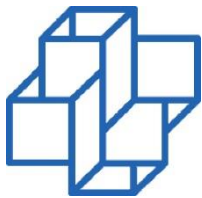




**Centro Federal de Educação Tecnológica
Celso Suckow da Fonseca
CEFET/RJ**

Music Project 2016 Report



**National
Laboratory
Scientific
Computing**

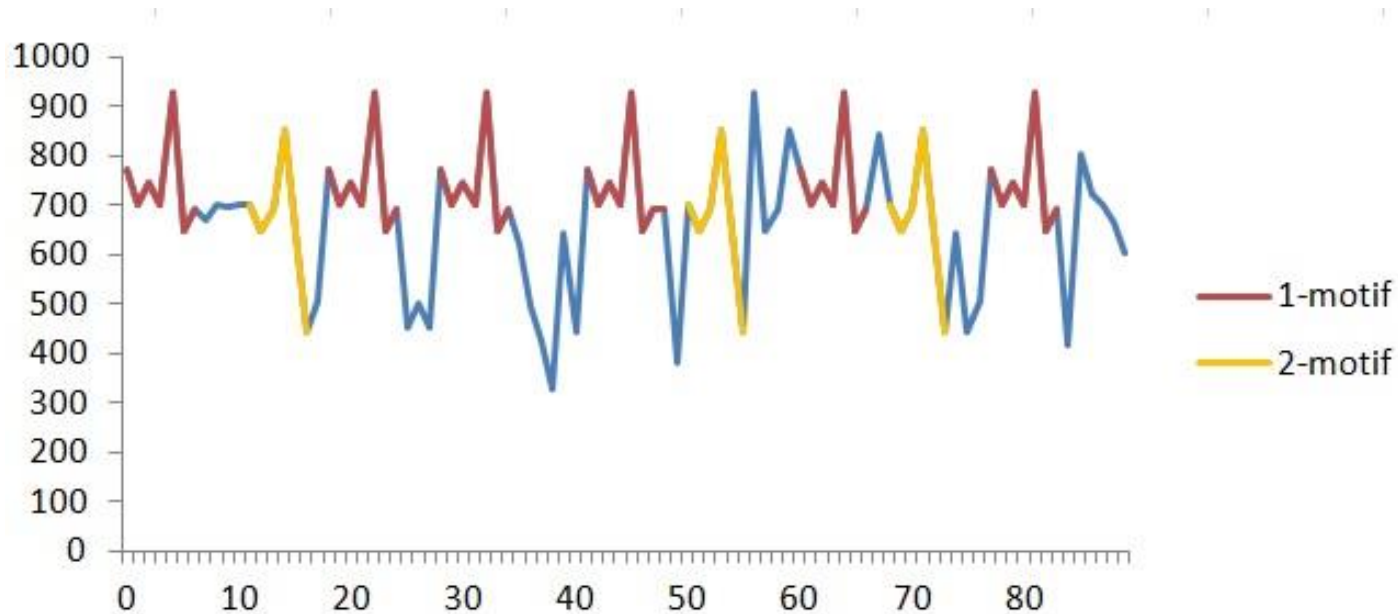


Eduardo Ogasawara

<http://eic.cefet-rj.br/~eogasawara>

Motifs in Time Series

- Time series express phenomenon of interest
- Identifying motifs (unknown patterns) in time series brings knowledge and enables predictions



- Similar studies are also being made by sequence mining researchers aiming to find patterns in time series

