

labestData: Biblioteca de Dados para Aprendizado de Estatística

PET Estatística UFPR
pet.estadistica.ufpr@gmail.br



Without data, you're just another person with an opinion.

– *W. Edwards Deming*

Ciência é conhecimento que entendemos tão bem que podemos ensiná-la para um computador. Todo o resto é arte.

– *Donald Knuth (1974)*

O que é o *labestData*?

- ▶ Pacote R com conjuntos de dados.
- ▶ Mais de 450 conjuntos de dados de 18 obras nacionais.
- ▶ Documentados e prontos para serem analisados.
- ▶ Em desenvolvimento pelo **PET Estatística UFPR**.



Por que fazer um pacote com conjuntos de dados?

- ▶ Análisar de dados é uma arte.
- ▶ Praticar é fundamental para seu domínio.
- ▶ Tornar os dados mais acessíveis estimula a prática.
- ▶ Dados de livros tem documentação, contexto didático e às vezes análise.
- ▶ Para praticar a análise de um mesmo tema, é necessário consultar vários livros.
- ▶ Digitar os dados não é uma etapa estimulante.




Objetivos


1. Externo: Contribuir a comunidade por reunir, organizar, manter e disponibilizar conjuntos de dados para o ensino-apredizado de Estatística.
2. Interno: Capacitar os membros do PET Estatística a desenvolver pacotes R e trabalhar com versionamento Git em equipe.





Livros da primeira etapa do *labestData*





 BANZATTO, D. A.; KRONKA, S. D. *Experimentação Agrícola*. 4. ed. Jaboticabal, SP: Funep, 2013.


 BARBIN, D. *Planejamento e Análise Estatística de Experimentos Agronômicos*. 2. ed. Londrina, PR: Mecenias, 2013.


 BARROS, W. S.; DIAS, L. A. S. *Biometria Experimental*. Viçosa, MG: Editora UFV, 2009.


 CHARNET, R. et al. *Análise de Modelos de Regressão Linear com Aplicações*. 2. ed. Campinas, SP: Editora Unicamp, 2008.


 COSTA, A. F. B.; EPPRECHT, E. K.; CARPINETTI, L. *Controle Estatístico de Qualidade*. 2. ed. São Paulo, SP: Atlas, 2010.


 COSTA, J. R. *Técnicas Experimentais aplicadas às Ciências Agrárias*. 2003. Seropédica, RJ: Embrapa Agrobiologia (Documentos 163). Disponível em: http://www.ica.ufmg.br/mestrado_doutorado/images/stories/arquivos_mestrado/experimentao%20agrícola.pdf.


 DEMÉTRIO, C. G. B.; ZOCCHI, S. S. *Modelos de Regressão*. 2011. Piracicaba, SP. ESALQ (apostila). Disponível em: <http://www.lce.esalq.usp.br/arquivos/aulas/2011/LCE5801/Apostila.pdf>.


 FARIA, J. C. *Notas de aulas expandidas: Metodologia e Estatística Experimental*. 2009. Ilhéus, BA. UESC (apostila). Disponível em: http://nbcgib.uesc.br/lec/download/faria/apostilas/CET076_12ed_1pf.pdf.






 FERREIRA, D. F. *Estatística Multivariada*. 2. ed. Lavras, MG: Editora UFLA, 2011.

 MANLY, B. F. *Métodos Estatísticos Multivariados: uma introdução*. Porto Alegre: Bookman, 2005.

 MINGOTI, S. A. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Belo Horizonte, MG: Editora UFMG, 2005.

 PAULA, G. A. *Modelos de regressão: com apoio computacional*. 2004. São Paulo, SP. IME-USP (apostila). Disponível em: https://www.ime.usp.br/~giapaula/texto_2013.pdf.

 Pimentel-Gomes, F. *Curso de Estatística Experimental*. 15. ed. Piracicaba, SP: FEALQ, 2009.

