

Análise dos Programas de Pós-Graduação: Ensino de Matemática (33002010223P9), Matemática (33002010005P1) e Matemática Aplicada (33002010006P8) - Grupo 4 - Relatório Final

Anderson Silva - 14/0130004 , Marco Antonio Kuroiwa - 16/0160464, Thiago da Silva Lima - 12/0136759

30/06/2019

Contents

1	Introdução	1
1.1	O que é Ciência?	1
1.2	Ciências no Brasil	2
2	CRISP-DM	3
3	Execução	4
3.1	Entendimento de Negócios	4
3.2	Entendimento dos dados	6
3.3	Preparação dos dados	6
3.4	Avaliação	8

1 Introdução

Este trabalho é o resultado de uma investigação dos Programas de Pós Graduação sugeridos na disciplina de Tópicos Avançados em Computadores, Turma D, onde o principal foco é a Ciência de Dados. Como Metodologia, o atual projeto adota o CRISP-DM, que é um modelo orientado na solução de problemas de Mineração de Dados.

1.1 O que é Ciência?

A ciência se caracteriza pela busca de conhecimento sistemático, isto é, organizado, ordenado, metódico e seguro dos fenômenos do mundo. Um dos objetivos básicos da ciência é, tornar o mundo compreensível, proporcionando ao homem, um meio de exercer o controle sobre a natureza. Assim, se desenvolve uma visão positiva da ciência.

Objetivos de uma pesquisa:

- Fazer uma contribuição inovadora para a Ciência
- Deve responder a uma pergunta – de interesse para a comunidade científica – ainda não respondida anteriormente – de relevância para o interesse social (caso de tecnologia) A parte mais difícil é: – achar a pergunta certa !

O método científico pode ser definido como um conjunto de procedimentos por meio dos quais um cientista consegue propor um conjunto de explicações para fenômenos, constituição e formação de materiais etc. De forma geral, o método científico pode apresentar as seguintes etapas:

1. **Observação:** É a etapa em que o pesquisador observa uma determinada matéria ou fenômeno.
2. **Elaboração do problema (fase do questionamento):** Nessa etapa, o cientista ou pesquisador elabora perguntas sobre o fenômeno ou material analisado, tais como: Por que esse fenômeno ocorre? Como esse fenômeno ocorre? Quais são os fatores que originaram esse fenômeno? Qual é a composição do material? Que substâncias formam esse material? Qual é a importância desse material?
3. **Hipóteses:** É a etapa em que o pesquisador responde às perguntas feitas na etapa anterior. Essas respostas podem ser pautadas em seu conhecimento prévio sobre materiais ou fenômenos semelhantes. A elaboração das hipóteses deve ser feita com muita cautela porque é por meio delas que a fase da experimentação será realizada, ou seja, elas serão o ponto de partida da experimentação.
4. **Experimentação:** Nessa etapa, experimentos e pesquisas bibliográficas são realizados com base nas hipóteses levantadas. O objetivo é encontrar a resposta para cada um dos questionamentos que foram elaborados. Cada cientista desenvolve essa etapa de acordo com os conhecimentos que possui e as práticas que são necessárias para o esclarecimento de cada hipótese.
5. **Análise dos resultados:** Após a fase da experimentação, o pesquisador analisa cada um dos resultados para verificar se eles são suficientes para explicar cada um dos problemas levantados e também se estão de acordo com as hipóteses. Caso os resultados não sejam satisfatórios, novas hipóteses podem ser levantadas para que novas experimentações ocorram.
6. **Conclusão:** A conclusão é a etapa em que o cientista verifica se os experimentos e pesquisas realizados respondem aos questionamentos levantados e permitem que ele faça afirmações acerca dos fenômenos ou materiais analisados.

1.2 Ciências no Brasil

Muitas iniciativas científicas existiram, porém sem obter grande êxito. Foi somente no período pós guerra que se foi discutido a importância estratégica do desenvolvimento tecnológico e estabelecidas leis de incentivo e investimento ao desenvolvimento tecnológico nacional.

No segundo pós-guerra, o Brasil conheceria uma ampliação de sua base industrial e o surgimento de um debate sobre a política científica. As questões candentes da realização da independência e do desenvolvimento econômico nacional exigiriam a autonomia tecnológica e a criação de polos de ciência. Em 1947, inseriu-se na constituição estadual paulista o percentual de 0,5% da arrecadação para a pesquisa científica (aumentada para 1% na constituição estadual de 1989). A comunidade científica começa sua articulação representativa com a criação da SBPC (Sociedade Brasileira para o Progresso da Ciência), em 1948 (entre outros, por José Reis e Maurício Rocha e Silva), a partir da decisão do governador de São Paulo Adhemar de Barros limitar a atividade do Instituto Butantã à produção de soros antiofídicos. A revista *Ciência e Cultura*, lançada em abril de 1949, será seguida, no mês de outubro, da primeira reunião, em Campinas com 104 participantes. Na década de 50, a SBPC terá importante papel na discussão da LDB Lei de Diretrizes e Bases da Educação, e numa polêmica com o governo federal acerca da necessidade de uma política científica que estimulasse a ciência básica, além da aplicada. Em 15/01/1951 é criado o Conselho Nacional de Pesquisas, e seu primeiro presidente foi o vice-almirante Álvaro Alberto da Motta e Silva. Um marco para a física no Brasil é a constituição, em 1949, do Centro Brasileiro de Pesquisas Físicas, dirigido por Cesar Lattes, que irá realizar pesquisas sobre raios cósmicos na Bolívia e estudos pioneiros sobre o méson π artificial. **Carneiro (2005)**

Tipos de pesquisa

- **Exploratória** A pesquisa exploratória procura explorar um problema, de modo a fornecer informações para uma investigação mais precisa. Elas visam uma maior proximidade com o tema, que pode ser construído com base em hipóteses ou intuições. É onde pesquisadores tentam explicar o que está acontecendo.
 - **Objetivo** Descobrir ideias e pensamentos.
 - **Processo** Não estruturado

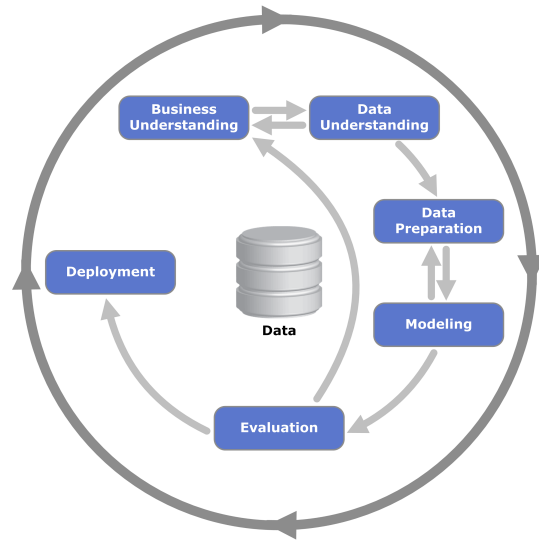


Figure 1: O ciclo de vida da mineração de dados. Fonte: IBM (2015, pág. 1)

- **Dados Qualitativo**
- **Coleta de dados** Pesquisas bibliográficas e estudos de caso são muito utilizados nas pesquisas exploratórias.
- **Descritiva** A pesquisa descritiva visa descrever algo. Para isso, é feita uma análise minuciosa e descritiva do objeto de estudo. Essa pesquisa não pode ter interferência do pesquisador.
 - **Objetivo** Descrever características e funções.
 - **Processo** Estruturado
 - **Dados Quantitativo**
 - **Coleta de dados** Apesar de também investirem na coleta e no levantamento de dados qualitativos, utiliza-se principalmente dados quantitativos.
- **Explicativa** A pesquisa explicativa é uma tentativa de conectar as ideias e fatores identificados para compreender as causas e efeitos de determinado fenômeno.
 - **Objetivo** Compreender causas e efeitos.
 - **Processo** Estruturado
 - **Dados Quantitativo**
 - **Coleta de dados** Baseada em métodos experimentais.

2 CRISP-DM

O atual projeto adota como Metodologia o CRISP-DM (Processo Padrão de Vários Segmentos de Mercados para Mineração de Dados), que é um modelo e metodologia que busca orientar os esforços da organização descrevendo as fases típicas e tarefas envolvidas do ciclo de vida em um projeto de mineração de dados (IBM, 2015). Acompanhe na Figura 1 abaixo o fluxograma padrão que representa uma visão geral das fases do CRISP-DM.

```
knitr::include_graphics("imgs/ciclo.png")
```

Como é possível observar, O CRISP-DM é um modelo composto por seis fases, estas, estão descritas abaixo:

- Business Understanding - Entendimento de Negócios: Trata-se da fase inicial do projeto, onde se investiga as metas de negócios da organização, reúne-se informações básicas e define critérios para o sucesso.
- Data Understanding - Entendimento dos Dados: Etapa de verificação dos dados disponíveis a fim de evitar problemas posteriores. Normalmente, esta é a etapa mais longa de um projeto (IBM, 2015).
- Data Preparation - Preparação de Dados: “Estima-se que a preparação de dados normalmente consome 50-70% do tempo e esforço do projeto”(IBM, 2015, p. 19).
- Modeling - Modelando: Nesta fase, o problema, as metas definidas no Entendimento de Negócios e os dados disponíveis definem as técnicas de modelagem a serem utilizadas. Estima-se que esta etapa seja realizada em várias iterações. Primeiramente, com os hiperparâmetros padrão e, nas iterações seguintes, refinando-os a fim de buscar por resultados cada vez mais satisfatórios.
- Evaluation - Avaliação: Fase onde se analisa e faz inferências sobre os esforços realizado. Para que seja possível cumprir esta etapa de forma objetiva, critérios de avaliação são estabelecidos para os modelos.
- Deployment - Implantação: Esta fase inclui o planejamento e monitoramento da implantação, além de tarefas de finalização, como produção do relatório final e revisão do projeto (IBM, 2015)

3 Execução

3.1 Entendimento de Negócios

3.1.1 Ensino de Matemática (33002010223P9)

O programa é um curso de pós-graduação do IME que se enquadra na modalidade de Mestrado, que confere à CAPES as atribuições de regular a oferta para propostas de tais cursos e avaliá-las. Recebeu nota 4 na avaliação prévia dessa instituição de fomento e teve início em agosto de 2012.

Este programa é um de modalidade profissional, diferente dos outros programas de pós-graduação da USP. Foi iniciado em 2012 e teve as primeiras dissertações defendidas em 2014 e 2015.

A finalidade do programa é a formação continuada em nível de pós-graduação de professores de Matemática do Ensino Básico. De nítido caráter público, o Mestrado Profissional em Ensino de Matemática é gratuito, em conformidade com o Regimento da Pós-Graduação da USP.

Curso	Nota do curso	Créditos em disciplinas	Créditos em trabalhos de conclusão para titulação	Equivalência hora-aula/créditos	Código
M	4	48	48	15	33002010223F1

Discentes	78
Docentes	29
Linhas de pesquisa	1
Disciplinas	15
Projetos de Pesquisa	12

O programa oferece as seguintes áreas de pesquisa:

- Tecnologias de Informática no Ensino de Matemática ou Estatística

- História da Matemática no Ensino Básico
- Transposição didática da Matemática para a sala de aula escolar
- Epistemologia da Matemática e Educação Básica
- Estatística no ensino fundamental e médio

3.1.2 Matemática (33002010005P1)

O Programa de Pós Graduação em Matemática (33002010005P1) do Instituto de de Estatística da USP é constituído por um curso de mestrado, iniciado em 1970. Acompanhe abaixo algumas informações extraídas da plataforma Sucupira:

Curso	Nota do curso	Créditos em disciplinas	Créditos em trabalhos de conclusão para titulação	Equivalência hora-aula/créditos	Código
M	5	48	48	15	33002010005M1
D	5	48	128	15	33002010005D2

Discentes	132
Docentes	61
Linhas de pesquisa	4
Disciplinas	152
Projetos de Pesquisa	89

De acordo com o site do Instituto de Matemática e Estatística (IME) da USP, o Programa está estruturado a partir de suas linhas de pesquisa e projetos e de suas áreas de concentração citadas abaixo. Cada área possui:

- Álgebra,
- Análise e Sistemas Dinâmicos,
- Geometria e Topologia Algébrica,
- Lógica, Conjuntos, Topologia geral e Combinatória.

3.1.3 Matemática Aplicada (33002010006P8)

O programa foi criado em 1970 para mestrado e doutorado. Sua única área de concentração é matemática aplicada. O programa conta com 5 linhas de pesquisa e 17 projetos de pesquisa. O programa permite a integração com outros programas do Instituto de Matemática e Estatística. Seu financiamento é feito por meio do CNPq, CAPES e FAPESP. Ambos os programas obtiveram nota 5 em todas as avaliações desde 1998. Com exceção dos anos de 2013 e 2014, quando obtiveram nota 4.

Curso	Nota do curso	Créditos em disciplinas	Créditos em trabalhos de conclusão para titulação	Equivalência hora-aula/créditos	Código
M	5	48	55	15	33002010006M8
D	5	40	136	15	33002010006D9

Discentes	110
Docentes	35
Linhas de pesquisa	5
Disciplinas	73

Discentes	110
Projetos de Pesquisa	17

O programa oferece as seguintes áreas de pesquisa:

- Equações Diferenciais e Aplicações
 - Dinâmica de Equações de Evolução
 - Dinâmica de Sistemas Hamiltonianos
 - Equações Diferenciais Parciais não Lineares
 - Mecânica Geométrica e Controle
 - Teoria Geométrica de EDPs e várias variáveis complexas
 - Teoria Qualitativa das Equações Diferenciais e suas Aplicações
- Física Matemática
 - Mecânica Estatística Rigorosa: Clássica e Quântica
 - Teoria Clássica e Quântica dos Campos
- Matemática Aplicada Computacional
 - Computação Gráfica
 - Métodos Numéricos e Mecânica dos Fluidos Computacional
 - Métodos Numéricos e Otimização
- Modelagem Matemática e Aplicações
 - Estatística Bayesiana, Otimização Estocástica e Sistemas Esparsos
 - Modelos Matemáticos Aplicados à Epidemiologia
 - Modelos Matemáticos em Genética
 - Modelos Matemáticos para Sistemas Sociais
- Sistemas Dinâmicos
 - Dinâmica e Geometria em Baixas Dimensões
 - Teoria Ergódica: Otimização Ergódica e Formalismo Termodinâmico

3.2 Entendimento dos dados

Os dados utilizados neste projeto foram obtidos com o auxílio da ferramenta e-lattes. Nela, é possível gerar análise dos programas de graduação a partir da lista de pesquisadores envolvidos. Para o atual projeto, foram utilizados três arquivos .csv contendo os nomes e ids dos pesquisadores dos três programas de graduação citados no capítulo anterior (Ensino de Matemática, Matemática e Matemática aplicada, da USP). EM seguida, essas listas foram processadas na plataforma, que retornou os arquivos .json que servirão de insumo para todas as análises realizadas. Os arquivos possuem informações sobre os pesquisadores, as publicações e as orientações. A fim de obtermos uma maior compreensão sobre os dados a serem manipulados e analisados.