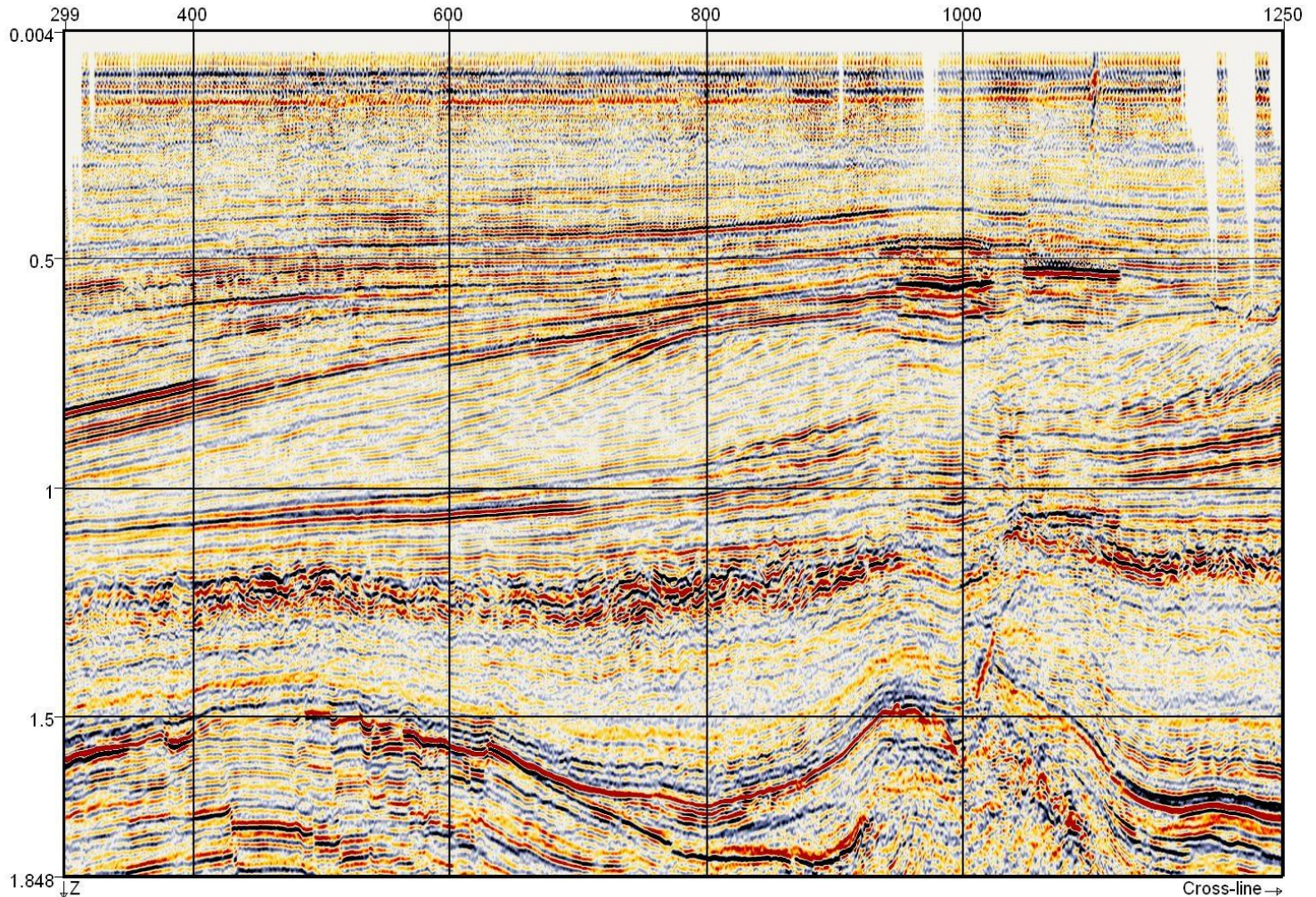
Spatial-Time Series

- Some phenomenon are modeled as set of time series, each one with a particular position
- Although many research is being made in motifs/patterns identification, few work is being made in spatial-time series
 - From motifs researchers (no other work)
 - From sequence mining researches (very few)
- Opens space for new approaches, algorithms and method trying to address motifs is spatial-time series

Seismic Traces Analysis (Netherlands dataset)

Space

Time



Crossline: 100 (951 time series with 462 observations)

Time Series and Sequences

Definition 1. A *time series* t is an ordered sequence of values in time [1], where each t_i is a value, $|t| = m$ is the number of elements in t , and t_m is the most recent value in t .

