

# Fundação Lemann Hackday

Guilherme Silveira  
Mauricio Aniche

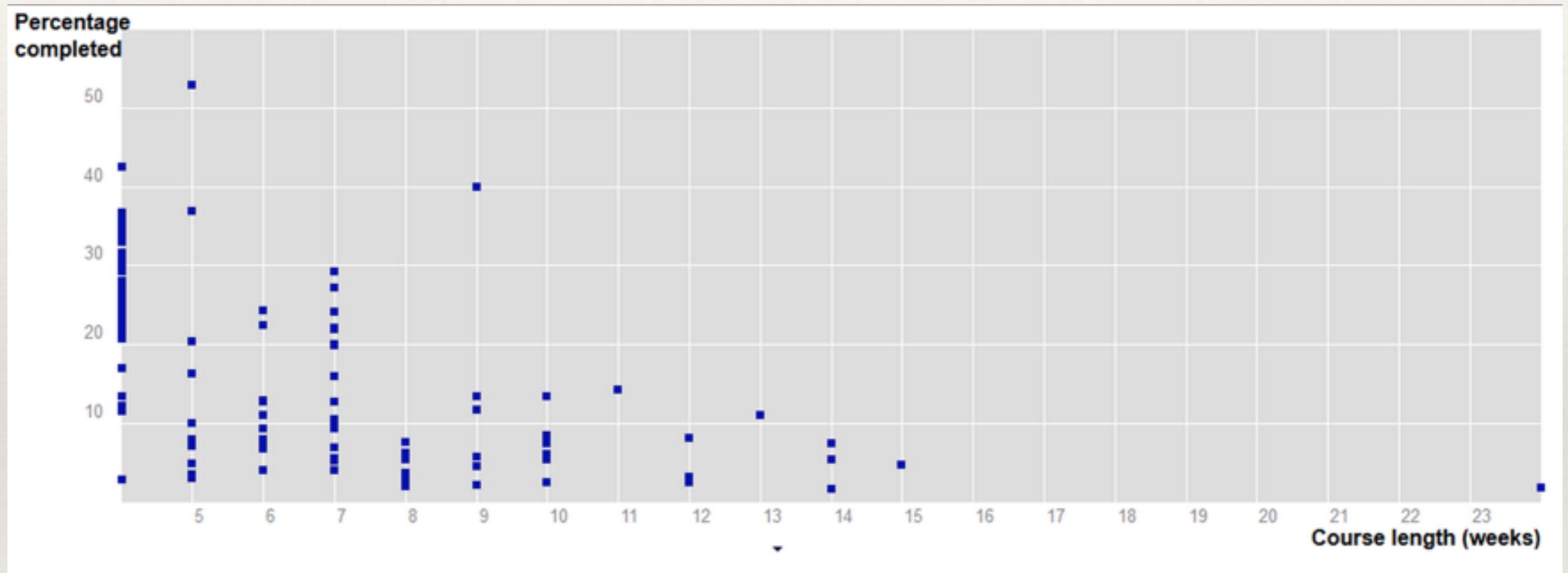
# Parte 1: gatópolis

# Dashboard...s



mas... pra quê?

# gráfico S2, e agora?



<https://moocmoocher.wordpress.com/2013/02/13/synthesising-mooc-completion-rates/>

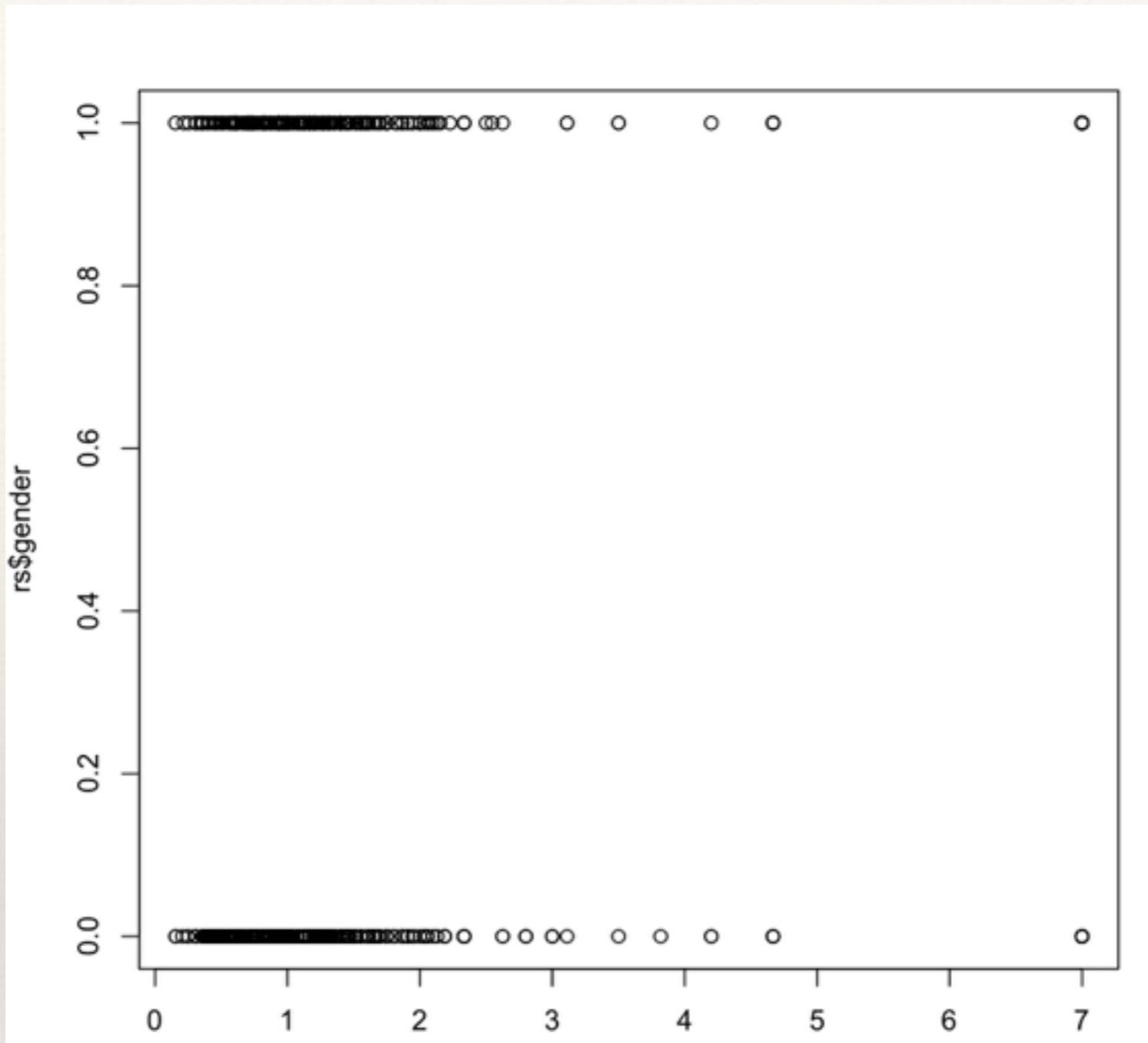
pra que um dashboard de alunos?  
pra que um dashboard de professores?  
pra que um dashboard de escolas?

explorando...

