

# Análise Exploratória da Malária no Brasil por Meio da Plataforma de Ciência de Dados Aplicada à Saúde

Lais Ribeiro Baroni<sup>1</sup>, Balthazar Paixão<sup>1</sup>, Alvaro Chrispino<sup>1</sup>, Gustavo Guedes<sup>1</sup>, Christovam Barcellos<sup>2</sup>, Marcel Pedroso<sup>2</sup>, Eduardo Ogasawara<sup>1</sup>

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

<sup>2</sup>Fiocruz - Fundação Oswaldo Cruz | Ministério da Saúde

# Introdução

- A malária é uma doença infecciosa causada por parasitas protozoários do gênero *Plasmodium* transmitida a partir da picada do mosquito vetor
- Os países tropicais e subtropicais constituem a área endêmica da doença
- A Amazônia Legal - que compreende os estados do AC, AP, AM, MT, PA, RO, RR, TO e MA- é a região mais susceptível à malária no Brasil
- Dentro da Amazônia Legal, a ocorrência da doença não é homogênea, variando de acordo com algumas características como fatores naturais, fatores geográficos e condições sociais

# Introdução

- Embora a taxa de incidência de malária no Brasil não seja tão alta como em alguns países africanos, é importante que esforços sejam feitos para o seu combate no Brasil
- Desde do início de 2018, a Organização Pan-Americana da Saúde (OPAS) já alertava quanto a um aumento de casos de malária na região das Américas e alertou quanto ao risco iminente de surtos
- O Relatório Mundial da Malária da OMS (2018) mostrou um aumento de 25% de casos de malária com relação ao ano anterior
- Um dos motivos atribuídos ao aumento é a negligência nos cuidados contra a malária

# Plataforma de Ciência de Dados Aplicada à Saúde (PCDAS)

- Criada para apoiar ao planejamento, monitoramento e avaliação de políticas públicas e serviços de saúde
- Desenvolvida pelo Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (Icict), da Fundação Oswaldo Cruz (Fiocruz)
- Conjunto de estratégias, ferramentas e técnicas para coleta, transformação e análise de grandes quantidades de dados voltados para saúde pública

# Plataforma de Ciência de Dados Aplicada à Saúde (PCDAS)

- Opções de acesso:
  - **Análise Visual:** disponibiliza visualização dos dados tratados pela equipe da Plataforma através de gráficos interativos
  - **Mineração de Dados e Análise Preditiva:** disponibiliza o R Studio Server para o usuário manipular e trabalhar com os dados
  - **Data Science Lab:** onde pode-se conhecer os projetos dos parceiros da Plataforma



# Plataforma de Ciência de Dados Aplicada à Saúde (PCDAS)

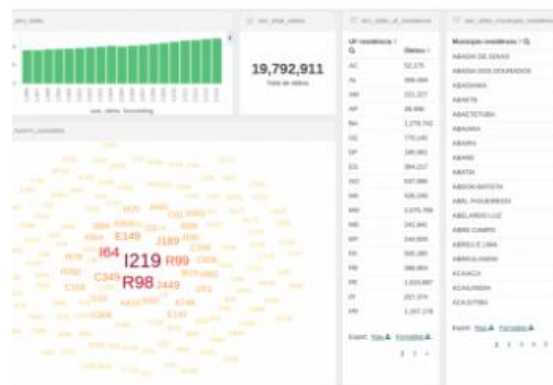
- Análise Visual de Dados

## Galeria visual de dados



### Nascimento

Dados do Sistema de Informações sobre Nascidos Vivos (SINASC) do Ministério da Saúde.



### Mortalidade

Dados do Sistema de Informações sobre Mortalidade (SIM) do Ministério da Saúde.