$$t = \langle t_1, t_2, \dots, t_m \rangle, t_i \in \mathbb{R}$$

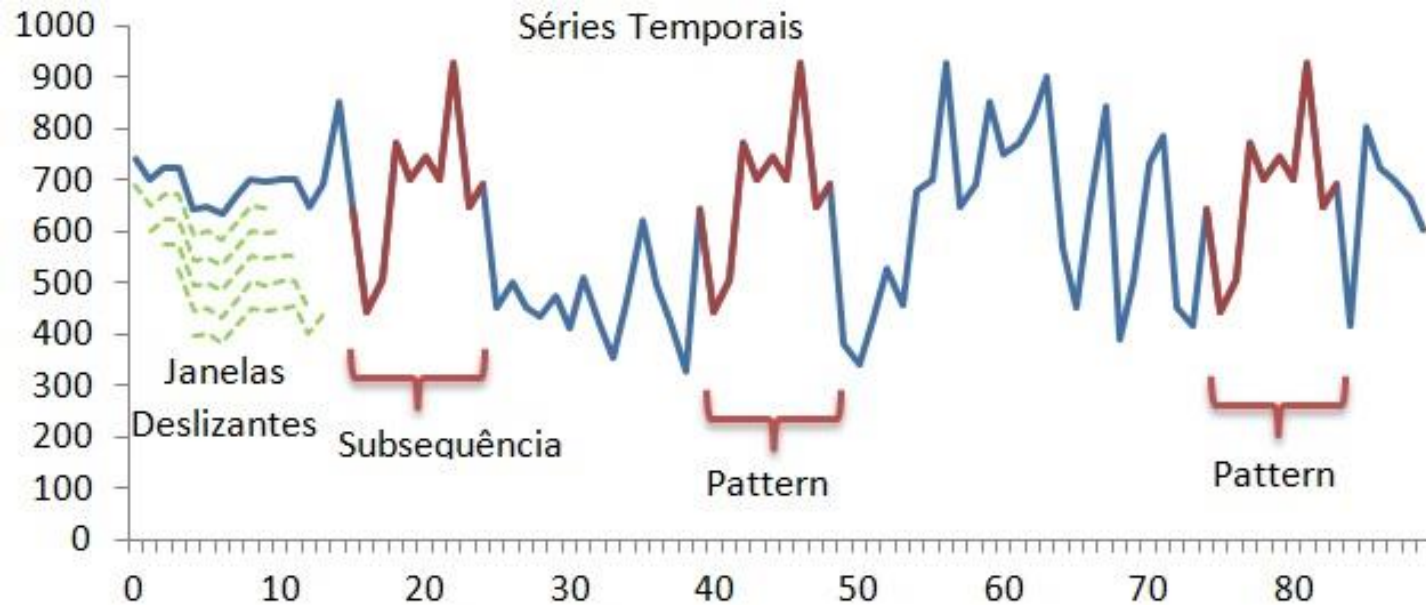
Definition 2. The p -th *sub sequence* [2] of size n in a time series t , represented as $t^{p,n}$, is an ordered sequence of values $\langle t_p, t_{p+1}, \dots, t_{p+n-1} \rangle$, where $|t^{p,n}| = n$ and $1 \leq p \leq |t| - n$.

