

Intro_pt

Dr. Darren Norris (dnorris75@gmail.com)

31 de Março de 2020

Apresentação

Aqui vamos pesquisar desmatamento entre 2000 - 2018 ao longo de 276 quilômetros de rios na Amazônia Brasileira. As pesquisas fazem parte de atividades desenvolvidos no projeto [Where is my Turtle](https://myturtlebrazil.wixsite.com/whereismyturtle) : <https://myturtlebrazil.wixsite.com/whereismyturtle> .

- Objetivo não é de apresentar detalhes sobre os cálculos/métodos estatísticas ou os funções no [R](#).
- Mas, sim, o objetivo é de apresentar um exemplo mostrando as capacidades e opções para desenvolver e integrar pesquisas científicas no ambiente estatística de [R](#)

Com menos de 100 linhas de código vamos: 1) carregar dados; 2) olhar mapas (SIG); 3) gerar tabelas com resumos; 4) visualizar resultados em gráficos; e 5) rodar análises estatísticas.

Pacotes

Para fazer todo isso em menos de 100 linhas, precisamos dos seguintes pacotes, que deve estar instalado antes: [plyr](#), [tidyverse](#), [sf](#), [mapview](#), [sjPlot](#), [sjmisc](#), [sjlabelled](#), [interactions](#).

Portanto, deve instalar os pacotes necessários antes de começar:

```
install.packages(c("plyr", "tidyverse", "raster", "sf", "mapview",  
                  "sjPlot", "sjmisc", "sjlabelled", "interactions"))
```

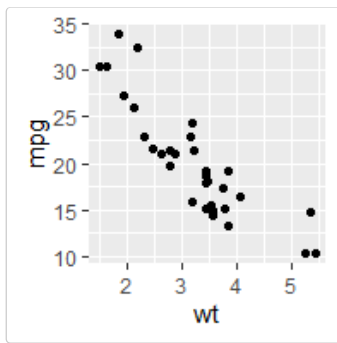
Carregar pacotes:

```
library(plyr)  
library(tidyverse)  
library(sf)  
library(mapview)  
library(sjPlot)  
library(sjmisc)  
library(sjlabelled)  
library(interactions)
```

Agora podemos fazer um gráfico com função “ggplot” (pacote [ggplot2](#)), que faz parte do “tidyverse”. Mais exemplos no [R cookbook](#) : <http://www.cookbook-r.com/Graphs/> .

Primeiro gráfico com os dados “mtcars” no pacote ggplot2:

```
ggplot2::ggplot(mtcars, aes(x = wt, y = mpg)) + geom_point()
```



Para entender melhor pode verificar ajudar no R:

```
?mtcars  
?geom_point
```

Mais exemplos com [Scatterplots](#) aqui: [http://www.cookbook-r.com/Graphs/Scatterplots_\(ggplot2\)/](http://www.cookbook-r.com/Graphs/Scatterplots_(ggplot2)/)

1) Carregar arquivos com dados de SIG.

Baixar arquivo com os dados (formato “GPKG”, tamanho 54.3 MB).

Link: <https://github.com/darrennorris/gisdata/blob/master/inst/vector/rivers.GPKG> . Lembrando-se de salvar o arquivo (“rivers.GPKG”) em um local conhecido no seu computador.

Agora avisar R sobre onde ficar o arquivo. O código abaixo vai abrir uma nova janela, e você deve buscar e selecionar o arquivo “rivers.GPKG”:

```
meuSIG <- file.choose()
```

Agora vamos olhar o que tem no arquivo.

```
sf::st_layers(meuSIG)
```

Existem camadas diferentes com pontos e linhas:

```
#> Driver: GPKG
#> Available layers:
#>   layer_name geometry_type features fields
#> 1 midpoints      Point        52      15
#> 2 centerline   Line String     52      15
#> 3 forestloss    Point    276086     12
```

1.1) Carregar dados (vectors)

O código abaixo vai carregar os dados e criar 3 objetos “rsm”, “rsl” e “fl”. Agora temos dados com: pontos cada 5 km ao longo os rios (rsm) ; linha central de rios (“rsl”) e pontos cada metro ao longos os rios (“fl”).

```
rsm <- sf::st_read(meuSIG, layer = "midoints")
rsl <- sf::st_read(meuSIG, layer = "centerline")
fl <- sf::st_read(meuSIG, layer = "forestloss")
```

