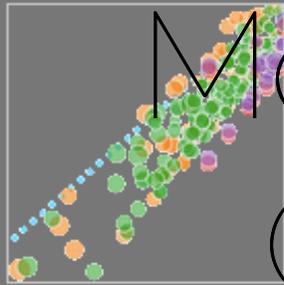




# PlanECO 2018

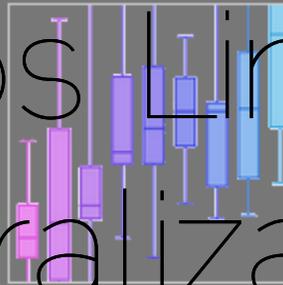
Line and Scatter Plots



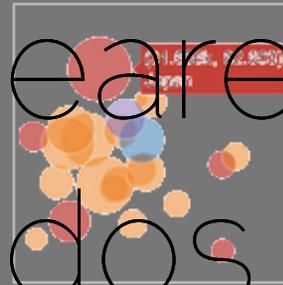
Bar Charts



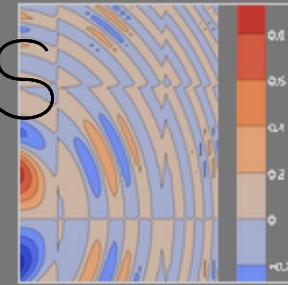
Box Plots



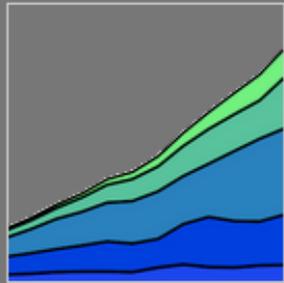
Bubble Charts



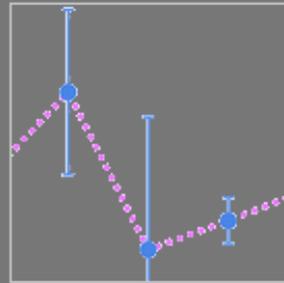
Contour Plots



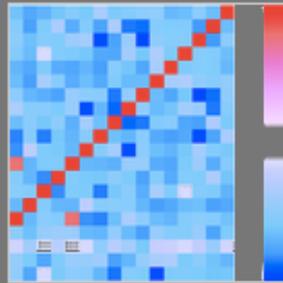
Filled Area Plots



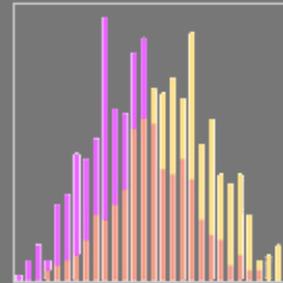
Error Bars



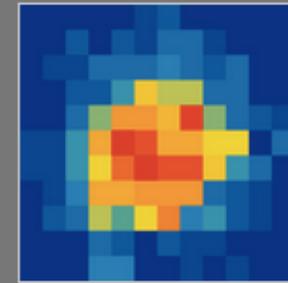
Heatmaps



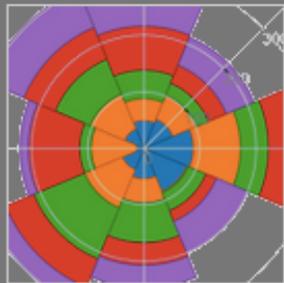
Histograms



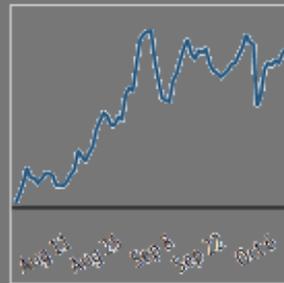
2D Histograms



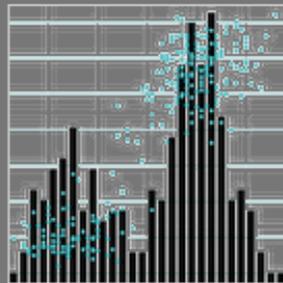
Polar Charts



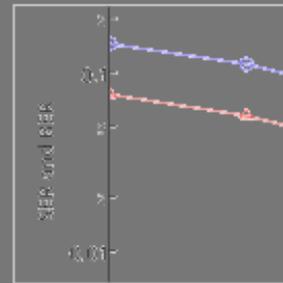
Time Series



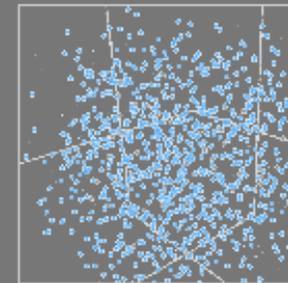
Multiple Chart Types



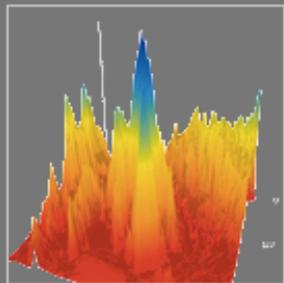
Log Plots



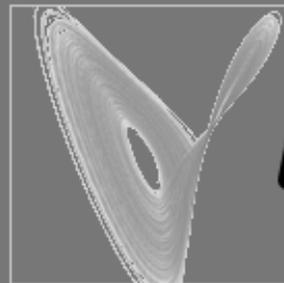
3D Scatter Plots



3D Surface Plots



3D Line Plots



# Modelos Lineares Generalizados

# PIAnEco

# Modelos Lineares Generalizados (GLMs)

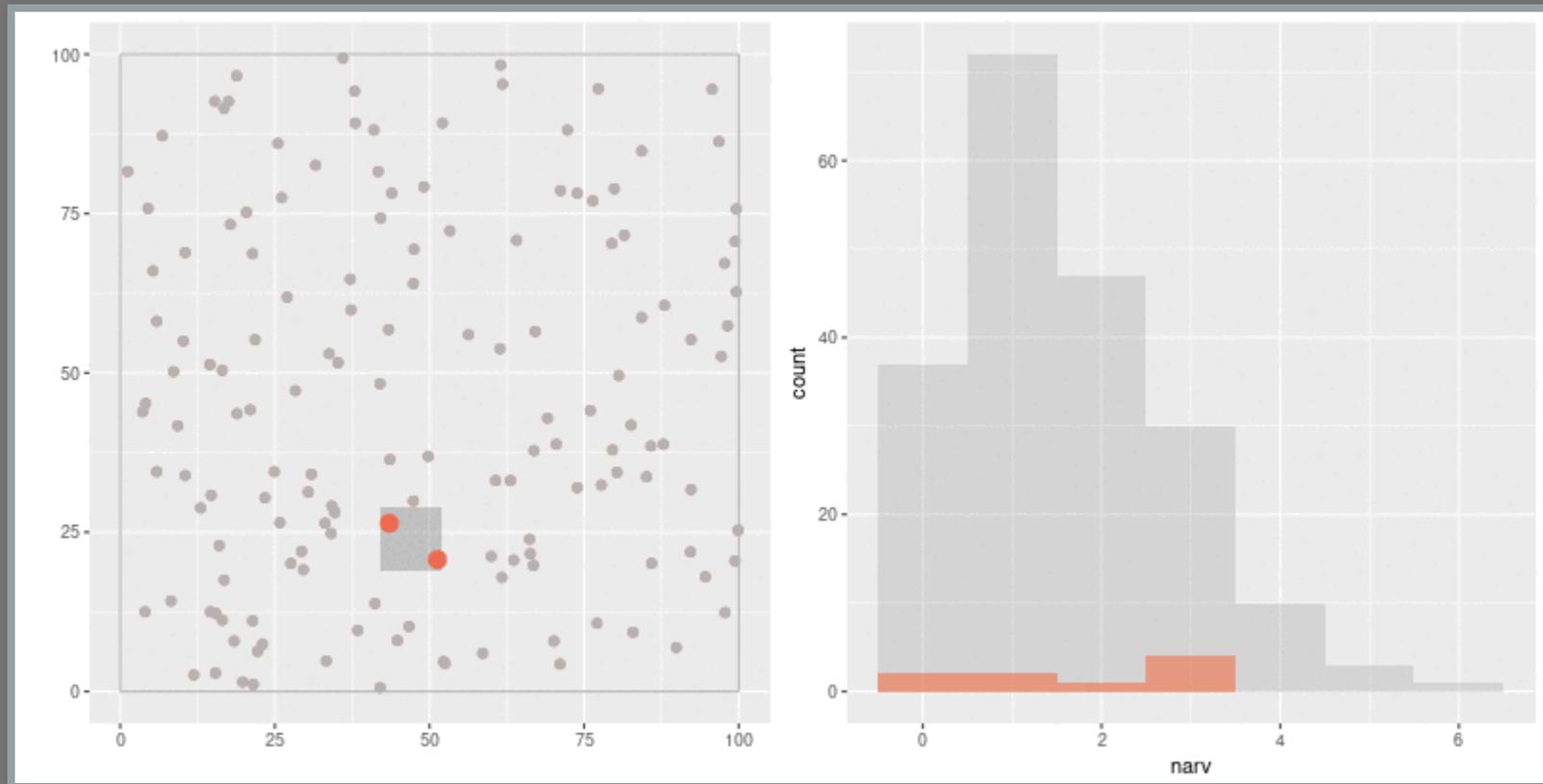
## Conceitos

- estrutura do erro
- preditora linear
- função de ligação
- inverso da função de ligação