$$t^{p,n} = \text{subseq}(t, p, n)$$

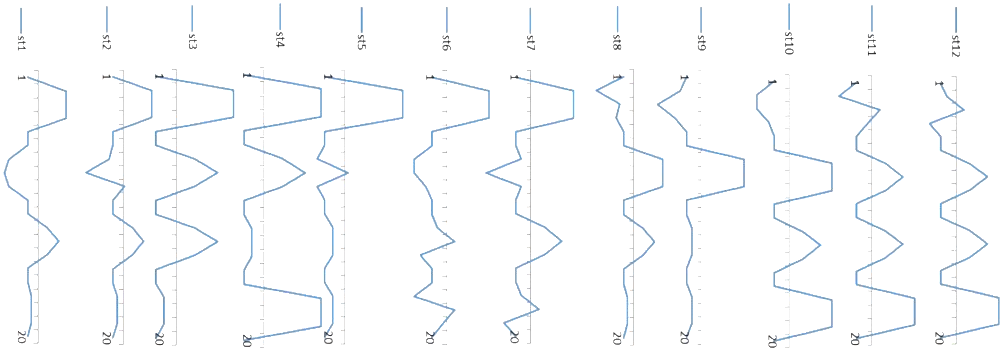
Motif in Time Series

Definition 5 Let $q = \langle q_1, q_2, \dots, q_n \rangle$ and $t = \langle t_1, t_2, \dots, t_m \rangle$ be two time series, such that $|q| = n$, $|t| = m$, and $m > n$. q is **included** in t ($q < t$) if and only if $\exists w_i \in W$ such that $w_i = q$.

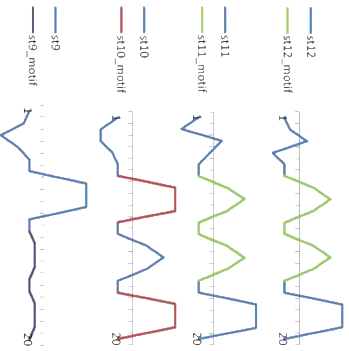
Definition 6 Given two time series q and t , q is a **motif** [Mueen, 2014] with support σ , if and only if q is included in t at least σ times. Formally, given time series q and t such that $W = sw(t, |q|)$, motif $(q, t, \sigma) \iff \exists R \subseteq W$, such that $\forall w_i \in R, w_i = q \wedge |R| \geq \sigma$.



Intuition of limitations of current approaches in spatial-time series



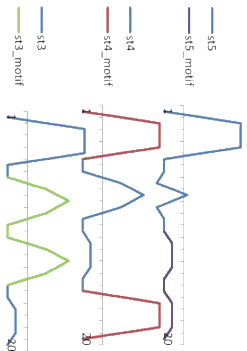
Motif Discovery Algorithm



st8 - No Motif found

st7 - No Motif found

st6 - No Motif found



st2 - No Motif found

st1 - No Motif found

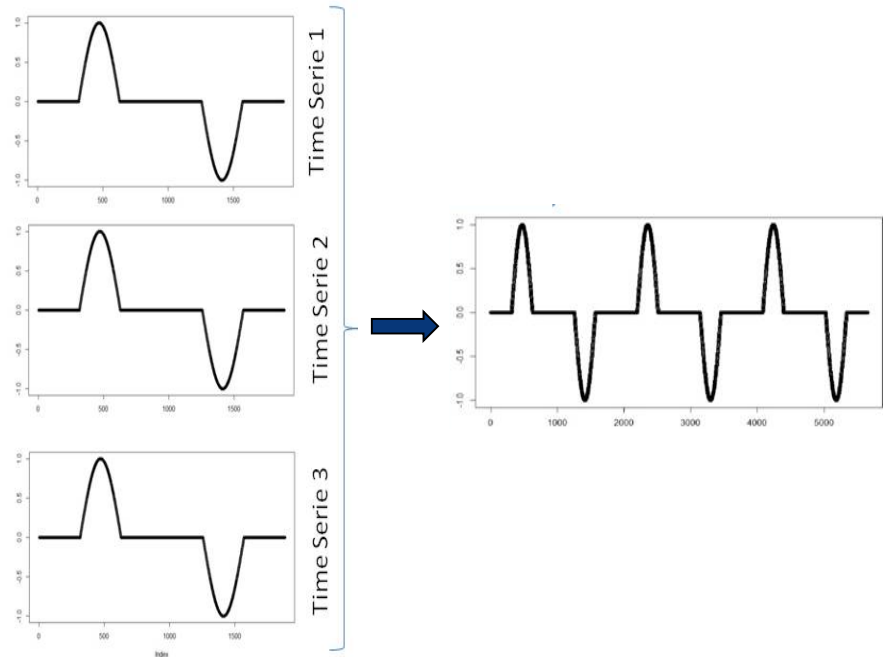
Spatial Time Motif

Definition 8 Let σ and κ be two support values such that $\sigma \geq \kappa$. A subsequence q is a **spatial-time motif** iff q is included at least σ times in D and q occurs in at least κ different spatial-time series.