2) Mapas

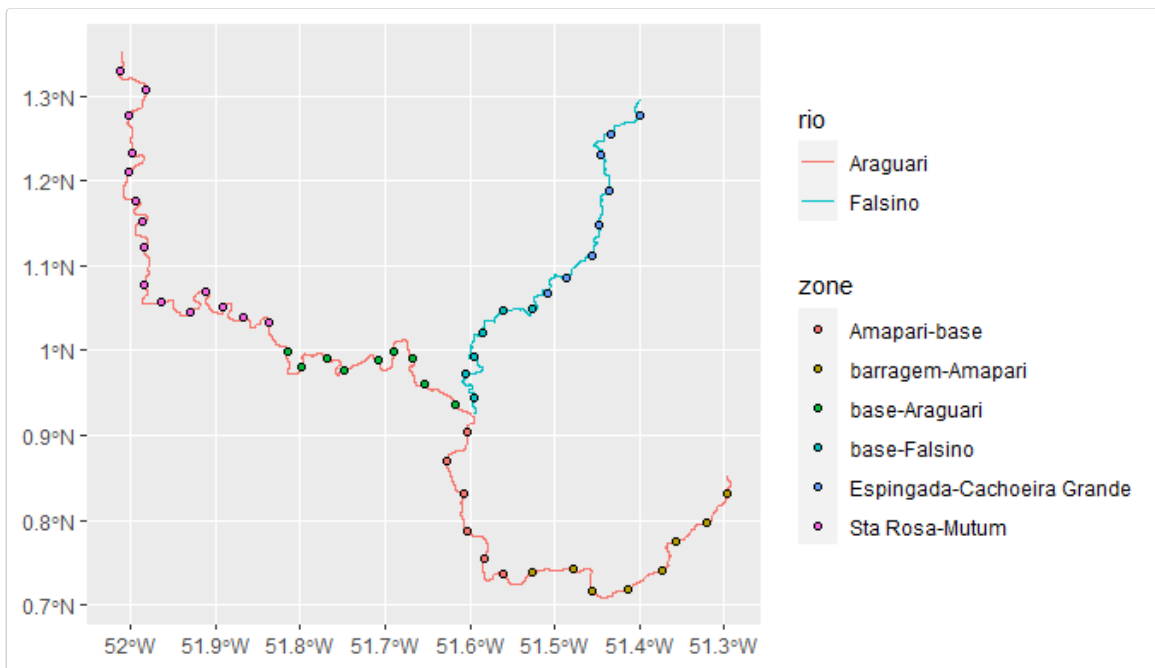
2.1) Mapa com pontos cada metro de rio

Objeto “fl” com 276086 pontos - muitos pontos cerca de 5 minutos para concluir....

```
ggplot2::ggplot(fl) + geom_sf(data = fl, aes(color=zone))
```

2.2) Mapa com linha central e pontos de rios em trechos de 5km

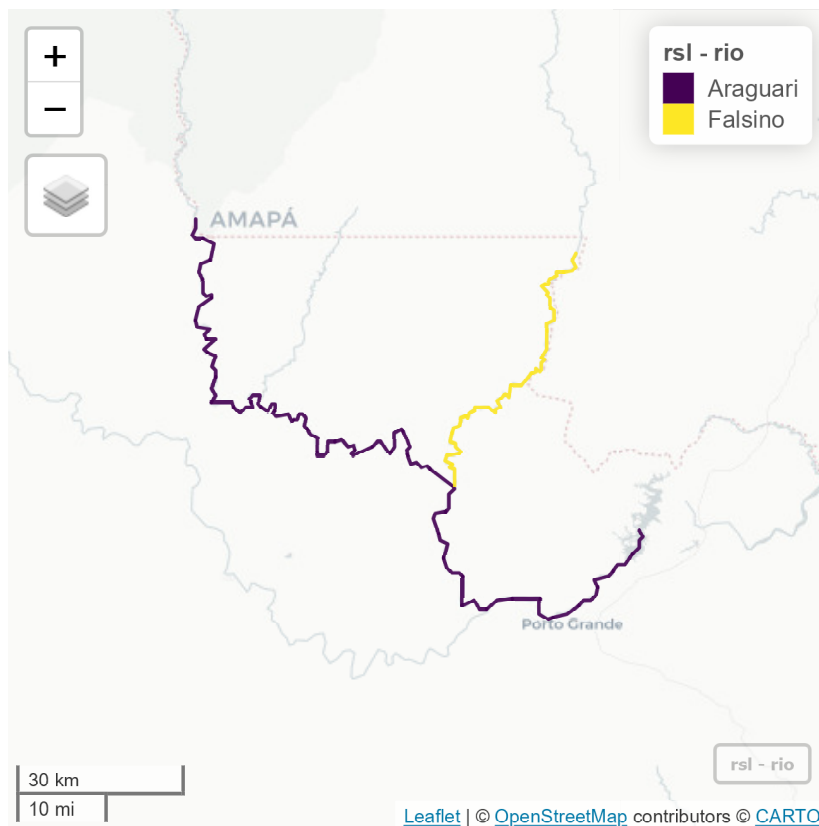
```
ggplot(rsl) +  
  geom_sf(aes(color=rio)) +  
  geom_sf(data = rsm, shape=21, aes(fill=zone))
```