3.3 Preparação dos dados

- Importando bibliotecas

```
library(tm)
library(SnowballC)
library(wordcloud)
library(jsonlite) #Importado para lidar com arquivos com extensão JSON
library(listviewer) #Importado para lidar com listas
library(ggplot2) #Importado para realizar visualizações
library(tidyr) #Importado par utilizar funções relacionadas a dataframes
library(dplyr)
library(readxl)
library(stringr)
```

```
library(tidyverse)
library(listviewer)
library(igraph)
#upload de arquivo com funções para transformar listas em Data Frames e objeto igraph
source("elattes.ls2df.R") #Métodos do arquivo "elattes.ls2df.R" também foi utilizado na transformação d
```

- Importando os datasets dos 3 programas de pós graduação

```
perfil_ens <- fromJSON("./Ensino de Matemática (33002010223P9)/profile.json")
public_ens <- fromJSON("./Ensino de Matemática (33002010223P9)/publication.json")
advise_ens <- fromJSON("./Ensino de Matemática (33002010223P9)/advise.json")

perfil_mat <- fromJSON("./Matemática (33002010005P1)/profile.json")
public_mat <- fromJSON("./Matemática (33002010005P1)/publication.json")
advise_mat <- fromJSON("./Matemática (33002010005P1)/advise.json")

perfil_apl <- fromJSON("./MATEMÁTICA APLICADA (33002010006P8)/profile.json")
public_apl <- fromJSON("./MATEMÁTICA APLICADA (33002010006P8)/publication.json")
advise_apl <- fromJSON("./MATEMÁTICA APLICADA (33002010006P8)/advise.json")
```

- Número de pesquisadores

```
print(paste('Ensino de Matemática =', length(perfil_ens)))
```

```
## [1] "Ensino de Matemática = 29"
```

```
print(paste('Matemática =', length(perfil_mat)))
```

```
## [1] "Matemática = 61"
```

```
print(paste('Matemática Aplicada =', length(perfil_apl)))
```

```
## [1] "Matemática Aplicada = 35"
```

- Exemplo de pesquisador do arquivo profile.json

```
names(perfil_ens[["0348490713417429"]])
```

```
## [1] "nome" "resumo_cv"
## [3] "areas_de_atuacao" "endereco_profissional"
## [5] "producao_bibliografica" "orientacoes_academicas"
## [7] "senioridade"
```

- Primeiro nível dos arquivos public.json

```
names(public_ens)
```

```
## [1] "PERIODICO"
## [2] "LIVRO"
## [3] "CAPITULO_DE_LIVRO"
## [4] "TEXTOS_EM_JORNAIS"
## [5] "EVENTO"
## [6] "ARTIGO_ACEITO"
## [7] "DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA"
```

- Primeiro nível dos arquivos advise.json

```
names(advise_ens)
```

```
## [1] "ORIENTACAO_EM_ANDAMENTO_DE_POS_DOUTORADO"
## [2] "ORIENTACAO_EM_ANDAMENTO_DOUTORADO"
## [3] "ORIENTACAO_EM_ANDAMENTO_MESTRADO"
## [4] "ORIENTACAO_EM_ANDAMENTO_GRADUACAO"
## [5] "ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA"
## [6] "ORIENTACAO_CONCLUIDA_POS_DOUTORADO"
## [7] "ORIENTACAO_CONCLUIDA_DOUTORADO"
## [8] "ORIENTACAO_CONCLUIDA_MESTRADO"
## [9] "OUTRAS_ORIENTACOES_CONCLUIDAS"
```

- Exemplo de conteúdo no glimpse no ano 2012 de orientação de doutorado em andamento

```
glimpse(advise_apl[["ORIENTACAO_EM_ANDAMENTO_DOUTORADO"]][["2012"]])
```

```
## Observations: 3
## Variables: 13
## $ natureza          <chr> "Tese de doutorado", "Tese de dout...
## $ titulo            <chr> "Integrabilidade em Sistemas Hamil...
## $ ano               <chr> "2012", "2012", "2012"
## $ id_lattes_aluno   <chr> "", "", "8286622946368668"
## $ nome_aluno        <chr> "Julio Cezar de Oliveira Andrade",...
## $ instituicao        <chr> "Instituto de Matemática e Estatís...
## $ curso             <chr> "Matemática Aplicada", "Matemática...
## $ codigo_do_curso   <chr> "90000003", "33020060", "33020060"
## $ bolsa            <chr> "SIM", "SIM", "NAO"
## $ agencia_financiadora <chr> "Conselho Nacional de Desenvolvime...
## $ codigo_agencia_financiadora <chr> "002200000000", "045000000000", ""
## $ nome_orientadores <list> ["Manuel Valentim de Pera Garcia"...
## $ id_lattes_orientadores <list> ["9893531212718568", "65476308390..."
```

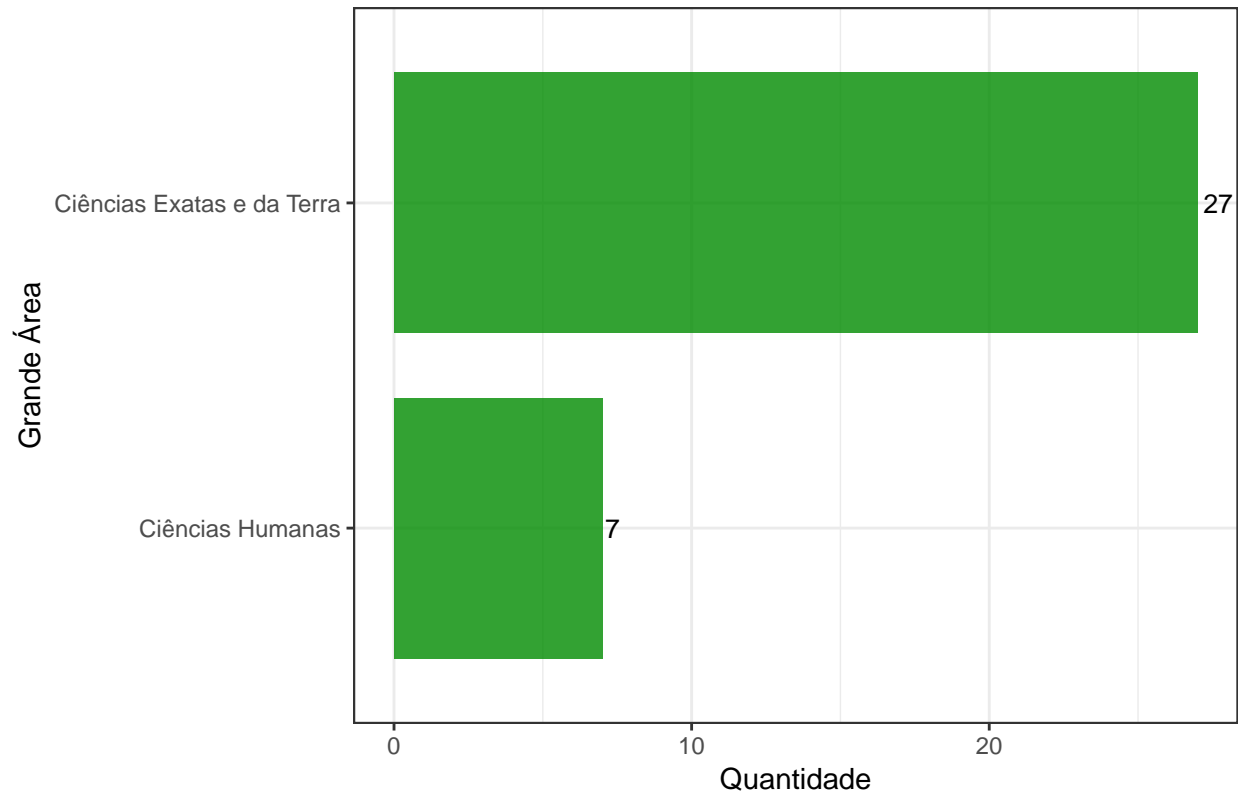
3.4 Avaliação

3.4.1 Pesquisadores por grande área

- Ensino de Matemática

```
perfil_ens %>%
  sapply(function(x) unique(x$areas_de_atuacao$grande_area)) %>%
  unlist() %>% table() %>% sort() %>% as.data.frame() %>% filter(!. == "") %>%
  ggplot(aes(x = ., y = Freq)) + geom_col(fill = "green4",alpha=0.8,width=0.8) + coord_flip() + geom_text(
  labs(title = "Número de Pessoas por Grande Área Atuação", y="Quantidade",x="Grande Área") + theme_bw()
  scale_x_discrete(labels = c('CIENCIAS_DA_SAUDE' = 'Ciências da Saúde',
                              'CIENCIAS_BIOLOGICAS' = 'Ciências Biológicas',
                              'CIENCIAS_HUMANAS' = 'Ciências Humanas',
                              "CIENCIAS_EXATAS_E_DA_TERRA" = "Ciências Exatas e da Terra",
                              "CIENCIAS_SOCIAIS_APLICADAS" = "Ciências Sociais Aplicadas",
                              "CIENCIAS_AGRARIAS" = "Ciências Agrárias",
                              "OUTROS" = "Outros",
                              "ENGENHARIAS" = "Engenharias",
                              "LINGUISTICA_LETRAS_E_ARTES" = "Linguística, Letras e Artes"))
```

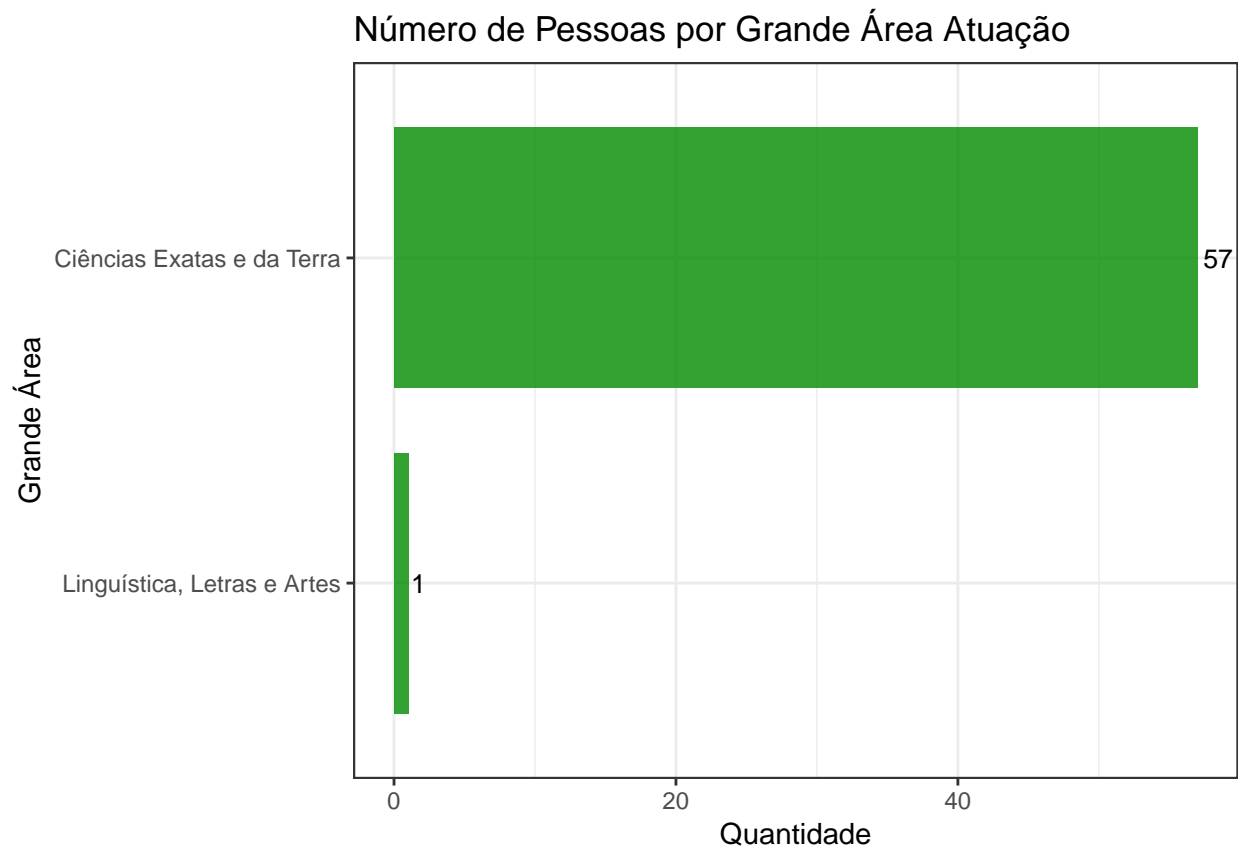

Número de Pessoas por Grande Área Atuação



- Matemática

```

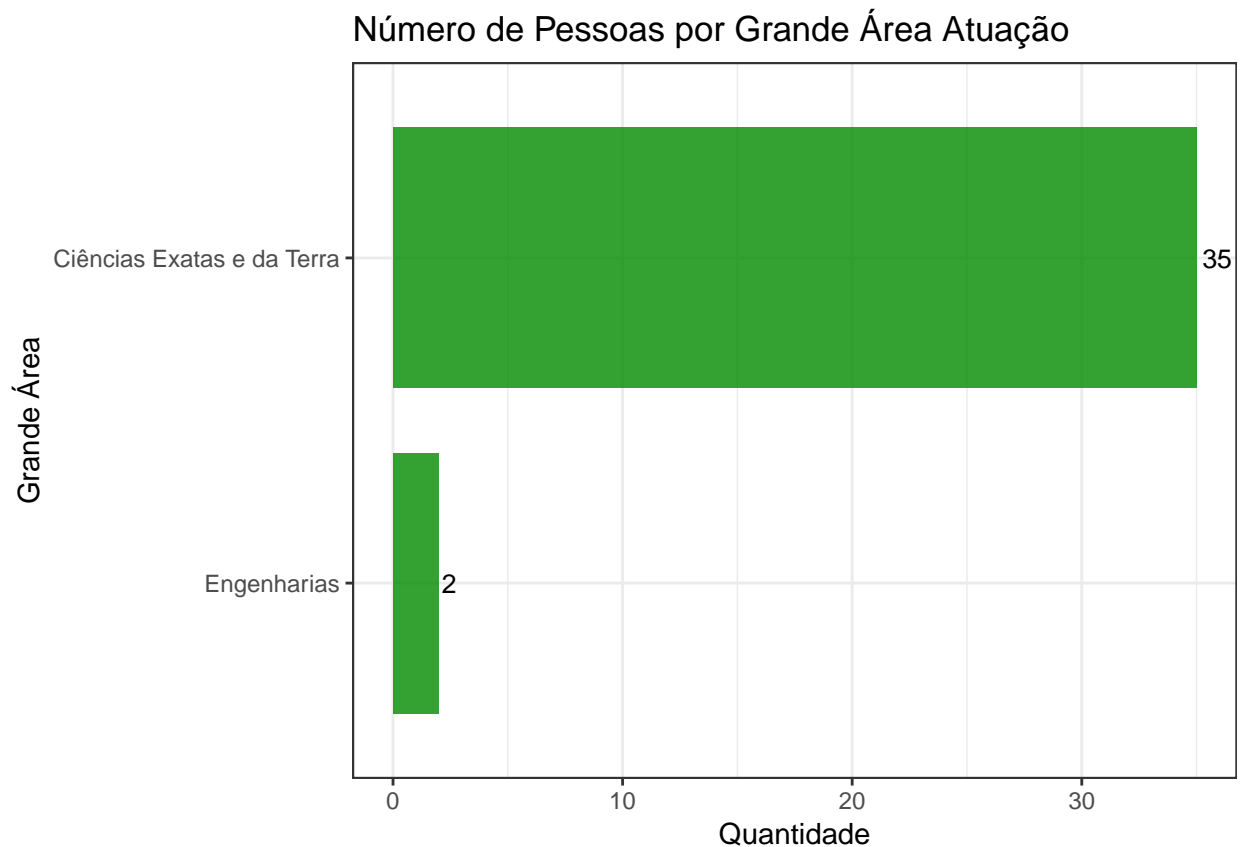
perfil_mat %>%
  sapply(function(x) unique(x$areas_de_atuacao$grande_area)) %>%
  unlist() %>% table() %>% sort() %>% as.data.frame() %>% filter(!. == "") %>%
  ggplot(aes(x = ., y = Freq)) + geom_col(fill = "green4",alpha=0.8,width=0.8) + coord_flip() + geom_text(
  labs(title = "Número de Pessoas por Grande Área Atuação", y="Quantidade",x="Grande Área") + theme_bw(
  scale_x_discrete(labels = c('CIENCIAS_DA_SAUDE' = 'Ciências da Saúde',
    'CIENCIAS_BIOLOGICAS' = 'Ciências Biológicas',
    'CIENCIAS_HUMANAS' = 'Ciências Humanas',
    'CIENCIAS_EXATAS_E_DA_TERRA' = "Ciências Exatas e da Terra",
    'CIENCIAS_SOCIAIS_APLICADAS' = "Ciências Sociais Aplicadas",
    'CIENCIAS_AGRARIAS' = "Ciências Agrárias",
    'OUTROS' = "Outros",
    'ENGENHARIAS' = "Engenharias",
    'LINGUISTICA_LETRAS_E_ARTES' = "Linguística, Letras e Artes"))
  
```



- Matemática Aplicada

```

perfil_apl %>%
  sapply(function(x) unique(x$areas_de_atuacao$grande_area)) %>%
  unlist() %>% table() %>% sort() %>% as.data.frame() %>% filter(!. == "") %>%
  ggplot(aes(x = ., y = Freq)) + geom_col(fill = "green4",alpha=0.8,width=0.8) + coord_flip() + geom_text(
  labs(title = "Número de Pessoas por Grande Área Atuação", y="Quantidade",x="Grande Área") + theme_bw()
  scale_x_discrete(labels = c('CIENCIAS_DA_SAUDE' = 'Ciências da Saúde',
                              'CIENCIAS_BIOLOGICAS' = 'Ciências Biológicas',
                              'CIENCIAS_HUMANAS' = 'Ciências Humanas',
                              "CIENCIAS_EXATAS_E_DA_TERRA" = "Ciências Exatas e da Terra",
                              "CIENCIAS_SOCIAIS_APLICADAS" = "Ciências Sociais Aplicadas",
                              "CIENCIAS_AGRARIAS" = "Ciências Agrárias",
                              "OUTROS" = "Outros",
                              "ENGENHARIAS" = "Engenharias",
                              "LINGUISTICA_LETRAS_E_ARTES" = "Linguística, Letras e Artes"))
  
```



A quantidade de pessoas por grande área de atuação nos mostra uma grande concentração de pessoas em Ciências Exatas e da Terra como era de se esperar ao avaliar os cursos de matemática: Ensino de Matemática 79,41%, Matemática 98,28% e Matemática Aplicada 94,59%. Pelo fato do programa de Ensino de Matemática ter uma parte voltada ao ensino, ele tem uma concentração maior que os outros programas em Ciências Humanas 20,59%. E o programa de Matemática Aplicada tem um contato, ainda que pequeno, com Engenharias 5,41%.