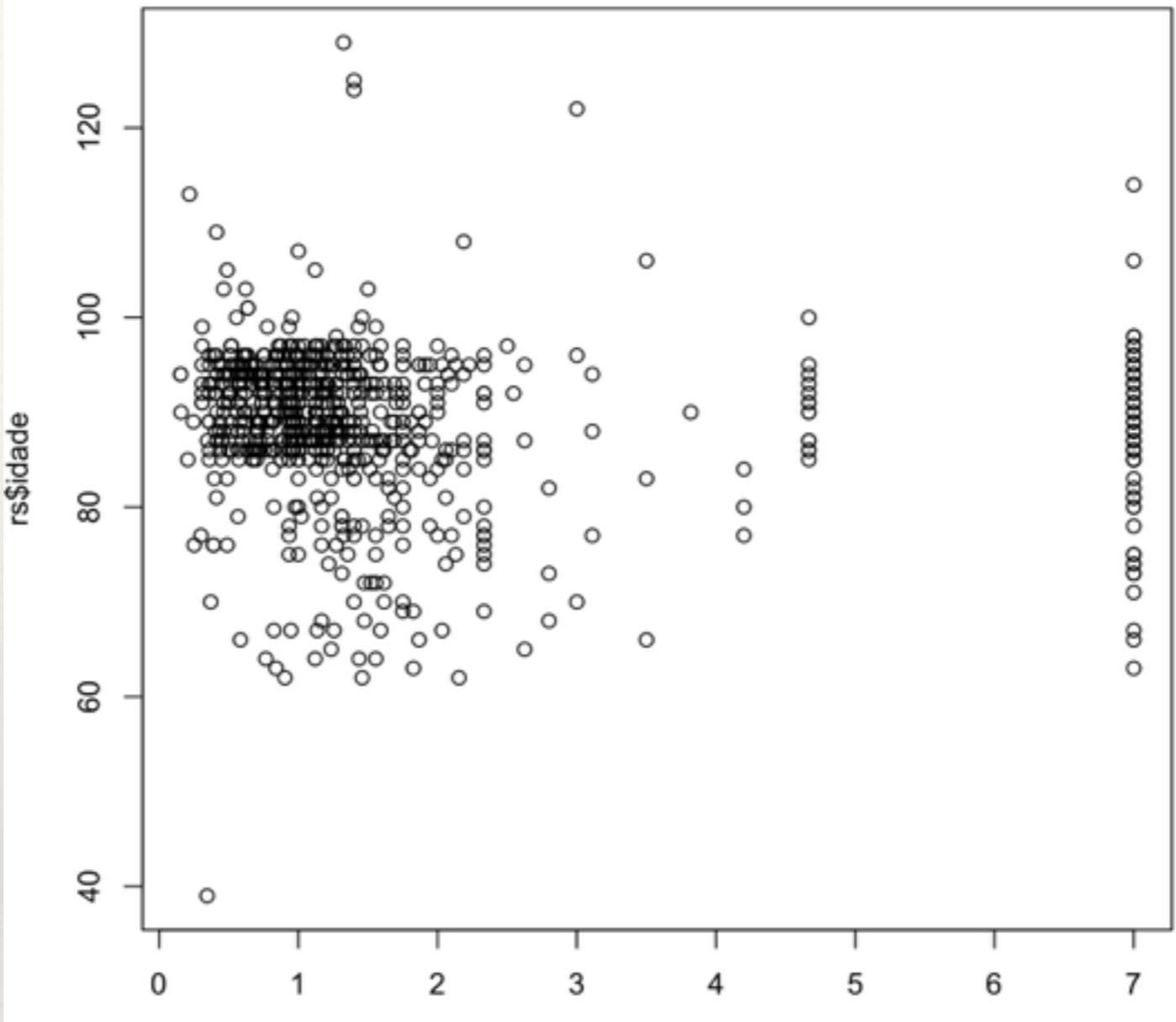
explorando...

Rápido!

quantos dias por semana  
o aluno estuda?



vezes por semana POR gênero



mas e aí?

o que a escola faz com um dashboard?

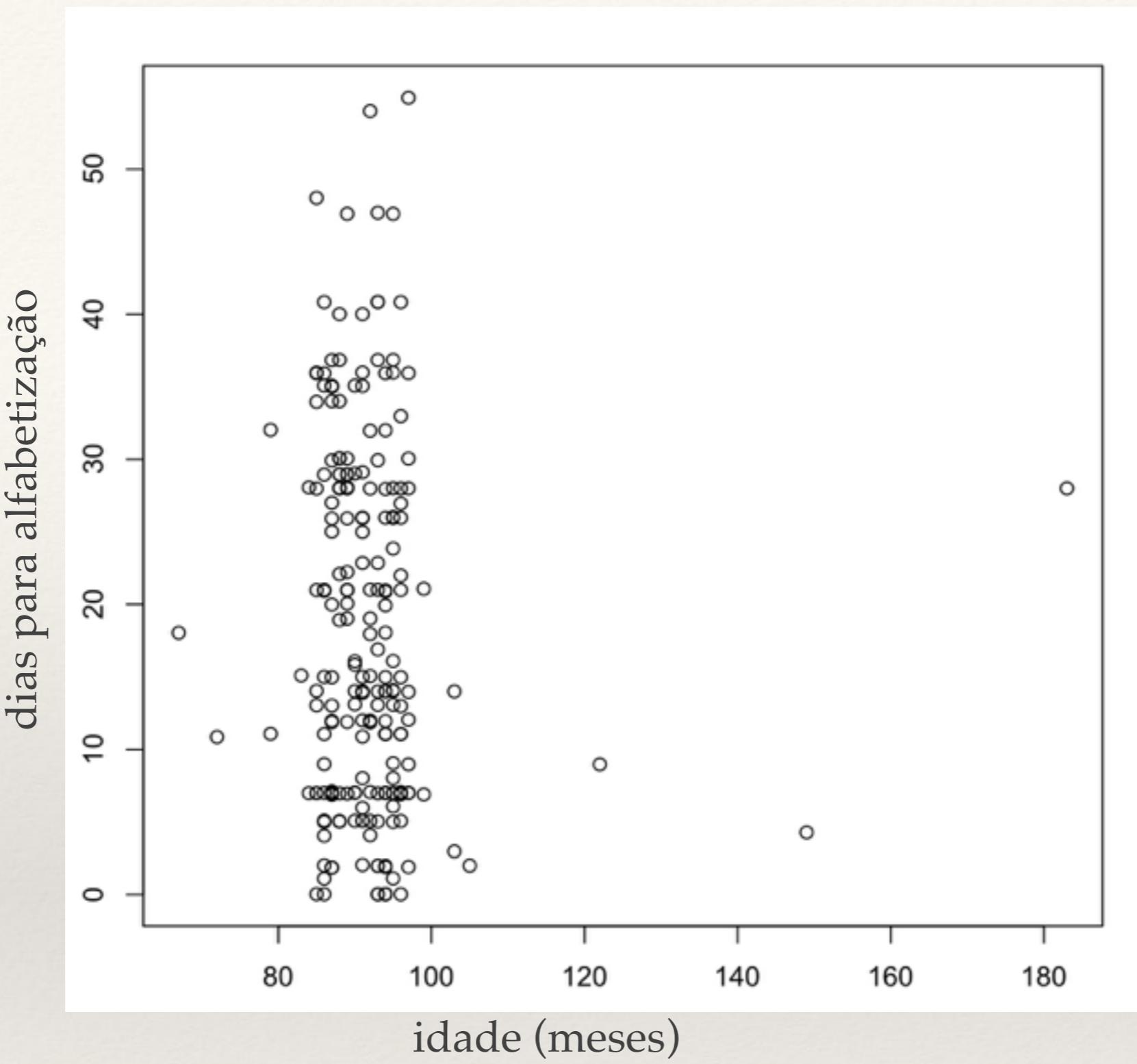
**OBJETIVO** do gatópolis?

**quero ALFABETIZAR!**

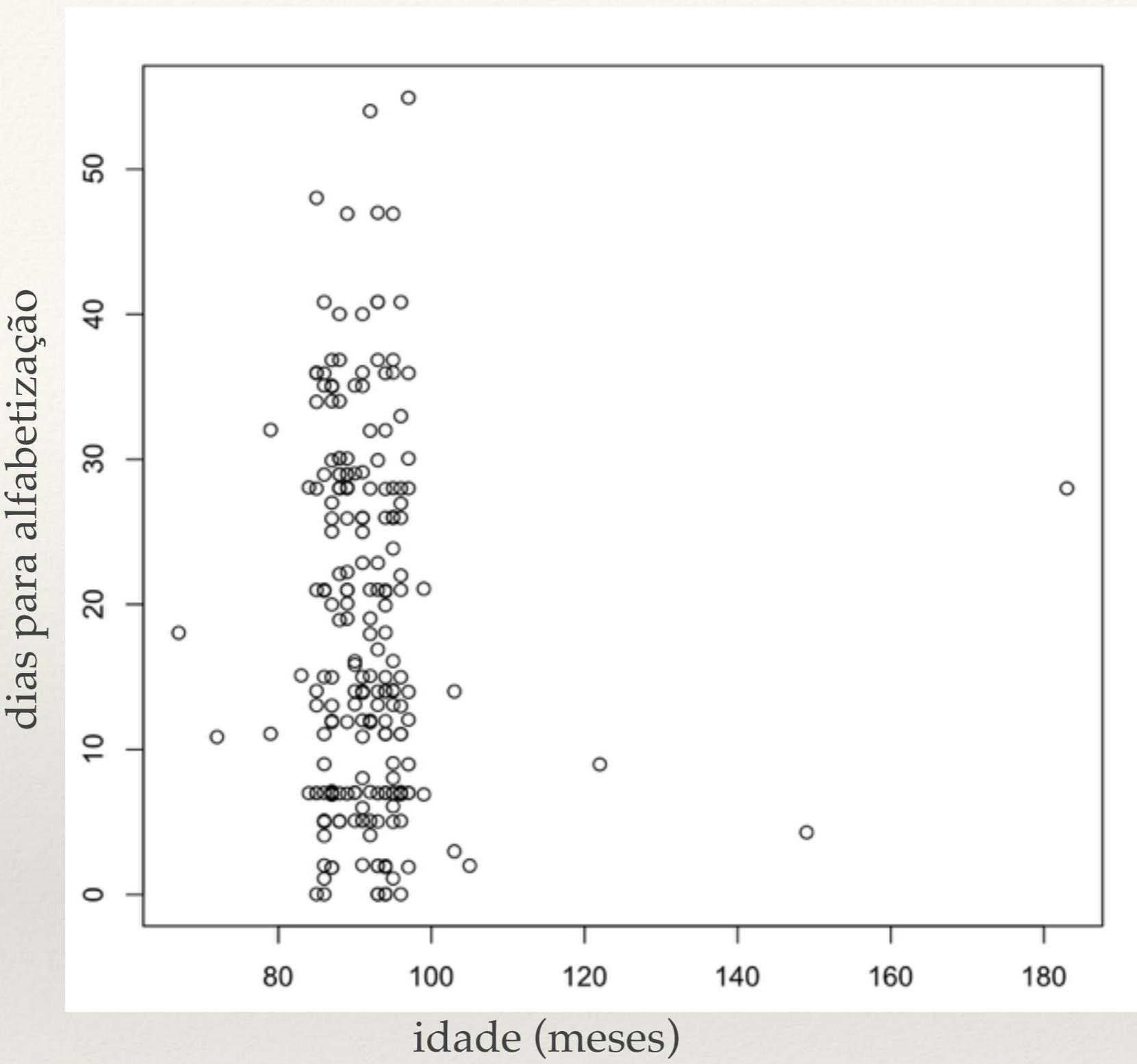
quero ALFABETIZAR!