		sslice = 4				sslice = 4				sslice = 4				
		st1	st2	st3	st4	st5	st6	st7	st8	st9	st10	st11	st12	
tslice = 10	#	1	b	b	b	b	b	b	b	b	b	b	b	b
	2	e	e	e	e	e	e	e	b	b	b	b	b	
	3	e	e	e	e	e	e	e	a	b	a	a	b	
	4	e	e	e	e	e	e	e	b	a	a	d	d	
	5	b	b	b	b	b	b	b	a	a	c	c	a	
	6	b	b	b	b	b	b	b	b	b	b	b	b	
	7	a	b	d	d	b	e	b	e	e	e	d	d	
	8	a	a	e	e	c	b	b	e	e	e	e	e	
	9	a	c	d	d	b	b	a	e	e	e	d	d	
	10	b	b	b	b	b	b	b	b	b	b	b	b	
tslice = 10	11	b	b	b	b	b	b	b	b	b	b	b	b	
	12	d	d	d	b	b	e	d	d	b	d	d	d	
	13	e	e	e	b	b	d	e	e	b	e	e	e	
	14	d	d	d	b	b	a	d	d	b	d	d	d	
	15	b	b	b	b	b	b	b	b	b	b	b	b	
	16	b	b	b	d	b	b	b	b	b	b	b	b	
	17	b	b	b	e	b	a	d	b	b	e	e	e	
	18	b	b	b	e	b	c	c	b	b	e	e	e	
	19	b	b	b	e	b	a	b	b	b	e	e	e	
	20	b	b	b	b	b	d	b	b	b	b	b	b	

Figure V.4: Toy dataset partitioned into blocks

Combined time series



Combined Spatial-Time Series

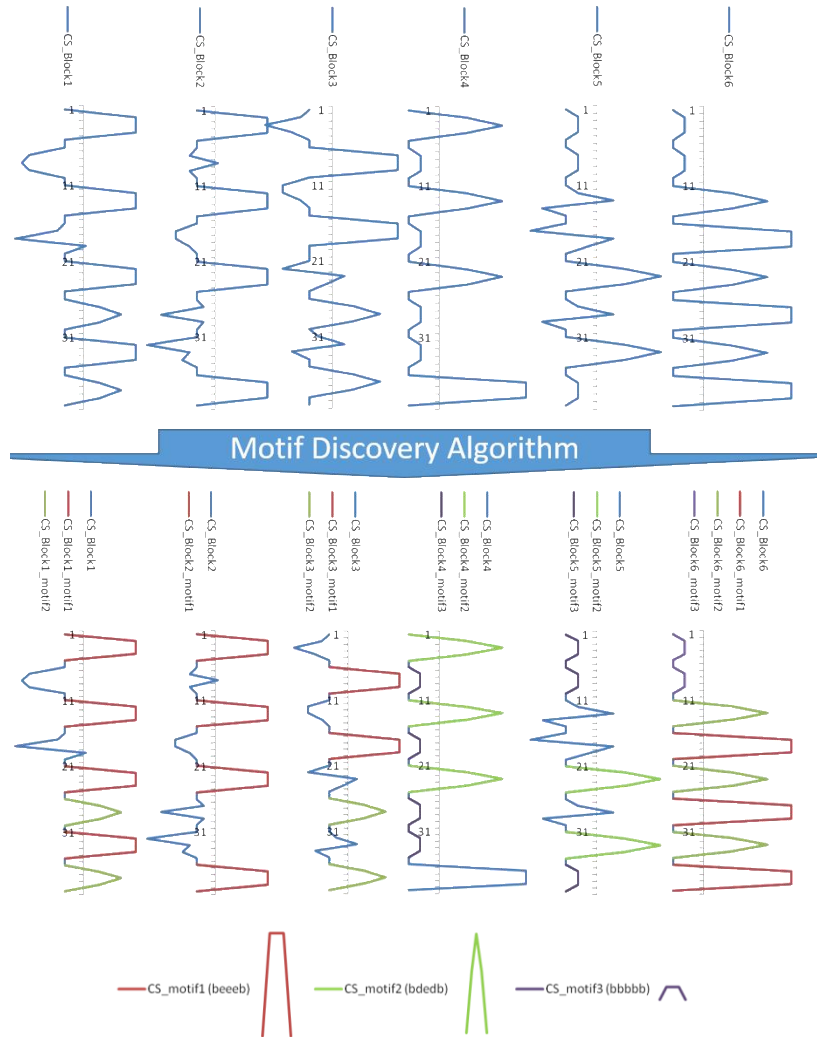
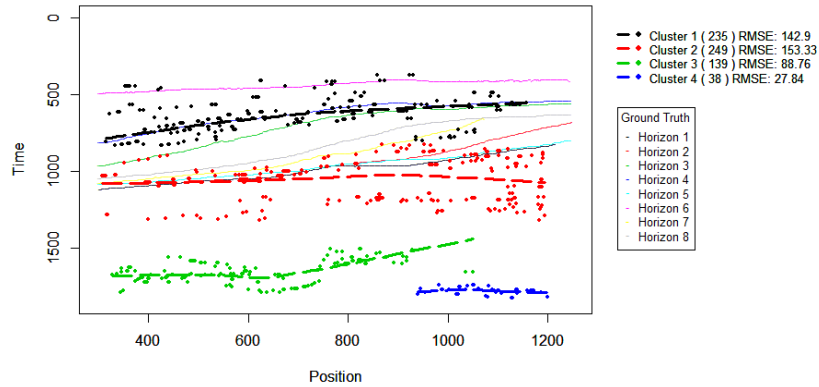