3.4.2 Distribuição dos pesquisadores por área de atuação

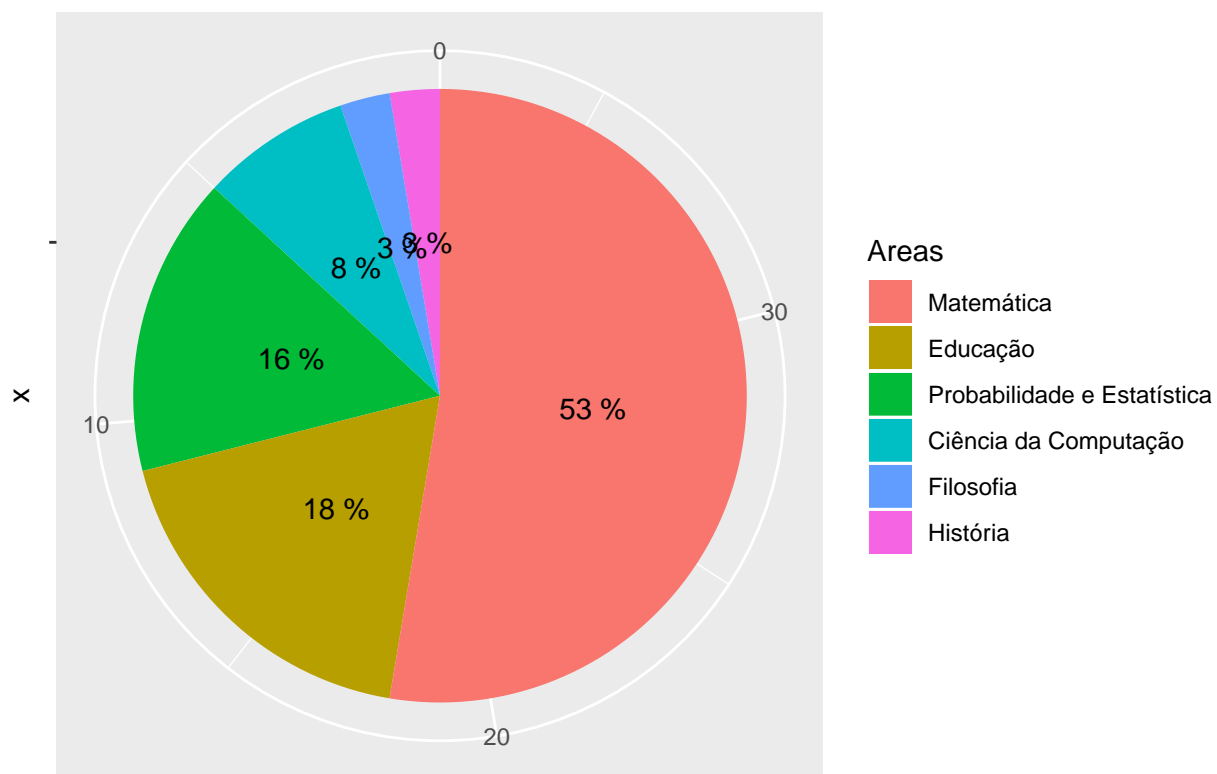
- Ensino de Matemática

```
areas_atuacao_ens <- perfil_ens%>%
  sapply(function(x) unique(x$areas_de_atuacao$area)) %>%
  unlist() %>% table() %>% sort(decreasing = TRUE) %>%
  as.data.frame() %>% filter(!. == "")

quantidade <- sum(areas_atuacao_ens$Freq)
areas_atuacao_ens <- mutate(areas_atuacao_ens, percent = round(areas_atuacao_ens$Freq/quantidade * 100,

colnames(areas_atuacao_ens) <- c("Areas", "Quantidade", "Porcentagem")

ggplot(areas_atuacao_ens, aes(x="", y=Quantidade, fill=Areas))+
  geom_bar(width = 1, stat = "identity") +
  coord_polar("y", start = 0, direction = -1) +
  geom_text(data = areas_atuacao_ens,
            aes(x = "", y=Quantidade, label = paste(Porcentagem, "%")),
            position = position_stack(vjust = 0.5))
```



Quantidade

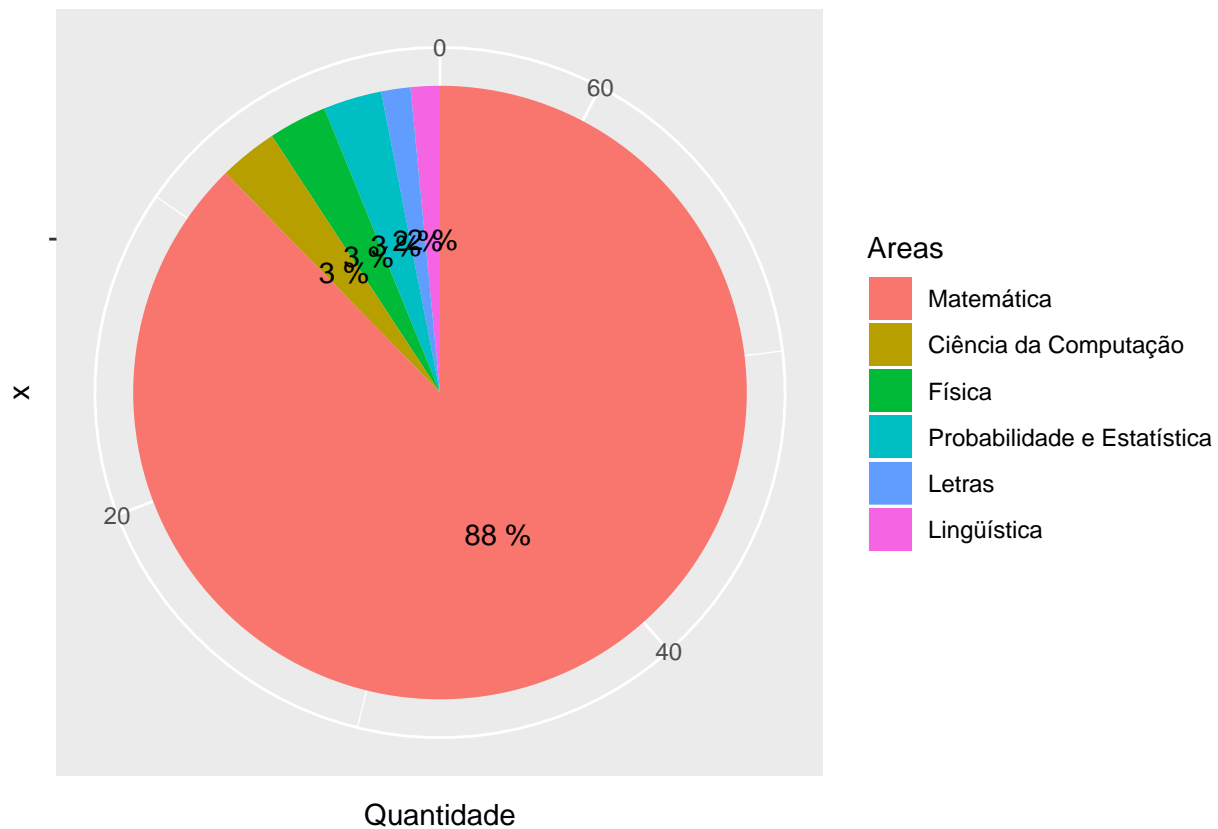
- Matemática

```
areas_atuacao_mat <- perfil_mat%>%
  sapply(function(x) unique(x$areas_de_atuacao$area)) %>%
  unlist() %>% table() %>% sort(decreasing = TRUE) %>%
  as.data.frame() %>% filter(!. == "")

quantidade <- sum(areas_atuacao_mat$Freq)
areas_atuacao_mat <- mutate(areas_atuacao_mat, percent = round(areas_atuacao_mat$Freq/quantidade * 100,

colnames(areas_atuacao_mat) <- c("Areas", "Quantidade", "Porcentagem")

ggplot(areas_atuacao_mat, aes(x="", y=Quantidade, fill=Areas))+
  geom_bar(width = 1, stat = "identity") +
  coord_polar("y", start = 0, direction = -1) +
  geom_text(data = areas_atuacao_mat,
    aes(x = "", y=Quantidade, label = paste(Porcentagem, "%")),
    position = position_stack(vjust = 0.5))
```



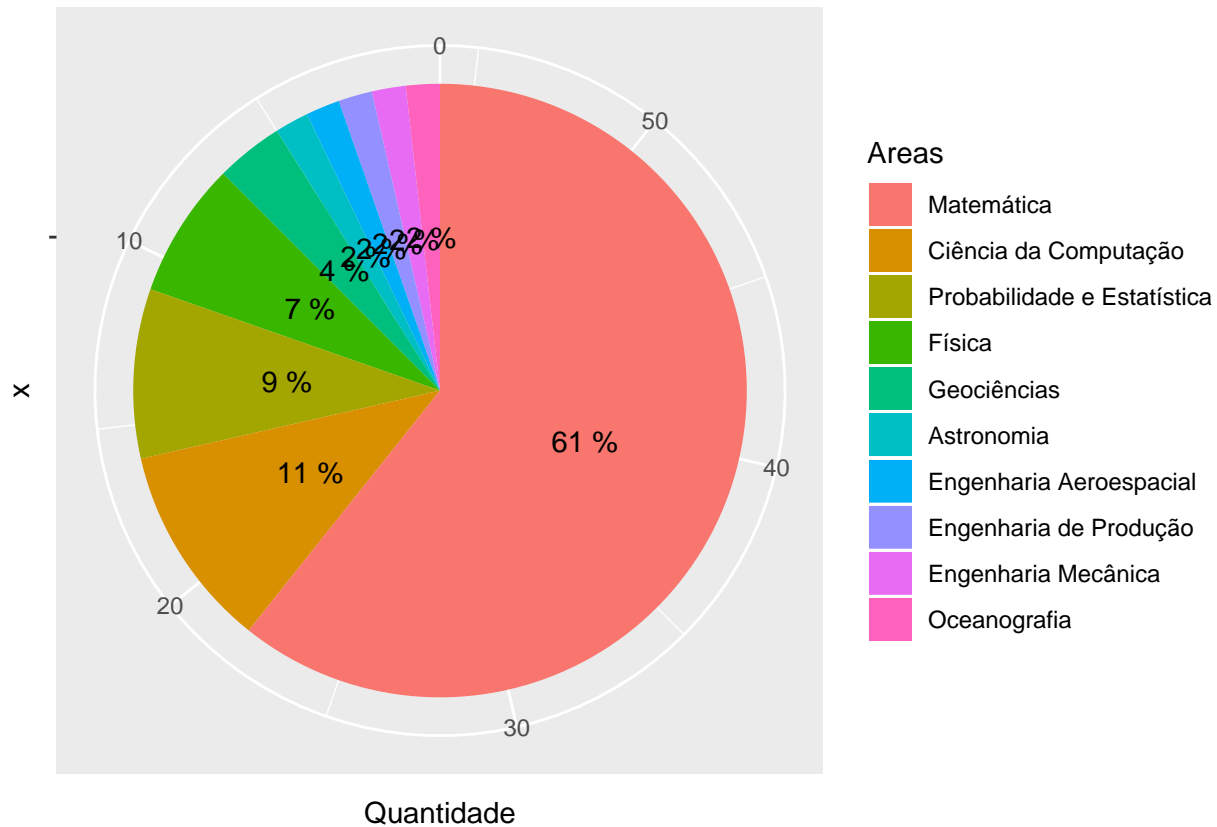
- Matemática Aplicada

```
areas_atuacao_apl <- perfil_apl%>%
  sapply(function(x) unique(x$areas_de_atuacao$area)) %>%
  unlist() %>% table() %>% sort(decreasing = TRUE) %>%
  as.data.frame() %>% filter(!. == "")

quantidade <- sum(areas_atuacao_apl$Freq)
areas_atuacao_apl <- mutate(areas_atuacao_apl, percent = round(areas_atuacao_apl$Freq/quantidade * 100,

colnames(areas_atuacao_apl) <- c("Areas", "Quantidade", "Porcentagem")

ggplot(areas_atuacao_apl, aes(x="", y=Quantidade, fill=Areas))+
  geom_bar(width = 1, stat = "identity") +
  coord_polar("y", start = 0, direction = -1) +
  geom_text(data = areas_atuacao_apl,
    aes(x = "", y=Quantidade, label = paste(Porcentagem, "%")),
    position = position_stack(vjust = 0.5))
```



A quantidade de pesquisadores atuando na área de Educação é discrepante na pós de Ensino de Matemática em comparação com os outros programas analisados. Este resultado é convergente com o previsto, dada a natureza do programa em questão.

3.4.3 Distribuição dos pesquisadores por especialidade (as 7 mais frequentes)

- Ensino de Matemática

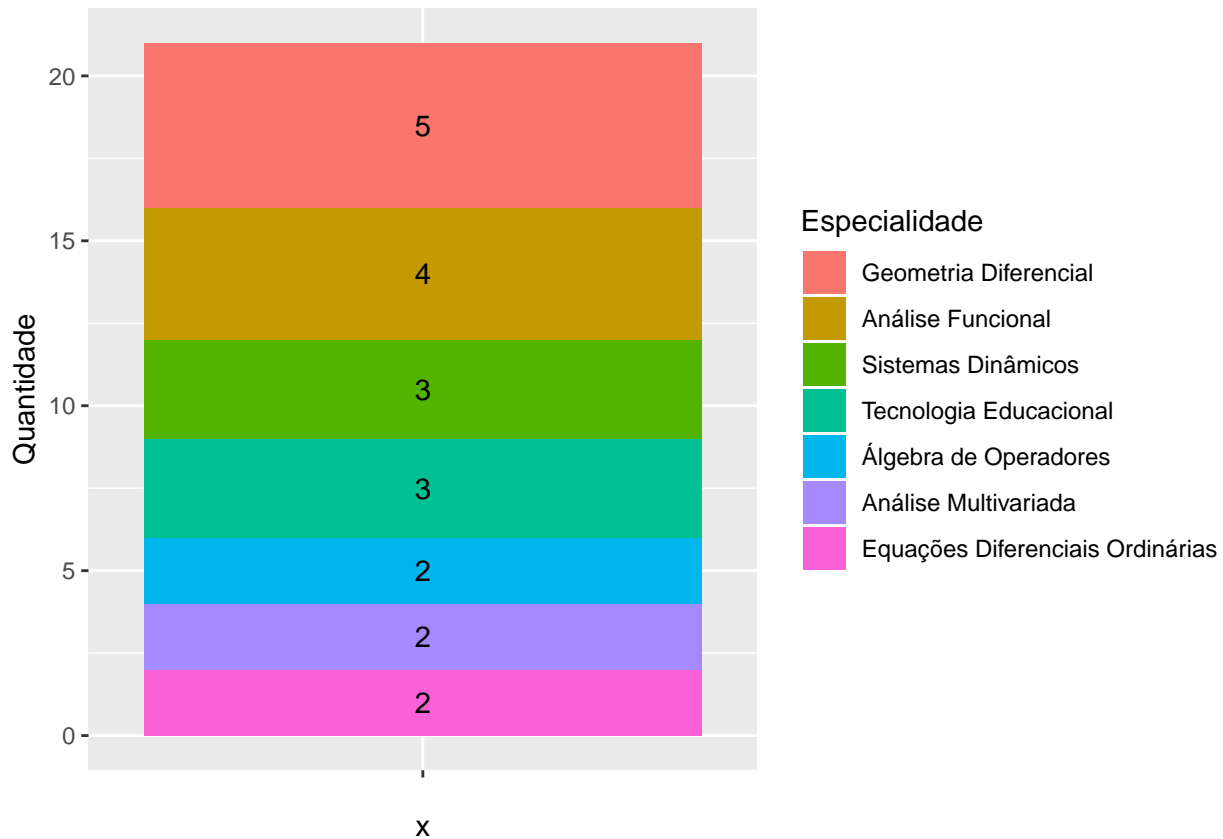
```
especialidades_frequentes <- perfil_ens%>%
  sapply(function(x) unique(x$areas_de_atuacao$especialidade)) %>%
  unlist() %>% table() %>% sort(decreasing = TRUE) %>%
  as.data.frame() %>% filter(!. == "") %>% head(7)

quantidade <- sum(especialidades_frequentes$Freq)

especialidades_frequentes <- mutate(especialidades_frequentes, percent = round(especialidades_frequentes$Freq / quantidade * 100))

colnames(especialidades_frequentes) <- c("Especialidade", "Quantidade", "Porcentagem")

ggplot(especialidades_frequentes, aes(x="", y=Quantidade, fill=Especialidade))+
  geom_bar(width = 1, stat = "identity") +
  #coord_polar("y", start = 0, direction = -1) +
  geom_text(data = especialidades_frequentes,
            aes(x="", y=Quantidade, label = Quantidade),
            position = position_stack(vjust = 0.5))
```



