

Análise inicial dos arquivos gerados pelo e-Lattes

LEMBRAR DE CONFIGURAR O DIRETÓRIO NO COMANDO setwd()

Pacotes para serem ativados

```
library(tidyverse) #Importado para manipulação de tibbles
library(jsonlite) #Importado para carga dos arquivos JSON para o R
library(listviewer) #Importado para análise dos arquivos JSON
library(igraph) #Importado para manipulação de grafo
library(dplyr) #Importado para uso do Operador Pipe
library(tidyr) #Importado para uso da função spread()
library(ggplot2) #Importado para visualizações com ggplot()
library(stringr) #str_to_upper

#CONFIGURAR - DIRETÓRIO DOS .R
setwd("~/Repository/DataScience/DS4A-BioAni-BioMic-BioMol-PatMol")
source("elattes.ls2df.R")

#CONFIGURAR - DIRETÓRIO DOS .JSON (Arquivos eLattes)
perfil <- fromJSON("240BiologiaAnimal/240profile.json")
public <- fromJSON("240BiologiaAnimal/240publication.json")
orient <- fromJSON("240BiologiaAnimal/240advise.json")
graphl <- fromJSON("240BiologiaAnimal/240graph.json")

#Arquivos Sucupira? Pesquisar
#res.area <- fromJSON("UnBPosGeral/researchers_by_area.json") #NÃO UTILIZADO
df.prog <- read.table("UnBPosGeral/prof_prog.csv", sep = ",",
                      colClasses = "character", encoding = "UTF-8", header = TRUE)

#SEPARAÇÃO DOS CAMPOS DE DF.PROG - Depende dos arquivos Sucupira
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")

#####
#Análise do arquivo perfil

#jsonedit(perfil)
#jsonedit(public)

#Número de Docentes encontrados
length(perfil)

## [1] 19

#Atributos presentes em um dos pesquisadores
glimpse(perfil[[1]], width = 30)

## List of 7
## $ nome : chr "Renato Caparroz"
```

```

## $ resumo_cv          : chr "Formado em Zootecnia pela Universidade Estadual Paulista Júlio de Mesquita
## $ areas_de_atuacao   :'data.frame':  3 obs. of  4 variables:
##   ..$ grande_area    : chr [1:3] "CIENCIAS_BIOLOGICAS" "CIENCIAS_BIOLOGICAS" "CIENCIAS_BIOLOGICAS"
##   ..$ area           : chr [1:3] "Genética" "Genética" "Zoologia"
##   ..$ sub_area       : chr [1:3] "Genética Animal" "Genética Molecular e de Microorganismos" ""
##   ..$ especialidade: chr [1:3] "" "" ""
## $ endereco_profissional :List of 8
##   ..$ instituicao: chr "Universidade de Brasília"
##   ..$ orgao      : chr "Instituto de Ciências Biológicas"
##   ..$ unidade    : chr "Departamento de Genética e Morfologia"
##   ..$ DDD        : chr "61"
##   ..$ telefone   : chr "31073070"
##   ..$ bairro     : chr "Asa Norte"
##   ..$ cep        : chr "70910900"
##   ..$ cidade     : chr "Brasília"
## $ producao_bibliografica :List of 3
##   ..$ CAPITULO_DE_LIVRO:'data.frame':  2 obs. of  13 variables:
##     .. ..$ tipo          : chr [1:2] "Capítulo de livro publicado" "Capítulo de livro publicado"
##     .. ..$ titulo_do_capitulo : chr [1:2] "Contribuição da análise de marcadores moleculares para a caracterização de espécies"
##     .. ..$ titulo_do_livro   : chr [1:2] "PLANO DE AÇÃO NACIONAL PARA A CONSERVAÇÃO DO GÊNERO AMAZONIA"
##     .. ..$ ano             : chr [1:2] "2011" "2017"
##     .. ..$ doi             : chr [1:2] "" ""
##     .. ..$ pais_de_publicacao : chr [1:2] "Brasil" "Brasil"
##     .. ..$ isbn           : chr [1:2] "999999999999X" "9788583320685"
##     .. ..$ nome_da_editora   : chr [1:2] "" "Imprell Gráfica e Editora"
##     .. ..$ numero_da_edicao_revisao: chr [1:2] "" "1"
##     .. ..$ organizadores    : chr [1:2] "Onildo João Marini-Filho" "Fábio José Viana Costa; Juliana"
##     .. ..$ paginas          : chr [1:2] " - " "141 - 258"
##     .. ..$ autores         :List of 2
##     .. ..$ autores-endogeno :List of 2
##   ..$ EVENTO          :'data.frame':  19 obs. of  11 variables:
##     .. ..$ natureza      : chr [1:19] "RESUMO" "RESUMO" "RESUMO" "RESUMO" ...
##     .. ..$ titulo        : chr [1:19] "Avaliação da eficiência de cinco locos microssatélite para espécies de peixes"
##     .. ..$ nome_do_evento : chr [1:19] "2o Congresso de Genética do Centro-oeste" "56o Congresso Brasileiro de Genética"
##     .. ..$ ano_do_trabalho : chr [1:19] "2010" "2010" "2010" "2011" ...
##     .. ..$ pais_do_evento : chr [1:19] "Brasil" "Brasil" "Brasil" "Brasil" ...
##     .. ..$ cidade_do_evento: chr [1:19] "Goiânia" "Guarujá" "Goiânia" "Cusco" ...
##     .. ..$ doi           : chr [1:19] "" "" "" "" ...
##     .. ..$ classificacao : chr [1:19] "REGIONAL" "NACIONAL" "REGIONAL" "INTERNACIONAL" ...
##     .. ..$ paginas       : chr [1:19] " - " "81 - 81" "1 - 1" "368 - 369" ...
##     .. ..$ autores       :List of 19
##     .. ..$ autores-endogeno:List of 19
##   ..$ PERIODICO        :'data.frame':  13 obs. of  10 variables:
##     .. ..$ natureza      : chr [1:13] "COMPLETO" "COMPLETO" "COMPLETO" "COMPLETO" ...
##     .. ..$ titulo        : chr [1:13] "Isolation and Characterization of Microsatellite Loci for Neotropical Fishes"
##     .. ..$ periodico     : chr [1:13] "Journal of Heredity" "Genetics and Molecular Biology (Imprints)"
##     .. ..$ ano           : chr [1:13] "2010" "2011" "2011" "2011" ...
##     .. ..$ volume        : chr [1:13] "101" "34" "34" "38" ...
##     .. ..$ issn          : chr [1:13] "00221503" "14154757" "14154757" "03014851" ...
##     .. ..$ paginas       : chr [1:13] "385 - 389" "161 - 164" "348 - 352" "000 - " ...
##     .. ..$ doi           : chr [1:13] "10.1093/jhered/esp116" "10.1590/S1415-47572010005000112" "" ""
##     .. ..$ autores       :List of 13
##     .. ..$ autores-endogeno:List of 13
## $ orientacoes_academicas:List of 4