### Cadastro Nacional de Estabelecimentos de Saúde

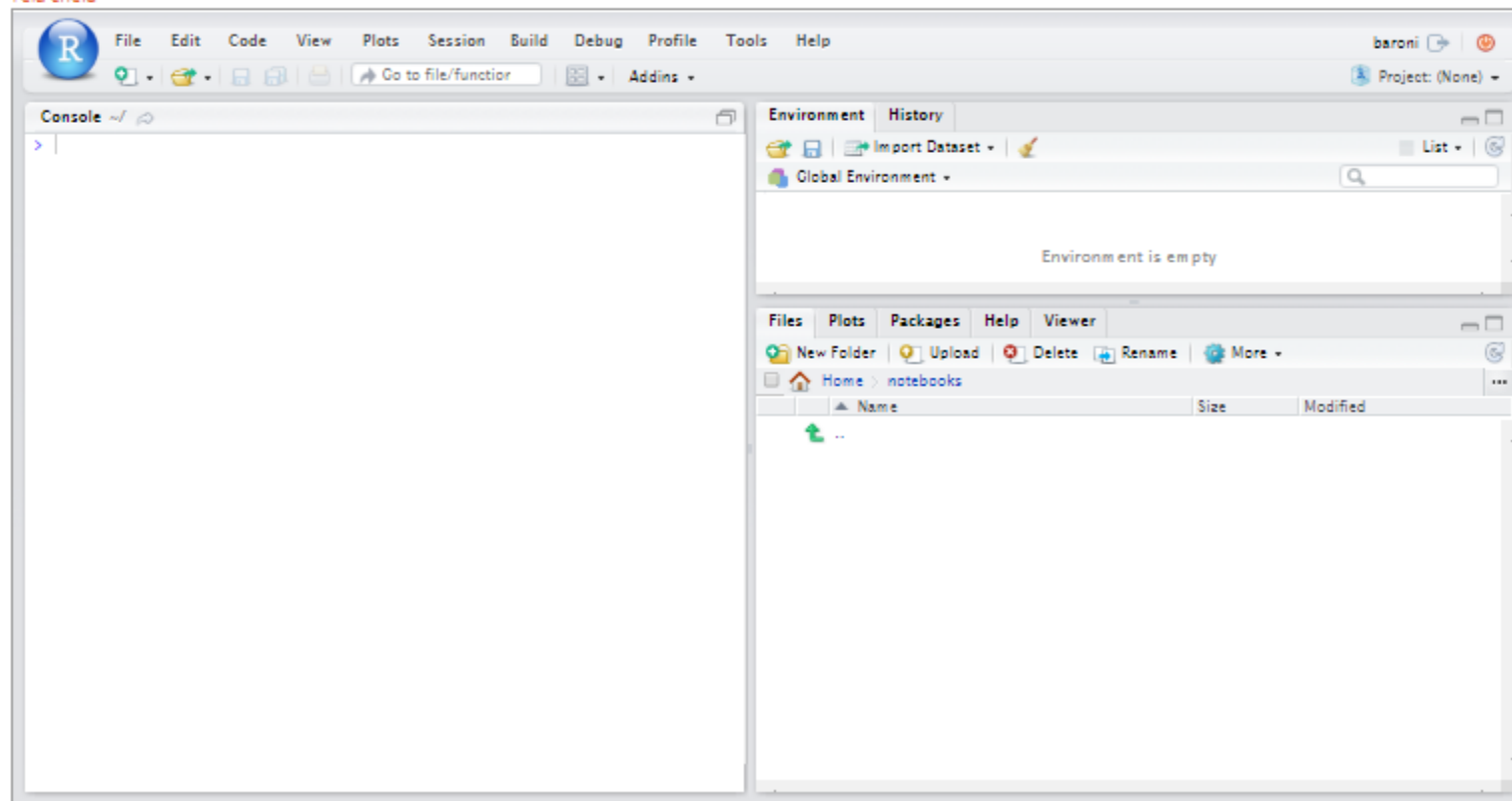
Dados do Cadastro Nacional de Estabelecimentos de Saúde (CNE) reunidos pela Plataforma de Ciência de Dados aplicada à Saúde (PCDaS).