- Matemática

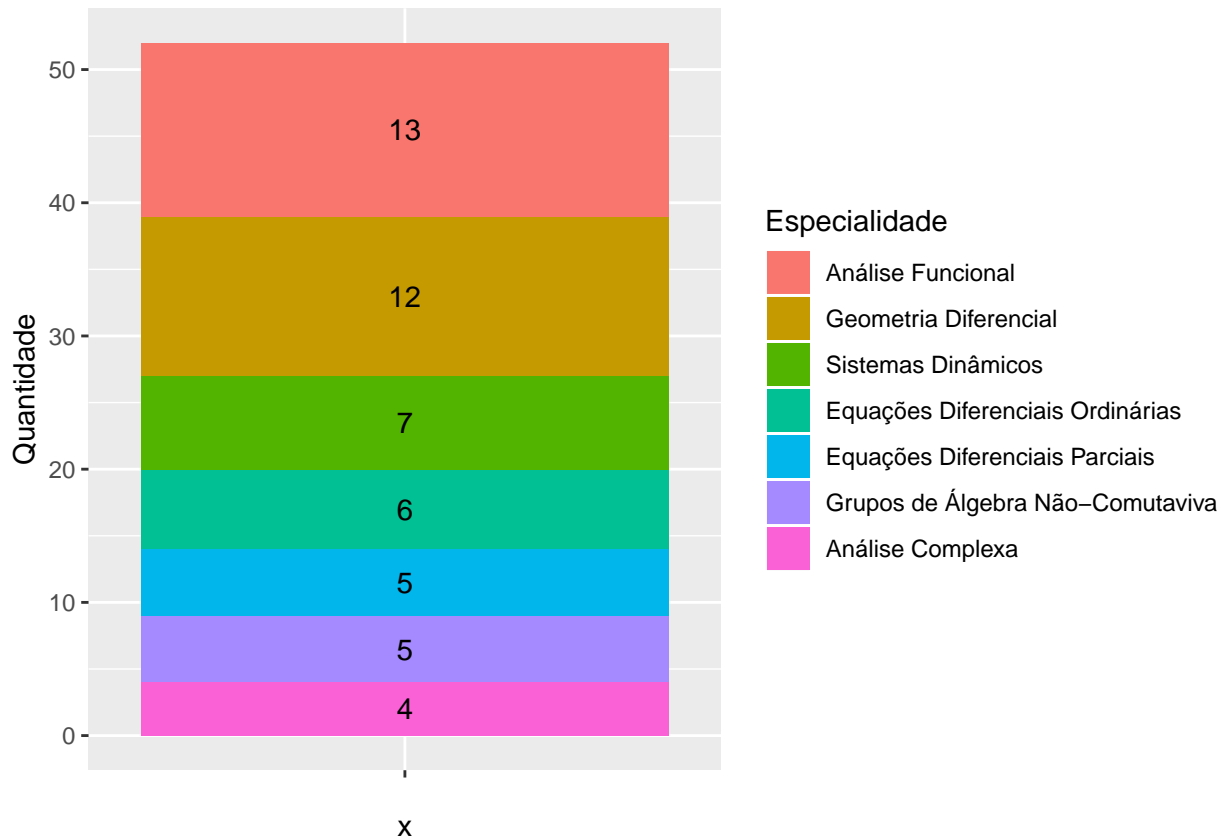
```
especialidades_frequentes <- perfil_mat%>%
  sapply(function(x) unique(x$areas_de_atuacao$especialidade)) %>%
  unlist() %>% table() %>% sort(decreasing = TRUE) %>%
  as.data.frame() %>% filter(!. == "") %>% head(7)

quantidade <- sum(especialidades_frequentes$Freq)

especialidades_frequentes <- mutate(especialidades_frequentes, percent = round(especialidades_frequentes$Freq / quantidade, 2))

colnames(especialidades_frequentes) <- c("Especialidade", "Quantidade", "Porcentagem")

ggplot(especialidades_frequentes, aes(x="", y=Quantidade, fill=Especialidade))+
  geom_bar(width = 1, stat = "identity") +
  #coord_polar("y", start = 0, direction = -1) +
  geom_text(data = especialidades_frequentes,
            aes(x="", y=Quantidade, label = Quantidade),
            position = position_stack(vjust = 0.5))
```



- Matemática Aplicada

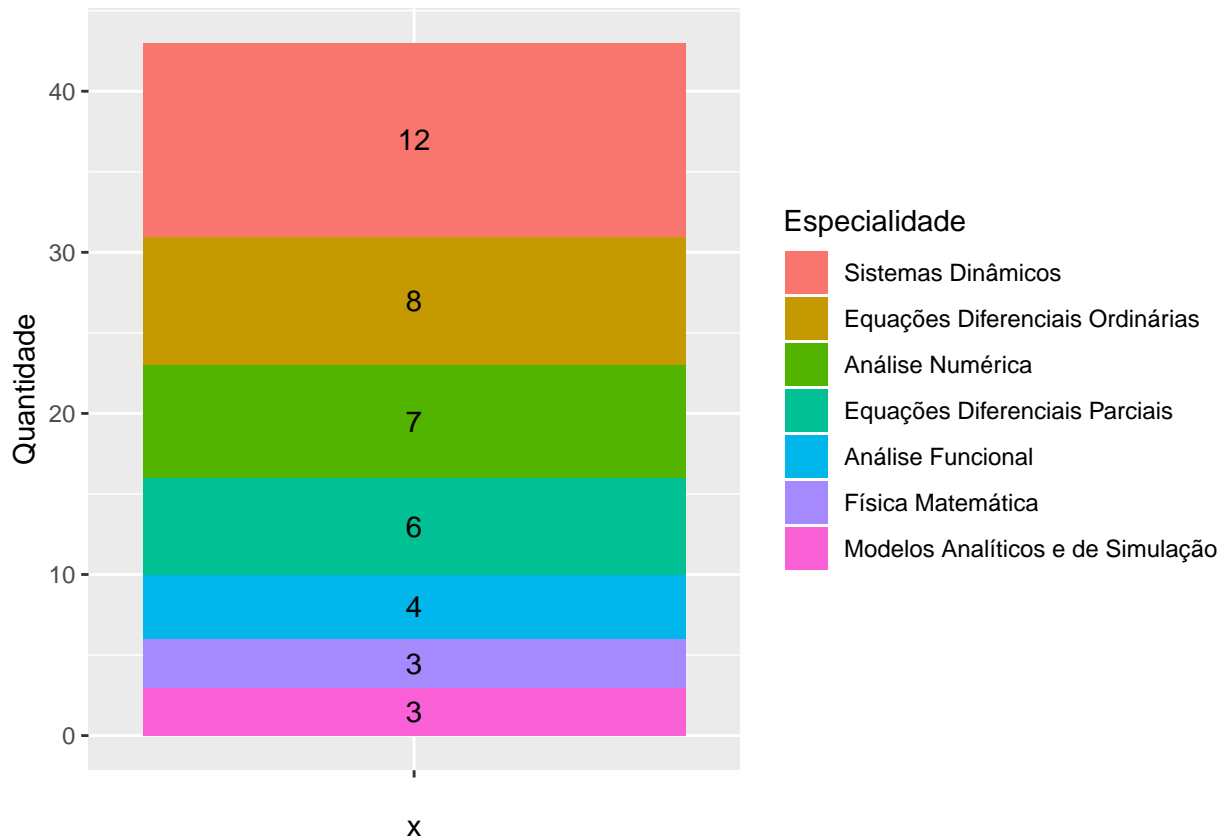
```
especialidades_frequentes <- perfil_apl%>%
  sapply(function(x) unique(x$areas_de_atuacao$especialidade)) %>%
  unlist() %>% table() %>% sort(decreasing = TRUE) %>%
  as.data.frame() %>% filter(!. == "") %>% head(7)

quantidade <- sum(especialidades_frequentes$Freq)

especialidades_frequentes <- mutate(especialidades_frequentes, percent = round(especialidades_frequentes$Freq / quantidade, 2))

colnames(especialidades_frequentes) <- c("Especialidade", "Quantidade", "Porcentagem")

ggplot(especialidades_frequentes, aes(x="", y=Quantidade, fill=Especialidade))+
  geom_bar(width = 1, stat = "identity") +
  #coord_polar("y", start = 0, direction = -1) +
  geom_text(data = especialidades_frequentes,
            aes(x="", y=Quantidade, label = Quantidade),
            position = position_stack(vjust = 0.5))
```

Como foi possível observar as especialidades mais frequentes nos programas de Ensino de Matemática e Matemática são sistemas dinâmicos, equações diferenciais ordinárias e análise funcional. O programa de Matemática aplicada diverge um pouco dos demais nas especialidades mais frequentes. O de Ensino de matemática conta com a especialidade de tecnologia educacional que não aparece nos outros programas.

3.4.4 Subáreas mais frequentes

- Ensino de Matemática

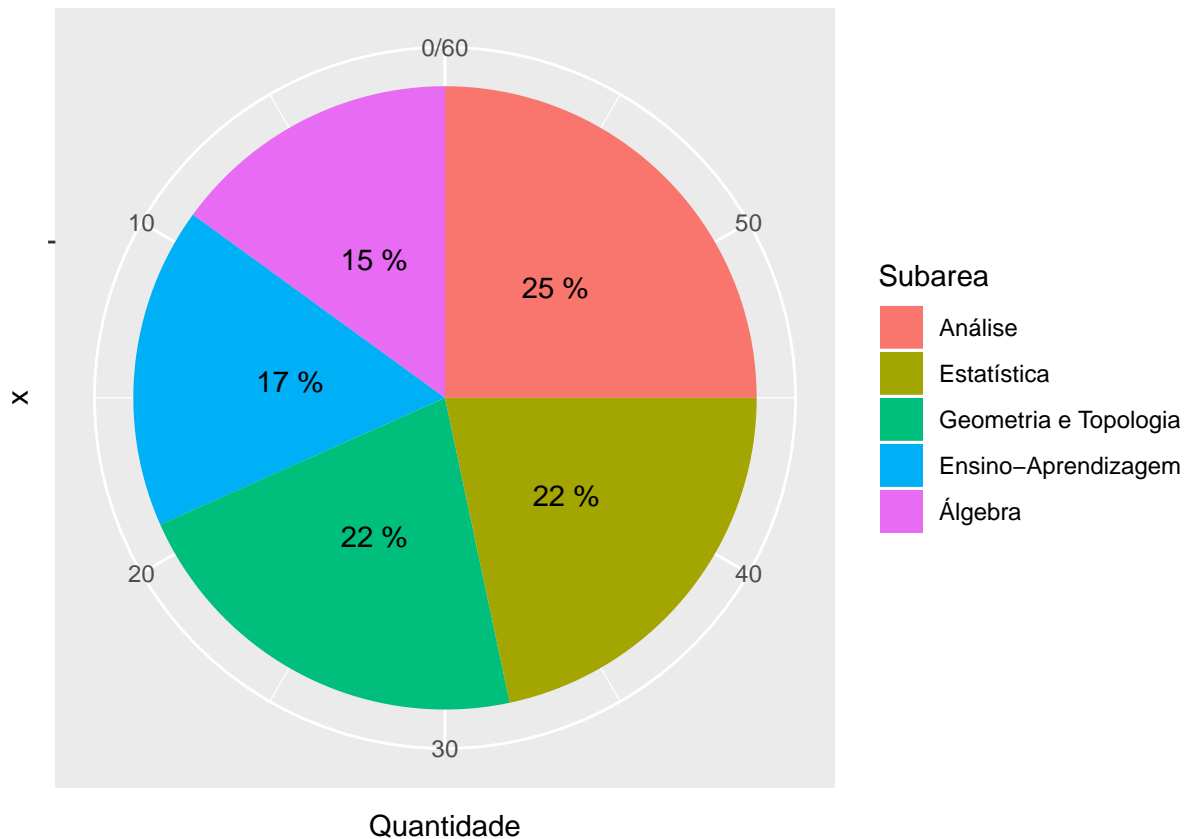
```
subarea <- perfil_ens %>%
  sapply(function(x) (x$areas_de_atuacao$sub_area)) %>%
  unlist() %>% table() %>%
  sort(decreasing = TRUE) %>% as.data.frame() %>%
  filter(!. == "") %>% head(5)

quantidade <- sum(subarea$Freq)

subarea <- mutate(subarea, percent = round(subarea$Freq/quantidade * 100, 0))

colnames(subarea) <- c("Subarea", "Quantidade", "Porcentagem")

ggplot(subarea, aes(x="", y=Quantidade, fill=Subarea))+
  geom_bar(width = 1, stat = "identity") +
  coord_polar("y", start = 0, direction = -1) +
  geom_text(data = subarea,
            aes(x="", y=Quantidade, label = paste(Porcentagem, "%")),
            position = position_stack(vjust = 0.5))
```



- Matemática

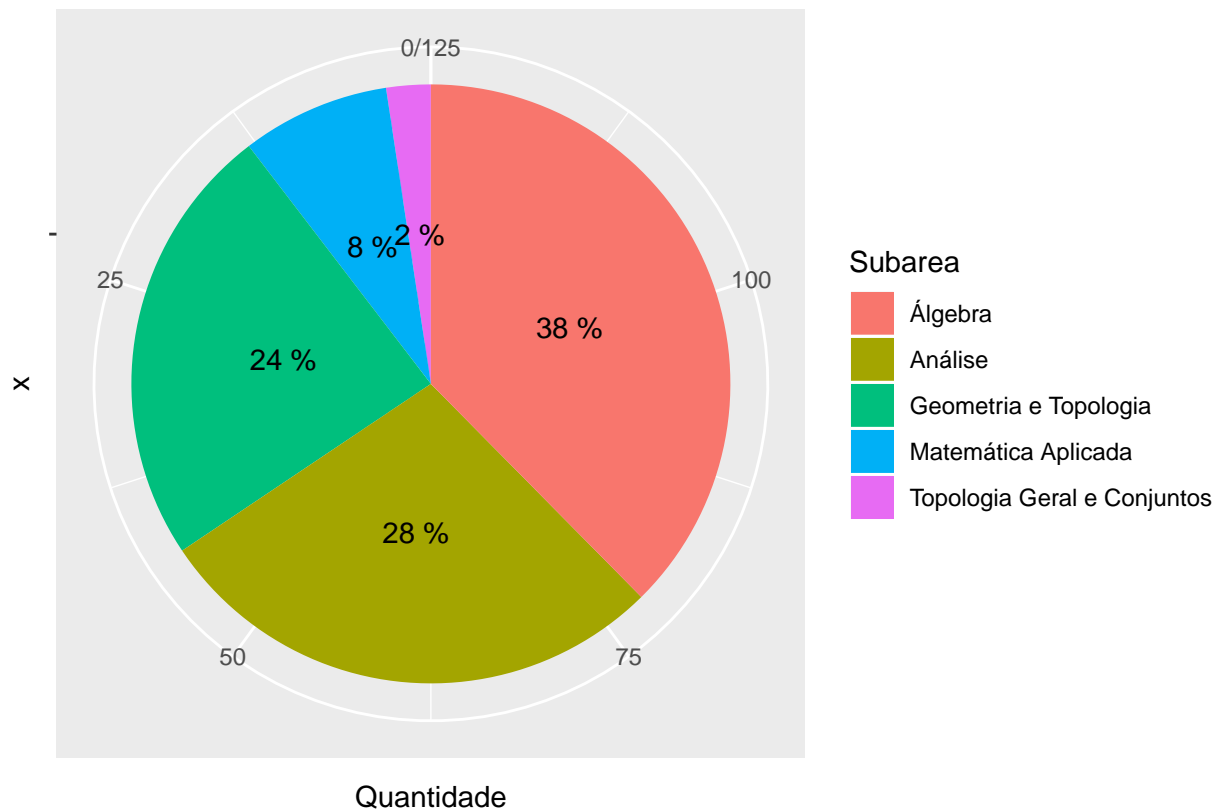
```
subarea <- perfil_mat %>%
  sapply(function(x) (x$areas_de_atuacao$sub_area)) %>%
  unlist() %>% table() %>%
  sort(decreasing = TRUE) %>% as.data.frame() %>%
  filter(!. == "") %>% head(5)

quantidade <- sum(subarea$Freq)

subarea <- mutate(subarea, percent = round(subarea$Freq/quantidade * 100, 0))

colnames(subarea) <- c("Subarea", "Quantidade", "Porcentagem")

ggplot(subarea, aes(x="", y=Quantidade, fill=Subarea))+
  geom_bar(width = 1, stat = "identity") +
  coord_polar("y", start = 0, direction = -1) +
  geom_text(data = subarea,
    aes(x = "", y=Quantidade, label = paste(Porcentagem, "%")),
    position = position_stack(vjust = 0.5))
```



