

Definir JSONS

```
library(jsonlite) #Importado para carga dos arquivos JSON para o R
perfil <- fromJSON("240BiologiaAnimal/240profile.json")
public <- fromJSON("240BiologiaAnimal/240publication.json")
orient <- fromJSON("240BiologiaAnimal/240advise.json")
graph1 <- fromJSON("240BiologiaAnimal/240graph.json")
df.prog <- read.table("UnBPosGeral/prof_prog.csv", sep = ",",
                     colClasses = "character", encoding = "UTF-8", header = TRUE)
```

CRISP-DM Fase 2 - Entendimento dos Dados

A segunda parte do CRISP-DM consiste no entendimento dos dados. Para realizar análises significativas com os **datasets** disponíveis, é essencial ter um bom entendimento sobre a forma que estão organizados.

Os arquivos utilizados são provenientes da plataforma Elattes e compilam informações sobre a produção científica dos professores do **Programa de Pós-graduação em Biologia Animal** no período de 2010 a 2017.

Os **datasets** que serão trabalhados consistem em: perfil profissional; orientações de mestrado e doutorado realizadas; produções bibliográficas e redes de colaboração entre os pesquisadores.

Arquivos Analisados

Os arquivos com informações sobre os pesquisadores do Programa de **Biologia Animal**:

- **240BiologiaAnimal/240profile.json**: apresenta dados sobre o **perfil** de todos os pesquisadores.
- **240BiologiaAnimal/240publication.json**: apresenta dados sobre as **publicações** e **produções bibliográficas** geradas por todos os pesquisadores.
- **240BiologiaAnimal/240advise.json**: apresenta dados sobre **orientações de mestrado e doutorado** feitas por todos os pesquisadores.
- **240BiologiaAnimal/240graph.json**: apresenta dados sobre **produções bibliográficas colaborativas** feitas entre os pesquisadores.

Análise estrutural dos dados

Para continuar com as análises, as seguintes bibliotecas são selecionadas:

```
#library(tidyverse) #Importado para manipulação de tibbles
library(listviewer) #Importado para análise dos arquivos JSON
library(igraph) #Importado para manipulação de grafo
```

```
##
## Attaching package: 'igraph'

## The following objects are masked from 'package:stats':
##
##   decompose, spectrum

## The following object is masked from 'package:base':
##
```

```
##      union
library(dplyr) #Importado para uso do Operador Pipe

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:igraph':
##
##      as_data_frame, groups, union
## The following objects are masked from 'package:stats':
##
##      filter, lag
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
```

```
library(tidyr) #Importado para uso da função spread()
```

```
##
## Attaching package: 'tidyr'
## The following object is masked from 'package:igraph':
##
##      crossing
```

```
library(ggplot2) #Importado para visualizações com ggplot()
```

```
setwd("~/Repository/DataScience/DS4A-BioAni-BioMic-BioMol-PatMol")
source("elattes.ls2df.R")
```

Após a importação e definição das bibliotecas utilizadas, podemos utilizar funções de tais pacotes para análise sistêmica dos dados, mas para isto, estes devem ser primeiramente descritos por meio de funções de descrição do pacote `dplyr` que facilita a visualização e manipulação dos dados:

SEPARACAO DOS CAMPOS DE DF.PROG

```
df.prog <- df.prog %>% separate(idLattes.Docente.Categoria.Grande.Area.Area.de.Avaliacao.Codigo.AreaPos,
                               c("idLattes", "Docente", "Categoria", "GrandeArea", "AreaDeAvaliacao", "Codigo", "AreaPos"),
                               sep = ";", extra = "drop", fill = "right")
```

Número de docentes na base e a construção das listas dos mesmos:

```
length(perfil)
```

```
## [1] 19
```

```
ProfileList <- list()
for (i in 1:length(perfil)) {
  ProfileList <- rbind(ProfileList, perfil[[i]]$nome)
}
```

Análise das listas:

Número de áreas de atuação cumulativas:

```
sum(sapply(perfil, function(x) nrow(x$areas_de_atuacao)))
```

```
## [1] 78
```

Numero de areas de atuacao por pessoa

```
table(unlist(sapply(perfil, function(x) nrow(x$areas_de_atuacao))))
```

```
##
```

```
## 2 3 4 5 6
```

```
## 2 3 7 5 2
```

Numero de pessoas por grande area

```
table(unlist(sapply(perfil, function(x) (x$areas_de_atuacao$grande_area))))
```

```
##
```

```
##          CIENCIAS_AGRARIAS          CIENCIAS_BIOLOGICAS
```

```
##                      2                      68
```

```
## CIENCIAS_EXATAS_E_DA_TERRA          CIENCIAS_HUMANAS
```

```
##                      2                      5
```

```
##          ENGENHARIAS
```

```
##                      1
```

Numero de pessoas que produziram os específicos tipos de producao

```
table(unlist(sapply(perfil, function(x) names(x$producao_bibliografica))))
```

```
##
```

```
##          ARTIGO_ACEITO
```

```
##                      3
```

```
##          CAPITULO_DE_LIVRO
```

```
##                      14
```

```
## DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA
```

```
##                      1
```

```
##          EVENTO
```

```
##                      15
```

```
##          LIVRO
```

```
##                      4
```

```
##          PERIODICO
```

```
##                      19
```

```
##          TEXTO_EM_JORNAIS
```

```
##                      4
```

Numero de publicacoes por tipo

```
sum(sapply(perfil, function(x) length(x$producao_bibliografica$ARTIGO_ACEITO$ano)))
```

```
## [1] 6
```

```
sum(sapply(perfil, function(x) length(x$producao_bibliografica$CAPITULO_DE_LIVRO$ano)))
```

```
## [1] 41
```

```
sum(sapply(perfil, function(x) length(x$producao_bibliografica$LIVRO$ano)))
```

```
## [1] 5
```

```
sum(sapply(perfil, function(x) length(x$producao_bibliografica$PERIODICO$ano)))
```

```
## [1] 681
```

```

sum(sapply(perfil, function(x) length(x$producao_bibliografica$TEXTOS_JORNAIS$ano)))

## [1] 4

Numero de pessoas por quantitativo de producoes por pessoa 0 = 1; 1 = 2...

table(unlist(sapply(perfil, function(x) length(x$producao_bibliografica$ARTIGO_ACEITO$ano))))

##
## 0 2
## 16 3

table(unlist(sapply(perfil, function(x) length(x$producao_bibliografica$CAPITULO_DE_LIVRO$ano))))

##
## 0 1 2 3 4 5 9
## 5 2 7 1 2 1 1

table(unlist(sapply(perfil, function(x) length(x$producao_bibliografica$LIVRO$ano))))

##
## 0 1 2
## 15 3 1

table(unlist(sapply(perfil, function(x) length(x$producao_bibliografica$PERIODICO$ano))))

##
## 10 11 13 18 19 21 23 26 27 33 40 44 57 68 103 104
## 1 1 1 1 3 1 1 2 1 1 1 1 1 1 1 1

table(unlist(sapply(perfil, function(x) length(x$producao_bibliografica$TEXTOS_JORNAIS$ano))))

##
## 0 1
## 15 4

Numero de producoes por ano

table(unlist(sapply(perfil, function(x) (x$producao_bibliografica$ARTIGO_ACEITO$ano))))

##
## 2016 2017
## 1 5

table(unlist(sapply(perfil, function(x) (x$producao_bibliografica$CAPITULO_DE_LIVRO$ano))))

##
## 2010 2011 2012 2013 2014 2015 2016 2017
## 7 9 3 4 3 4 5 6

table(unlist(sapply(perfil, function(x) (x$producao_bibliografica$LIVRO$ano))))

##
## 2013 2014 2016
## 2 1 2

table(unlist(sapply(perfil, function(x) (x$producao_bibliografica$PERIODICO$ano))))

##
## 2010 2011 2012 2013 2014 2015 2016 2017
## 81 81 81 90 90 76 91 91