agora sim posso criar algo que alfabetize

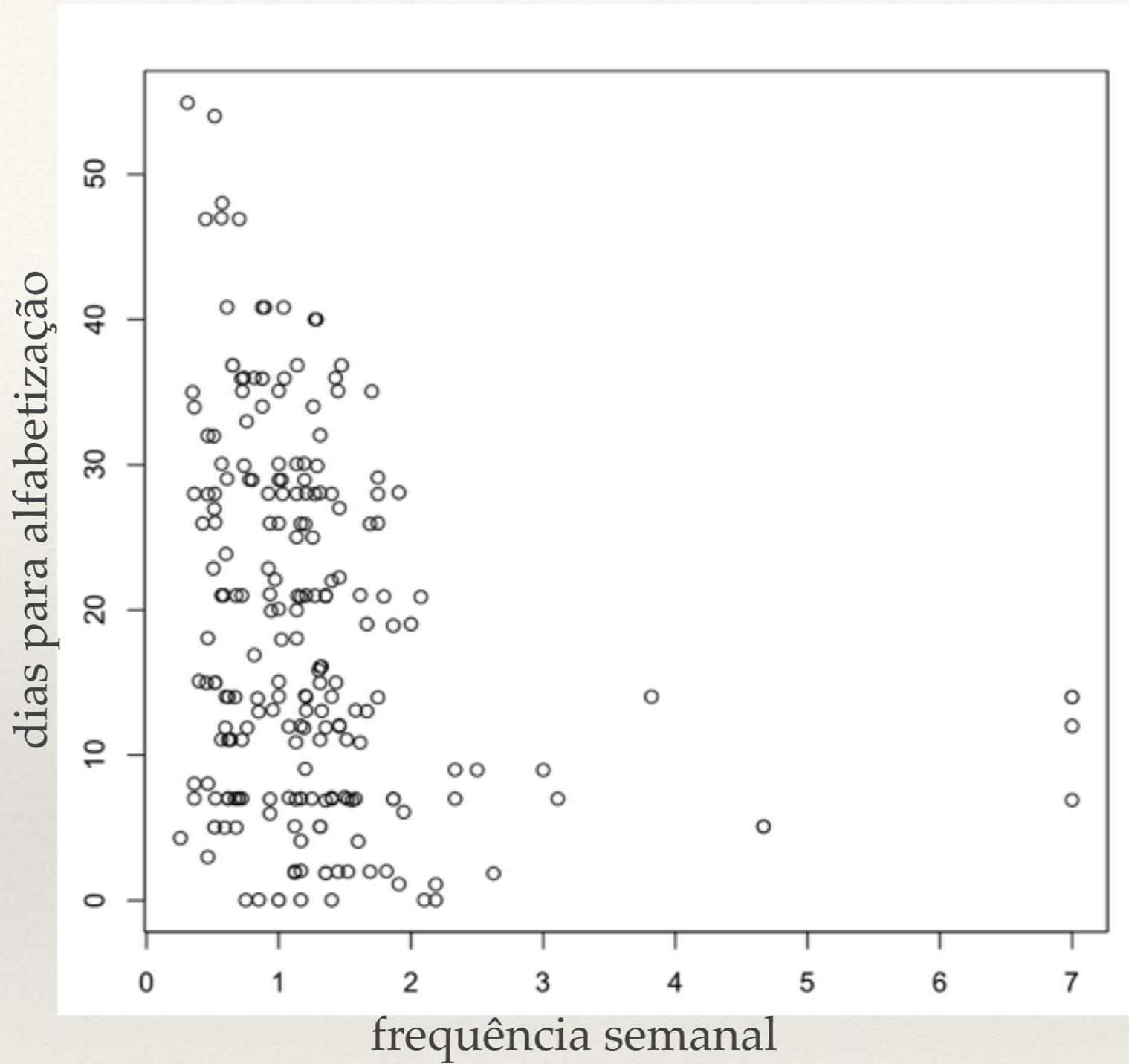
p: qual o tempo de alfabetização?



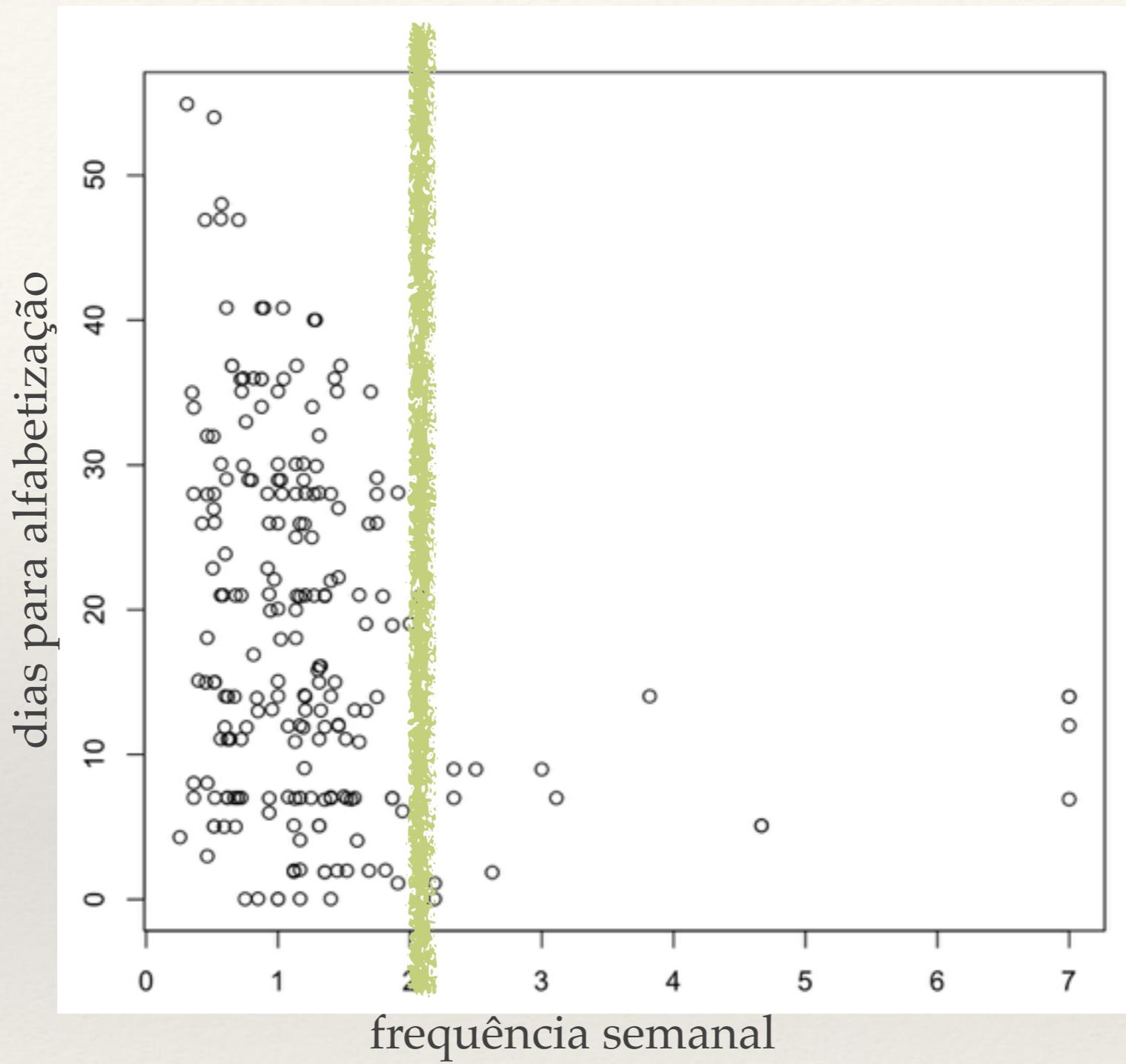
idade em meses por tempo para  
alfabetização



