



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

# Saúde do Trabalhador

## Padrões em Afastamentos e Desligamentos

---

Ivair Nobrega Luques

[Ivair.luques@eic.cefet-rj.br](mailto:Ivair.luques@eic.cefet-rj.br)

# Motivação

Fiocruz – Ensp - Cesteh

Cesteh – Centro de Saúde do Trabalhador e Ecologia Humana

Prover informação que apoie a tomada de decisão do Cesteh no planejamento de ações em defesa da saúde do trabalhador.

# Base de Dados

## RAIS – Relatório Anual de Informações Sociais

- Disponibilizada por ftp pelo MTPS.
- Informações agrupadas por Estado/Ano em arquivos txt
- Existem dados no período de 1985 a 2015
- Atualmente gera cerca de 70 milhões de registros/ano

# Estrutura do Arquivo

- 45 atributos
- Relevantes para o Cesteh neste momento
  - Causa 1º Afastamento /ano
  - Causa 2º Afastamento /ano
  - Causa 3º Afastamento /ano
  - Causa do Desligamento
  - Classificação Brasileira de Ocupações
  - Sexo
  - Faixa Etária
  - Raça / Cor
  - Faixa tempo de emprego
  - Número de Funcionários

# Objetivo deste Trabalho

- Identificar padrões de relação entre os diversos tipos de afastamento e algumas informações sobre o trabalhador, informadas na RAIS.
- Identificar padrões de relação entre os diversos tipos de desligamento e algumas informações sobre o trabalhador, informadas na RAIS.