```

```
table(unlist(sapply(perfil, function(x) (x$producao_bibliografica$TEXTO_EM_JORNAIS$ano))))
```

```
##
## 2012 2013 2014 2016
##    1    1    1    1
```

Numero de pessoas que realizaram diferentes tipos de orientacoes

```
length(unlist(sapply(perfil, function(x) names(x$orientacoes_academicas))))
```

```
## [1] 102
```

Numero de pessoas por tipo de orientacao

```
table(unlist(sapply(perfil, function(x) names(x$orientacoes_academicas))))
```

```
##
##          ORIENTACAO_CONCLUIDA_DOUTORADO
##                                18
##          ORIENTACAO_CONCLUIDA_MESTRADO
##                                19
##          ORIENTACAO_CONCLUIDA_POS_DOUTORADO
##                                7
##          ORIENTACAO_EM_ANDAMENTO_DOUTORADO
##                                16
## ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA
##                                10
##          ORIENTACAO_EM_ANDAMENTO_MESTRADO
##                                14
##          OUTRAS_ORIENTACOES_CONCLUIDAS
##                                18
```

Numero de orientacoes concluidas

```
sum(sapply(perfil, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTRADO$ano)))
```

```
## [1] 136
```

```
sum(sapply(perfil, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTORADO$ano)))
```

```
## [1] 92
```

```
sum(sapply(perfil, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_POS_DOUTORADO$ano)))
```

```
## [1] 39
```

Numero de pessoas por quantitativo de orientacoes por pessoa 0 = 1; 1 = 2...

```
table(unlist(sapply(perfil, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTRADO$ano))))
```

```
##
## 1 2 5 6 7 8 9 10 11 12 13
## 1 2 3 2 2 3 1 1 2 1 1
```

```
table(unlist(sapply(perfil, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTORADO$ano))))
```

```
##
## 0 1 2 3 5 6 7 8 10 13
## 1 4 2 2 1 2 2 3 1 1
```

```

table(unlist(sapply(perfil, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_POS_DOUTORADO$ano))),
      ##
      ## 0 2 3 5 6 7 14
      ## 12 2 1 1 1 1 1

Numero de orientacoes por ano

table(unlist(sapply(perfil, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTRADO$ano)))),
      ##
      ## 2010 2011 2012 2013 2014 2015 2016 2017
      ## 17 14 21 26 20 13 19 6

table(unlist(sapply(perfil, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTORADO$ano)))),
      ##
      ## 2010 2011 2012 2013 2014 2015 2016 2017
      ## 2 15 15 4 11 15 11 19

table(unlist(sapply(perfil, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_POS_DOUTORADO$ano)))),
      ##
      ## 2010 2011 2012 2013 2014 2015 2016
      ## 3 4 9 5 9 3 6

```

Análise dos dataframes:

Arquivo Profile por Currículo: Extrai perfis dos professores:

```

perfil.df.professores <- extrai.perfis(perfil)

```

extrai producao bibliografica de todos os professores

```

perfil.df.publicacoes <- extrai.producoes(perfil) %>%
  select(tipo_producao, everything()) %>% arrange(tipo_producao)

```

extrai orientacoes

```

perfil.df.orientacoes <- extrai.orientacoes(perfil) %>%
  select(id_lattes_orientadores, natureza, ano, orientacao, everything())

```

extrai areas de atuacao

```

perfil.df.areas.de.atuacao <- extrai.areas.atuacao(perfil) %>%
  select(idLattes, everything())

```

cria arquivo com dados quantitativos para analise

```

perfil.df <- data.frame()
perfil.df <- perfil.df.professores %>%
  select(idLattes, nome, resumo_cv, senioridade) %>%
  left_join(
    perfil.df.orientacoes %>%
      select(orientacao, idLattes) %>%
      filter(!grepl("EM_ANDAMENTO", orientacao)) %>%
      group_by(idLattes) %>%
      count(orientacao) %>%

```

```

    spread(key = orientacao, value = n),
  by = "idLattes") %>%
left_join(
  perfil.df.publicacoes %>%
    select(tipo_producao, idLattes) %>%
    filter(!grepl("ARTIGO_ACEITO", tipo_producao)) %>%
    group_by(idLattes) %>%
    count(tipo_producao) %>%
    spread(key = tipo_producao, value = n),
  by = "idLattes") %>%
left_join(
  perfil.df.areas.de.atuacao %>%
    select(area, idLattes) %>%
    group_by(idLattes) %>%
    summarise(n_distinct(area)),
  by = "idLattes")

glimpse(perfil.df)

```

```

## Observations: 19
## Variables: 15
## $ idLattes          <chr> "1612292306950738", "18...
## $ nome              <chr> "Renato Caparroz", "Mon...
## $ resumo_cv          <chr> "Formado em Zootecnia p...
## $ senioridade        <chr> "9", "8", "9", "9", "9"...
## $ ORIENTACAO_CONCLUIDA_DOUTORADO <int> 6, 1, 6, 7, NA, 1, 1, 2...
## $ ORIENTACAO_CONCLUIDA_MESTRADO  <int> 8, 6, 10, 1, 9, 5, 2, 5...
## $ ORIENTACAO_CONCLUIDA_POS_DOUTORADO <int> NA, NA, NA, NA, NA, NA,...
## $ OUTRAS_ORIENTACOES_CONCLUIDAS  <int> 18, 9, 17, NA, 18, 9, 2...
## $ CAPITULO_DE_LIVRO    <int> 2, NA, 4, 9, 1, 4, 2, N...
## $ DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA <int> NA, NA, NA, NA, NA, NA,...
## $ EVENTO               <int> 19, 14, 12, NA, 11, 10,...
## $ LIVRO                 <int> NA, NA, NA, 1, NA, NA, ...
## $ PERIODICO             <int> 13, 10, 21, 68, 18, 44,...
## $ TEXTO_EM_JORNAIS      <int> NA, NA, NA, NA, NA, 1, ...
## $ `n_distinct(area)`    <int> 2, 3, 1, 1, 4, 2, 3, 1,...

```

CRISP-DM Fase 3 - Preparação dos Dados

A terceira fase do CRISP é conhecida por ser a parte de preparação dos dados. Tal fase possui como característica a execução de atividades para construir o conjunto final de dados a partir dos dados brutos iniciais. Pode-se separar essa etapa em cinco momentos que serão descritos nessa seção.

- Seleção dos dados
- Limpeza dos dados
- Construção dos dados
- Integração dos dados
- Formatação dos dados

Na etapa de seleção dos dados a entrada é o conjunto de dados bruto e nela ocorre a decisão dos dados a serem usados para análise. Os critérios incluem relevância para as metas de mineração de dados, qualidade e restrições técnicas, como limites no volume de dados ou tipos de dados. Então vem a fase da limpeza que

recebe a seleção de dados úteis efetuada anteriormente e é efetuado um aumento na qualidade dos dados para o nível exigido pelas técnicas de análise selecionadas. Aqui pode haver o uso de técnicas mais elaboradas, como a estimativa de dados ausentes por modelagem e inserção de padrões adequados.