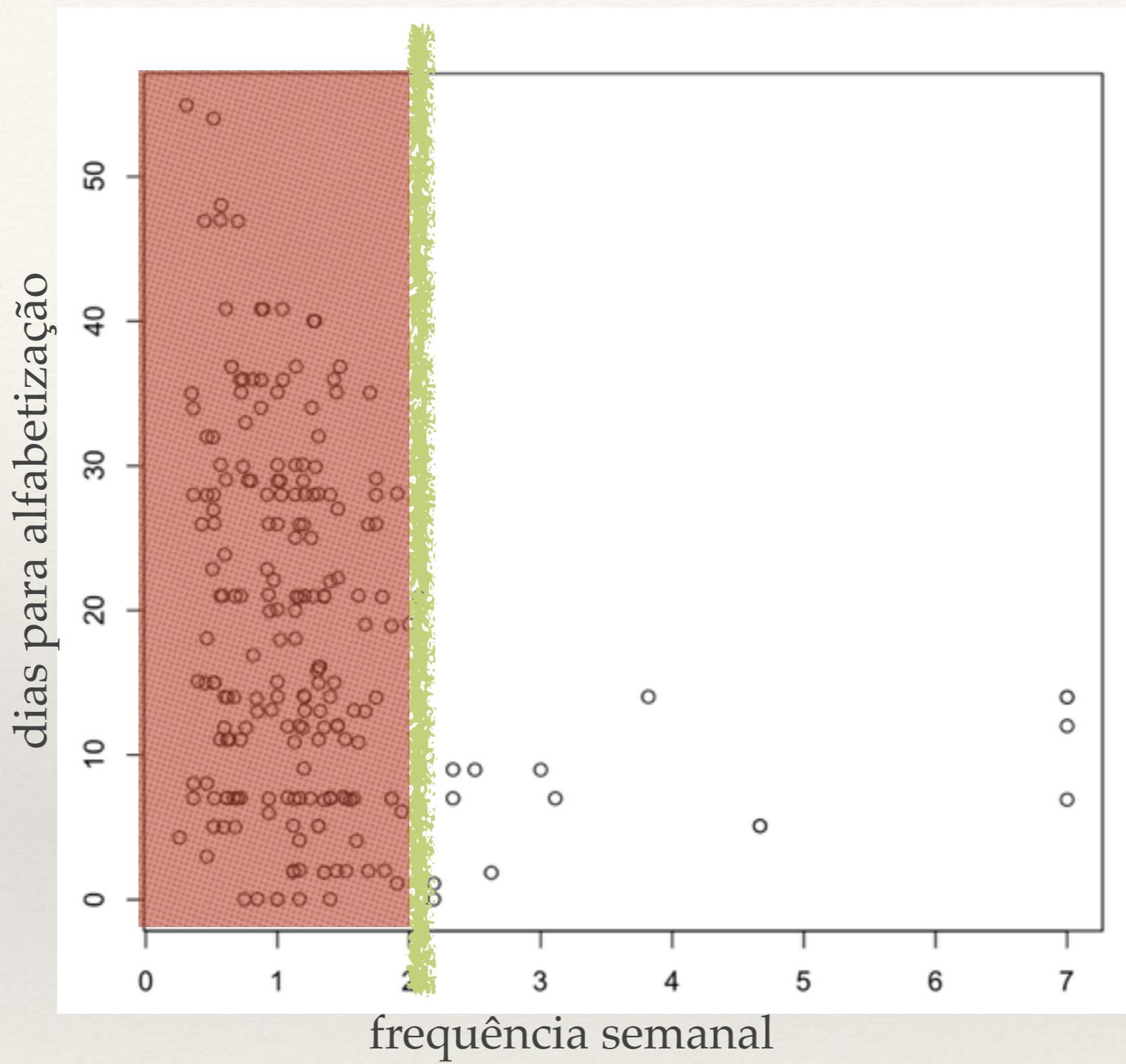
interessante: idade parece não afetar alfabetização!

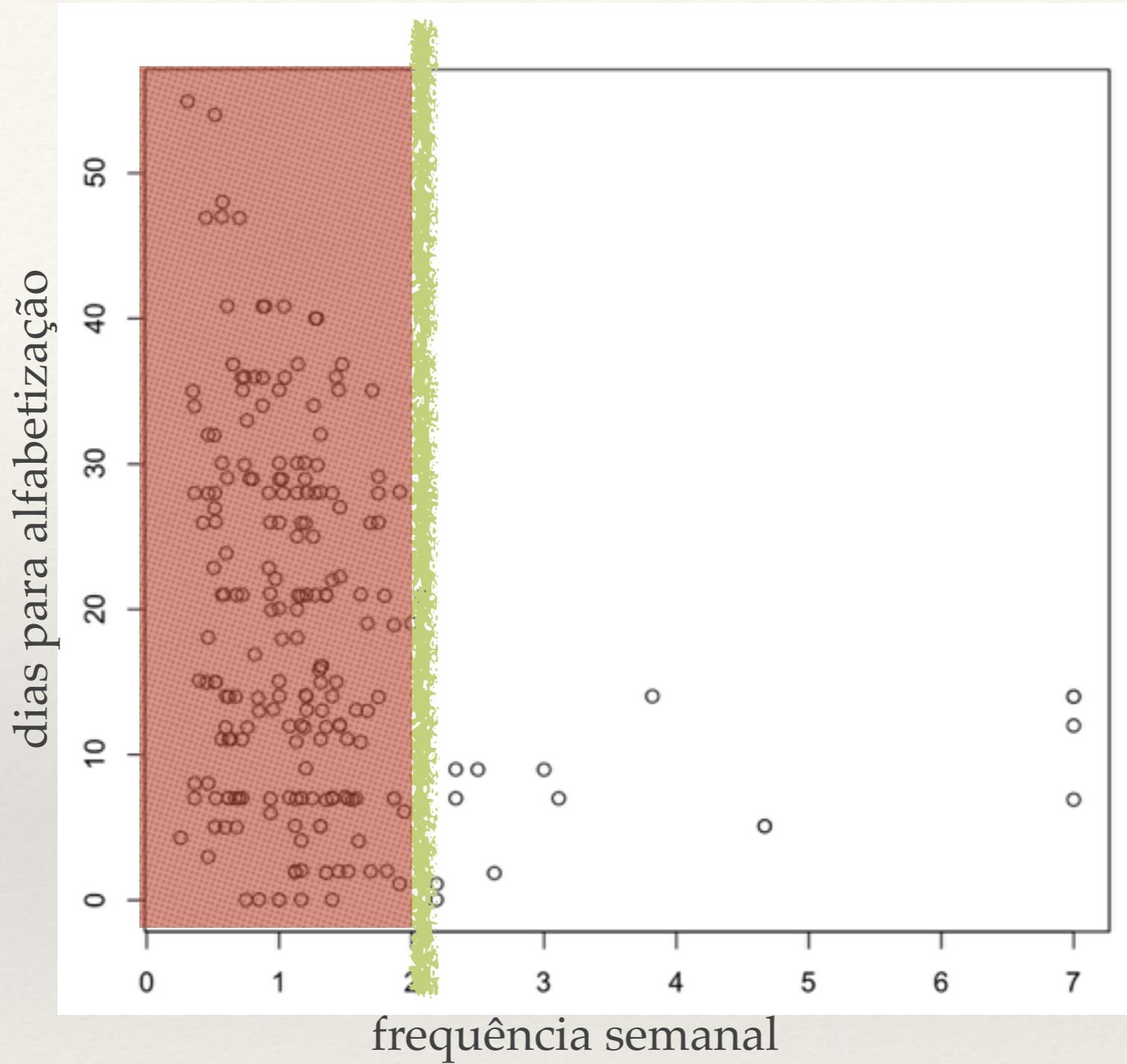


# freq. semanal por alfabetização



# freq. semanal por alfabetização





```
> summary(x)
   Min. 1st Qu. Median      Mean 3rd Qu.      Max.
0.02059 7.02000 16.90000 18.98000 28.05000 54.93000
> |
```

```
> summary(y)
   Min. 1st Qu. Median      Mean 3rd Qu.      Max.
0.02465 5.08700 7.00200 8.42600 13.00000 20.90000
|
```

1 x ou menos => 7 a 28 dias

2 x ou mais => 5 a 13 dias

```
> x <- rs$tempo[rs$dias_por_semana<2]
> y <- rs$tempo[rs$dias_por_semana >= 2]
> wilcox.test(x, y)
```

Wilcoxon rank sum test with continuity correction

```
data: x and y
W = 2893, p-value = 0.0002101
alternative hypothesis: true location shift is not equal to 0
```

```
> x <- rs$tempo[rs$dias_por_semana<2]  
> y <- rs$tempo[rs$dias_por_semana >= 2]  
> wilcox.test(x, y)
```