# Análise Exploratória de Dados

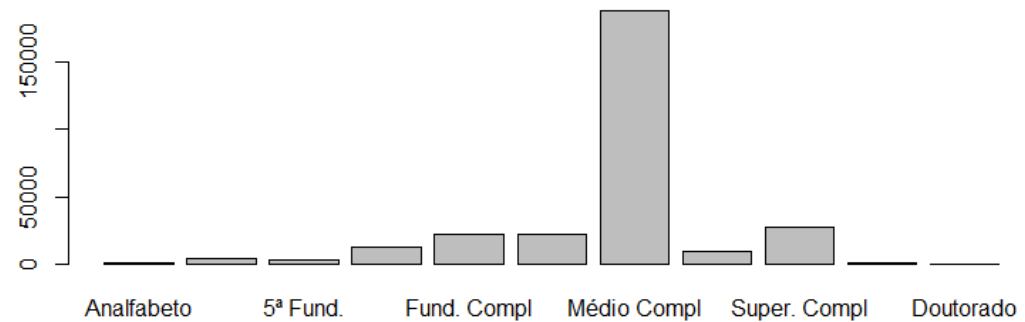


Fig 1 – Grau de Instrução

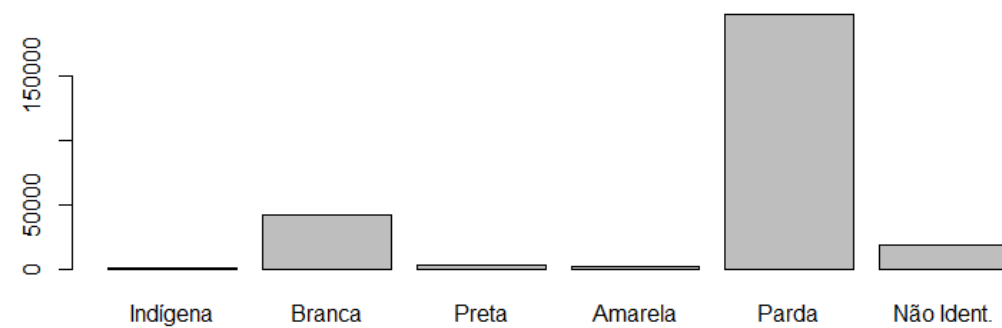


Fig 2 -Raca / Cor

# Análise Exploratória de Dados

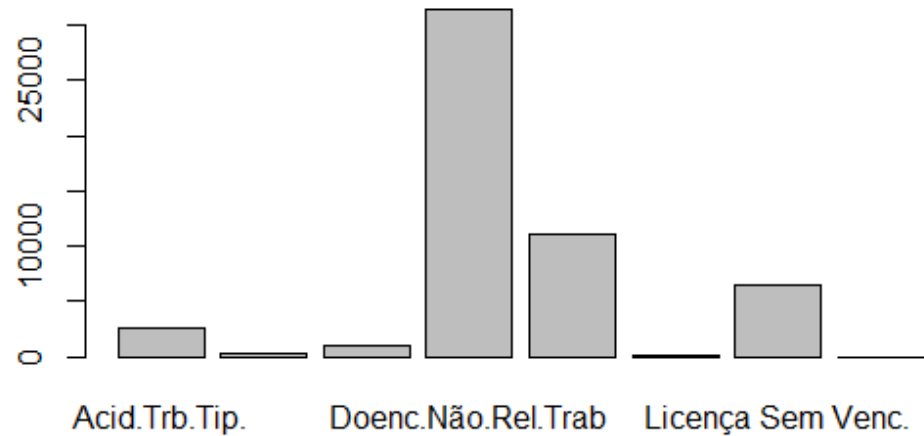


Fig 3 - Afastamentos

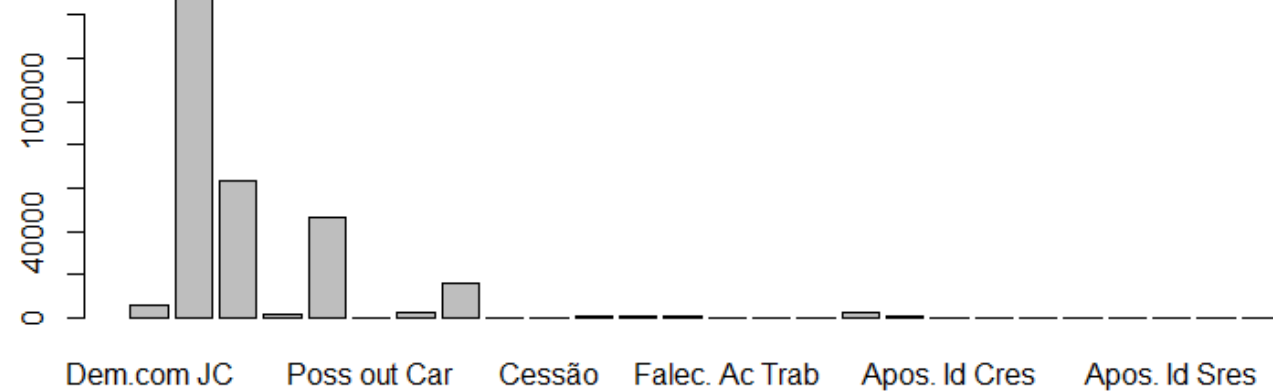


Fig 4 - Desligamentos

# Pré-Processamento

- Limpeza de Dados
  - Eliminar os registros que não contenham informações preenchidas de 1º afastamento e de 2º afastamento e de 3º afastamento e de desligamentos
- Redução de Dados
  - Redução de dimensionalidade -> Redução de atributos sem relevância para o Cesteh
- Adaptações nas observações
  - Gerar a partir de uma observação duas ou mais observações



## Dados de 2015 no Amazonas

MotivDesliga	sexo	raca	Idade	Escolar
Dem.sem JC :150.203	masc.:197.565	Indígena : 1.024	30 a 39:110.099	Médio Compl :212.407
Term. Contr. : 63.051	fem. :135.208	Branca : 47.682	18 a 24: 74.241	Super. Compl : 35.608
Desl.sem JC : 46.232		Preta : 3.501	25 a 29: 66.240	Fund. Compl : 25.245
Trans s/ Onus: 16.003		Amarela : 2.492	40 a 49: 50.346	Médio Incompl.: 24.658
Dem.com JC : 5.907		Parda :224.446	50 a 64: 27.238	6ª a 9ª Fund. : 14.648
(Other) : 9.735		Não Ident.: 20.083	65 ou +: 2.703	Super.Incompl.: 10.759
NA's : 41.642		NA's : 33.545	(Other): 1.906	(Other) : 9.448

NumFunc	Afastamento	CBO
250 a 499:51.034	Sem Afasta. :279.621	521 : 30.214
50 a 99 :44.965	Doenc.Não.Rel.Trab: 31.367	411 : 25.627
500 a 999:39.804	Licença Mater. : 11.039	731 : 25.193
1000 ou +:38.644	Licença Sem Venc. : 6.524	514 : 21.358
100 a 249:31.045	Acid.Trb.Tip. : 2.655	513 : 14.278
(Other) :78.852	Doenc.Relac.Trab : 1.097	(Other):216.069
NA's :48.429	(Other) : 470	NA's : 34

# Mineração de Dados

- Será utilizada a abordagem de Padrões Frequentes
  - ✓ Algoritmo APRIORI - Identificar Regras de Associação

Regras de Associação : Encontrar elementos que implicam na presença de outros elementos numa mesma transação.