O terceiro passo é a construção dos dados. Essa tarefa inclui operações de preparação de dados construtivos, como a produção de atributos derivados, novos registros ou valores transformados para atributos existentes. A penúltima atividade é a integração dos dados. Este é o momento no qual as informações são combinadas de vários bancos de dados, tabelas ou registros para criar novos registros ou valores. Por fim, ocorre a tarefa de formatação dos dados, que é a realização de modificações na estrutura dos dados de forma que as operações planejadas possam ser efetuadas de forma conveniente.

Para tornar a análise mais fácil de ser feita e até mesmo para possibilitar a realização de comparações ao final, os mesmos procedimentos foram realizados para os três programas de pós-graduação. Além disso, é importante ressaltar que as variáveis e estruturas montadas foram nomeados de forma mnemônica permitindo a distinção de diferentes programas e aspectos, como orientações, publicações, entre outros.

Para a limpeza dos dados temos:

Analise dos dados no formato lista

numero de Publicacoes em periódicos

```
sum(sapply(public$PERIODICO, function(x) length(x$natureza)))
```

```
## [1] 529
```

anos analisados

```
names(public$PERIODICO)
```

```
## [1] "2010" "2011" "2012" "2013" "2014" "2015" "2016" "2017"
```

20 revistas mais publicadas

```
head(sort(table(as.data.frame(unlist(
  (sapply(public$PERIODICO, function(x) unlist(x$periodico)))
)), decreasing = TRUE),20)
```

```
##
##                               Plos One
##                               18
##                               Toxicon (Oxford)
##                               14
##                               Genetics and Molecular Biology (Impresso)
##                               11
##                               Journal of Biomedical Nanotechnology
##                               8
##                               Genetics and Molecular Research
##                               7
##                               International Journal of Nanomedicine (Online)
##                               7
##                               Protein and Peptide Letters
##                               7
## Revista de la Universidad Industrial de Santander. Salud
##                               7
##                               Journal of Nanobiotechnology
##                               6
##                               Scientific Reports
```



```
## 6
## Theriogenology
## 6
## Behavioural Brain Research
## 5
## Frontiers in Behavioral Neuroscience
## 5
## Frontiers in Pharmacology
## 5
## Genética na Escola
## 5
## Journal of Nanomedicine & Nanotechnology
## 5
## Nanomedicine
## 5
## Small Ruminant Research
## 5
## Bulletin of Environmental Contamination and Toxicology
## 4
## Journal of Nanoscience and Nanotechnology (Print)
## 4
```

Analise dos dados no formato DF

```
public.periodico.df <- pub.ls2df(public, 1) #artigos
public.livros.df <- pub.ls2df(public, 2) #livros
public.eventos.df <- pub.ls2df(public, 5) #eventos
```

Publicacao por ano

```
table(public.periodico.df$ano)
```

```
##
## 2010 2011 2012 2013 2014 2015 2016 2017
## 70 68 60 70 70 53 66 72
```

20 revistas mais publicadas Mesma visao que anterior mas agora trabalhando no DataFrame

```
head(sort(table(public.periodico.df$periodico), decreasing = TRUE), 20)
```

```
##
## Plos One
## 18
## Toxicon (Oxford)
## 14
## Genetics and Molecular Biology (Impresso)
## 11
## Journal of Biomedical Nanotechnology
## 8
## Genetics and Molecular Research
## 7
## International Journal of Nanomedicine (Online)
## 7
## Protein and Peptide Letters
## 7
## Revista de la Universidad Industrial de Santander. Salud
## 7
```

```
##          Journal of Nanobiotechnology
##          6
##          Scientific Reports
##          6
##          Theriogenology
##          6
##          Behavioural Brain Research
##          5
##          Frontiers in Behavioral Neuroscience
##          5
##          Frontiers in Pharmacology
##          5
##          Genética na Escola
##          5
##          Journal of Nanomedicine & Nanotechnology
##          5
##          Nanomedicine
##          5
##          Small Ruminant Research
##          5
##          Bulletin of Environmental Contamination and Toxicology
##          4
##          Journal of Nanoscience and Nanotechnology (Print)
##          4
```

Orientacao

Analise dos dados em formato lista

Numero de Orientacoes Mestrado e Doutorado

```
sum(sapply(orient$ORIENTACAO_CONCLUIDA_DOUTORADO, function(x) length(x$natureza))) +
  sum(sapply(orient$ORIENTACAO_CONCLUIDA_MESTRADO, function(x) length(x$natureza)))
```

```
## [1] 220
```

Analise dos dados no formato DF

```
orient.posdoutorado.df <- ori.ls2df(orient, 6) #pos-Doutorado concluido
orient.doutorado.df <- ori.ls2df(orient, 7) #Doutorado concluido
orient.mestrado.df <- ori.ls2df(orient, 8) #Mestrado concluido

orient.df <- rbind(rbind(orient.posdoutorado.df, orient.doutorado.df), orient.mestrado.df)
```

Grafo

```
g <- g.ls2ig(graph1)
df <- as.data.frame(V(g)$name); colnames(df) <- "Idlattes"
df <- left_join(df, df.prog, by = c("Idlattes" = "idLattes")) #

## Warning: Column `Idlattes`/`idLattes` joining factor and character vector,
## coercing into character vector
```

Apenas para fins de análise inicial, foram retiradas as observacoes com duplicacao de pesquisadores no caso de haver professores em mais de um programa

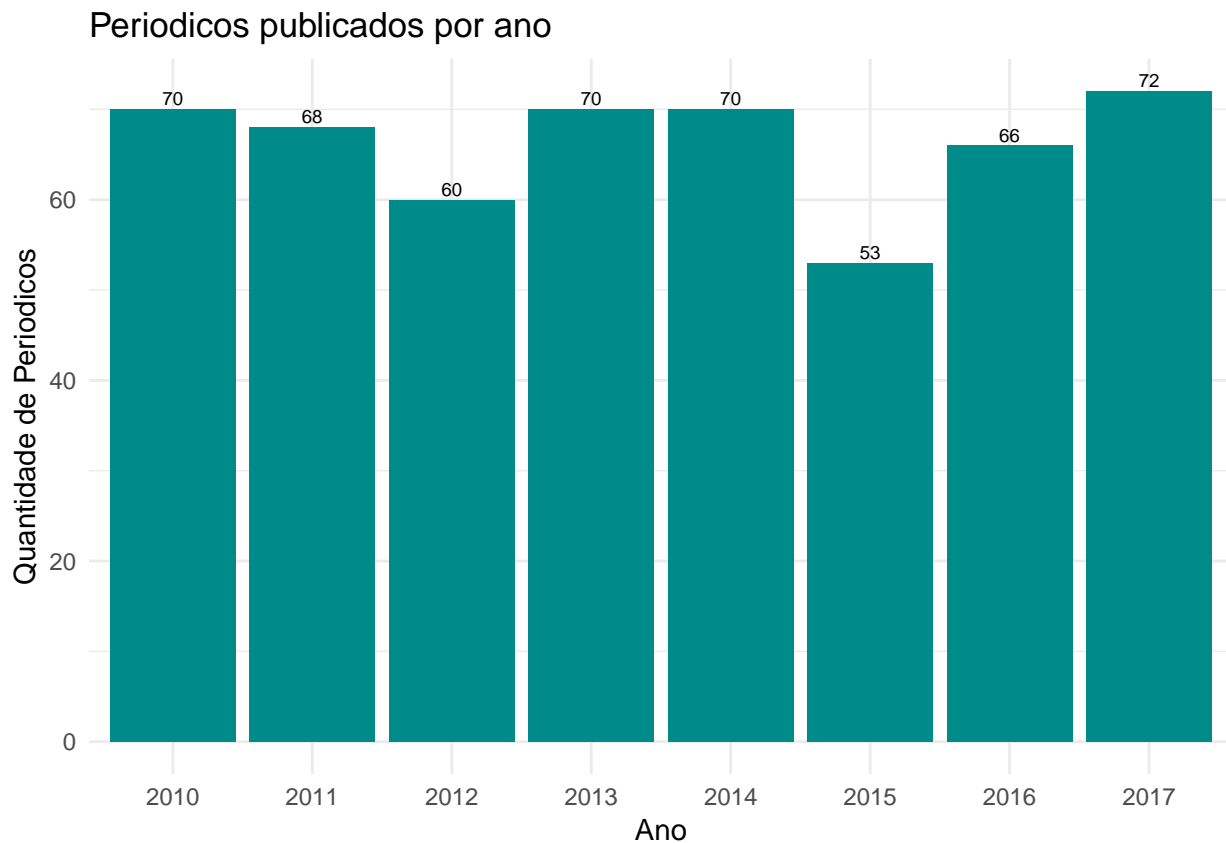
```
df <- df %>% group_by(Idlattes) %>%
  slice(1L)
V(g)$programa <- df$Programa
V(g)$orient_dout <- perfil.df$ORIENTACAO_CONCLUIDA_DOUTORADO
V(g)$orient_mest <- perfil.df$ORIENTACAO_CONCLUIDA_MESTRADO
V(g)$publicacao <- perfil.df$PERIODICO
V(g)$eventos <- perfil.df$EVENTO
```