```

```

## ..$ ORIENTACAO_CONCLUIDA_DOUTORADO : 'data.frame': 6 obs. of 13 variables:
## ..$ natureza : chr [1:6] "Tese de doutorado" "Tese de doutorado" "Tese de dou
## ..$ titulo : chr [1:6] "DESENVOLVIMENTO E UTILIZAÇÃO DE MARCADORES MICROSSA
## ..$ ano : chr [1:6] "2011" "2012" "2015" "2015" ...
## ..$ id_lattes_aluno : chr [1:6] "" "3120545955849198" "8940212399170232" "8940212399
## ..$ nome_aluno : chr [1:6] "Dimas Oliveira Santos" "Cássia Alves Lima" "Amanda
## ..$ instituicao : chr [1:6] "Universidade Estadual Paulista Júlio de Mesquita Fi
## ..$ curso : chr [1:6] "Zootecnia" "Biologia Animal" "Biologia Animal" "Bio
## ..$ codigo_do_curso : chr [1:6] "60029790" "51500450" "51500450" "51500450" ...
## ..$ bolsa : chr [1:6] "NAO" "SIM" "SIM" "SIM" ...
## ..$ agencia_financiadora : chr [1:6] "" "Coordenação de Aperfeiçoamento de Pessoal de Nív
## ..$ codigo_agencia_financiadora: chr [1:6] "" "045000000000" "045000000000" ...
## ..$ nome_orientadores :List of 6
## ..$ id_lattes_orientadores :List of 6
## ..$ ORIENTACAO_CONCLUIDA_MESTRADO : 'data.frame': 8 obs. of 13 variables:
## ..$ natureza : chr [1:8] "Dissertação de mestrado" "Dissertação de mestrado"
## ..$ titulo : chr [1:8] "Estrutura social do veado-campeiro (Ozotocerus bezo
## ..$ ano : chr [1:8] "2011" "2012" "2014" "2015" ...
## ..$ id_lattes_aluno : chr [1:8] "" "8940212399170232" "1736902373766697" "2084644920
## ..$ nome_aluno : chr [1:8] "Aline Meira Bonfim Mantellatto" "Amanda Vaz Rocha"
## ..$ instituicao : chr [1:8] "Universidade Estadual Paulista Júlio de Mesquita Fi
## ..$ curso : chr [1:8] "Genética e Melhoramento Animal" "Ecologia e Evoluçã
## ..$ codigo_do_curso : chr [1:8] "34180044" "60021144" "51500450" "51500450" ...
## ..$ bolsa : chr [1:8] "SIM" "SIM" "NAO" "NAO" ...
## ..$ agencia_financiadora : chr [1:8] "Fundação de Amparo à Pesquisa do Estado de São Paul
## ..$ codigo_agencia_financiadora: chr [1:8] "037700000002" "045000000000" "" "" ...
## ..$ nome_orientadores :List of 8
## ..$ id_lattes_orientadores :List of 8
## ..$ ORIENTACAO_EM_ANDAMENTO_DOUTORADO: 'data.frame': 5 obs. of 13 variables:
## ..$ natureza : chr [1:5] "Tese de doutorado" "Tese de doutorado" "Tese de dou
## ..$ titulo : chr [1:5] "FILOGENIA MOLECULAR DE SERPENTES NEOTROPICAIS DO GRU
## ..$ ano : chr [1:5] "2014" "2015" "2015" "2016" ...
## ..$ id_lattes_aluno : chr [1:5] "1736902373766697" "" "0919072891217484" "5963240497
## ..$ nome_aluno : chr [1:5] "Diogo dos Santos Nascimento" "Gislaine Aparecida Fe
## ..$ instituicao : chr [1:5] "Universidade de Brasília" "Universidade de Brasília
## ..$ curso : chr [1:5] "Biologia Animal" "Biologia Animal" "Biologia Animal
## ..$ codigo_do_curso : chr [1:5] "51500450" "51500450" "51500450" "51500450" ...
## ..$ bolsa : chr [1:5] "SIM" "SIM" "SIM" "SIM" ...
## ..$ agencia_financiadora : chr [1:5] "Coordenação de Aperfeiçoamento de Pessoal de Nível
## ..$ codigo_agencia_financiadora: chr [1:5] "045000000000" "045000000000" "045000000000" "045000
## ..$ nome_orientadores :List of 5
## ..$ id_lattes_orientadores :List of 5
## ..$ OUTRAS_ORIENTACOES_CONCLUIDAS : 'data.frame': 18 obs. of 13 variables:
## ..$ natureza : chr [1:18] "INICIACAO_CIENTIFICA" "INICIACAO_CIENTIFICA" "INIC
## ..$ titulo : chr [1:18] "Biologia reprodutiva e proporção sexual do papagai
## ..$ ano : chr [1:18] "2010" "2010" "2010" "2011" ...
## ..$ id_lattes_aluno : chr [1:18] "" "" "" "" ...
## ..$ nome_aluno : chr [1:18] "Luma Tatiana Silva Castro" "Giovana Colnaghi Abdal
## ..$ instituicao : chr [1:18] "Universidade Federal de Goiás" "Universidade Feder
## ..$ curso : chr [1:18] "Medicina Veterinária" "Medicina Veterinária" "Ciênc
## ..$ codigo_do_curso : chr [1:18] "90000005" "90000005" "90000008" "90000008" ...
## ..$ bolsa : chr [1:18] "NAO" "NAO" "SIM" "SIM" ...
## ..$ agencia_financiadora : chr [1:18] "" "" "Conselho Nacional de Desenvolvimento Científ
## ..$ codigo_agencia_financiadora: chr [1:18] "" "" "002200000000" "002200000000" ...

```

```

##    .. ..$ nome_orientadores      :List of 18
##    .. ..$ id_lattes_orientadores  :List of 18
##    $ senioridade                  : chr "9"

#Constroi lista de nomes de docentes/pessoas
ProfileList <- list()
for (i in 1:length(perfil)) {
  ProfileList <- rbind(ProfileList, perfil[[i]]$nome)
}

##Analise dos dados em formato list
# Numero de areas de atuacao cumulativo
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 especificos 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_EM_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_EM_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$producao_bibliografica$TEXTOS_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_DOUTORADOS$ano)),
##
## 0 1 2 3 5 6 7 8 10 13
## 1 4 2 2 1 2 2 3 1 1
table(unlist(sapply(perfil, function(x) length(x$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 dados em formato Data Frame
#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)
perfil.df.publicacoes$tipo_producao[grepl("DEMAIS", perfil.df.publicacoes$tipo_producao)] <- "OUTRAS_PROD"

perfil.df.publicacoes %>% select(pais_de_publicacao) %>% distinct()

## pais_de_publicacao
## 1 <NA>
## 2 Brasil
## 3 Inglaterra
## 4 Holanda
## 5 Estados Unidos
## 6 Alemanha
## 7 Croácia
## 8 Espanha
## 9 Grã-Bretanha
## 10 México

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

```

```
mutate(situacao = ifelse(grepl("CONCLUIDA", orientacao), "Concluída", "Em andamento")) %>%
mutate(Natureza = case_when(grepl("MESTRADO", str_to_upper(natureza)) ~ "Mestrado",
                             grepl("PÓS-DOCTORADO", str_to_upper(natureza)) ~ "Pós-doutorado",
                             grepl("DOCTORADO", str_to_upper(natureza)) ~ "Doutorado",
                             grepl("INICIACAO", str_to_upper(natureza)) ~ "Iniciação Científica",
                             grepl("INICIAÇÃO", str_to_upper(natureza)) ~ "Iniciação Científica",
                             TRUE ~ "Outras naturezas"))

perfil.df.orientacoes %>% select(natureza, Natureza) %>% distinct() %>% arrange(Natureza)
```

```
##                                     natureza
## 1                                     Tese de doutorado
## 2                                     INICIACAO_CIENTIFICA
## 3                                     Iniciação Científica
## 4                                     Dissertação de mestrado
## 5                                     ORIENTACAO-DE-OUTRA-NATUREZA
## 6                                     TRABALHO_DE_CONCLUSAO_DE_CURSO_GRADUACAO
## 7 MONOGRAFIA_DE_CONCLUSAO_DE_CURSO_APERFEICOAMENTO_E_ESPECIALIZACAO
## 8                                     Supervisão de pós-doutorado
##                                     Natureza
## 1                                     Doutorado
## 2 Iniciação Científica
## 3 Iniciação Científica
## 4                                     Mestrado
## 5                                     Outras naturezas
## 6                                     Outras naturezas
## 7                                     Outras naturezas
## 8                                     Pós-doutorado
```

```
str_to_upper("pós-doutorado")
```

```
## [1] "PÓS-DOCTORADO"
```

```
#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(num_areas = n_distinct(area)),
  by = "idLattes")

glimpse(perfil.df)

## Observations: 19
## Variables: 15
## $ idLattes          <chr> "1612292306950738", "186941...
## $ nome              <chr> "Renato Caparroz", "Monica ...
## $ resumo_cv         <chr> "Formado em Zootecnia pela ...
## $ senioridade       <chr> "9", "8", "9", "9", "9", "9...
## $ ORIENTACAO_CONCLUIDA_DOUTORADO <int> 6, 1, 6, 7, NA, 1, 1, 2, 5,...
## $ ORIENTACAO_CONCLUIDA_MESTRADO <int> 8, 6, 10, 1, 9, 5, 2, 5, 13...
## $ ORIENTACAO_CONCLUIDA_POS_DOUTORADO <int> NA, NA, NA, NA, NA, NA, NA,...
## $ OUTRAS_ORIENTACOES_CONCLUIDAS <int> 18, 9, 17, NA, 18, 9, 2, 6,...
## $ CAPITULO_DE_LIVRO <int> 2, NA, 4, 9, 1, 4, 2, NA, N...
## $ EVENTO           <int> 19, 14, 12, NA, 11, 10, 4, ...
## $ LIVRO             <int> NA, NA, NA, 1, NA, NA, NA, ...
## $ OUTRAS_PRODUCOES <int> NA, NA, NA, NA, NA, NA, NA,...
## $ PERIODICO         <int> 13, 10, 21, 68, 18, 44, 26,...
## $ TEXTO_EM_JORNAIS <int> NA, NA, NA, NA, NA, 1, NA, ...
## $ num_areas         <int> 2, 3, 1, 1, 4, 2, 3, 1, 3, ...

####
###Publicacao
##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"