# Estrutura do Erro

## Contagem de árvores em Parcelas

Completa Aleatoriedade Espacial



# GLMs: conceitos

## Função de ligação

- a função de ligação é aplicada na **Esperança** (valor esperado) da variável resposta aleatória:

$$\eta = g(E(Y))$$

## Preditor Linear

- o preditor  $\eta$  é a resposta linear da soma de uma ou mais ( $p$ ) preditoras

$$\eta = \alpha + \sum_{j=1}^p \beta_j x_{ij}$$

## Função inversa

- a função inversa da ligação retorna à escala da preditora

$$Y = g(\eta)^{-1}$$

# Função de ligação: canônica

- dependendo da estrutura de erro há uma função de ligação padrão

resposta	resíduos	ligação
contínua	gaussiano	identidade
contagem	poisson	log
proporção	binomial	logit
binária	binomial	logit

LMs são GLMs com função de ligação identidade

$$\eta = I(E(Y))$$
$$\eta = E(Y)$$

# GLMs x Transformar LMs

Modelo Transformado de LMs

$$f(\hat{Y} + \epsilon)$$

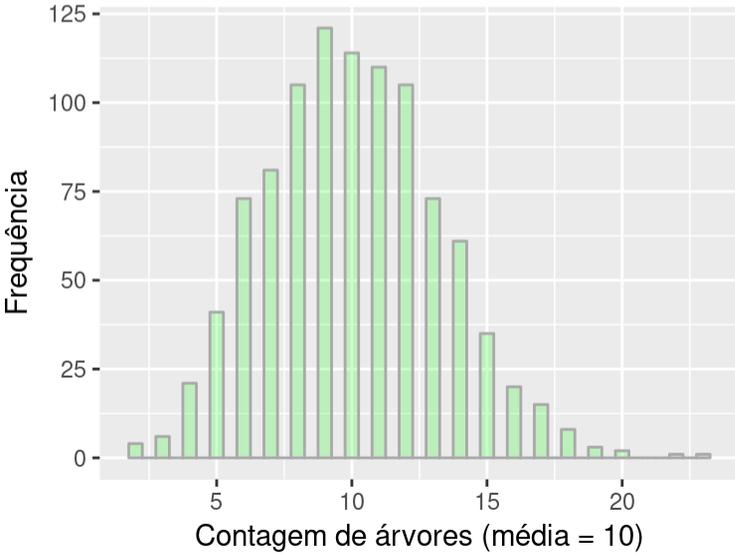
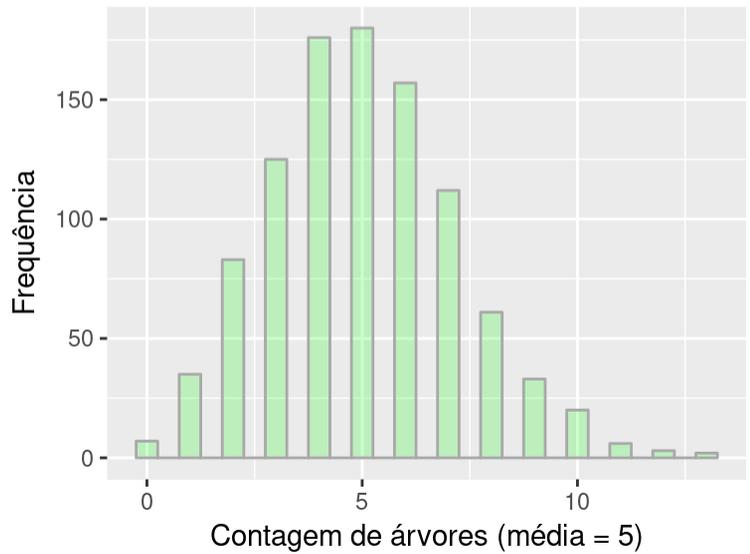
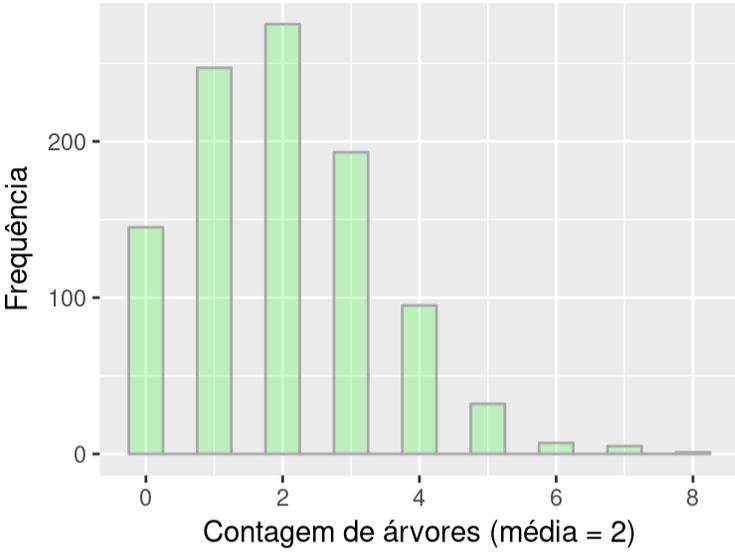
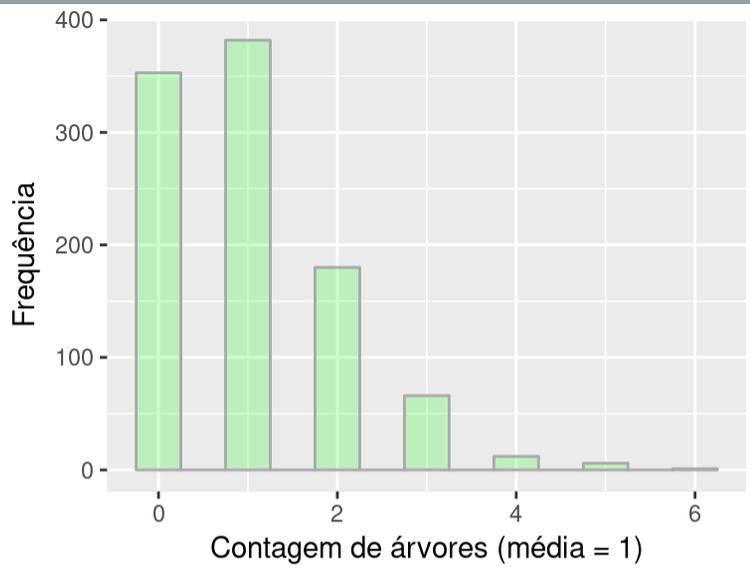
- transforma o valor observado (predito mais o resíduo)

Função de Ligação

$$f(\hat{Y})$$

- tranforma apenas a expepança da variável resposta
- preserva a estrutura do resíduo
- resíduo pode ser diferente da Gaussiana





# Dados de Contagem

- contagem é limitada no zero
- variância não é constante (aumenta com a média)
- erro não tem distribuição normal
- valores inteiros (não contínuos)