Figure V.5: Motif Discovery Algorithm to Combined Series

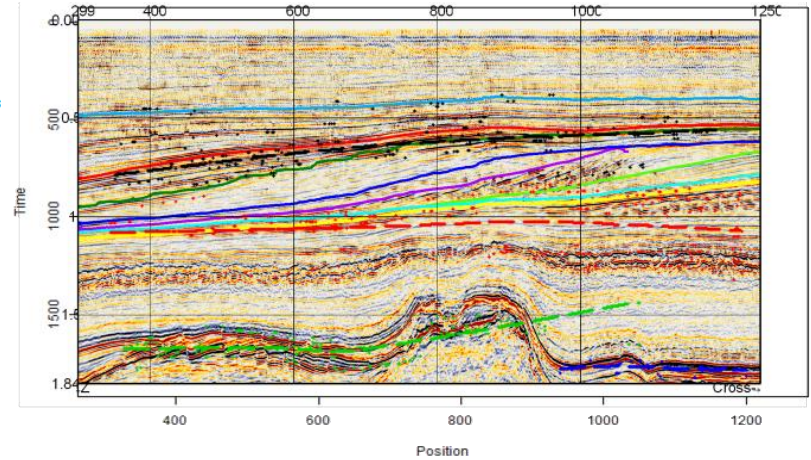
Combined Series Algorithm

```
1: function STMOTIF( $D, w, a, sslice, tslice, \sigma, \kappa, rf$ )
2:    $D_s \leftarrow norm\_sax(D, a)$ 
3:    $b \leftarrow partition(D_s, sslice, tslice)$ 
4:   for each  $b_i \in b$  do
5:      $t \leftarrow combine(b_i)$ 
6:      $motifs \leftarrow identify(t, w) \cup motifs$ 
7:   end for
8:    $cand\_motifs = aggregate(motifs)$ 
9:    $st\_motifs = evaluate(cand\_motifs, \sigma, \kappa)$ 
10:   $topst\_motifs = rank(st\_motifs, rf)$ 
11:  return  $topst\_motifs$ 
12: end function
```