```

```

#BLOCO ABAIXO REQUER BASE DE DADOS df.prog
df <- left_join(df, df.prog, by = c("Idlattes" = "idLattes")) #

#Apenas para fins de analise 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

#Indice de bolsas

bolsas.df <- orient.df %>% filter(bolsa == "SIM") %>%
  group_by(natureza, ano) %>%
  summarise(num_bolsas = n()) %>%
  inner_join((orient.df %>%
    group_by(natureza, ano) %>%
    summarise(num_orientes = n())), by = c("natureza", "ano")) %>%
  mutate(ratio = num_bolsas/num_orientes)

```

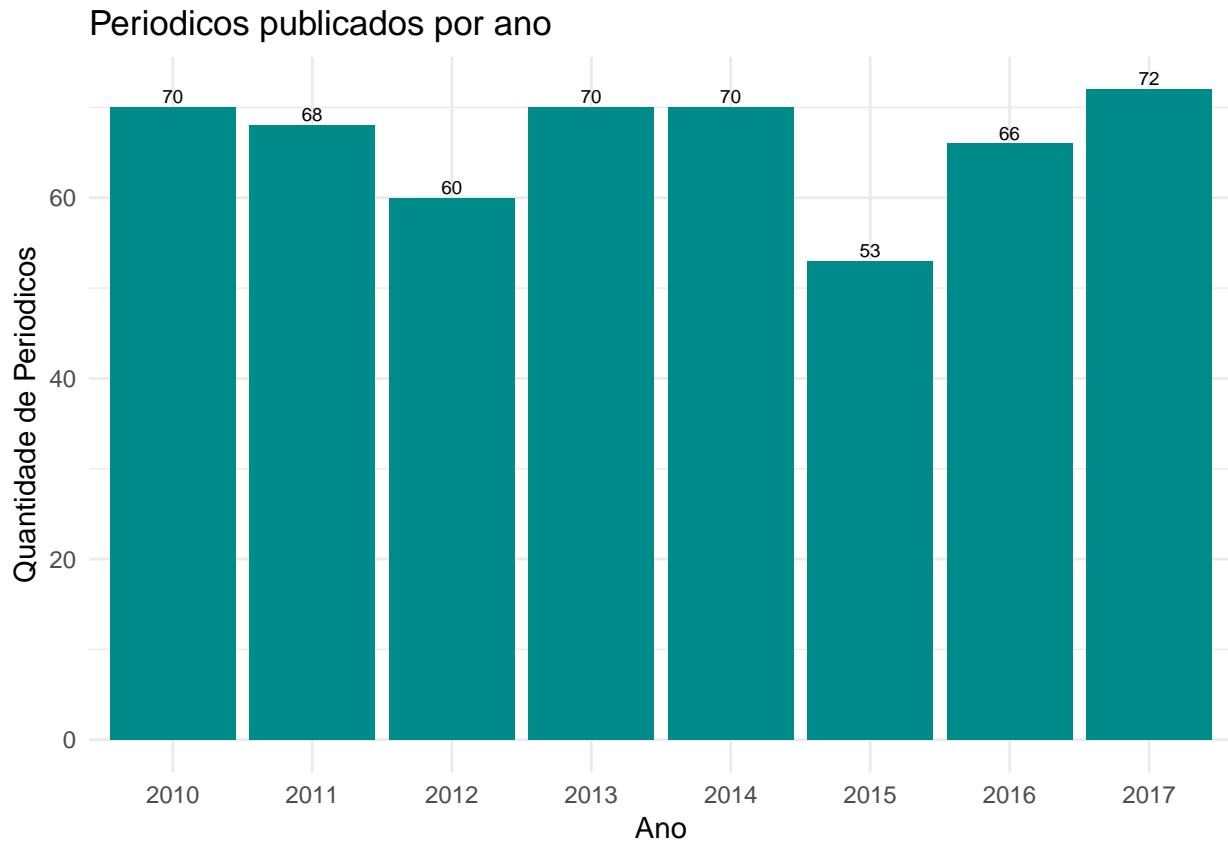
Visualizacao

Grafico de barras; periodicos por ano

```

public.periodico.df %>%
  group_by(ano) %>%
  summarise(Quantidade = n()) %>%
  ggplot(aes(ano, 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")

```



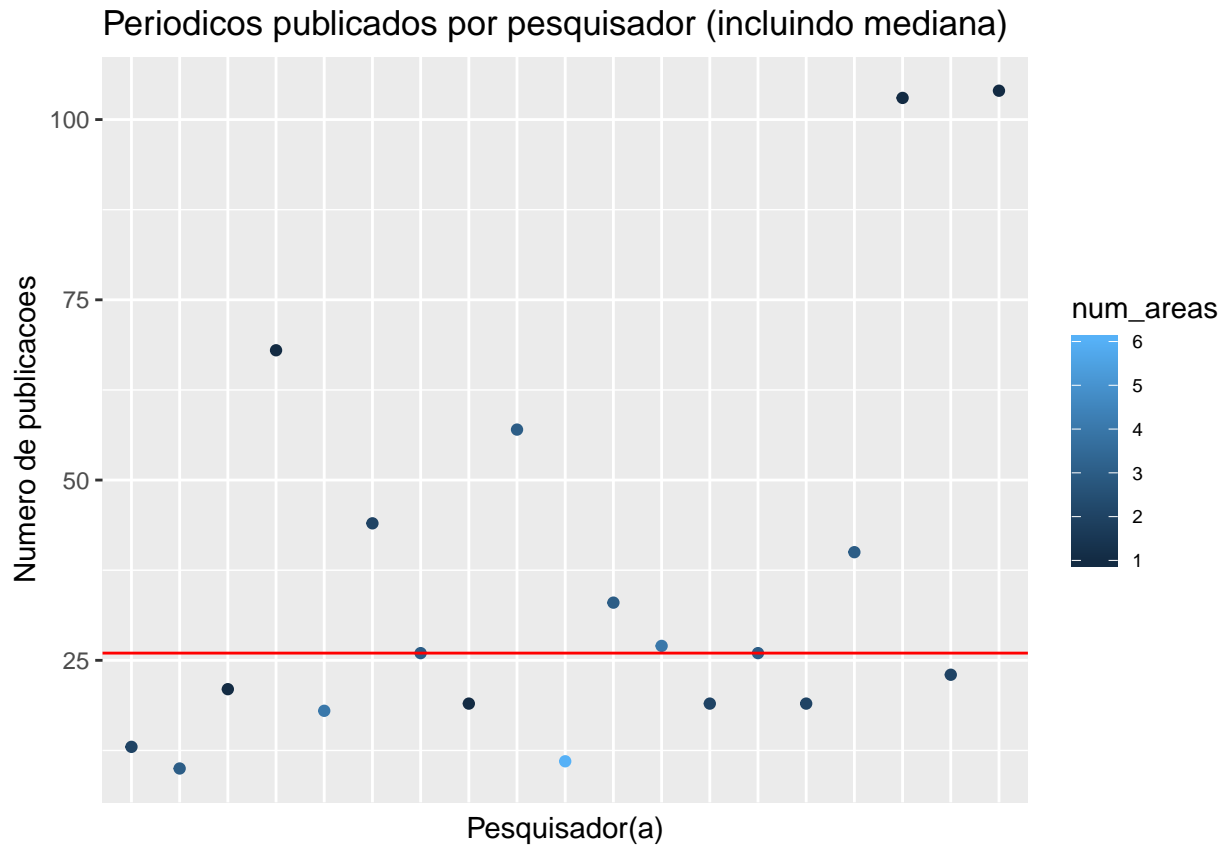
Sob uma perspectiva de publicações em periódicos, o Programa de Pós-Graduação de Biologia Animal apresentou pouco crescimento entre 2010 e 2017, registrando de 60 a 70 publicações por ano em grande parte da amostra.