2.3) Interativo (funcione somente com internet)

Mostrando agora com fundo de mapas “base” (OpenStreetMap/ESRI etc)

```
mapview(rsl, zcol = "rio") + mapview(rsm, zcol = "zone")
```



Mais exemplos com mapas e dados espaciais no R:

[sf e ggplot2](https://www.r-spatial.org/r/2018/10/25/ggplot2-sf.html) : <https://www.r-spatial.org/r/2018/10/25/ggplot2-sf.html>

[Capítulo 8](#) no livro [Geocomputation with R](#) : <https://geocompr.robinlovelace.net/adv-map.html>

3) Resumo de dados na tabela de atributos

3.1) Desmatamento por rio e por seis zonas

Agora vamos fazer alguns resumos com os dados na tabela de atributos. No arquivo “fl” temos dados mostrando valores de [desmatamento](https://science.sciencemag.org/content/342/6160/850), “Global Forest Change” desenvolvido por Hansen e co-autores 2013 (<https://science.sciencemag.org/content/342/6160/850>) e obtidos de no site https://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.6.html.

Os valores calculados são uma proporção de área desmatado nos anos 2000-2018 em raios de 500 metros, 1 quilometro, 5 quilômetros e 10 quilômetros. Os dados em pontos cada metro ao longo de 276 quilômetros de rios (276086 pontos) . Vamos calcular comprimento de rio e valor media de desmatamento em seis zonas diferentes:

```
plyr::ddply(data.frame(fl), .(rio, zone), summarise,
  rio_km = round((length(dsmt500) / 1000),1),
  fl_500m = round(mean(dsmt500),3),
  fl_1km = round(mean(X1KM_ds),3),
  fl_5km = round(mean(X5km_ds),3),
  fl_10km = round(mean(X10km_d),3)
)

#>      rio                zone rio_km fl_500m fl_1km fl_5km fl_10km
#> 1 Araguari      Amapari-base   32.5   0.092  0.075  0.031  0.032
#> 2 Araguari      barragem-Amapari 42.7   0.167  0.216  0.186  0.178
#> 3 Araguari      base-Araguari  47.4   0.007  0.004  0.002  0.003
#> 4 Araguari      Sta Rosa-Mutum  78.6   0.009  0.008  0.004  0.005
#> 5 Falsino       base-Falsino   32.2   0.015  0.012  0.005  0.004
#> 6 Falsino Espingada-Cachoeira Grande 42.8   0.002  0.001  0.001  0.003
```

3.2) Desmatamento por rio e por 52 subzonas

Agora, repetindo o mesmo processo, mas acrescentando resumos para 52 subzonas (cerca de 5 km cada) para um análise mais preciso.

```
dfsubzona <- plyr::ddply(data.frame(fl), .(rio, zone, subz_id), summarise,
  afl_500m = round(mean(dsmt500),3),
  bfl_1km = round(mean(X1KM_ds),3),
  cfl_5km = round(mean(X5km_ds),3),
  dfl_10km = round(mean(X10km_d),3)
)
dfsubzona$zone <- relevel(dfsubzona$zone, ref="barragem-Amapari")
```