-  RAMALHO, M. A. P.; FERREIRA, D. F.; OLIVEIRA, A. C. *Experimentação em Genética e Melhoramento de Plantas*. 2. ed. Lavras, MG: Editora UFLA, 2005.
-  RAMOS, E. M. L. S.; ALMEIDA, S. S.; ARAÚJO, A. R. *Controle Estatístico da Qualidade*. 1. ed. Porto Alegre, RS: Bookman, 2013.
-  STORCK, L. et al. *Experimentação Vegetal*. 3. ed. Santa Maria, RS: Editora UFSM, 2011.
-  VIEIRA, S. *Estatística Experimental*. 2. ed. São Paulo, SP: Atlas, 1999.
-  ZIMMERMANN, F. J. *Estatística Aplicada à Pesquisa Agrícola*. 1. ed. Santo Antônio de Goiás, GO: Embrapa Arroz e Feijão, 2004.

Organização dos conjuntos de dados

Nomeação

BanzattoQd3.4.1

PaulaEg1.12.2

BarbinEx8

PimentelPg269

Documentação

Contexto e descrição de todas as variáveis.

Classificação

Keywords classificam os dados por tema.

DIC	PSS	RL
DBC	GE	RM
DQL	BAF	RP
FAT2	AnaFat	contagem
FAT3	AnaDisc	binomial
FATADI	AnaClust	sensorial
FRAC	ACP	
PS	ACC	

Exemplos

Análise descritiva dos dados.

Interface Shiny

```
library(labestData)
```

```
labestDataView()
```

labestData: Conjuntos de dados para Ensino de Estatística

PET-Estatística UFPR - Versão 0.0.16.451

Dados disponíveis **Baixe os dados:**

Escolha um dataset Download tsv



labestData: Conjuntos de dados para Ensino de Estatística

PET-Estatística UFPR - Versão 0.0.16.451

Dados disponíveis

FerreiraEx7.4.1 ▲

FerreiraEx7.4.1

FerreiraEx8.5.1

FerreiraEx9.7.2

FerreiraEx7.4.1

ManlyTb10.2

ManlyTb10.4

Avaliação de

ManlyTb1.1

ManlyTb1.2

Description

ManlyTb1.3

Dados referentes a um estudo com animais avaliados antes e depois de terem sido submetidos a uma dieta balanceada. O interesse do estudo é avaliar se houve efeito significativo da dieta nas variáveis peso e teor de proteína.

Format

Um `data.frame` com 12 observações e 3 variáveis.

`fase`

Fator que indica a qual fase do estudo a observação pertence. Antes ou depois da dieta.

`peso`

Baixe os dados:

Download tsv

Tabela de dados

R Documentation

labestData: Conjuntos de dados para Ensino de Estatística

PET-Estatística UFPR - Versão 0.0.16.451

Dados disponíveis

FerreiraEx7.4.1 ▼

Baixe os dados:

Download tsv

Documentação

Tabela de dados

	fase	peso	teor
1	antes	250	10
2	antes	300	12
3	antes	350	13
4	antes	320	15
5	antes	400	9
6	antes	320	11
7	depois	280	12
8	depois	320	16
9	depois	360	13
10	depois	380	18
11	depois	410	15
12	depois	350	12

Como instalar o *labestData*?

```
library(devtools)

# Do GitLab.
url <- "https://gitlab.c3sl.ufpr.br/pet-estatistica/labestData.git"
install_git(url = url, branch = "master")

# Do GitHub.
install_github(repo = "pet-estatistica/labestData",
               branch = "master")
```

Como citar

```
citation("labestData")
```

```
##  
## Para citar o pacote 'labestData' use:  
##  
##   PET Estatística UFPR (2016). labestData: Conjuntos de  
##   Dados para Ensino de Estatística. R package version  
##   0.0-16.425.  
##  
## A BibTeX entry for LaTeX users is  
##  
##   @Manual{,  
##     title = {labestData: Conjuntos de Dados para Ensino de Estatística},  
##     author = {PET Estatística UFPR},  
##     year = {2016},  
##     note = {R package version 0.0-16.425},  
##     url = {http://gitlab.c3sl.ufpr.br/pet-estatistica/labestData, https://github.com/pet-estatistica/labestData},  
##   }
```

