

Estatística Descritiva

Prof. Dr. José Henrique Hildebrand **Grisi** Filho

Universidade de São Paulo

Roteiro da aula

- ◊ Medidas de tendência central
- ◊ Medidas de dispersão
- ◊ Tabelas
- ◊ Gráficos

Medidas de tendência central (de localização ou de posição)

- ◊ média aritmética (ou simplesmente média)
- ◊ mediana
- ◊ moda

Média aritmética

- ◆ mais comum
- ◆ somam-se todos os valores e divide-se pelo número total de observações
- ◆ desvantagem: é influenciada por “outliers” (dados discrepantes)
- ◆ é apropriada se a distribuição dos dados é simétrica

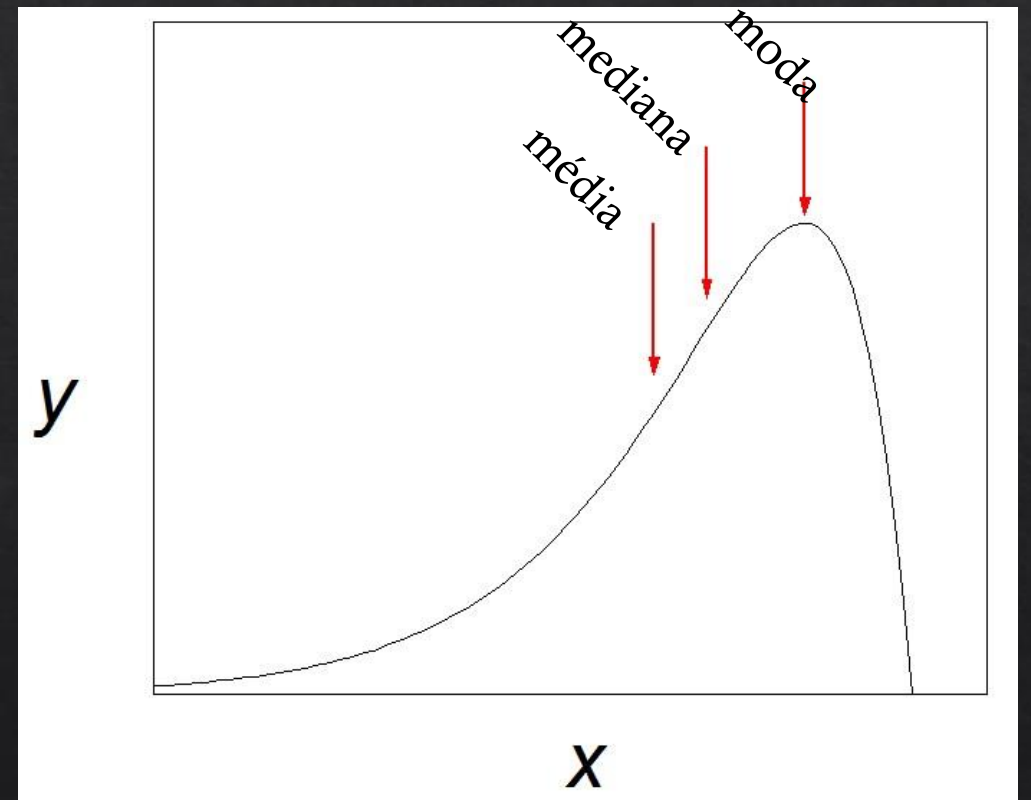
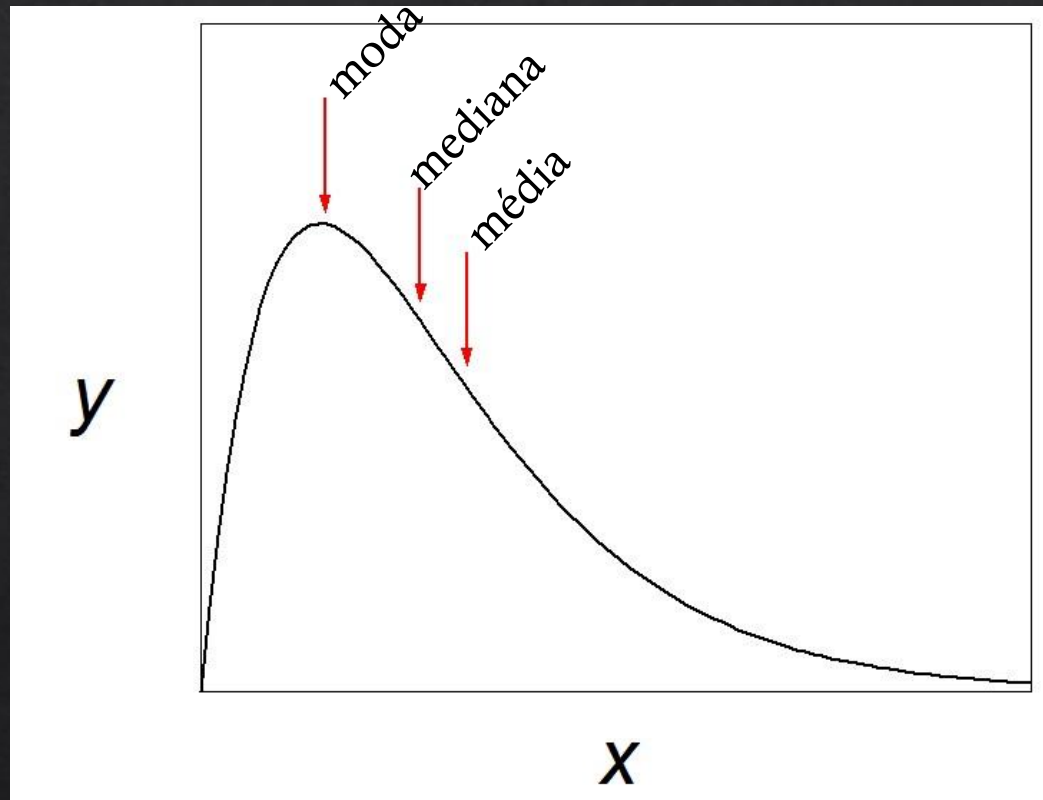
Mediana

- ◇ é o valor central em um conjunto de N observações colocadas em ordem de magnitude crescente (ou decrescente)
- ◇ mesmo número de observações acima e abaixo da mediana
- ◇ vantagem da mediana: não é afetada por “outliers” ou se a distribuição de dados é assimétrica

Moda

- ♦ Bastante conhecida, pouco utilizada
- ♦ É a observação mais comum (que ocorre com maior frequência) em um conjunto de dados
- ♦ Desvantagem: a moda é determinada sem se considerar a maioria das observações
- ♦ Há distribuições com mais de uma moda

Distribuições assimétricas



Medidas de dispersão (ou espalhamento)

- ◊ Amplitude
- ◊ Amplitude interquartil
- ◊ Variância
- ◊ Desvio Padrão

Amplitude

- ◇ Diferença entre o maior e o menor dado observado
- ◇ Simples de calcular
- ◇ Não é uma boa medida de dispersão, porque seu cálculo se baseia nos valores extremos da amostra e não em todos os dados

Amplitude interquartil

- ◊ $Q=Q3-Q1$: É a amplitude de valores que abrange os dados centrais (50%) das observações (diferença entre o terceiro e o primeiro quartil)
- ◊ não é influenciada pela presença de outliers
- ◊ ignora a maioria das observações

Variância

- ◊ Variância mede a dispersão dos dados em torno da média
- ◊ Utiliza todas as observações
- ◊ É uma medida sensível de dispersão

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{N - 1}$$

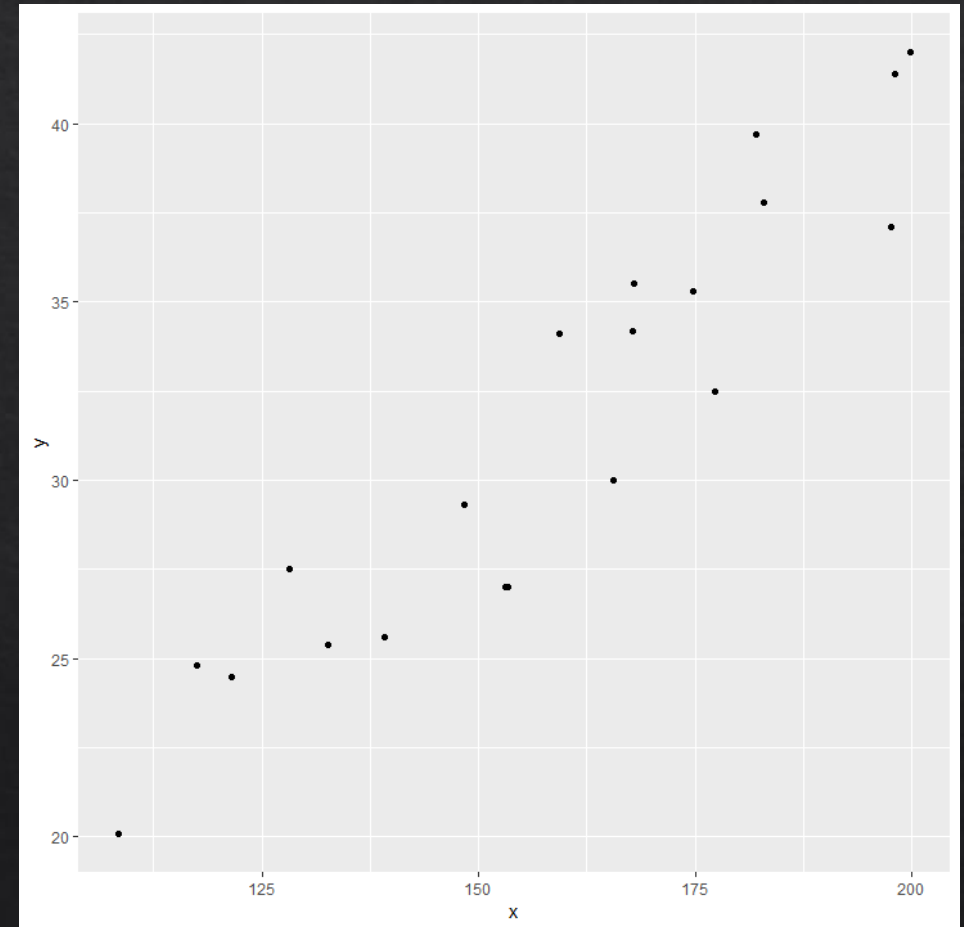
Desvio padrão (s)

