



Detecção de Anomalias Frequentes no Transporte Rodoviário Urbano

Ana Beatriz Cruz¹, João Ferreira¹, Diego Carvalho¹, Eduardo Mendes³,
Esther Pacitti⁴, Rafaelli Coutinho¹, Fabio Porto², Eduardo Ogasawara¹

¹ CEFET/RJ - Centro Federal de Educação Tecnológica Celso Suckow da Fonseca

² LNCC - Laboratório Nacional de Computação Científica

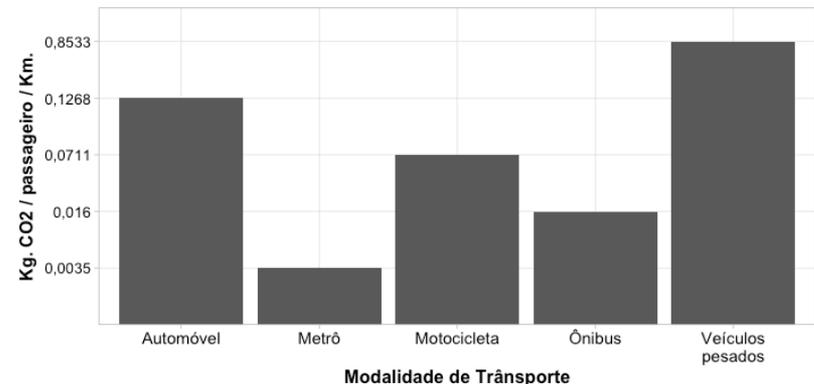
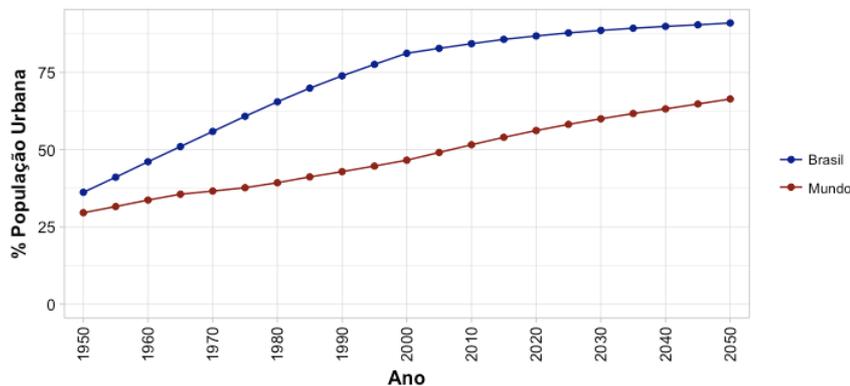
³ FGV - Fundação Getúlio Vargas

⁴ Inria & University of Montpellier



Introdução

- Aumento da população urbana
- Problemas relacionados a mobilidade
 - Muitos veículos rodoviários, trânsito intenso e poluição
 - Engarrafamentos crônicos
 - Acidentes e engarrafamentos anômalos
- Quais eventos são sistêmicos e quais são anômalos?