Função de ligação

$$\log(\alpha + \sum \beta_i x_i)$$

# GLM: modelo de contagem

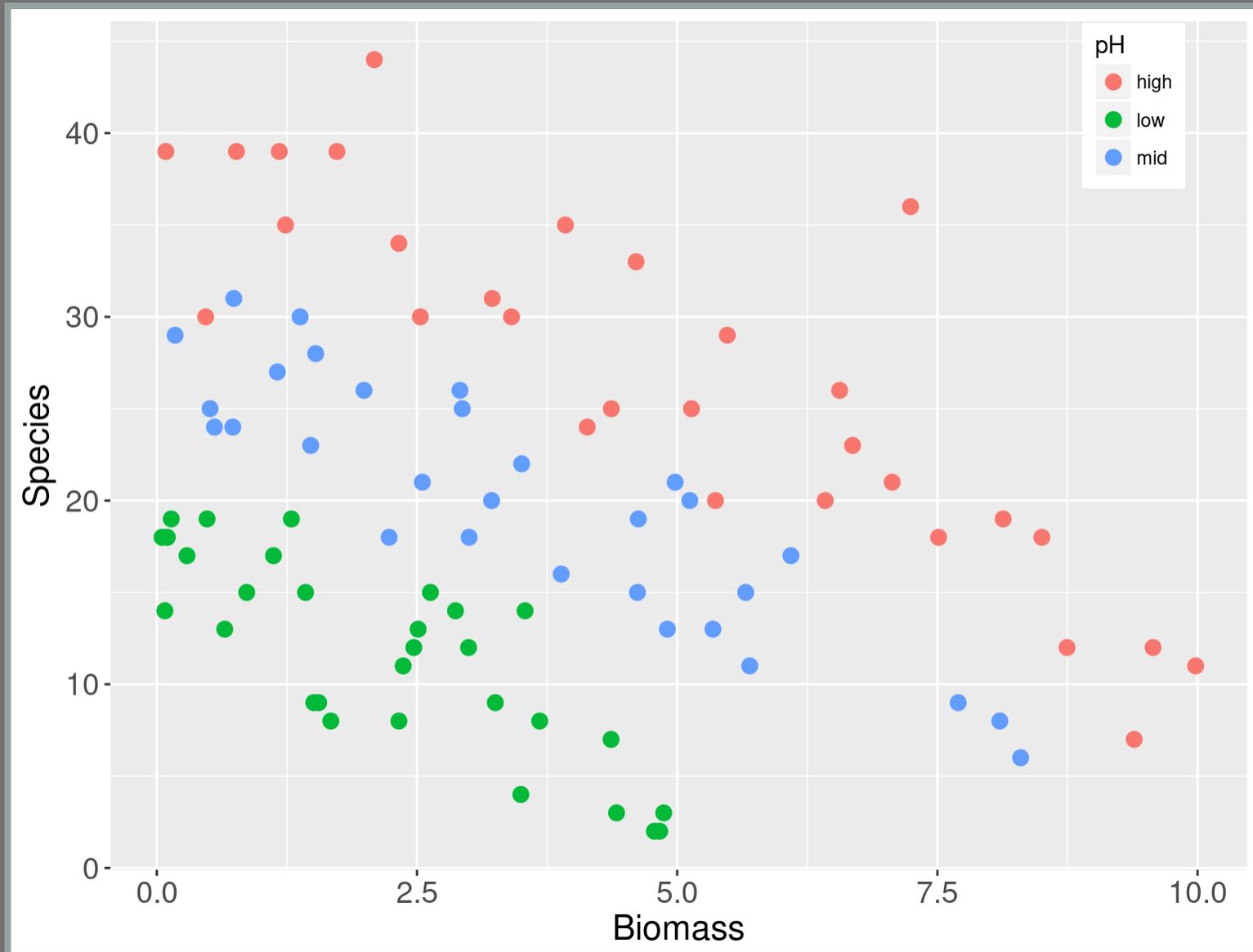
- faça o modelo cheio usando a família de ligação `poisson(log)`
- avalie o sobre-dispersão do erro pela razão: `Residual deviance/degrees of freedom`
- se o valor da razão for muito maior que 1, ajuste o modelo cheio novamente com a família `quasipoisson`
- se a sobredispersão persistir uma alternativa é modelar o resíduo com a binomial negativa (um parâmetro a mais relacionado à agregação)
- compare os modelos simplificados com o mais complexo usando `anova`
  - com `poisson` use o argumento `test = "Chisq"`
  - com `quasipoisson` use o argumento `test = "F"`
- retenha o modelo mínimo adequado
- retorne os coeficientes e preditos do modelo para escala original (antilog)

# Exemplo: Contagem de espécies

- Biomassa e ph do solo estão relacionados à coexistência de espécies

	pH	Biomass	Species
1	high	0.4692972	30
2	high	1.7308704	39
3	high	2.0897785	44
31	mid	0.1757627	29
32	mid	1.3767783	30
33	mid	2.5510426	21
61	low	0.1008479	18
62	low	0.1385961	19
63	low	0.8635151	15

# Gráfico: riqueza de espécies



# Modelo

```
glm01 <- glm(Species ~ Biomass + pH + Biomass:pH,  
  family = poisson, data= arv)  
anova(glm01, test = "Chisq")
```

```
## Analysis of Deviance Table  
##  
## Model: poisson, link: log  
##  
## Response: Species  
##  
## Terms added sequentially (first to last)  
##  
##  
##           Df Deviance Resid.  Df Resid. Dev  Pr  
## NULL                89      452.35  
## Biomass             1    44.673    88      407.67 2.3  
## pH                  2   308.431    86       99.24 < 2  
## Biomass:pH         2    16.040    84       83.20 0.0  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.0
```

# *summary(glm01)*

```
##
## Call:
## glm(formula = Species ~ Biomass + pH + Biomass:
##      data = arv)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4978  -0.7485  -0.0402   0.5575   3.2297
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|
## (Intercept)    3.76812    0.06153   61.240 < 2e
## Biomass        -0.10713    0.01249  -8.577 < 2e
## pHlow         -0.81557    0.10284  -7.931 2.18e
## pHmid         -0.33146    0.09217  -3.596 0.000
## Biomass:pHlow -0.15503    0.04003  -3.873 0.000
## Biomass:pHmid -0.03189    0.02308  -1.382 0.166
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.0
##
## (Dispersion parameter for poisson family taken
##
```

# *summary(glm01)*

```
# (Dispersion parameter for poisson family taken  
to be 1)  
# Null deviance: 452.346 on 89 degrees of  
freedom  
# Residual deviance: 83.201 on 84 degrees of  
freedom  
# AIC: 514.39
```

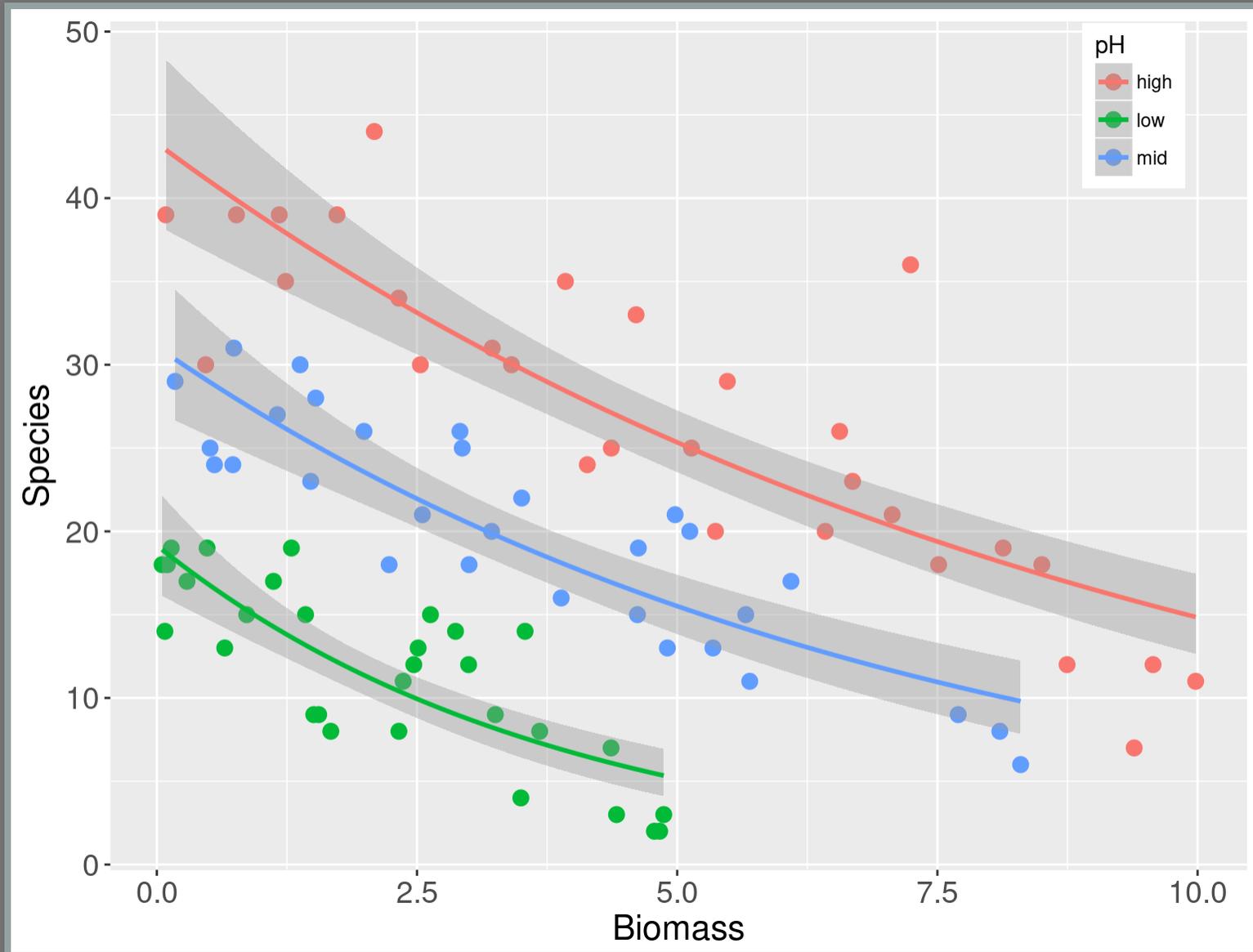
# Simplificando o Modelo

ligação da família Poisson: Qui-quadrado

```
glm02 <- glm(Species ~ Biomass + pH, family =  
  poisson, data= arv)  
anova(glm01, glm02, test = "Chisq")
```

```
## Analysis of Deviance Table  
##  
## Model 1: Species ~ Biomass + pH + Biomass:pH  
## Model 2: Species ~ Biomass + pH  
##   Resid. Df Resid. Dev Df Deviance   Pr(>Chi)  
## 1         84      83.201  
## 2         86      99.242 -2    -16.04 0.0003288 **  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.0
```

