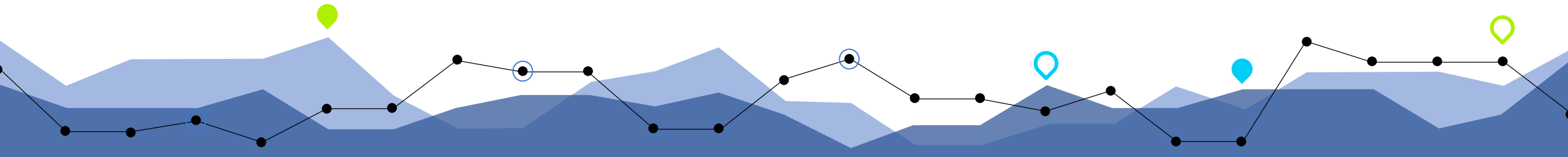


# Estatística e Introdução à Bioestatística

IMD0920/DEB0512/DIM0122 TI na Saúde  
Profs Drs. Julio Melo, Edgard Faria, Beatriz  
Stranski, Anna Giselle, Fabrícia Cavalcanti

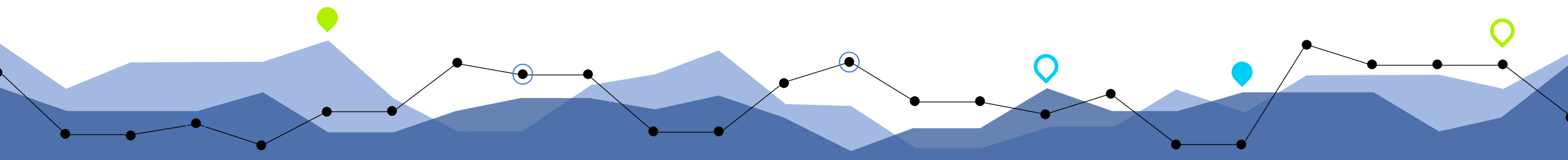
# Roteiro

- **Estatística**
  - **Funções de Densidade de Probabilidade**
- **Bioestatística**
- **Hands-on Zika virus database**
- **Inferência estatística**



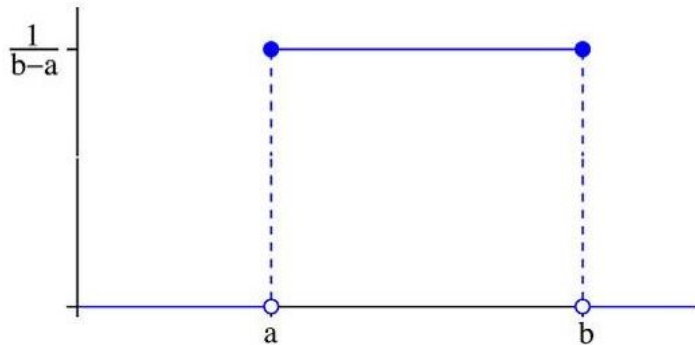
# Estatística

- **Função de densidade de probabilidade**
- Função gerada pelo histograma das variáveis aleatórias, onde a amplitude representa a probabilidade da variável ser igual aquele valor. Definido por  $P(x = x_0)$  ou  $f(x_0)$ .
- **Distribuição de variáveis aleatórias**
- Funções de densidade de probabilidade que são largamente conhecidas na ciencia.
- Distribuição uniforme
- Distribuição binomial
- Distribuição normal
- Distribuição de Poisson
- etc



