s choropleth maps, showing the variation of blue shades from 2012 to 2017. The predominant pattern presents the states in the north of the map in light blue shades, while the states in the south are in darker blue shades. . . . .	73
Figure 48 – Small multiples and heatmaps detailing the distribution of LACs in São Paulo state from 2012 to 2017, revealing a very homogeneous behavior over the whole range of years. . . . .	73
Figure 49 – Small multiples and heatmaps detailing the distribution of LACs in Minas Gerais state from 2012 to 2017. The second most populous state in Brazil is not as homogeneous as São Paulo. . . . .	74
Figure 50 – Mesoregions of Minas Gerais state. Triângulo Mineiro/Alto Paranaíba mesoregion presents a darker shade of blue compared to Metropolitana de Belo Horizonte, in which the capital city of the state is located. . . .	75
Figure 51 – Small multiples and heatmaps detailing the distribution of LACs in Amazonas state from 2012 to 2017. The small maps shows that Centro Amazonense mesoregion is responsible for the majority of LACs reports. The heat maps shows the distribution of LACs for each day from 2012 to 2017, and the workdays/weekends patterns can be better seen in the stripes forming diagonal lines from top to bottom. . . . .	76
Figure 52 – (A) shows the glyph associated with Amazonas state and (B) shows its associated values. A significant reduction in the number of LACs could be observed in this state both in the glyph, which presents a sharp point related to 2012, and in the tooltip, that confirms this reduction over the years. . . . .	76

Figure 53 – A detailed view of Alagoas state from 2012 to 2017 showing the peculiar pattern of LACs in the beginning/end of each year. This pattern can be confirmed by analyzing both visualizations in our proposed system. 77

---

## List of Tables

Table 1 – IBGE’s data set attributes description, with Brazil’s territorial division used for geographical context. . . . .	50
--	----



---

## Acronyms list

**BFLPO** Brazilian Federal Labor Prosecution Office

**CAT** Comunicação de Acidente de Trabalho

**GDP** Gross Domestic Product

**INFOVIS** Information Visualization

**LA** Labor Accident

**LAC** Labor Accident Communication

**ODSST** Observatório Digital de Saúde e Segurança do Trabalho

**VA** Visual Analytics



---

# Contents

<b>1</b>	<b>INTRODUCTION . . . . .</b>	<b>29</b>
<b>1.1</b>	<b>Objectives and Contributions . . . . .</b>	<b>31</b>
1.1.1	Hypothesis . . . . .	31
1.1.2	Contributions . . . . .	31
<b>1.2</b>	<b>Thesis Organization . . . . .</b>	<b>32</b>
<b>2</b>	<b>FUNDAMENTALS . . . . .</b>	<b>33</b>
<b>2.1</b>	<b>Information Visualization . . . . .</b>	<b>33</b>
<b>2.2</b>	<b>Visualization of Time Varying Data . . . . .</b>	<b>34</b>
<b>2.3</b>	<b>Governmental Data Communication . . . . .</b>	<b>41</b>
2.3.1	Visualization of Governmental Data . . . . .	41
2.3.2	Brazilian Federal Labor Prosecution Office and the ODSST . . . . .	43
<b>2.4</b>	<b>Final Considerations . . . . .</b>	<b>46</b>
<b>3</b>	<b>PROPOSAL AND SYSTEM . . . . .</b>	<b>49</b>
<b>3.1</b>	<b>BFLPO's Data Set . . . . .</b>	<b>49</b>
<b>3.2</b>	<b>Design Considerations . . . . .</b>	<b>50</b>
3.2.1	Dendrogram Composition . . . . .	51
3.2.2	Choropleth Composition . . . . .	54
3.2.3	Visualizations' Structures . . . . .	57
<b>3.3</b>	<b>System Description . . . . .</b>	<b>58</b>
3.3.1	The Dendrogram Composition . . . . .	59
3.3.2	The Choropleth Composition . . . . .	61
3.3.3	Implementation Details . . . . .	62
<b>4</b>	<b>ANALYSIS RESULTS . . . . .</b>	<b>65</b>
<b>4.1</b>	<b>Analysis Procedure: Strategy's Evaluation . . . . .</b>	<b>65</b>
<b>4.2</b>	<b>Dendrogram Composition . . . . .</b>	<b>66</b>

4.3 Choropleth Composition . . . . . 73

4.4 Final Considerations . . . . . 77

5 CONCLUSION . . . . . 79

BIBLIOGRAPHY . . . . . 83



I hereby certify that I have obtained all legal permissions from the owner(s) of each third-party copyrighted matter included in my thesis, and that their permissions allow availability such as being deposited in public digital libraries.

student name and signature



---

## Introduction

Labor accidents (LAs) represent a serious problem to the society. Around the world, per year, the numbers exceed 374 million of LAs recorded, resulting in more than 2.78 million of deaths. It is estimated that the amount of money spent due to poor safety work conditions is about 3.94% of the global GDP(ILO, 2019). In Brazil, from 2012 to 2018, more than 4.5 million of LAs were recorded and among them approximately 16 thousand resulted in death. For social security the expenditures were about 29 billion of Brazilian reais(Smartlab de Trabalho Decente MPT - OIT, 2017). LAs can impact all sectors of the society, resulting in physical or psychological damages to the employee, loss of manpower, expenses with compensations, and fines to the employer, as well as social security expenditures with compensations and hospitalizations to the State.

In Brazil, the Brazilian Federal Labor Prosecution Office (BFLPO) is the institutional agency responsible for the inspection of labor conditions when there is a public interest, aiming to regulate and mediate the relationship between employees and employers (Ministério Público do Trabalho, 2018). One of its main duties is the supervision and control of labor health and security, and to do it effectively, it keeps records of the LAs using documents named *Comunicação de Acidente de Trabalho* or Labor Accident Communication (LAC), which are documents that should be filled by the companies to communicate the occurrence of a LA, occupational disease or death(INSS, 2018).

Labor accident data relates two important aspects: geographical and temporal information. The geographical information provides a good comprehension of a country's situation. Each geographical position reveal differences or similarities related to several associated particularities. This sheds light on the understanding of LAs' evolution since some factors related to population size, economic activity, efficiency in inspection policies, etc, may impact the type and frequency of accidents occurrence. The LA data is usually collected in a specific locality with some time specification, such as date and hour, and this temporal information is the main characteristic in the analysis of LAs evolution, since it reveals how the accidents are distributed over the years and also can reveal patterns, anomalous behaviors, seasonalities, as well as the impact of public policies. Considering

the time, governmental data can be seen as temporal series whose observation is important to study the past, to understand the present and to make predictions about the future, as well as for comprehending phenomenon behaviors in specific spans of time. From that arises the need to understand the LA data evolution in order to identify patterns in the occurrence of LAs and to better understand LA occurrence behavior, providing means for creation of public policies to improve work conditions in the country.

In a country of continental proportion, such as Brazil, the amount of governmental data produced is huge and, as the data is collected over time, this data volume daily grows. There are different approaches that can be applied to understand huge data volumes and its evolution. The manual analysis is arduous, demands a considerable amount of time and is susceptible to mistakes. Automatic approaches employing machine learning techniques, while often effectively applied to big data volumes, usually results in processes in which the user is not able to interfere, except for parameters' empirical manipulation, which may impair users to understand data behavior and infer reasons to justify this behavior.

In Brazil, the BFLPO and the International Labor Organization (ILO) provides the *Observatório Digital de Saúde e Segurança no Trabalho* (ODSST), which is a website containing information about LAs in two main approaches, statistical, with some gross values, and visual, with some charts. They also publicly provide the raw data using spreadsheets in which each line presents a record of one LAC gathered at some specific period. The exploration process and data analysis available in the website is very limited and is not suitable to reveal complex patterns, specially for analysis of evolution of LAs' occurrence, since it presents only simple bar charts without interactions, significantly reducing the data mining ability by the user.

Information Visualization (INFOVIS) is a research area that develops data exploration strategies by mapping abstract data to visual representations to enhance the cognitive capacity of the user (MUNZNER, 2014). Visual representations and interactions may be valuable to analytical tasks in the sense that they accelerate rapid insight into complex data (THOMAS; COOK., 2005), revealing particular characteristics and providing means to the user to make inferences about it, performing such tasks intuitively and effectively.

INFOVIS is a potential tool to reveal patterns and trends in the data gathered by the BFLPO, representing an important approach to understand the variation over time in this huge amount of data. Analyzing the evolution in the LA data behavior allows, among other tasks, to understand the period of occurrence of the events, seasonalities and anomalous behaviors, which may have impact on the decision making by the diverse administrative sectors of society and government. A visual analysis strategy can also benefit the population, since it communicates, in a simple and clear manner, the information within the data, enabling the monitoring of government actions and increasing their awareness about the locality they live, in order to encourage their participation in creating safety policies. Thus, the analysis of the LA data behavior over time becomes a

relevant research task.

Moreover, there are several different INFOVIS techniques that can provide different perspectives of a given data set, in this sense the combination of techniques well known in the literature could be of good use to potentialize information extraction.

## 1.1 Objectives and Contributions

The analysis of LAs is an important task performed by the BFLPO, but even though they lack of strategies that could potentialize this analysis. In this sense the objective of this research work is to develop a visual temporal analysis strategy of labor accident data, employing INFOVIS techniques. The specific objectives are:

- ❑ To setup layouts for different perspectives analysis of the temporal evolution in the BFLPO's LACs data set;
- ❑ To combine different interactive layouts attending to the proposed structure to create visualizations that potentialize information extraction;
- ❑ To develop a system to evaluate the proposed strategy;

### 1.1.1 Hypothesis

A visual temporal exploration strategy, using interactive Information Visualization techniques, better communicates the temporal evolution of labor accidents behavior to the user enhancing the comprehension of the accidents occurrence distribution over time.

### 1.1.2 Contributions

The contributions of this work are:

- ❑ A strategy for temporal visual analysis of LAs data that combines a set of INFOVIS techniques that provide different perspectives of data, potencializing the information extraction;
- ❑ A temporal visual analysis system that communicates the patterns and anomalies within Brazil's LA data, allowing the analysis of the distribution of the accidents over time and the comparisons of the LA evolution from different perspectives;
- ❑ Suitable visual metaphors combination that characterize the occurrence of LAs in Brazil, providing to the government strategic information to guide the creation and/or improvement of public policies, as well as to evaluate their impact over time in different geographic regions.

- An instrument to simplify the communication of LA data to the population, granting real transparency of governmental information to the citizens, and allowing them to better identify the problems, as well as to comprehend the impacts of the public policies over time, playing their role as government inspectors.

## 1.2 Thesis Organization

This thesis is organized as follow: chapter 2 presents the basic concepts related to this research, as well as the related works to the main topics addressed. Chapter 3 presents the strategy we proposed using INFOVIS techniques and the design considerations for the system that we implemented to test our strategy. Chapter 4 presents the results of using our system in the analysis of the LA data provided by the BFLPO regarding temporal context. At last, we present chapter 5 with our final considerations.

---

## Fundamentals

This chapter presents some fundamentals and discusses the main topics related to this research work. First, we present the Information Visualization research area, and focus on some techniques that perform the visualization of temporal data. We also introduce the governmental data and present discussions on strategies for visual analysis of this type of data. Finally, we detail the BFLPO, as well as the labor accident repository and the already developed computational tool that they provide for analysis.

### 2.1 Information Visualization

Information Visualization (INFOVIS) is a research area that studies techniques to create visual representation from abstract data to aid people to do tasks more effectively. One of the main goals of the INFOVIS is to investigate which techniques are more suitable to associate a data set with a specific task. It is used to enhance humans cognitive ability in tasks where they can not be substituted (MUNZNER, 2014).

The INFOVIS is based on the most keen human sense, the vision. Due to the great communication channel between the brain and the eyes, a large amount of visual information is quickly processed. Other advantage is the information technology for communication, which provides more resources to explore the vision than to any other human sense (MUNZNER, 2014).

Figure 1 presents the classic visual analysis process, described in 5 stages (LIU et al., 2014), which are:

1. **Data Transformation and analysis:** The data are studied and pre-processed to be adequately structured for the analysis. Examples of pre-processing tasks include missing values processing, sumarization, among others;
2. **Filtering:** The data are filtered and subsets are selected to compose each specific visualization, according to the nature of the data;

3. **Mapping:** The chosen subsets are encoded in visual representations that consider geometric primitives as lines, points, circles or more complex visual metaphors, with different attributes as color, size, texture, etc;
4. **Rendering:** The created representation is transformed in an image to be presented to the user;
5. **UI Control/Controls:** A set of interactions are added to the rendered image to allow the user to do the visual exploration of the data, showing different perspectives.

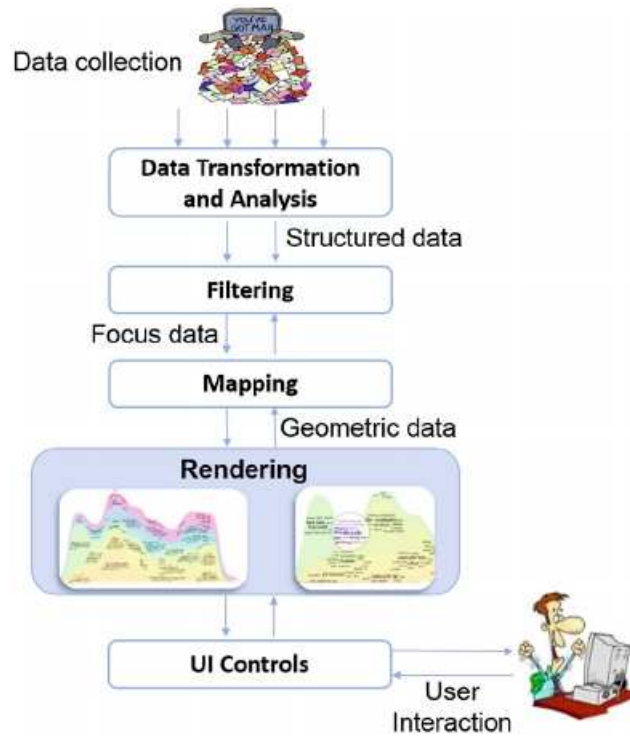


Figure 1: Pipeline of the visual analysis process using INFOVIS.

Source: (LIU et al., 2014).

INFOVIS provides a variety of visual representations that can be combined with interactions in myriad ways to address different needs. Visual analysis can benefit several different areas such as Medicine (LEE et al., 2010; KOLESNICHENKO et al., 2019), Education (LU; WEI, 2010; TERVAKARI et al., 2014), and Engineering (PARKINS; MCCUTCHEON, 1999; DEWEN; PING, 2011; WANG et al., 2012), among others, to perform different tasks such as data exploration. In the next sections we present some studies related to the main focus of this research work that use from simple to more sophisticated INFOVIS techniques and the particular case of the BFLPO system.



## 2.2 Visualization of Time Varying Data

Time varying data is data observed and collected over time (LEE et al., 2015). The time attribute can vary in scales, from nanoseconds to centuries. The main goal in time varying data analysis is the understanding of data evolution, in order to properly address the phenomena described by the data considering the date of occurrence of the events. Beyond the difficult of deal with a large amount of data, another challenge is the structure of the data, if it is irregular, unstructured or scattered (JANG; EBERT; GAITHER, 2012). The analysis of time varying data is important for both the study of the past and for predictive calculations, as well as to comprehend specific phenomena's behaviors over specific periods of time. This section presents applications of information visualization techniques with focus in time varying data analysis in different scenarios.

Lin-spiration (GRAELLS; JAIMES, 2012) is an interesting visualization which is composed of temporal series and interaction tools that allow data exploration. This visualization uses a linear layout with curled spirals in both ends aiming to provide two perspectives of the data. The central part of the layout, where the series are shown straight, focus on a specific period of the data under analysis. On the other hand, spirals concentrate all the rest remaining time series information, allowing the user to have an idea about the data not focused at a specific moment. The visualization allows multiple time series to be analyzed simultaneously, in a stacked way and also presents a scroll bar for each time series so one can control what should be on focus.

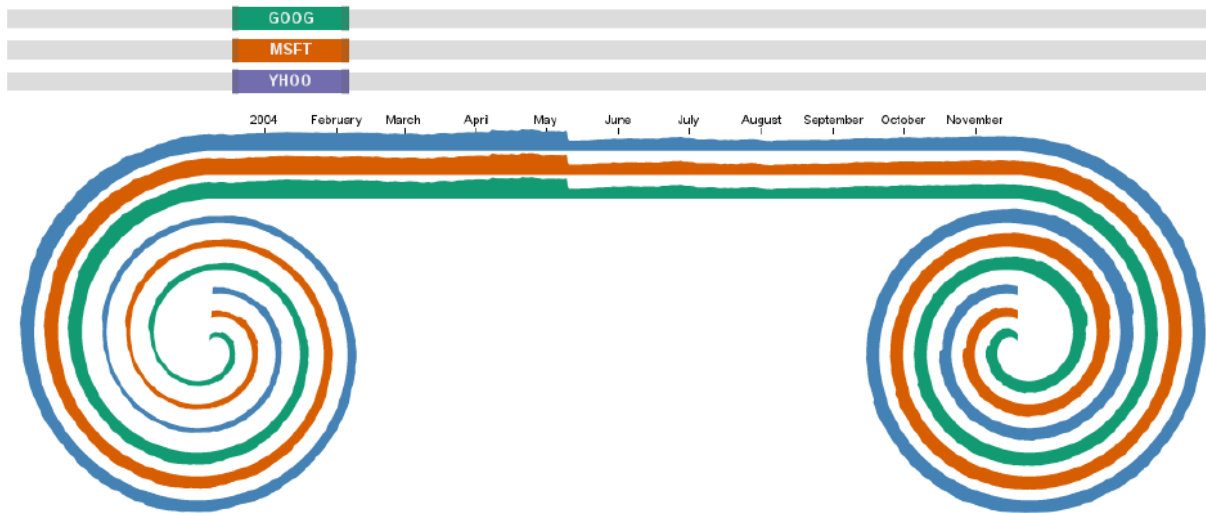


Figure 2: Representation of the time series of the Lin-spiration.

Source: (GRAELLS; JAIMES, 2012)

Figure 2 shows the complete visualization with three time series representing stock market data. The layout employs scroll bars that represents each stock market: YHOO in green, NSFT in orange and GOOG in purple. The center of the image shows the behavior between February to November, and the curled ends contain data from previous/posterior

periods, providing an overview of how much data was measured, and how it behaves over time. As the scroll bars are moved, the visualization also moves, rolling and unrolling the spirals. This layout favors the perception of nuances in the data and allows comparisons of different time series and different periods of time.

Another aspect that can be explored in a temporal data is the hierarchy and a variety of approaches incorporate this hierarchical information in a combination of different layouts. In (LI et al., 2013) INFOVIS techniques are applied to the study of index pollution in the province of Shandong in China. This proposal is very similar to one presented earlier by Burch, Beck e Diehl (2008). Both visualizations consist in a tree, combined with a linear structure for the temporal aspect. This temporality is represented in both works as different bar charts that are shown for each leaf node of a tree, as can be seen in Figure 3 and Figure 4. Other charts are used to summarize measured values as sector charts, and thumbnails show a tiny representation of the existence of a transaction related to its node for overview purposes. Visualizing temporal data organized in hierarchies is an interesting approach that can benefit comparisons and identification of patterns and trends in different levels of abstraction, as shown by the authors of both works. This structure presents potential for deep investigation of the temporality using filters, scale change, and coordination with other layouts. However, a simple tree that does not show the leaf nodes aligned may be an obstacle that can decrease the perception of the nuances impairing the comparison.