- ◇ Raiz quadrada da variância
- ◇ apresenta a mesma dimensão que os dados originais
- ◇ muito útil para distribuições simétricas (p.ex. gaussiana)

Tabelas e Gráficos

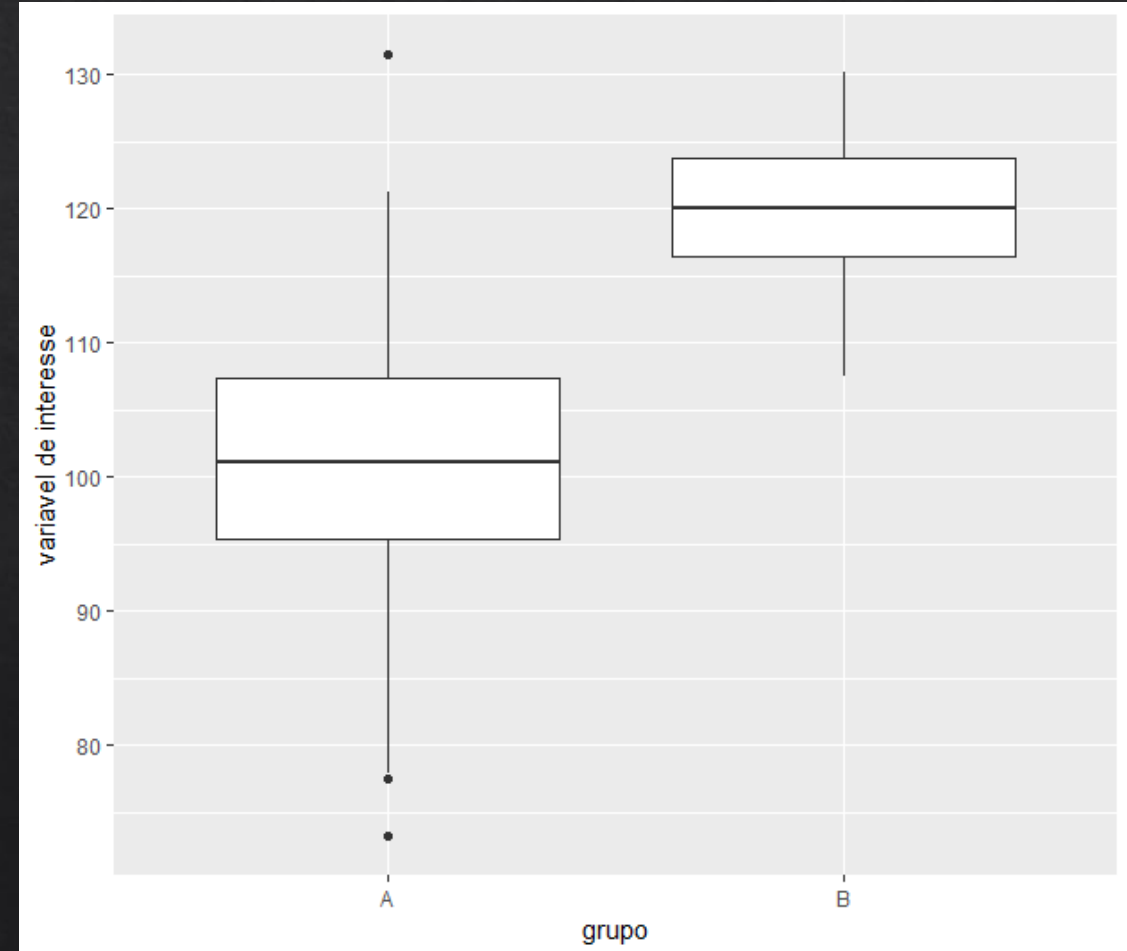
id	x	y
A	165.6	30.0
B	197.6	37.1
C	121.5	24.5
D	174.7	35.3
E	153.1	27.0
F	117.5	24.8
G	139.2	25.6
H	182.0	39.7
I	153.4	27.0
J	148.4	29.3
K	199.9	42.0
L	182.9	37.8
M	128.2	27.5
N	177.3	32.5
O	132.6	25.4
P	108.4	20.1
Q	198.0	41.4
R	167.8	34.2
S	159.3	34.1
T	168.0	35.5

id	x	y
P	108.4	20.1
C	121.5	24.5
F	117.5	24.8
O	132.6	25.4
G	139.2	25.6
E	153.1	27.0
I	153.4	27.0
M	128.2	27.5
J	148.4	29.3
A	165.6	30.0
N	177.3	32.5
S	159.3	34.1
R	167.8	34.2
D	174.7	35.3
T	168.0	35.5
B	197.6	37.1
L	182.9	37.8
H	182.0	39.7
Q	198.0	41.4
K	199.9	42.0



Tabelas e Gráficos

Grupo	Média	Desvio Padrão	Mediana	IQR
A	100,8	9,8	101,1	11,9
B	119,9	5,5	120,0	7,4



Tabelas X Gráficos

Tabelas

- ◇ Dados precisos
- ◇ Grande esforço mental para comparar coisas e encontrar padrões

Gráficos

- ◇ Dados pouco precisos
- ◇ Pouco esforço mental para comparar coisas ou encontrar padrões

Tabelas

- ◇ Dicas para construção de uma tabela:
 - ◇ título (informativo, conciso, sem ambigüidade)
 - ◇ cabeçalho curto para cada linha e coluna
 - ◇ incluir unidades de medida
 - ◇ números com grau de acurácia adequado
 - ◇ é mais fácil olhar informações dispostas em colunas do que em linhas

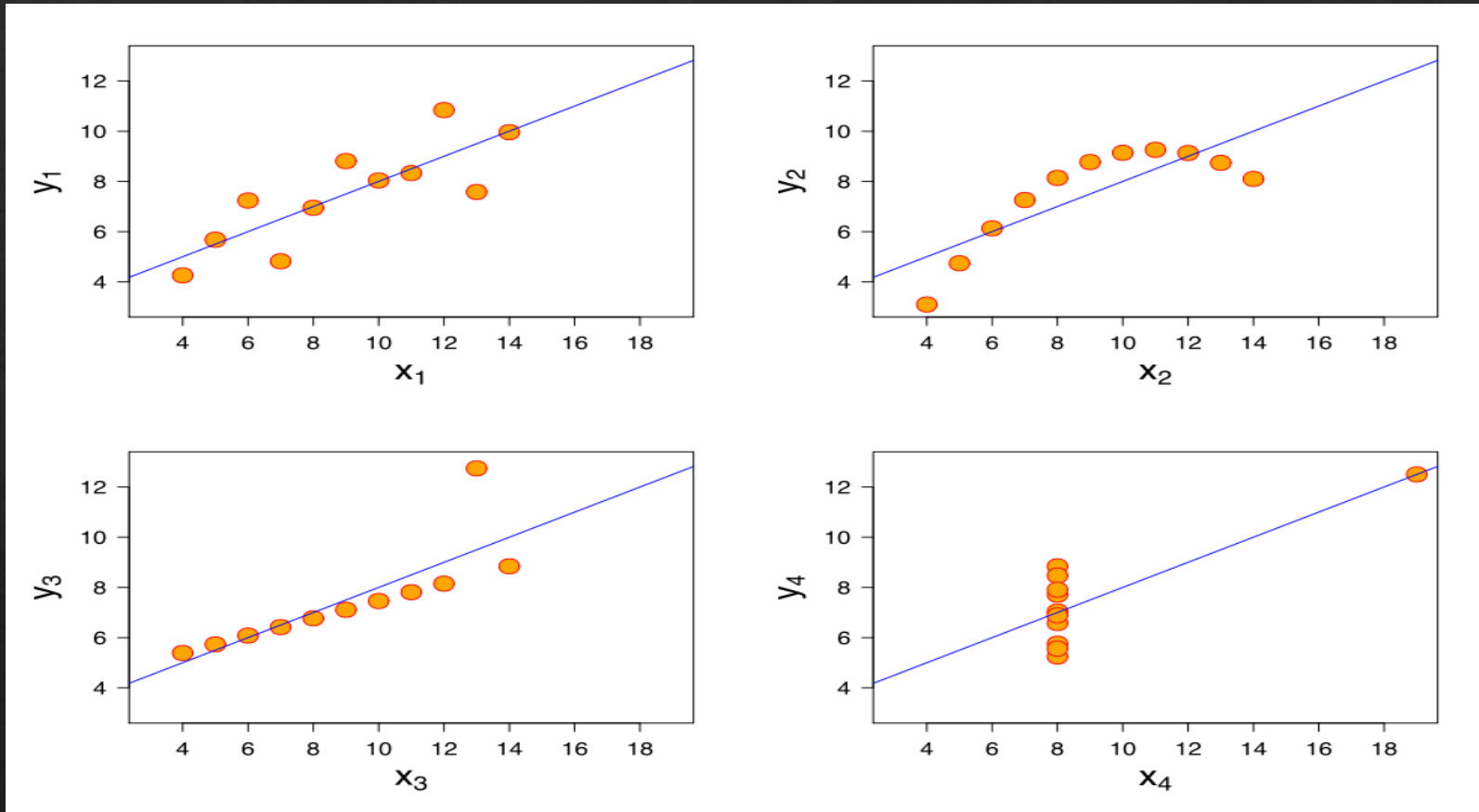
Tabelas

Tabela 1 – Número e percentual de alimentos segundo a origem, incriminados em surtos de doenças transmitidas por alimentos (Paraná, 1998)

Origem do alimento	N	%
Mista	72	66,1
Animal	32	29,4
Vegetal	5	4,6
Total	109	100,0

(Zoli JA *et al.*, *Higiene Alimentar*, vol. 16, n. 95, pp. 62-71 (2002))

Quarteto de Anscombe



<http://blog.revolutionanalytics.com/2017/05/the-datasaurus-dozen.html>

https://en.wikipedia.org/wiki/Anscombe%27s_quartet

Gráficos

- ◆ Dicas para fazer um gráfico:
 - ◆ simplicidade
 - ◆ título (informativo, conciso, sem ambigüidade)
 - ◆ nomear eixos, segmentos, barras, usando legendas que descrevam os símbolos usados
 - ◆ apresentar as unidades
 - ◆ a escolha do tipo de gráfico deve assegurar que toda informação relevante seja mostrada