# Estatística

- **Distribuição de probabilidade para variáveis aleatórias**



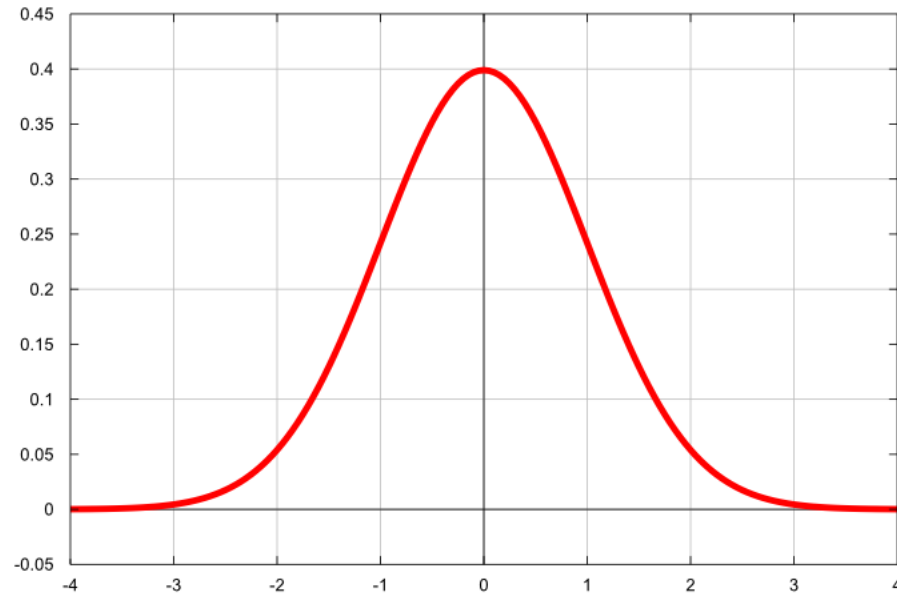
$$\bar{X} = \frac{a + b}{2}$$

$$S^2 = \frac{1}{12} (b - a)^2$$

$$P(x=x_0) = \begin{cases} \frac{1}{b-a} & \text{se } a < x_0 < b \\ 0 & \text{caso contrário} \end{cases}$$

# Estatística

- **Distribuição de probabilidade para variáveis aleatórias**



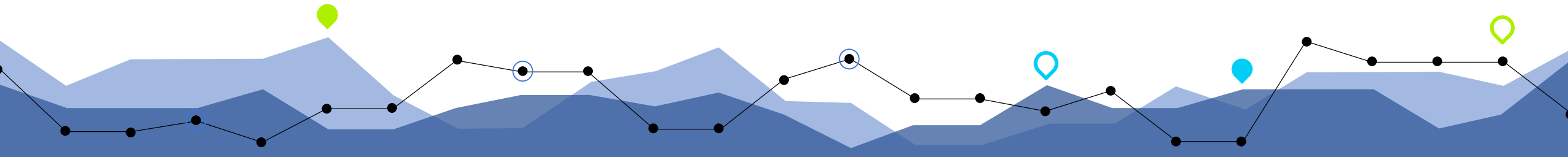
$$\bar{X} = \mu$$

$$S^2 = \sigma^2$$

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\mu-x)^2}{2\sigma^2}}$$

# Bioestatística

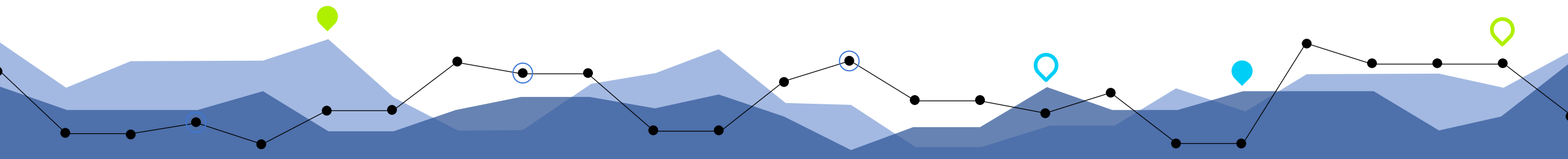
- **A bioestatística é caracterizada pelos dados e não pela ciencia**
- **Bases de dados biológicos (que na maioria das vezes são numéricas).**
- **Medias, medianas, desvio padrão e variancia passam a ter significado biologico ao invés de puramente numérico.**
- **Estudos passam a requerer maior espaço amostral devido à grande diversidade e aleatoriedade dos processos biológicos.**



## Bioestatística

- **Base de dados Zika Virus – Epidemic<sup>1</sup>**
- Reporta alguns casos reportados de Zika Virus a partir de 2015.
- Base contem informações como local, quantidade de casos reportados, período do report e outros.

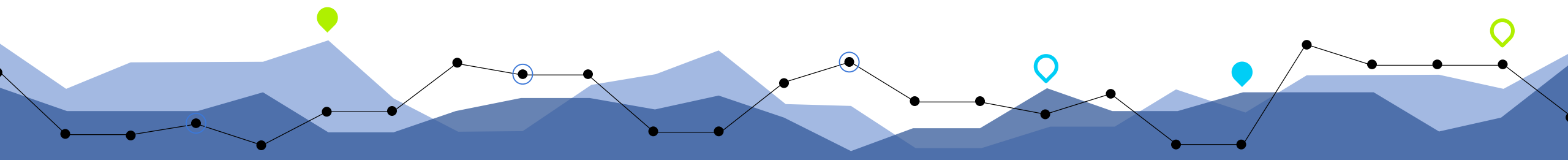
Fonte 1: <https://www.kaggle.com/cdc/zika-virus-epidemic>



# Bioestatística

- **Base de dados Zika Virus – Epidemic<sup>1</sup>**
- report\_date
- location/location\_type
- data\_field
- value, unit