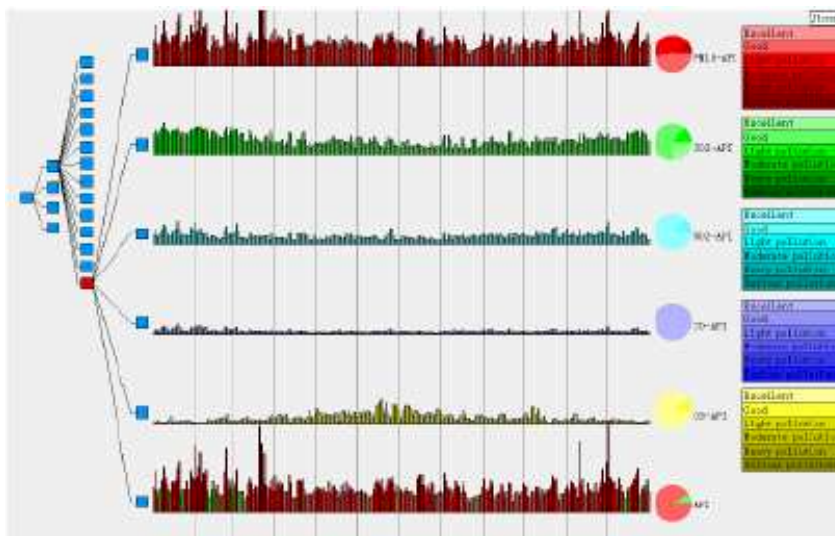


Figure 3: Layout proposed by Li et al. (2013) displays a tree-like diagram presenting five pollutant indexes and the air pollution index(API). The root node represents the province of Shandong, the next layer of nodes represents the cities being monitored, the third layer represents the monitor stations of air pollution and the last layer represents the pollutants measured in the station. Each bar chart represents one pollutant index measured during the year 2010 and the API.

Source: (LI et al., 2013)

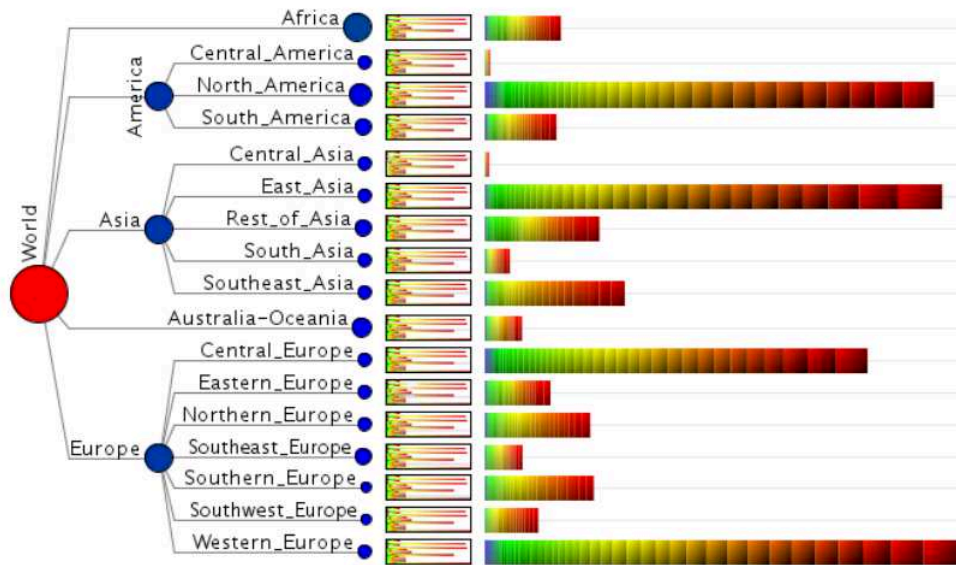


Figure 4: Layout proposed by Burch, Beck e Diehl (2008) presenting a tree-like diagram in which each node is a region of the world. A thumbnail structure and a bar chart is associated to each node.

Source: (BURCH; BECK; DIEHL, 2008).

The work developed by Morawa et al. (2014) deals with hierarchy of time series, such as decades, years or months. This work proposes a graph structure over a timeline with a structure named **Time Shadow**. Each node of the graph represents an event and has a shadow, which is a semi-transparent layer behind the node that indicates the duration of the event. The edges of the graph represent the relationship between the events, and with them it is possible to determine whether one event has influence over another event. Figure 5 shows an example with part of a graph placed over a timeline over a span of years around 1970 to 1990. The nodes are representing different persons and each node has a Time Shadows in elliptical shape.

The authors also propose interactions with the timeline, as shown in the Figure 6. Figure 6-a shows that it is possible to enlarge the representation space of some period, in order to show with more details the events represented by the graph, that occurred at the moment. This interaction can also be used to reduce the space of a non interesting area, giving space to enlarge others. Figure 6-b shows that it is possible to change the scale of the timeline in specific periods, changing the scale of just an isolated span in the timeline. In Figure 6-c, one can compress specific periods of time to allow non adjacent periods to be shown side by side. The example shows a scale of centuries, where 1700 is hidden to show 1600 beside 1800. These proposed interactions are interesting ways to deal with visual timeline representations since there is no need of a fixed scale. A further idea could be the addition of summarization structures when the time is compressed, in order to provide new ways to find patterns.

Soriano-Vargas et al. (2017) propose a visualization system to analyze the ionospheric

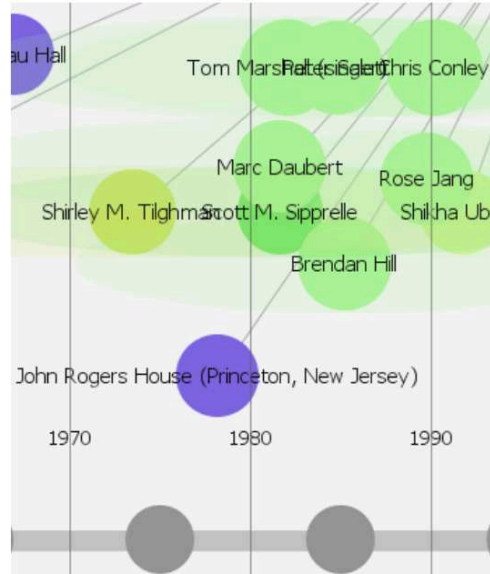


Figure 5: Disposal of the nodes of the graph over the timeline.

Source: (MORAWA et al., 2014).

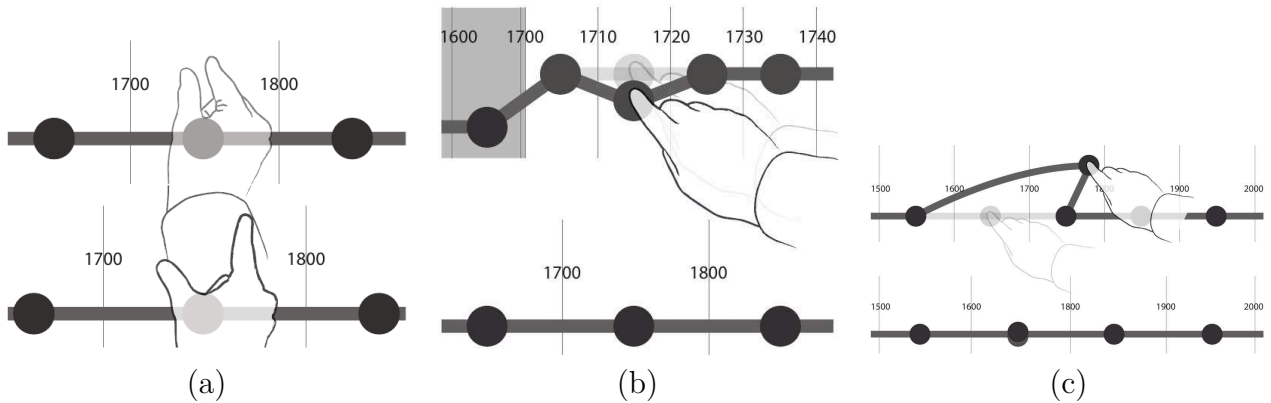


Figure 6: (a) Interaction to zoom in or out on a specific period. (b) Interaction of transitioning among different time scales in specific periods. (c) Interaction to hide some specific period to allow comparison of non adjacent periods.

Source: (MORAWA et al., 2014)

scintillation phenomenon named **Time Matrix**. The matrix shows individual values in its cells that were recorded and aggregated considering some time unit. In the same time that it is possible to find individual values, the matrix itself also forms visual patterns, which can be visually meaningful, once this values are mapped to colors in a single hue. An example of their Time Matrix can be seen in Figure 7 that presents a set of small multiples. Each Time Matrix is encoding observations of a whole month (1-31 January, 2014), and each row of a Time Matrix shows the observations along one day. In this visualization, they present the set of matrices as small multiples of complete matrices and a specific variable is depicted for each matrix in a specific period. This matrices can show time spans in which there are several missing data, colored in gray, and in this case the shapes formed inside each time matrix can be difficult to compare. The time matrix



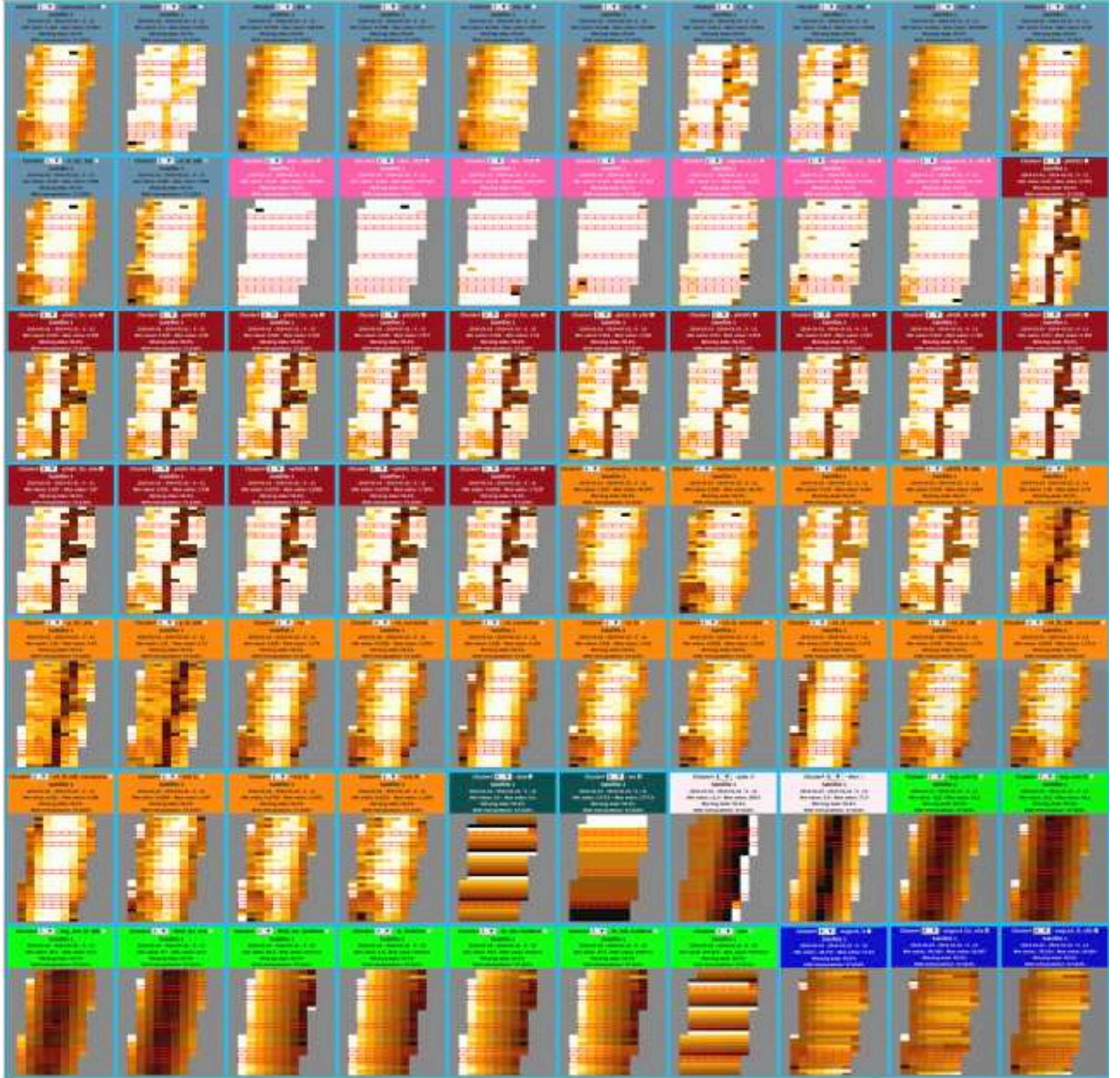


Figure 7: Small multiples of Time Matrices depicting observations of ionospheric scintillation variables. Each Time Matrix represents the period 1-31 January, 2014, and each line of the Time Matrix represent the observations of one day.

Source: (SORIANO-VARGAS et al., 2017).

approach aids the identification of elements with similar temporal behavior.

Other works add context information to the visualizations. Sharmin et al. (2015) propose 4 types of visualization: space-time, time, contextual, and event-based. All of them are time-based and show the stress of an individual during different moments of the day. The authors suggest the use of filters to set the time period to be exhibited. All the visualizations in this work show a period of one day, but they also suggest that more days can be shown, for instance, all the week days. They present a summarization and when expanded, details on demand can be shown. Figure 8 presents one of their the contextual visualization, in which the relationship between the stress and the executed

activity considering the time is shown. Depicting the context in a visualization is an interesting approach to improve the comprehension about data evolution, as it combines the computational power of the machine with the specialist knowledge in the tool to communicate information in a simple and clear way.

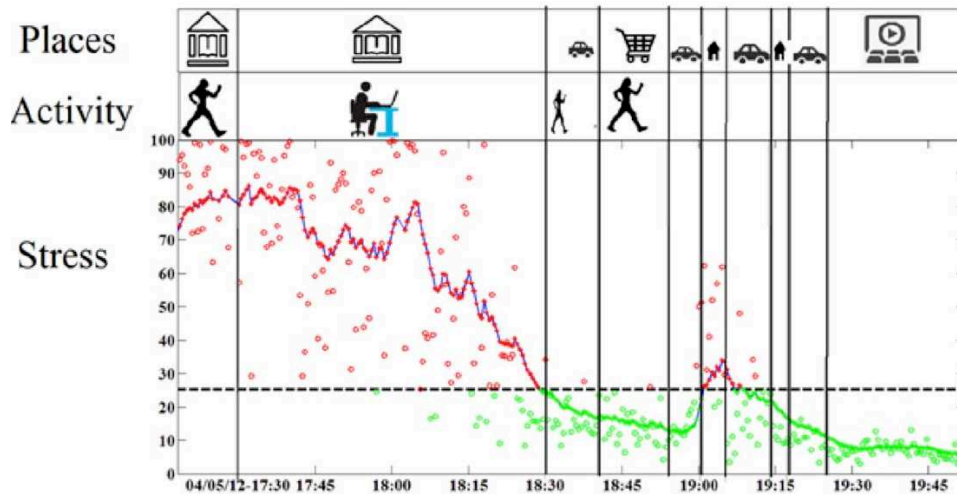


Figure 8: Context visualization in (SHARMIN et al., 2015) displaying the stress level of an individual according to her activities during the time span of one day.

Source: (SHARMIN et al., 2015)

In (BADAM et al., 2016) the context is used in a tool that combines visualization of time series with statistic analysis metrics. This tool presents the evolution of a time series until some specific period, then the rest of the series is presented as branches that simulate future behavior. These branches are calculated observing the predictions of a computational model applied to the time series data and it is up to the specialists the evaluation of the events' context, and selection, based on their knowledge, of which branch presents higher probability to occur.

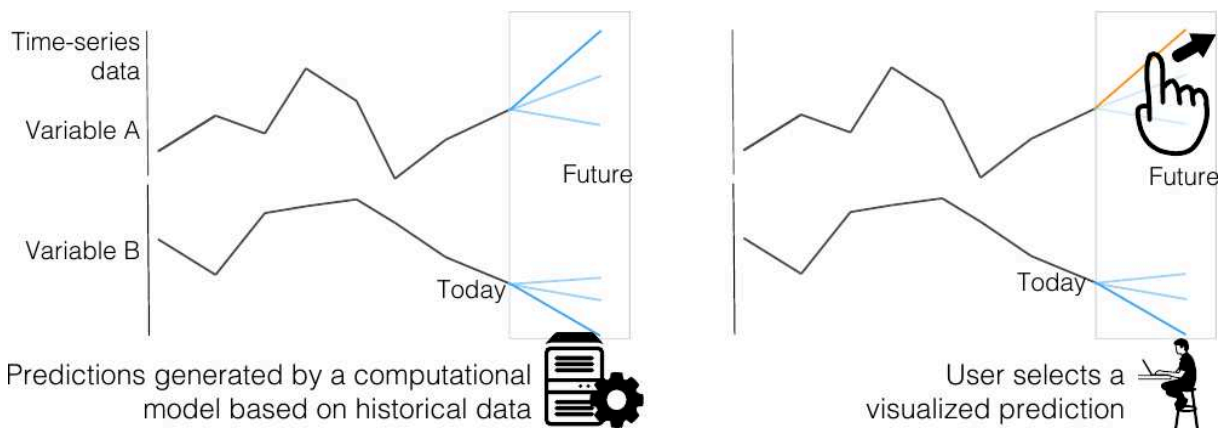


Figure 9: Context visualization about stock exchange prediction in (BADAM et al., 2016).

Source: (BADAM et al., 2016)

The example shown in Figure 9 illustrates a stock exchange prediction scenario. When the specialist choose one branch, new predictions are done based on the previous choice. This approach is promising to be applied on labor accidents analysis, since the specialists could evaluate the current behavior of a city based on several factors such as its locality, economic activity, etc. Also, based in the predictions, users may determine which public policy may produce better results on a specific situation.

## 2.3 Governmental Data Communication

The definition of governmental data consists in all data of any type or format that is collected, created, received, maintained, or disseminated by any government entity, independently of its physical form, storage or condition of use (State of Minnesota - Department of Health, 2019). Governmental data concern about different sectors of the society all over a country, which implies the creation of voluminous data sets.

Basically almost all governmental data is public, which implies in the government creating methods to make them available to the population. However, the analysis and understanding of a large amount of data is already a challenge for specialists in data science, and for non-specialists, the population, it is a much more tough task. In (MENDONÇA; MACIEL; FILHO, 2014) the authors highlight that it is important to make governmental data available to the population, but this data must be provided in a way that allows experts and/or citizens to extract meaningful and strategical information from them. It is thus necessary the use of proper structures and tools to address data analysis. The use of new technologies, such as INFOVIS techniques, to present governmental data can improve government-population communication, and the understanding of the data can improve the population's trust in the government (YING; XIALING; WEI, 2017; DÍAZ; AEDO; HERRANZ, 2014). The analysis of governmental data using INFOVIS techniques can help specialist of the government in identifying problems, creating adequate public policies and in decision making, and it can also engage the population in inspecting the government actions for better social conditions. In the next section we present some works that use INFOVIS in the analysis of governmental data from different sectors of the society.

### 2.3.1 Visualization of Governmental Data

The use of INFOVIS techniques is an interesting solution to deliver governmental data in a more affordable way for both specialists and population, and such representations can be often found in official news and websites. Some examples are the projects developed in Africa(International Food Policy Research Institute (IFPRI) and Datawheel, 2017), United States(DATAWHEEL, 2016) and Chile(DATAWHEEL, 2018) that provide information about different sectors of the society using graphics and maps. In (HOXHA;



BRAHAJ; VRANDEČIĆ, 2011) visual strategies are employed to provide to the population information about economy, education, health, among other sectors, using simple line charts, bar charts, sectors charts and even treemaps. In Brazil, a similar approach named DataViva(DATAVIVA, 2013) aims to collaborate in the development of public policies, academic researches and to encourage public and private investments, in order to help in the solution of society problems, using official data. DataViva also allows some interactions with its layouts, but the offered layouts are significantly limited, and do not offer any coordination among themselves, which make harder the identification and comprehension of more complex patterns.