Resultados e visualizações:

Foram escolhidos alguns resultados, em relação aos dados encontrados durante o processo, para serem plotados em gráficos.

Grafico de barras; periodicos por ano

```
public.periodico.df %>%
  group_by(ano) %>%
  summarise(Quantidade = n()) %>%
  ggplot(aes(x = ano, y = Quantidade)) +
  geom_bar(position = "stack", stat = "identity", fill = "darkcyan") +
  ggtitle("Periodicos publicados por ano") +
  geom_text(aes(label=Quantidade), vjust=-0.3, size=2.5) +
  theme_minimal() + labs(x="Ano", y="Quantidade de Periodicos")
```



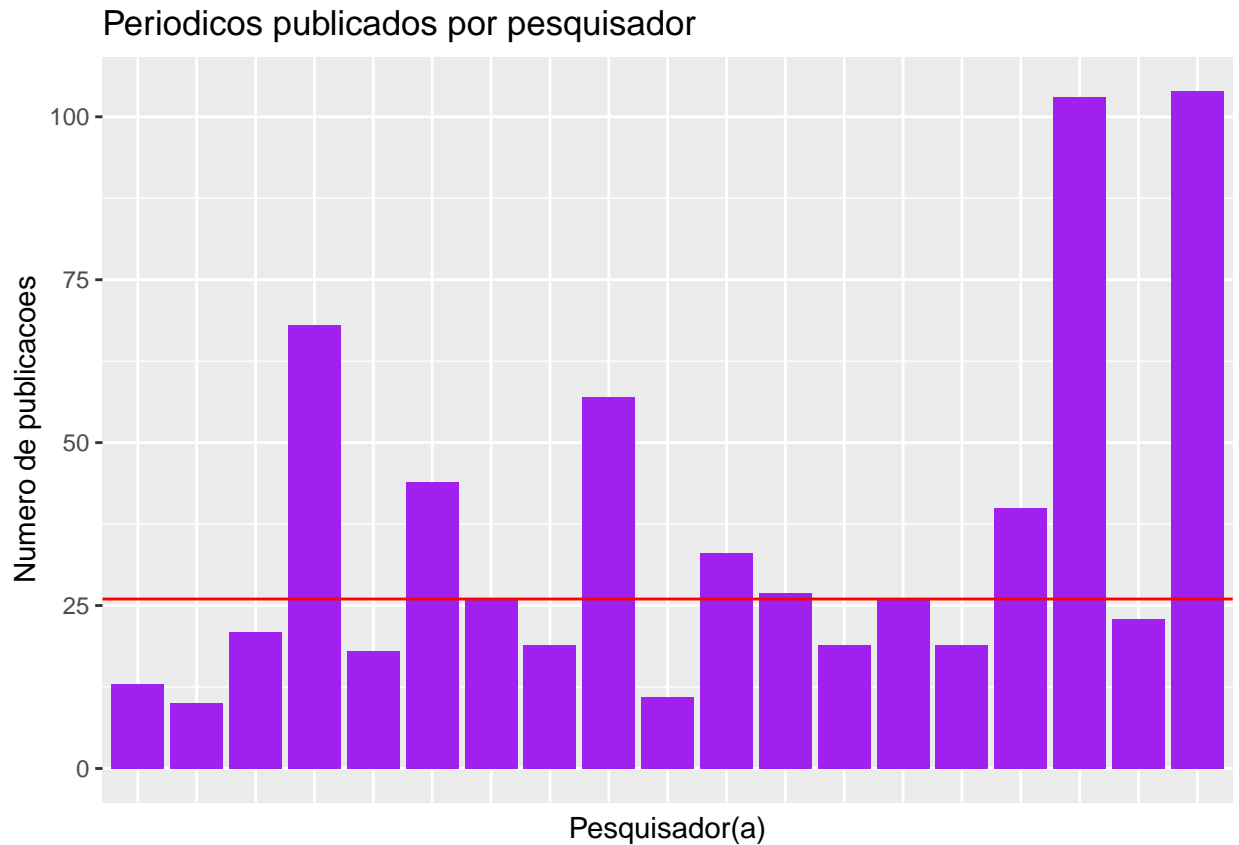
Sob uma perspectiva de publicações em periódicos, o Programa de Pós-Graduação de Biologia Animal mantém-se estável com flutuações normais, mas apresentou uma queda considerável de 2014 para 2015, diminuindo o número de publicações em cerca de 18% mas recuperando em anos posteriores.

Quantidade de periodicos publicados por professor(a) entre 2010 e 2017 - by Jonas

```

perfil.df %>%
  ggplot(aes(idLattes,PERIODICO)) +
  geom_col(fill = "purple") +
  ggtitle("Periodicos publicados por pesquisador") +
  theme(legend.position="right",legend.text=element_text(size=7)) +
  guides(fill=guide_legend(nrow=5, byrow=TRUE, title.position = "top")) +
  labs(x="Pesquisador(a)",y="Numero de publicacoes") +
  theme(axis.text.x=element_blank(), axis.ticks.x=element_blank()) +
  geom_hline(yintercept = sum(perfil.df %>% summarize(x = median(PERIODICO))), color = "red")

```



O gráfico acima demonstra o total de publicações para cada pesquisador, demonstrando relativa homogeneidade entre a quantidade de pesquisadores em média, porém há aqueles que deslocam a média para cima por estarem fora do desvio padrão acima ou abaixo, sendo que há destaque para dois pesquisadores que tem um número de publicação que corresponde cerca de quatro vezes o valor da média.