Equipe *labestData*

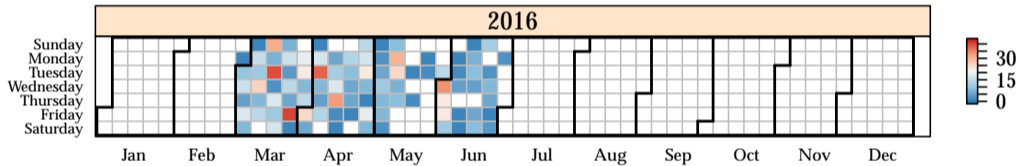
Membro	Vínculo
Alcides Conte Neto	Voluntário
Altamiro Antonio Basiewics	Petiano
Ângela Luiza Cunha Legey	Petiana
Bruna Davies Wundervald	Petiana
Bruno Geronymo	Petiano
Daniel Ikenaga	Petiano
Eduardo Elias Ribeiro Junior	Petiano *
Gabriel Sartori Klostermann	Petiano
Jhenifer Caetano Veloso	Petiana
Mônica Ludmila Hintz De Oliveira	Petiana
Paula Alessandra Zeizer Dimas	Petiana
Cesar Augusto Taconeli	Professor *
Walmes Marques Zeviani	Tutor *

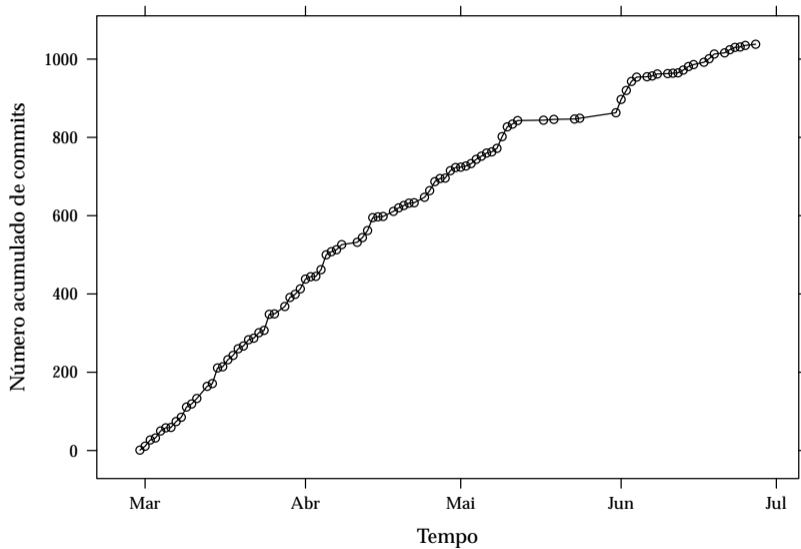