Quantidade de periodicos publicados por professor(a) entre 2010 e 2017 Numero de areas de pesquisa do professor

```

perfil.df %>%
  ggplot(aes(idLattes,PERIODICO, color = num_areas)) +
  geom_point() +
  ggtitle("Periodicos publicados por pesquisador (incluindo mediana)") +
  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")

```



Investigando o comportamento dos pesquisadores do programa, os professores foram ordenados pelo número total de publicações em periódicos entre 2010 e 2017, cuja mediana foi apresentada pela linha vermelha do gráfico acima, que sinaliza 26 publicações no período total. O gráfico indica que, dos 19 pesquisadores incluídos na base de dados, dois totalizaram publicações que duplicaram este valor e outros dois quadruplicaram este valor, com mais de 104 publicações em periódicos.

publicacao de livros por pais/ano

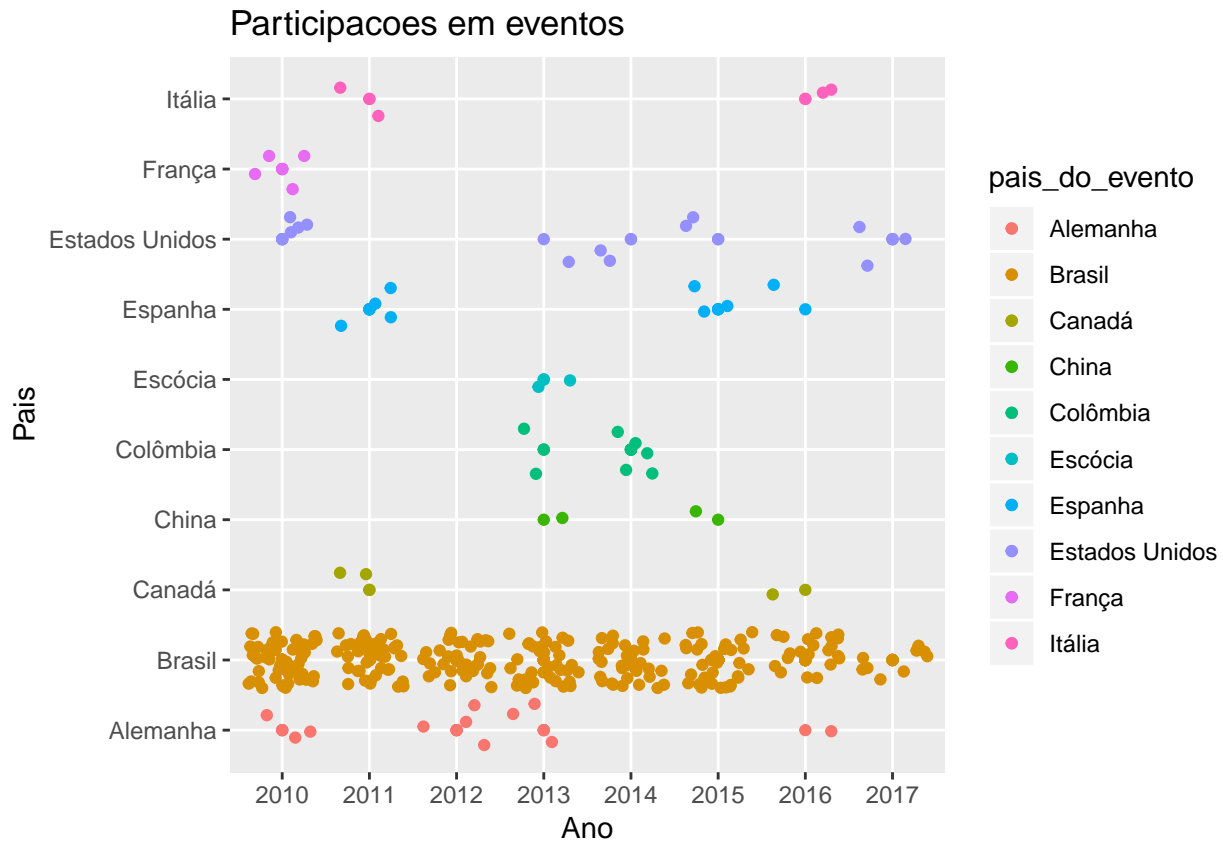
```
public.livros.df %>%
  group_by(ano, pais_de_publicacao) %>%
  ggplot(aes(x=ano, y=pais_de_publicacao, color= pais_de_publicacao)) +
  ggtitle("Livros publicados por ano") +
  xlab("Ano") + ylab("Pais") + geom_point() + geom_jitter()
```



Apenas três anos apresentaram registros de publicação de livros por parte do Programa no período de 2010 a 2017. Entretanto, uma observação notável é um maior número de publicações em países estrangeiros que no Brasil, o que sinaliza um bom grau de internacionalização do programa.