Gráficos - Dados qualitativos

- ◆ Para dados qualitativos, cada observação pertence a uma dentre várias categorias ou classes distintas.
- ◆ São utilizados, entre outros:
 - gráfico de barras (comprimento)
 - gráfico de setores (ângulo)

Gráfico de setores (pizza, torta)

Figura 1 - Porcentagem de surtos de salmonelose e respectivos alimentos incriminados (Paraná 1996-1997)

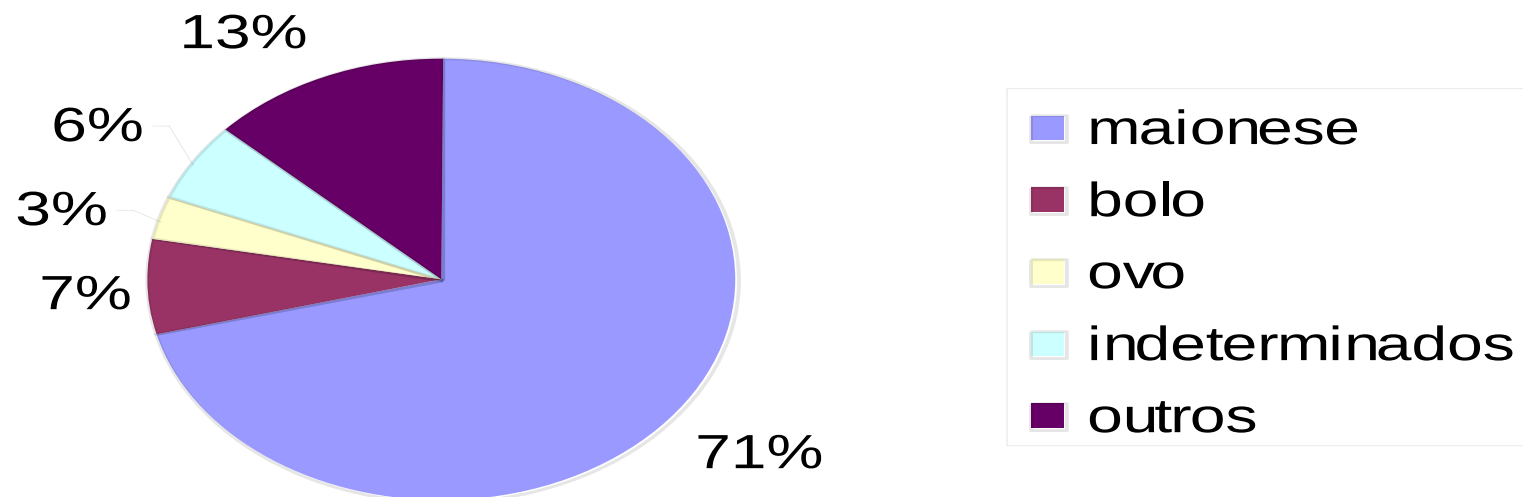
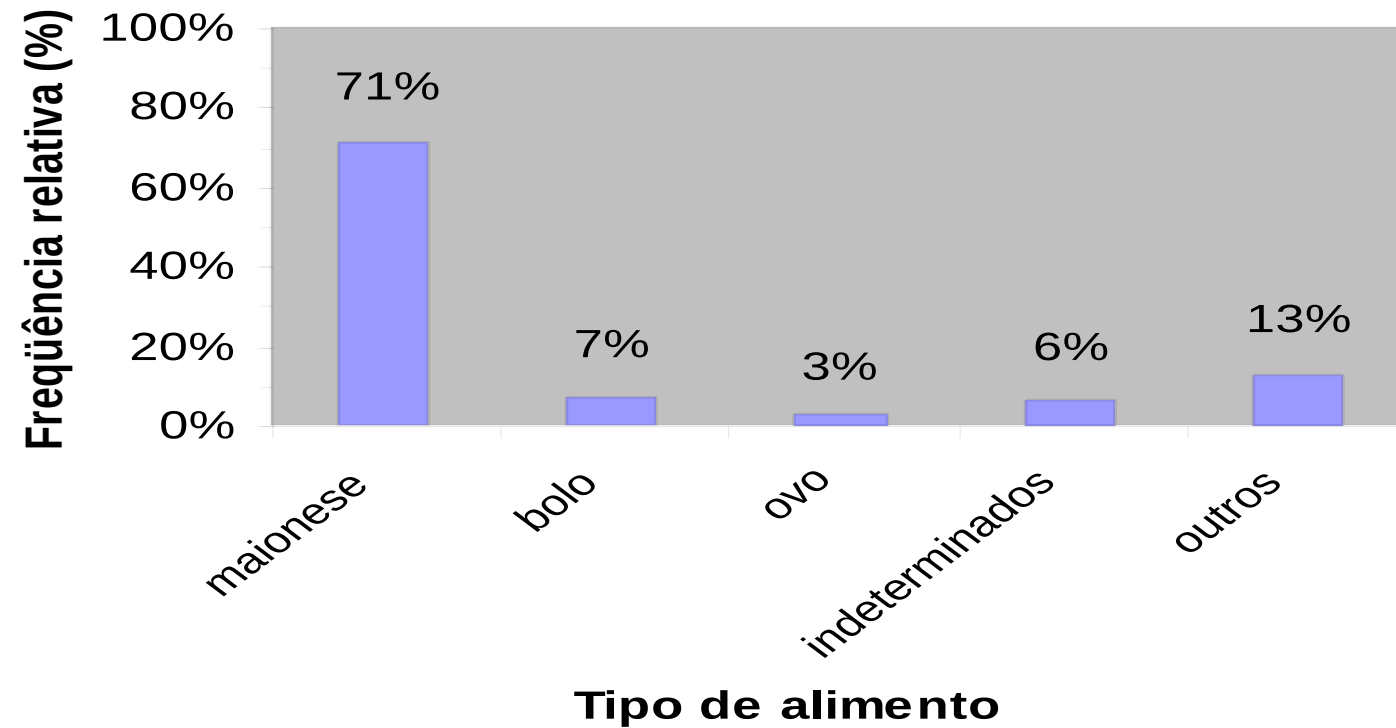


Gráfico de barras

Figura 2- Porcentagem de surtos de salmonelose e respectivos alimentos incriminados (Paraná 1996-1997)



Pizza X Barras

Figura 1 - Porcentagem de surtos de salmonelose e respectivos alimentos incriminados (Paraná 1996-1997)

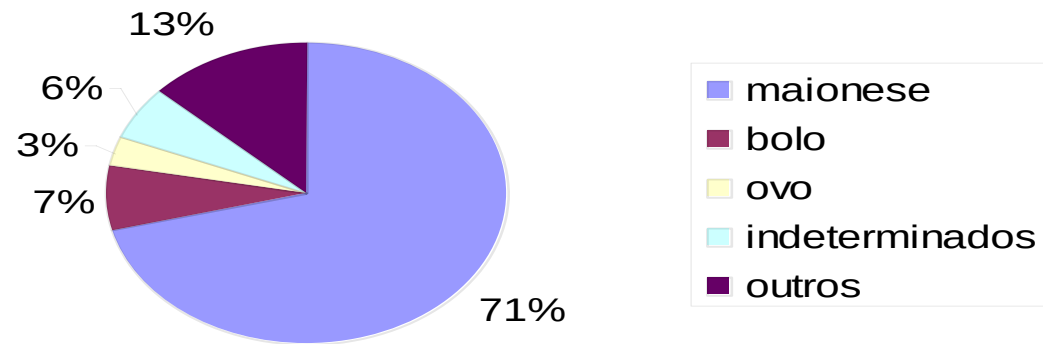
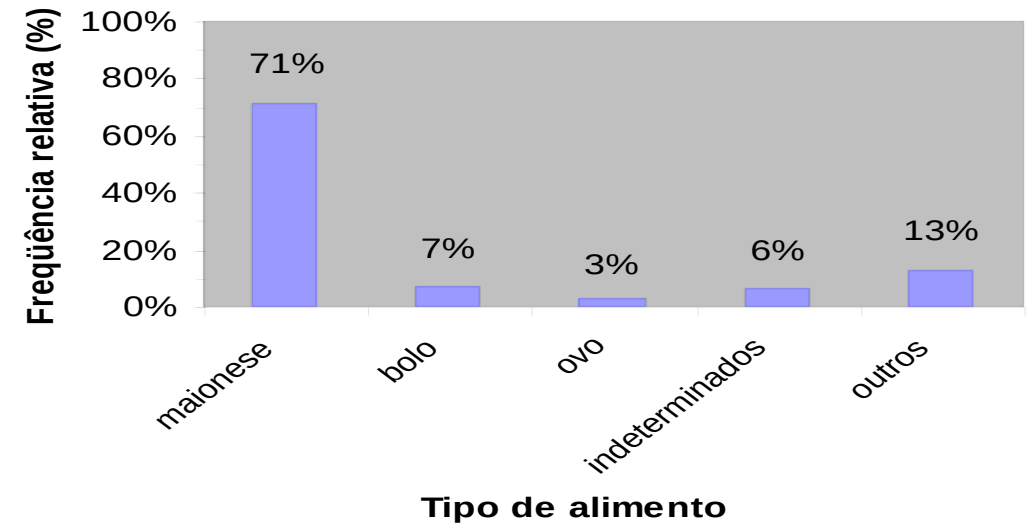
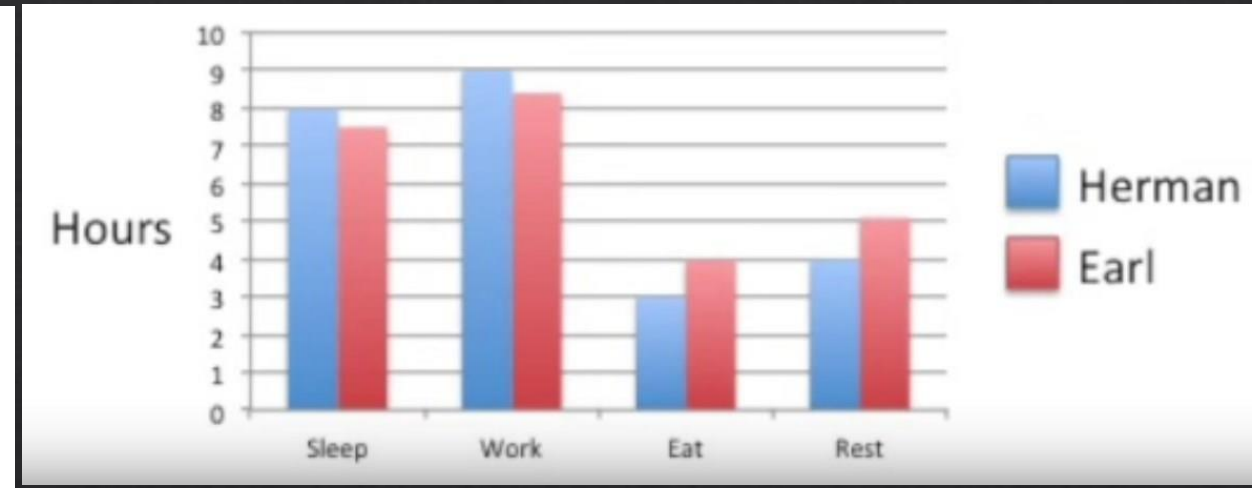