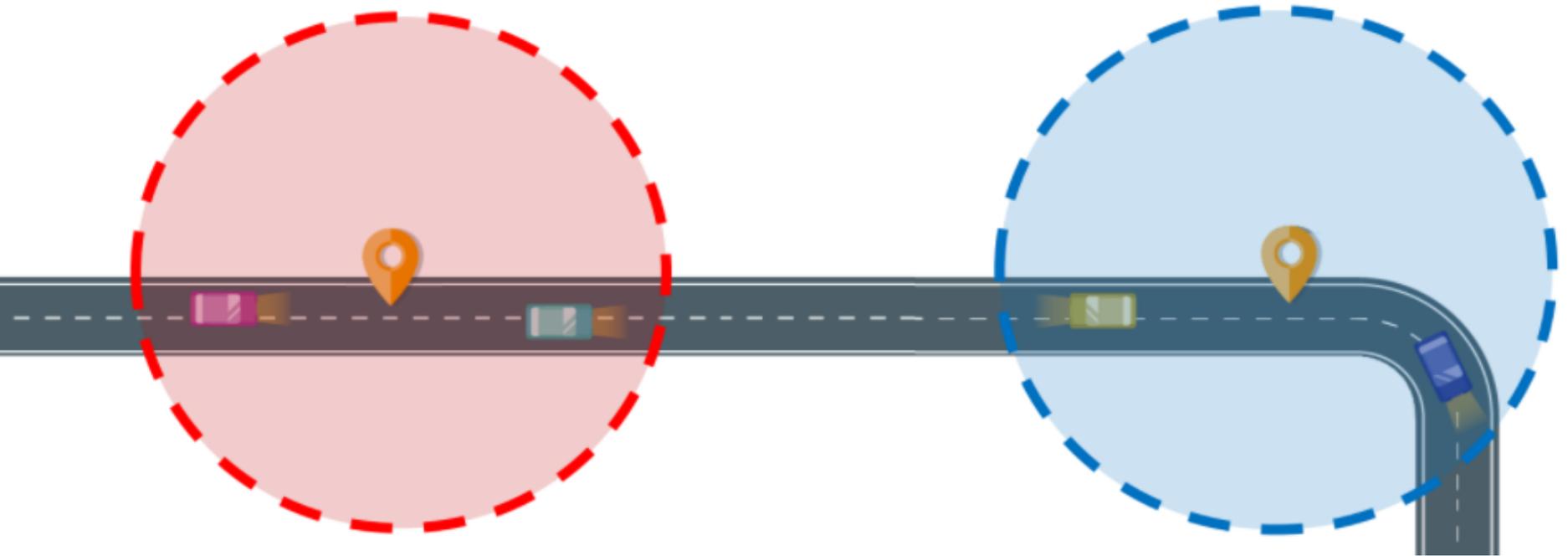


Estudos de comportamento de trânsito

- **Perspectiva de trajetória**
 - Compreensão do comportamento do trânsito visando minimizar impactos de engarrafamentos
 - Ferreira et al. [2013] usaram dados de GPS de taxi NY para ter análise visual das trajetórias
 - Verhein and Chawla [2008] usaram dados de celulares, RFID e satélites para identificar trajetórias de trânsito intenso
 - Pan et al. [2013] analisaram as variações nas movimentações das trajetórias para observar comportamento anômalos em rotas
- **Perspectiva de agregações espaço-temporais**
 - Associação de observação de objetos móveis a objetos permanentes
 - Andrienko e Andrienko [2008,2011] propuseram métodos de agregação espaço-temporal para trajetórias visando formar mapas de fluxo do espaço urbano

Agregação espaço-temporal

- Carência de estudos que analisam a mobilidade por uma perspectiva de agregações de séries espaço-temporais
 - Propicia uma melhor compreensão?
 - Propicia uma compreensão complementar?



Trajectoria

- Dados gerados por sensores como GPS e RFID
- Observação:
 - $\hat{tr}_o = (p, v)$
 - $\hat{tr}_o.v \in \mathbb{R}$
 - $\hat{tr}_o.p \in \mathbb{R}^2$
- Trajetória:
 - $tr_k = \langle \hat{tr}_1, \dots, \hat{tr}_{co} \rangle$
- Visão Individualizada do objeto que se move no espaço
- *Dataset* de trajetórias:
 - $TR = \cup_{k=1}^{ck} tr_k$



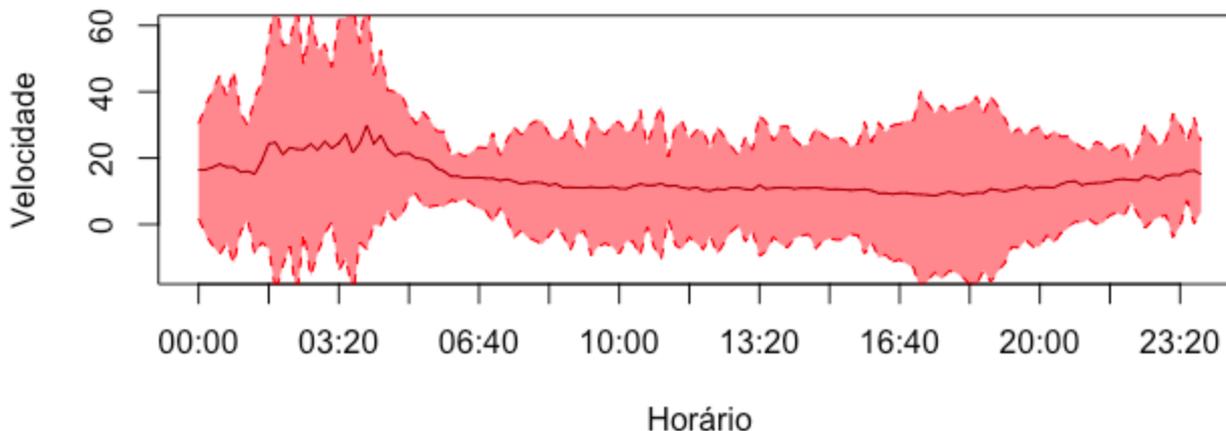
Agregação de séries espaço-temporais de objetos permanentes a partir das séries de trajetórias