Estatísticas do *labestData*

Número total de commits: 1048

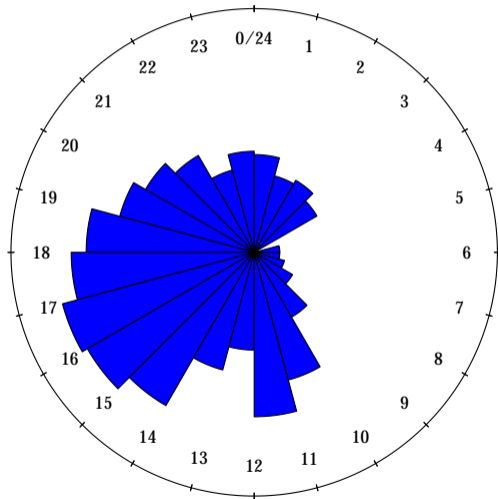
Número total de obras: 18

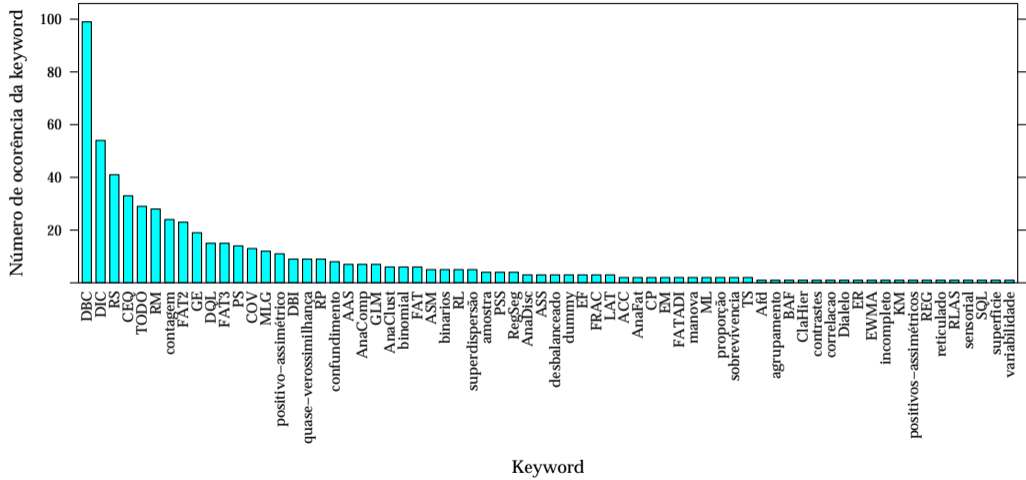
Calendar Heat Map of Values



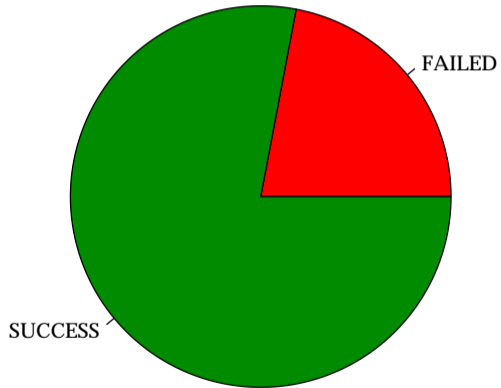


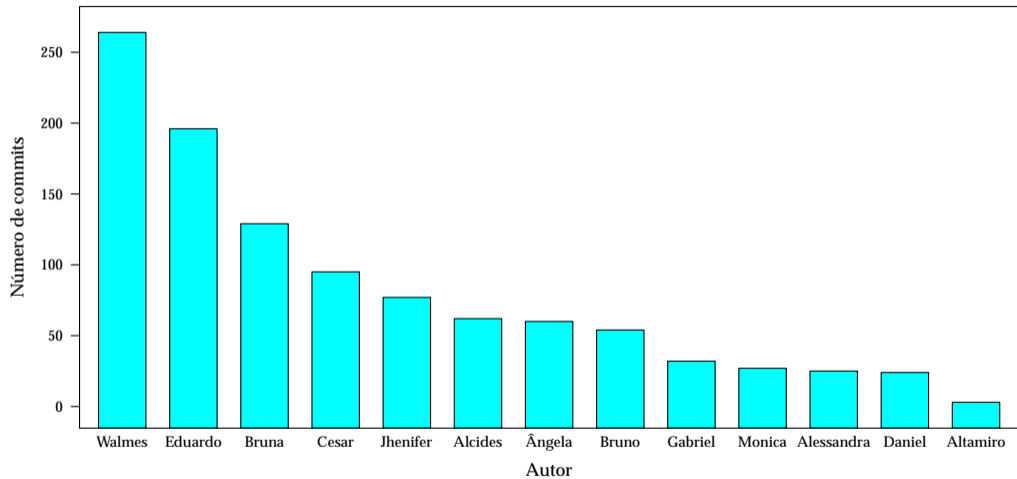
Commits por hora



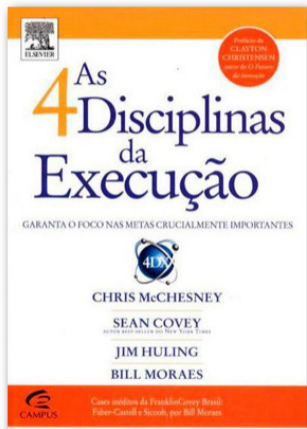


Status do build





Fluxo de trabalho



Referências

Covey, S., J. Huling, and C. Mcchesney. 2014. *As 4 Disciplinas Da Execução: Garanta O Foco Nas Metas Crucialmente Importantes*. Elsevier Brasil.