Evaluation Using Seismic Datasets



(a) Result - Best ranked motif



(b) Result over seismic picture

Figure VI.5: Combined series experiment result

Summary of research

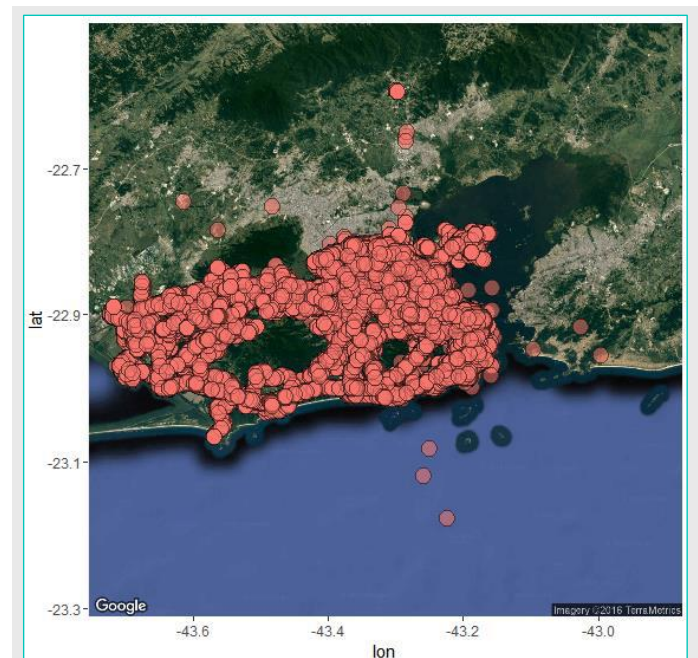
- Spatial-Time Motif Identification
 - Master Student: Murillo Dutra (defended on July 2016)
 - New spatial-time motif algorithm
 - Experimental evaluation: seismic dataset
 - Expected Target: Find seismic horizons
 - Current goals:
 - Write paper
- Spatial Sequence Mining
 - Master Student: Riccardo Campisano (defense on March 2017)
 - First paper submitted to SBBD 2016
 - Experimental evaluation: seismic dataset
 - Expected target: Find seismic faults
 - Current Goal: Develop new algorithm to find tight patterns

Current Results

- One Master Degree Murillo Dutra (July 2016)
 - Co-advised by Fabio Porto
- SBBD Paper: Spatial Sequential Pattern Mining for Seismic Data
 - Riccardo Campisano, Fabio Porto, Esther Pacitti, Florent Masegla, Eduardo Ogasawara

Next Steps: Research Proposal for Urban Mobility

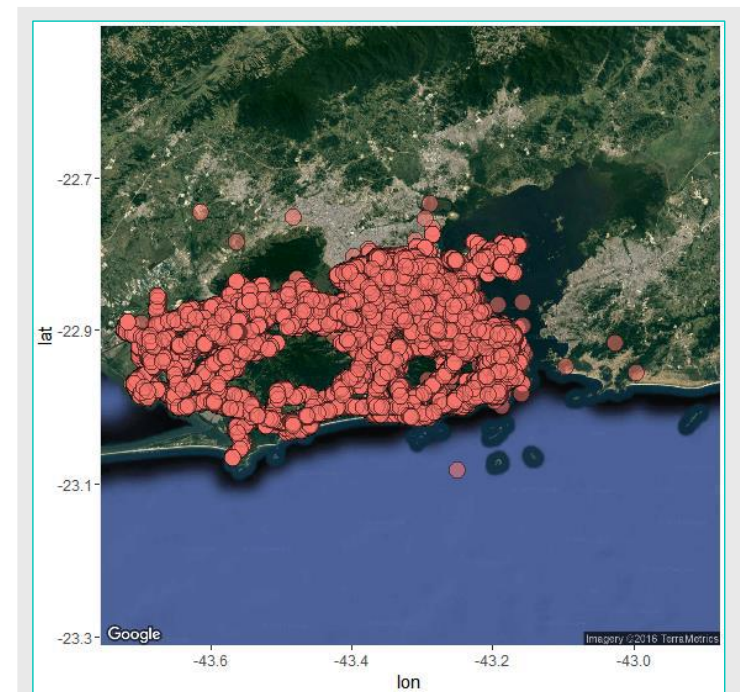
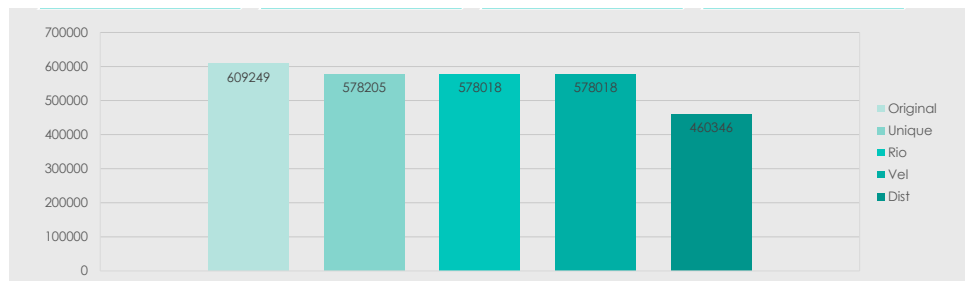
- Spatial-Time Motif Identification for Urban Mobility (IoT)
 - Trajectory data using Buses of Rio de Janeiro
 - All buses send its position every minute