# Gráfico do nosso modelo



# Predito pelo modelo: preditor linear

- alto pH, biomassa 0.47 e 0.75

```
## (Intercept)      Biomass  
## 3.7681236    -0.1071298
```

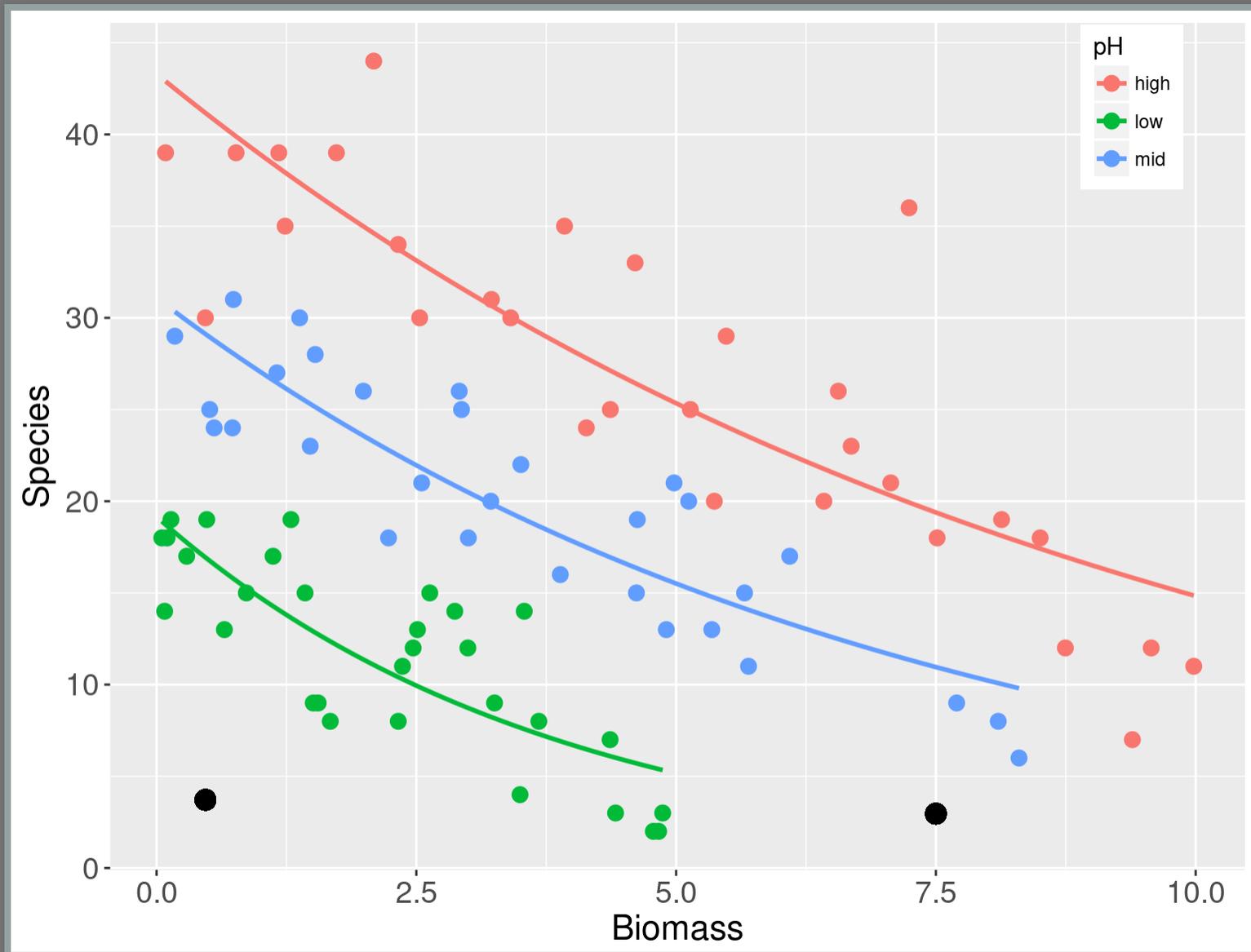
$$3.77 - 0.107 * 0.47$$

```
## [1] 3.717773
```