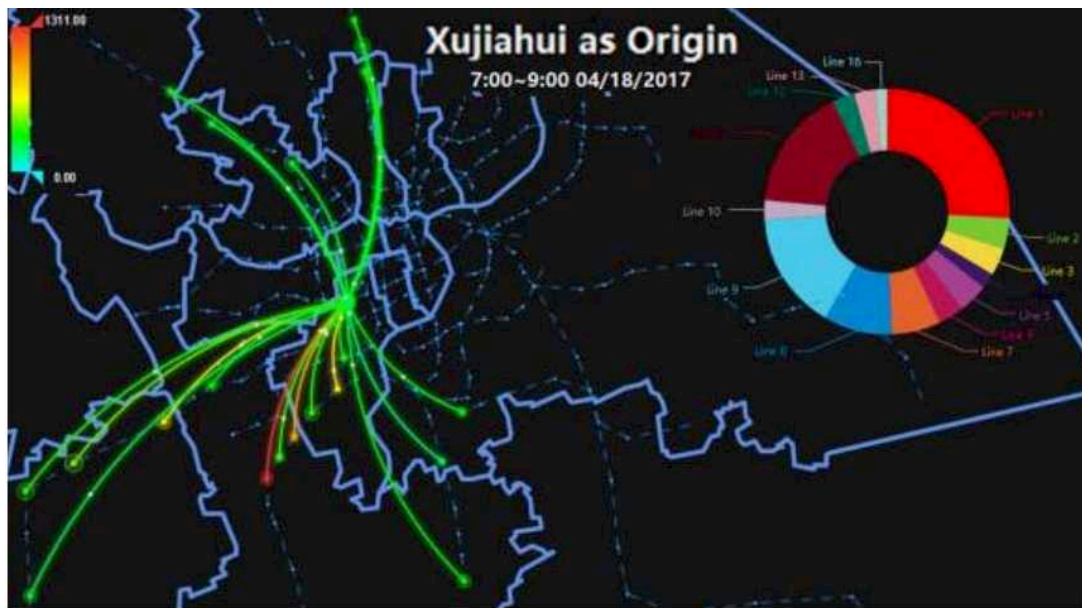


Figure 10: Visualization of people flow departing from the station of Xujiahui during two hours of observation. The arches indicate the destination of the passengers, the colors indicate the volume of passengers and the sector chart beside indicates which subway lines were more used and the proportion of use among them.

Source: (ZHIYUAN et al., 2017)

Websites that uses simple INFOVIS techniques are important, as the visual representation facilitate to communicate strategic information to the user. However, when one has many layouts to analyze, relate them to extract more information, or information from different aspects, may not be straightforward. In this sense, some works propose to combine different layouts in order to provide different perspectives of the data. In (ZHIYUAN et al., 2017) it is proposed the use of various Information Visualization techniques, as well as combinations among them, aiming to understand the flow of passengers in a subway of Shanghai. For example, Figure 10 shows two layouts together, a map and a sector chart. The map shows the subway stations with Xujiahui station as the departing point, and from there a set of edges indicate the destination of the passengers. The edges colors, as in a heatmap, indicate the volume of people that used the routes, in a range of two hours,



from 7am to 9am, in April, 18th, 2017. The sector chart shows the subway lines that were used and the proportion of their use. This work also proposes additional layouts, as histograms, bubble and line charts, to better understand the flow in more than 300 stations in Shanghai. Considering the space-time nature of the data, this is an attractive proposal, since the combination of a set of techniques can offer a broader view of the data, in which the evolution of events can be directly associated with the locality in which it happened, and the evolution in different localities can be compared. In the map, the main destiny is depicted by the warmer colors and the sectors chart indicates which is the main train line used.

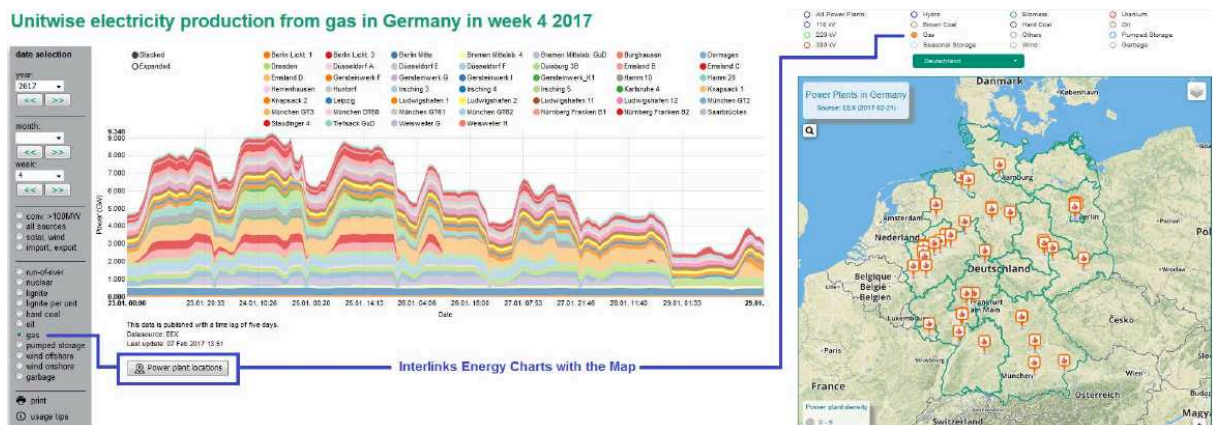


Figure 11: Visualization of German power plants data.

Source: (RODRIGUES et al., 2017)

A study on the energy sector in Germany (RODRIGUES et al., 2017) also employs layouts coordination. The system provides map visualizations to indicate the geolocation of each power plant in Germany, and a ThemeRiver to show changes in consumption of each type of energy over time, allowing to perform summarization and comparisons, as shown in Figure 11. Some options can be selected in a side panel, than the power plants selected are marked in the map and the ThemeRiver is generated. Due to the limited number of colors, the authors use tooltips in order to not confuse the analysis. One of the goals of this work is to make information about energy consumption clearer to the population to increase their knowledge and incite them in changing attitude

In the United States of America, a visualization (GROEGER; GRABELL; COTTS, 2015) shows the mean of restitution to be paid to a worker per accident that results in amputation of some body part for each state of the country. The visualization is basically comprised of one entire body for each state and the body parts' sizes are represented proportionally to the rest of the body according to the restitution paid for each. Thus, as shown in Figure 12, it is easy to find which state pays more, or which body part has better compensation, but the visualization can lead us to many other questions, as “why one state pays more than other?”, “is it because the state has better public policy or is it related to the economic activities?”.

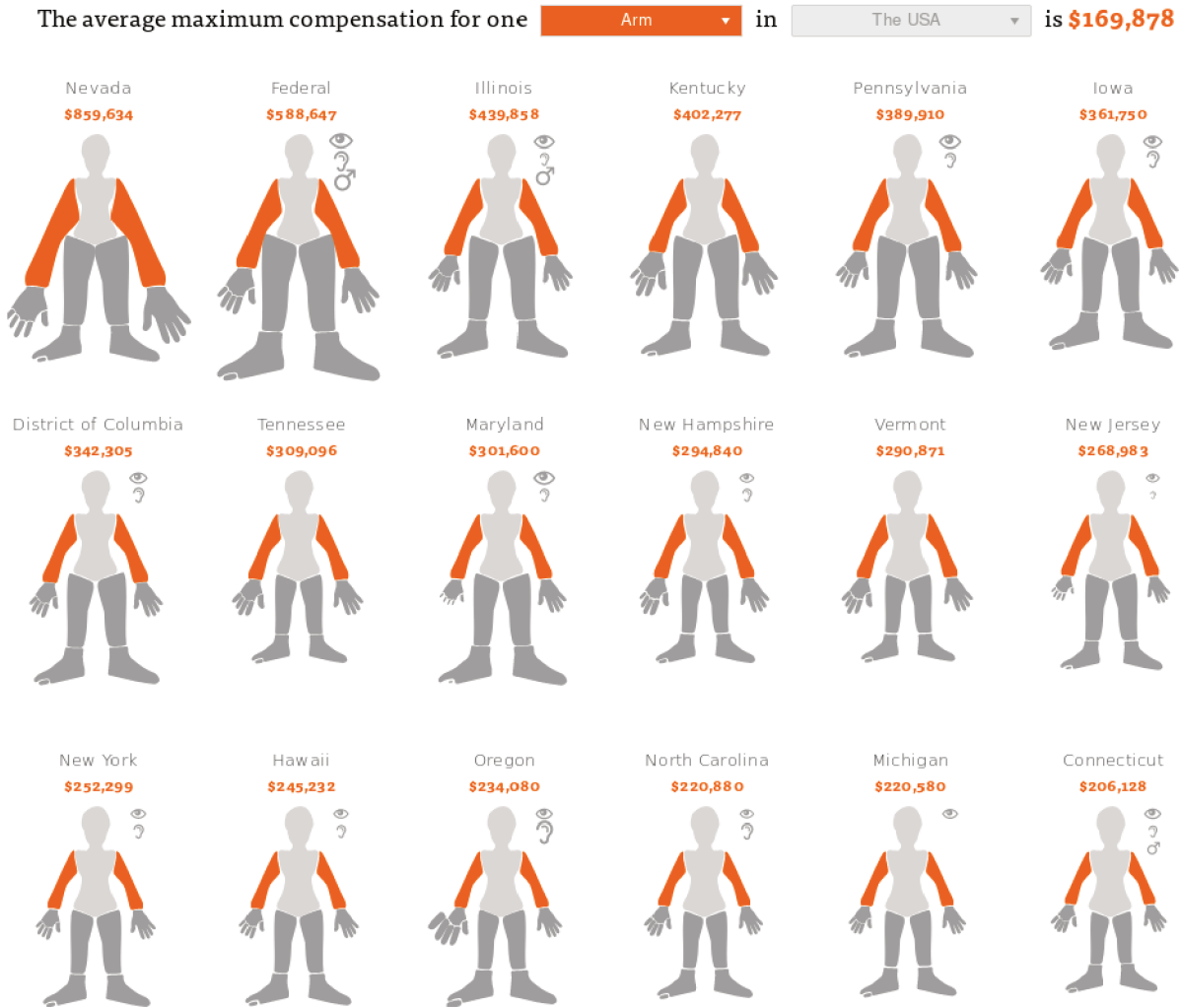


Figure 12: Representation of the mean compensation paid by the loss of one arm in the states of the USA.

Source: (GROEGER; GRABELL; COTTS, 2015)

### 2.3.2 Brazilian Federal Labor Prosecution Office and the ODSST

The Brazilian Federal Labor Prosecution Office (BFLPO) is an official government agency which is responsible for supervising labor conditions in Brazil, and one of its assignments is to inspect the conditions of health and safety at work. Each labor accident (LA) occurred in Brazil has its own *Comunicação de Acidente de Trabalho (CAT)*, or Labor Accident Communication (LAC), which is a document to describe the accident in details, including date, time and locality where the accident occurred, among others. From 2012 to 2018, a massive volume of data relative to LA was collected by the BFLPO and made publicly available in spreadsheet formats<sup>1</sup>. We believe that an effective representation of this data makes it more understandable to the population and helps domain experts to make decisions that may prevent accidents in Brazilian localities. In this sense, INFOVIS techniques represent a potential tool to provide this representation, creating in-

<sup>1</sup> Only data from 2012 to 2017 are available to download at the moment of this work's development.

teractive layouts that better represent this data. The BFLPO provides a digital platform named *Observatório Digital de Saúde e Segurança no Trabalho (ODSST)* (Smartlab de Trabalho Decente MPT - OIT, 2017), in which the LA data may be analyzed using a set of basic charts, and simple maps visualizations with summarization of the types of accident, type of local where they occur, role of the worker, economic sector, missing work days, among other aspects.



Figure 13: Bubble map of Brazil provided by ODSST. The bubbles indicate the localization of a LA occurrence and the size of the bubbles indicates the amount of missing work days due to the the accident.

Source: (Smartlab de Trabalho Decente MPT - OIT, 2017)

Some of the visual analysis tools provided by ODSST are presented as follow. Figure 13 shows the map of Brazil with bubbles that indicate the localities that reported LAs and the size of these bubbles indicate the volume of LAs. Some filters can be used to concentrate the analysis in specific geographical regions. Although maps are good visual metaphors, in this case the use of bubbles may result in cluttering and impair the analysis. In Figure 13, for example, the bubble representing the state of São Paulo overlaps many others close localities, which makes harder the analysis of these areas. Another limitation is to visually compare the size of the bubbles represented in the map, in order to compare

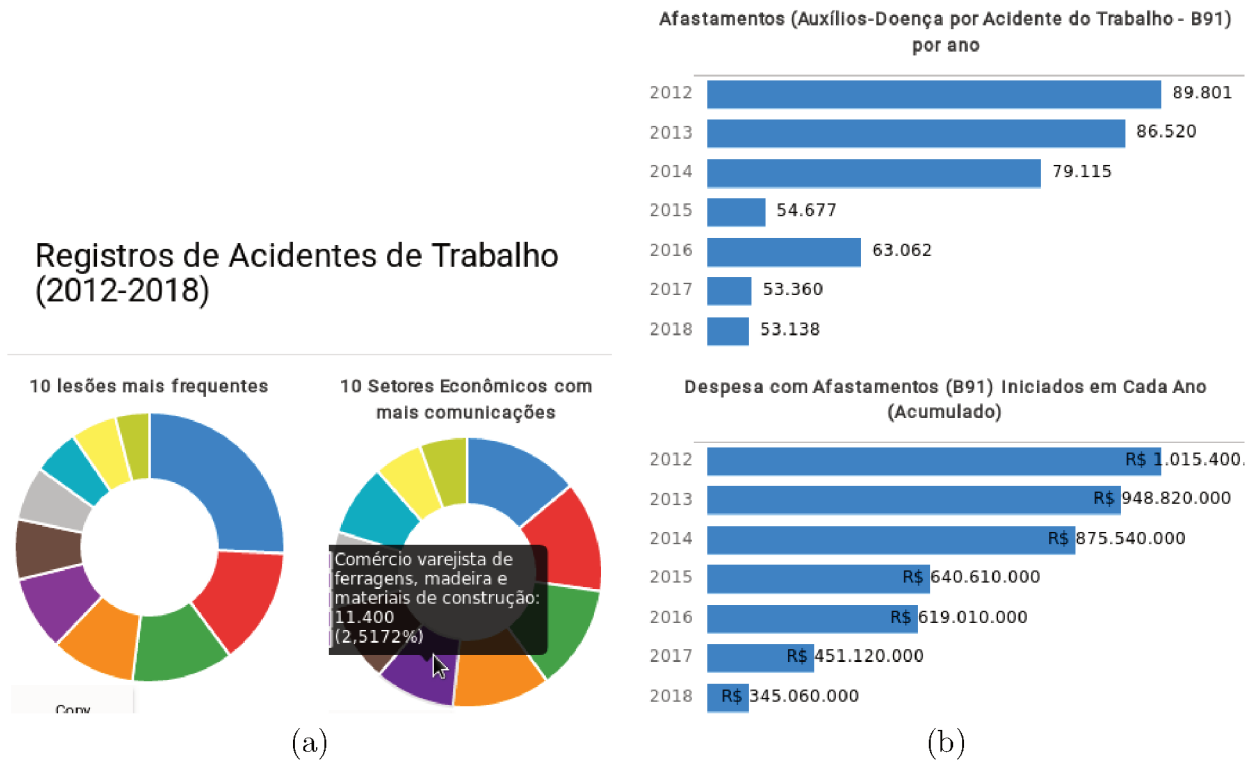


Figure 14: Basic charts provided by ODSST. (a) The sector chart indicates the 10 more frequent injuries and the 10 economic sectors with more reports of LA between 2012 to 2018, in São Paulo state. (b) The bar chart above indicates the accumulated amount of missing work days by year due to injured employee removal. The chart below indicates the spent volume with expenses arising from removals.

Source: (Smartlab de Trabalho Decente MPT - OIT, 2017)

different localities regarding the parameter under analysis.

Figure 14 shows simple sectors charts and bar charts with data from São Paulo state between 2012 to 2018. In Figure 14-a it is possible to see in the sectors chart the ten most frequent injuries registered in LACs. In Figure 14-b the bar chart shows the annual volume of records of illness leave requests and the chart for the annual illness expenditure. The charts only provide basic quantitative values, and it is hard to comprehend the reasons that lead to the exposed results. It is also hard to relate the data shown by the two charts, reinforcing the necessity of a more sophisticated interactive layouts for an effective analysis and exploration. In this sense, it would be interesting to create and apply appropriated visual metaphors that could better communicate the data, revealing more information quickly.

## 2.4 Final Considerations

In this section we discussed INFOVIS as a suitable approach for governmental data analysis and we focused on temporal data analysis, presenting works that use INFOVIS to

show data from different private and public sectors of the society. The analysis of temporal data is important to understand evolution. There are several different forms of visual temporal data analysis presenting different concepts among the works we discussed, such as Lin-spiration with overview and focus, context visualizations to help the interpretation of the data, evolution in hierarchical structures and predictions, and timelines to show events occurrences. However, the strategies of works focusing on visual temporal analysis of governmental data are less sophisticated, and often provide simple charts, such as bar charts, sector charts, line charts, etc., which impair the understanding of more complex patterns.

We also presented the ODSST, a Brazilian platform for LA data analysis. But as the other governmental data analysis systems, it also presents limitations related to the capability of revealing complex patterns. No other works could be found concerning about labor accident analysis. Our proposal aims to address this problem providing a strategy that uses INFOVIS to create a temporal analysis tool.



## Proposal and System

This research work proposes a strategy for temporal visual exploration and analysis of labor accident data. We propose two visualizations for this analysis, named **Dendrogram Composition** and **Choropleth Composition**, aiming to provide effective information extraction in few steps, to be used both by domain experts and by the population. The proposed visualizations were designed to attend the particularities of the data set provided by the BFLPO, in order to highlight strategic visual patterns present in the data from this repository. In this chapter we describe the data set and necessary pre-processing tasks performed, the design decisions taken for each visualization and a developed system for the proposal evaluation.

### 3.1 BFLPO's Data Set

The BFLPO's data set<sup>1</sup> is publicly provided by the BFLPO containing LA data collected from 2012 to 2017. Each record stores information of one LAC, properly anonymised. Since our analysis focus on the temporal data aspect, the simple existence of a LAC with specified locality and date is enough to be considered as a valid entry, even when the remaining attribute values were not properly informed. For this reason, only the date and geographical attributes were considered. In order to improve the context giving more precise information to the BFLPO's data set, as well as to retrieve the complete Brazilian hierarchical structure, the IBGE's official territorial division was added to the data, according to the information presented in Table 1, which considers the definition of 2016<sup>2</sup>. With the complete hierarchy defined, for each year the amount of LACs for the localities in all hierarchical levels were calculated and used to compose the visualizations.

The IBGE's population estimations<sup>3</sup> for each city in Brazil, yearly measured from 2012 to 2017, were also used. As the BFLPO data set stores LA data related to each city only,

<sup>1</sup> <https://smartlab.mpt.mp.br>

<sup>2</sup> [http://geoftp.ibge.gov.br/organizacao\\_do\\_territorio/estrutura\\_territorial/divisao\\_territorial/2016/](http://geoftp.ibge.gov.br/organizacao_do_territorio/estrutura_territorial/divisao_territorial/2016/)

<sup>3</sup> <https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?edicao=17283&t=downloads>

Table 1: IBGE’s data set attributes description, with Brazil’s territorial division used for geographical context.

Geographical Division	Hierarchy Level	Description	# of Categories
Region	1	Division of Brazil’s states in groups based on similarities to help the interpretation of statistics, without political autonomy.	5
State	2	Subdivisions of the country with political autonomy but with some subordination to the country’s government.	27
Mesoregion	3	Subdivision of the states to help the interpretation of statistics, without political autonomy.	137
Microregion	4	Subdivision of the mesoregions to help the interpretation of statistics, without political autonomy.	554
City	5	Administrative subdivision of the states with political autonomy but subordinate to the power of the states and the country.	5570

we summarized the values to compute the information of higher hierarchical levels. We also normalized the data considering the proportion of LACs to the population size in each locality, for each year, in order to get more accurate information regarding the relationship between population and LA occurrence, as well as the profile of specific localities in terms of these information.