publicacao de livros por pais/ano

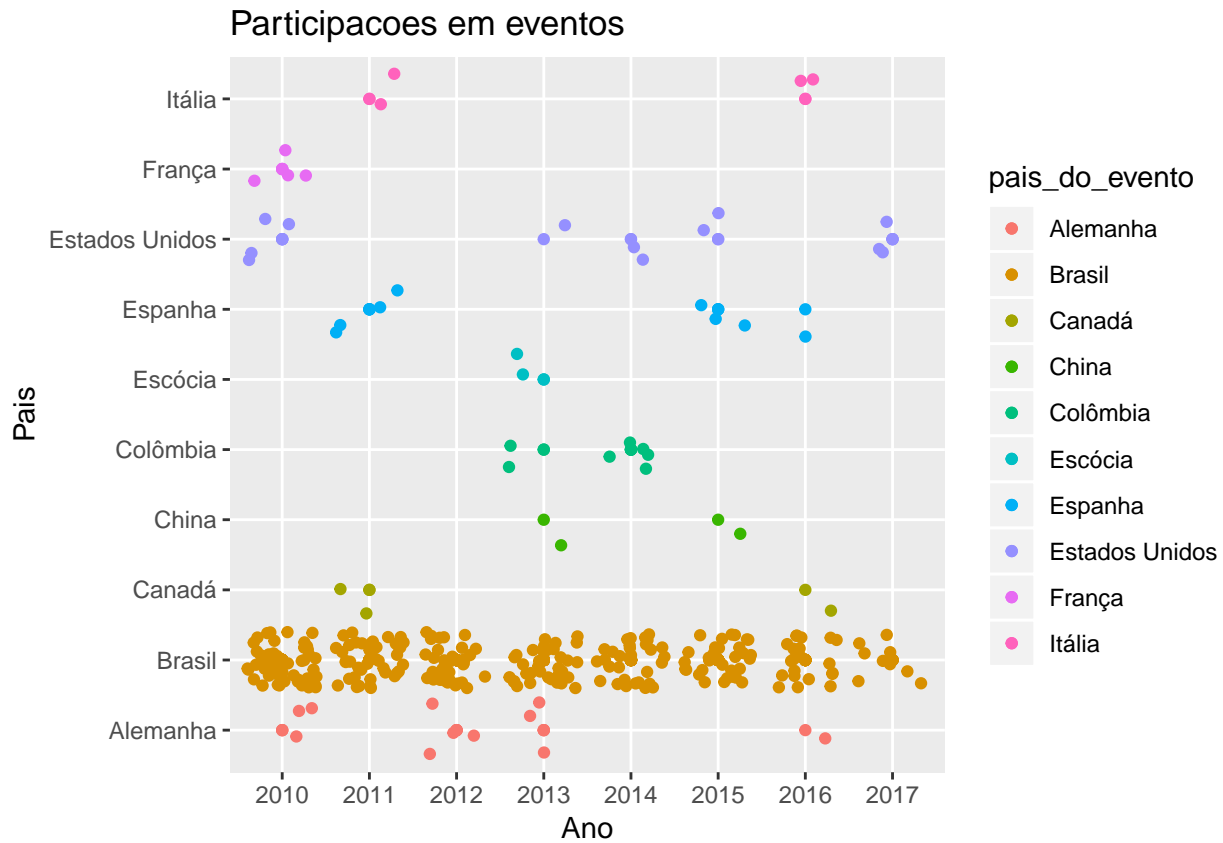
```
public.livros.df %>%
  group_by(ano,pais_de_publicacao) %>%
  ggplot(aes(x=ano,y=pais_de_publicacao, color= pais_de_publicacao)) +
  ggtitle("Livros publicados por ano") +
  xlab("Ano") + ylab("Pais") + geom_point() + geom_jitter()
```



Denota-se homogeneidade na distribuição de livros publicados para cada país de publicação, onde os estados unidos se equiparam ao Brasil quanto a número de livros publicados para o PPG em Biologia Animal

Eventos nacionais e internacionais

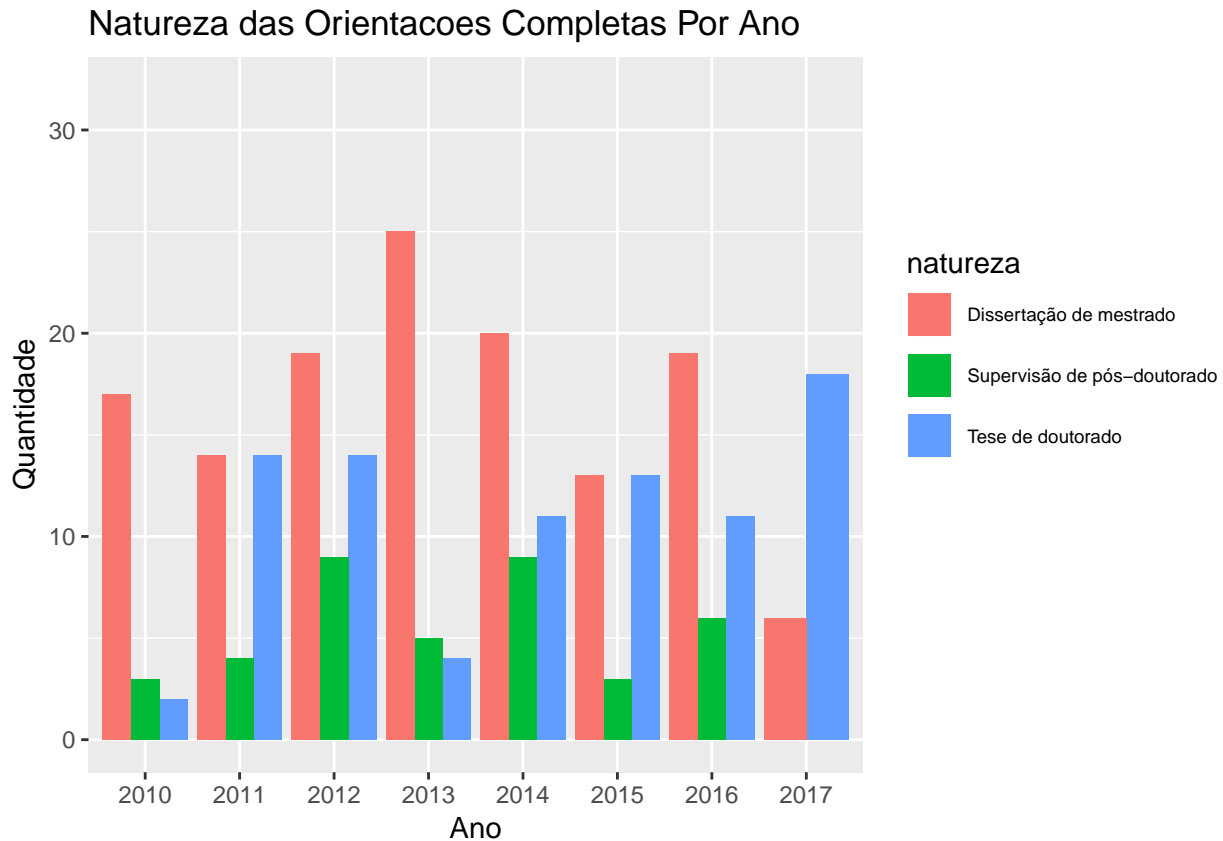
```
public.eventos.df %>%
  filter(pais_do_evento %in%
         c(names(head(sort(table(public.eventos.df$pais_do_evento)
                             , decreasing = TRUE), 10)))) %>%
  group_by(ano_do_trabalho,pais_do_evento) %>%
  ggplot(aes(x=ano_do_trabalho,y=pais_do_evento, color= pais_do_evento)) +
  ggtitle("Participacoes em eventos") +
  xlab("Ano") + ylab("Pais") + geom_point() + geom_jitter()
```



Considerando eventos, o Programa de Biologia Animal teve comparecimento maior em eventos no Brasil que em países estrangeiros. Ainda assim, o gráfico mostra que o programa esteve presente em eventos de sede internacional em todos os anos registrados na base de dados, principalmente nos Estados Unidos.