- Partição do espaço S em c_i regiões
 - $S = \bigcup_{i=1}^{c_i} \hat{s}_i$
 - \hat{s}_i é uma região do espaço: $\hat{s}_i = [(x_{l_i}, y_{l_i}), (x_{u_i}, y_{u_i})]$
- Partição do tempo T em c_j intervalos
 - $T = \bigcup_{j=1}^{c_j} \hat{t}_j$
 - \hat{t}_j é um intervalo de tempo: $\hat{t}_j = [(j - 1) \cdot ma + 1, j \cdot ma]$
- Dataset ST de séries espaço-temporais:
 - $ST = \bigcup_{i=1}^{c_i} st_i, st_i = \langle \hat{u}_{i,1}, \dots, \hat{u}_{i,c_j} \rangle$
- Agregação
 - $\forall \hat{s}_i \in S, \hat{t}_j \in T \left(\hat{u}_{i,j} = \Gamma_{\theta}(tr_k \cdot \hat{tr}_o \cdot v) \left(\sigma_{o \in \hat{t}_j, tr_k \cdot \hat{tr}_o \cdot p \in \hat{s}_i} tr_k \cdot \hat{tr}_o \right) \right)$

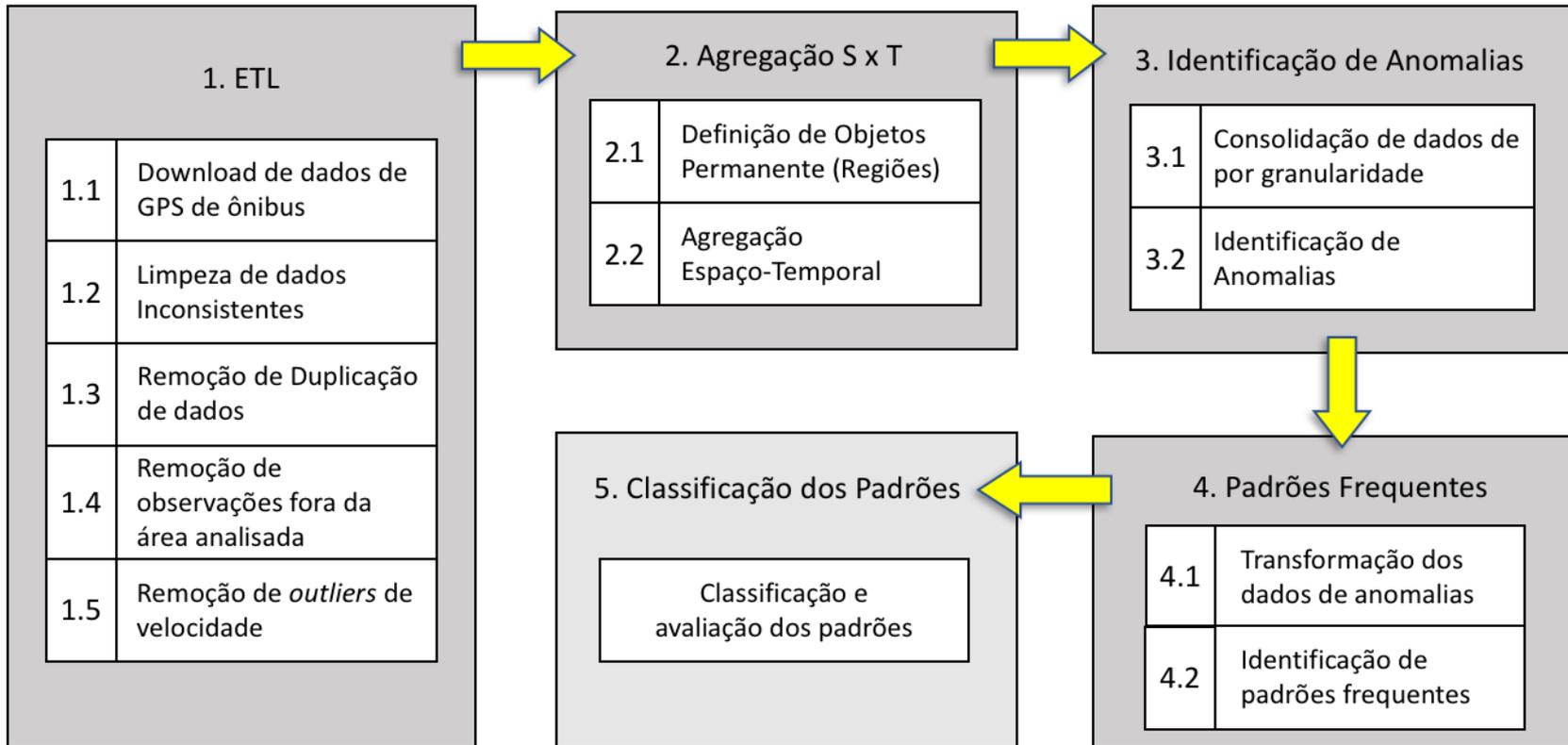


Valor esperado, intervalo típico e anomalias

- Considere ma o intervalo de particionamento do tempo em minutos
 - Um dia tem obs observações: $obs = \frac{1440}{ma}$, ma divisor de 1440
- Para $w \in [1, obs]$ as observações $\hat{u}_{i,w+d \cdot ma}$ corresponde a observações no mesmo horário w nos diferentes dias d
- O valor esperado na posição i e no horário w para um intervalo de dias D é $\overline{st}_{i,w,D}$
- O intervalo típico $IT_{i,w,D}$:
 - $IT_{i,w,D} = [q_1 - 3 \times IQR, q_3 + 3 \times IQR]$
- Anomalia são observações em ST no intervalo de dias D que não sejam típicas



Metodologia



Análise dos padrões gerais e mensais