# Plataforma de Ciência de Dados Aplicada à Saúde (PCDAS)

- Mineração de Dados e Análise Preditiva

## RStudio Server

Tela cheia



# Plataforma de Ciência de Dados Aplicada à Saúde (PCDAS)

- O serviço disponibilizado é gratuito e aberto para qualquer pessoa
- Condição: disponibilizar o código – mantendo o ideal de ciência aberta para todos
- Infraestrutura tecnológica:

## **2 servidores para gestão do cluster**

4 Processadores Intel Xeon E5-2630 v3 com 8 Núcleos e 2.4GHz cada  
256 GB memória RDIMM total (128 GB por servidor)

## **4 servidores para armazenamento de dados**

40 Terabytes de armazenamento  
8 Processadores Intel Xeon E5-2630 v3 com 8 Núcleos e 2.4GHz cada  
64 GB memória RDIMM total (16 GB por servidor)

**10 Gigabit conexão internet**



# Proposta

- Realizar uma análise holística dos dados de malária na Amazônia Legal a partir dos recursos disponíveis na PCDaS
- Conduzir uma análise exploratória preliminar direcionada para o Sistema de Informações de Vigilância Epidemiológica módulo malária (SIVEP-Malária) entre os anos de 2009 a 2015
- A partir dessa análise, levantar perguntas de pesquisas que podem ajudar na compreensão e no combate da malária no Brasil

# Fontes de Dados

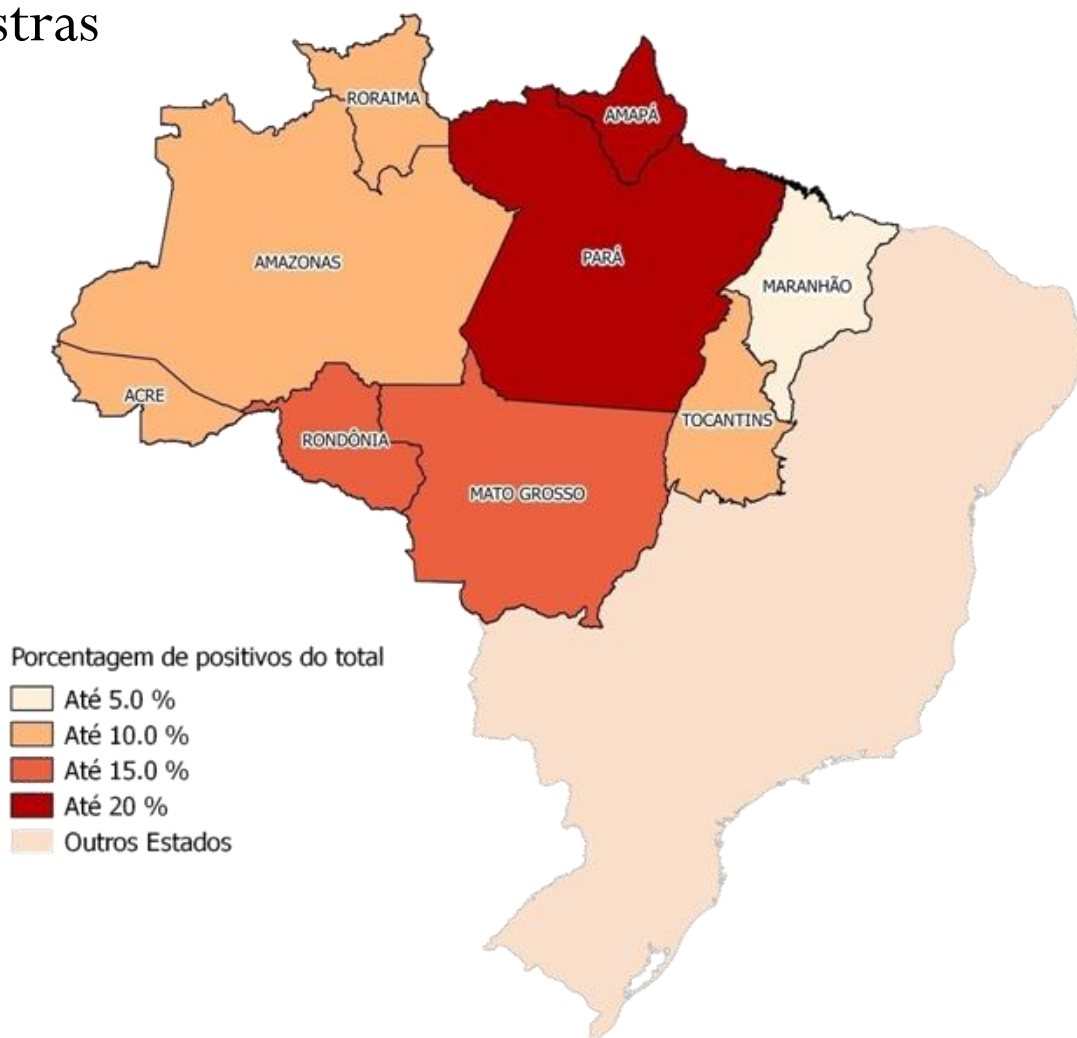
- SIVEP-Malária
  - Objetivo: monitoramento e administração da malária
  - Implantado em 2003
  - Período considerado: 2009 a 2015
  - 30 atributos
  - 15.764.287 registros, sendo cerca de 12% casos positivos de malária
- Censo do IBGE 2010