Eventos nacionais e internacionais

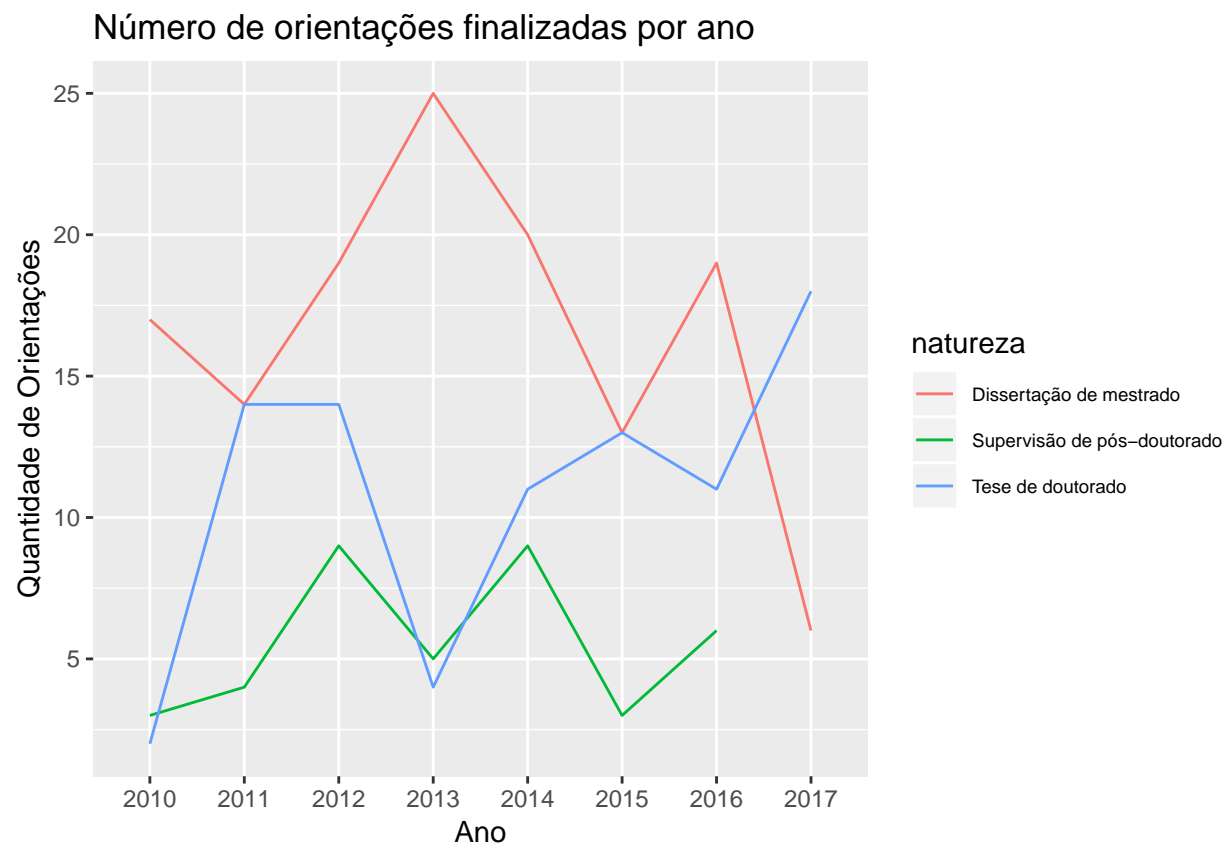
```
public.eventos.df %>%
  filter(pais_do_evento %in%
    c(names(head(sort(table(public.eventos.df$pais_do_evento)
      , decreasing = TRUE), 10)))) %>%
  group_by(ano_do_trabalho,pais_do_evento) %>%
  ggplot(aes(x=ano_do_trabalho,y=pais_do_evento, color= pais_do_evento)) +
  ggtitle("Participacoes 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 pelo menos um evento de sede internacional em todos os anos registrados na base de dados.

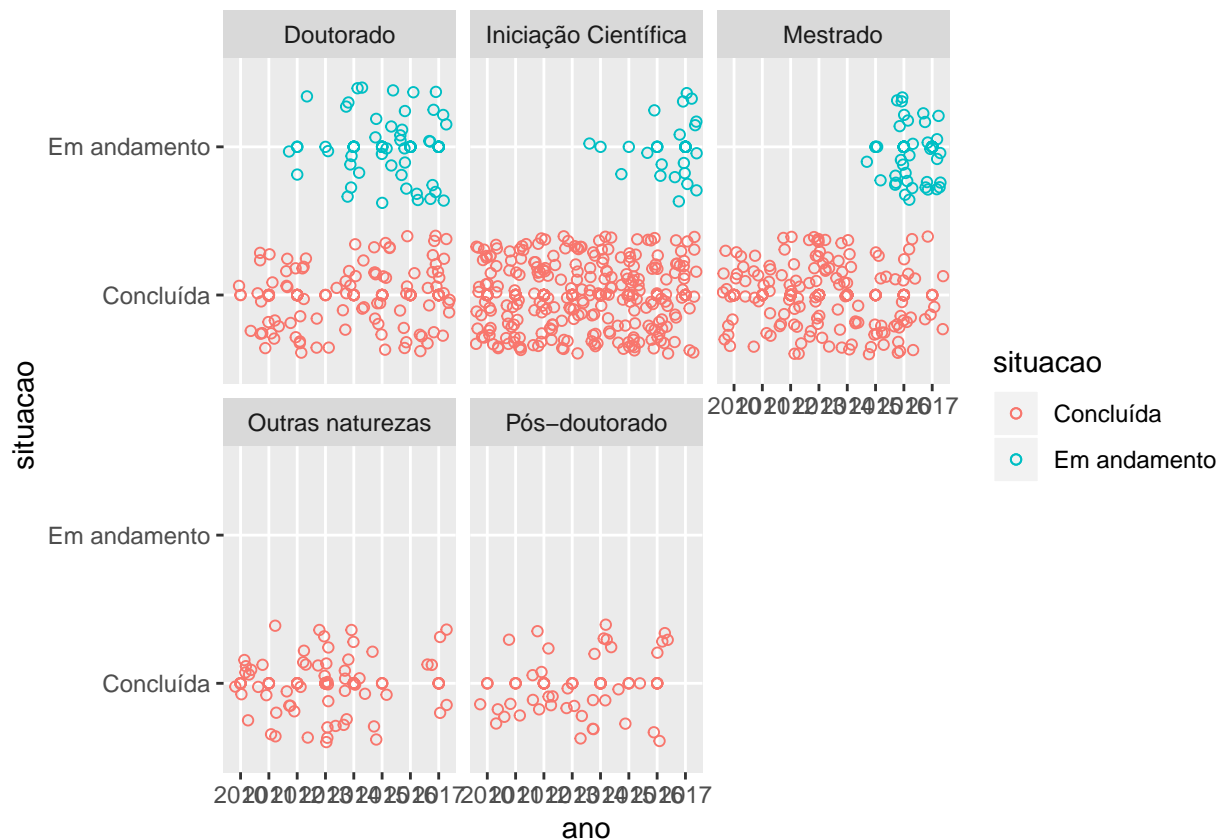
Orientacoes completas por ano e natureza

```
orient.df %>%
  group_by(ano, natureza) %>% summarise(total = n()) %>%
  ggplot(aes(x=ano,y=total,group=natureza,color=natureza)) +
  geom_line() +
  ggtitle("Número de orientações finalizadas 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 Orientações")
```

Evolução temporal de orientações em andamento e concluídas

```
perfil.df.orientacoes %>% group_by(ano, situacao) %>%
  ggplot(aes(x=ano,y=situacao,color=situacao)) +
  geom_point(shape = 1) + geom_jitter(shape = 1) + facet_wrap(. ~ Natureza)
```



Observando a evolução do número de orientações completas ao longo dos anos, percebe-se que o Programa de Pós-Graduação cresceu consideravelmente nas naturezas de doutorado e pós-doutorado após 2010. Entretanto, não há um comportamento linear na evolução do número de orientações finalizadas. A partir de 2011, todos os anos pares apresentaram um aumento no número de pós-doutorados enquanto os anos ímpares seguintes mostram retrocesso neste número. Além disso, o ano de 2017 apresentou um aumento considerável no número de teses de doutorado, mas uma queda vertiginosa nas demais naturezas de orientações - sugerindo uma mudança brusca nas bases do programa.

Índice de bolsas entre naturezas e evolução temporal