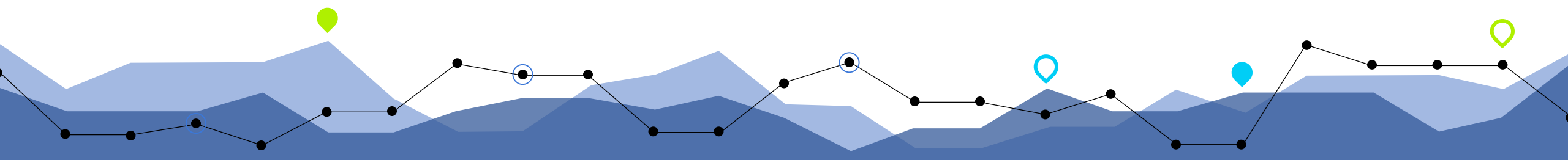
Fonte 1: <https://www.kaggle.com/cdc/zika-virus-epidemic>





## Bioestatística- Exercícios

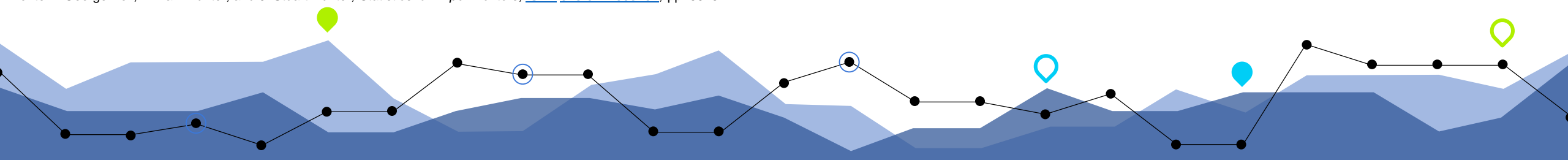
1. Mostre, através de histogramas, os países com mais reports na base de dados.
2. Construa 3 histogramas que mostre, por Mes, as incidencias de Zika nos países onde houve maior incidencia de Zika no ano de 2016.
3. Mostre, através de Box Plot, os numérios de casos por região no Brasil. (Use os estados como sendo as amostras, para cada boxplot).
4. Crie um histograma que mostre a incidencia total de casos no brasil por estado e por mes. Através desse histograma construa a função de densidade de probabilidade que mostra a distribuição de casos de zika por estado e por mes.



## Teste estatístico

- **Testes usados para determinar se determinados dados descrevem ou não determinada característica:**
  - Hipótese  $H_0$  vs  $H_1$
  - Probabilidade de que, em uma determinada população, a característica testada seja observada na maioria dos subconjuntos amostrais possíveis.
  - Na maioria dos casos a distribuição *pai* de onde as amostras vem, é considerada com sendo normal, o que é uma suposição válida pelo teorema do limite central<sup>1</sup>.

Fonte 1: George Box, William Hunter, and J. Stuart Hunter, *Statistics for Experimenters*, [ISBN 978-0471093152](https://doi.org/10.1002/9780471093152), pp. 66–67.

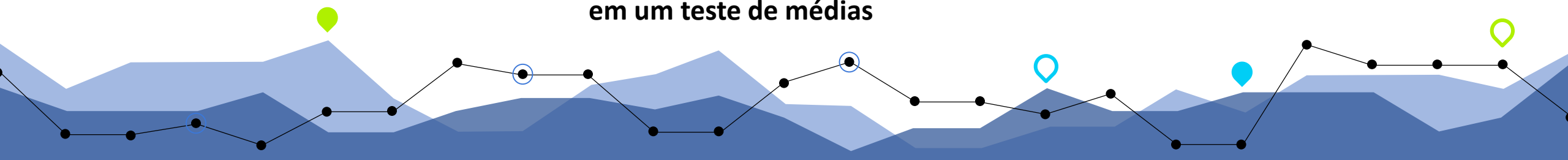


## Teste estatístico

- **Teste t-student**
- Usa uma distribuição bem conhecida pra testar variáveis dentro de um determinado intervalo.
- Define um valor de  $p$  que indica a probabilidade de erro dentro do grupo amostral escolhido.

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

Equação para determinação de  $t$ ,  
em um teste de médias



## Teste estatístico

1. Calcule a média do aparecimento de Zika no Brasil no ano de 2016, em cada estado brasileiro reportado na base de dados do Zika Virus. Calcule, através desses valores, a média de casos por 100 mil habitantes em cada estado.
2. Verifique, através de teste estatístico se os valores calculados condizem com a média *real* proposta pelo ministério da saúde<sup>1</sup> para cada região do Brasil: Centro Oeste 113,4 casos/100 mil habitantes, seguida do Nordeste (53,5); Sudeste (41,4); Norte (36,0); Sul (6,1).

Fonte 1: <http://www.brasil.gov.br/saude/2016/04/saude-divulga-pimeiro-balanco-com-casos-de-zika-no-pais>