## 3.2 Design Considerations

The temporal and hierarchical geographical organization are the main aspects that the proposed visualizations are intended to highlight. The temporal layouts are designed to reveal the evolution of the LACs over time, such as anomalous behavior and seasonality, among other situations. The geographical layouts aid the understanding of Brazil’s LA profiles, and provide navigation of localities, context, and comparison considering the locality’s particularities. Hierarchical layouts help to better characterize each locality in Brazil, in order to provide more specific and detailed information about them, and also to show how the lower hierarchical levels contribute to the superior levels’ profile.

The following subsections detail the chosen layouts and their contribution to the analysis strategy proposed by this research work. In the end of each subsection the resulting visualization is presented. The details of the combinations is discussed in the section 3.3.



### 3.2.1 Dendrogram Composition

The main idea of this visualization is to provide multi-level comparison. In this section we describe the layouts that compound the first visualization, the Dendrogram Composition, which is the dendrogram combined with bar charts and radial glyphs. We also describe and motivate the visual decisions, as well as necessary adaptations to the data.

#### 3.2.1.1 Dendrogram

A dendrogram is a type of node-link diagram containing one root node, internal nodes and leaf nodes. It differs from a tree because it displays all its leaf nodes aligned, which allows the analysis of information associated with the leaf nodes in different hierarchical levels.

Figure 15 shows an example of a dendrogram representing the geographic hierarchy of Brazil. The nodes are displayed by circles which are filled in white when they are expanded showing the subsequent hierarchy, or filled in blue when they are not. Interactions with a node can display/hide the subsequent hierarchy, in this way, leaf-nodes can show any hierarchical level. Thus, in this figure the branch with white filled nodes are displaying all hierarchy levels: country(root) - region(1) - state(2) - mesoregion(3) - microregion(4) - city(5), while the blue filled nodes are showing levels that can be expanded.

In our proposal, this layout was combined with summarization structures, that were added to the leaf nodes, in order to aid the analysis of different hierarchy levels and the comparison among them, as well as navigating through the localities, which is efficient to keep track of the path during the data exploration.

#### 3.2.1.2 Bar Charts

A bar chart is a quantitative graphic that represent different measured variables for comparison, encoding values in its height or length. The creation of this layout is assigned to Willian Playfer, in his *Commercial and Political Atlas*, from 1786(BENIGER; ROBYN, 1978). For this research work, we used bar charts combined to the dendrogram (see section 3.2.1.1), one bar chart associated with each leaf node. The bar charts encode a period of one year, each bar encoding the amount of LAs registered in a single day, as can be seen in Figure 16.

The scale of all the bar chart associated with the dendrogram's nodes are normalized by the scale of the root-node's bar chart, *Brasil*, which summarizes the total amount of LA in the entire country in a single selected year. The goal of the bar charts in our strategy is to highlight the distribution variation of LAs over a year for all localities and not to present precise data for analysis in place. Thus, the increase/decrease of LACs in different localities can be promptly noticed and compared. Figure 17 show three aligned bars charts of the distribution of LACs in different localities. In this figure, the size of

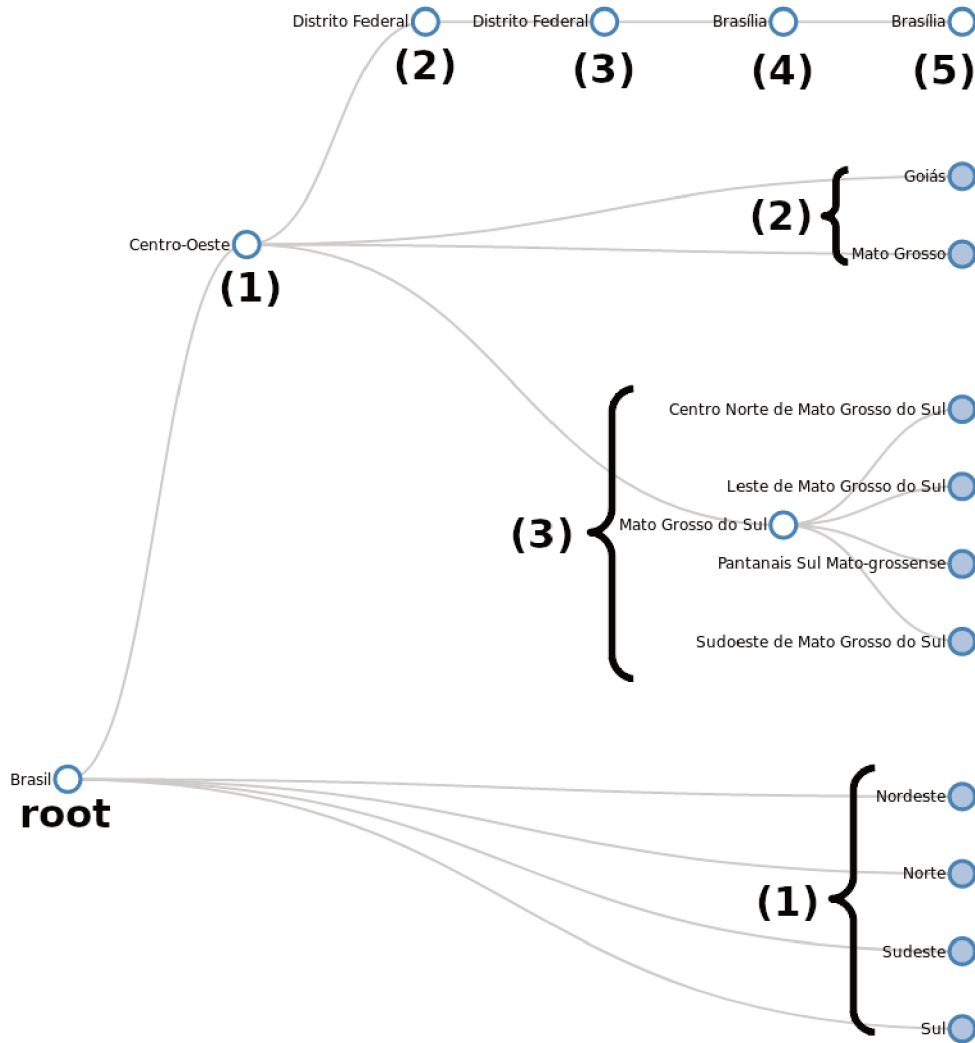


Figure 15: Dendrogram showing Brazil's geographical territorial division, in which the nodes represent (1) region, (2) state, (3) mesoregion, (4) microregion, and (5) city, all aligned.

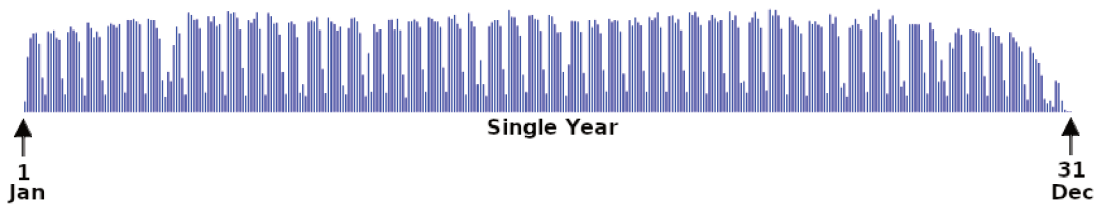


Figure 16: Bar chart showing the daily amount of LACs in the period of one year.

the bars does not differ too much but sequence of tall bars followed by short sequences of small bars can be seen in all the three bar charts, suggesting periodicity on the LA occurrence.

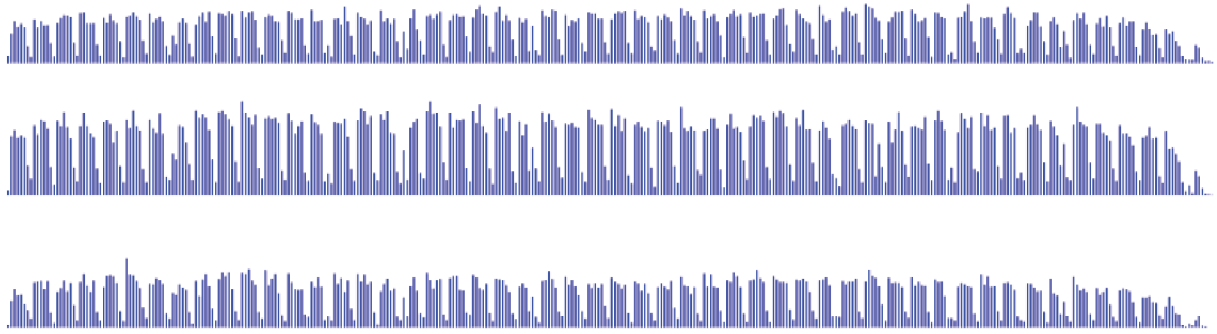


Figure 17: Three different localities' Bar charts showing the distributions of LACs in a single year. The alignment of bar charts favors behavior comparison among them.

### 3.2.1.3 Radial Glyph

A glyph is a compact visual representation of a piece of data, small and independent, in which the visual attributes of a graphical entity are dictated by one or more attributes of a data record (WARD, 2008; BORGO et al., 2013). They are a type of visual sign that differs from other types of signs such as icons, indices, and symbols, but that can make use of such visual features.

Our strategy uses a radial glyph, which is a set of equiangular hidden axes, each representing one attribute, and in each axis a point represents the specific attribute value associated with a data record (CHAMBERS et al., 1998). Then, a line intercepts all axes on each of these points, forming a geometric shape. In our case each axis refers to one year, from 2012 to 2017, encoding the amount of LAs occurred in each locality in these years. An example of a possible resulting hexagonal glyph can be seen in Figure 18.

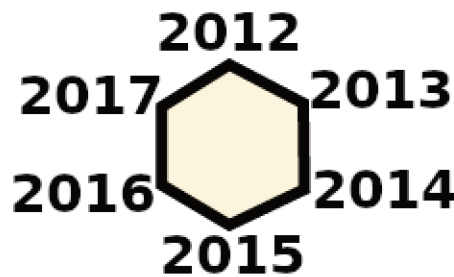


Figure 18: Clockwise arrangement of the glyph's edges considering the years from 2012 to 2017. The lines create a geometric shape.

With this glyph one identifies the years that registered more/less LAs, if LACs increased or decreased over the years, as well as meaningful differences in the occurrences between the years, considering a specific locality. We used this layout in the same way as the bar charts, associating it with each dendrogram's leaf nodes. The resulting layout allows quick identification of subtle variations in its shape over the years for a single locality and the comparison of the glyph's shapes of different localities, as can be seen in

Figure 19. A glyph has 6 axes, so homogeneous hexagons present well balanced values over the years. On the other hand, abnormal and/or behavior differences are promptly perceived by variations on the shapes of the glyphs.

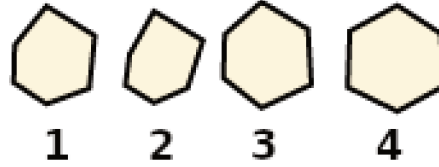


Figure 19: Comparison strategy using radial glyphs. Glyph 2 quickly catches the attention due to its non-homogeneous shape. Glyphs 1, 3 and 4 are more homogeneous.

The combination of the associated layouts that compound the dendrogram composition visualization can be seen in Figure20.

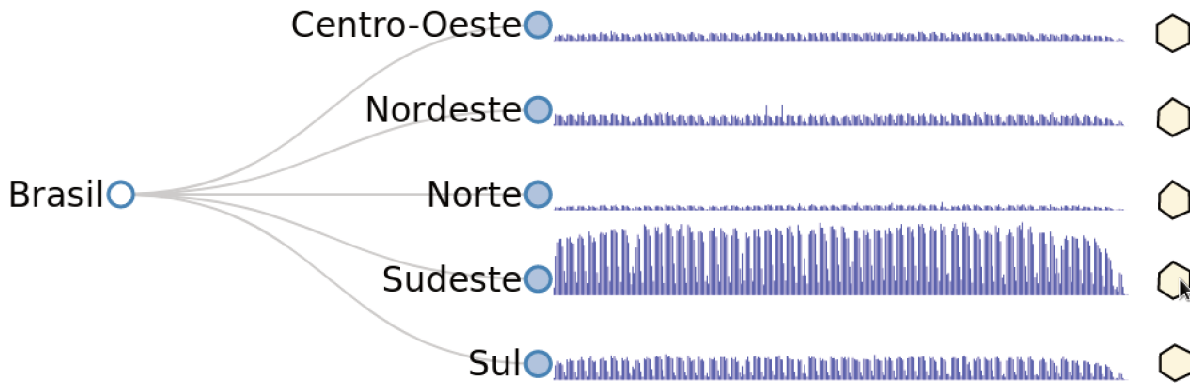


Figure 20: Dendrogram composition visualization: A dendrogram showing Brazil's hierarchy based on the territorial division, the bar charts of LACs in a single year and glyphs with an overview of the amount of LACs in the six years.

### 3.2.2 Choropleth Composition

In this section we describe the layouts that compound the Choropleth Composition visualization, which is a choropleth map combined with small multiples and heat maps. We describe and motivate the visual decisions, as well as necessary adaptations to the data.

#### 3.2.2.1 Choropleth Map

A choropleth map is a thematic map divided in areas, in which the areas are colored/shaded or filled according to a measurement, in a way to highlight an associated characteristic. This representation first appeared in the beginning of the 19th century

authored by Baron Pierre Charles Dupin, and might be the first statistical map in history (FRIENDLY, 2009). Figure 21 shows an example of Brazil's choropleth map divided in its states, encoding the proportion of LAs registered in a specific year. Choropleth maps are adequate to represent proportions and the readily recognizable map format offers geographical context to the information contained therein, which is useful when dealing with LACs, since each occurrence is associated with a specific locality. This visualization may improve the understanding of LA occurrences distribution within the country, specially when the distribution is not homogeneous. It also serves as an easy navigation tool for the localities.

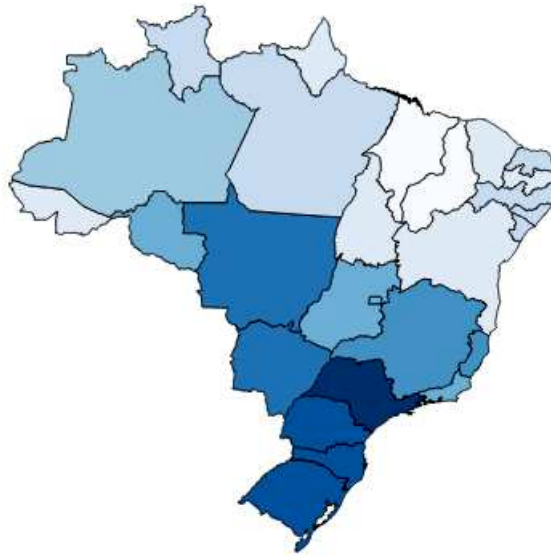


Figure 21: Choropleth map showing the proportion of LACs to the population size for each Brazil's state in the year 2012.

### 3.2.2.2 Small Multiples

Small multiples is a technique based on the repetition of charts that shows the combination of a set of variables indexed by another chosen variable, in which each chart shows a different perspective of a data set, allowing comparisons, identification of patterns, anomalies and evolution (TUFTE, 2001). Small multiples are normally used in temporal contexts (FUCHS et al., 2013) to compare evolution, because the time attribute of a data set is commonly used in the charts as a variable to index other attributes of the same data set.

Figure 22 shows an example of small multiples depicting maps and using time to index the percentage of out-of-wedlock births in France between 1990 and 2013. The different orange shades in the map indicate the increase/decrease of this percentages. In this figure it is possible to quickly identify the growth in out-of-wedlock births due to the orange shades becoming darker over the years.

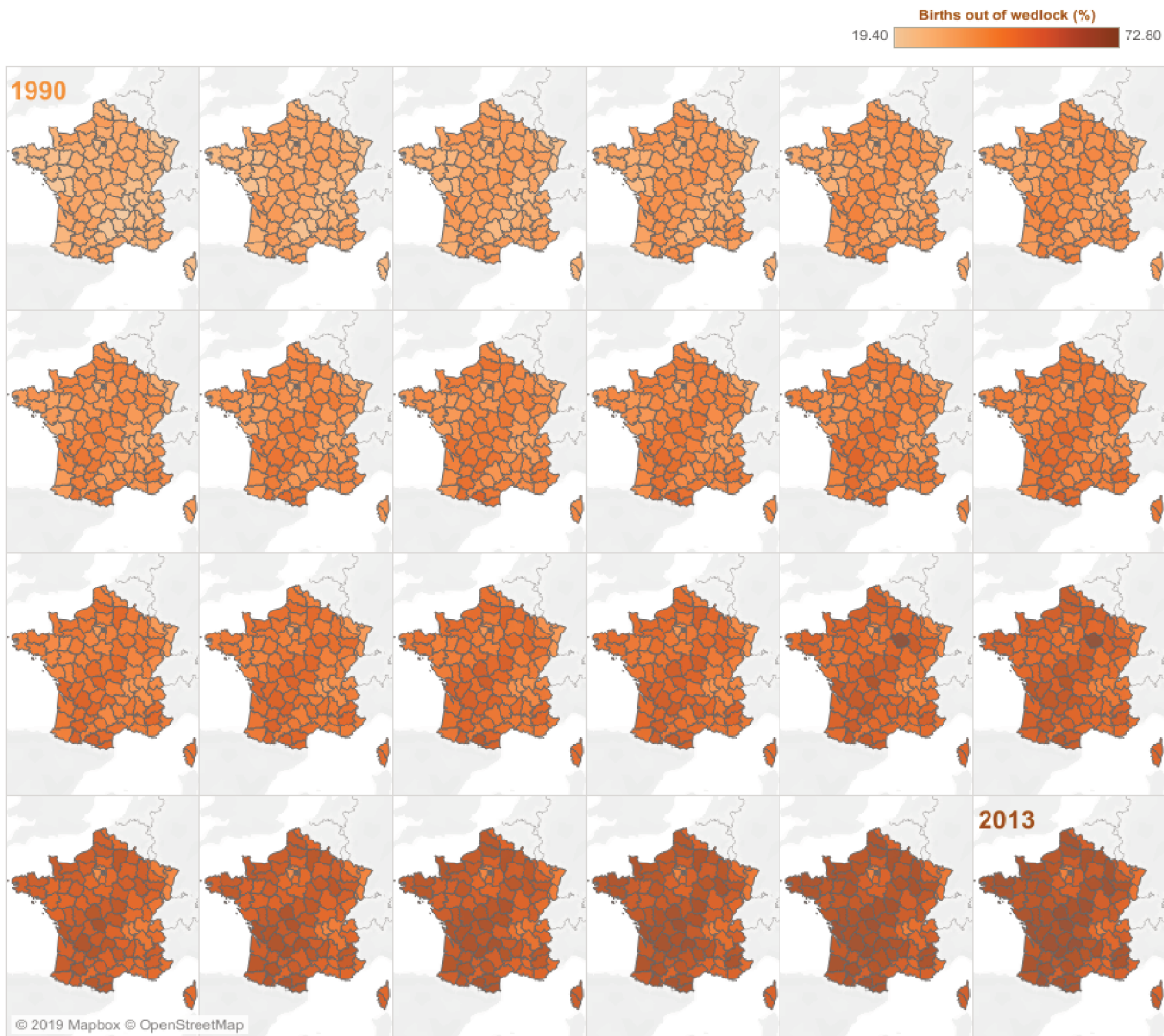


Figure 22: Small multiples showing the evolution of the percentage of out-of-wedlock births in France in a range of 14 years.

Source: <https://public.tableau.com>

In this research we are comparing the number of LACs in Brazil and our small multiples also uses maps. The divisions represent the states of the country and the color reflect the proportion of LACs to the population size in those areas. The main goal of this coloring scheme is to provide an overview of the evolution, and at the same to guide the user to find concentration areas, as the smallest modification in the distribution of the colors are quickly noticed. Basically, the idea is not to provide data details, but to provide quick perception and meaningful changes in multiple views of the same data.