Visualizar tabela, aproveitando funções disponíveis no pacote “sjPlot”:

```
sjPlot::tab_df(dfsubzona)
```

rio	zone	subz_id	afl_500m	bfl_1km	cfl_5km	dfl_10km
Araguari	Amapari-base	10	0.14	0.12	0.03	0.05
Araguari	Amapari-base	11	0.14	0.10	0.03	0.02
Araguari	Amapari-base	12	0.09	0.09	0.02	0.01
Araguari	Amapari-base	13	0.04	0.05	0.02	0.01
Araguari	Amapari-base	14	0.09	0.06	0.01	0.01
Araguari	Amapari-base	9	0.05	0.05	0.07	0.09

Araguari	barragem-Amapari	1	0.16	0.38	0.23	0.16
Araguari	barragem-Amapari	2	0.11	0.14	0.16	0.20
Araguari	barragem-Amapari	3	0.06	0.13	0.19	0.21
Araguari	barragem-Amapari	4	0.08	0.13	0.20	0.23
Araguari	barragem-Amapari	5	0.46	0.34	0.19	0.20
Araguari	barragem-Amapari	6	0.21	0.29	0.20	0.17
Araguari	barragem-Amapari	7	0.15	0.19	0.18	0.13
Araguari	barragem-Amapari	8	0.11	0.13	0.14	0.12
Araguari	base-Araguari	15	0.00	0.00	0.01	0.01
Araguari	base-Araguari	16	0.00	0.00	0.00	0.00
Araguari	base-Araguari	17	0.01	0.01	0.00	0.00
Araguari	base-Araguari	18	0.03	0.01	0.00	0.00
Araguari	base-Araguari	19	0.00	0.00	0.00	0.00
Araguari	base-Araguari	20	0.00	0.00	0.00	0.00
Araguari	base-Araguari	21	0.00	0.00	0.00	0.00
Araguari	base-Araguari	22	0.00	0.00	0.00	0.00
Araguari	base-Araguari	23	0.00	0.00	0.00	0.00
Araguari	Sta Rosa-Mutum	24	0.00	0.00	0.00	0.00
Araguari	Sta Rosa-Mutum	25	0.00	0.00	0.00	0.01
Araguari	Sta Rosa-Mutum	26	0.00	0.00	0.01	0.01
Araguari	Sta Rosa-Mutum	27	0.10	0.06	0.01	0.01
Araguari	Sta Rosa-Mutum	28	0.02	0.04	0.01	0.01
Araguari	Sta Rosa-Mutum	29	0.00	0.00	0.01	0.01
Araguari	Sta Rosa-Mutum	30	0.00	0.00	0.00	0.01
Araguari	Sta Rosa-Mutum	31	0.00	0.00	0.00	0.00
Araguari	Sta Rosa-Mutum	32	0.00	0.00	0.00	0.00
Araguari	Sta Rosa-Mutum	33	0.00	0.00	0.00	0.00
Araguari	Sta Rosa-Mutum	34	0.00	0.00	0.00	0.00
Araguari	Sta Rosa-Mutum	35	0.00	0.00	0.00	0.00
Araguari	Sta Rosa-Mutum	36	0.00	0.00	0.00	0.00
Araguari	Sta Rosa-Mutum	37	0.00	0.00	0.00	0.00
Araguari	Sta Rosa-Mutum	38	0.00	0.00	0.00	0.00
Falsino	base-Falsino	39	0.01	0.01	0.01	0.01
Falsino	base-Falsino	40	0.02	0.02	0.01	0.00
Falsino	base-Falsino	41	0.05	0.04	0.01	0.00

Falsino	base-Falsino	42	0.01	0.00	0.00	0.00
Falsino	base-Falsino	43	0.00	0.00	0.00	0.00
Falsino	base-Falsino	44	0.00	0.00	0.00	0.00
Falsino	Espingada-Cachoeira Grande	45	0.00	0.00	0.00	0.00
Falsino	Espingada-Cachoeira Grande	46	0.00	0.00	0.00	0.00
Falsino	Espingada-Cachoeira Grande	47	0.00	0.00	0.00	0.00
Falsino	Espingada-Cachoeira Grande	48	0.00	0.00	0.00	0.00
Falsino	Espingada-Cachoeira Grande	49	0.00	0.00	0.00	0.00
Falsino	Espingada-Cachoeira Grande	50	0.00	0.00	0.00	0.00
Falsino	Espingada-Cachoeira Grande	51	0.00	0.00	0.00	0.00
Falsino	Espingada-Cachoeira Grande	52	0.00	0.00	0.00	0.00