```
bolsas.df %>%
  ggplot(aes(x=ano,y=ratio*100,color=natureza, group=natureza)) +
  geom_line() + #facet_wrap(. ~ natureza) + #CASO PREFIRA DIVIDIR AS NATUREZAS
  ggtitle("Porcentagem de orientações contempladas por bolsas") +
  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="Índice de bolsas")
```

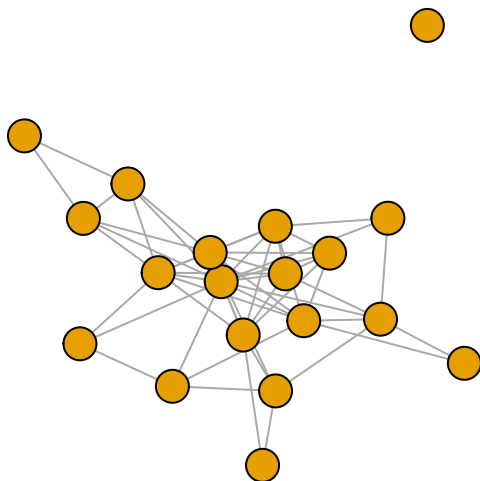
Porcentagem de orientações contempladas por bolsas



É possível perceber que o número de bolsas oferecidas para o Programa não apresentou um comportamento linear ao longo dos anos. Além disso, as teses de pós-doutorado se mostram a natureza de pesquisa melhor contemplada pelas agências financiadoras, visto que quase todas as observações receberam bolsa. Por fim, é possível observar que os anos com maior número de orientações de mestrado e doutorado também consistem nos anos com maior número de bolsas distribuídas entre o programa, mas a ocorrência de pesquisas realizadas sem este apoio financeiro foi marcante em todo o período.

Grafo de proximidade entre pesquisadores do Programa de Pós-Graduação

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

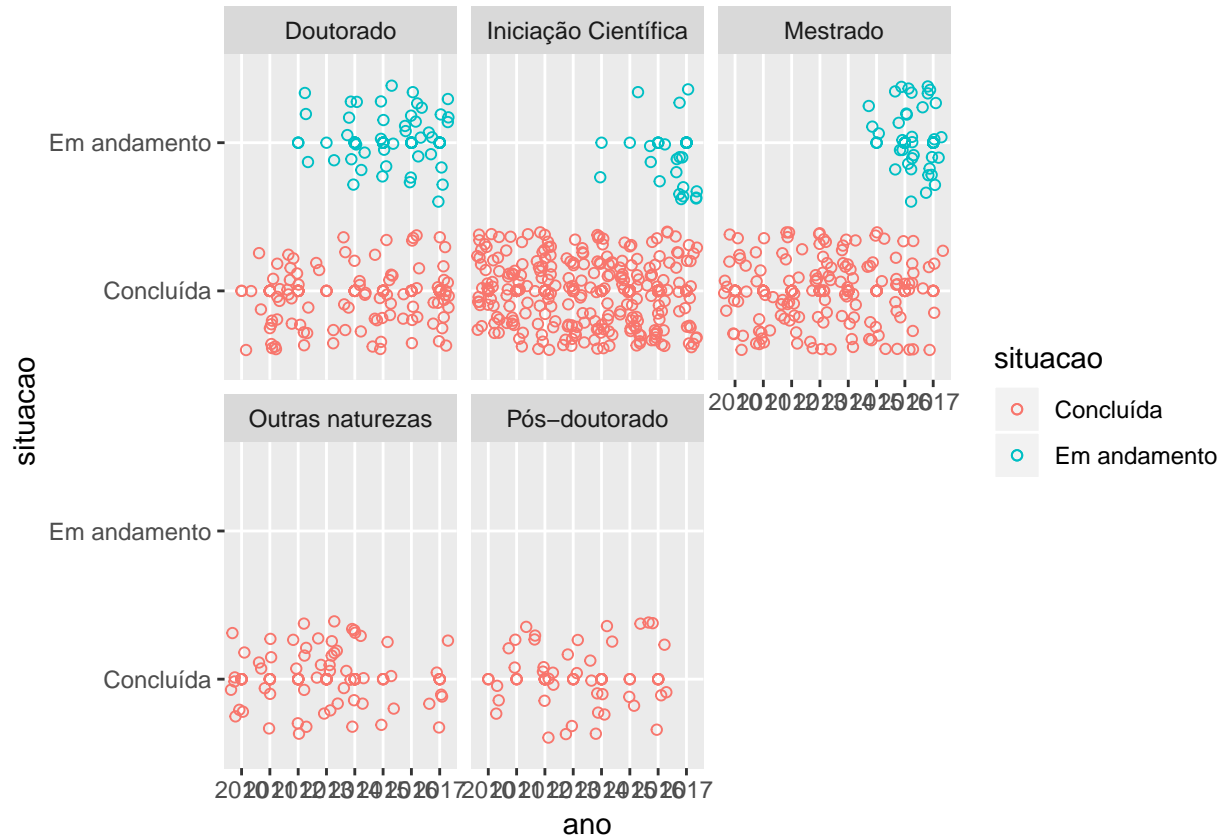


O grafo acima representa os pesquisadores do Programa de Pós-Graduação em seus vértices e a existência de cooperação entre eles em suas arestas. Portanto, é possível observar uma considerável cooperação entre os

membros do Programa, mas alguns pesquisadores relatam baixa diversidade em suas colaborações no período (2010 a 2017).

Natureza das orientações por tipo de orientação, ano e andamento

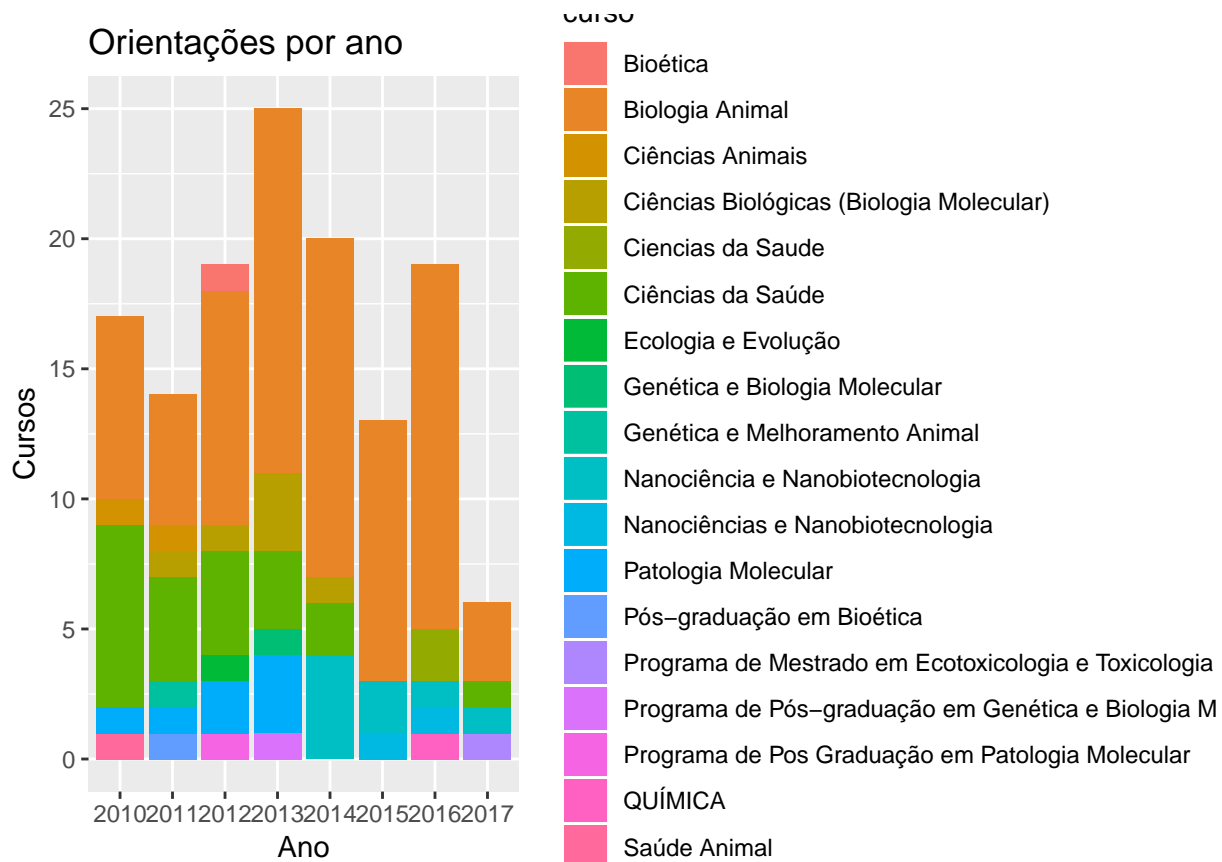
```
perfil.df.orientacoes %>% group_by(ano, situacao) %>%
  ggplot(aes(x=ano, y=situacao, color=situacao)) +
  geom_point(shape = 1) + geom_jitter(shape = 1) + facet_wrap(. ~ Natureza)
```



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 - NÃO CONSEGUI LEGENDA PRA TODOS OS CURSOS

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



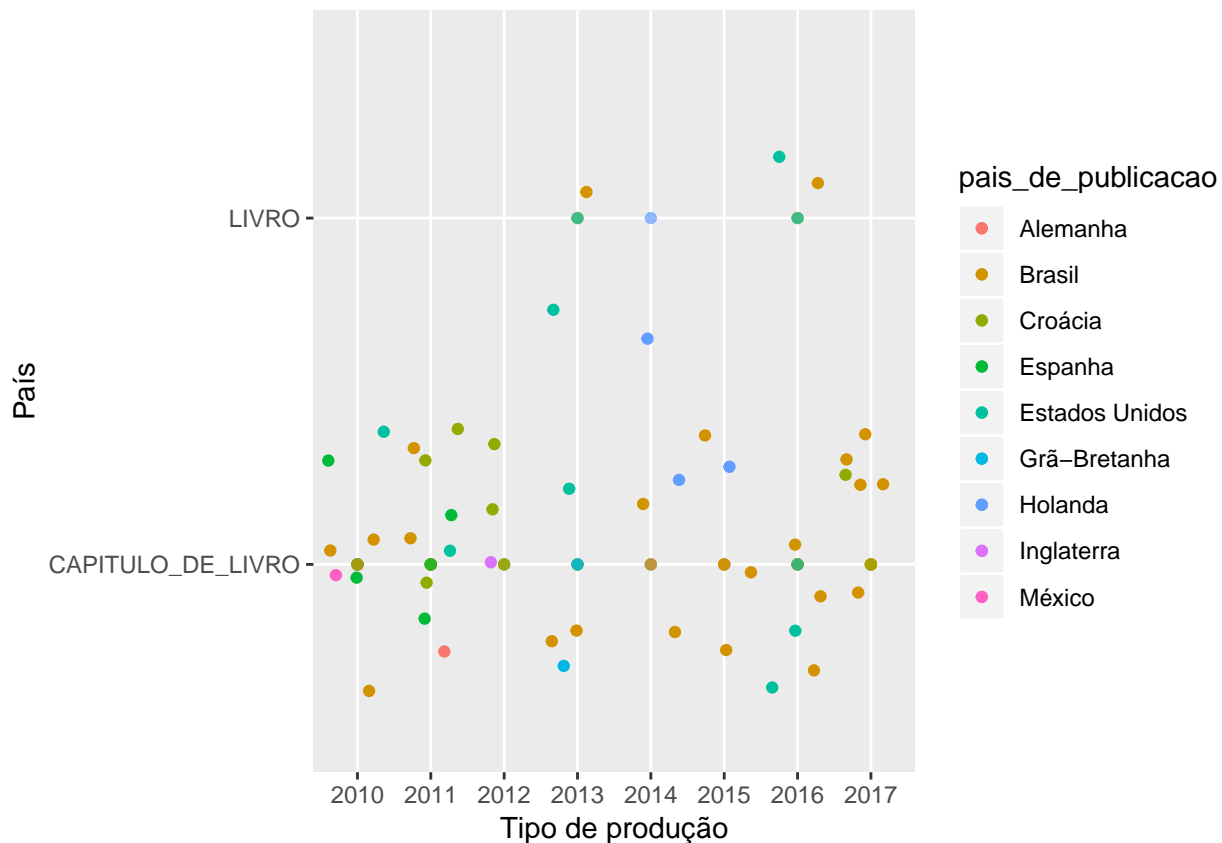
Dentre os mestrados nos programas de pós graduação estudados, pode se perceber o grande domínio em volume da pós graduação em Biologia animal, que na maioria dos anos corresponde a quase metade das orientações de pós graduação por ano dentre os PPG estudados.

Publicações de capítulos de livros por ano/país