$$3.77 - 0.107 * 0.75$$

```
## [1] 2.96465
```





# Predito pelo modelo: preditor linear

- alto pH, biomassa 0.47 e 7.5
- antilog:  $\exp()$

$$\exp(3.717)$$

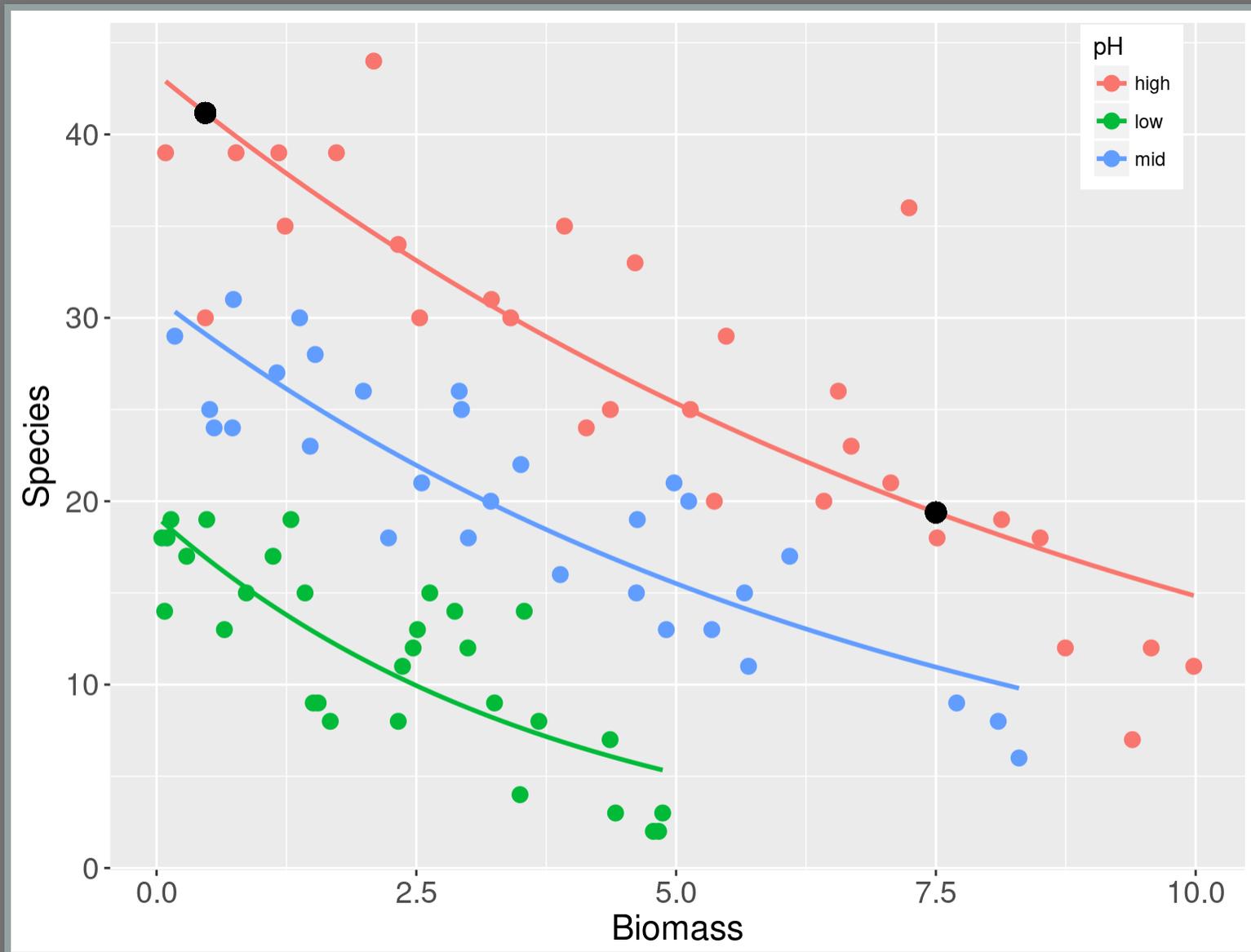
```
## [1] 41.17258
```

$$\exp(2.964)$$

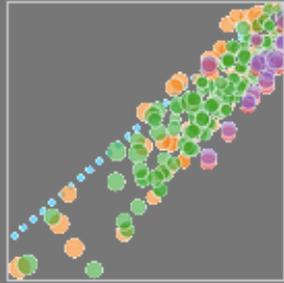
```
## [1] 19.38792
```

Predito: função inversa

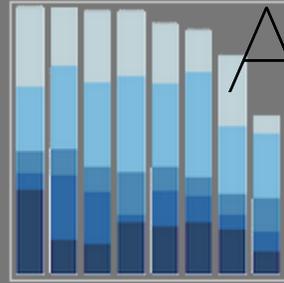
$$\exp(E_{(y)})$$



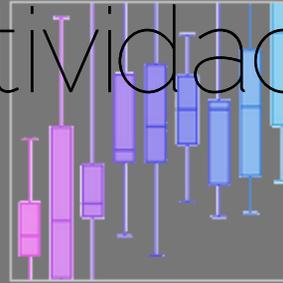
Line and Scatter Plots



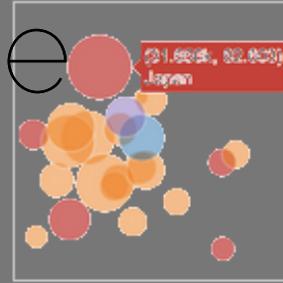
Bar Charts



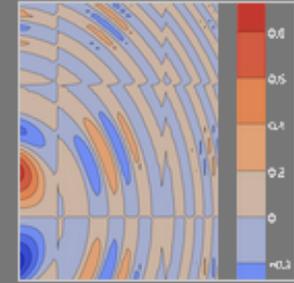
Box Plots



Bubble Charts

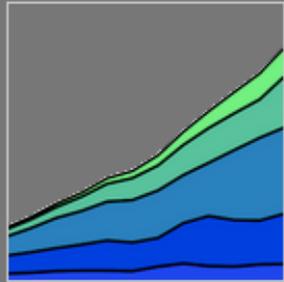


Contour Plots

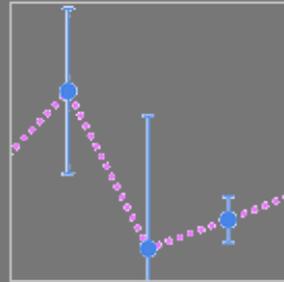


Atividade

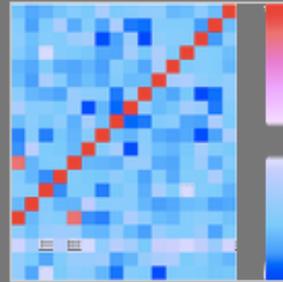
Filled Area Plots



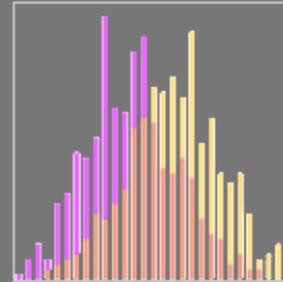
Error Bars



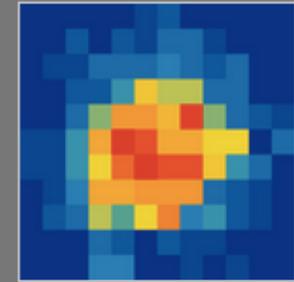
Heatmaps



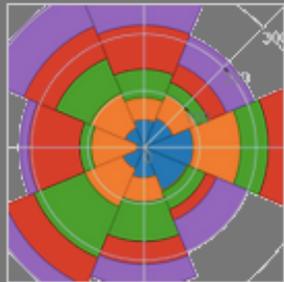
Histograms



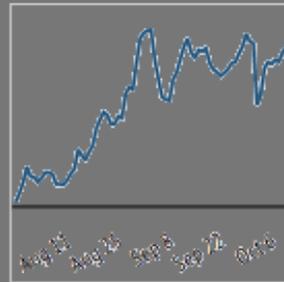
2D Histograms



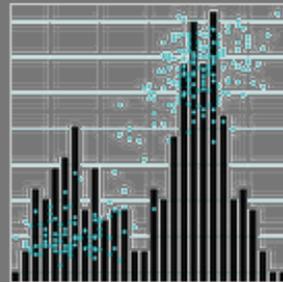
Polar Charts



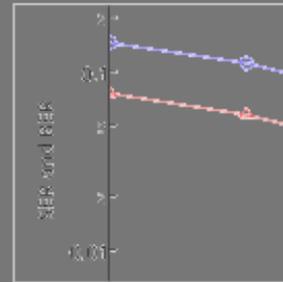
Time Series



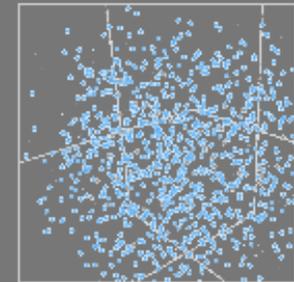
Multiple Chart Types



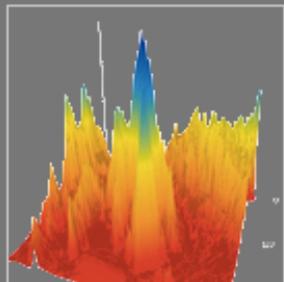
Log Plots



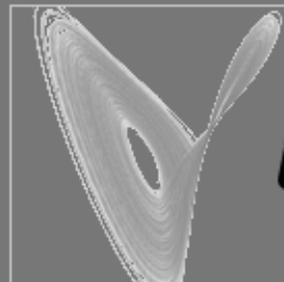
3D Scatter Plots



3D Surface Plots



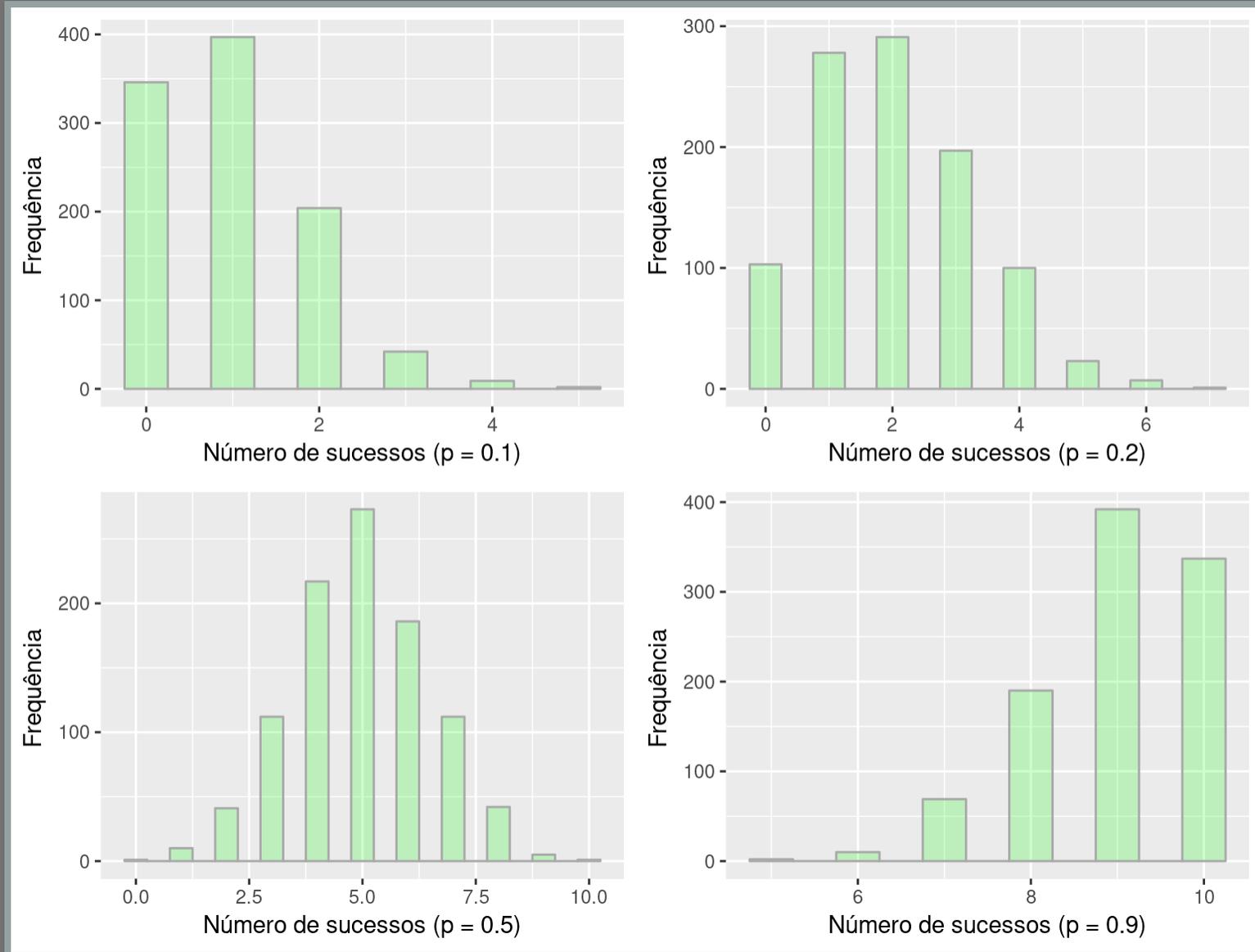
3D Line Plots



PIAnEco

# GLM Binomial

# GLM Binomial: estrutura do erro



# GLM Binomial: logit

Log da Razão de Chance (log(Odds Ratio))