Figura 2- Porcentagem de surtos de salmonelose e respectivos alimentos incriminados (Paraná 1996-1997)





"Bar Charts Are Better than Pie Charts"

StatQuest with Josh Starmer

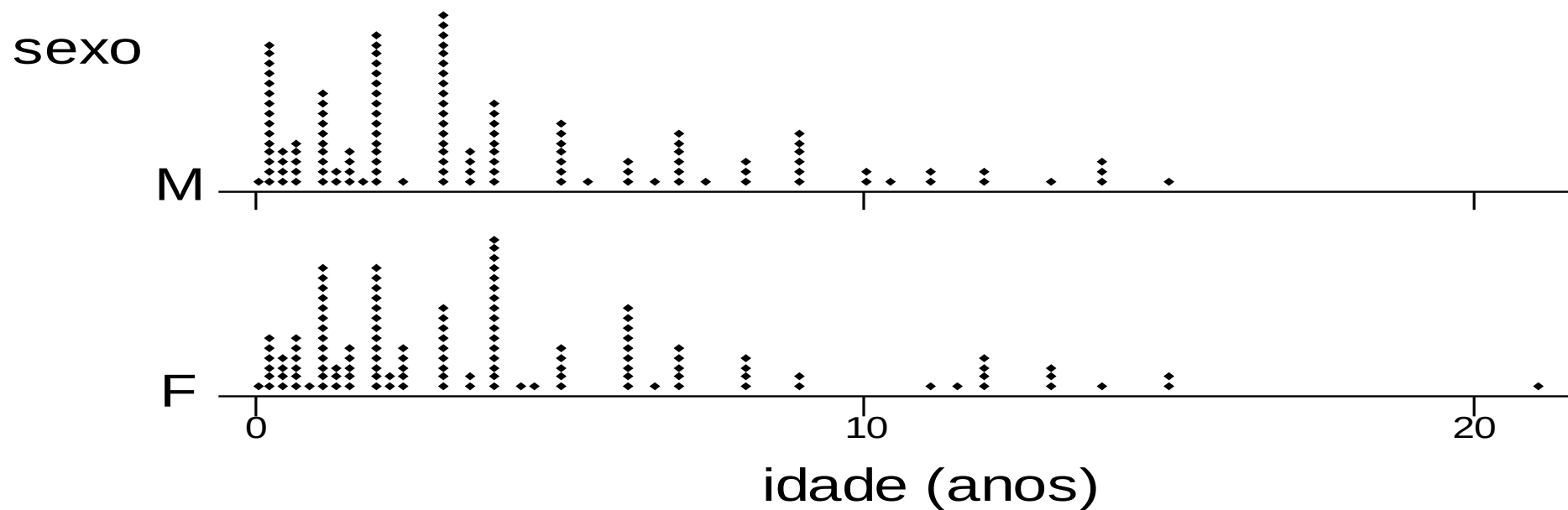
https://www.youtube.com/watch?v=RiEZ_hEf96A

Gráficos - Dados quantitativos

- ◊ Diagrama de pontos
- ◊ Histogramas
- ◊ Boxplot
- ◊ Gráficos de dispersão

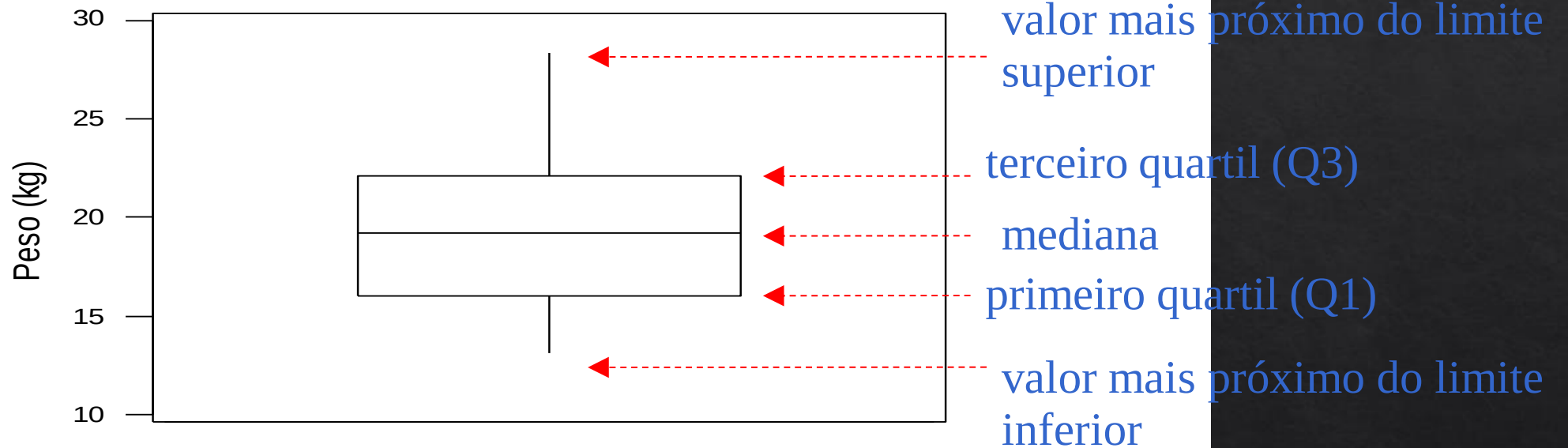
Diagrama de pontos

Figura 3(a): Idades de cães, segundo o sexo, obtidas em levantamento feito em uma escola (Shimozako, 2002)



“Boxplot”

Figura 4: Boxplot para peso de 48 cães



Limite superior: $Q3 + 1,5 (Q3 - Q1)$

“Outliers” (dados discrepantes): dados fora dos limites superior e inferior, indicados com asteriscos

“Boxplot”

Figura 3(b): Idades de cães, segundo o sexo, obtidas em levantamento feito em uma escola (Shimozako, 2002)