- Matemática Aplicada

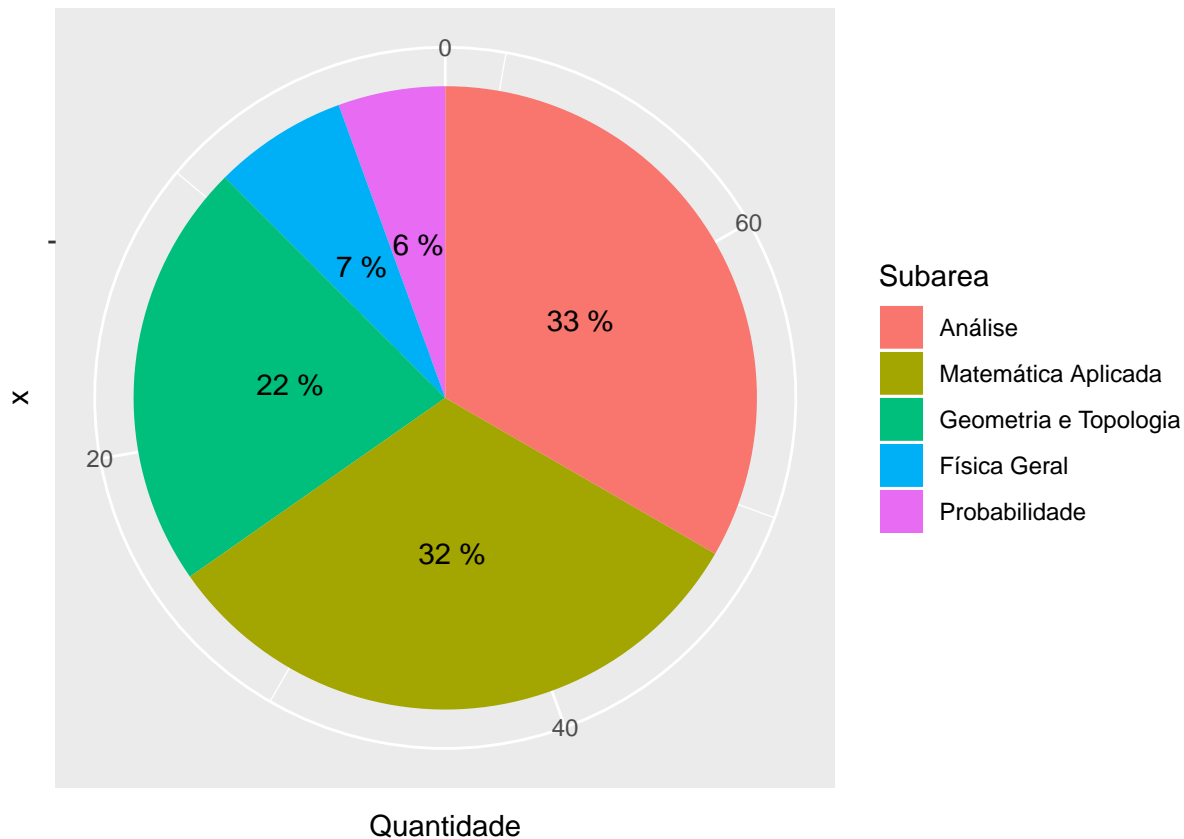
```
subarea <- perfil_apl %>%
  sapply(function(x) (x$areas_de_atuacao$sub_area)) %>%
  unlist() %>% table() %>%
  sort(decreasing = TRUE) %>% as.data.frame() %>%
  filter(!. == "") %>% head(5)

quantidade <- sum(subarea$Freq)

subarea <- mutate(subarea, percent = round(subarea$Freq/quantidade * 100, 0))

colnames(subarea) <- c("Subarea", "Quantidade", "Porcentagem")

ggplot(subarea, aes(x="", y=Quantidade, fill=Subarea))+
  geom_bar(width = 1, stat = "identity") +
  coord_polar("y", start = 0, direction = -1) +
  geom_text(data = subarea,
    aes(x="", y=Quantidade, label = paste(Porcentagem, "%")),
    position = position_stack(vjust = 0.5))
```



Ao analisar as subáreas mais frequentes dos programas de graduação, é possível notar a presença constante da subárea “Análise”, sendo a mais frequente em dois dos três programas. Observando as principais diferenças entre os programas, Ensino de Matemática é o único que apresenta a “Ensino-Aprendizagem”, enquanto Matemática aplicada é o único com “Física Geral”.

3.4.5 Total de publicações em Periódicos

- Ensino de Matemática

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

```
## [1] 201
```

- Matemática

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

```
## [1] 726
```

- Matemática Aplicada

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

```
## [1] 303
```

Considerando que o programa Ensino de Matemática não se trata de um mestrado academico, já é esperado que o número de suas publicações tenha menor comparado aos outros programas. Ainda existe uma grande discrepância entre os artigos publicados pelo programa Matemática e Matemática aplicada. Porém a natureza do programa Matemática Aplicada está em desenvolvimento não necessário academico para artigos.

3.4.6 Cinco revistas mais publicadas

- Ensino de Matemática

```
head(sort(table(as.data.frame(unlist
(sapply(public_ens$PERIODICO, function(x) unlist(x$periodico)))
)), decreasing = TRUE),5)
```

```
##
##      Revista do Professor de Matemática
##                                17
##      Communications in Algebra
##                                6
##      Journal of Algebra (Print)
##                                6
##      Educação Matemática Pesquisa (Online)
##                                4
##      Journal of Algebra and its Applications
##                                4
```

- Matemática

```
head(sort(table(as.data.frame(unlist
(sapply(public_mat$PERIODICO, function(x) unlist(x$periodico)))
)), decreasing = TRUE),5)
```

```
##
##      Journal of Algebra (Print)
##                                32
##      Communications in Algebra
##                                22
##      Journal of Mathematical Analysis and Applications (Print)
##                                19
##      Proceedings of the American Mathematical Society
##                                18
##      Topology and its Applications
##                                15
```

- Matemática Aplicada

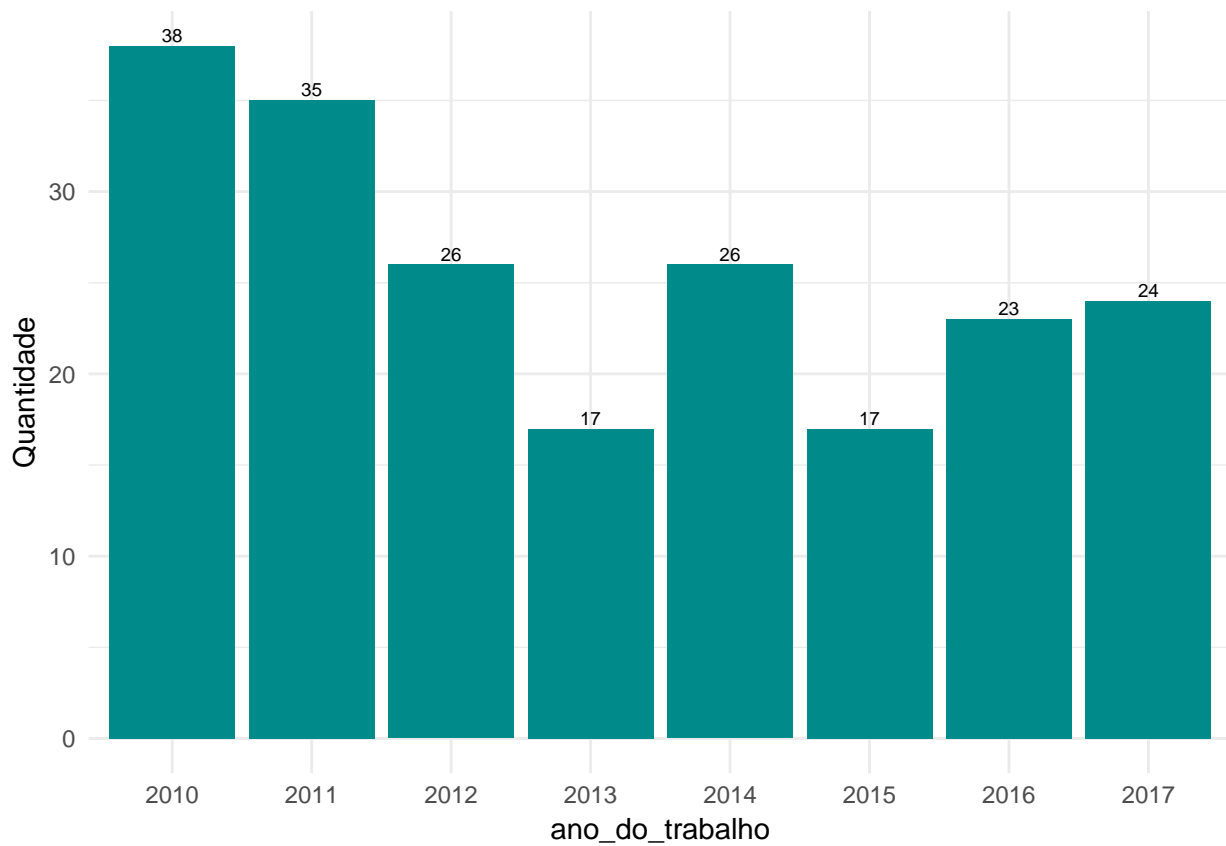
```
head(sort(table(as.data.frame(unlist
(sapply(public_apl$PERIODICO, function(x) unlist(x$periodico)))
)), decreasing = TRUE),5)
```

```
##
##      Qualitative Theory of Dynamical Systems
##                                8
##      Computational Optimization and Applications
##                                7
##      Journal of Mathematical Physics
##                                7
##      AIP Conference Proceedings
##                                6
##      Discrete and Continuous Dynamical Systems
##                                5
```

3.4.7 Número de eventos por ano

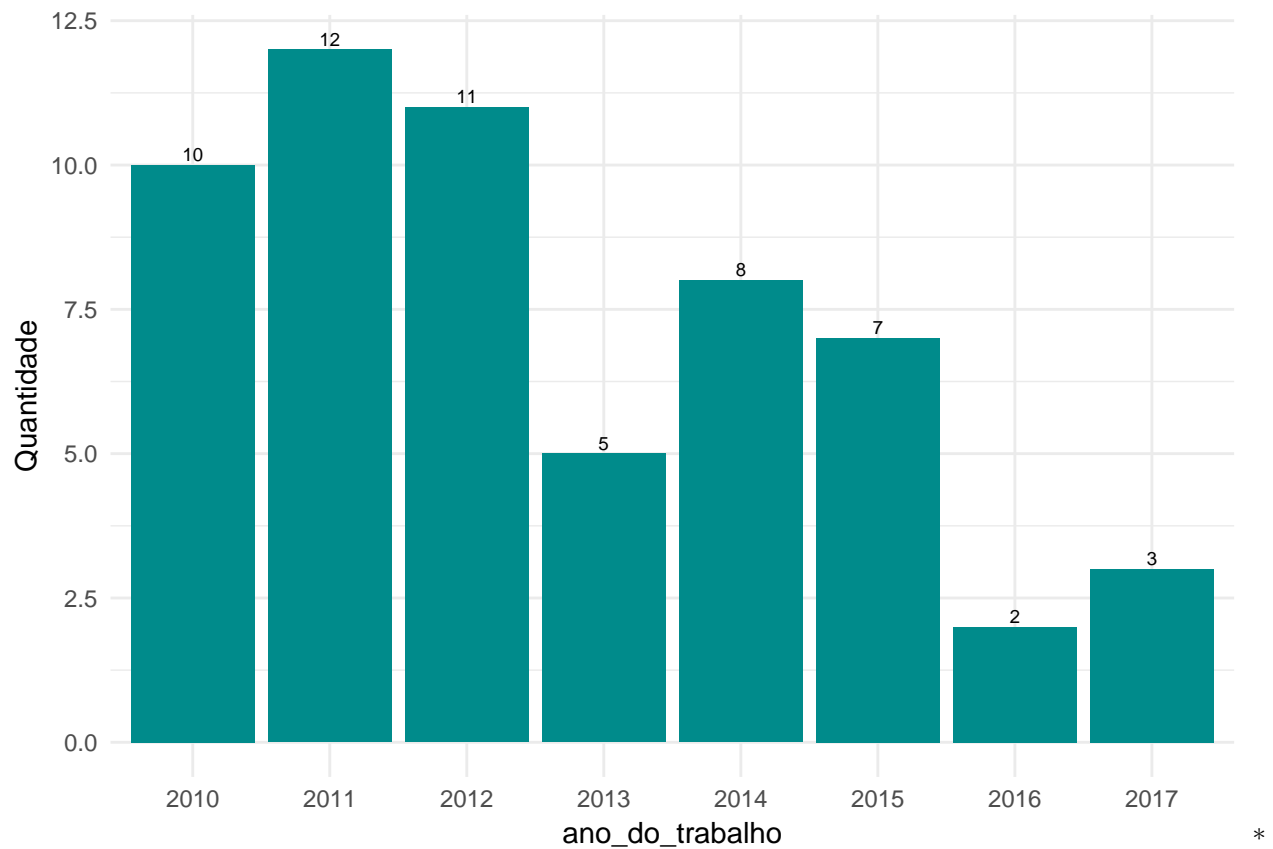
- Ensino de Matemática

```
public.eventos.df <- pub.ls2df(public_ens, 5)
public.eventos.df %>%
  group_by(ano_do_trabalho) %>%
  summarise(Quantidade = n()) %>%
  ggplot(aes(x = ano_do_trabalho, y = Quantidade)) +
  geom_bar(position = "stack", stat = "identity", fill = "darkcyan") +
  geom_text(aes(label=Quantidade), vjust=-0.3, size=2.5) +
  theme_minimal()
```



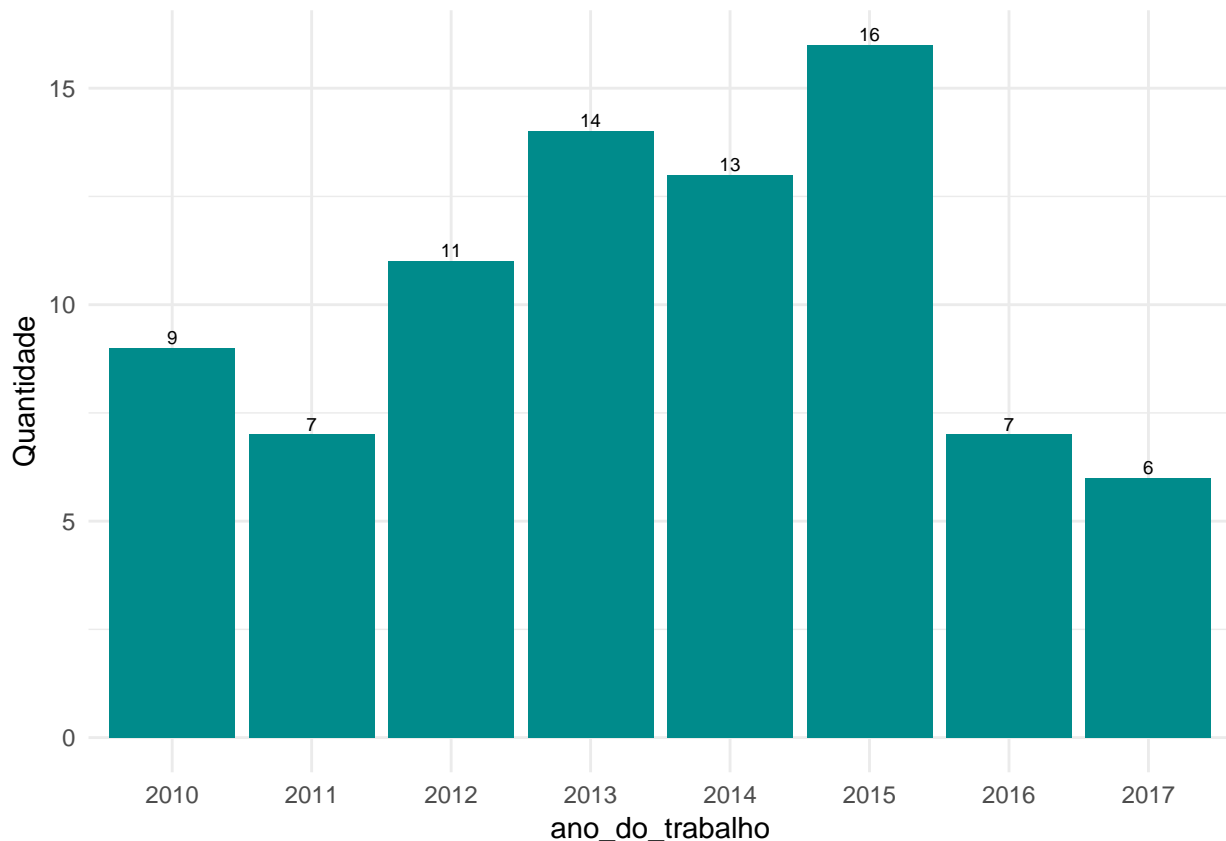
Matemática

```
public.eventos.df <- pub.ls2df(public_mat, 5)
public.eventos.df %>%
  group_by(ano_do_trabalho) %>%
  summarise(Quantidade = n()) %>%
  ggplot(aes(x = ano_do_trabalho, y = Quantidade)) +
  geom_bar(position = "stack", stat = "identity", fill = "darkcyan") +
  geom_text(aes(label=Quantidade), vjust=-0.3, size=2.5) +
  theme_minimal()
```



Matemática Aplicada

```
public.eventos.df <- pub.ls2df(public_apl, 5)
public.eventos.df %>%
  group_by(ano_do_trabalho) %>%
  summarise(Quantidade = n()) %>%
  ggplot(aes(x = ano_do_trabalho, y = Quantidade)) +
  geom_bar(position = "stack", stat = "identity", fill = "darkcyan") +
  geom_text(aes(label=Quantidade), vjust=-0.3, size=2.5) +
  theme_minimal()
```