Exportar tabela com resumos ("dfsubzona") em formato "html" para word:

```
tab_df(dfsubzona, file = "subzona.html")
```

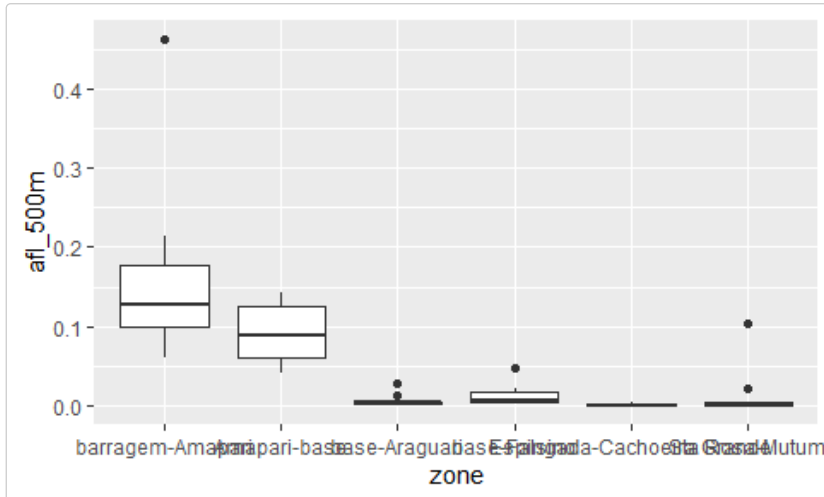
Exportar tabela com resumos ("dfsubzona") em formato para excel:

```
write.csv2(dfsubzona, file = "subzona.csv", row.names=FALSE)
```


4) Gráficos

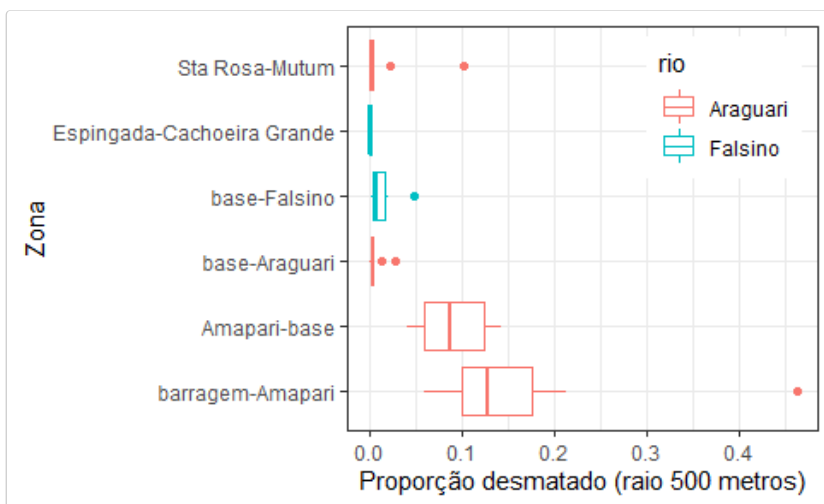
4.1) mostrando as diferenças em desmatamento entre zonas

```
ggplot2::ggplot(dfsubzona, aes(x = zone, y = afl_500m)) +
  geom_boxplot()
```



Agora, através de ajustes no código, vamos fazer um gráfico mais elegante e mais informativo.....

```
ggplot2::ggplot(dfsubzona, aes(x = zone, y = afl_500m)) +
  geom_boxplot(aes(color=rio)) +
  ylab("Proporção desmatado (raio 500 metros)") + xlab("Zona") +
  coord_flip() +
  theme_bw() + theme(legend.position=c(0.8, 0.8))
```