transacoes_dia	transacoes_dia_semana	transacoes_num_semana	transacoes_mes
- data (DD-MM-YYYY)	- dia_semana: string	- num_semana: string	- mes: string
- velocidade: string	- velocidade: string	- velocidade: string	- velocidade: string
- horário: string	- horário: string	- horário: string	- horário: string
- bairro: string	- bairro: string	- bairro: string	- bairro: string
- tipo: string	- tipo: string	- tipo: string	- tipo: string

Modelagem de anomalias mensais e anuais



Processo de análise

Anomalias anuais e mensais

Algorithm 2 *identify_and_classify_rules* (*anomalies_per_month*, *sup_pattern*, *conf_pattern*, *sup_classification*, *conf_classification*)

```
1: yearly_anomalies  $\leftarrow$  union(anomalies_per_month)
2: yearly_rules  $\leftarrow$  arules(yearly_anomalies, sup_pattern, conf_pattern)
3: expected_rules  $\leftarrow$  expected_rules(yearly_rules, sup_classification, conf_classification)
4: months  $\leftarrow$  get_month_list(anomalies_per_month)
5: monthly_rules  $\leftarrow$  []
6: for m in months do
7:   m_anomalies  $\leftarrow$  get_month_list(anomalies_per_month[m])
8:   m_rules  $\leftarrow$  arules(m_anomalies)
9:   monthly_rules  $\leftarrow$  append(m_rules)
10: end for
11: unexpected_rules  $\leftarrow$  unexpected_rules(monthly_rules, yearly_rules)
```

Regras esperadas

Algorithm 3 *expected_rules* (*general_rules*, *sup_min*, *lift_min*)