Wilcoxon rank sum test with continuity correction

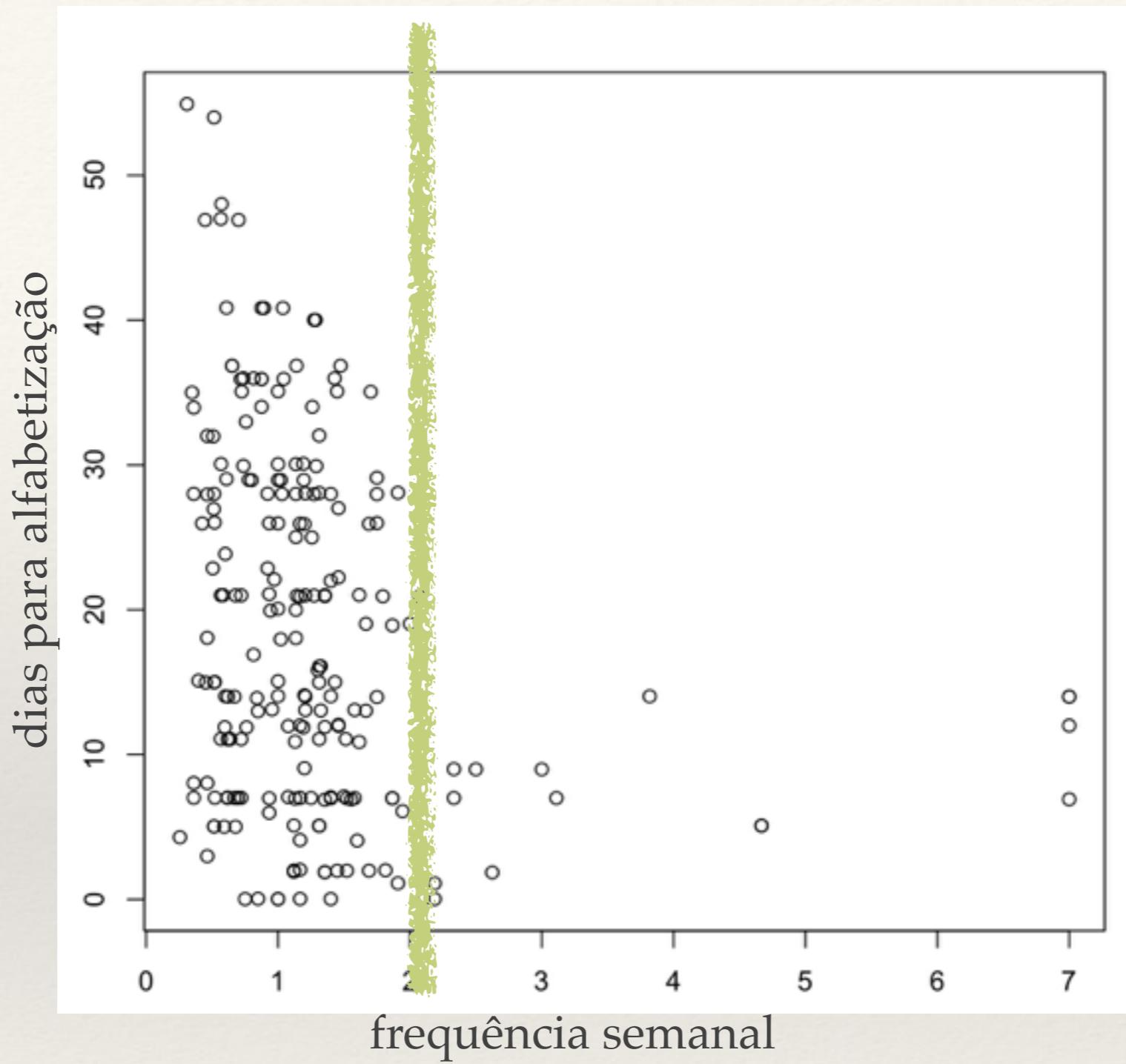
```
data: x and y  
W = 2893, p-value = 0.0002101  
alternative hypothesis: true location shift is not equal to 0
```

```
> x <- rs$tempo[rs$dias_por_semana<2]  
> y <- rs$tempo[rs$dias_por_semana >= 2]  
> wilcox.test(x, y)
```

Wilcoxon rank sum test with continuity correction

```
data: x and y  
W = 2893, p-value = 0.0002101  
alternative hypothesis: true location shift is not equal to 0
```

SIM. alfabetização foi mais rápida.



**ação: estudar 2 ou mais dias/semana**

---

# Para aumentar a frequência...

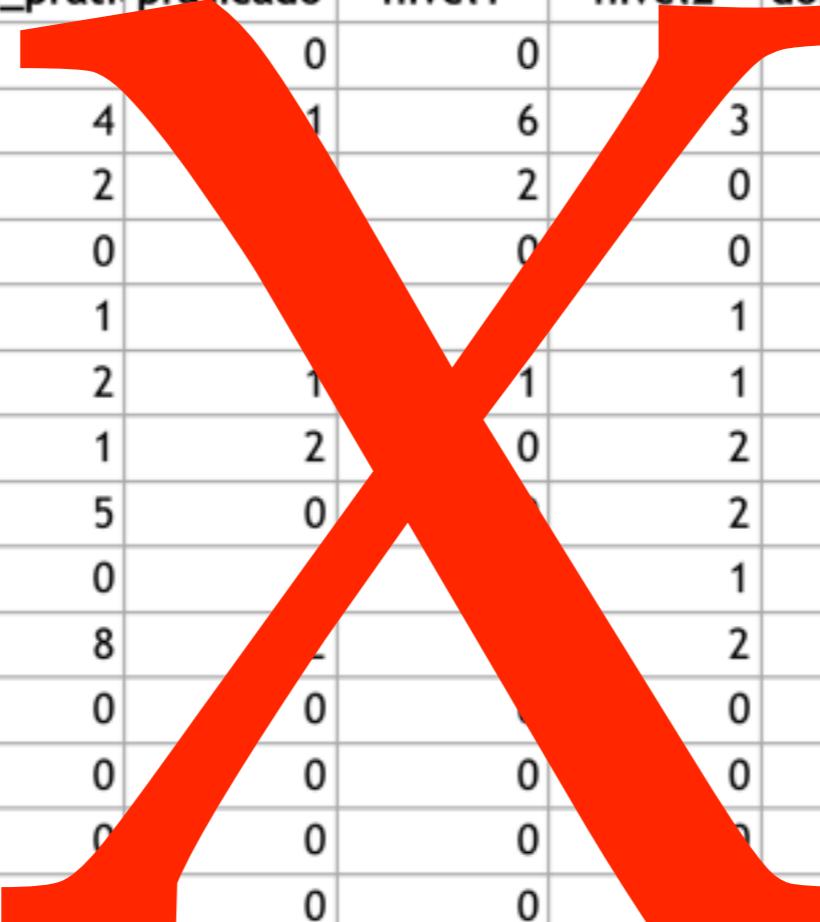
---

- ❖ Dashboard do professor: quais alunos estão estudando menos de 3 vezes por semana. Me dê o contato dos alunos e pais.
- ❖ Dashboard da escola: quais professores tem o menor numero de alunos estudando
- ❖ Automático pro aluno: se não está no ritmo, pinga no celular para abrir o jogo

# Parte 2: khan

n_dificuldade	cisica_pratica	praticado	nivel1	nivel2	dominado	pontos	perciseminut
3	3	0	0	0	0	0	10.83
1	4	1	6	3	11	5240	51.95
2	2	3	2	0	9	8375	45.98
0	0	0	0	0	0	0	0.00
0	1	0	0	1	8	4987	38.23
1	2	1	1	1	8	5238	31.38
1	1	2	0	2	3	2825	10.40
0	5	0	0	2	9	4719	29.68
4	0	0	0	1	3	4225	29.37
4	8	2	0	2	10	27660	31.50
1	0	0	0	0	7	3875	9.72
0	0	0	0	0	0	0	0.00
0	0	0	0	0	0	0	0.00
2	2	0	0	1	6	1250	32.72
3	3	0	0	0	16	4225	36.25

quando um nível sobe,  
o outro diminui



n	dificuldade	cisas_praticadas	praticado	nivel1	nivel2	dominado	pontos	exerciseminuto
3	3	0	0	0	0	0	0	10.83
1	4	1	1	6	3	11	5240	51.95
2	2	2	1	2	0	9	8375	45.98
0	0	0	0	0	0	0	0	0.00
0	1	1	1	1	1	8	4987	38.23
1	2	1	1	1	1	8	5238	31.38
1	1	2	0	0	2	3	2825	10.40
0	5	0	0	0	2	9	4719	29.68
4	0	0	0	0	1	3	4225	29.37
4	8	2	2	2	2	10	27660	31.50
1	0	0	0	0	0	7	3875	9.72
0	0	0	0	0	0	0	0	0.00
0	0	0	0	0	0	0	0	0.00
2	2	0	0	0	0	6	1250	32.72
3	3	0	0	0	0	16	4225	36.25

quando um nível sobe,  
o outro diminui



**escola**

e a escola nisso?

quanto mais tempo em exercício,  
o khan avalia mais precisamente  
as habilidades do aluno

quantos exercícios são necessários  
pra dominar um assunto?