### 3.2.2.3 Heat Map

Heat map is a visual representation in which an area is divided in regions and each region is colored/shaded according to a measured value. Heat maps can be applied to maps, matrices, geometric shapes, images, etc. When using a matrix the resulting layout

is then called heat map table, organizing variable values in columns and rows, creating a mosaic that highlights patterns. Figure 23 shows an example of a heat map table encoding a summarization of the number of LACs per day over a year, in which the blue shade is proportional to the associated value. The produced mosaic clearly shows more concentration of dark shades in the first/lasts columns, indicating a greater occurrence of LACs at the beginning and at the end of the year.

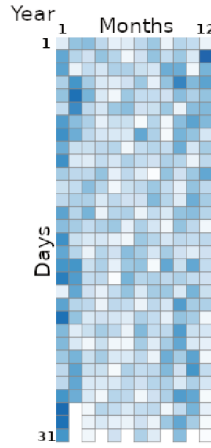


Figure 23: A heat map table showing the period of one year. The months are presented as columns and the days as rows. Each cell present a different blue shade according to the number of LACs in the corresponding day.

We propose a different organization for the heat map, named **Stripes**. In this layout the days are displayed side-by-side and the years are vertically arranged, forming a mosaic that allows direct comparison and understanding of the LACs evolution within a year and yearly. An example of this heat map is shown in Figure 24, in which the patterns of dark blue shades are horizontally concentrated in the beginning and in the end of the stripes, and vertically, the concentrations are in the upmost stripes.

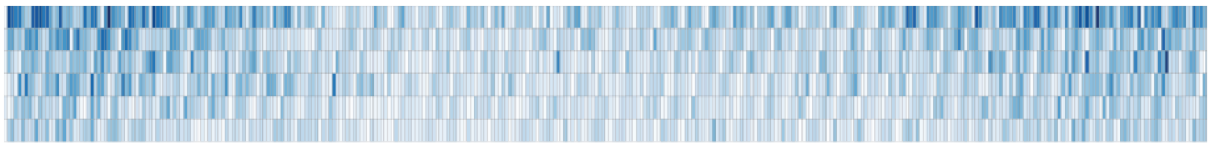


Figure 24: A heat map displayed as stripes, with days presented side-by-side and years, from 2012 to 2017, vertically arranged. The amount of LACs are presented in different blue shades, with low values associated to light shades.

The combination of the associated layouts that compound the choropleth composition visualization can be seen in Figure25.

### 3.2.3 Visualizations' Structures

Our proposed strategy comprises two visualizations, the Dendrogram Composition and the Choropleth Composition. The goal of the Dendrogram Composition visualization is to



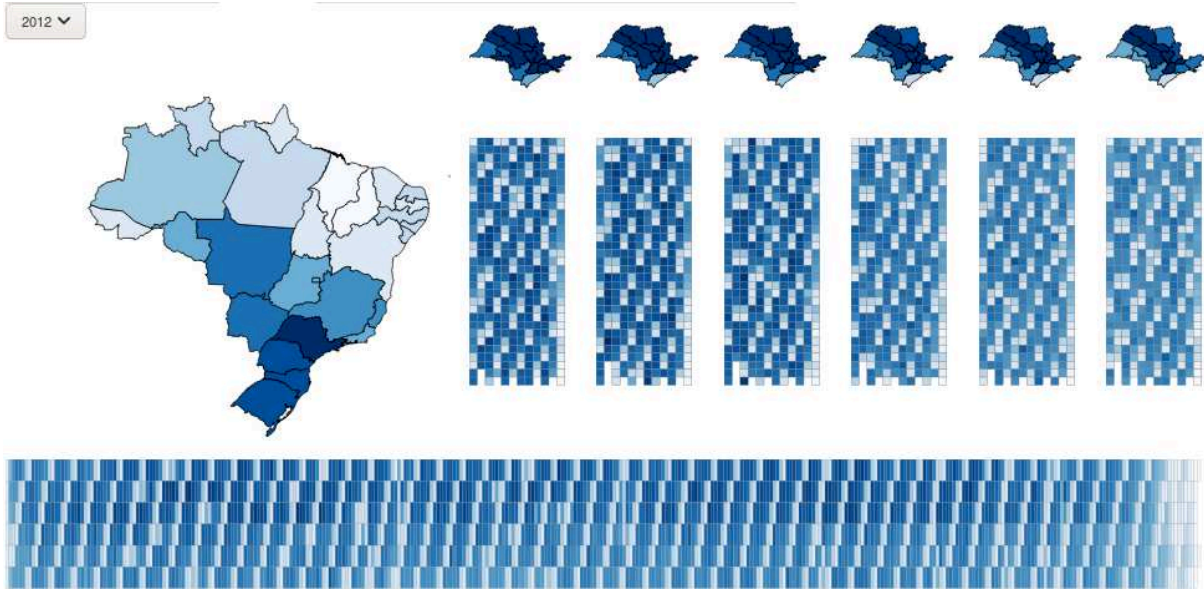


Figure 25: Choropleth composition visualization: A choropleth map presenting the states of Brazil, small multiples of choropleth maps presenting a specific state of the country and two different representations of heat maps.

provide multi-level comparisons of the evolution of LACs in Brazil during a single year. The goal of the Choropleth Composition visualization is to provide geographical context and comparison of LACs evolution within a specific chosen locality in a range of years. The strategy to develop both visualizations consists in the combination of largely employed INFOVIS techniques in order to take advantage of the different perspective they provide to potentialize information extraction. Figures 26 and 27 show the visualizations' diagrams of our strategy. The hard line boxes represent this visualization's main structures for each visualization. The dotted line boxes represent the INFOVIS techniques chosen to present the aspects of the defined structures. Both the proposed structures and the INFOVIS techniques were chosen to attend specific needs of the BFLPO's data set, but the strategy can be adapted to other types of data sets and problems that share similarities with the LACs data set and can be presented as our defined structure. The elements described in the dotted line boxes can be seen as modules that can be replaced for other INFOVIS techniques in order to better adapt the strategy to other specific problem's needs or even to improve the results for the LAs problem.

### 3.3 System Description

In order to evaluate our proposed strategy, we developed a system that implements it providing two visualizations to help the user to understand the data. The system allows the user to interact with the visualizations in order to get different perspectives of the data, navigating through Brazil's territorial divisions, comparing and analyzing the local-



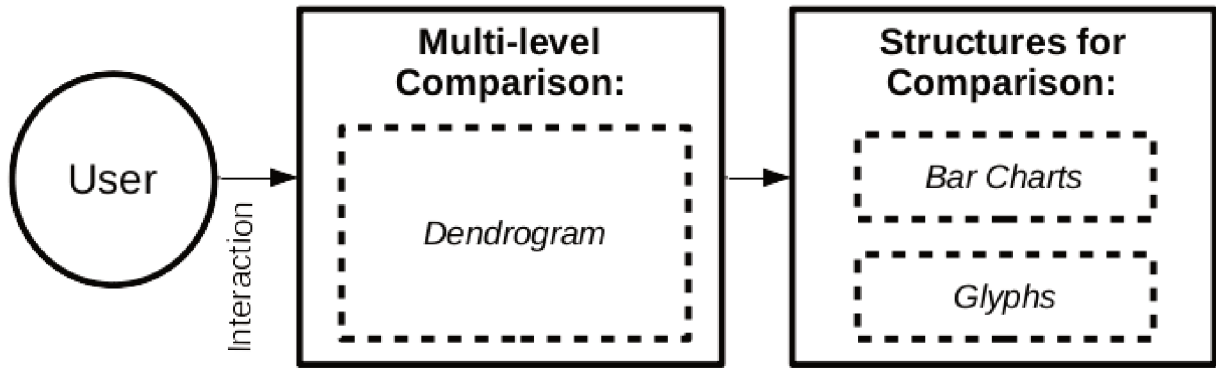


Figure 26: Diagram of the Dendrogram Composition visualization. The structure in this strategy is shown in hard line boxes, comprised of multi-level perspective structure and comparison structures for evolution analysis in a single chosen year. The elements in the dotted line boxes shows the INFOVIS techniques chosen to represent each structure. The user first interacts with the multi-level structure(dendrogram), and through it she is able to interact with the attached layouts.

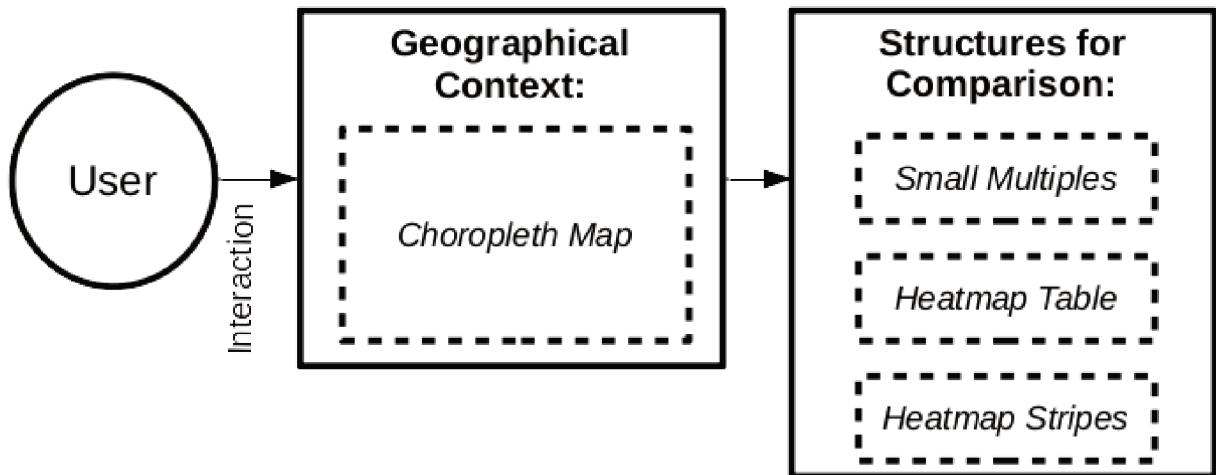


Figure 27: Diagram of the Choropleth Composition visualization. The structure in this strategy is shown in hard line boxes, comprised of geographical context structure and temporal comparison structures for evolution analysis. The user first interacts with the geographical context structure then she is able to interact with the attached layouts. The elements in the dotted line boxes shows the INFOVIS techniques chosen to represent the each structure.

ities over the years, among other tasks. In this section we describe the two visualizations that compose our system, the Dendrogram Composition and the Choropleth composition, also the interactions for data exploration and the tools used for its implementation, including programming languages and libraries, as well as how these tools were employed to construct the visualizations.

### 3.3.1 The Dendrogram Composition

The Dendrogram Composition is a combination of a dendrogram, bar charts and radial glyphs. The resulting visualization is presented in Figure 28. Using the dendrogram it is possible to navigate in Brazil’s geographical territorial division hierarchically. A click in any dendrogram’s node show/hide the next/previous level, and for all current leaf nodes shown, a bar chart and a glyph is associated.

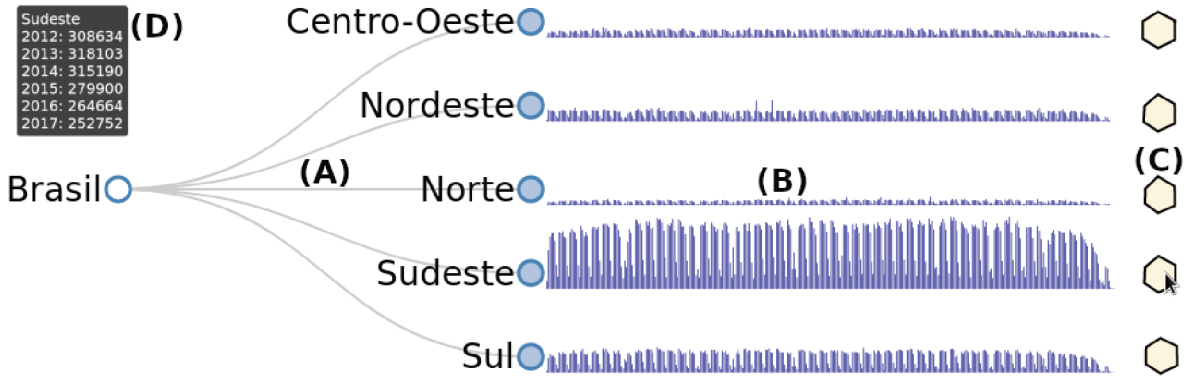


Figure 28: The resulting Dendrogram Composition visualization.(A) shows the dendrogram, showing the regions of Brazil in its first hierarchical level, the five regions. (B) shows the bar charts aligned representing data from a single year. (C) shows the glyphs aligned beside the bar charts. (D) shows a tooltip, which is the result of the interaction with a glyph that shows the raw values encoded in each axes.

The bar charts present the period of one year, with the amount of LACs for each day producing shapes corresponding to the evolution over the year. The height of the bars of a node’s bar chart correspond to the summarization of the amount of LACs of all its sub-levels in each respective day. Since all the bar charts follow the root node’s scale, *Brasil*, lower levels tend to be smaller. For this reason, we added interactions that help the users to scroll up/down over a node to change the considered scale for the correspondent bar chart, increasing/decreasing the bars height. A middle-click reset the scale to the original value. It is important to highlight that all the distribution’s proportion does not change. An example of this interaction is presented in Figure 29.



Figure 29: Scale changes over a node’s bar chart to highlight patterns. The two nodes refer to the same locality, (1) presents the normal scale and (2) presents the scale change.

The glyphs present the amount of LACs for six years, resulting in a geometric shape that guide the exploration to specific years. These glyphs use their own scale and thus

have the same size, with variations only in their shapes. Users can hover over the glyphs to show a tooltip with the amount of LACs in each year. Figure 28D presents the tooltip with the values associated with the *Sudeste* node.

The analysis of the bar charts and glyphs, guided by the dendrogram, help the user to identify the specific localities that present more or less LA registered, to compare the period of occurrences, to understand the distribution of LACs over a single year and over a range of years, to identify anomalies, to understand the behavior of specific localities and how lower levels in the hierarchy contributes to the patterns seen in the higher levels, among other tasks.

### 3.3.2 The Choropleth Composition

The Choropleth Composition is a combination of a choropleth map with small multiples presenting six choropleth maps, a set of six heat map tables and six heat map stripes. The resulting visualization can be see in Figure 30.

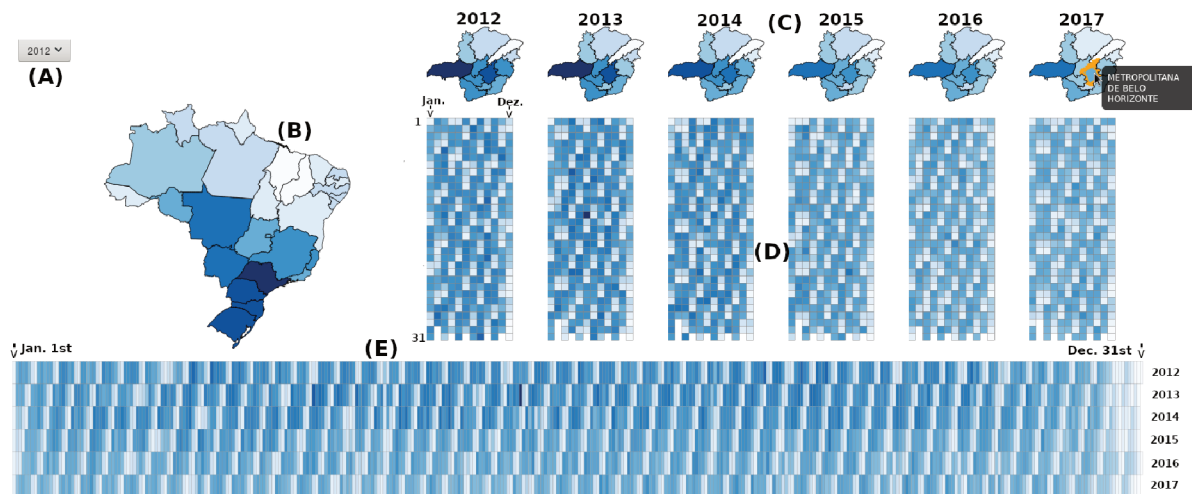


Figure 30: The resulting Choropleth Composition visualization. (A) a dropdown menu where a year, out of six, is selected to color the big choropleth map. (B) Brazil's choropleth map divided in states and colored according to the specified year proportion of LACs to population size. (C) small multiples of six choropleth maps of one selected state divided in its mesoregions also according to the proportion. (D) six heat map tables each forming individual mosaics. (E) six heat map stripes that together form a mosaic in for a different perspective.

A dropdown menu is used to select the year to be analyzed in the visualization. The bigger choropleth map shows the proportion of LACs to the population size for each Brazilian state in a chosen year. The blue shade is associated with the correspondent value, in which light/dark shades of blue are associated with low/high values, respectively. This map shows the distribution of LACs and also provide visual guidance for navigation.

By clicking in a state of interest, the correspondent small multiples and heat maps are shown.

The small multiples show six smaller maps of the selected state, one per each year. These maps are divided in mesoregions, providing information about how each of them contributes to the pattern observed in their state and in the in the bigger choropleth map. By hovering on each region in any map, users are able to see the associated locality's name, as presented in Figure 31.



Figure 31: Hover over the mesoregion Metropolitana de Belo Horizonte in the state of Minas Gerais.

We also use six heat map tables and stripes, one for each year, considering the amount of LACs for each day. The heat map tables present the months in its columns and the days in its rows, and the tables are shown side-by-side (see Figure 30D). The heat map stripes present the days of an entire year side-by-side, and years vertically arranged (see Figure 30E). The cells in both the heat map table and heat map stripes are shaded according to the amount of LACs per day, where light/dark shades of blue are associated with low/high values, respectively. The shades produce mosaics in which patterns can be quickly perceived. Hovering on a cell shows the exact value encoded.

Analysis of the heat map tables provide comparisons within a year, for example to verify in which month period the LACs are more frequent, and their behavior in different months. The overall comparison of the tables can also provide quick understanding of the evolution of LACs. The analysis of the heat map stripes aids the direct comparison of days for all the six years in analysis, and also helps to understand the evolution of the LACs, considering the period of occurrences.

The analysis of specific regions on the choropleth maps helps the identification of localities with proportionally more/less LACs and can aid the understanding of the reasons for such behavior. The analyses of the heat maps tables and stripes help the user to understand the distribution of LACs over the year by the shade's distribution in the maps.

### 3.3.3 Implementation Details

We developed a web system for a temporal visual analysis of the data provided by the BFLPO repository using a set of libraries and toolkits described in this section. Figure 32 shows the diagram of the system components as well as the communication among them.

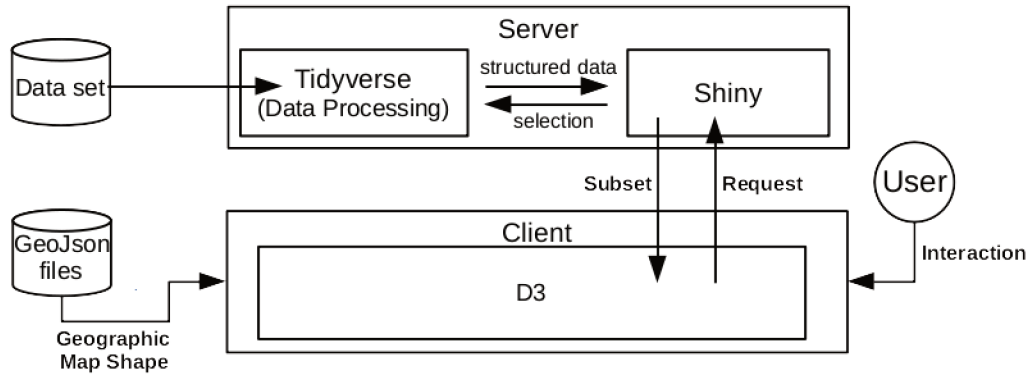


Figure 32: System overview.