```

perfil.df.publicacoes %>%
  filter((tipo_producao %in% c('LIVRO', 'CAPITULO_DE_LIVRO'))) %>%
  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.

Perfil-Areas - Questao 12

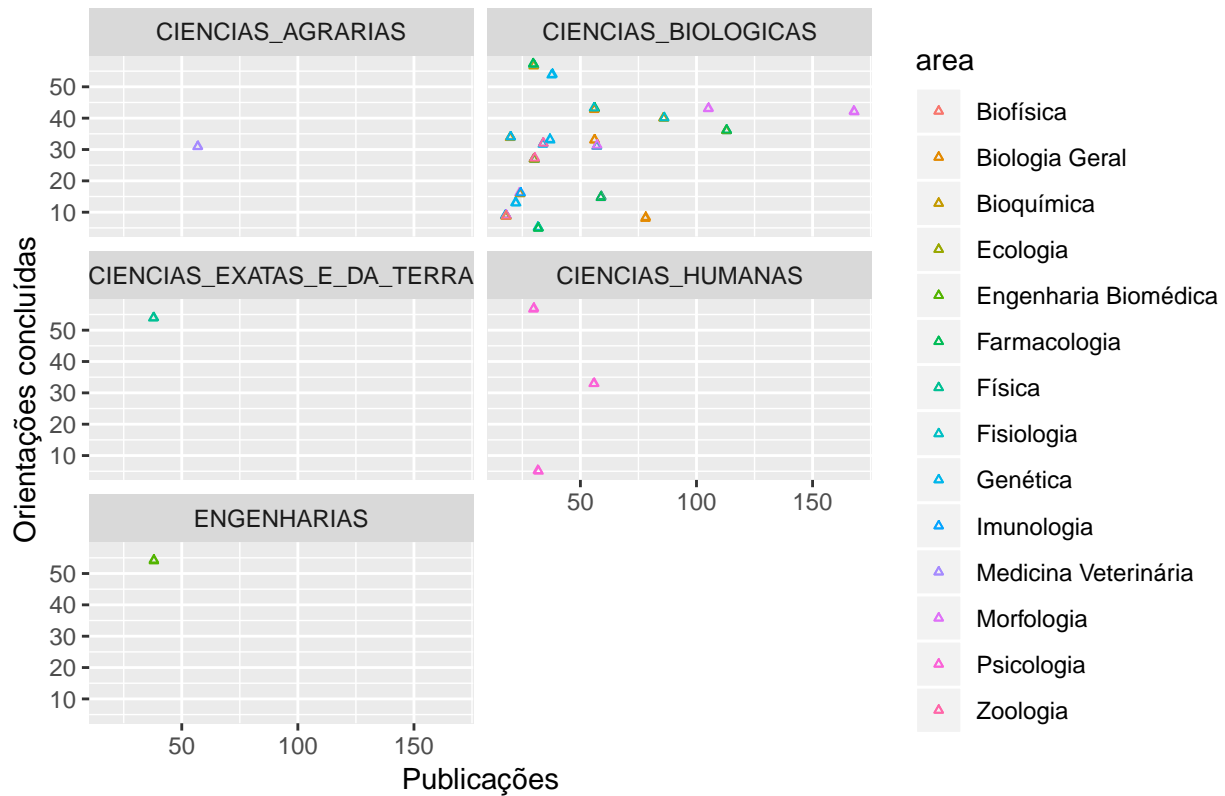
```

perfil.areas <- perfil.df.areas.de.atuacao %>%
  left_join(perfil.df, by = "idLattes") %>%
  rowwise() %>% #realizar sum() corretamente
  mutate(orientacoes_concluidas = sum(ORIENTACAO_CONCLUIDA_DOUTORADO,
    ORIENTACAO_CONCLUIDA_POS_DOUTORADO, ORIENTACAO_CONCLUIDA_MESTRADO,
    OUTRAS_ORIENTACOES_CONCLUIDAS, na.rm = TRUE)) %>%
  mutate(publicacoes = sum(CAPITULO_DE_LIVRO, EVENTO, PERIODICO,
    LIVRO, TEXTO_EM_JORNAIS, OUTRAS_PRODUCOES, na.rm = TRUE)) %>%
  select(idLattes, grande_area, area, sub_area, especialidade, orientacoes_concluidas, publicacoes)
class(perfil.areas) <- c("tbl_df", "data.frame") #desfazer rowwise

#Graficos ignorando especialidade e subarea
perfil.areas %>%
  select(-sub_area, -especialidade) %>%
  distinct() %>%
  group_by(publicacoes) %>%
  ggplot(aes(publicacoes, orientacoes_concluidas, color = area)) +
  geom_point(shape = 2, size = .8) + geom_jitter(shape = 2, size = .8) +
  ggtitle('Relação de Orientações Concluídas x Publicações') +
  labs(x='Publicações',y='Orientações concluídas') + facet_wrap(. ~ grande_area, ncol = 2)