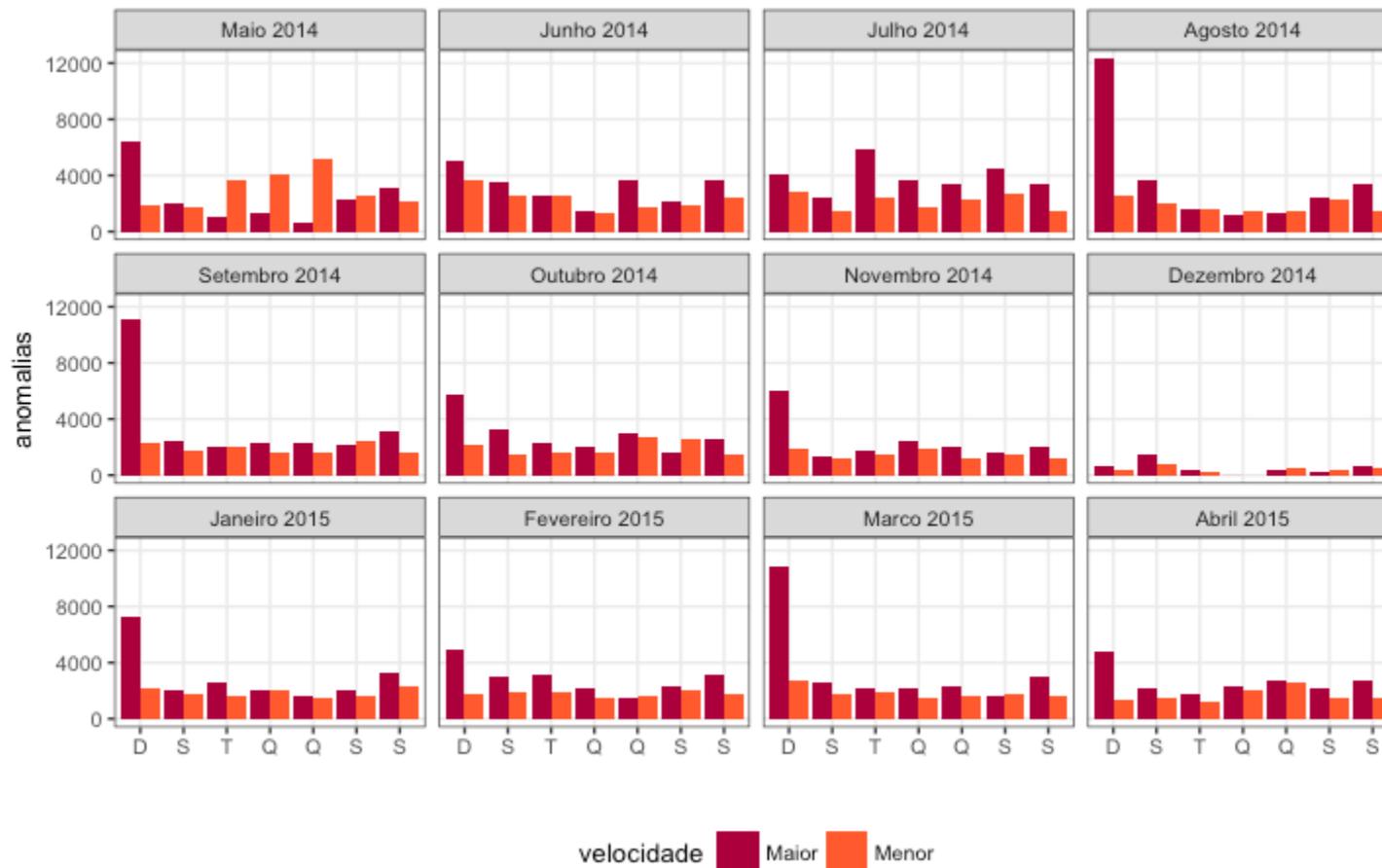
```
1: expected_rules  $\leftarrow$  []
2: for i in general_rules do
3:   supi  $\leftarrow$  get_support(i)
4:   lifti  $\leftarrow$  get_lift(i)
5:   nconditioni  $\leftarrow$  get_number_of_terms(i)
6:   if (nconditioni == 1 and supi  $\geq$  sup_min and lifti  $\geq$  lift_min) then
7:     expected_rules  $\leftarrow$  append(i)
8:   end if
9: end for
```