R(R Core Team, 2018) is an important tool in the development of the visual web system for LA data analysis. It is a statistical and programming language suitable to be used in massive data because it provides many libraries to help the users to make sense of their data. R is adequate to be used for computational processes using the large BFLPO data set. Two important R libraries/packages used are Tidyverse(WICKHAM, 2017) and Shiny(CHANG et al., 2017).

Tidyverse comprehends a variety of packages designed specifically for data science, providing operations for data transformation, filtering, summarization, and organization. We used it for LA data pre-processing and preparation, in order to produce the appropriate data format used in each visualization, including saving the resulting data in R format (.rds) to speed up the reading of them by the system.

The web system was developed using the Shiny package, both the front-end, the user interface, and the back-end, the server. Shiny is a package to build interactive applications from R and it is responsible for the communication and data exchange between server and user interface. In the user interface, it is responsible for the design of the web page and also for the addition of HTML, CSS and JavaScript elements, among other coding languages, to be integrated within R. The server is responsible, among other functions, to send the data computed in R to the functions or scripts that will use these data and to keep the communication with such scripts.

JavaScript is a portable language which requires only a web browser to execute the associated functionalities and the D3(BOSTOCK; OGIEVETSKY; HEER, 2011) is a JavaScript library that provides a set of functions to aid the development of interactive information visualization charts. D3 can read a data set or receive it from other integrated applications, and then handle this data by binding it to Document Object Model(DOM) to apply data-driven transformations. For this research work, D3 scripts were developed and integrated with R, in a way that the user interface in R read and load the scripts, and the server sends the specific data set to each script to generate the interactive visualizations. GeoJSON(LUIZ, 2014) files with Brazil's geographic information<sup>4</sup> were used in D3 scripts

<sup>4</sup> <https://github.com/fititnt/gis-dataset-brasil>

for map generation in the Choropleth Composition.

## Analysis Results

This chapter presents the results of applying the proposed system in a temporal visual analysis of labor accident data. The BFLPO (Smartlab de Trabalho Decente MPT - OIT, 2017) provides a repository with LA data, daily collected from 2012 to 2017, with more than 3 million anonymized LACs. The Geographical hierarchy used in the analysis is the official Brazilian territorial division (see Table 1). The analysis procedure, interesting patterns and events identified in each visualization are also discussed. We present the analysis using the visualizations individually, as well as in combination, in order to explore their complementary capabilities. To understand and explain the patterns, trends and/or anomalies found, we used the complete original data set provided by the BFLPO, governmental websites and official external sources, such as national and local news.

### 4.1 Analysis Procedure: Strategy's Evaluation

This section details the analysis procedure adopted to evaluate our strategy, which was performed by visually analyzing/exploring the visualizations, identifying meaningful patterns/situations, and exploring/explaining the possible reasons for the occurrences. We combined the BFLPO's and IBGE's data sets and used subsets of this data, as described in section 3.1, considering the attributes geography, date and proportion of LACs to population size, focusing on the main important aspect to this research work, the temporal information, to generate the system's visualizations.

The exploration process using the Dendrogram Composition can be performed in a free way exploring elements that quickly catch the user's attention. All the analysis we discuss in this section started by first identifying global trends and patterns in the bar charts and glyphs, then we proceeded going further in the hierarchy and repeating carefully this analysis for all descendant levels. In deeper hierarchical levels, the bars tend to diminish their size, which are occasioned by the distribution of the amounts of LACs of a level in its sub-levels, so the rescaling interaction is often necessary.

The exploration process in the Choropleth Composition is completely based on the

blue shades. A first step in the adopted analysis procedure for this visualization was to determine the year to be analyzed, and then analyze Brazil's map according to the blue shades of its states, which measures a proportion of LACs to population size, in order to find macro patterns. The second step was to analyze the states whose blue shade most deviate to the others, which naturally catches the attention. When a state was chosen, we analyzed the small multiples associated to the state in order to understand the distributions of the LACs proportion over the state's areas, and the heat maps in order to understand the distributions over the years.

## 4.2 Dendrogram Composition

Figure 33 shows the bar charts of the number of LACs in all five Brazilian geographic regions in 2012. A first overall analysis shows that the Sudeste region presents the highest number of accidents, followed by the Sul region. Norte region presents the lowest rates for this year. A significant difference among regions can be clearly perceived at a first glance. However, the associated glyphs are very homogeneous, as can be seen in Figure 34, which means that they keep a very similar hexagonal shape, still presenting a reduction in number of LACs over time in all regions. These macro patterns, such as periodicity, proportion, and size in the bar charts, and glyphs' shapes and proportions, were first verified in 2012, but can be seen occurring throughout the years.

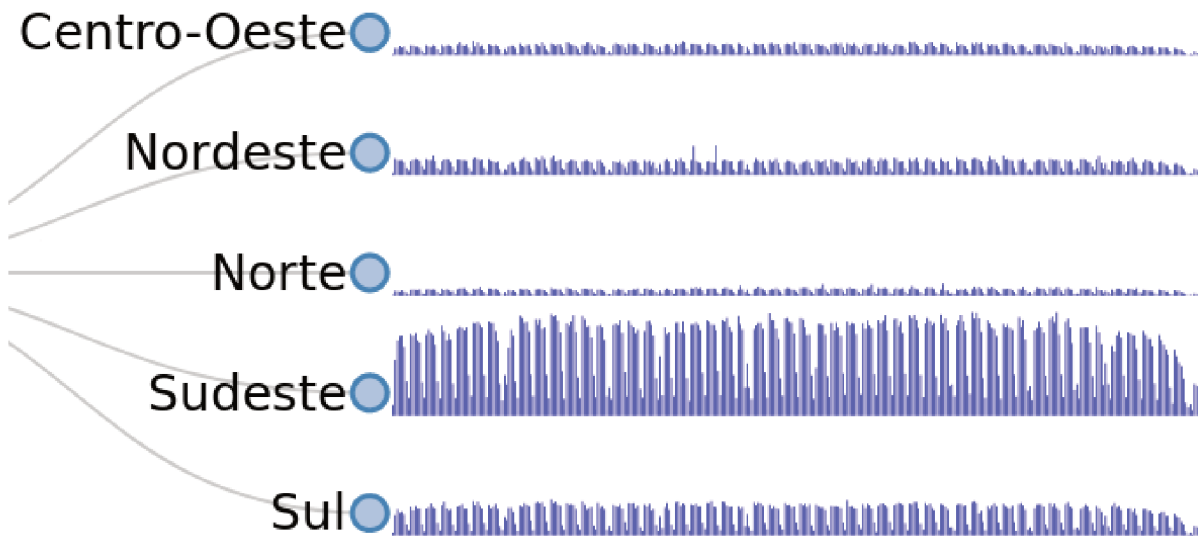


Figure 33: Bar charts of the five Brazilian geographic regions.

It is possible to investigate the causes for a node's pattern occurrence by expanding it. This expansion will reveal how a pattern is distributed further down in the hierarchy. This analysis may reveal scenarios in which an event comes solely from a specific location further down in the hierarchy, representing a localized event, or scenarios in which the



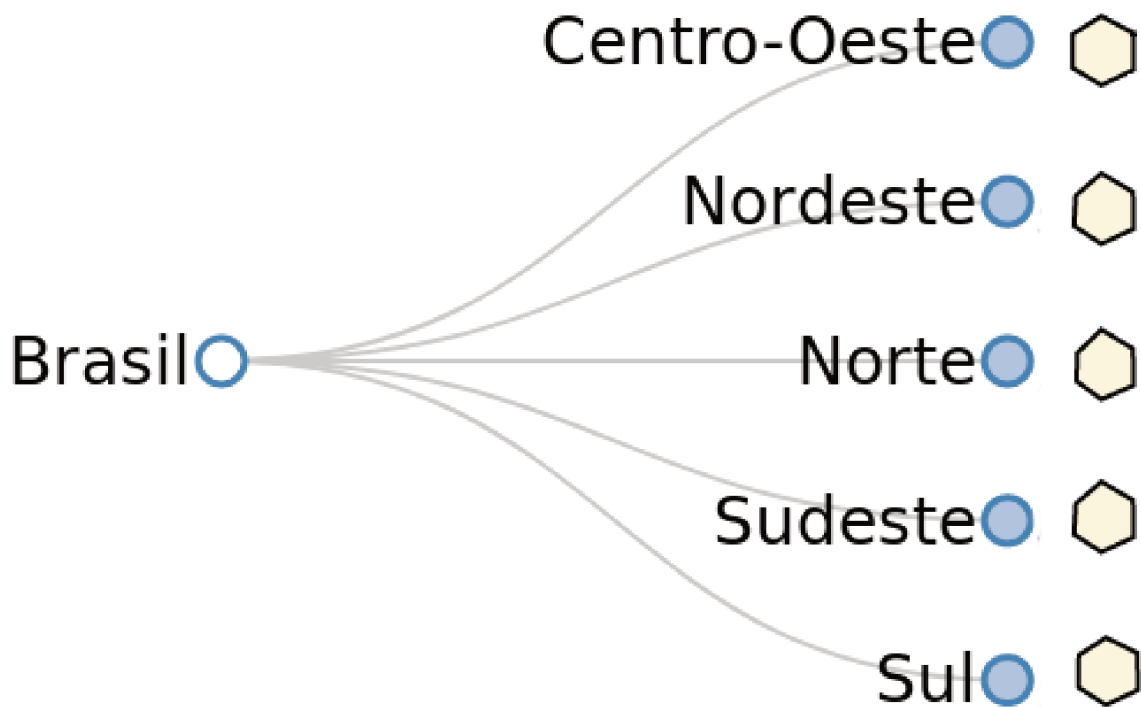


Figure 34: Dendrogram regions' glyphs. Higher hierarchical levels present glyphs with similar shapes, close to a perfect hexagon, since they summarize the amount of LACs of all its sub-levels. This homogeneity makes it harder to get meaningful analysis about their LACs' distribution.

event occurs in a distributed way among the sub-level nodes, representing a regional phenomenon. This summarization is also present in the glyphs. Examples of these phenomena are described below.

When analyzing the Nordeste's node in the year 2012, it is possible to notice that this region's bar chart presents two peaks, highlighted in Figure 35. When expanding this bar chart, in Figure 36, one notice that the peaks are associated with Bahia state.



Figure 35: Nordeste's bar chart distribution in 2012, showing two high peaks. The first relates to May 15th, and the second to May 25th.



Figure 36: Bahia state's bar chart distribution, with associated glyph, showing that the two high peaks observed in Nordeste, in Figure 35, are associated to this state.

A further expansion reveals that these peaks are associated to two distinct mesoregions, Metropolitana de Salvador and Sul Baiano, shown in Figure 37. The first peak is associated to the city of Salvador, in Metropolitana de Salvador mesoregion, and occurred in May 15th, in which 159 employees from civil construction had an accident related to food ingestion. It represents a localized event pattern that does not strongly influence the glyph shape, because Salvador presents a high accident rate over the year. A smaller peak could also be noticed in January 19th, when again 56 civil construction employees were also involved in an accident related to food ingestion. The visualization promptly revealed a situation that deserves a deeper investigation, in this case a single city registering high number of LACs in specific days. The visualization can guide the analysts to directly concentrate the analysis in strategic information. Thus, these events' investigations can be done by checking, in the BFLPO's complete data set, the LACs' information related to the date and localities mentioned above, which can reveal if the accidents occurred in the same company, or in similar working conditions, among other aspects. This analysis, as well as its findings, is not straightforward to perform otherwise.

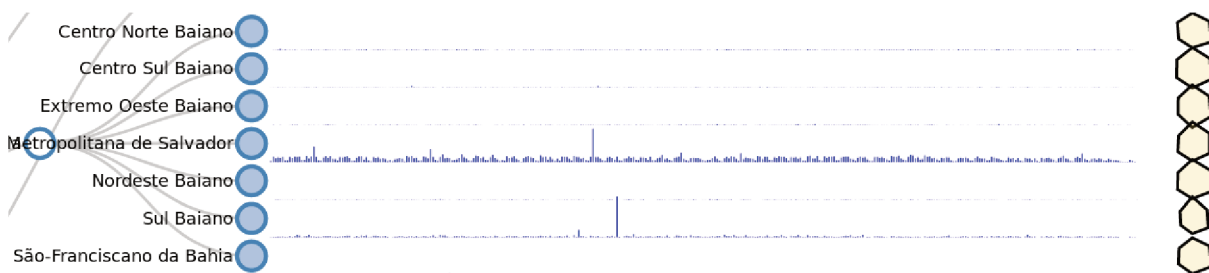


Figure 37: Expansion of Bahia state's node showing its mesoregions' bar charts. The two high peaks observed in Bahia come from Metropolitana de Salvador mesoregion and Sul Baiano mesoregion.

The second peak is associated to the city of Itamarajú, from the mesoregion Sul Baiano, also seen in Figure 37, which registered 243 accidents in 2012, 208 only in May 25th, possibly due to a serious traffic accident, as reported by a local newspaper<sup>1</sup>. The distribution of LACs in the city of Itamarajú produces a localized event pattern that is reflected in the shape of the mesoregion's glyph, with a sharp point for 2012, shown in Figure 38.



Figure 38: The glyph of Itamarajú revealing a high peak in 2012.

<sup>1</sup> <https://teixeirabahianews.blogspot.com/2012/05/tragedia-entre-itamaraju-e-teixeira-de.html>

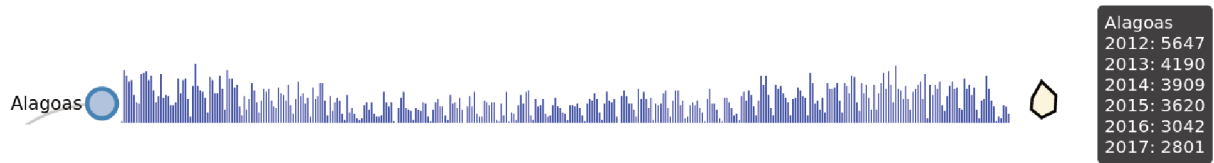


Figure 39: Distribution of LACs in Alagoas state in 2012 showing a seasonal pattern with higher bars in the beginning and end of the chart.

The analysis of the glyphs' shapes also reveal interesting scenarios. Figure 39 shows the scaled bar chart and the glyph associated with Alagoas state. The glyph presents one sharp point, indicating 2012 with the higher number of LACs registered. Also in the glyph, it is possible to notice a gradual reduction of occurrences, detailed on the tooltip. This reduction was around 50.4% from 2012 to 2017, whereas the national average reduction was 15%. When going downward in the hierarchy, it can be noticed that the shape of the Alagoas' glyph is primarily determined by a mesoregion named Leste Alagoano (see Figure 40), which presents a very similar shape.



Figure 40: Bar chart of Leste Alagoano. This mesoregion presents the pattern observed in Alagoas state more pronounced than the other mesoregions in the state. Its glyph also indicates reduction of LACs over the years.

Alagoas state's bar chart was also investigated for all years, and the result is shown in Figure 41.

An interesting seasonal pattern can be noticed, in which high rates are observed during the beginning and end of the year, and repeated in every subsequent year, being strongly determined by the Leste Alagoano mesoregion previously mentioned. As indicated by the glyph, the pattern of the state is distributed among the microregions, suggesting a regional event. Figure 42 shows the patterns of Alagoas state's microregions in their bar charts.

This behavior was noticed in no other states, only in a mesoregion named Mata Pernambucana, located in Pernambuco state. The dendrogram's leaf nodes organization facilitates the comparison between localities from different levels, enhancing the analysis process. Figure 43 shows the comparison between Leste Alagoano and Mata Pernambucana mesoregions, which are geographically adjacent. By deeper exploring the Mata Pernambucana mesoregion, one can notice that the pattern is mostly determined by a microregion named Mata Meridional Pernambucana, which is geographically adjacent to Alagoas state. This geographical proximity may explain the similarity of the patterns observed in these two localities.

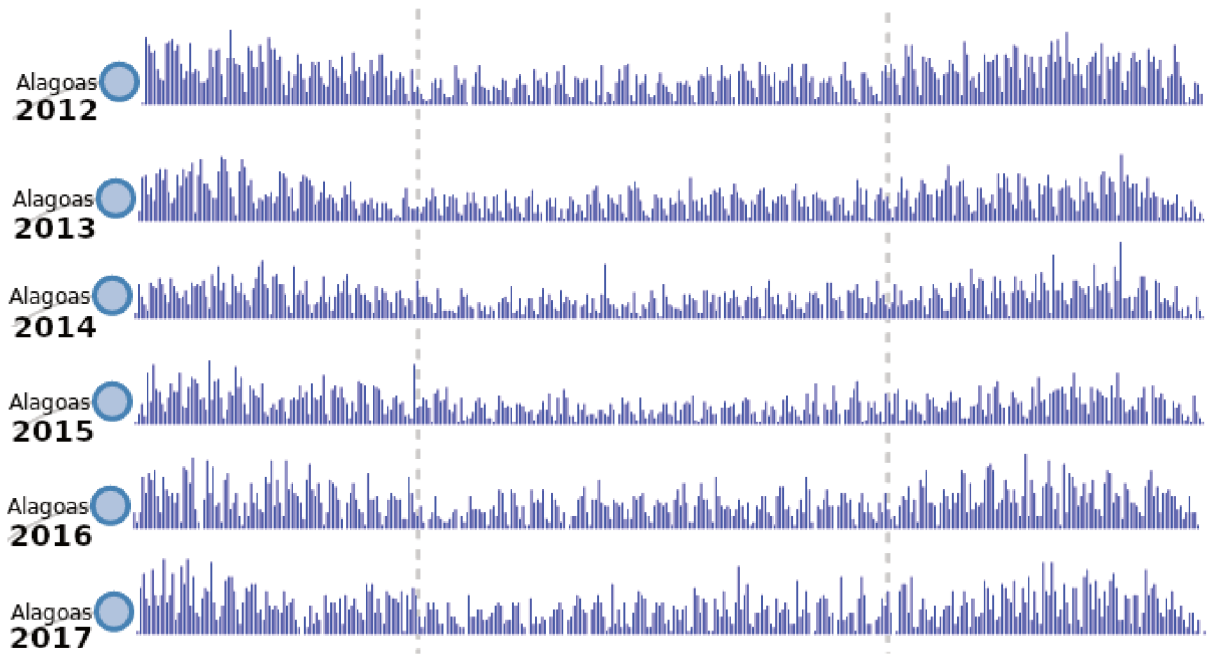


Figure 41: Alagoas state's bar charts from 2012 to 2017. An increasing in the amount of LACs in the begin/end of each year can be noticed.

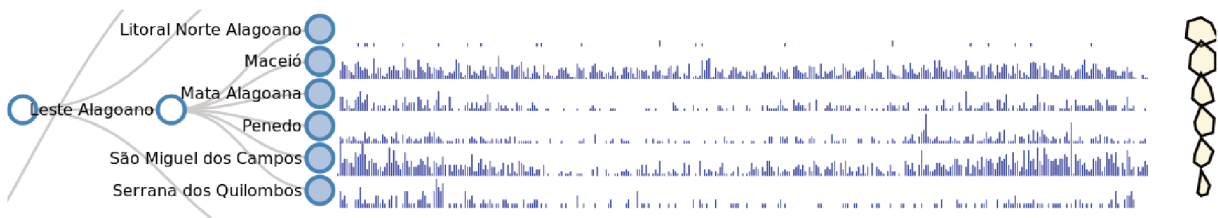


Figure 42: Leste Alagoano's microregions analysis, with bar charts and associated glyphs. At this level, it represents a regional event, since the behavior is observed in all these microregions.

When analyzing the characteristics of these localities, guided by the information in BFLPO's complete data set, we verified that they strongly participate on the sugar cane cultivation/processing, and most of the LACs occurred on raw sugar production activity. The bar charts suggest an association between the accidents and the sugar cane culture cycle. The number of accidents associated to the cultivation period are high only in Pernambuco state. In the same period the LACs in Alagoas are associated with sugar cane processing. The behavior in both localities are similar regarding raw sugar production.