# Metodologia

- Pré-processamento de dados
  - Objetivo: melhorar a qualidade dos dados, seja em exatidão, integridade, consistência ou interpretabilidade
  - Técnicas : limpeza de dados (remoção de valores inconsistentes) e transformação de dados (valores traduzidos de código para texto)
- Integração de dado
  - dados do Censo do IBGE
  - Objetivos: melhor observar o efeito da malária de acordo com as características da população
  - A integração de dados enriquece a análise, possibilitando uma visão diferenciada das ocorrências, podendo colocar a análise exploratória em um contexto diferenciado
- Análise Exploratória

# Análise Exploratória dos Dados

- Percentagem de amostras que confirmaram a suspeita de malária

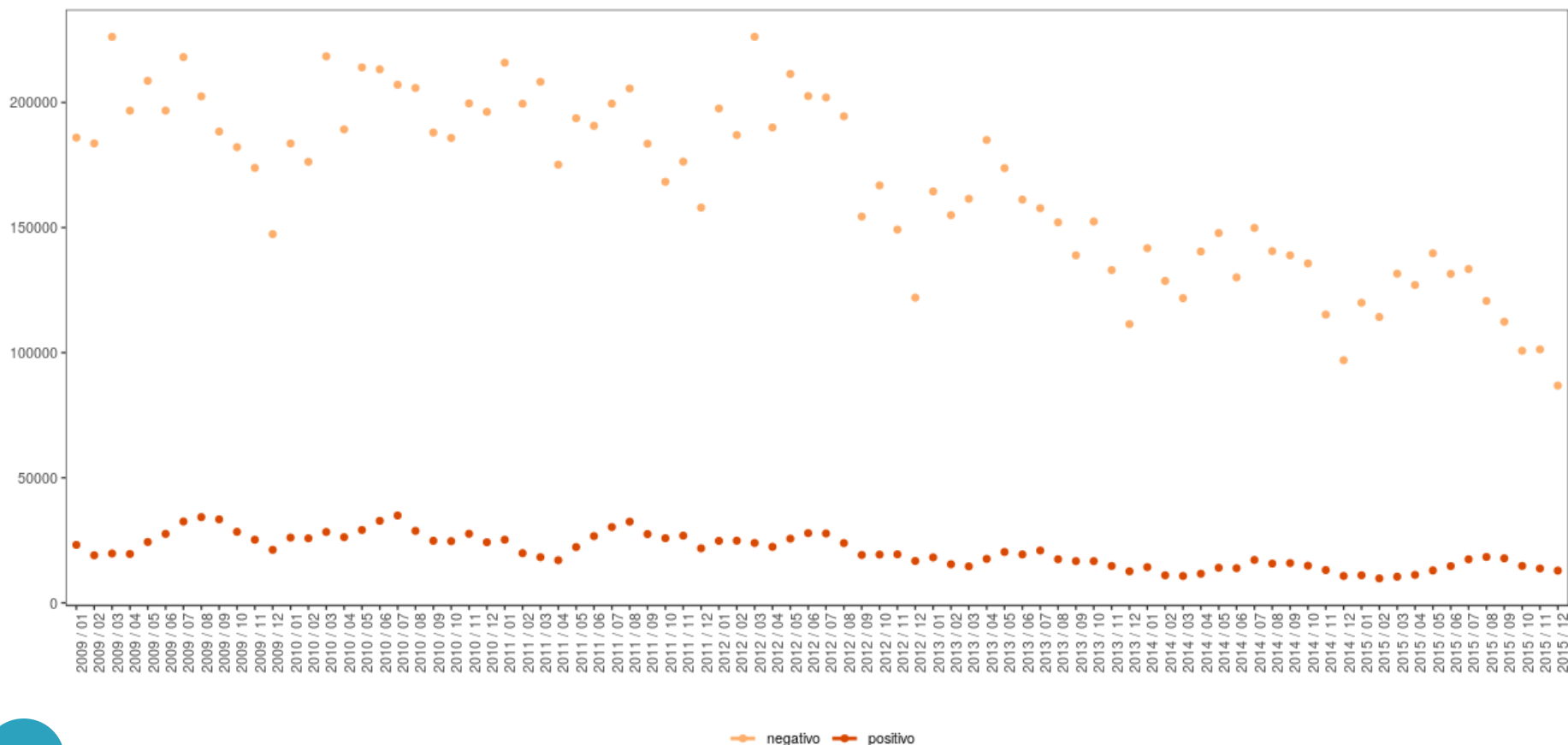


# Análise Exploratória dos Dados

- Disparidade entre os estados na proporção número de casos de malária por número de atendimentos
- Perguntas:
  - Os estados que apresentam proporções mais altas fazem pouco exame?
  - Os estados que apresentam proporções baixas atingem os grupos de risco?