Orientacoes completas por ano e natureza

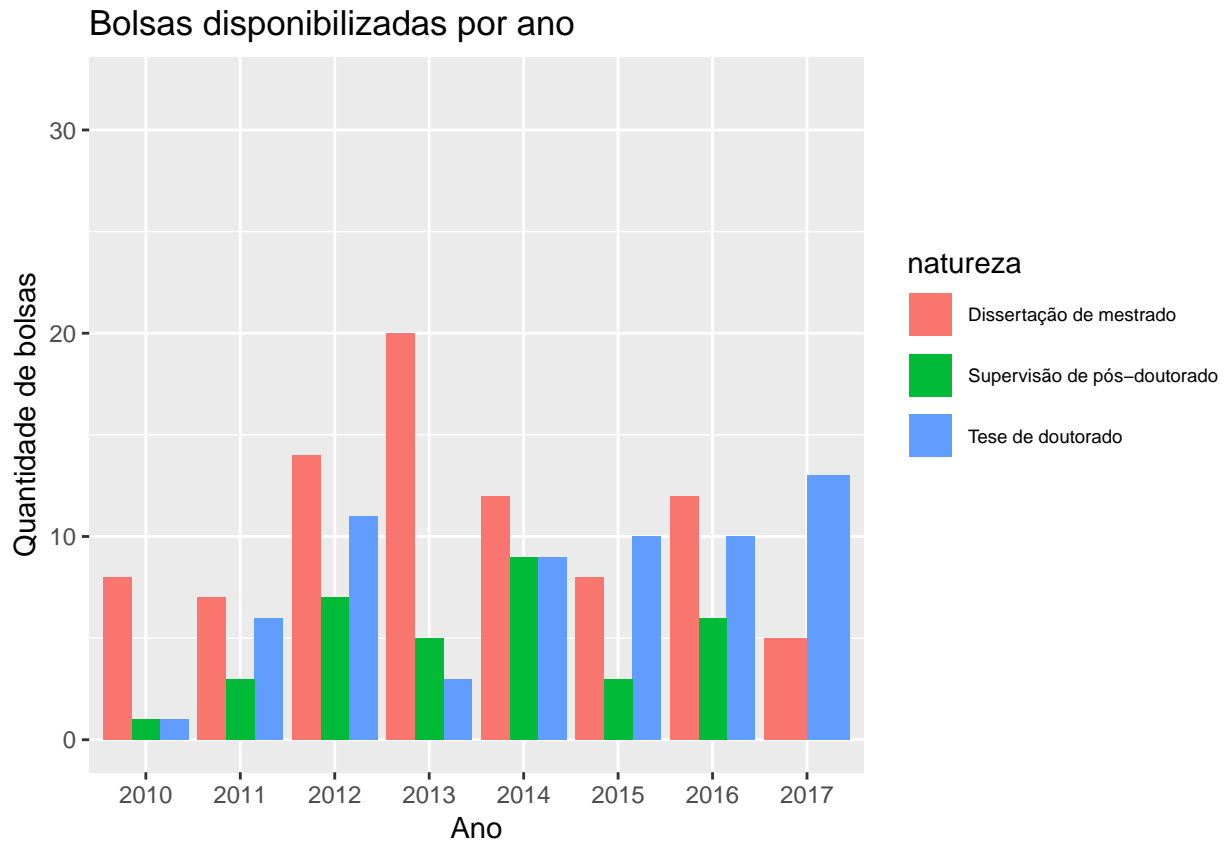
```
ggplot(orient.df, aes(ano, fill=natureza)) +
  geom_bar(stat = "count", position="dodge") +
  ggtitle("Natureza das Orientacoes Completas Por Ano") +
  theme(legend.position="right", legend.text=element_text(size=7)) +
  guides(fill=guide_legend(nrow=5, byrow=TRUE, title.position = "top")) +
  labs(x="Ano", y="Quantidade") + scale_y_continuous(limits = c(0, 32))
```



Observando a evolução do número de orientações completas ao longo dos anos, percebe-se que o Programa de Pós-Graduação cresceu consideravelmente na natureza de mestrado após 2010, com uma regressão em 2016. Entretanto, não há um comportamento linear na evolução do número de orientações finalizadas. 2014 despontou no número de dissertações de mestrado, mas este comportamento não se repetiu posteriormente. Enquanto as orientações de mestrado do Programa parecem bem estabelecidas, as supervisões de pós-doutorado ainda parecem incipientes, tendo em vista que não despontaram em nenhum dos anos pesquisados e chegaram a zero registros em 2017.

Bolsas distribuidas por ano - by Jonas

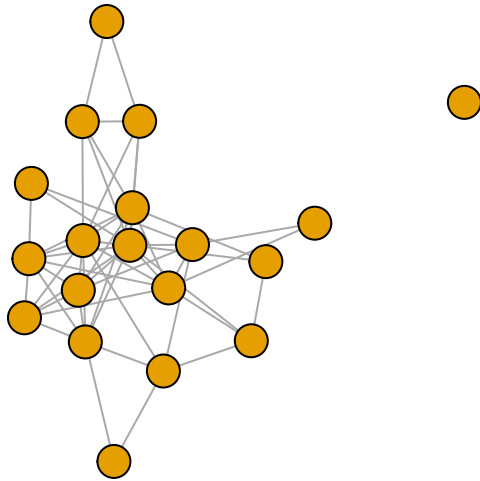
```
orient.df %>% filter(bolsa == "SIM") %>%
ggplot(aes(ano, fill=natureza)) +
  geom_bar(stat = "count", position = "dodge") +
  ggtitle("Bolsas disponibilizadas por ano") +
  theme(legend.position="right", legend.text=element_text(size=7)) +
  guides(fill=guide_legend(nrow=5, byrow=TRUE, title.position = "top")) +
  labs(x="Ano", y="Quantidade de bolsas") + scale_y_continuous(limits = c(0, 32))
```

Comparando os gráficos de orientações completas e de bolsas, é possível perceber que o número de bolsas oferecidas para o Programa acompanhou o total de orientações de maneira satisfatória ao longo dos anos. Durante todo o período, todas as naturezas apresentaram um índice de pelo menos 50% de bolsas, chegando a 100% em alguns casos. Além disso, as teses de pós-doutorado se mostram a natureza de pesquisa melhor contemplada pelas agências financiadoras, visto que apenas uma das observações não recebeu bolsa. Por fim, é possível observar que o pico no número de orientações de mestrado registrado em 2014 não foi tão expressivo em número de bolsas, caracterizando um ano com elevado número de alunos não bolsistas.