$$\eta = \log\left(\frac{E(Y)}{1 - E(Y)}\right)$$

$$\eta = \log\left(\frac{a + bx}{1 - a + bx}\right)$$

$$\eta = \frac{e^{a+bx}}{1 - e^{a+bx}}$$

$$x \rightarrow \infty; y_p \rightarrow 1$$

$$x \rightarrow -\infty; y_p \rightarrow 0$$

# GLM Binomial: odds ratio

Razão de Chance

- probabilidade : sucessos/tentativas
- chance: sucessos/falhas

$n$  = tentativas

$p$  = sucessos

$q$  = falha

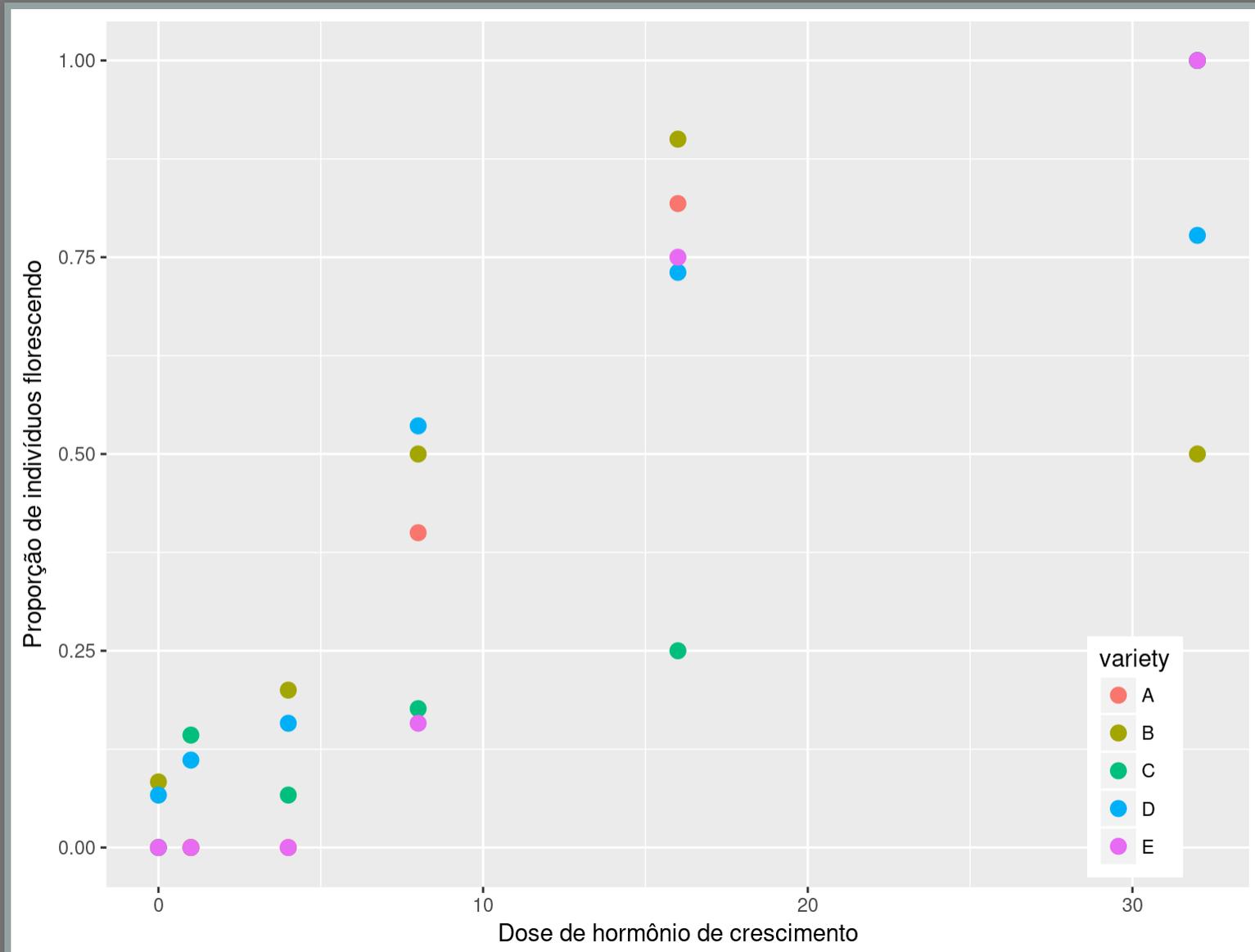
$q = n - p$

$$\log\left(\frac{p}{q}\right) = a + bx$$

# GLM binomial: florescer

flowered	number	dose	variety
0	12	1	A
0	17	4	A
0	17	1	B
3	15	4	B
2	14	1	C
1	15	4	C
2	18	1	D
3	19	4	D

# Gráfico dos dados



# Variável preditora (sucesso, falhas)

sucesso	falha
0	12
0	17
4	6
9	2
10	0
0	17
3	12
6	6
9	1
9	9
2	12
1	14
3	14
5	15
15	0
2	16
3	16

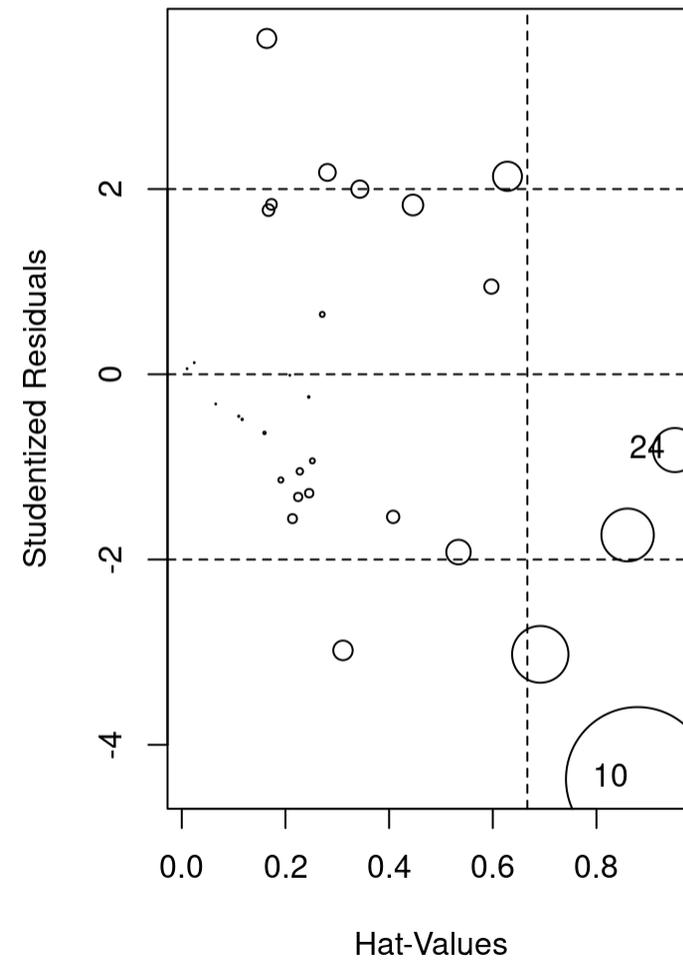
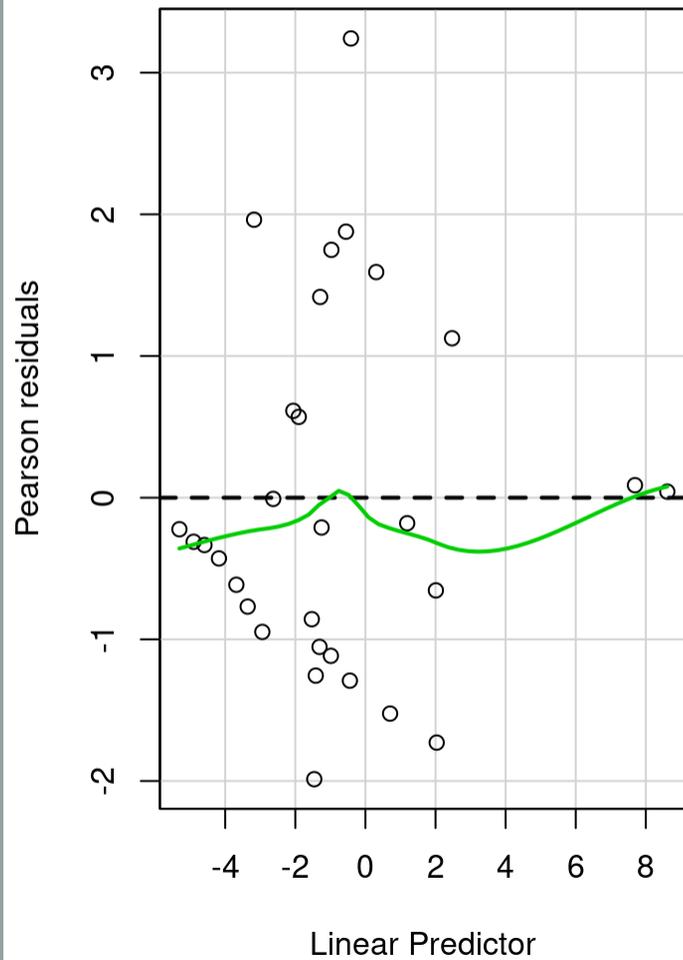
sucesso	falha
15	13
19	7
21	6
0	13
0	15
3	16
15	5
17	0
0	11
1	11
0	17
1	14
0	10