# Análise Exploratória dos Dados

- Evolução da quantidade de casos positivos e negativos

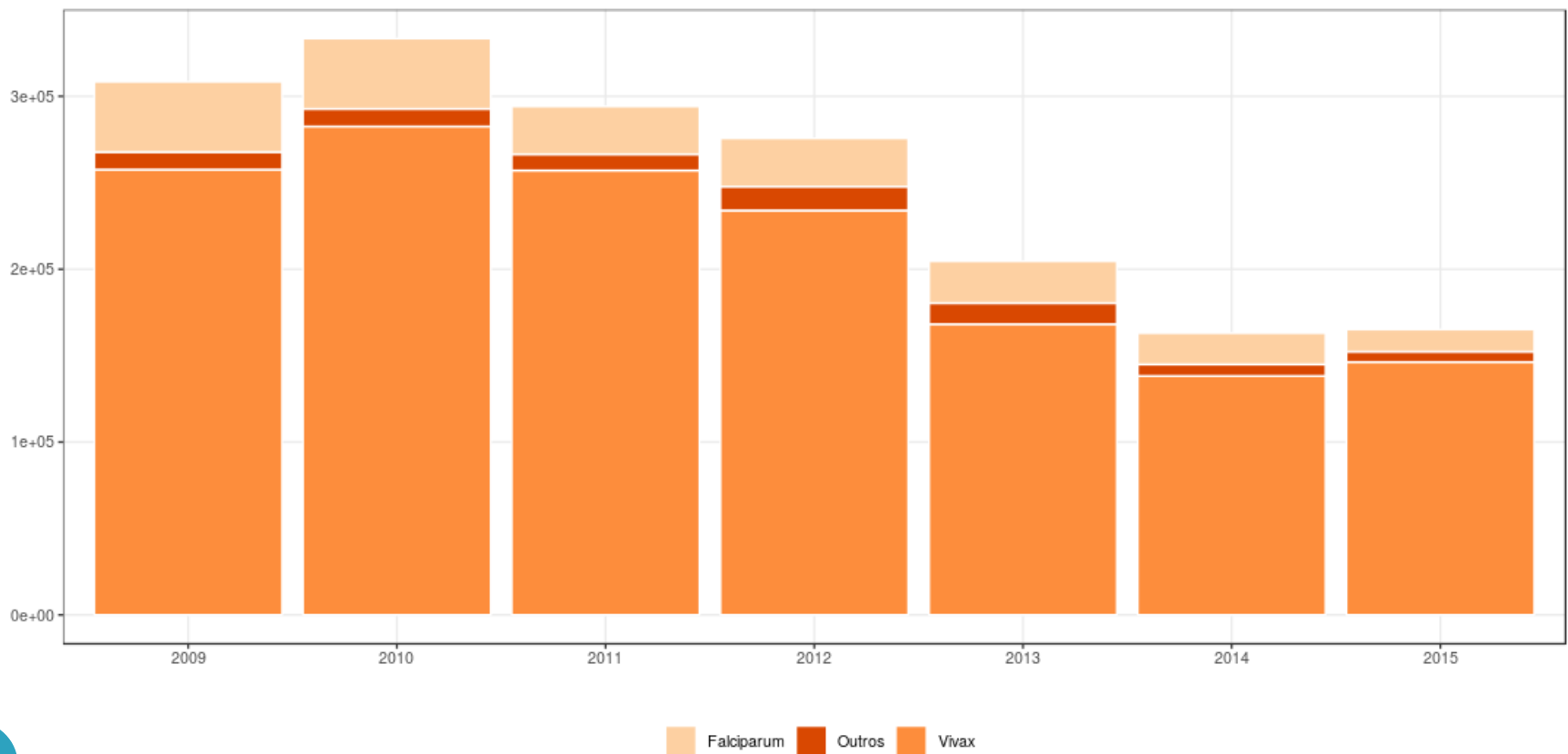


# Análise Exploratória dos Dados

- A prevalência malária está relacionada a eventos de chuva, mais frequentes no verão
- No entanto, ao observar o gráfico percebemos que há tanto mais exames feitos (refletidos nos pontos plotados para casos negativos) quanto mais ocorrências de casos positivos de malária durante os meses do inverno.
- Perguntas:
  - Quais os fatores envolvidos no aumento de número de exames feitos durante o inverno?