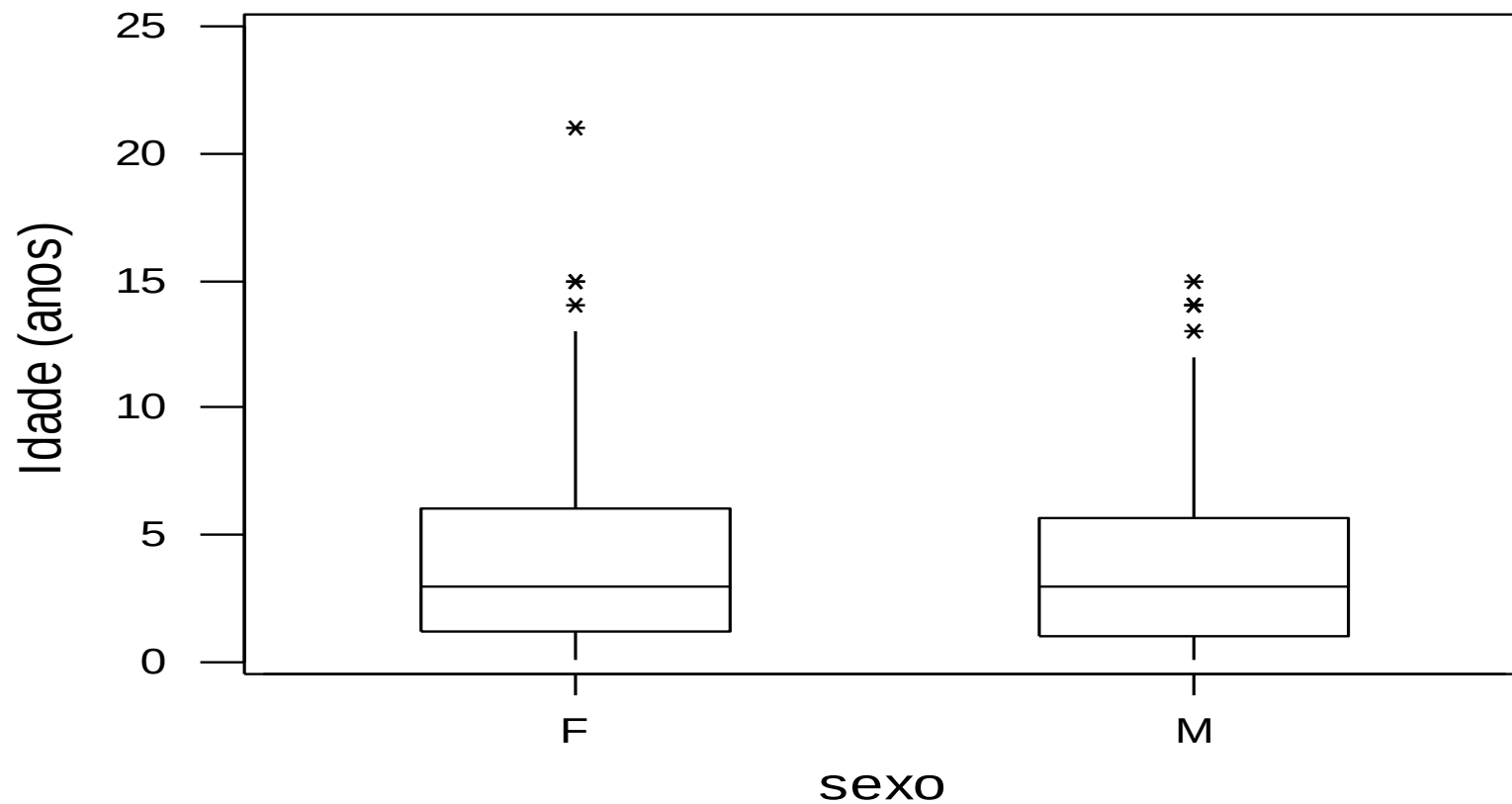
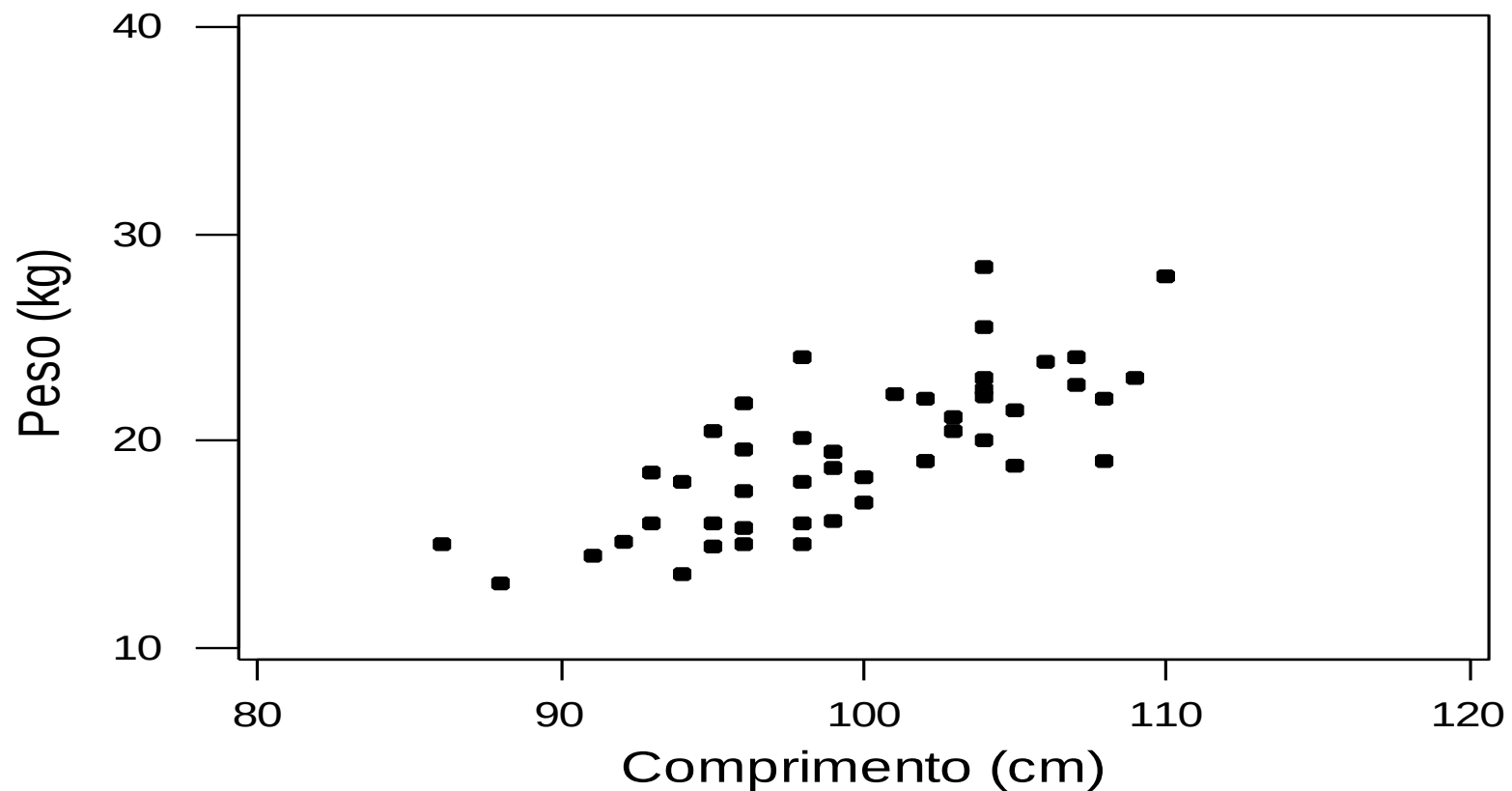
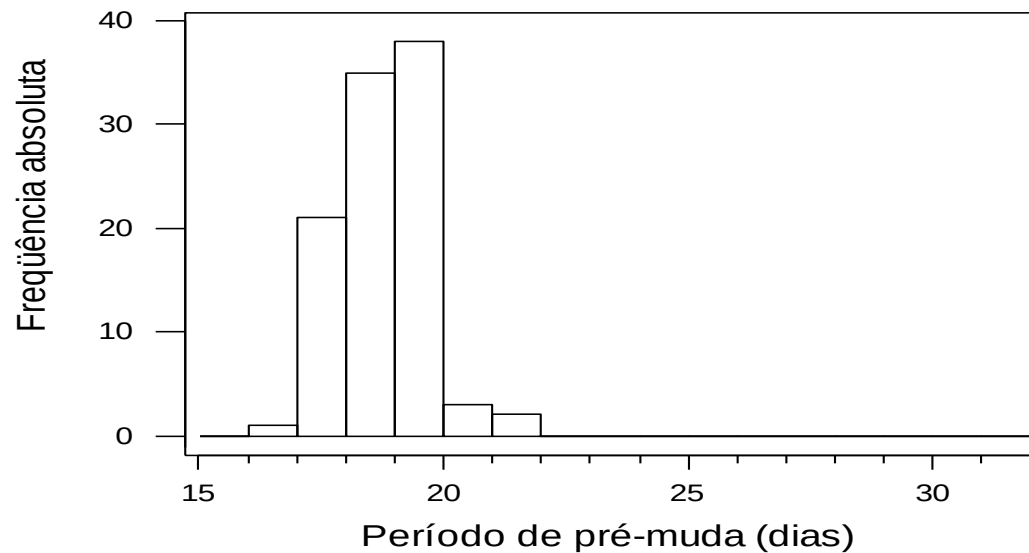
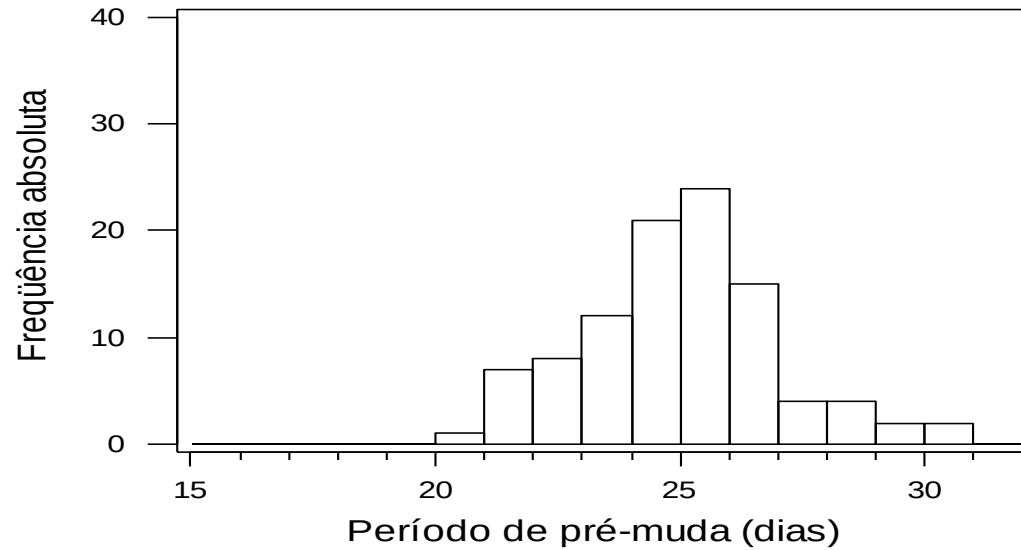