Sutherland, J. 2016. *Scrum - a Arte de Fazer O Dobro de Trabalho Na Metade Do Tempo*. LEYA BRASIL.

Princípios do desenvolvimento coletivo

Princípios

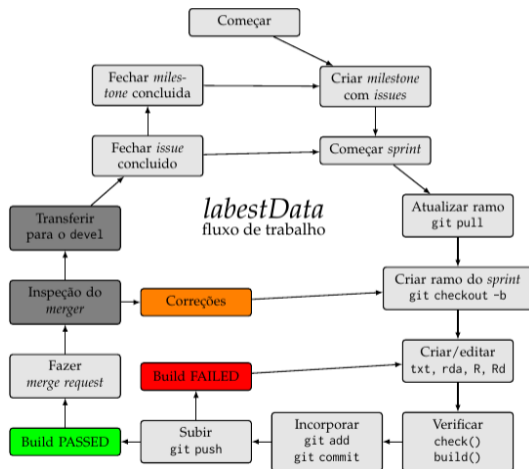
- ▶ Base semanal de desenvolvimento.
- ▶ Reuniões de *sprint* no mesmo horário toda semana.
- ▶ Autonomia dos membros na escolha das tarefas.
- ▶ Transparência e conhecimento coletivo das tarefas.
- ▶ Conhecimento do progresso e das dificuldades.
- ▶ Hierárquia horizontal ou sem hierárquia.

Benefícios

- ▶ Aumento do engajamento dos membros.
- ▶ Aumento da comunicação entre os membros.
- ▶ Aumento da padronização e qualidade do produto.
- ▶ Superação coletiva dos obstáculos.
- ▶ Diminuição dos riscos de prazos.

Ciclo de desenvolvimento

- ▶ Parte web: GitLab - C3SL
- ▶ Parte local: Git + R



No GitLab

- ▶ GitLab mantido pelo C3SL.
- ▶ Cada obra é uma *milestone*.
- ▶ Os *issues* na *milestone* são as tarefas de cada semana (*sprint*).
- ▶ Contribuições são feitas em um ramo autor-numerado: e.g. seumadruga171.
- ▶ São feitos *Merge Request* a cada final de *sprint*.



[⟨gitlab.c3sl.ufpr.br⟩](http://gitlab.c3sl.ufpr.br)

pet-estastica / labestData · Milestones

This project Search

Open Closed All + New Milestone

Milestone	Issues	Merge Requests	Completion
Bryan J.F. Manly	8	8	68%
Demetrio & Zocchi	5	5	100%
Faria	5	3	37%
Ferreira	6	6	100%
Paula	10	10	70%
Barros & Dias	6	1	14%
Charnet	11	10	66%
Epprecht	6	5	54%

Each milestone entry includes an 'Edit' button, a 'Close Milestone' button, and a 'Delete' button. The 'Barros & Dias' milestone has a note: '(Expired) expired on May 6, 2016'. The 'Charnet' milestone also has a note: '(Expired) expired on May 6, 2016'.

<https://gitlab.c3sl.ufpr.br/pet-estastica/labestData/milestones>

pet-estatística / labestData ▾ · Milestones

This project Search 🔍

Expired Milestone #7 · expired on May 6, 2016 Close Milestone Edit Delete

Pimentel-Gomes

Adicionar todas as tabelas do livro

Curso de estatística experimental / Frederico Pimentel-Gomes. -- 15. ed. -- Piracicaba: FEALQ, 2009.