$f(\text{exercícios}) = \text{domínio do aluno}$

$f(\text{exercícios}) = \text{domínio do aluno}$

(usamos uma regressão linear simplérrima)

qual a velocidade de crescimento  
da maestria para uma escola?

```
> x <- lm(formula = rs2$dominado ~ rs2$exerciseminutes)
> x

Call:
lm(formula = rs2$dominado ~ rs2$exerciseminutes)

Coefficients:
  (Intercept)  rs2$exerciseminutes
    0.8548                  0.1663

> █
```

qual a velocidade de crescimento  
da maestria para uma escola?

```
> x <- lm(formula = rs2$dominado ~ rs2$exerciseminutes)
> x

Call:
lm(formula = rs2$dominado ~ rs2$exerciseminutes)

Coefficients:
(Intercept) rs2$exerciseminutes
              0.8548                  0.1663
```



qual a velocidade de crescimento  
da maestria para uma escola?

```
> x <- lm(formula = rs2$dominado ~ rs2$exerciseminutes)
> x

Call:
lm(formula = rs2$dominado ~ rs2$exerciseminutes)

Coefficients:
(Intercept) rs2$exerciseminutes
              0.8548                  0.1663
```

>

qual a velocidade de crescimento  
da maestria para uma escola?

$f_{\text{escola}1}(\text{exercícios}) = \text{domínio do aluno}$

$f_{\text{escola}2}(\text{exercícios}) = \text{domínio do aluno}$

...

$f_{\text{escola}n}(\text{exercícios}) = \text{domínio do aluno}$

o número de exercícios varia  
por escola...

```
> library(RMySQL)
> con <- dbConnect(MySQL(), dbname="hackday_khan", user='root', password='')
>
> sql <- "select distinct escola from khan"
> rs <- dbGetQuery(con, sql)
> for(i in 1:length(rs$escola)) {
+ escola <- rs$escola[i]
+
+ sql <- paste("select * from khan where escola = '", escola, "'", sep="")
+ rs2 <- dbGetQuery(con, sql)
+
+ x <- lm(formula = rs2$dominado ~ rs2$exerciseminutes)
+ cat(paste(escola, "", x$coefficients[1], "", x$coefficients[2], "\n", sep=""), file=arquivo, append=TRUE)
+ }
```

qual a velocidade de crescimento  
da maestria para todas elas?

EM MONTEIRO LOBATO	4.66386957063584e-16	1.33630289532294
E.M. PROF BADIA DE FARIA	0.339066064677893	0.409285715091525
EM PREF. NOÉ ALVES FERREIRA	-0.224388064058562	0.394717910233839
E.M. PROF. MARIA HOSTIM DA COSTA	0.360647442800162	0.350942739930668
MARES RIOS E MATAS	0.132149978987653	0.350257670300959
E.M. PROFESSORA ÁUREA ALVIM TÓFFOLI	0.232013535711005	0.340357562945786
EMEF JOSÉ PARENTE PRADO	8.68388710030213e-17	0.338345864661653
E.M. PROF. VILNA CORREA PRETTI	9.44177441700835e-17	0.335401643468053
E.E CARLOS MAXIMILIANO	0.0511673538043102	0.333852372383043
EMEF ANTÔNIO MENDES CARNEIRO	0.0105645126476804	0.333737820264195
SOLAR ORIENTE E SEC XXI	0.246066334320678	0.324041353867363
E.M. PROF. ILKA MULLER DE MELLO	2.1991152953533	0.304508091658811
ELVIRA PAOLILO BRAIDO	-0.948440204447929	0.303093963444351
EMEIEF JOÃO DE BARROS	0.194106205855361	0.298585073775857
CIE MARIA DIRCE DE CARVALHO DAVILA	0.178411643727526	0.29405124481508
E. M. REINEIRO DONATO PASTINA	0.0499764361224101	0.293250876714476
EB MUN EDITH WILLECKE	0.500964140223755	0.290529541665532
CENTRO INTEGRADO CHAPEUZINHO VERMELHO	-0.467124967648268	0.287128433911058
EMEB PHILOMENA SALVIA ZUPARDO	1.02317640480128	0.273586862070672
CENTRO RIO TAMISA	0.075543047958859	0.269309914807641
EMEF PROF WALDOMIRO GALLO	1.61130634232136	0.264187492181617
EMEIEF PROFª M. DAURIA S. OLIVEIRA	-0.434307781101961	0.261428654049277
EMEF. VALERIANO FONSECA	0.488887448030564	0.25297669865062
EMEF JOSE BENIGO GOMES	-0.421574868689122	0.252899049094415
JOSÉ DOMINGOS DE ALCÂNTARA	-0.0400428275935152	0.242747917379103
EMEF NAZARETH DE SIQUEIRA RANGEL BARBOSA	1.20167209101219	0.240384744953038
E.M DERMEVAL AROUCA PROF.	0.650579759510648	0.238034960692615
CEMUS I E V – TURMAS DE EJA	0.245413920694737	0.236851320193156
E.M. PROF. NEUSA MARIA REBELLO VIEIRA	2.66827802384999	0.236481748021977