# GLM Binomial: ajustando o modelo

(sucesso, falha) ~ dose + variety + dose:variety

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: yb
##
## Terms added sequentially (first to last)
##
##
##              Df Deviance Resid.  Df Resid. Dev
## NULL                    29    303.350
## dose                    1    197.098
## variety                  4     9.483
## dose:variety            4    45.686
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.0
```

# Diagnóstico do modelo



```
##          StudRes          Hat          CookD
## 10 -4.3712682  0.8793205  14.023488
## 24 -0.8177008  0.9513937   1.309788
```

# GLM Binomial: ajustando o modelo

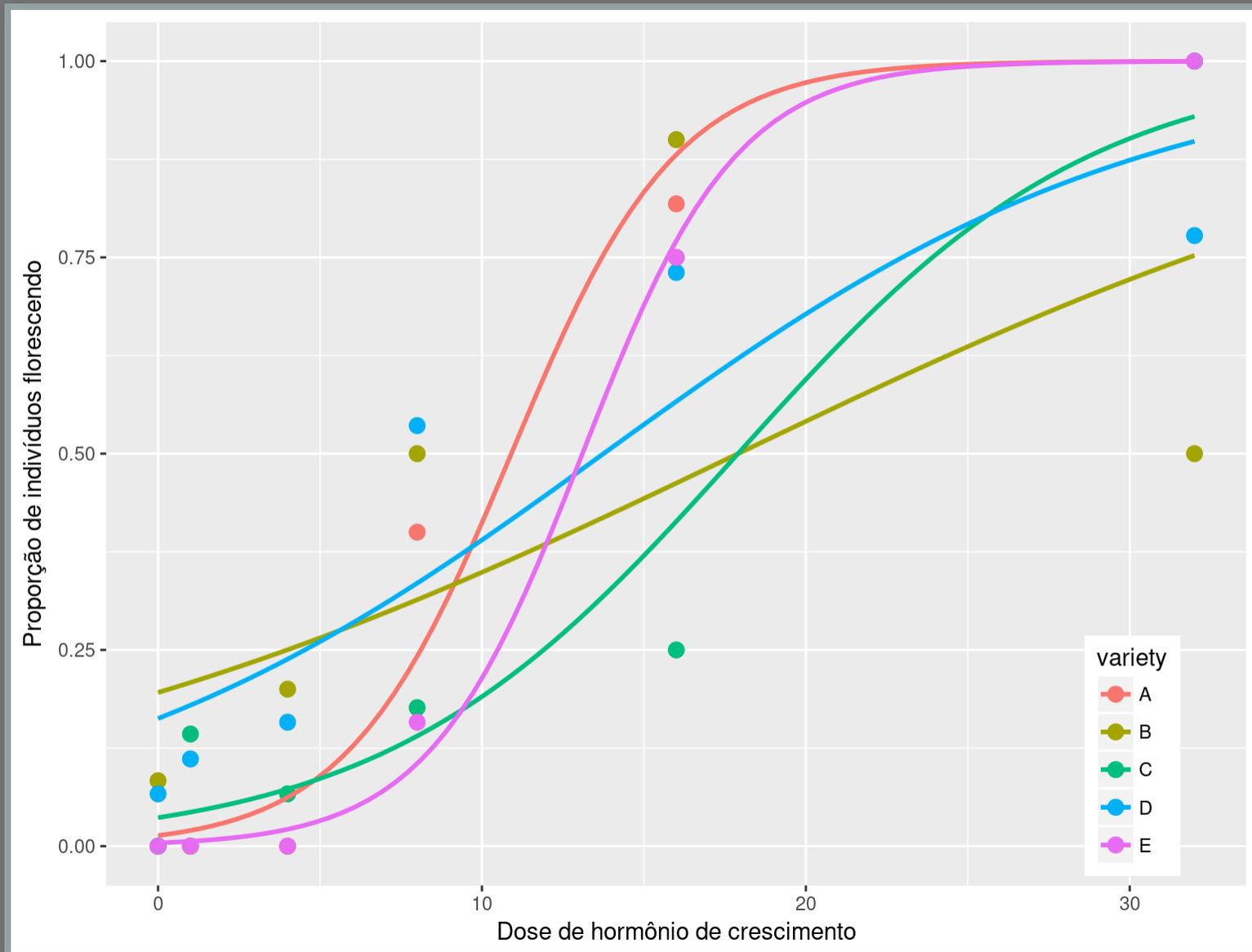
```
##
## Call:
## glm(formula = yb ~ dose + variety + dose:variety,
##      data = flor)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6648  -1.1200  -0.3769   0.5735   3.3299
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.59165    1.03215  -4.449  8.64e-05
## dose          0.41262    0.10033   4.113  3.91e-04
## varietyB     3.06197    1.09317   2.801  0.00547
## varietyC     1.23248    1.18812   1.037  0.29911
## varietyD     3.17506    1.07516   2.953  0.00311
## varietyE    -0.71466    1.54849  -0.462  0.64411
## dose:varietyB -0.34282    0.10239  -3.348  0.00077
## dose:varietyC -0.23039    0.10698  -2.154  0.03141
## dose:varietyD -0.30481    0.10257  -2.972  0.00277
## dose:varietyE -0.00649    0.13292  -0.049  0.96111
## ---
```

# GLM Binomial: simplificando o modelo

```
bin02 <- glm(yb ~ dose + variety, data=flor,  
            family=binomial)  
anova(bin01, bin02, test = "Chisq")
```

```
## Analysis of Deviance Table  
##  
## Model 1: yb ~ dose + variety + dose:variety  
## Model 2: yb ~ dose + variety  
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)  
## 1      20      51.083  
## 2      24      96.769 -4   -45.686 2.863e-09 **  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.0
```

# GLM Binomial: gráfico do modelo



# Predito pelo modelo: linear

```
predict(bin01)
```

```
##          1          2          3
## -4.1790352 -2.9411862 -1.2907208  2.0
##          7          8          9
## -1.2504982 -0.9713115 -0.4129381  0.7
##         13         14         15
## -1.9013292 -0.4434907  2.4721863 -1.3
##         19         20         21
##  0.3082377  2.0330705 -4.9001833 -3.6
##         25         26         27
##  7.6897302 -4.5916515 -1.5296848 -3.3
```

# Predito pelo modelo: antilogit

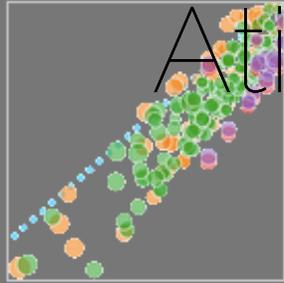
```
1 / (1 + 1 / exp(predict(bin01)))
```