Progress

10 issues: 4 open and 6 closed **80% complete** expired + New Issue Browse Issues

Issues 10 Merge Requests 10 Participants 1 Labels 1

Unstarted Issues (open and unassigned)	Ongoing Issues (open and assigned)	Completed Issues (closed)
0	4	6

Tabelas 40, 41, 42 e 43 do livro Curso de Estatística Experimental

#49 Planejamento de Experimentos 🌱

Tabelas 36, 37, 38 e 39 do livro Curso de Estatística Experimental

#48 Planejamento de Experimentos 🌱

Tabelas 32, 33, 34 e 35 do livro Curso de Estatística Experimental

#47 Planejamento de Experimentos 🌱

Tabelas 27, 28, 29, 30 e 31 do livro Curso de Estatística Experimental

Tabelas 22, 23, 24, 25 e 26 do livro Curso de Estatística Experimental

#45 Planejamento de Experimentos 🌱

Tabelas 17, 18, 19, 20 e 21 do livro Curso de Estatística Experimental

#44 Planejamento de Experimentos 🌱

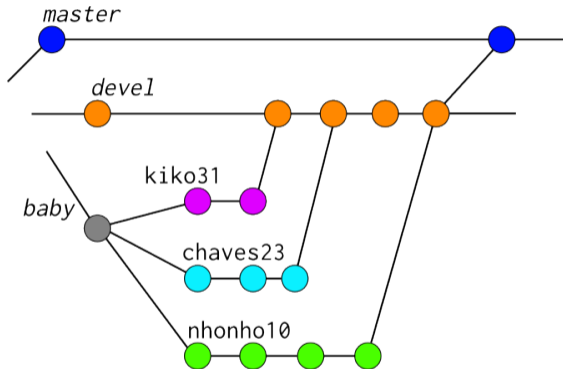
Tabelas 12, 13, 14, 15 e 16 do livro Curso de Estatística Experimental

#43 Planejamento de Experimentos 🌱

Tabelas 7, 8, 9, 10 e 11 do livro Curso de Estatística Experimental

No Git

- ▶ master: ramo estável.
- ▶ devel: ramo de desenvolvimento.
- ▶ baby: ramo de partida.
- ▶ chiquinha123: ramos autor-numerados de *sprint*.



No R com devtools e roxygen2

devtools

- ▶ Ferramentas de desenvolvimento.
- ▶ `use_data()`, `document()`,
`check_man()`, `check()`, `build()`,
`install()`, `install_git()`,
`install_github()`, `use_vignette()`,
`build_vignettes()`, etc.
- ▶ Facilita o processo de verificação e construção do pacote.

roxygen2

- ▶ Documentação nos arquivos R.
- ▶ Um *parser* transforma em Rd e escreve no NAMESPACE.
- ▶ Cada dataset gera um
 - ▶ `data-raw/*.txt`
 - ▶ `data/*.rda`
 - ▶ `R/*.R`
 - ▶ `man/*.Rd`

```

#' @name DiasEx9.6.4
#' @title Produção de Frutos de Variedades de Manga
#' @description Experimento instalado em delineamento inteiramente
#'   casualizado, onde foram estudadas 5 variedades de manga com cada
#'   parcela constituída de 3 arvores. Cada variedade teve 6
#'   repetições. Foi avaliada a produção de frutos por parcela.
#' @format Um {data.frame} com 30 observações e 3 variáveis, em que
#'
#' |describe{
#'
#' |item{\code{varied}}{Fator categórico que representa as variedades de
#'   manga.}
#'
#' |item{\code{rept}}{Inteiro que identifica as repetições.}
#'
#' |item{\code{frut}}{Total de frutos por parcela, sendo que uma parcela
#'   tem 3 plantas.}
#'
#' }
#' @keywords DIC contagem
#' @source Dias, L. A. S., Barros, W. S. (2009). Biometria
#'   Experimental. Viçosa, MG: UFV. (Exercício 4, Cap. 9, pág. 260)
#' @examples
#'
#' library(lattice)
#'
#' data(DiasEx9.6.4)
#' str(DiasEx9.6.4)
#'
#' xtabs(~varied, data = DiasEx9.6.4)
#' unstack(DiasEx9.6.4, frut ~ varied)
#'
#' xyplot(frut ~ reorder(varied, frut), data = DiasEx9.6.4,
#'   xlab = "Variedades de manga",
#'   ylab = "Número de frutos por parcela")
NULL