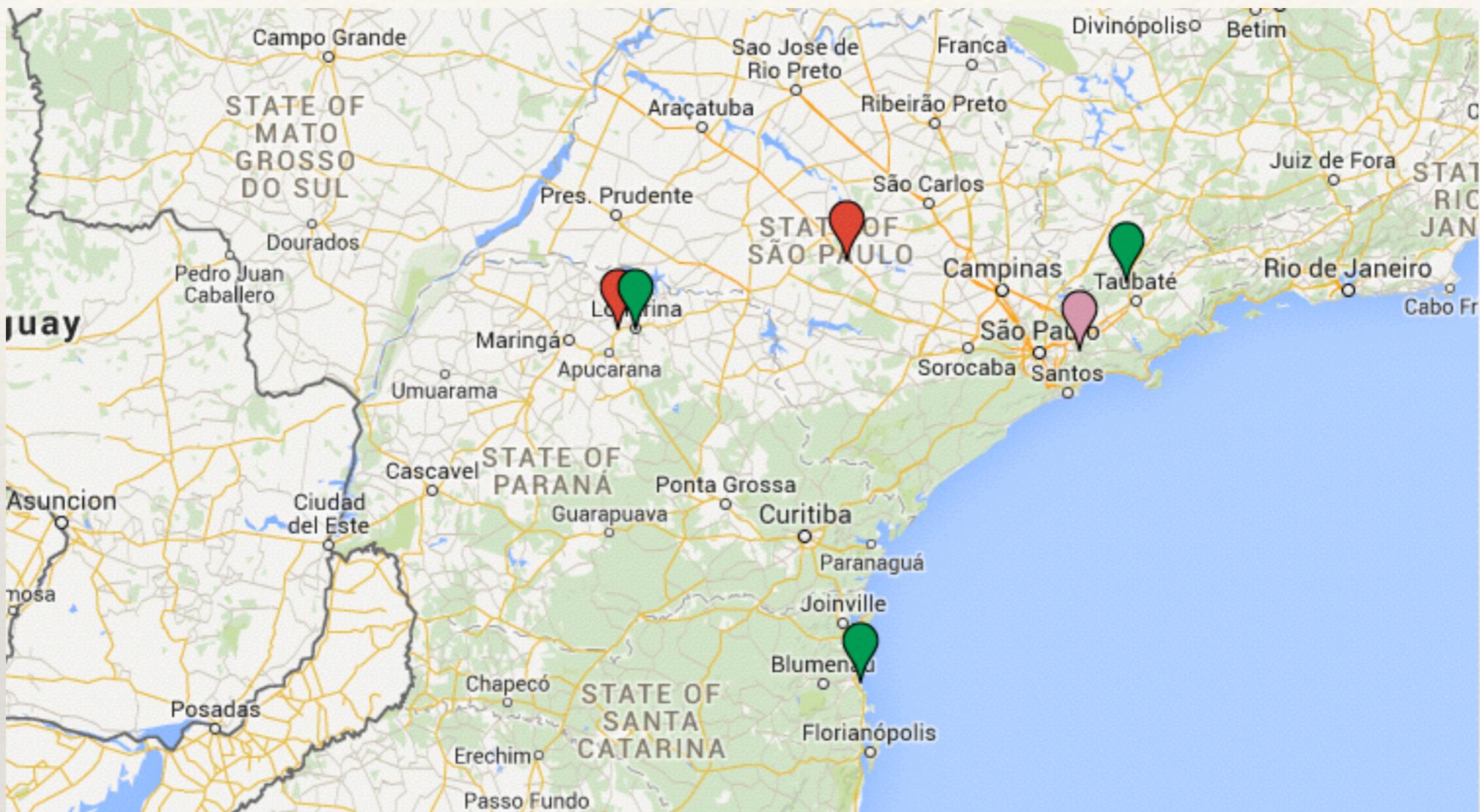
a mais rápida permite

crescer mais rápido



EM MONTEIRO LOBATO	4.66386957063584e-16	1.33630289532294
E.M. PROF BADIA DE FARIA	0.339066064677893	0.409285715091525
EM PREF. NOÉ ALVES FERREIRA	-0.224388064058562	0.394717910233839
E.M. PROF. MARIA HOSTIM DA COSTA	0.360647442800162	0.350942739930668
MARES RIOS E MATAS	0.132149978987653	0.350257670300959
E.M. PROFESSORA ÁUREA ALVIM TÓFFOLI	0.232013535711005	0.340357562945786
EMEF JOSÉ PARENTE PRADO	8.68388710030213e-17	0.338345864661653
E.M. PROF. VILNA CORREA PRETTI	9.44177441700835e-17	0.335401643468053
E.E CARLOS MAXIMILIANO	0.0511673538043102	0.333852372383043
EMEF ANTÔNIO MENDES CARNEIRO	0.0105645126476804	0.333737820264195
SOLAR ORIENTE E SEC XXI	0.246066334320678	0.324041353867363
E.M. PROF. ILKA MULLER DE MELLO	2.1991152953533	0.304508091658811
ELVIRA PAOLILO BRAIDO	-0.948440204447929	0.303093963444351
EMEIEF JOÃO DE BARROS	0.194106205855361	0.298585073775857
CIE MARIA DIRCE DE CARVALHO DAVILA	0.178411643727526	0.29405124481508
E. M. REINEIRO DONATO PASTINA	0.0499764361224101	0.293250876714476
EB MUN EDITH WILLECKE	0.500964140223755	0.290529541665532
CENTRO INTEGRADO CHAPEUZINHO VERMELHO	-0.467124967648268	0.287128433911058
EMEB PHILOMENA SALVIA ZUPARDO	1.02317640480128	0.273586862070672
CENTRO RIO TAMISA	0.075543047958859	0.269309914807641
EMEF PROF WALDOMIRO GALLO	1.61130634232136	0.264187492181617
EMEIEF PROFª M. DAURIA S. OLIVEIRA	-0.434307781101961	0.261428654049277
EMEF. VALERIANO FONSECA	0.488887448030564	0.25297669865062
EMEF JOSE BENIGO GOMES	-0.421574868689122	0.252899049094415
JOSÉ DOMINGOS DE ALCÂNTARA	-0.0400428275935152	0.242747917379103
EMEF NAZARETH DE SIQUEIRA RANGEL BARBOSA	1.20167209101219	0.240384744953038
E.M DERMEVAL AROUCA PROF.	0.650579759510648	0.238034960692615
CEMUS I E V – TURMAS DE EJA	0.245413920694737	0.236851320193156
E.M. PROF. NEUSA MARIA REBELLO VIEIRA	2.66827802384999	0.236481748021977

a mais rápida permite  
crescer mais rápido



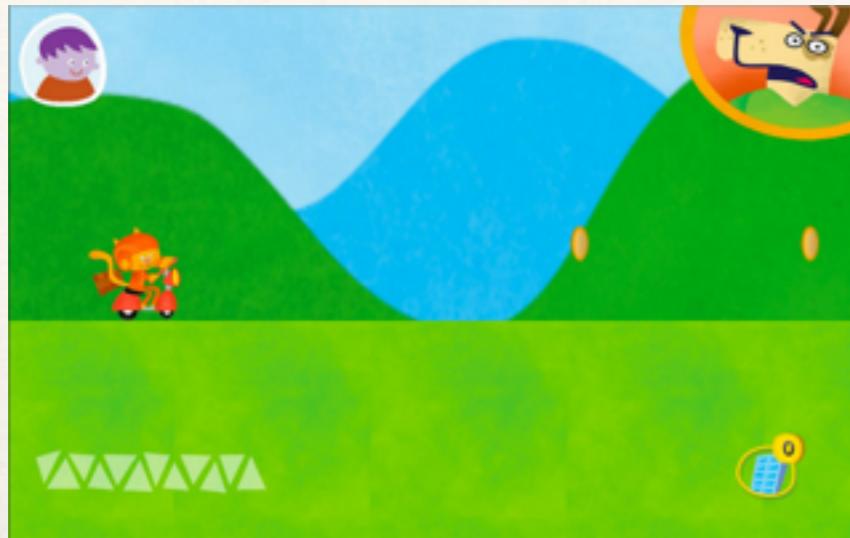
cruza com dados geográficos...

as “piores” podem estudar  
o que potencializou as “melhores”

as “piores” podem estudar  
o que potencializou as “melhores”

dado as limitações de todo tipo de rank, claro...

concluindo...



alfabetizaremos  
mais rápido



potencializaremos  
as escolas



---

# obrigado

@guilhermecaelum  
@mauricioaniche

---

slides extras

poderia ser uma falha...

de tanto procurar você encontra uma  
coincidência que é estatisticamente  
significante

```
> wilcox.test(x, y)
```

```
Wilcoxon rank sum test with continuity correction  
data: x and y  
W = 7244, p-value = 0.001301  
alternative hypothesis: true location shift is not equal to 0
```

mesmo 1 ou 2 vezes por semana tem significância estatística

# quantidade de acertos?

count(asserts_percentage)	asserts_percentage
16879	0
1307	25
25772	50
88375	100
1	150
24	200
7	300
1	400
1	866
3	2600

10 rows in set (0.06 sec)

vamos limpar, como nos outros

# quantidade de acertos?

count(asserts_percentage)	asserts_percentage
1	50
87525	100
1	150
24	200
7	300
1	400
1	866
3	2600

sem asserts=0

# quantidade de acertos?

count(asserts_percentage)	asserts_percentage
1	50
87525	100

sem asserts=0, sem asserts\_percentage>100

# quantidade de acertos?

count(asserts_percentage)	asserts_percentage
1	50
87525	100

sem asserts=0, sem asserts\_percentage>100

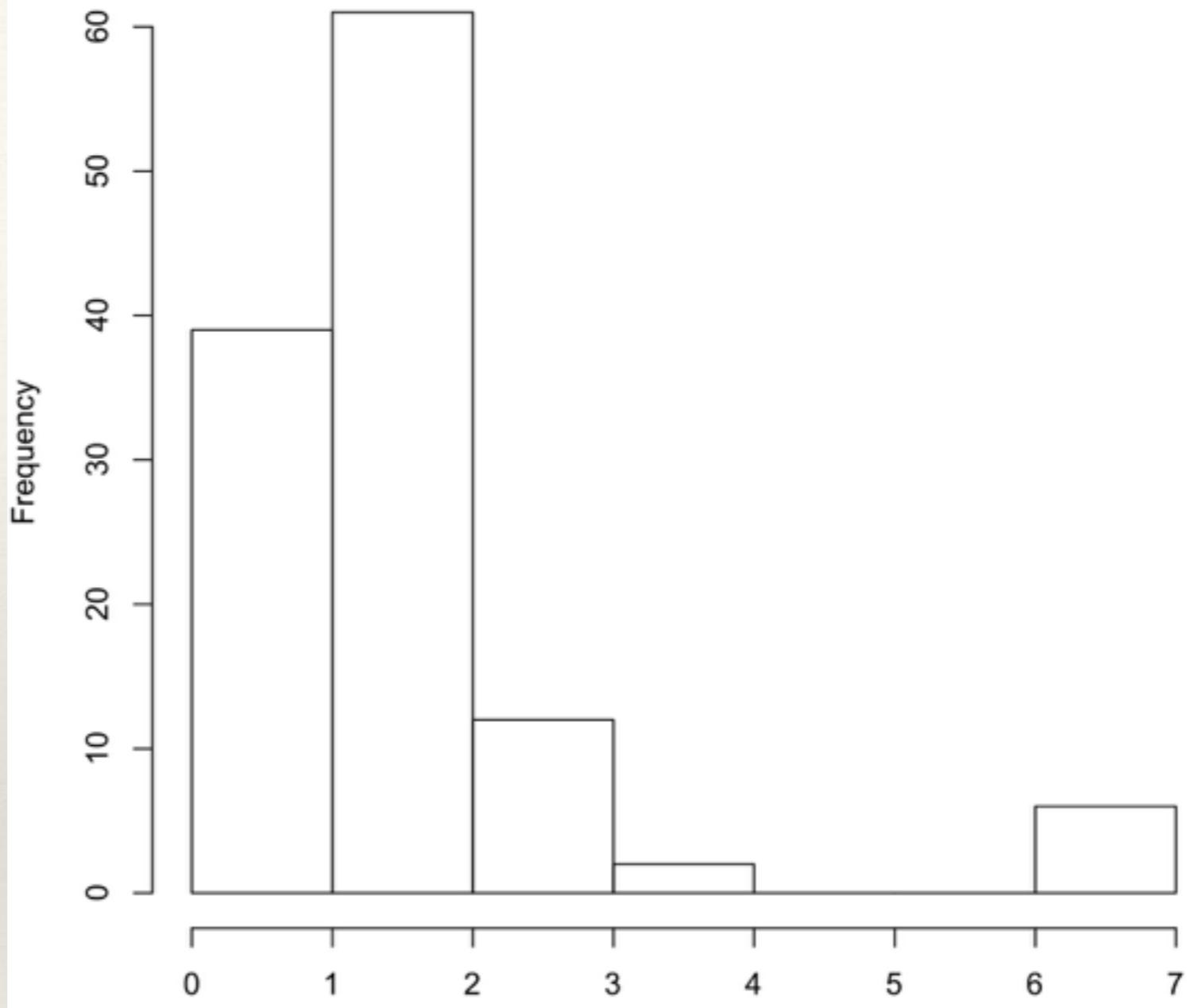
todo mundo acerta...