  - Quais os fatores envolvidos no aumento de número de casos de malária durante o inverno?

# Análise Exploratória dos Dados

- Distribuição do tipo de malária nos sete anos de estudo



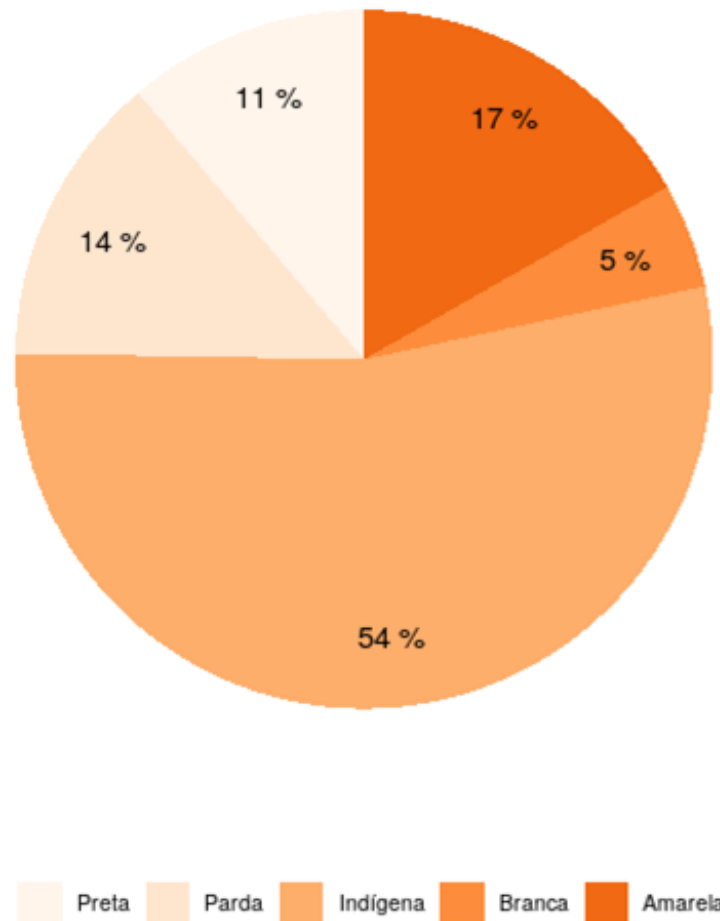


# Análise Exploratória dos Dados

- Perguntas
  - Há algum padrão envolvido na ocorrência de cada tipo de malária?
  - Casos de malária para os outros tipos de plasmódio que não Vivax estão localizados em regiões específicas da Amazônia Legal?
  - Os casos de malária de outros tipos são provenientes de casos importados para o Brasil?