```

```

# Generated by roxygen2: do not edit by hand
# Please edit documentation in R/DiasEx9.6.4.R
\name{DiasEx9.6.4}
\alias{DiasEx9.6.4}
\title{Produção de Frutos de Variedades de Manga}
\format{Um {data.frame} com 30 observações e 3 variáveis, em que
}
\describe{
  \item{\code{varied}}{Fator categórico que representa as variedades de
    manga.}
  \item{\code{rept}}{Inteiro que identifica as repetições.}
  \item{\code{frut}}{Total de frutos por parcela, sendo que uma parcela
    tem 3 plantas.}
}
\source{
  Dias, L. A. S., Barros, W. S. (2009). Biometria
    Experimental. Viçosa, MG: UFV. (Exercício 4, Cap. 9, pág. 260)
}
\description{
  Experimento instalado em delineamento inteiramente
  casualizado, onde foram estudadas 5 variedades de manga com cada
  parcela constituída de 3 arvores. Cada variedade teve 6
  repetições. Foi avaliada a produção de frutos por parcela.
}
\examples{
}
library(lattice)

data(DiasEx9.6.4)
str(DiasEx9.6.4)

xtabs(~varied, data = DiasEx9.6.4)
unstack(DiasEx9.6.4, frut ~ varied)

xyplot(frut ~ reorder(varied, frut), data = DiasEx9.6.4,
  xlab = "Variedades de manga",
  ylab = "Número de frutos por parcela")
}
\keyword{DIC}
\keyword{contagem}

```


Guia de estilo de código

O jogo dos 7 erros

```
# Lê os dados.
da <- read.table("~/dados_dissertacao.R",
                 header = TRUE, sep = "\t")
str(da)

# Transforma variáveis para fator.
da <- within(da, {
  fonte <- factor(fonte)
  bloco <- factor(bloco)
})

library(lattice)

# Visualização dos dados.
xyplot(y ~ dose, groups = fonte, data = da,
       xlab = "Dose de calcário",
       ylab = "Produção de milho")

# Ajuste do modelo com interação e aditivo.
m1 <- lm(y ~ bloco + fonte * dose, data = da)
m0 <- lm(y ~ bloco + fonte + dose, data = da)

# Quadro de ANOVA.
anova(m1)
```