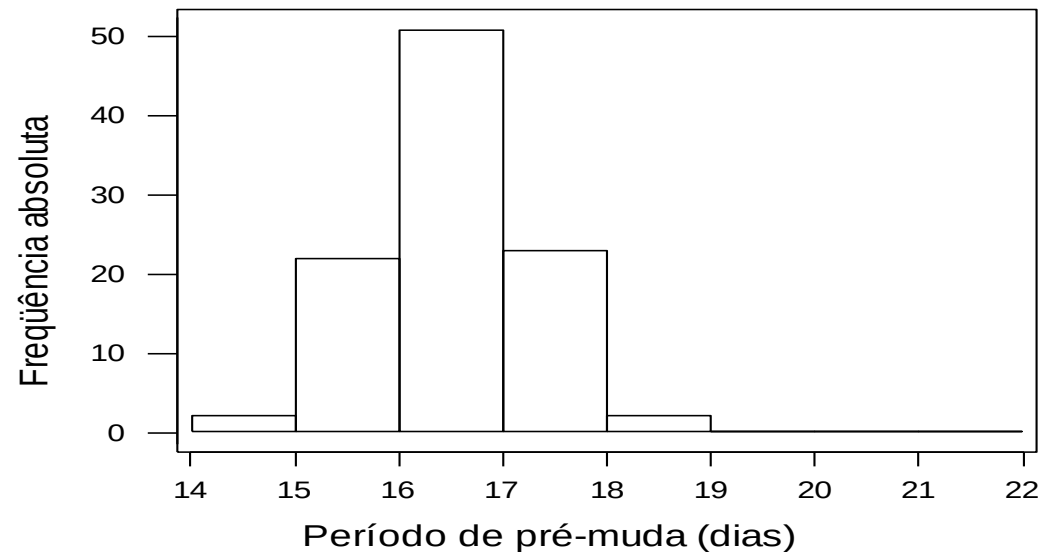
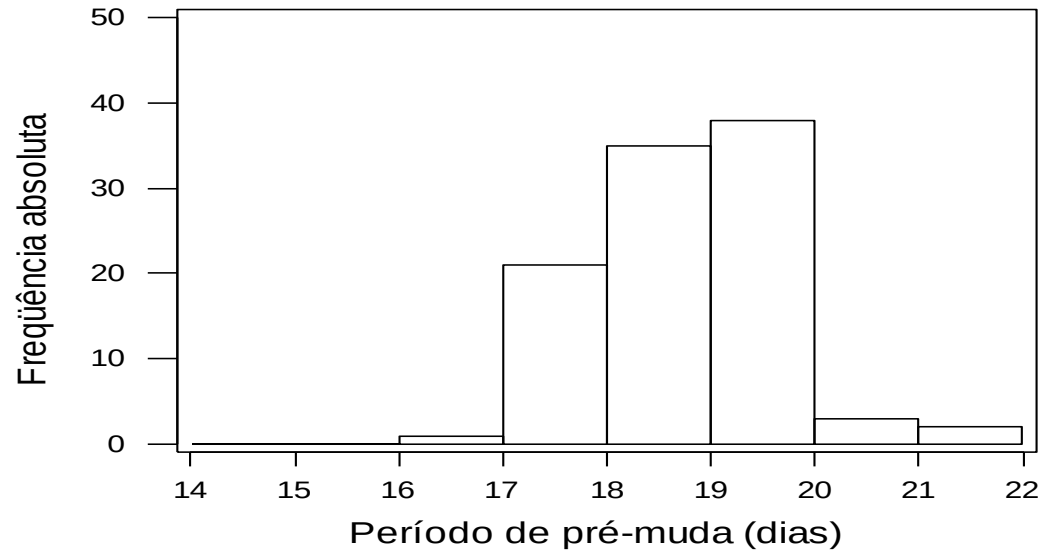
Diagrama de dispersão

Figura 5: Diagrama de dispersão para comprimento e peso de 48 cães

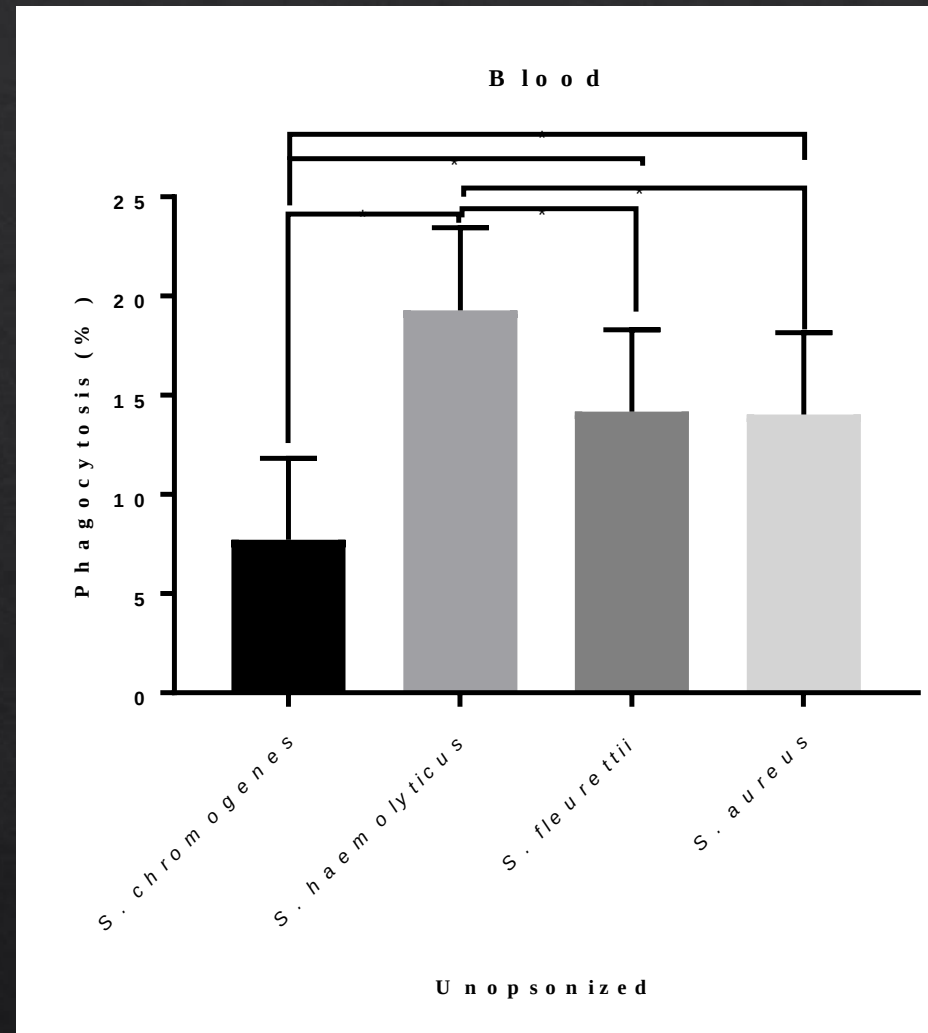




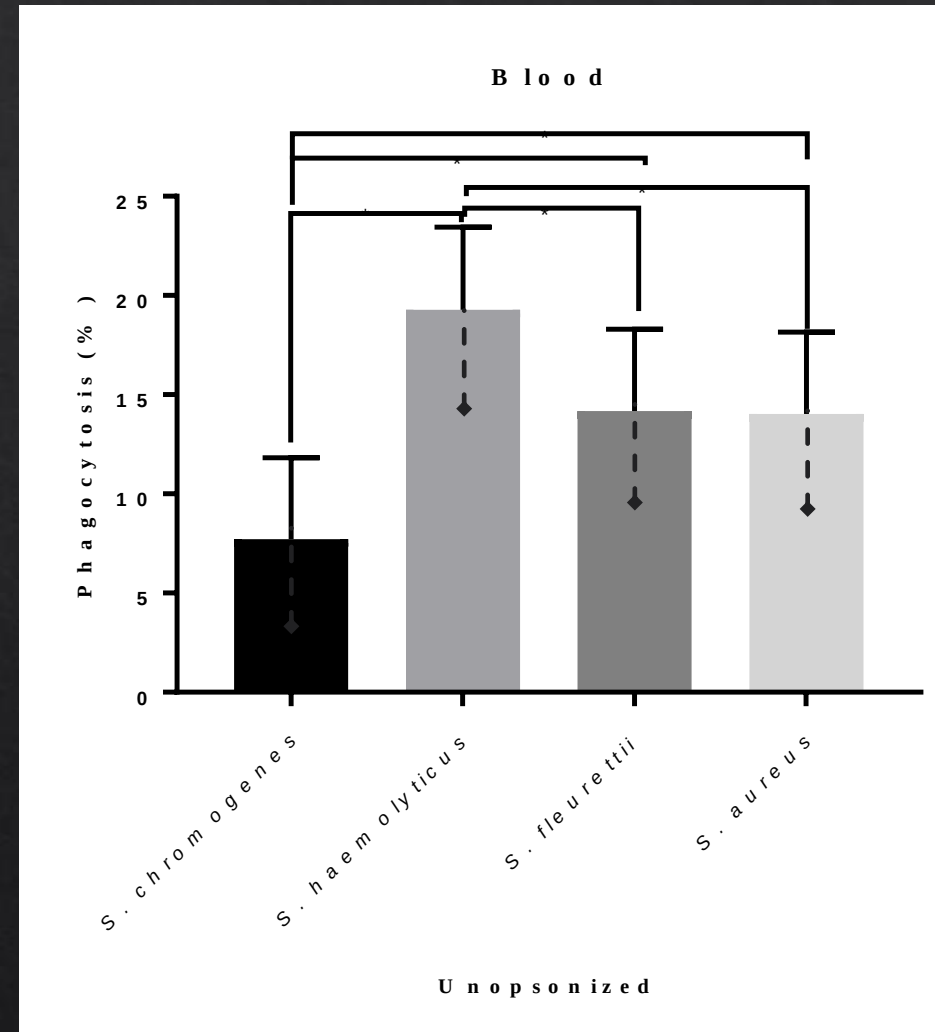
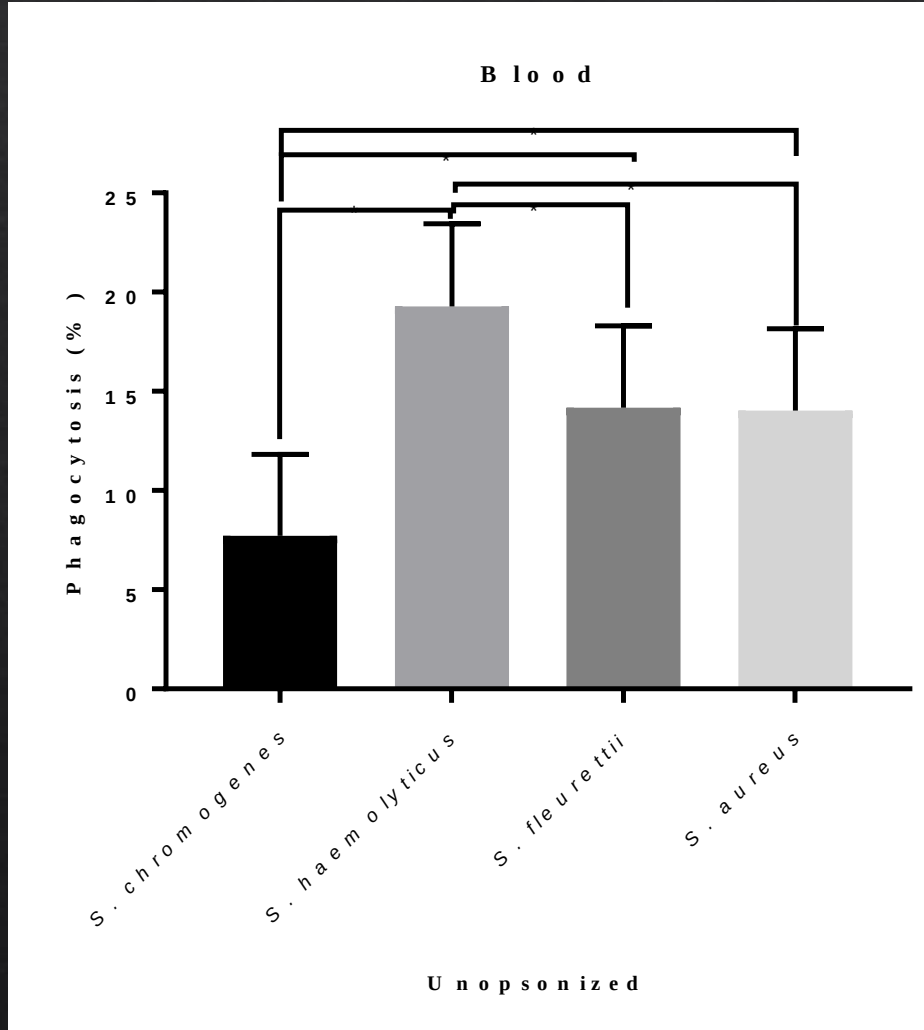
de pré-muda de ninfa
patos (dados
empíricos)