# Análise Exploratória dos Dados

- Proporções dos índices criados pela divisão da quantidade de indivíduos afetados pela malária segundo o SIVEP pela quantidade absoluta de indivíduos segundo o censo do IBGE de 2010, para cada raça



# Análise Exploratória dos Dados

- Gráfico é relevante para enfatizar a importância da integração de dados
- Caso a quantidade de indivíduos de cada raça não fosse considerada segundo os dados do IBGE, o gráfico representaria as porcentagens dos números absolutos de casos notificados e, dessa forma, a raça parda predominaria (com 70%) e perderíamos a noção da maior incidência entre a população indígena
- Perguntas:
  - Quais os principais fatores envolvidos na predominância de casos de malária na raça indígena?
  - Os esforços para atendimento preventivo e curativos estão sendo suficientes para essa população?

# Análise Exploratória dos Dados

- Proporção de casos na Amazônia Legal agrupados por gênero

Gênero	AC	AP	AM	MA	MT	PA	RO	RR	TO
Masculino	56%	62%	59%	64%	77%	62%	64%	61%	72%
Feminino	44%	38%	41%	36%	23%	38%	36%	39%	28%

# Análise Exploratória dos Dados

- Perguntas:
  - Os homens estão mais expostos às áreas com alta incidência do mosquito vetor?
  - A atividade ocupacional do indivíduo influencia nessa proporção de gênero?
  - A variação das proporções homem/mulher variam de estado para estado pela relação com a disponibilidade de trabalho em cada estado?

# Análise Exploratória dos Dados

- Associação entre o tipo de detecção e a percepção dos sintomas para casos positivos de malária

Sintoma	Deteccção Ativa	Deteccção Passiva
Sim	84%	98%
Não	16%	2%

# Análise Exploratória dos Dados

- Perguntas:
  - A detecção ativa é importante para encontrar (e, posteriormente, diagnosticar e tratar) indivíduos que apresentam casos assintomáticos de malária?
  - Quais os motivos que levam à procura de atendimento médico por demanda espontânea quando o paciente não sente sintoma?
    - Isto está relacionado à questões trabalhistas?
    - Isto pode estar relacionado à casos onde o paciente vive em uma região endêmica?

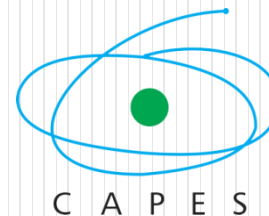
# Conclusão

- A análise exploratória feita na Plataforma viabilizou a elaboração de perguntas de pesquisa
- Todas as considerações levantadas na análise exploratória tratam-se de conjecturas ou hipóteses
  - A comprovação pode se dar por uma análise mais aprofundada ou integrada desses dados



# Conclusão

- Muitas das perguntas levantadas trazem aspectos espaço-temporais como *proxy* para um melhor entendimento da malária no Brasil
  - Tal fato requer a integração com fontes adicionais de dados, como, por exemplo, informações de regiões de saúde
- Pretendemos fazer uso no PCDaS de técnicas de mineração de dados como, por exemplo, padrões frequentes
- Idealiza-se a possibilidade de trabalhar com dados mais recentes do SIVEP-malária



# Análise Exploratória da Malária no Brasil por Meio da Plataforma de Ciência de Dados Aplicada à Saúde

Lais Ribeiro Baroni<sup>1</sup>, Balthazar Paixão<sup>1</sup>, Alvaro Chrispino<sup>1</sup>, Gustavo Guedes<sup>1</sup>, Christovam Barcellos<sup>2</sup>, Marcel Pedroso<sup>2</sup>, Eduardo Ogasawara<sup>1</sup>

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

<sup>2</sup>Fiocruz - Fundação Oswaldo Cruz | Ministério da Saúde

Email para contato: [lais.baroni@eic.cefet-rj.br](mailto:lais.baroni@eic.cefet-rj.br)