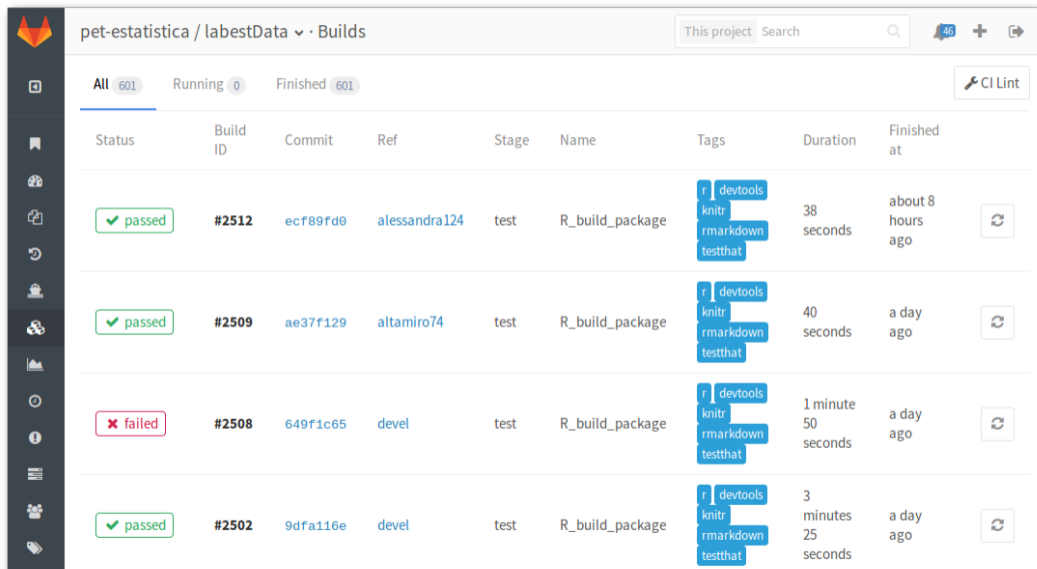
```
## Lê os dados
meus.dados=read.table("/home/walmes/dados_Dissertacao.R", h = T, sep = "\t")
str(da)
meus.dados = within(meus.dados,
{
  Fonte = factor(Fonte); Bloco = factor(Bloco)
})
require(lattice)

### Visualiza os dados
xyplot(y~dose, groups=fonte, data=da
       , xlab='Dose de calcário',
       , ylab='Produção de milho')

### Ajuste
mod.int <- lm(y~Bloco+Fonte*dose, data=meus.dados)
mod.adi <- lm(y~bloco+fonte+dose, data=meus.dados)

anova(mod.adi)
```

Integração contínua



The screenshot displays a CI/CD dashboard for the project 'pet-estastica / labestData'. The interface includes a top navigation bar with the project name, a search box, and notification icons. Below the navigation, there are filters for 'All' (601), 'Running' (0), and 'Finished' (601) builds. A 'CI Lint' button is visible in the top right. The main area contains a table of build records with columns for Status, Build ID, Commit, Ref, Stage, Name, Tags, Duration, and Finished at. Each row also includes a refresh icon.

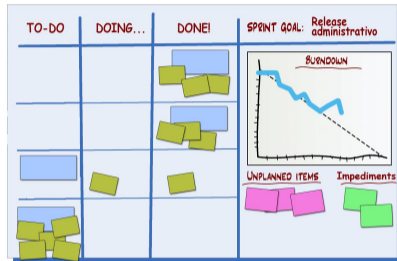
Status	Build ID	Commit	Ref	Stage	Name	Tags	Duration	Finished at
passed	#2512	ecf89fd0	alessandra124	test	R_build_package	r devtools knitr rmarkdown testthat	38 seconds	about 8 hours ago
passed	#2509	ae37f129	altamiro74	test	R_build_package	r devtools knitr rmarkdown testthat	40 seconds	a day ago
failed	#2508	649f1c65	devel	test	R_build_package	r devtools knitr rmarkdown testthat	1 minute 50 seconds	a day ago
passed	#2502	9dfa116e	devel	test	R_build_package	r devtools knitr rmarkdown testthat	3 minutes 25 seconds	a day ago

Obras para a fase 2 do *labestData*



Modificações no ciclo de trabalho

- ▶ Usar um quadro SCRUM.
- ▶ Implementar verificação de estilo de código na IC com o pacote lint.
- ▶ Padronizar a forma de fazer a citação das obras na documentação.
- ▶ Escrever vinhetas para os dados disponíveis.
- ▶ Para as obras prontas, passar para um só arquivo R as documentações.
- ▶ Meta: ter **1000** datasets até 7 de Outubro de 2016.



Considerações finais

- ▶ Usem e abusem do *labestData*.
- ▶ Para sugestões, colaboração ou relato de bugs:
[〈https://github.com/pet-estatistica/labestData〉](https://github.com/pet-estatistica/labestData).

Considerações finais

- ▶ Usem e abusem do *labestData*.
- ▶ Para sugestões, colaboração ou relato de bugs:
[〈https://github.com/pet-estatistica/labestData〉](https://github.com/pet-estatistica/labestData).

Obrigado