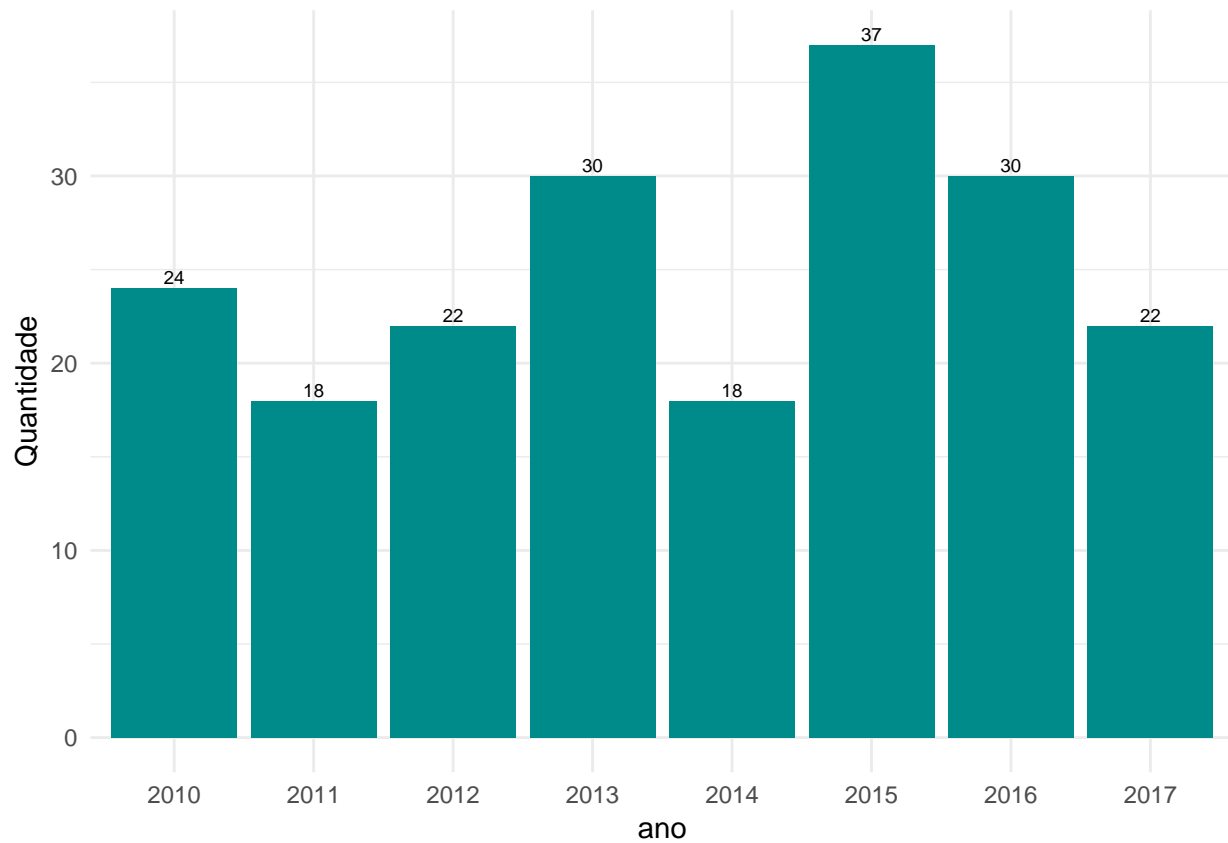


Fazendo uma comparação geral, nota-se que o programa de Ensino de Matemática possui um número geral maior de eventos em relação aos outros programas, mesmo não sendo o que possui mais pesquisadores. No ano em que há o menor número de eventos, 2013, quando foram realizados 17 eventos, ainda assim é uma quantidade maior do que os anos de mais eventos dos outros programas. Portanto, podemos concluir a partir desses gráficos que este programa enfatiza a realização de eventos.

3.4.8 Número de publicações por ano

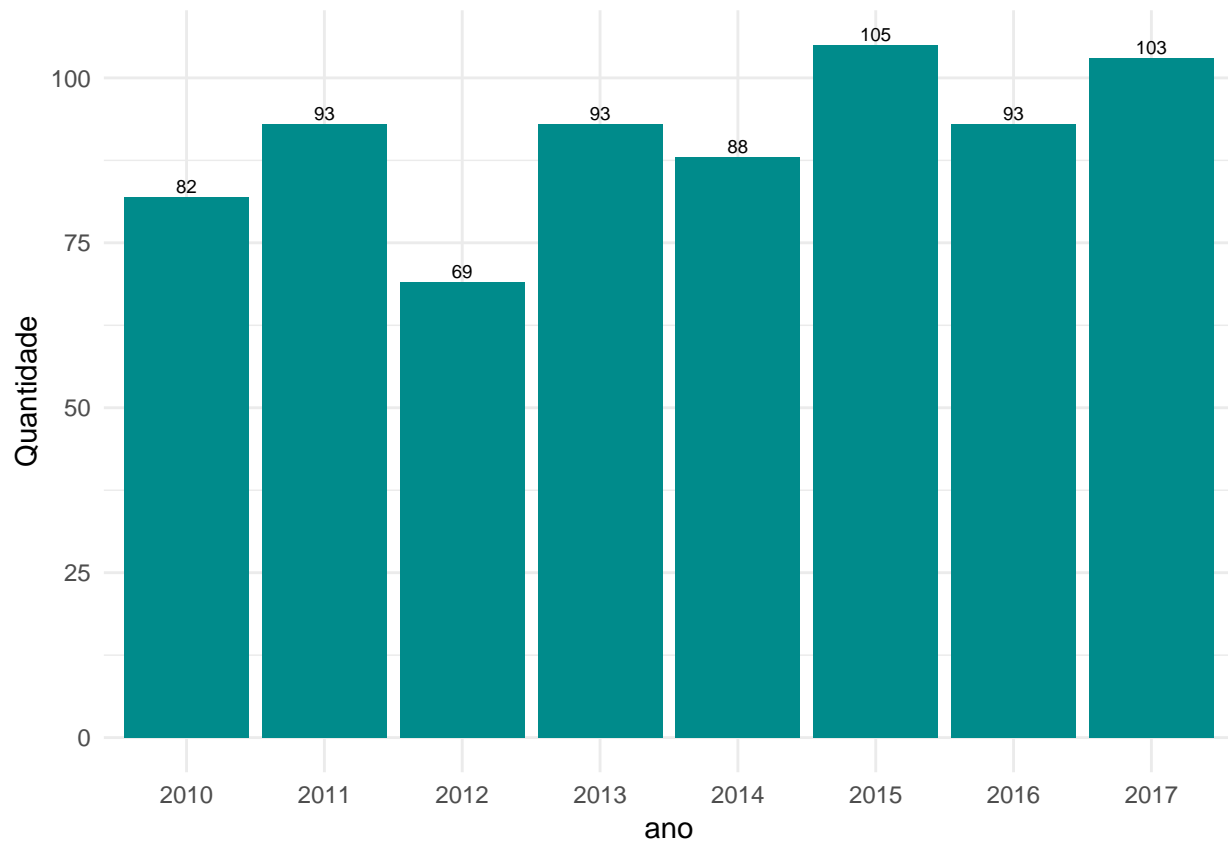
- Ensino de Matemática

```
public.periodico.df <- pub.ls2df(public_ens, 1)
public.periodico.df %>%
  group_by(ano) %>%
  summarise(Quantidade = n()) %>%
  ggplot(aes(x = ano, y = Quantidade)) +
  geom_bar(position = "stack", stat = "identity", fill = "darkcyan") +
  geom_text(aes(label=Quantidade), vjust=-0.3, size=2.5) +
  theme_minimal()
```

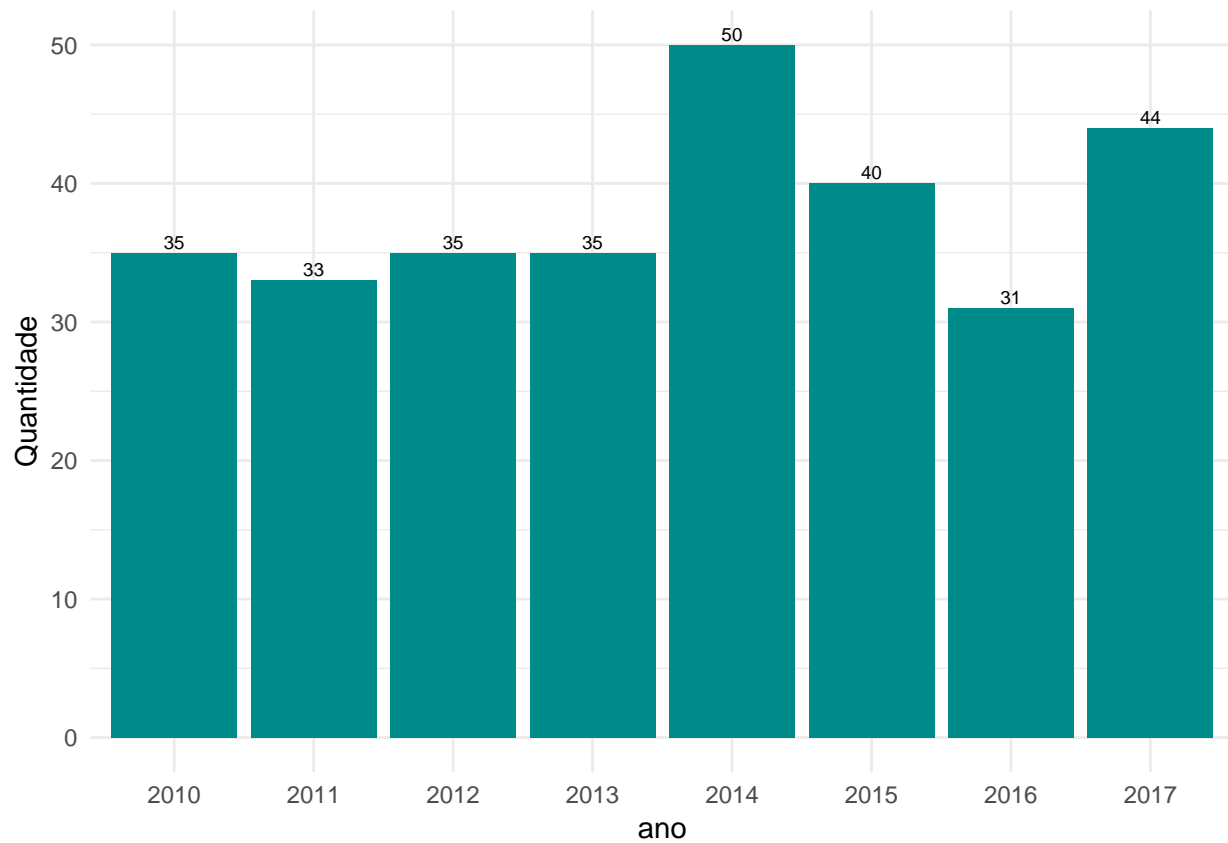
- Matemática

```
public.periodico.df <- pub.ls2df(public_mat, 1)
public.periodico.df %>%
  group_by(ano) %>%
  summarise(Quantidade = n()) %>%
  ggplot(aes(x = ano, y = Quantidade)) +
  geom_bar(position = "stack", stat = "identity", fill = "darkcyan") +
  geom_text(aes(label=Quantidade), vjust=-0.3, size=2.5) +
  theme_minimal()
```



- Matemática Aplicada

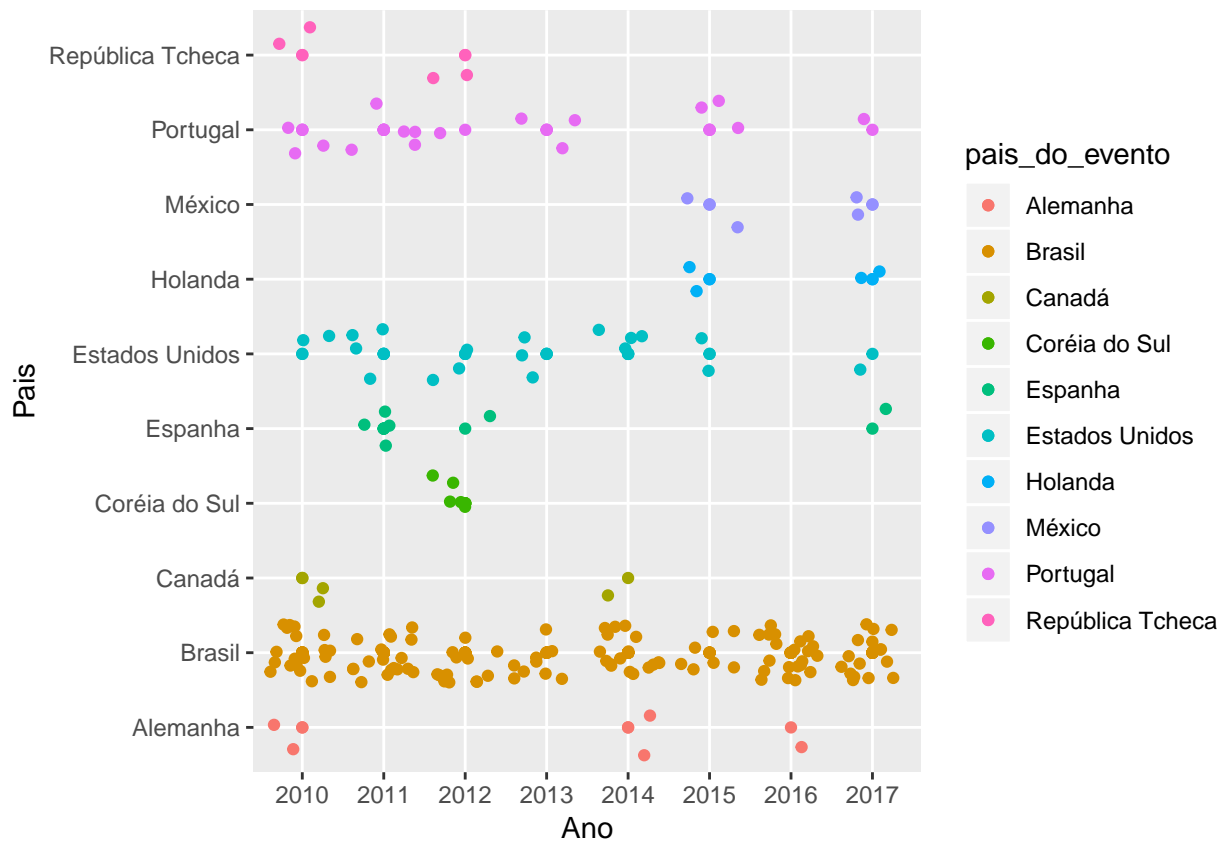
```
public.periodico.df <- pub.ls2df(public_apl, 1)
public.periodico.df %>%
  group_by(ano) %>%
  summarise(Quantidade = n()) %>%
  ggplot(aes(x = ano, y = Quantidade)) +
  geom_bar(position = "stack", stat = "identity", fill = "darkcyan") +
  geom_text(aes(label=Quantidade), vjust=-0.3, size=2.5) +
  theme_minimal()
```