```

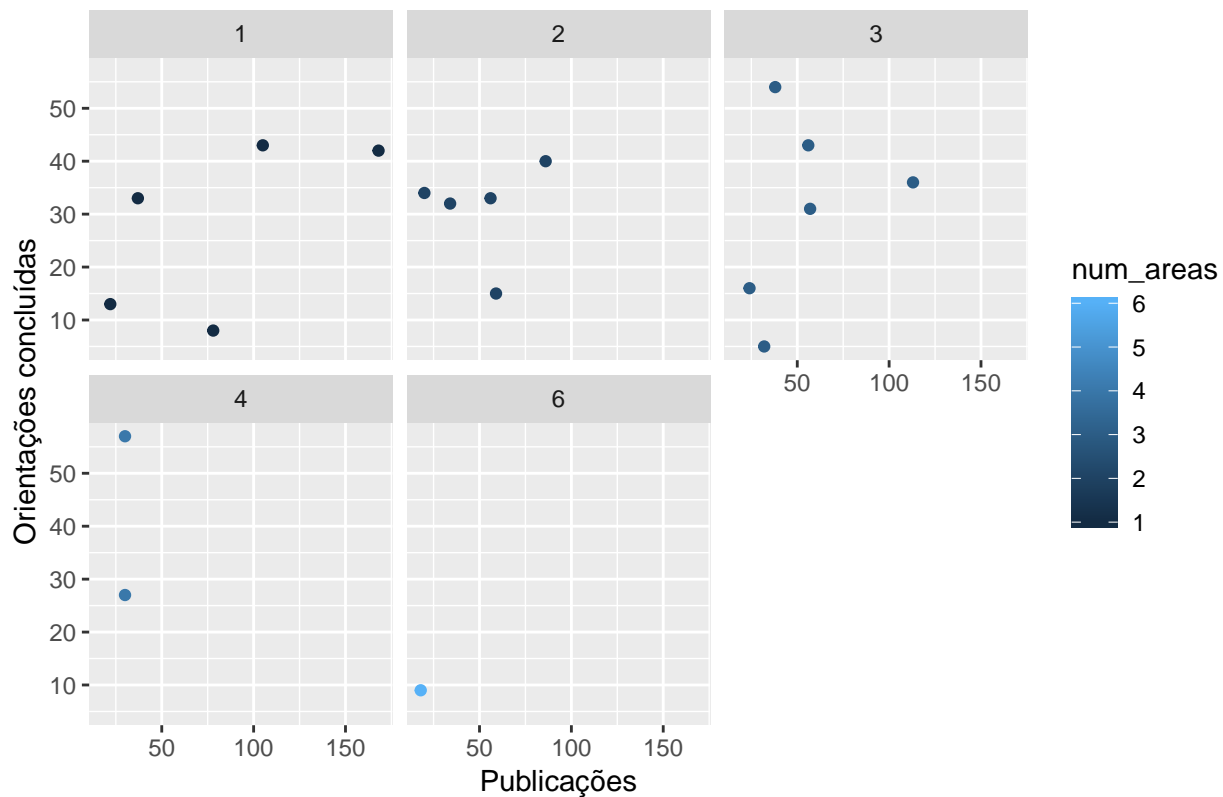
Relação de Orientações Concluídas x Publicações



Relação de produção-orientação pelo número de áreas do pesquisador Quem trabalha em mais áreas diferentes publica/orienta mais?

```
perfil.areas %>%
  select(-sub_area, -especialidade) %>%
  distinct() %>%
  group_by(idLattes, orientacoes_concluidas, publicacoes) %>%
  summarise(num_areas = n()) %>%
  ggplot(aes(publicacoes, orientacoes_concluidas, color = num_areas)) +
  geom_point() +
  ggtitle('Orientações e Publicações pelo Número de Áreas') +
  labs(x='Publicações',y='Orientações concluídas') + facet_wrap(. ~ num_areas)
```

Orientações e Publicações pelo Número de Áreas

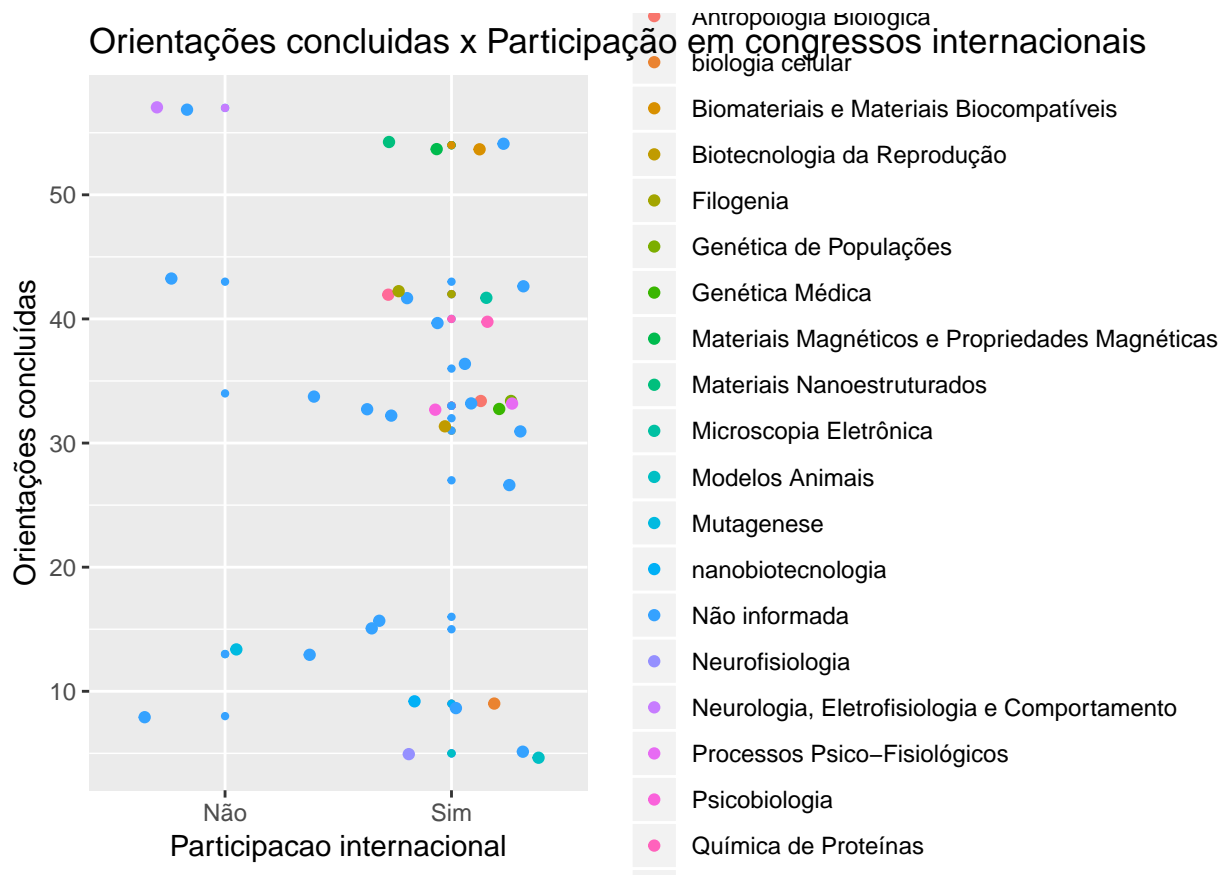


Perfil-Areas - Questao 14

```
public.eventos.df$`autores-endogeno` <- gsub('\\s+', '', public.eventos.df$`autores-endogeno`)
especialidade.orient <- public.eventos.df %>%
  filter(classificacao == "INTERNACIONAL") %>%
  select(`autores-endogeno`) %>%
  separate_rows(`autores-endogeno`, sep = ";") %>%
  distinct() %>%
  arrange(`autores-endogeno`) %>% mutate(internacional = "Sim") %>%
  right_join(perfil.df, by = c("autores-endogeno" = "idLattes")) %>%
  left_join(perfil.areas, by = c("autores-endogeno" = "idLattes")) %>%
  select(`autores-endogeno`, internacional, especialidade, orientacoes_concluidas) %>% distinct() %>%
  rename(idLattes = `autores-endogeno`)

especialidade.orient[is.na(especialidade.orient)] <- 0 #Remove NAs do dataframe
especialidade.orient$especialidade <- sub("^$", "Não informada", especialidade.orient$especialidade)
especialidade.orient$internacional <- sub(0, "Não", especialidade.orient$internacional)

especialidade.orient %>%
  ggplot(aes(internacional, orientacoes_concluidas, color = especialidade)) +
  geom_point(size = .8) + geom_jitter() +
  ggtitle('Orientações concluídas x Participação em congressos internacionais') +
  labs(x='Participacao internacional',y='Orientações concluídas')
```

Grafo - Questao 15

```
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
E(g)$eventos <- perfil.df$EVENTO
plot(g, vertex.label = NA, vertex.color=V(g)$publicacao, edge.color=E(g)$eventos)
```