The further one goes down in the hierarchy, the more distinguishable the glyphs tend to be, because while the bars tend to diminish in size, the glyphs keep the proportion throughout the levels and their shape tend to be more discernible. A specific situation highlighted by the glyphs is associated to a microregion named São Miguel do Araguaia, located in Goiás state, Centro-Oeste region. It presented an inverse behavior when compared to the rest of the same state and the entire country over the years, characterized by an overall growth in the number of LACs. Figure 44 presents the glyph of this microregion

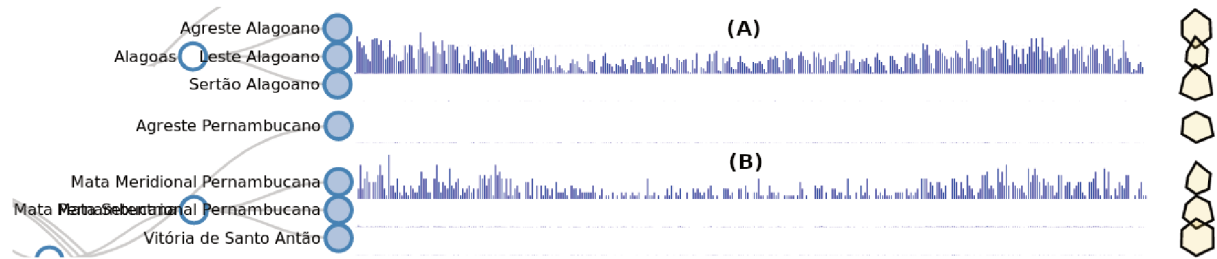


Figure 43: Comparative Analysis of bar charts from Alagoas' mesoregions and Mata Pernambucana's microregions. Similar patterns could be observed in two different hierarchical levels: (A) in a mesoregion of Alagoas named Leste Alagoano, and (B) in a microregion of Pernambuco named Mata Meridional Pernambucana.

and the associated tooltip with LA amounts showing a significant increase throughout the years.

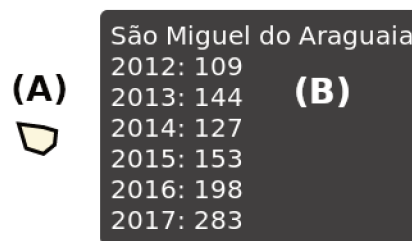


Figure 44: Analysis of São Miguel do Araguaia's associated glyph (A), and the respective absolute values (B). This microregion presented a growth in the number of registered LACs.

A deeper investigation in the data set revealed that the accidents were related to the two largest economic activities on this area, gold mining in the city of Crixás, and animal slaughtering in the other cities of the microregion, which may suggest an improvement in the inspections of work conditions, and so better report of LAs, but more probable, it may be suggesting an overall decline in working conditions associated to these activities. It is an example that illustrates how the visualizations may guide the authorities in the investigation of uncommon situations regarding LACs occurrence patterns.

Another interesting pattern was observed in a glyph of the same region, but in the Mato Grosso state, microregion of Norte Araguaia, city of Vila Rica. Figure 45 presents the glyph and the tooltip associated with this city, revealing that the number of accidents occurred in 2013 was five times larger than in the other years.

A number of 77 LACs out of 108, registered during 2013, are related to pecuary activity, which is the main economic activity of this locality. One can expect, as a normal behavior, such volume appearing throughout all the years, but this specific period may evoke a deeper investigation by the authorities on the causes of this odd situation. A possible explanation may suggest that the work conditions are not very well inspected in

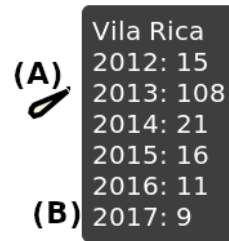


Figure 45: (A) shows the glyph associated with the city of Vila Rica and (B) shows its associated values. The number of LACs registered in 2013 is significantly bigger than in the other years.

this locality, and just for a single year, 2013, this inspections were more rigid.

The visualization also helps to compare geographically distant localities, in order to investigate similar behaviors, common characteristics or to show contrasts. The city of Ananindeua is located in the Norte region of Brazil, in Pará state, and a city named Paranaguá is located in the state of Paraná, in Sul region. These two far apart cities are not directly related, but the visualization allows their comparison. Guided by their similar glyphs, when analyzing their tooltips, for example, and their official populational estimation<sup>2 3</sup> we could see that Ananindeua has a population size about three times bigger than Paranaguá, but registered about 35% less LACs. Figure 46 shows their associated glyphs and tooltips with the amounts of LACs for each year.

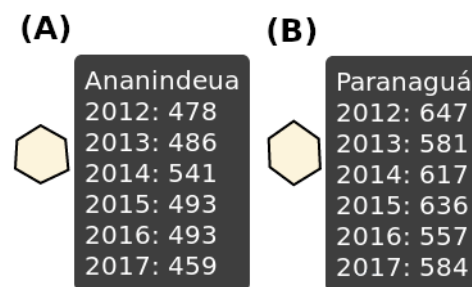


Figure 46: Comparative analysis of Ananindeua (A) and Paranaguá (B) glyphs. Ananindeua is three times bigger in population size than Paranaguá, but in Paranaguá the amount of registered LACs are 35% bigger.

Small cities reporting a considerable number of LACs or big cities reporting few are situations that deserve attention and a deeper investigation, since LAs may not being properly reported. Tooltips may aid situations in which visible pattern get harder to be perceived, providing useful summarizations.

<sup>2</sup> <https://cidades.ibge.gov.br/brasil/pa/ananindeua/panorama>

<sup>3</sup> <https://cidades.ibge.gov.br/brasil/pr/paranagua/panorama>



## 4.3 Choropleth Composition

The first view is Brazil's map colored according to the year 2012, which is the default option of the system. Figure 47, presents the maps for all the years under analysis. One can easily notice that the shades of the states in the south region are darker than in the north part of the maps. It is also possible to see a reduction in the proportion of LACs over the years due to the shades that became lighter, but keeping the pattern in all states.

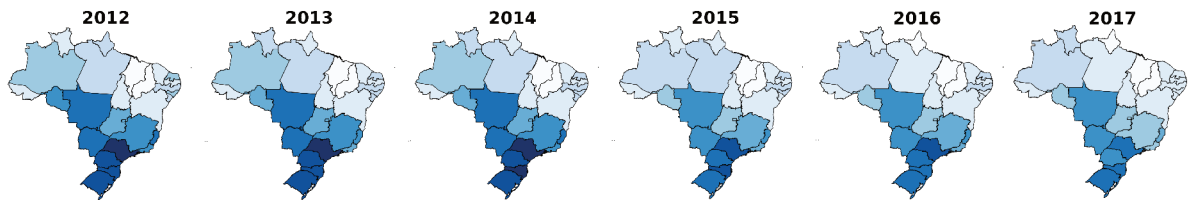


Figure 47: Brazil's choropleth maps, showing the variation of blue shades from 2012 to 2017. The predominant pattern presents the states in the north of the map in light blue shades, while the states in the south are in darker blue shades.

Figure 47 also shows that the state of São Paulo clearly presents the highest proportion of LAs registered, represented by the darker shades in the maps over the years. A detailed analysis of this state is shown in Figure 48.

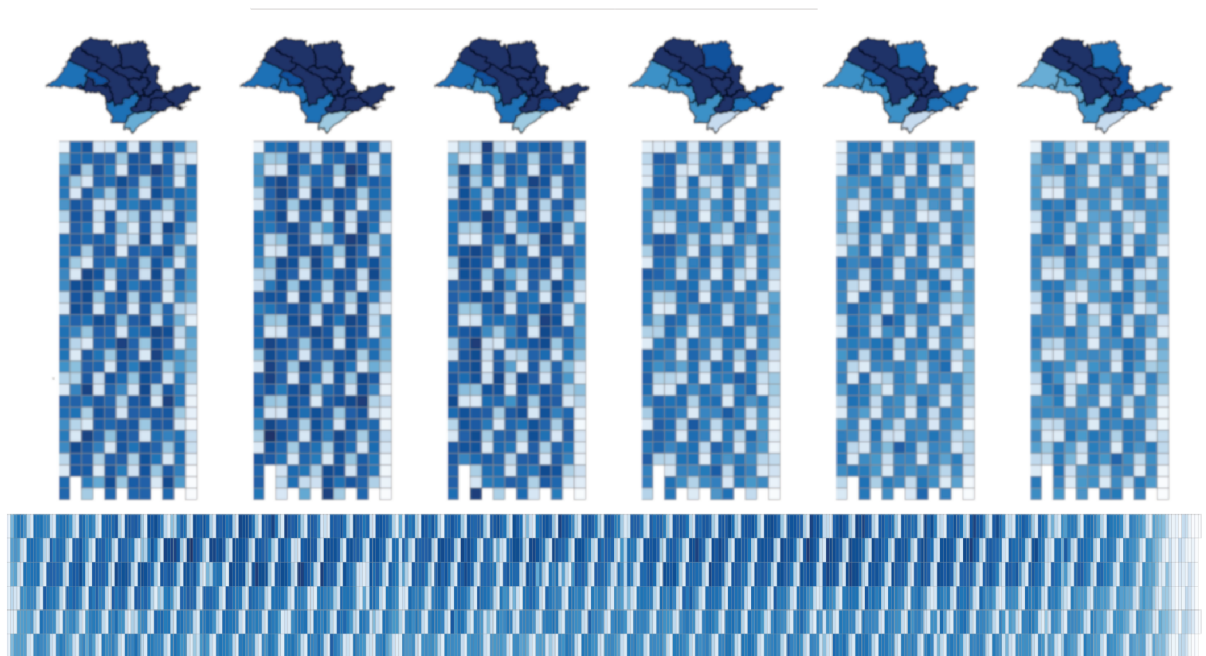


Figure 48: Small multiples and heatmaps detailing the distribution of LACs in São Paulo state from 2012 to 2017, revealing a very homogeneous behavior over the whole range of years.

The associated choropleth small multiples layouts show that almost all the mesoregions present dark hue values, becoming slightly lighter in subsequent years, suggesting that better safety measures may have been taken. The heatmap stripes illustrate the accident

occurrence behavior over the years, indicating a decrease in the number of LAs in the end of December. This behavior may be associated to the fact that in this period employees usually are on vacations, or in an break between Christmas and New Year. The accidents are well distributed over the remaining periods of the year in this state, suggesting an homogeneous behavior.

The same analysis were performed for Minas Gerais state, as shown in Figure 49.

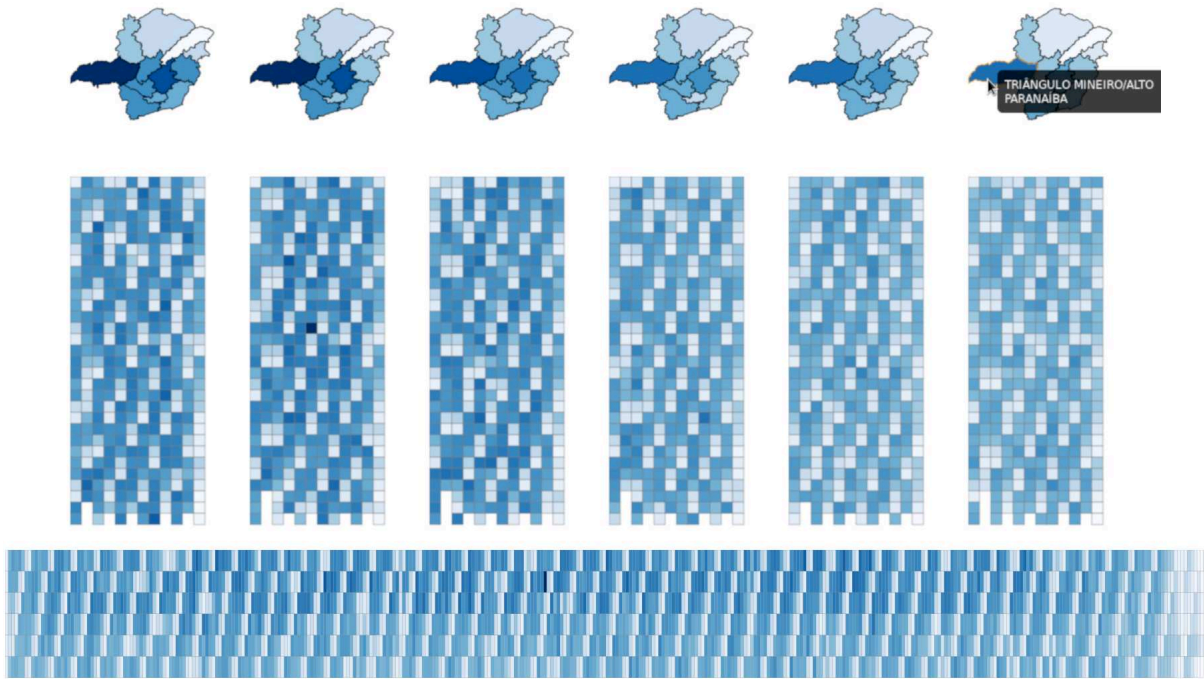


Figure 49: Small multiples and heatmaps detailing the distribution of LACs in Minas Gerais state from 2012 to 2017. The second most populous state in Brazil is not as homogeneous as São Paulo.

Although Minas Gerais is the second most populous Brazilian state (São Paulo is the first), the layout presents significantly lighter shades than São Paulo for all the years. There is also no homogeneity in its associated mesoregions, which present less accidents than São Paulo, except for a mesoregion named Triângulo Mineiro/Alto Paranaíba.

It is expected, for each state, that mesoregions in which the correspondent capital city is located to present the largest proportion of LACs, as they concentrate more companies and thus more employees. The visualization shows two mesoregions that concentrate higher number of LACs. However, the mesoregion named Metropolitana de Belo Horizonte, which contains the capital city of Minas Gerais state, Belo Horizonte, presents lighter shades than the Triângulo Mineiro/Alto Paranaíba mesoregion, as shown in Figure 50. Maybe this behavior is related to the fact that the largest city in this mesoregion, Uberlândia, is a logistic center for the entire southern half of Brazil. This odd situation could be quickly addressed by the visualization using the small multiples maps, that clearly pointed a scenario that was not expected as a normal behavior.



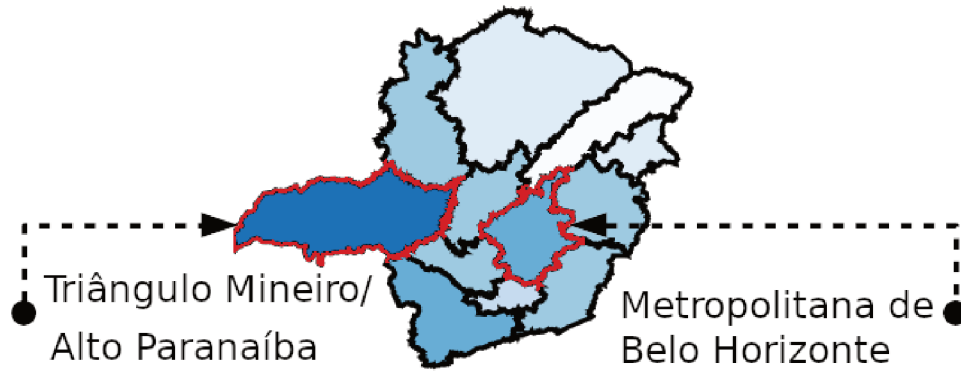


Figure 50: Mesoregions of Minas Gerais state. Triângulo Mineiro/Alto Paranaíba mesoregion presents a darker shade of blue compared to Metropolitana de Belo Horizonte, in which the capital city of the state is located.

The heatmap stripes also reveal another interesting pattern observed in São Paulo and Minas Gerais states, which is the difference in the number of accidents when considering workdays and weekends, reflected by the shades difference between these periods, forming diagonal alternate lines. On the other hand, several states from Norte and Nordeste regions present a more homogeneous pattern, presenting lighter shades. An example can be seen in Figure 51 which shows Amazonas state, from Norte region. If one pays closer attention to the heatmaps, it is possible to notice that the workdays/weekends patterns are still there, but are harder to see due to the light shades. The absence of clear patterns may suggest a low number of LACs registered in Norte and Nordeste regions. It is known that these regions are the poorer regions in Brazil, and present many problems related to decent work conditions and informal employment. Thus, in this case these patterns also suggest a LACs' under-reporting issue, and the the layouts are useful in revealing situations that deserve a better attention by the government, regarding LA inspection and reporting.

Amazonas state present a curious case. In this locality half of the cities are concentrated in one mesoregion, named Centro Amazonense, which is the only mesoregion whose shades change more significantly over the years in the small multiples maps, as can be seen in Figure 51. In 2016 and 2017 in this mesoregion, the blue shade became lighter, while in the other mesoregions, one can see some light blue shades, suggesting an increasing homogeneity in the state as a whole in the recent years. Figure 48 depicts the values related to the Amazonas' glyph in the Dendrograms Composition. In the glyph of Amazonas state we can see that the number of LACs reduced over the years. It may indicate an improvement in the inspection of work conditions, because in this situation there was a reduction in the number of LACs, while the proportion in the state as a whole became more homogeneous. The combination of Amazonas' behavior analysis in both visualizations is an example of the how their combination can yield more understanding over specific scenarios.

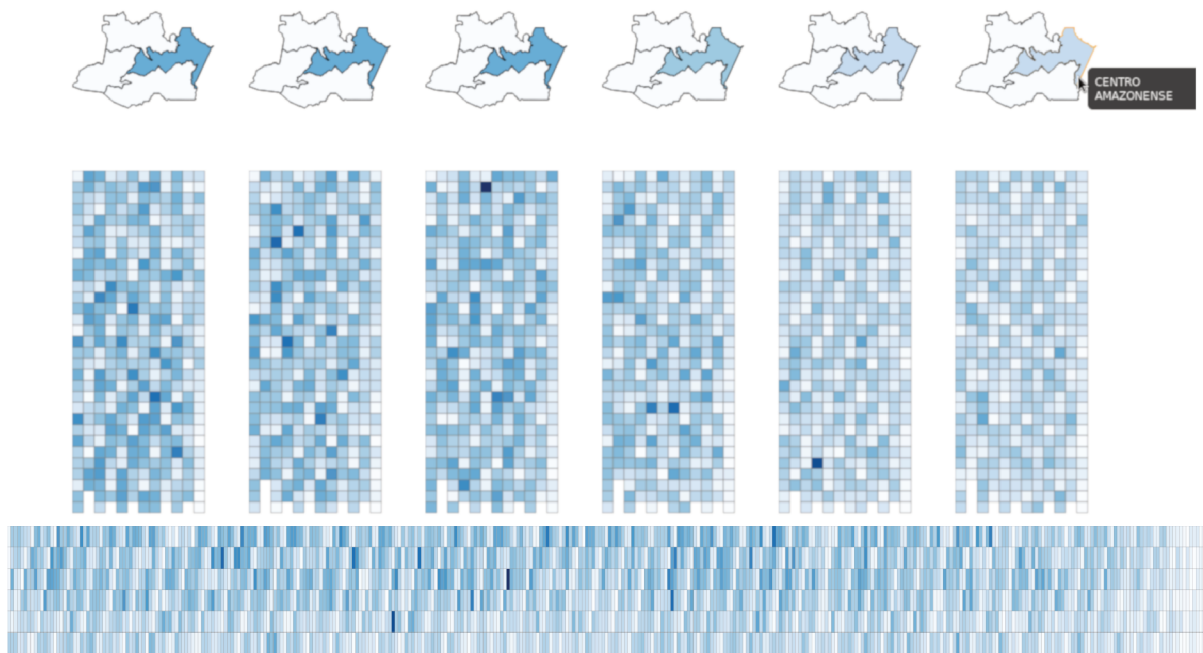


Figure 51: Small multiples and heatmaps detailing the distribution of LACs in Amazonas state from 2012 to 2017. The small maps shows that Centro Amazonense mesoregion is responsible for the majority of LACs reports. The heat maps shows the distribution of LACs for each day from 2012 to 2017, and the workdays/weekends patterns can be better seen in the stripes forming diagonal lines from top to bottom.