3.4.9 Eventos por país por ano

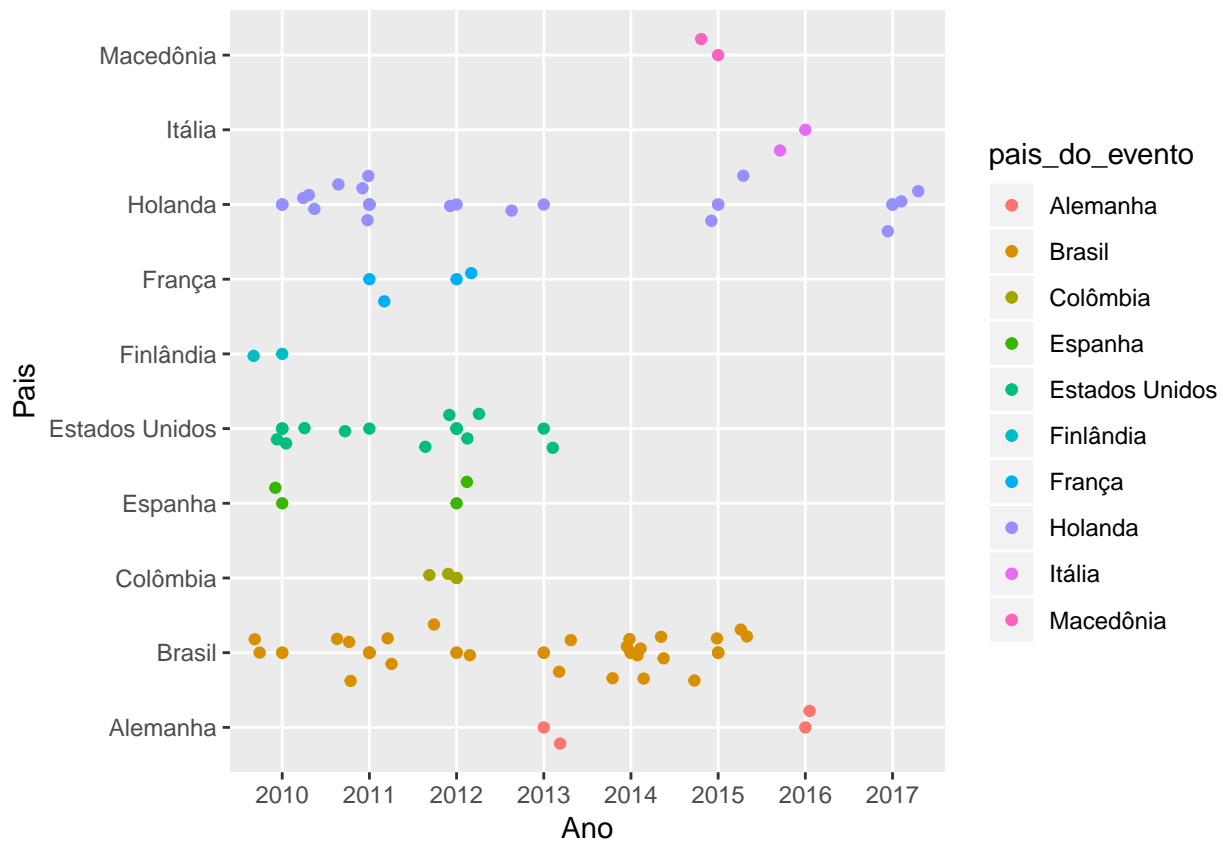
- Ensino de Matemática

```
public.eventos.df <- pub.ls2df(public_ens, 5)
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)) +
  xlab("Ano") + ylab("Pais") + geom_point() + geom_jitter()
```



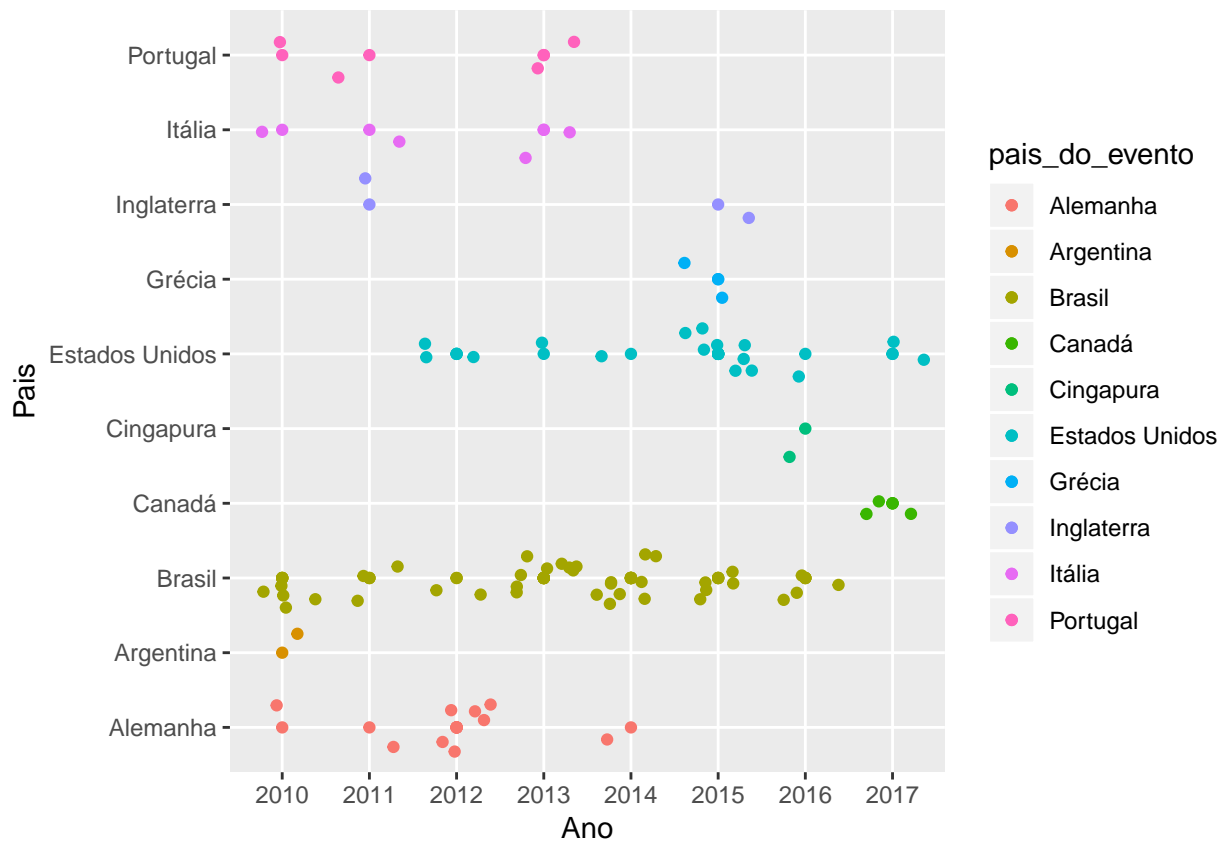
- Matemática

```
public.eventos.df <- pub.ls2df(public_mat, 5)
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)) +
  xlab("Ano") + ylab("Pais") + geom_point() + geom_jitter()
```



- Matemática Aplicada

```
public.eventos.df <- pub.ls2df(public_apl, 5)
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)) +
  xlab("Ano") + ylab("Pais") + geom_point() + geom_jitter()
```

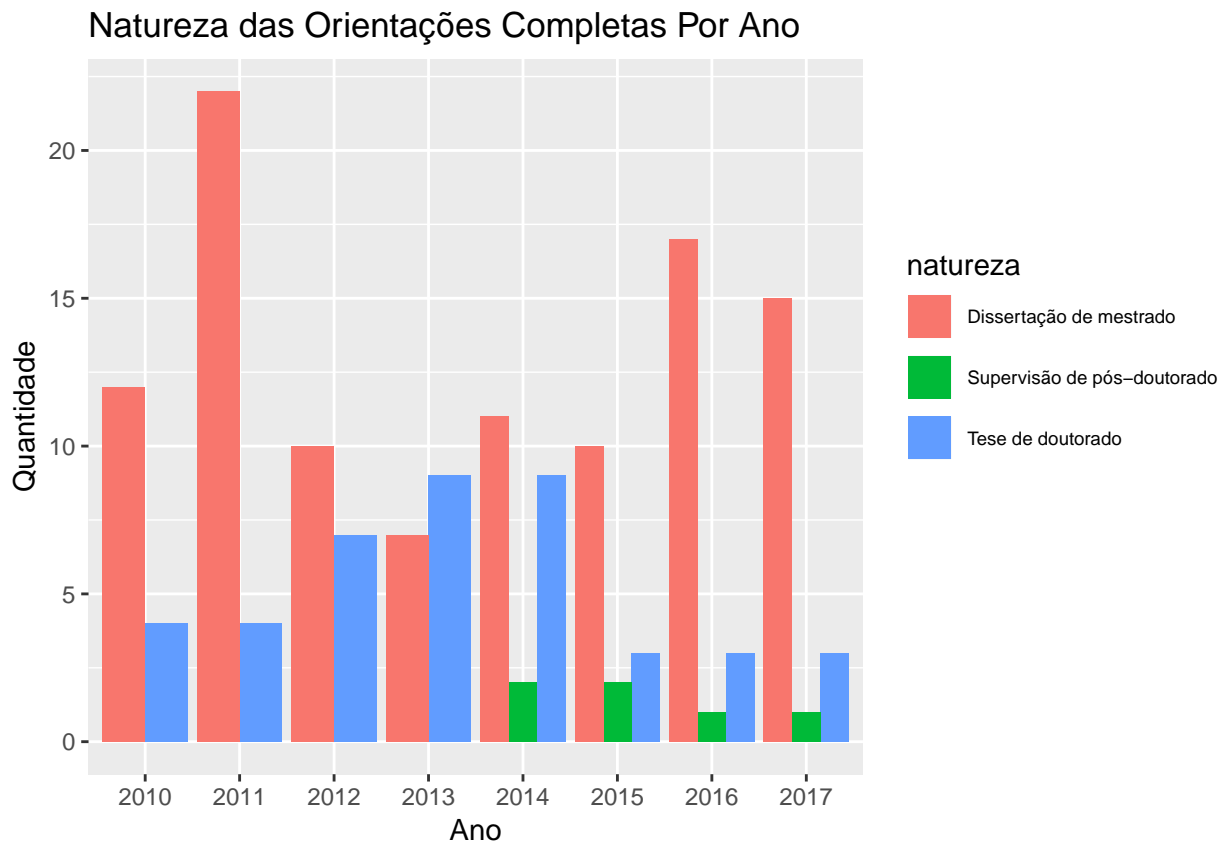


No programa de Ensino de Matemática, nota-se que no ano de 2016, houve uma interrupção nos eventos de todos os países, com exceção do Brasil e Alemanha. Já no programa de Matemática, uma forte diminuição em 2014, e também uma falta de eventos no próprio Brasil a partir de 2016, além da ausência de Portugal, que por conta da língua, está presente nos outros programas. Para Matemática Aplicada, os eventos também sofreram queda, tornando o ano de 2017 sem eventos no Brasil, tendo apenas alguns nos Estados Unidos e Cingapura.

3.4.10 Natureza das orientações por ano

- Ensino de Matemática

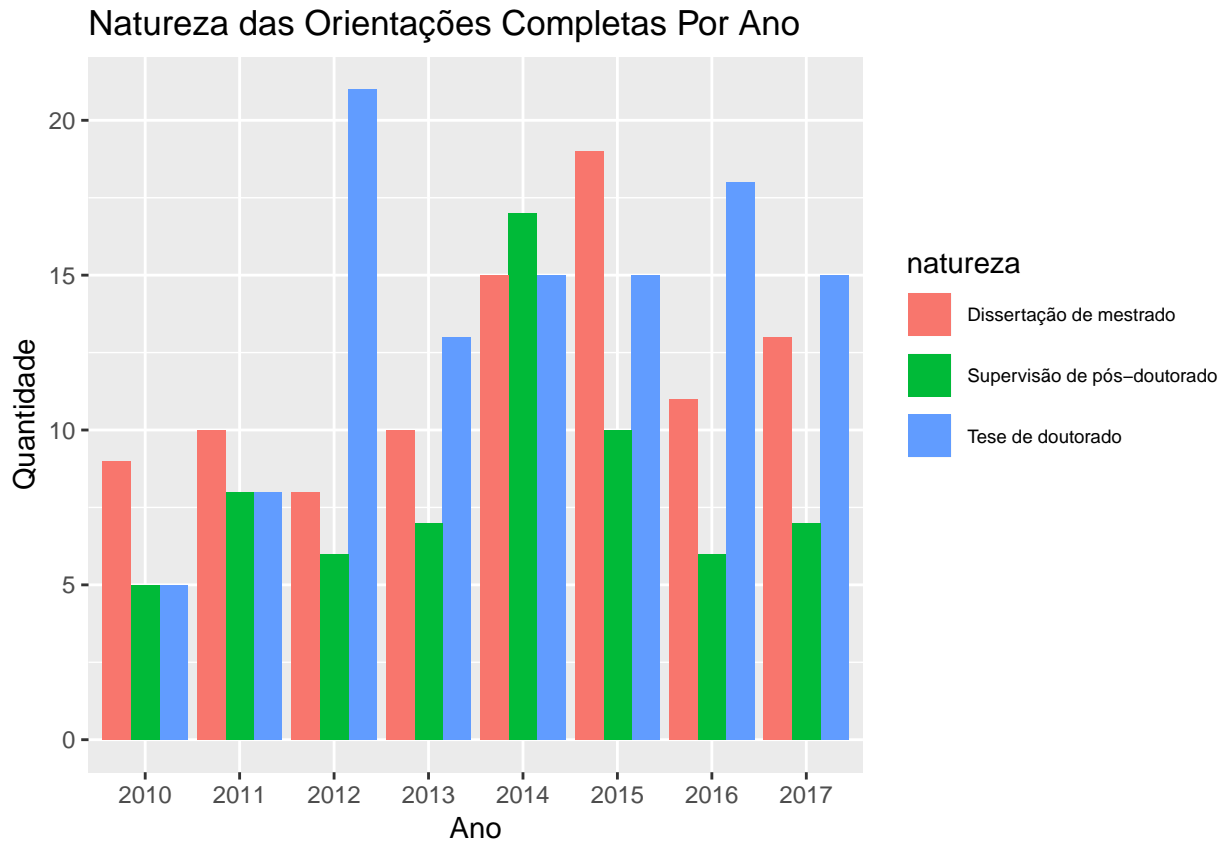
```
orient.posdoutorado.df <- ori.ls2df(advice_ens, 6) #pos-Doutorado concluído
orient.doutorado.df <- ori.ls2df(advice_ens, 7) #Doutorado concluído
orient.mestrado.df <- ori.ls2df(advice_ens, 8) #Mestrado concluído
orient.df <- rbind(rbind(orient.posdoutorado.df, orient.doutorado.df), orient.mestrado.df)
ggplot(orient.df, aes(ano, fill=natureza)) +
  geom_bar(stat = "count", position="dodge") +
  ggtitle("Natureza das Orientações 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")
```



A partir deste gráfico, nota-se que a produção de orientações de pós-doutorando iniciou-se apenas em 2014. Percebe-se também uma constante diminuição em sua quantidade no decorrer dos anos, até 2017. Em 2011, temos uma quantidade de mestrados muito maior do que nos outros anos ou do que os outros programas analisados.

- Matemática

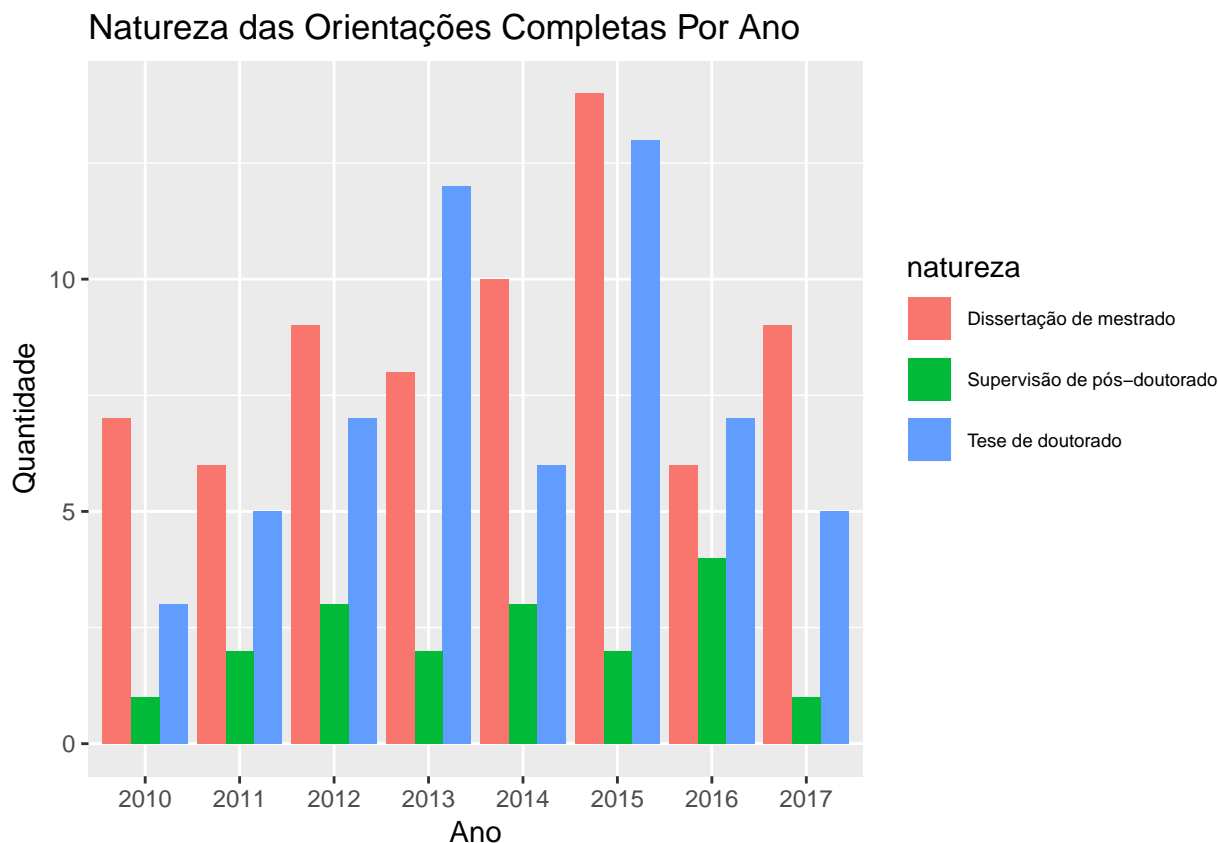
```
orient.posdoutorado.df <- ori.ls2df(advice_mat, 6) #pos-Doutorado concluído
orient.doutorado.df <- ori.ls2df(advice_mat, 7) #Doutorado concluído
orient.mestrado.df <- ori.ls2df(advice_mat, 8) #Mestrado concluído
orient.df <- rbind(rbind(orient.posdoutorado.df, orient.doutorado.df), orient.mestrado.df)
ggplot(orient.df, aes(ano, fill=natureza)) +
  geom_bar(stat = "count", position="dodge") +
  ggtitle("Natureza das Orientações 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")
```



O programa de Matemática possui uma quantidade de orientações maior e mais distribuída do que os outros programas, como é possível notar neste gráfico. Além disso, a produção de teses do ano de 2012 se destaca em relação aos outros anos e também em relação aos outros programas.