Regras não esperadas

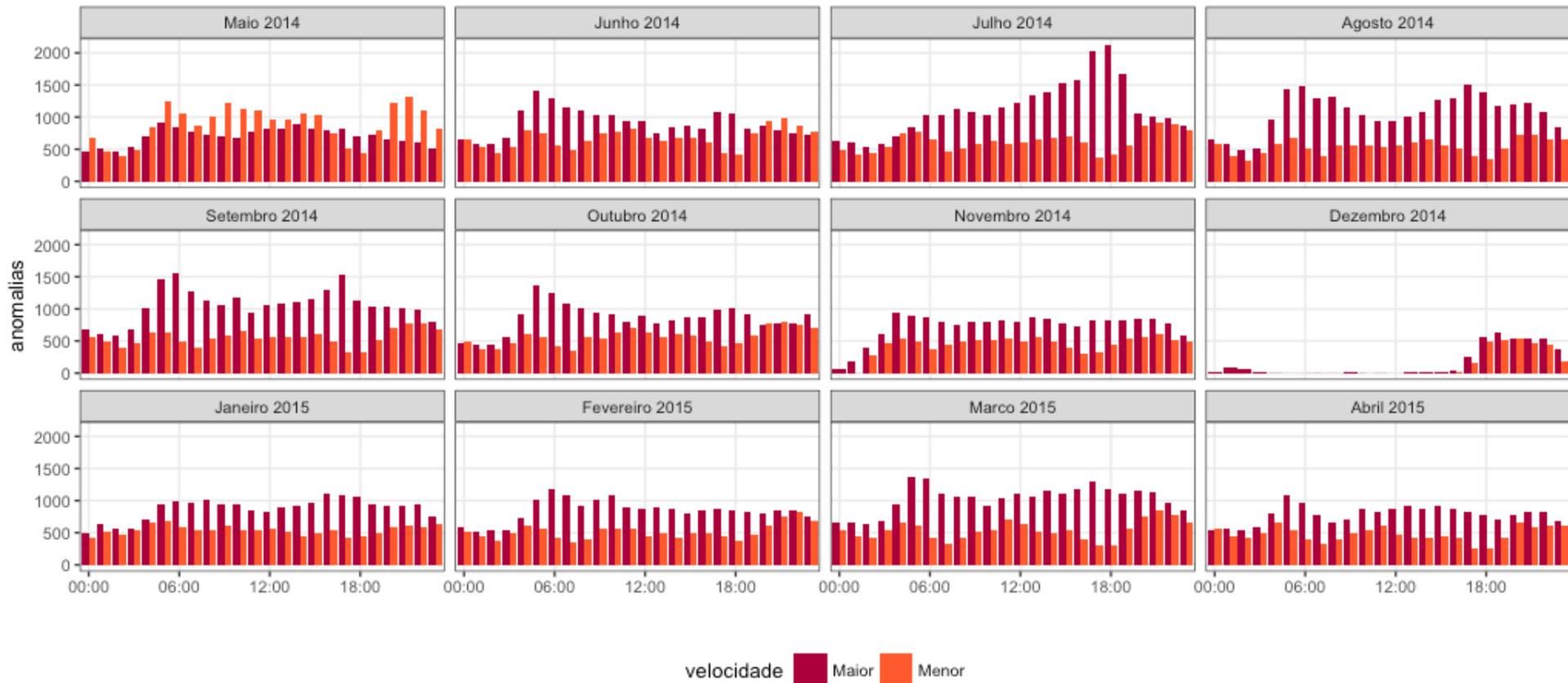
Algorithm 4 *unexpected_rules* (*specific_rules*, *general_rules*)

```
1: unexpected_rules  $\leftarrow$  []
2: for j in specific_rules do
3:   condj  $\leftarrow$  get_condition(j)
4:   conseqj  $\leftarrow$  get_consequence(j)
5:   expectCondj  $\leftarrow$  0
6:   expectConseqj  $\leftarrow$  0
7:   for i in general_rules do
8:     condi  $\leftarrow$  get_condition(i)
9:     nj  $\leftarrow$  get_number_of_terms(condj)
10:    weightj  $\leftarrow$  1/n
11:    for k in condj do
12:      if (condj[k] == condi) then
13:        expectCondj  $\leftarrow$  expectCondj + weightj
14:      end if
15:    end for
16:    if expectCondj > 0 then
17:      conseqj  $\leftarrow$  get_consequence(j)
18:      if (conseqj == conseqi) then
19:        expectConseqj  $\leftarrow$  1
20:      end if
21:    end if
22:    expectednessj  $\leftarrow$  (expectCondj + expectConseqj)/2
23:    unexpectednessj  $\leftarrow$  1 - expectednessj
24:    if unexpectednessj = 1 then
25:      unexpected_rules  $\leftarrow$  append(j)
26:    end if
27:  end for
28: end for
```