Buses positions in Rio in one day

Urban Mobility: Steps Already done

- Data Collection and Cleaning
 - Each day, we collect approximately 600.000 observations
 - After cleaning and outlier removal we maintain approximately 450.000 observations

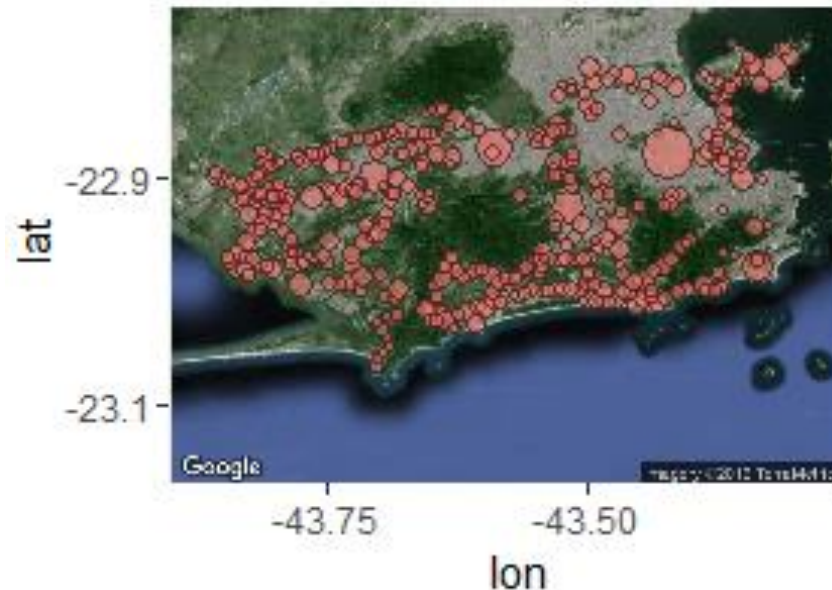


Proposal for Urban Mobility

- Explore spatial-time motif identification in *trajectory data*
 - Goal: Find relevant patterns of moving objects
 - Expected to develop new trajectory motif algorithm
- Explore spatial-time motif identification in buses stations aggregated data
 - Apply spatial-temporal aggregation to convert buses trajectory data into permanent (fixed position) spatial-time series placed on buses stations
 - Goal: Find relevant relevant patterns in different regions
 - Expected to develop new spatial-time motif algorithm

Proposal for Urban Mobility

- Spatial-Time Motif Identification for Urban Mobility (IoT)
 - Spatial-temporal aggregation
 - Aggregate trajectory data into buses stations
 - New spatial-time motif algorithm
 - Find relevant patterns



New Students

- Ana Beatriz Cruz
 - Master Degree Student
 - ETL on Trajectory Data
 - Spatial-Temporal Aggregation
 - Initial implementation of spatial-time motifs
- Heraldo Borges
 - PhD Student
 - Study in more depth Spatial-Time Motifs
 - Possibility to work with Esther and Florent one year at Montpellier
 - Propose new algorithms
 - First Goal: Aid Murillo in submitting his paper