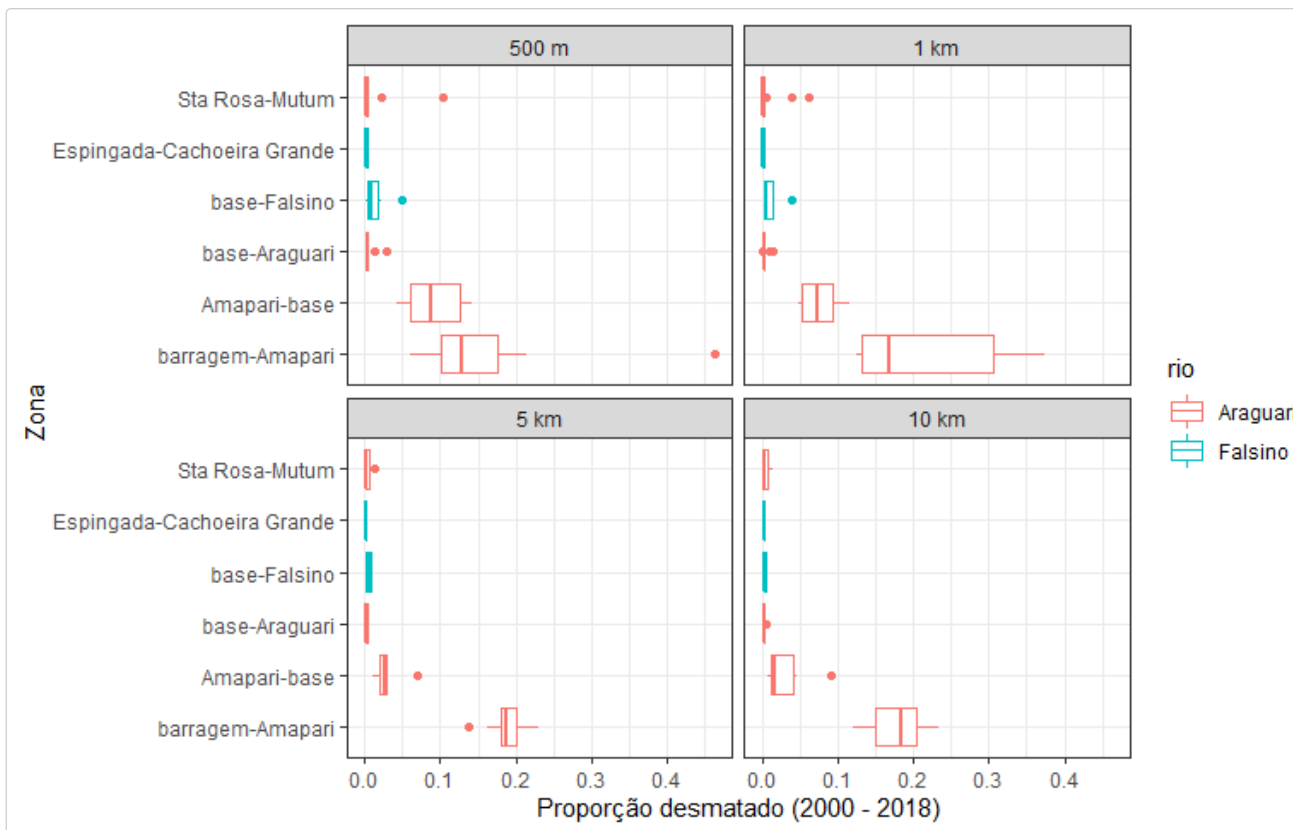
4.2) Gráfico comparando desmatamento entre raios de 500m, 1km, 5km e 10km

Primeiro preciso reorganizar os dados para facilitar apresentação visual:

```
baplot <- pivot_longer(dfsbzona, cols = c("afl_500m", "bfl_1km",
                                           "cfl_5km", "dfl_10km"),
                      names_to = "variable", values_to = "desmat")
baplot$variablef <- as.factor(baplot$variable)
levels(baplot$variablef) <- c("500 m", "1 km", "5 km", "10 km")
```

Agora o gráfico:

```
ggplot2::ggplot(baplot, aes(x = zone, y = desmat)) +
  geom_boxplot(aes(color=rio)) +
  ylab("Proporção desmatado (2000 - 2018)") + xlab("Zona") +
  coord_flip() +
  theme_bw() +
  facet_wrap(~variablef)
```



5) Análises estatística, comparando desmatamento entre zonas e raios

Veja guias mostrando análises estatísticas no R para entender melhor: [Capítulo 12](#) no livro “Ciência de Dados com R—Introdução.....”: <https://cdr.ibpad.com.br/modelos.html>

Modelo adequado para dados de proporção (valores de 0-1). Aqui um modelo comparando efeito de raio (500 m, 1 km, 5 km, 10 km) e zona sobre proporção de desmatamento:

```
glm1 <- glm(desmat ~ variablef + zone, family = quasibinomial, data = baplot)
```