Grafo de proximidade entre pesquisadores do Programa de Pós-Graduação - by Jonas

```
plot(g, vertex.label = NA)
```

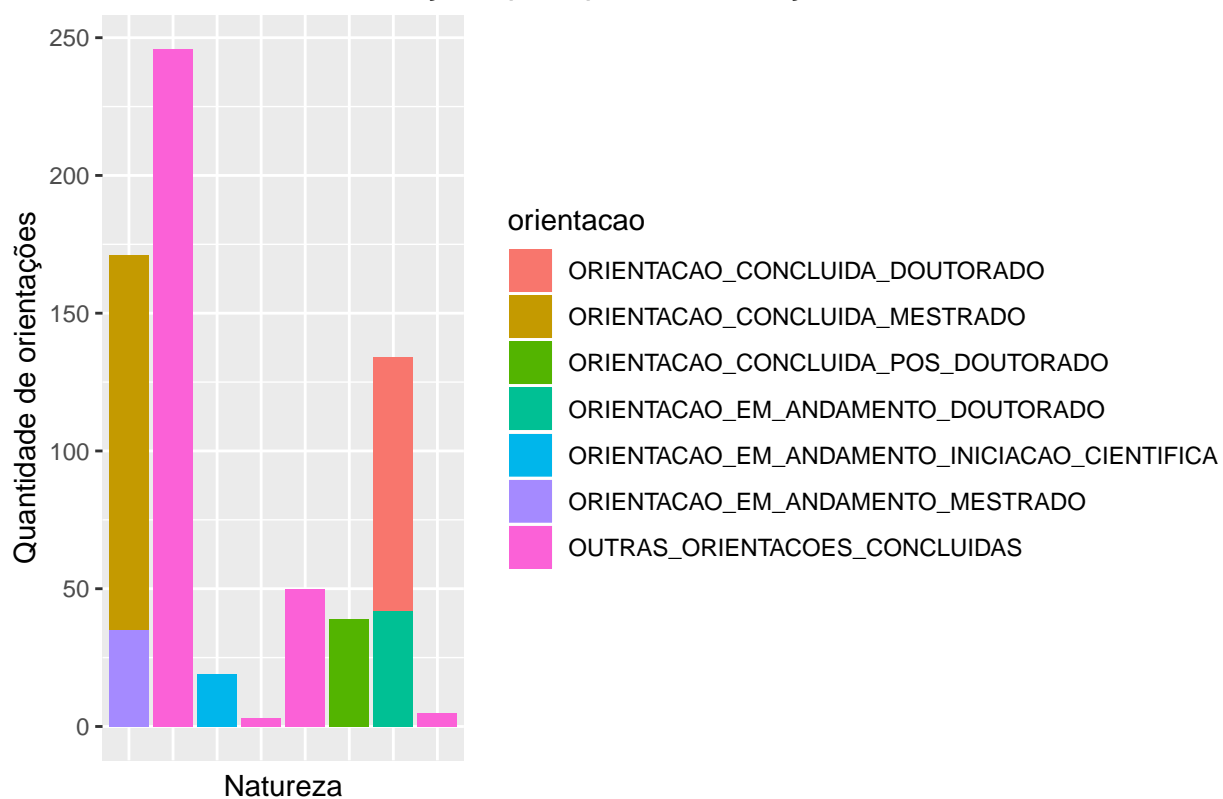


#O grafo acima representa os pesquisadores do Programa de Pós-Graduação em seus vértices #e a existência de cooperação entre eles em suas arestas. Portanto, é possível #observar que grande parte dos membros do Programa possui um grau de cooperação #pequeno, limitado a até dois outros pesquisadores.

Natureza das orientações por tipo de orientação

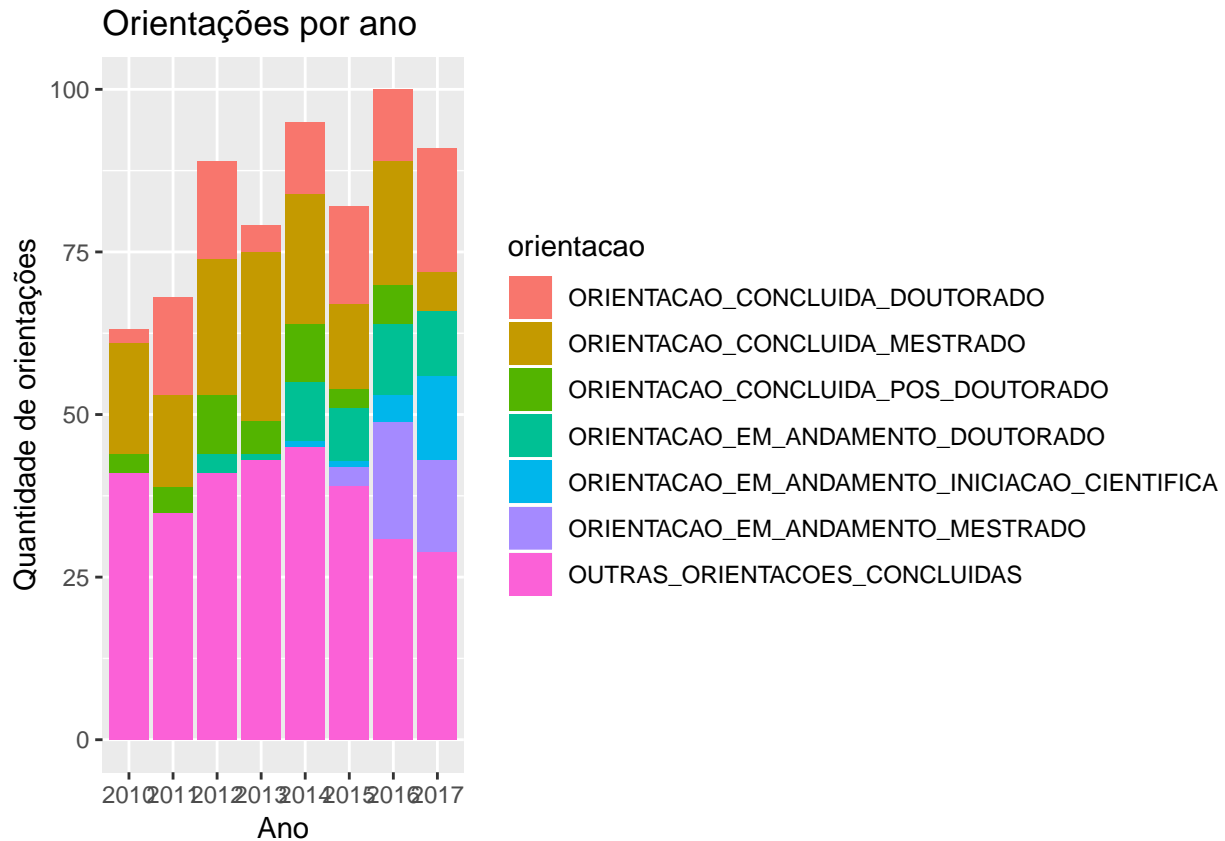
```
ggplot(perfil.df.orientacoes, aes(natureza,fill=orientacao)) +
  geom_bar(stat = 'count') +
  theme(legend.position = 'right') +
  ggtitle('Natureza das orientações por tipo de orientação') +
  labs(x='Natureza',y='Quantidade de orientações') +
  theme(axis.text.x=element_blank(),
        axis.ticks.x=element_blank())
```

Natureza das orientações por tipo de orientação



Orientações por ano:

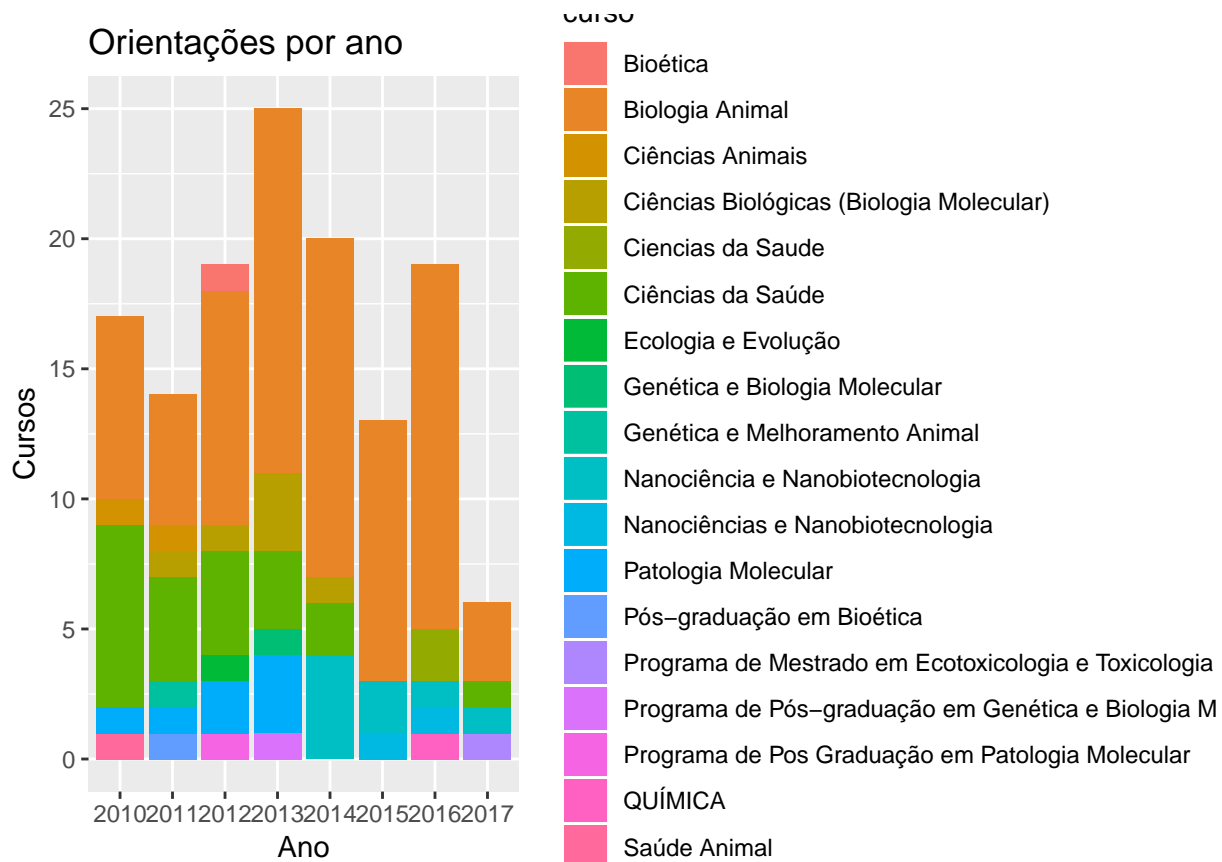
```
ggplot(perfil.df.orientacoes, aes(ano, fill=orientacao)) +
  geom_bar(stat = 'count') +
  ggtitle('Orientações por ano') +
  theme(legend.position = 'right') +
  labs(x='Ano', y='Quantidade de orientações')
```



A partir do número de orientações por ano e pela natureza destas, percebe-se que a quantidade de orientações de pós graduações no sentido restrito era pequena em proporção a outras orientações conduzidas por PPG por ano, até sua volta em 2015, onde o número de pós graduações no sentido restrito foram maiores que outras orientações.

Mestrados e cursos que mais ocorrem por ano

```
ggplot(orient.mestrado.df, aes(ano, fill=curso)) +
  geom_bar(stat = 'count') +
  ggtitle('Orientações por ano') +
  theme(legend.position = 'right') +
  labs(x='Ano', y='Cursos')
```



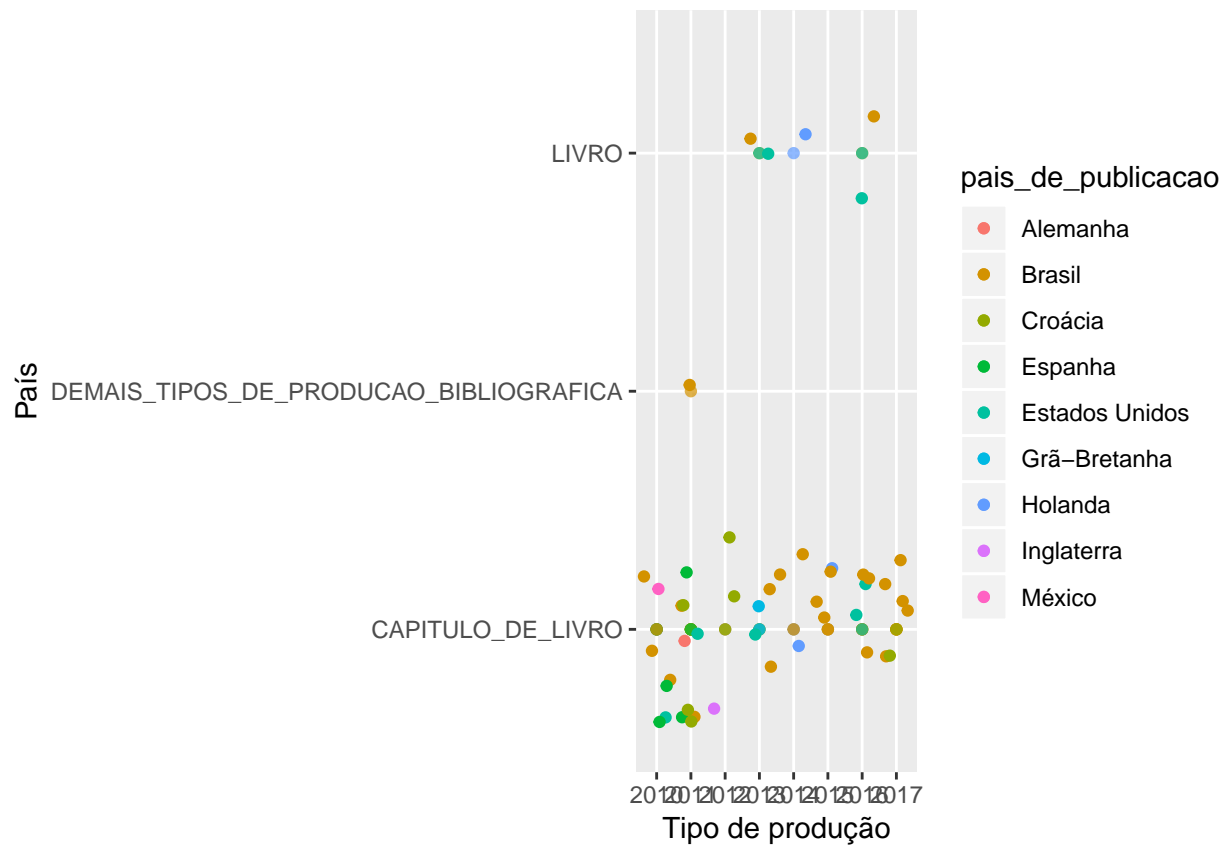
Dentre os mestrados nos programas de pós graduação estudados, pode se perceber o grande domínio em volume da pós graduação em Biologia Animal, que na maioria dos anos corresponde a quase metade das orientações de pós graduação por ano dentre os PPG estudados.

Publicações em países:

```

perfil.df.publicacoes %>%
  filter(!(tipo_producao %in% c('EVENTO', 'TEXTO_EM_JORNAIS', 'PERIODICO', 'ARTIGO_ACEITO'))) %>%
  group_by(tipo_producao, pais_de_publicacao) %>%
  ggplot(aes(ano, tipo_producao, col=pais_de_publicacao)) +
  geom_point(alpha = 0.7) + geom_jitter() +
  labs(x='Tipo de produção', y='País')

```



Observa-se deficiência de dados quanto ao país de publicação para periódicos, textos em jornais e artigos aceitos para os PPG analisados. Ainda assim, este gráfico demonstra bem a heterogeniedade das publicações de livros e/ou capítulos no Brasil e demais países.