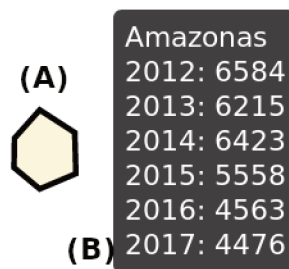


Figure 52: (A) shows the glyph associated with Amazonas state and (B) shows its associated values. A significant reduction in the number of LACs could be observed in this state both in the glyph, which presents a sharp point related to 2012, and in the tooltip, that confirms this reduction over the years.

Another interesting pattern could be seen when analyzing Alagoas state's heatmaps' configuration, which shows an increase on the number of LAs in the beginning/end of the years, with a slight decrease in 2017, as shown in Figure 53.

This pattern was already noticed in the Dendrogram Composition analysis presented in Section 4.2. However, the Choropleth Composition allowed the visual comparison among all considered years at once, in which the heatmaps quickly reveal the recurrence of this pattern. The small multiples layout also confirmed *Leste Alagoano* mesoregion as

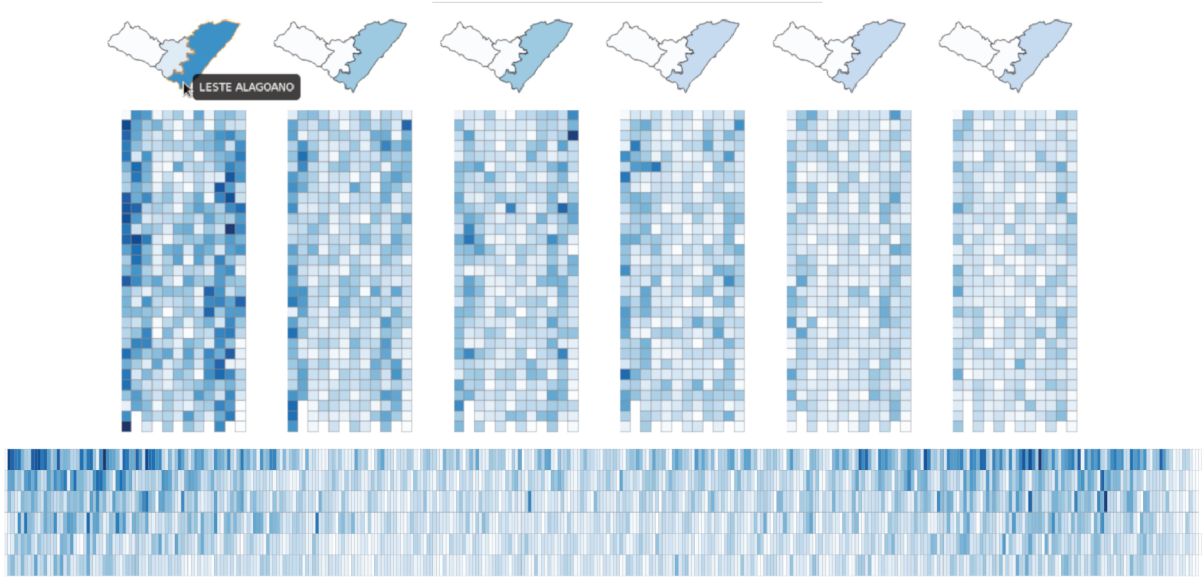


Figure 53: A detailed view of Alagoas state from 2012 to 2017 showing the peculiar pattern of LACs in the beginning/end of each year. This pattern can be confirmed by analyzing both visualizations in our proposed system.

the one that most contributes to this behavior, presenting the highest proportion of LACs registered in the state. This is another example that illustrates how the visualizations are complementary in providing a detailed comprehension on events involving LAs, and how they can be combined to enhance this analysis.

## 4.4 Final Considerations

In this chapter we presented the results obtained by using our system, which comprises two visualizations: Dendrogram Composition and Choropleth Composition. Using the visualizations we could identify patterns that recurrently occurred every year, as well as odd situations related to specific days with high numbers of LACs, which suggests collective accidents. However, we could also observe localities with an increase of LACs during a single year in which the LAs were well distributed all over the year. Furthermore, we observed localities with high proportion of LACs, which demand the attention of the authorities in order to help in reducing these numbers and to improve safety and security at work, and finally we also identified areas that are possibly lacking government attention and need inspection improvements. The system revealed patterns that would be difficult to find otherwise, effectively guiding the investigation of the data set.



---

## Conclusion

In this chapter we summarize the conclusions from this research project, outlining the main contributions, limitations and directions for future research. We proposed a visual strategy that combines a set of information visualization techniques to explore and analyze the evolution of labor accidents occurrences over time in Brazil. We also developed a computational system implementing this strategy, and performed experiments to evaluate its ability in revealing strategic patterns to aid decision making.

Using the visualizations we were able to identify seasonal patterns, such as activities regularly occurring in specific months every year, local/regional phenomena related to single localities or extended to a broader area, which are often related to same economic activity, odd occurrence behaviors in specific periods, such as a considerable increase/decrease of LACs in a specific period under analysis, similar distribution of accidents occurrences among localities with different economic activities or geographically distant, it was also possible to identify disparities among close localities with similar economic activities, patterns closely related to specific dates, such as holidays and weekends, among other things. The analysis of the layouts allowed the visualization of trends that would be difficult to identify by looking at the original tabular data. Both visualizations also proved their complementary capacity, one supporting some analysis made in the other.

The data used is publicly available and our strategy allows citizens to extract valuable information from it, creating an effective communication channel between government and population. The government can also benefit with this strategy, their specialists can understand and explain the patterns faster than a layman, getting even more meaningful information. It provides the comprehension of the labor situation in the localities and over time, allowing the government to direct their attention to specific localities to verify their work conditions and to promote a regular inspection. Moreover, it may provide means to improve the current public polices and/or create new ones, or to compare the current public polices in different localities and identify which is better to be applied to localities with same profile.

After performing the detailed analysis presented in Chapter 4, we were able to identify

some limitations in our proposed analysis strategy, as well as on the data itself. A crucial limitation related to the LACs data is that it strongly depends on the employee, the employer, or a supervisor to report the accident, and this process is very deficient, specially in localities from the Norte and Nordeste regions, in the sense that several LACs contain missing data, approximated data, or even no LAC is reported. This deficiency impacts on the layouts, that may not reflect the real situation of a specific locality, regarding labor accidents occurrence.

In our strategy, the Dendrogram Composition visualization does not allow the comparison of localities from different levels of the same hierarchy, such as a microregion with its corresponding state. For these tasks, users must combine the dendrogram from the first locality with the choropleth/small multiples from the Choropleth Composition visualization. It is also not trivial to compare the bar charts from far apart localities in the visual representation in the dendrogram. A reordering procedure of the nodes' position may facilitate the investigation of additional patterns and help the identification of groups of related localities, but finding the optimal order may not be trivial as well. Moreover, for localities presenting few registered accidents, it is difficult to identify meaningful patterns, and each accident may cause significant changes in the bar chart, which may confuse the analysis for non-experts. In this case, our tool should get the attention of the government's experts to localities/areas that lack effective inspection and data collect.

To improve the results of this research work and to extract even more information from this data, some possibilities could be explored as future work. We only considered the hierarchical and temporal information, as well as the number of registered accidents, but the LACs aggregate additional information that may be valuable to enhance the analysis. Moreover, we suggest additional interaction techniques for the system, including:

- ❑ Bar charts scale change, in the Dendrogram Composition, to allow exploration in specific periods;
- ❑ Exploration of other levels in the small multiples of the Choropleth Composition, in order to understand patterns formation in different levels;
- ❑ Alignment of the heatmap stripes by weekdays, in order to improve the comparison based on a different perspective in time organization;
- ❑ Show percentages of increase/decrease in the number of LACs and population size to improve proportion perception;
- ❑ Use a visual representation for the tooltip of the Dendrogram Composition to inform the evolution throughout the years.

We also intend to perform a user study both with common users and with experts from the Brazilian Federal Labor Prosecution Office. The experiment with common users

may reveal whether the system is being able to communicate the information about LAs to the population. The BFLPO's expert are the users who can most benefit from this analysis, experiment with them may help us to evaluate our strategy in terms of usability and decision making capabilities. Their feedback on our strategy may also contribute to the development of new functionalities directly guided by their specific needs.

The proposed strategy and the system were developed to attend the specific needs of the BFLPO regarding LAs, but we noticed that our strategy may benefit the analysis of other data repositories that bear the same characteristics, helping the decision making in different scenarios.





---

## Bibliography

- BADAM, S. K. et al. Timefork: Interactive prediction of time series. In: **Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems**. New York, NY, USA: ACM, 2016. (CHI '16), p. 5409–5420. ISBN 978-1-4503-3362-7. DOI 10.1145/2858036.2858150.
- BENIGER, J. R.; ROBYN, D. L. Quantitative graphics in statistics: A brief history. **The American Statistician**, Taylor & Francis, v. 32, n. 1, p. 1–11, 1978. DOI 10.1080/00031305.1978.10479235.
- BORGO, R. et al. Glyph-based visualization: Foundations, design guidelines, techniques and applications. **Eurographics State of the Art Reports**, 05 2013. DOI 10.2312/conf/EG2013/stars/039-063.
- BOSTOCK, M.; OGIEVETSKY, V.; HEER, J. D3: data-driven documents. **IEEE Transactions on Visualization and Computer Graphics**, v. 17, n. 12, p. 2301–2309, Dec 2011. ISSN 1077-2626. DOI 10.1109/TVCG.2011.185.
- BURCH, M.; BECK, F.; DIEHL, S. Timeline trees: Visualizing sequences of transactions in information hierarchies. In: **Proceedings of the Working Conference on Advanced Visual Interfaces**. New York, NY, USA: ACM, 2008. (AVI '08), p. 75–82. ISBN 978-1-60558-141-5. DOI 10.1145/1385569.1385584.
- CHAMBERS, J. M. et al. **Graphical Methods for Data Analysis**. [S.l.]: Chapman and Hall/Cole Publishing Company, 1998. 158-162 p. DOI 10.1201/9781351072304. ISBN 0-412-05271-7.
- CHANG, W. et al. **shiny: Web Application Framework for R**. [S.l.], 2017. R package version 1.0.5.
- DATAVIVA. **DataViva**. 2013. [Http://dataviva.info](http://dataviva.info).
- DATAWHEEL. **DataUSA**. 2016. [Https://datausa.io/](https://datausa.io/).
- \_\_\_\_\_. **DataChile**. 2018. [Https://es.datachile.io/](https://es.datachile.io/).
- DEWEN, S.; PING, S. Visualization of complicated objects for hydraulic engineering based on opengl. In: **Proceedings of 2011 International Conference on Electronic Mechanical Engineering and Information Technology**. [S.l.: s.n.], 2011. v. 1, p. 316–319. DOI 10.1109/EMEIT.2011.6022934.

DÍAZ, P.; AEDO, I.; HERRANZ, S. Understanding citizen participation in crisis and disasters: The point of view of governmental agencies. In: **Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces**. New York, NY, USA: ACM, 2014. (AVI '14), p. 395–397. ISBN 978-1-4503-2775-6. DOI 10.1145/2598153.2602227.

FRIENDLY, M. **Milestones in the history of thematic cartography, statistical graphics, and data visualization**. 2009. DOI 10.1007/3-540-28084-7\_4.

FUCHS, J. et al. Evaluation of alternative glyph designs for time series data in a small multiple setting. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. New York, NY, USA: ACM, 2013. (CHI '13), p. 3237–3246. ISBN 978-1-4503-1899-0. DOI 10.1145/2470654.2466443.

GRAELLS, E.; JAIMES, A. Lin-spiration: Using a mixture of spiral and linear visualization layouts to explore time series. In: **Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces**. New York, NY, USA: ACM, 2012. (IUI '12), p. 237–240. ISBN 978-1-4503-1048-2. DOI 10.1145/2166966.2167006.

GROEGER, L.; GRABELL, M.; COTTS, C. **Workers' Comp Benefits: How Much is a Limb Worth?** 2015. <https://projects.propublica.org/graphics/workers-compensation-benefits-by-limb>.

HOXHA, J.; BRAHAJ, A.; VRANDEČIĆ, D. Open.data.al: Increasing the utilization of government data in albania. In: **Proceedings of the 7th International Conference on Semantic Systems**. New York, NY, USA: ACM, 2011. (I-Semantics '11), p. 237–240. ISBN 978-1-4503-0621-8. DOI 10.1145/2063518.2063558.

ILO. **Safety and Health at work**. 2019. <http://www.ilo.org/global/topics/safety-and-health-at-work/lang-en/index.htm>.

INSS. **Comunicação de Acidente de Trabalho – CAT**. 2018. <https://www.inss.gov.br/servicos-do-inss/comunicacao-de-acidente-de-trabalho-cat/>.

International Food Policy Research Institute (IFPRI) and Datawheel. **Data Africa**. 2017. <https://DataAfrica.io>.

JANG, Y.; EBERT, D.; GAITHER, K. Time-varying data visualization using functional representations. **IEEE Transactions on Visualization and Computer Graphics**, IEEE Computer Society, v. 18, n. 3, p. 421–433, 1 2012. ISSN 1077-2626. DOI 10.1109/TVCG.2011.54.

KOLESNICHENKO, O. et al. Big data analytics of inpatients flow with diabetes mellitus type 1 : Revealing new awareness with advanced visualization of medical information system data. In: **2019 9th International Conference on Cloud Computing, Data Science Engineering (Confluence)**. [S.l.: s.n.], 2019. p. 191–196. DOI 10.1109/CONFLUENCE.2019.8776910.

LEE, S.-M. et al. A new visualization method for pairwise time-series data with random walk plot. In: **Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication**. New York, NY, USA: ACM, 2015. (IMCOM '15), p. 64:1–64:8. ISBN 978-1-4503-3377-1. DOI 10.1145/2701126.2701144.

- LEE, T.-Y. et al. Cyclestack: Inferring periodic behavior via temporal sequence visualization in ultrasound video. In: **PacificVis '10: Proceedings of IEEE Pacific Visualization Symposium 2010**. [S.l.: s.n.], 2010. p. 89–96. DOI 10.1109/pacificvis.2010.5429602.
- LI, N. et al. A method of hierarchical time-series data visualization. In: **Proceedings of the 6th International Symposium on Visual Information Communication and Interaction**. New York, NY, USA: ACM, 2013. (VINCI '13), p. 113–114. ISBN 978-1-4503-1988-1. DOI 10.1145/2493102.2493121.
- LIU, S. et al. A survey on information visualization: recent advances and challenges. **The Visual Computer**, v. 30, n. 12, p. 1373–1393, Dec 2014. ISSN 1432-2315. DOI 10.1007/s00371-013-0892-3.
- LU, W.; WEI, J. Dynamic visualization of evolutionary curricula model. In: **2010 International Conference on Educational and Information Technology**. [S.l.: s.n.], 2010. v. 2, p. V2–484–V2–488. DOI 10.1109/ICEIT.2010.5607621.
- LUIZ, E. R. **Geographic Information Systems (GIS) - Brasil Coleção de shape-files, GeoJSON e TopoJSON prontas para uso**. 2014. [Http://fititnt.github.io/gis-dataset-brasil/](http://fititnt.github.io/gis-dataset-brasil/).
- MENDONÇA, P. G. A. de; MACIEL, C.; FILHO, J. V. Visualizing aedes aegypti infestation in urban areas: A case study on open government data mashups. In: **Proceedings of the 15th Annual International Conference on Digital Government Research**. New York, NY, USA: ACM, 2014. (dg.o '14), p. 186–191. ISBN 978-1-4503-2901-9. DOI 10.1145/2612733.2612751.
- Ministério Público do Trabalho. **MPT - Ministério Público do Trabalho**. SAUN Quadra 5, Lote C, Torre A - Brasília - DF: [s.n.], 2018. [Https://mpt.mp.br/pgt/mpt-nos-estados](https://mpt.mp.br/pgt/mpt-nos-estados).
- MORAWA, R. et al. Combining timeline and graph visualization. In: **Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces**. New York, NY, USA: ACM, 2014. (ITS '14), p. 345–350. ISBN 978-1-4503-2587-5. DOI 10.1145/2669485.2669544.
- MUNZNER, T. **Visualization Analysis and Design**. Boca Raton: A K Peters/CRC Press, 2014. (AK Peters Visualization Series). DOI 10.1201/b17511. ISBN 978-1-4665-0893-4.
- PARKINS, F. L. M. T.; MCCUTCHEON, R. T. Visualisation of construction costs and techniques of employment-intensive road construction in developing areas. In: **1999 IEEE International Conference on Information Visualization (Cat. No. PR00210)**. [S.l.: s.n.], 1999. p. 225–229. DOI 10.1109/IV.1999.781563.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2018.
- RODRIGUES, N. et al. Visualization of time series data with spatial context: Communicating the energy production of power plants. In: **Proceedings of the 10th International Symposium on Visual Information Communication**

**and Interaction**. New York, NY, USA: ACM, 2017. (VINCI '17), p. 37–44. ISBN 978-1-4503-5292-5. DOI 10.1145/3105971.3105982.

SHARMIN, M. et al. Visualization of time-series sensor data to inform the design of just-in-time adaptive stress interventions. In: **Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing**. New York, NY, USA: ACM, 2015. (UbiComp '15), p. 505–516. ISBN 978-1-4503-3574-4. DOI 10.1145/2750858.2807537.

Smartlab de Trabalho Decente MPT - OIT. **Observatório Digital de Saúde e Segurança no Trabalho**. 2017. [Http://observatoriosst.mpt.mp.br](http://observatoriosst.mpt.mp.br).

SORIANO-VARGAS, A. et al. Visual analytics of time-varying multivariate ionospheric scintillation data. **Computers & Graphics**, v. 68, p. 96–107, 2017. ISSN 0097-8493. DOI 10.1016/j.cag.2017.08.013.

State of Minnesota - Department of Health. **The Minnesota Government Data Practices Act**. 2019.

<https://www.health.state.mn.us/communities/practice/resources/chsadmin/data-mgdpa.html>.

TERVAKARI, A.-M. et al. Usefulness of information visualizations based on educational data. In: **2014 IEEE Global Engineering Education Conference (EDUCON)**. [S.l.: s.n.], 2014. p. 142–151. DOI 10.1109/EDUCON.2014.6826081.

THOMAS, J. J.; COOK., K. A. Illuminating the path: An r&d agenda for visual analytics. In: \_\_\_\_\_. [S.l.]: National Visualization and Analytics Ctr, 2005. cap. 3, p. 69–104. ISBN 0769523234.

TUFTE, E. R. **The Visual Display of Quantitative Information**. [S.l.]: Graphics Pr, 2001. ISBN 0961392142.

WANG, H. et al. The design and implementation of the platform of interactive information visualization on aircraft product development. In: **2012 3rd International Conference on System Science, Engineering Design and Manufacturing Informatization**. [S.l.: s.n.], 2012. v. 1, p. 180–184. DOI 10.1109/ICSSEM.2012.6340702.

WARD, M. O. **Multivariate Data Glyphs: Principles and Practice**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. 179-198 p. DOI 10.1007/978-3-540-33037-0\_8. ISBN 978-3-540-33037-0.

WICKHAM, H. **tidyverse: Easily Install and Load the 'Tidyverse'**. [S.l.], 2017. R package version 1.2.1.

YING, Y.; XIALING, T.; WEI, T. Study on governmental cultural resources purchase management based on public information behavior big data. In: **Proceedings of the 8th International Conference on E-Education, E-Business, E-Management and E-Learning**. New York, NY, USA: ACM, 2017. (IC4E '17), p. 72–75. ISBN 978-1-4503-4821-8. DOI 10.1145/3026480.3026491.

ZHIYUAN, H. et al. Application of big data visualization in passenger flow analysis of shanghai metro network. In: **2017 2nd IEEE International Conference on Intelligent Transportation Engineering (ICITE)**. [S.l.: s.n.], 2017. p. 184–188. DOI 10.1109/ICITE.2017.8056905.