```
##           1           2           3  
## 0.015082316 0.050154735 0.215730827 0  
##           7           8           9  
## 0.222613918 0.274619176 0.398207832 0  
##           13          14          15  
## 0.129958109 0.390909521 0.922168829 0  
##           19          20          21  
## 0.576455053 0.884225776 0.007390197 0  
##           25          26          27  
## 0.999542708 0.010034395 0.178039802 0
```

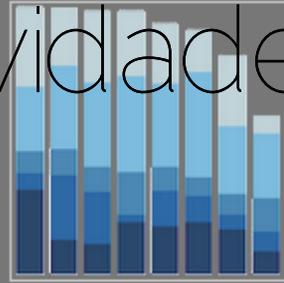
# GLM: ajuste de modelo proporção

- ajuste a variável resposta (sucesso, falha) ou proporção
- não transforme (sucesso,falha) em proporção a menos que o número de tentativas seja igual para todas as unidades amostrais
- use a família de ligação binomial(logit)
- avalie o sobre-dispersão do erro pela razão Residual deviance / degrees of freedom
- razão  $> 1$ , use família quasibinom
- busque o modelo mínimo adequado com anova
  - binomial use test = "Chisq"
  - quasipoisson use test = "F"
- retenha o modelo mínimo adequado
- retorne os coeficientes e preditos do modelo para escala original (antilogit)

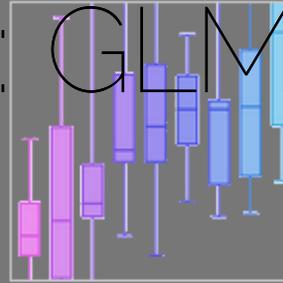
Line and Scatter Plots



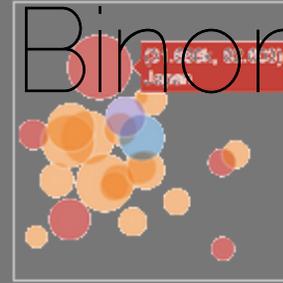
Bar Charts



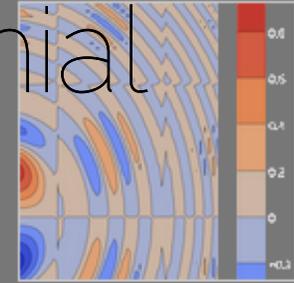
Box Plots



Bubble Charts

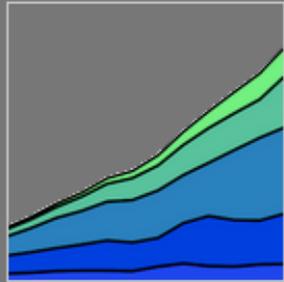


Contour Plots

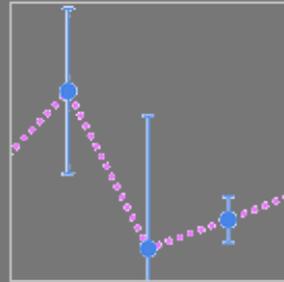


# Atividade: GLM Binomial

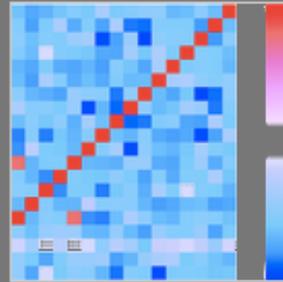
Filled Area Plots



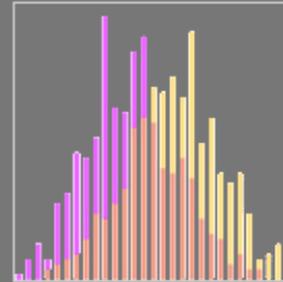
Error Bars



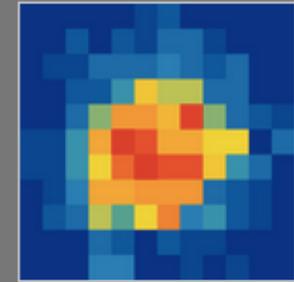
Heatmaps



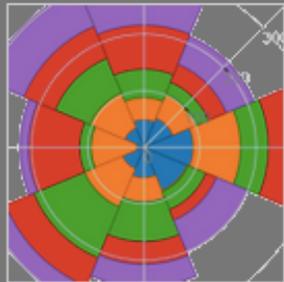
Histograms



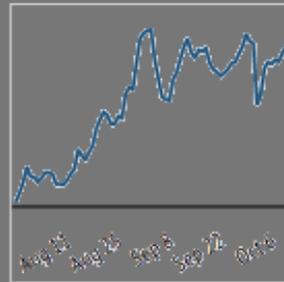
2D Histograms



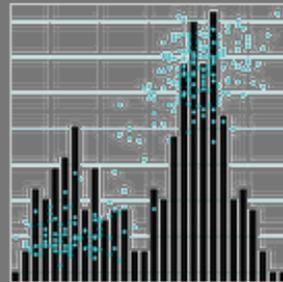
Polar Charts



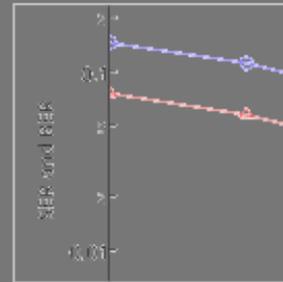
Time Series



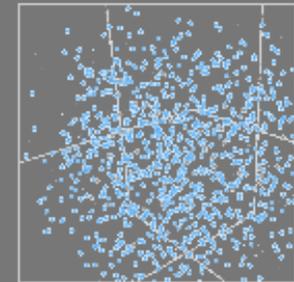
Multiple Chart Types



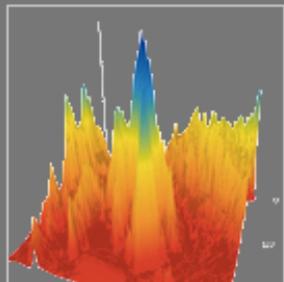
Log Plots



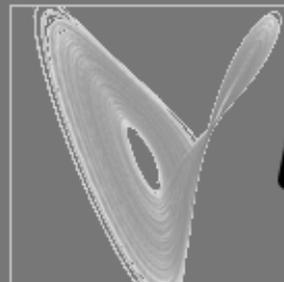
3D Scatter Plots



3D Surface Plots

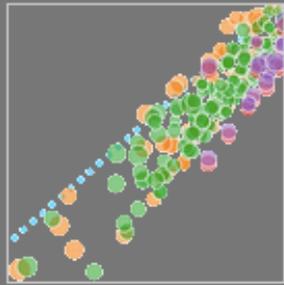


3D Line Plots

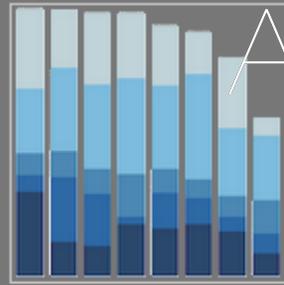


# PIAnEco

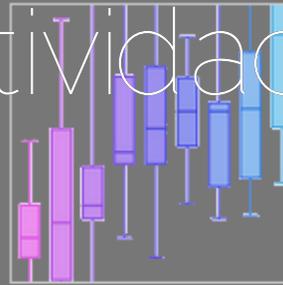
Line and Scatter Plots



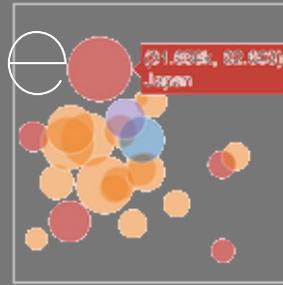
Bar Charts



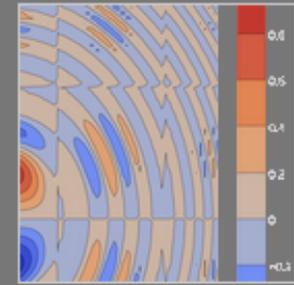
Box Plots



Bubble Charts

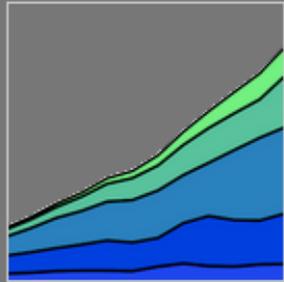


Contour Plots

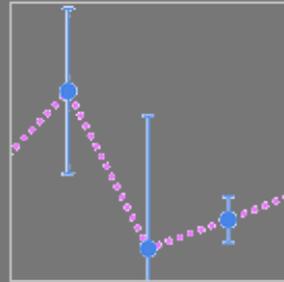


Atividade

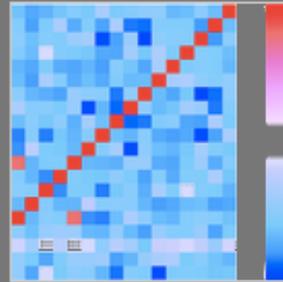
Filled Area Plots



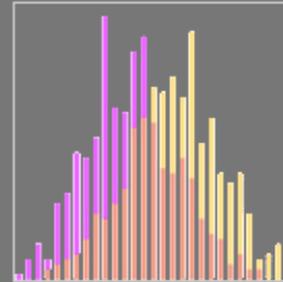
Error Bars



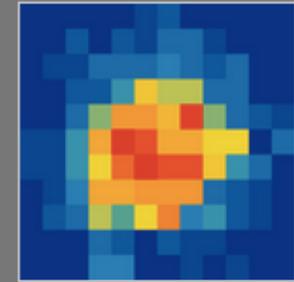
Heatmaps



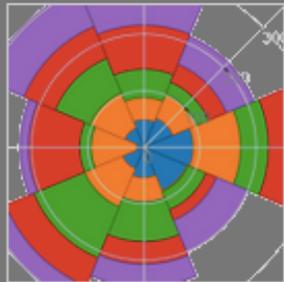
Histograms



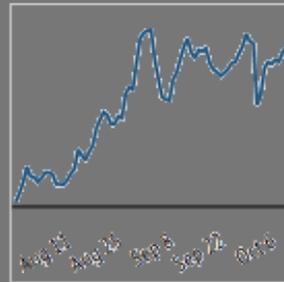
2D Histograms



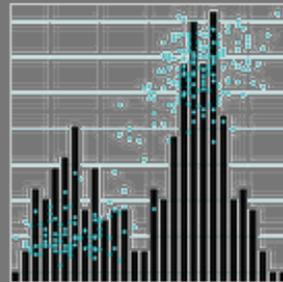
Polar Charts



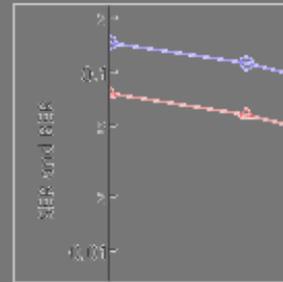
Time Series



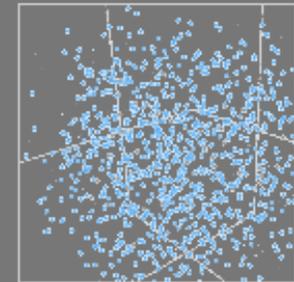
Multiple Chart Types



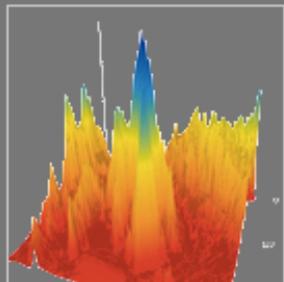
Log Plots



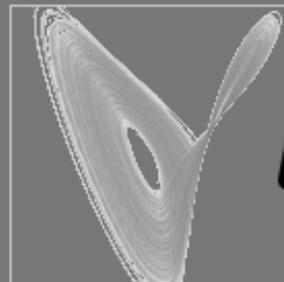
3D Scatter Plots



3D Surface Plots



3D Line Plots



PIAnEco