# todo mundo acerta?

count(asserts_percentage)	asserts_percentage
1	50
87525	100

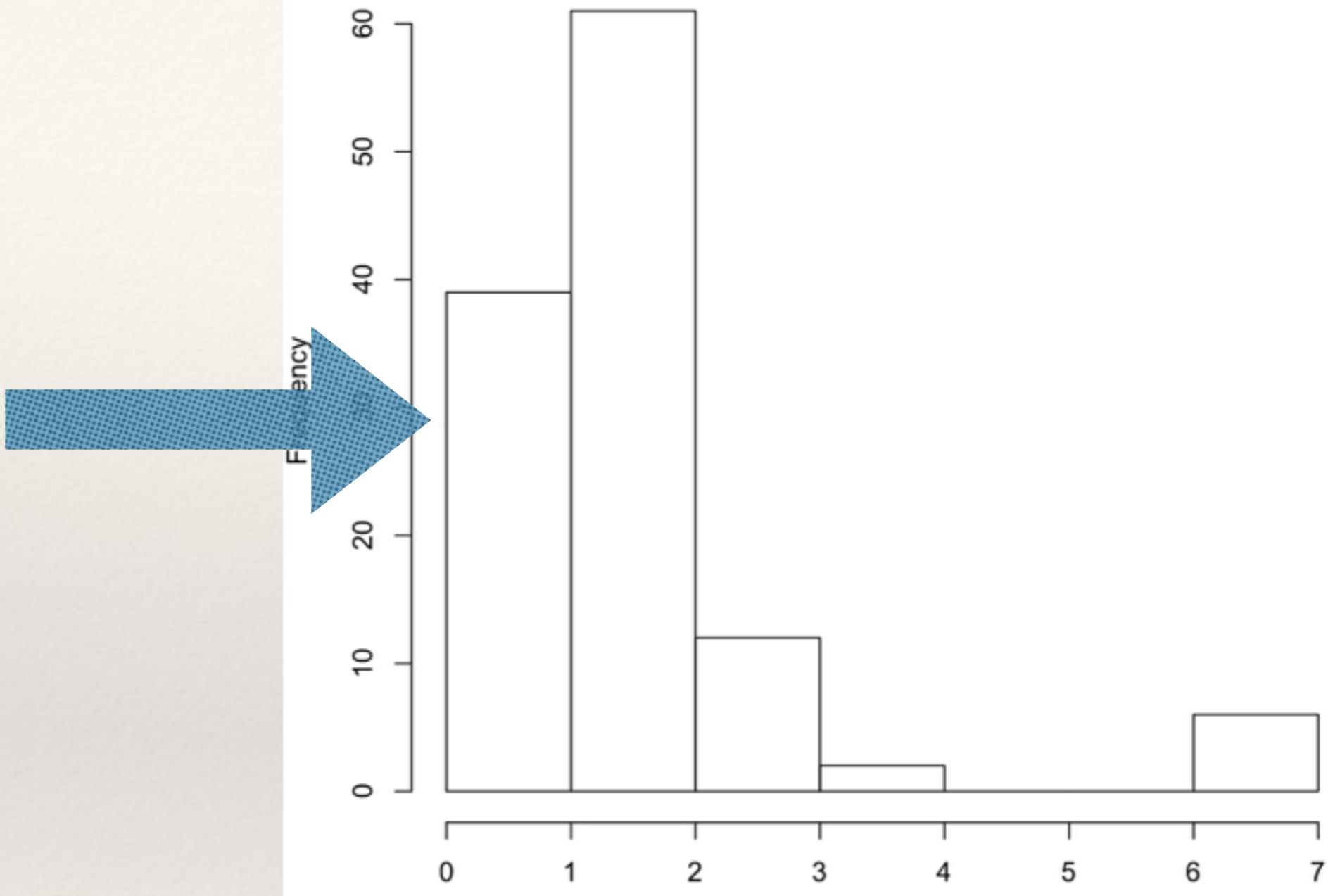
o mesmo “problema” que temos no Alura

Histogram of rs\$dias\_por\_semana



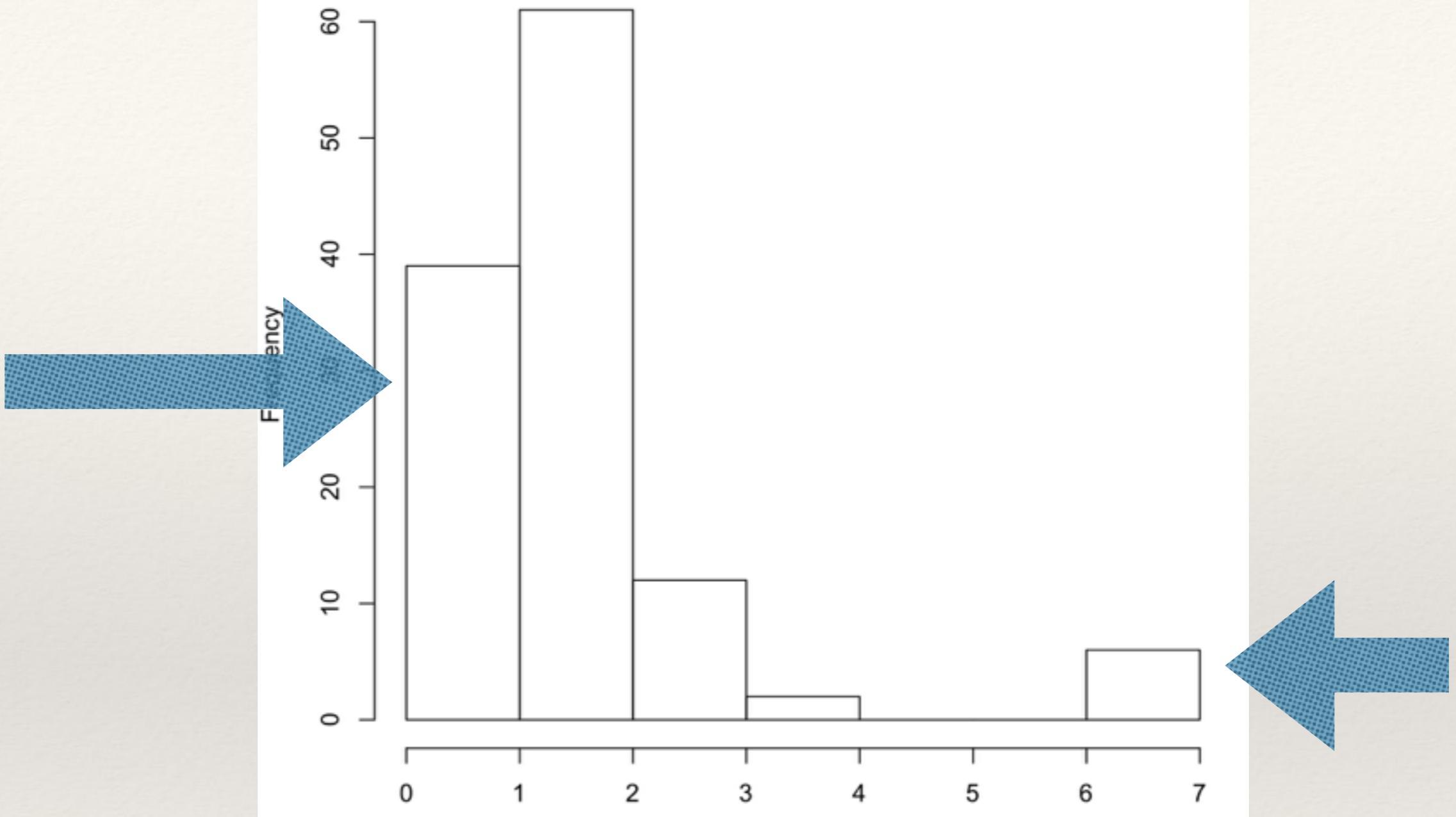
vezes por semana em uma escola

Histogram of rs\$dias\_por\_semana



vezes por semana em uma escola

Histogram of rs\$dias\_por\_semana



vezes por semana em uma escola