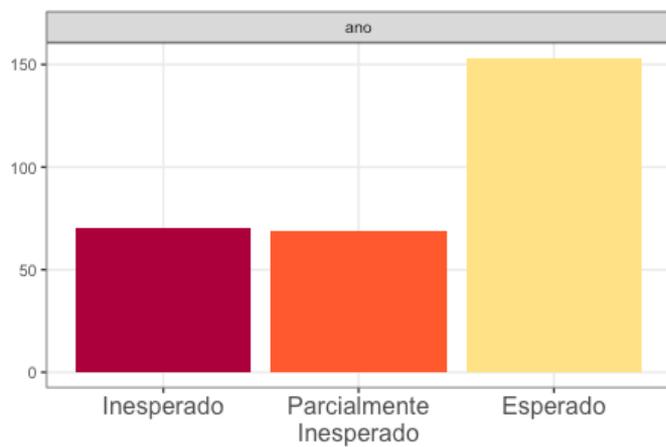
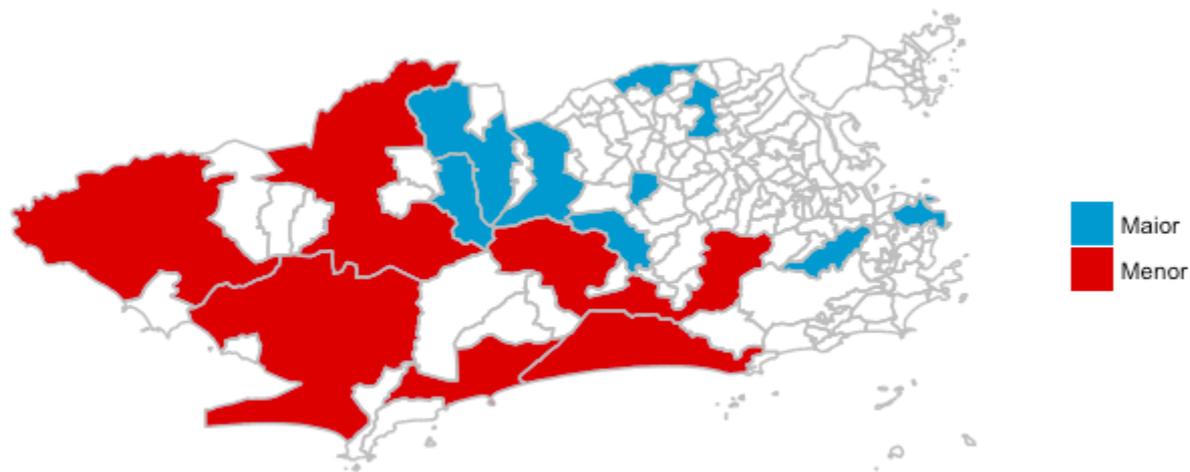
Anomalias por mês (por dia da semana)



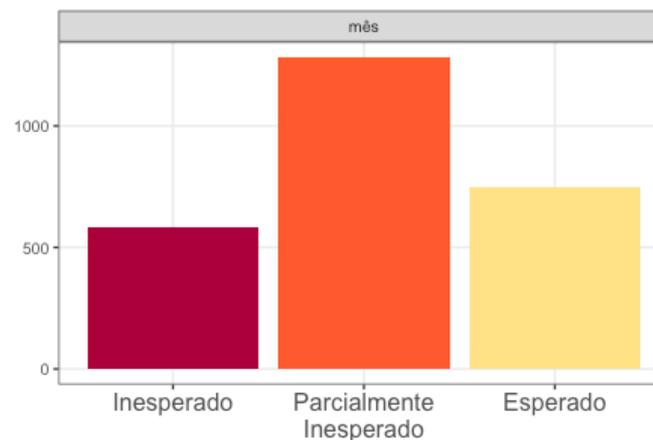
Anomalias por mês (por horário)



Regras produzidas



Por ano



Por mês

Conclusões

- Metodologia para identificar as anomalias no trânsito por meio de agregações espaço-temporais
- Classificação das regras de associação de anomalias como esperadas ou inesperadas
- A aplicação da identificação de regras esperadas:
 - Classifica anomalias como pontuais
 - Apoia o entendimento de comportamentos sistêmicos do trânsito
- Próximos passos
 - Explorar a semântica das regras
 - Em diferentes períodos



Contato



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

Agradecimentos