Agora, mostrando três formas diferentes de apresentar resumo do modelo: Primeiramente com base R:

```
summary(glm1)
#>
#> Call:
#> glm(formula = desmat ~ variablef + zone, family = quasibinomial,
#>      data = baplot)
#>
#> Deviance Residuals:
#>      Min       1Q   Median       3Q      Max
#> -0.40389 -0.07675 -0.03754  0.02427  0.60305
#>
#> Coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)      -1.3706     0.1163  -11.789  < 2e-16 ***
#> variablef1 km       0.1064     0.1526   0.697   0.4865
#> variablef5 km      -0.2529     0.1629  -1.552   0.1222
#> variablef10 km     -0.2763     0.1637  -1.687   0.0931 .
#> zoneAmapari-base   -1.3315     0.1482  -8.982  < 2e-16 ***
#> zonebase-Araguari  -4.0406     0.3994 -10.116  < 2e-16 ***
#> zonebase-Falsino   -3.2524     0.3337  -9.747  < 2e-16 ***
#> zoneEspingada-Cachoeira Grande -4.9015     0.6441  -7.610  1.07e-12 ***
#> zoneSta Rosa-Mutum -3.5434     0.2480 -14.290  < 2e-16 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> (Dispersion parameter for quasibinomial family taken to be 0.02251878)
#>
#> Null deviance: 21.0637 on 207 degrees of freedom
#> Residual deviance: 3.2369 on 199 degrees of freedom
#> AIC: NA
#>
#> Number of Fisher Scoring iterations: 9
```

E com funções disponíveis no pacote “sjPlot”:

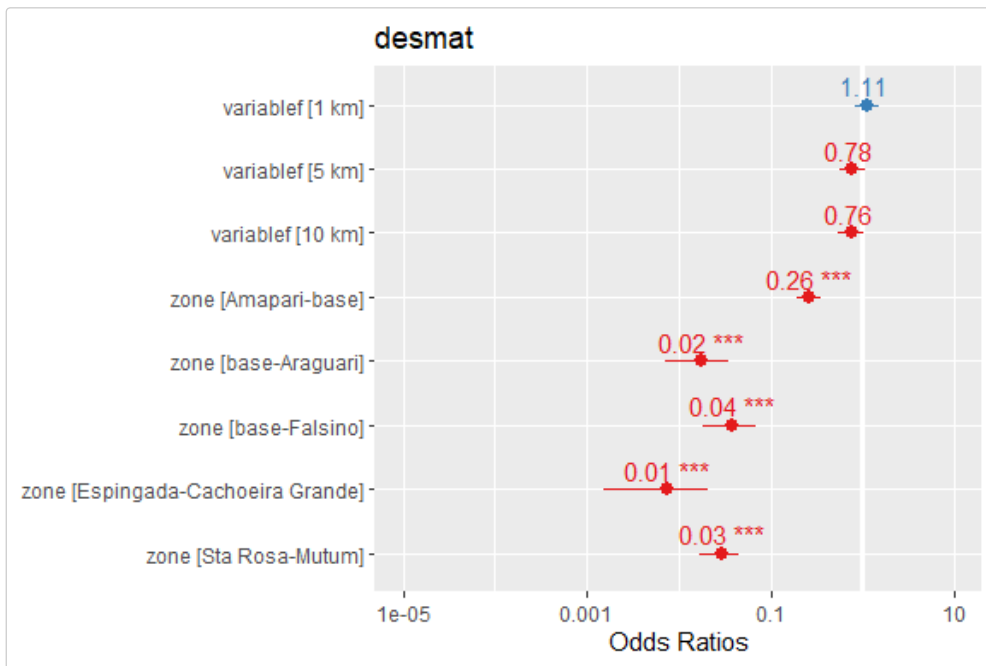
```
sjPlot::tab_model(glm1)
```

desmat			
Predictors	Odds Ratios	CI	p
(Intercept)	0.25	0.20 – 0.32	<0.001
variablef [1 km]	1.11	0.82 – 1.50	0.487
variablef [5 km]	0.78	0.56 – 1.07	0.122
variablef [10 km]	0.76	0.55 – 1.04	0.093
zone [Amapari-base]	0.26	0.20 – 0.35	<0.001

zone [base-Araguari]	0.02	0.01 – 0.04	<0.001
zone [base-Falsino]	0.04	0.02 – 0.07	<0.001
zone [Espingada-Cachoeira Grande]	0.01	0.00 – 0.02	<0.001
zone [Sta Rosa-Mutum]	0.03	0.02 – 0.05	<0.001
Observations	208		
R ² Tjur	0.007		