1ª Etapa : Encontrar os itemsets frequentes , com um valor de suporte  $\geq$  suporte mínimo

2ª Etapa: Gerar as regras a partir dos itemsets, com um valor de confiança  $\geq$  confiança mínima

- Parametrização: Suporte 0,1   Confiança 0,5

# Resultados Afastamentos

Afastamentos 332.773 observações

suporte	confiança	lift
Min. :0.1001	Min. :0.5158	Min. :0.874
1st Qu.:0.1164	1st Qu.:0.6012	1st Qu.:1.006
Median :0.1519	Median :0.6517	Median :1.079
Mean :0.2042	Mean :0.6660	Mean :1.109
3rd Qu.:0.2514	3rd Qu.:0.7240	3rd Qu.:1.141
Max. :0.6629	Max. :1.0000	Max. :1.813

Regra de Associação	Suporte	confiança	Lift
sexo=fem. => Afastamento=Licença Mater	0,1315096	1	1,11812948
sexo=masc. raca=Parda, Escolar=Médio Compl => Afastamento=Doenc.Não.Rel.Trab	0,1550459	0,7966167	1,13498827
Idade=50 a 64 => Afastamento=Doenc.Não.Rel.Trab	0,1039095	0,7796443	1,13211226
sexo=masc., Escolar=Médio Compl, Afastamento=Doenc.Não.Rel.Trab => raca=Parda	0,1550459	0,7729319	1,11659017
Idade=30 a 39, Afastamento=Doenc.Não.Rel.Trab => raca=Parda	0,1400888	0,7408218	1,1174663

# Resultados Desligamentos

Desligamentos 291.131 observações

support	confidence	lift
Min. :0.1022	Min. :0.5104	Min. :0.8909
1st Qu.:0.1221	1st Qu.:0.6104	1st Qu.:1.0037
Median :0.1472	Median :0.6518	Median :1.0354
Mean :0.2043	Mean :0.6481	Mean :1.0725
3rd Qu.:0.2489	3rd Qu.:0.6996	3rd Qu.:1.0720
Max. :0.6799	Max. :0.7373	Max. :2.5995

Regra de Associação	Suporte	confiança	Lift
Idade=30 a 39,Escolar=Médio Compl => MotivDesliga=Dem.sem JC	0,1226733	0,5677992	1,1005369
sexo=fem., Idade=18 a 25 => MotivDesliga=Dem.sem JC	0,1166279	0,5819721	1,1280076
sexo=masc.,Escolar=Sup Compl => MotivDesliga=Dem.sem JC	0,2154975	0,5511068	1,0681829
sexo=masc.,raca=Parda,Escolar=Médio Compl => MotivDesliga=Dem.sem JC	0,1580113	0,5709144	1,106575
TempoEmpr=01 => MotivDesliga=Term.	0,1335756	0,5629904	2,5995458

## Conclusões

- ✓ O uso de padrões frequentes é uma ferramenta útil na identificação de associações entre os dados e o grau de dependência entre os mesmos.
- ✓ Algumas das associações obtiveram destaque em razão do significativo volume de dados de determinado atributo.
- ✓ Existem inúmeras demandas do CESTEh que seriam atendidas com a aplicação de padrões frequentes

## Trabalhos futuros

- ✓ Realizar os experimentos deste trabalho com maior um número de dados.  
Ex: Informação de todos os estados no ano de 2015
- ✓ Realizar os experimentos deste trabalho com outros algoritmos de padrões frequentes , por exemplo o ECLAT
- ✓ Realizar Agrupamentos de afastamentos e desligamentos por CBO
- ✓ Identificar regras de associações apenas entre afastamentos e desligamentos

# Saúde do Trabalhador

## Padrões em Afastamentos e Desligamentos

Ivair Nobrega Luques  
[Ivair.luques@eic.cefet-rj.br](mailto:Ivair.luques@eic.cefet-rj.br)