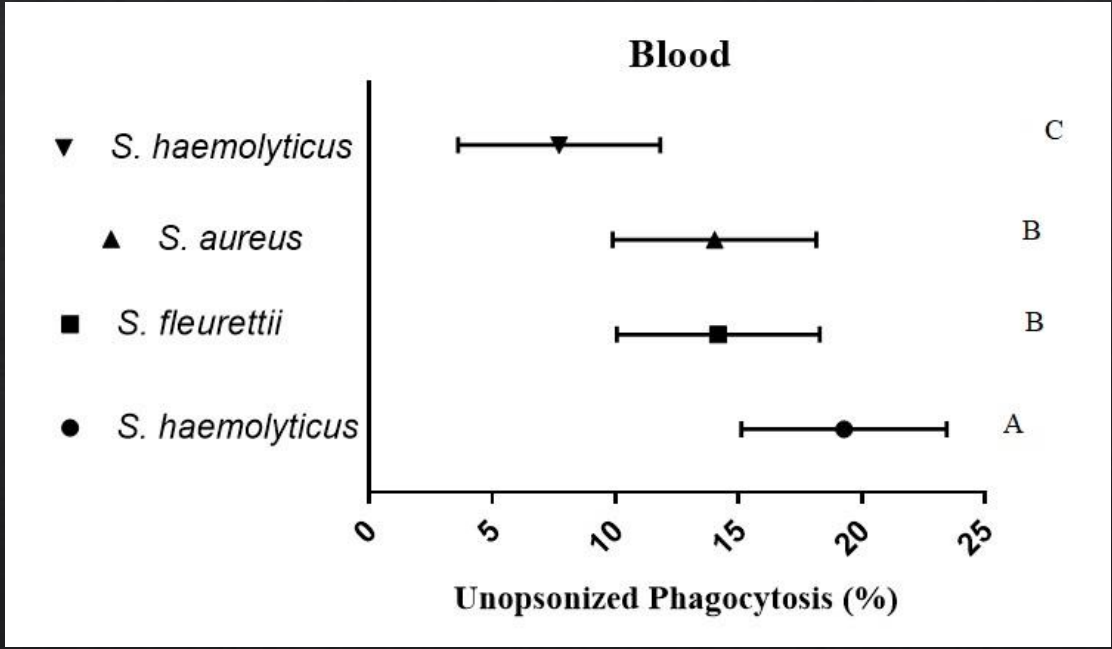
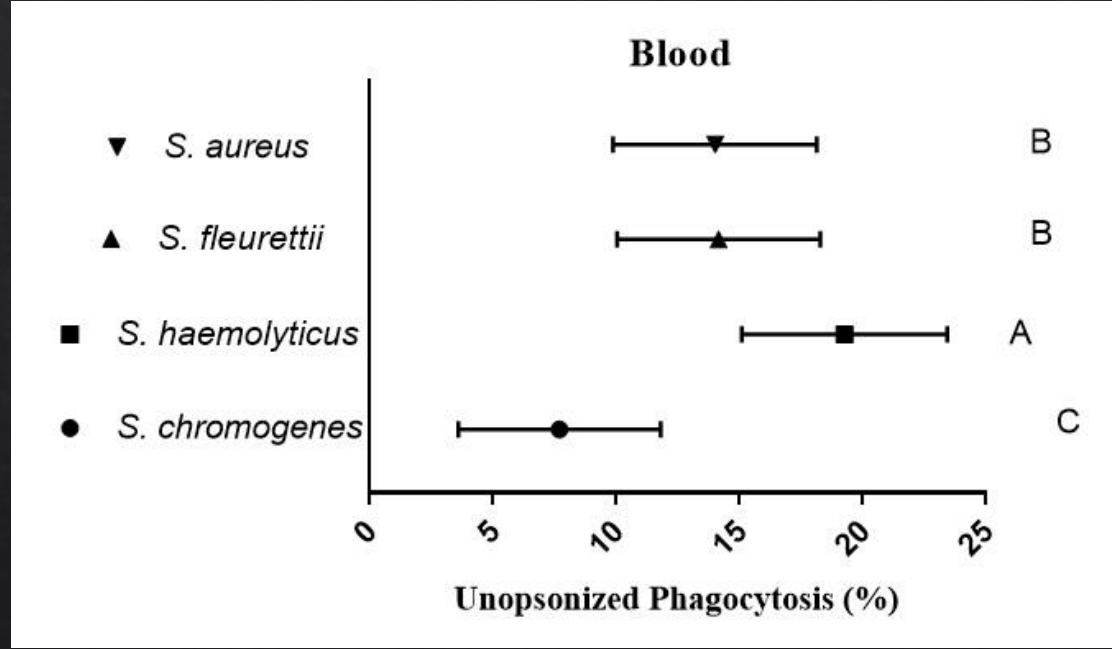


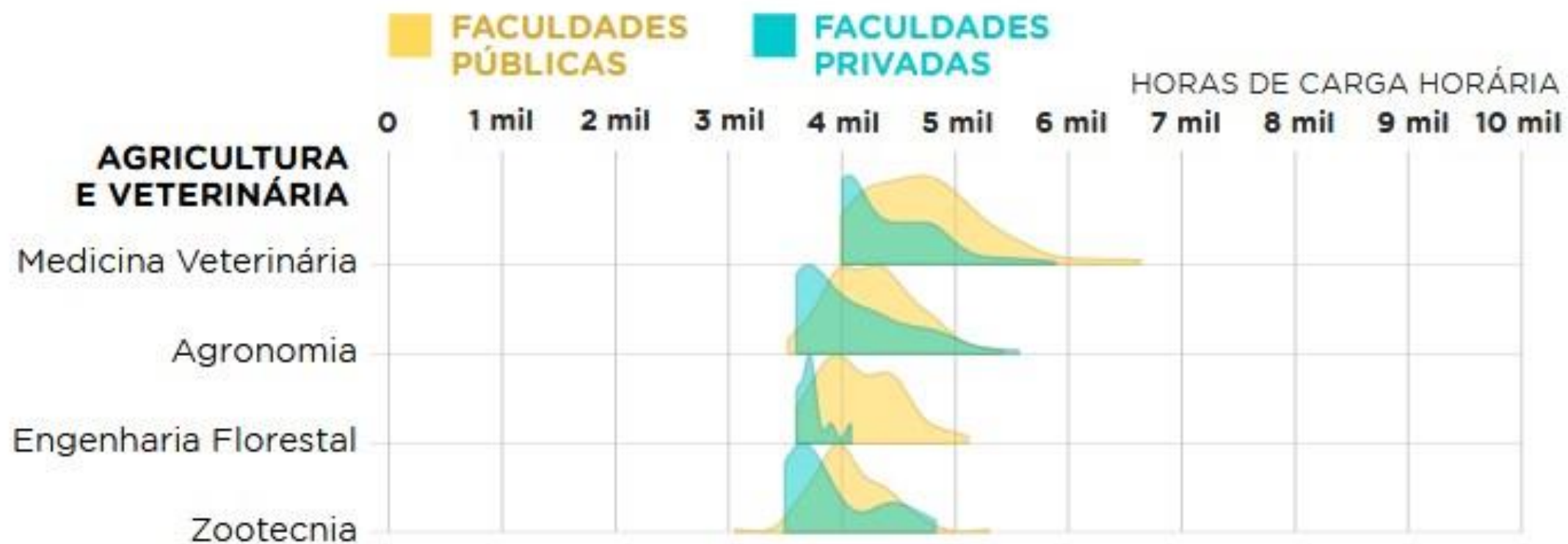
Período de pré-muda de ninfa
de carrapatos (dados
hipotéticos)



Problemas







talks should happen soon.

Question Of The Day

What should cost less: a
gallon of gas or a gallon of milk?

YES
43%



NO
57%



To record your answer to the Question
of the Day and other polls about today's
news, visit www.post-journal.com.