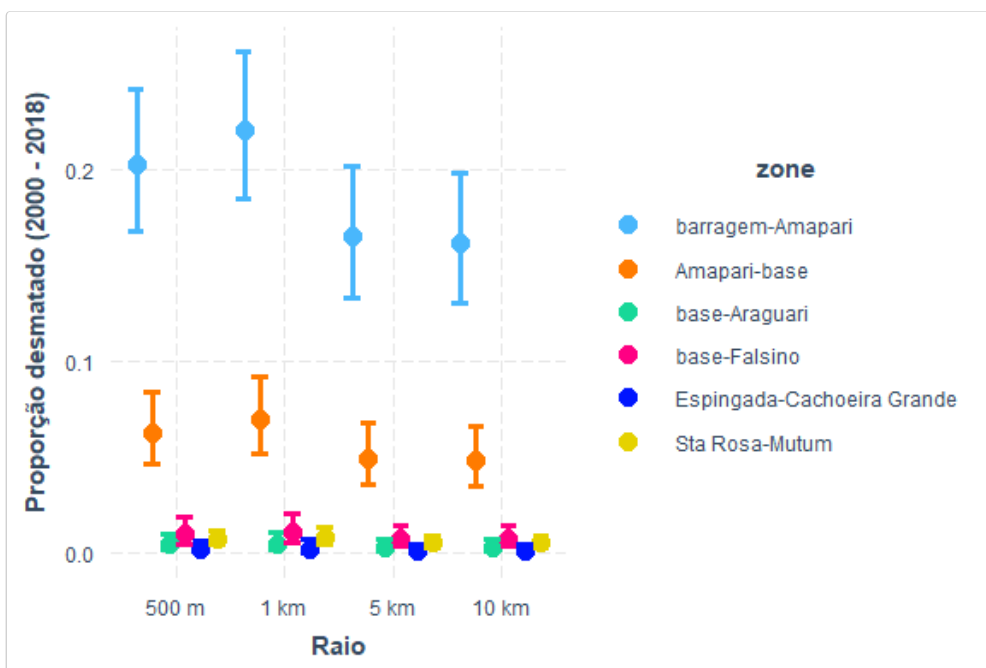
Gráfico de valores do modelo, no estilo “Forest-plot”

```
sjPlot::plot_model(glm1, show.values = TRUE, value.offset = .3)
```



E com função “cat_plot” disponível no pacote “interactions”:

```
interactions::cat_plot(glm1, pred = variablef, modx = zone, x.label = "Raio",  
  y.label = "Proporção desmatado (2000 - 2018)")
```



6) Finalizar.

Salvar uma cópia de dados para uso futuro.

```
save.image("intro.RData")
```

7) Leitura

Para quem buscar mais detalhes sobre as funções/métodos estatísticos no **R**:

Livro [Processamento e Análise de Dados](https://www.msperlin.com/padfeR/index.html) : <https://www.msperlin.com/padfeR/index.html>

Livro [Ciência de Dados com R](https://cdr.ibpad.com.br/) : <https://cdr.ibpad.com.br/>

Além disso, existem fontes variadas de ajudar online:

[cursos](https://pt.coursera.org/learn/r-programming) : <https://pt.coursera.org/learn/r-programming>;

[R-br](http://r-br.2285057.n4.nabble.com/) a lista Brasileira oficial de discussão do programa R: <http://r-br.2285057.n4.nabble.com/>,

Capítulos 1 a 3 no livro “Processamento e Análise de Dados.....”: <https://www.msperlin.com/padfeR/index.html>

Capítulos 1 a 2 no livro “Ciência de Dados com R–Introdução.....”: <https://cdr.ibpad.com.br/cdr-intro.pdf>

Inglês

Instruções de instalação: <https://moderndive.netlify.com/1-getting-started.html>

Com 3 aulas básicas: <https://rladiessydney.org/courses/ryouwithme/01-basicbasics-0/>

Foruns: [rcomputing](https://www.facebook.com/rcomputing/) : <https://www.facebook.com/rcomputing/> ;

<https://www.r-bloggers.com/> exemplo com PCA:

<https://www.r-bloggers.com/computing-and-visualizing-pca-in-r/>