BY
PRE

T
the
are

ry:

C
the

Am

Hor

by

the

Yor

the

Hor

day

C

I

Clas

St.

after

Nor

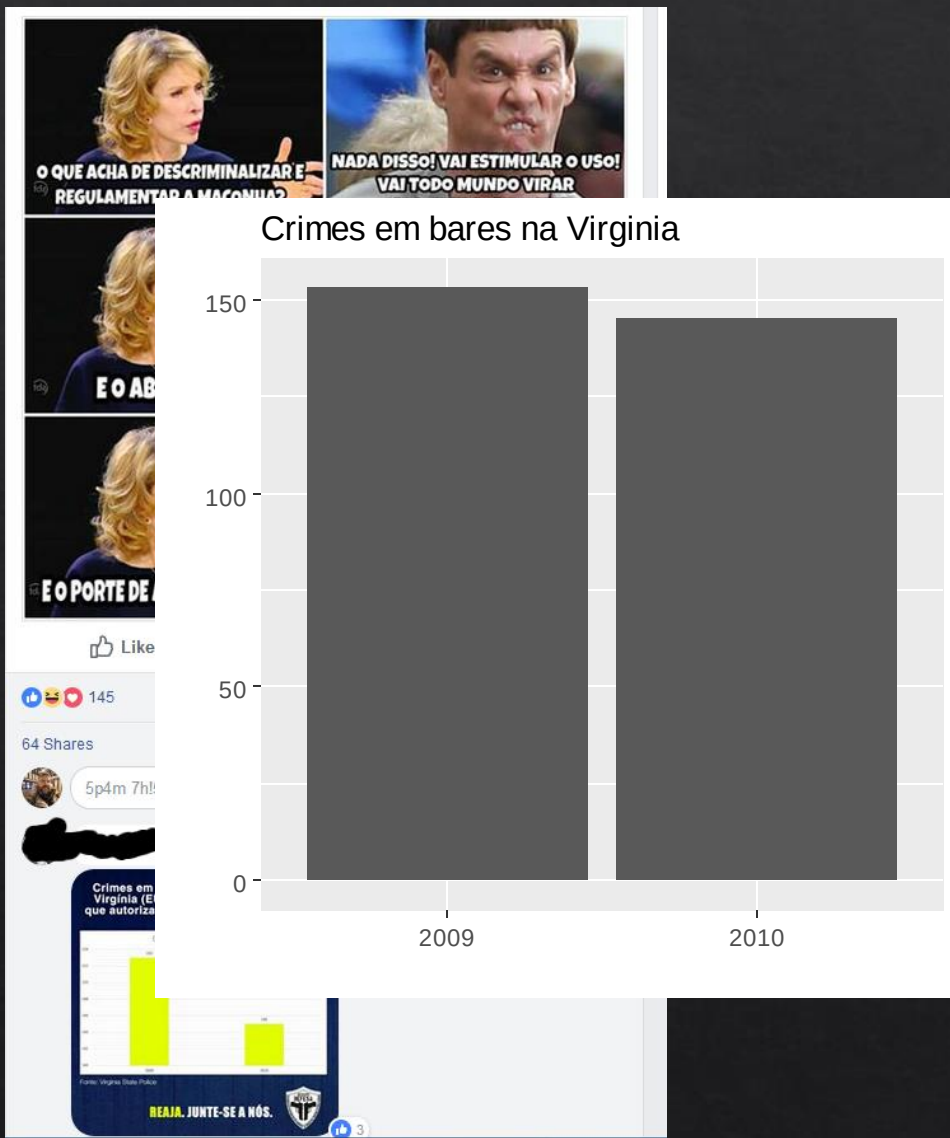
I

tific

astr

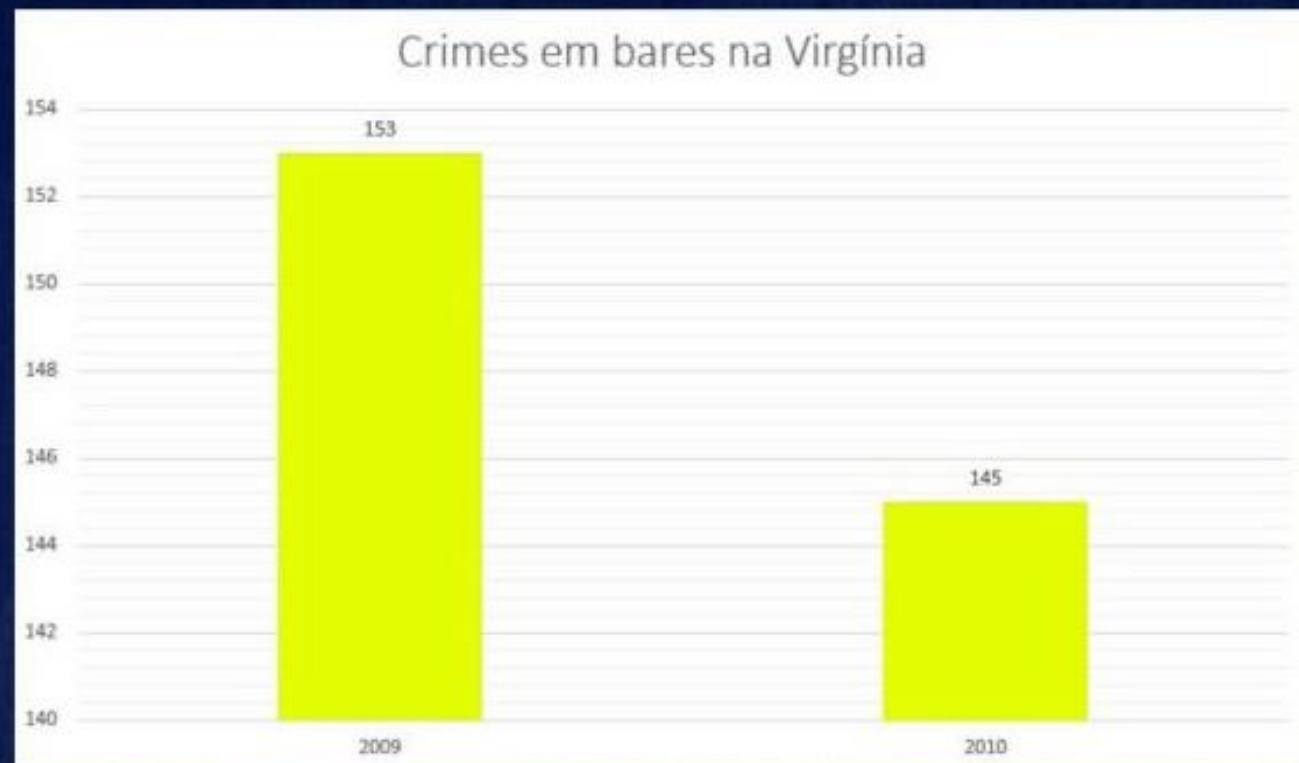
Gal

I



“Vocês tem achismo, eu tenho números”
– usuário da internet

Crimes em bares reduzem no estado de Virgínia (EUA) após promulgação de lei que autoriza porte de armas nesses locais



Fonte: Virginia State Police

REAJA. JUNTE-SE A NÓS.





PSDB SP

6 h • 🌐

Acelera, SP!!! 🙌🙌🙌🙌



DORIA
LIDERA
PESQUISA
APONTA IBOPE



FONTE: UOL MAIO'2018

VOTO
2013
El pueblo decide

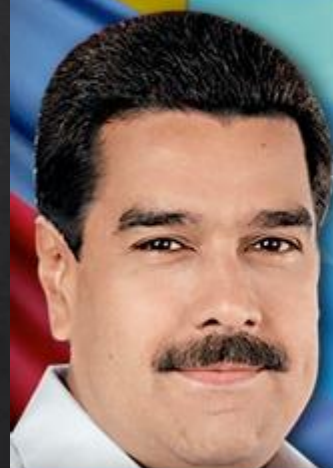
ELECCIÓN PRESIDENCIAL 2013



PORCENTAJES

50,66

% 49,07

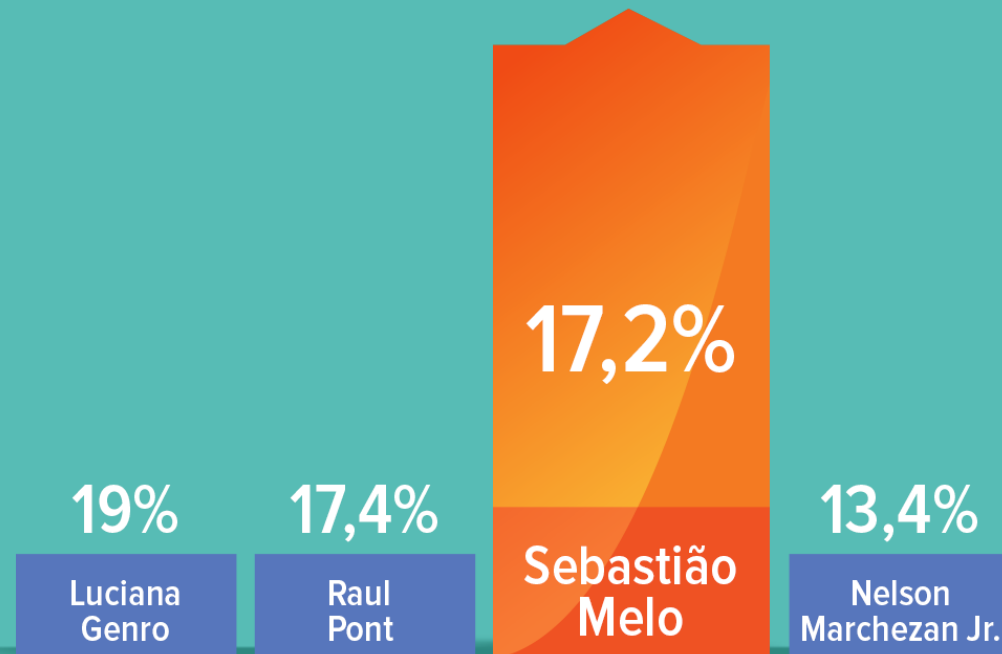


Nicolás Maduro Moros



Henrique Capriles Radonski

MELO É O QUE MAIS CRESCER NAS PESQUISAS!



Fonte: instituto Methodus



#MELO15

GRÁFICO

Como mentir com gráficos: 7 detalhes que podem te enganar

Gabriel Zanlorenssi, Rodolfo Almeida e Gabriel Maia 31 Mar 2018 (atualizado 22/Jun 15h13)



Escalas enganosas, declarações imprecisas e contas que não fecham: existem diversas maneiras de levar a interpretações erradas por meio de gráficos. Conheça algumas

<https://www.nexojornal.com.br/grafico/2018/03/31/Como-mentir-com-gr%C3%A1ficos-7-detalhes-que-podem-te-enganar>

GRÁFICO

Como mentir com gráficos: mais 7 detalhes que podem te enganar

Rodolfo Almeida e Gabriel Zanlorenssi 01 Abr 2019 (atualizado 04/Abr 13h11)



Escalas quebradas, intervalos suspeitos e eixos duplos: veja essas e outras maneiras de levar a interpretações erradas em visualizações de dados

<https://www.nexojornal.com.br/grafico/2019/04/01/Como-mentir-com-gr%C3%A1ficos-mais-7-detalhes-que-podem-te-enganar>

“A mediana não é a mensagem”
(Stephen Jay Gould)

<http://www.phoenix5.org/articles/GouldMessage.html>

http://www.cancerguide.org/median_not_msg.html