- Matemática Aplicada

```
orient.posdoutorado.df <- ori.ls2df(advice_apl, 6) #pos-Doutorado concluído
orient.doutorado.df <- ori.ls2df(advice_apl, 7) #Doutorado concluído
orient.mestrado.df <- ori.ls2df(advice_apl, 8) #Mestrado concluído
orient.df <- rbind(rbind(orient.posdoutorado.df, orient.doutorado.df), orient.mestrado.df)
ggplot(orient.df, aes(ano, fill=natureza)) +
  geom_bar(stat = "count", position="dodge") +
  ggtitle("Natureza das Orientações 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")
```

É possível notar que 2015 foi o ano de maior produção. Isso se deve ao grande número de teses de doutorado e de dissertações de mestrados concluídas.

3.4.11 Mineração de Texto

O BoW (Bag-of-Words) é um modelo de extração de características de texto simples e flexível. Ele se baseia no número de ocorrências de palavras de uma frase. Para isso, é construído um vetor com n elementos, onde n é o número de palavras do vocabulário considerado.

A fim de extrair conteúdo significativo dos textos, algumas técnicas de Processamento de Linguagem Natural (PLN) foram aplicadas, como remoção das stop words. As stop words (palavras vazias) são palavras que agregam pouco ou nenhum valor semântico. Geralmente, são as palavras mais comuns da língua, incluindo artigos, preposições, verbos de ligação, entre outras. Não existe um conjunto bem definido de quais palavras devem ser classificadas como palavras vazias. Naturalmente, esse conjunto depende do idioma em questão e em alguns casos, termos recorrentes do contexto, como jargões e gírias, também podem ser inclusos. No atual trabalho, foram consideradas as stop words em inglês e português.

```
palavras_vazias <- c(stopwords('en'), stopwords('pt'))
```

Outra técnica aplicada para a limpeza é a stemização. Do inglês, stemming, refere-se ao processo de reduzir as palavras à uma forma primitiva, como um radical. Este processo visa a generalização de pequenas variações nas palavras, como variações de gênero ou número. As palavras “trabalhador”, “trabalhadores” e “trabalhadora” tornam-se uma só. Para tal, foi utilizado o stemDocument do pacote SnowballC.

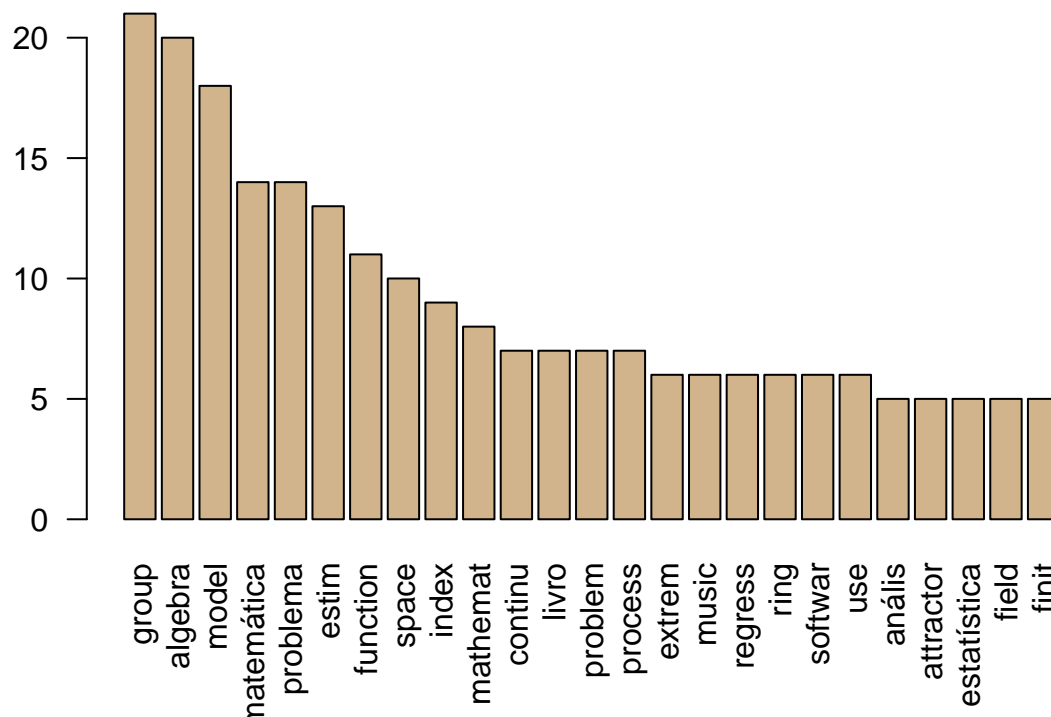
Além das técnicas citadas, outras estratégias de limpeza também foram aplicadas, como remoção de números, pontuações, espaços em branco excessivos e conversão dos termos para minúsculo. Como resultado, acompanhe as análises a seguir.

3.4.11.1 Palavras mais comuns nos títulos de publicações

- Ensino de Matemática

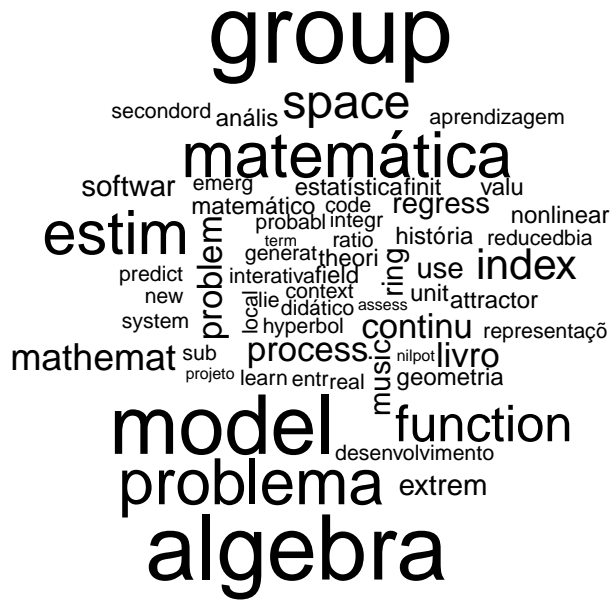
```
public.periodico.df <- pub.ls2df(public_ens, 1)
# capturando títulos e realizando limpeza
titulos <- public.periodico.df['titulo'] %>%
  tolower() %>%
  removePunctuation() %>%
  stripWhitespace %>%
  removeNumbers() %>%
  stemDocument(language = "english") %>%
  removeWords(palavras_vazias)%>%
  VectorSource() %>%
  VCorpus()
# Exibindo gráfico
titulos_tdm <- TermDocumentMatrix(titulos)
titulos_m <- as.matrix(titulos_tdm)
term_frequency <- rowSums(titulos_m)
term_frequency <- sort(term_frequency, decreasing = TRUE)
barplot(term_frequency[1:25], col = "tan", las = 2, main = "Palavras mais
frequentes nos títulos das publicações")
```

**Palavras mais
frequentes nos títulos das publicações**



```
# Exibindo wordcloud
term_vec <- names(term_frequency)
wordcloud(term_vec, term_frequency, max.words = 60, scale=c(3,.1))
title(main="Nuvem de palavras dos títulos dos artigos")
```

Nuvem de palavras dos títulos dos artigos

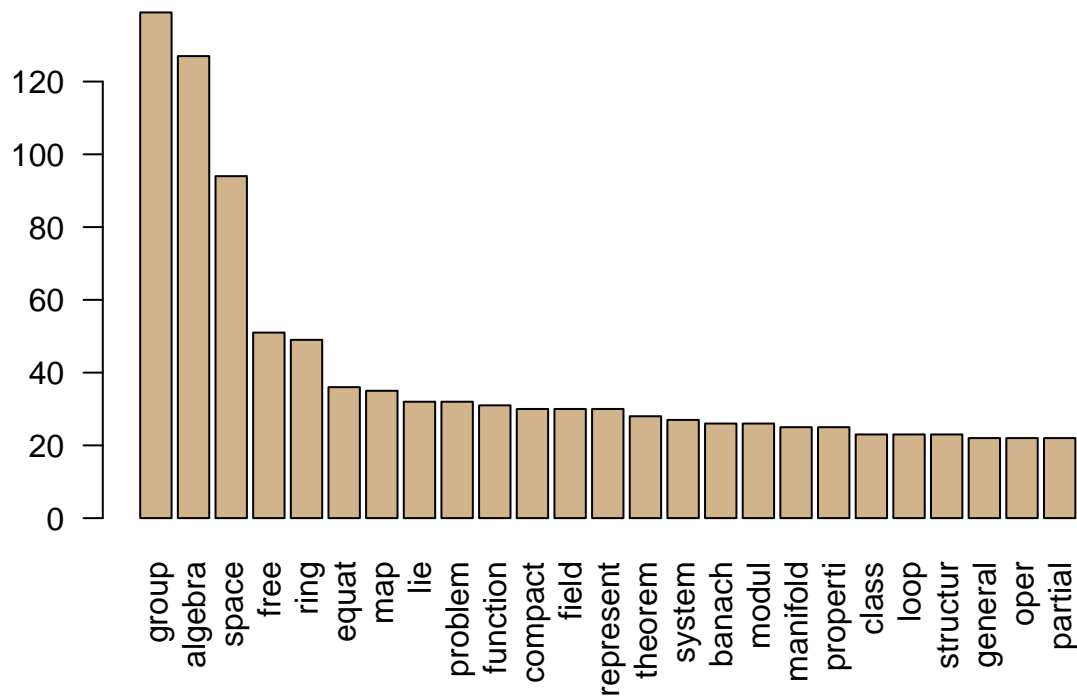


- Matemática

```
public.periodico.df <- pub.ls2df(public_mat, 1)
# capturando títulos e realizando limpeza
titulos <- public.periodico.df['titulo'] %>%
  tolower() %>%
  removePunctuation() %>%
  stripWhitespace %>%
  removeNumbers() %>%
  stemDocument(language = "english") %>%
  removeWords(palavras_vazias)%>%
  VectorSource() %>%
  VCorpus()

# Exibindo gráfico
titulos_tdm <- TermDocumentMatrix(titulos)
titulos_m <- as.matrix(titulos_tdm)
term_frequency <- rowSums(titulos_m)
term_frequency <- sort(term_frequency, decreasing = TRUE)
barplot(term_frequency[1:25], col = "tan", las = 2, main = "Palavras mais
frequentemente nos títulos das publicações")
```

Palavras mais frequentes nos títulos das publicações



- Matemática Aplicada

```
# Exibindo wordcloud
```

```
term_vec <- names(term_frequency)
```

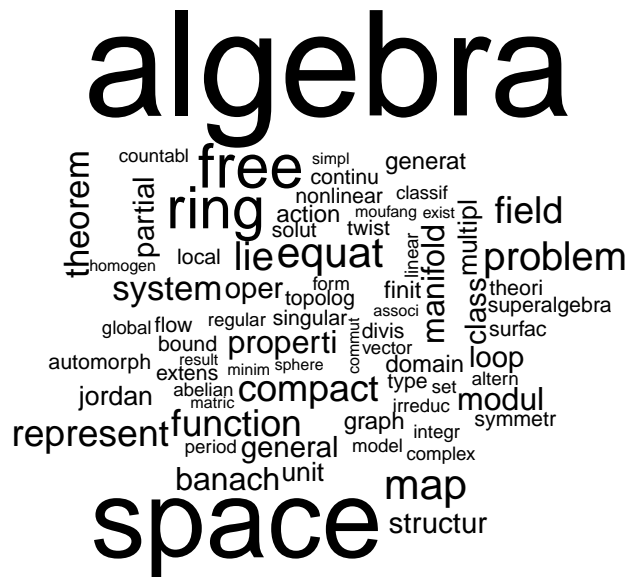
```
wordcloud(term_vec, term_frequency, max.words = 75, scale=c(5,.1))
```

```
## Warning in wordcloud(term_vec, term_frequency, max.words = 75, scale =
```

```
## c(5, : group could not be fit on page. It will not be plotted.
```

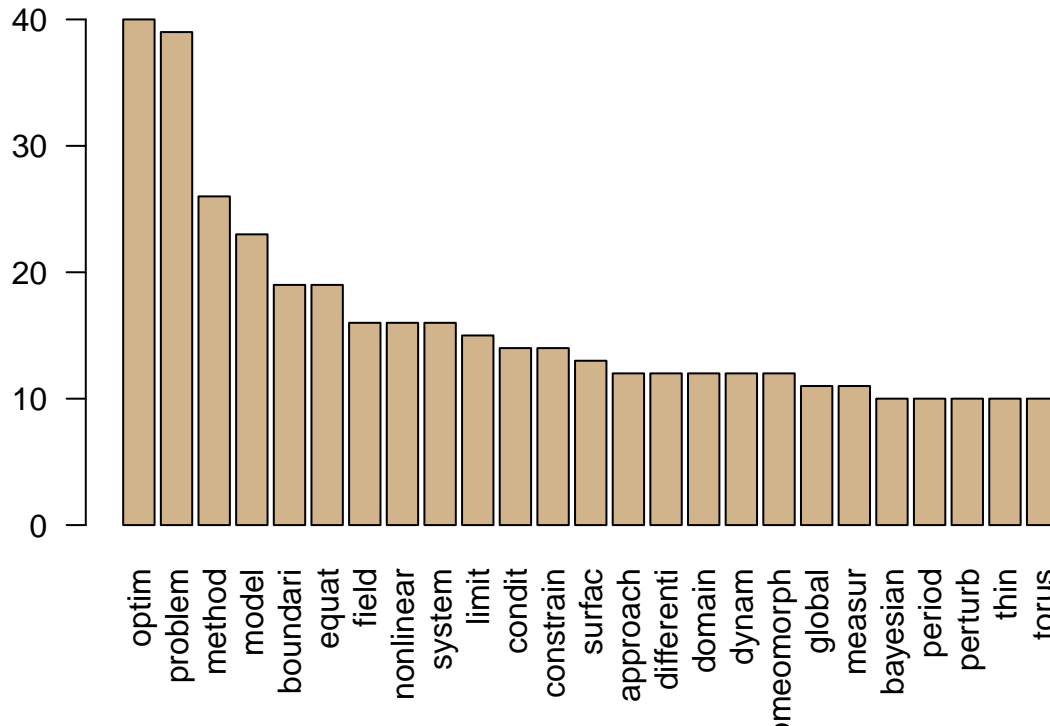
```
title(main="Nuvem de palavras dos títulos dos artigos")
```

Nuvem de palavras dos títulos dos artigos

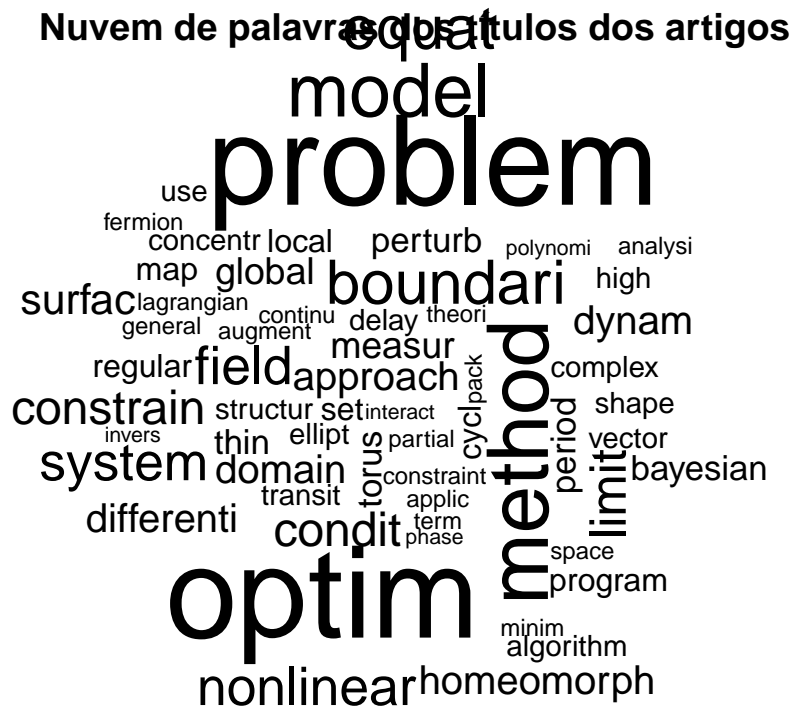


```
public.periodico.df <- pub.ls2df(public_apl, 1)
# capturando títulos e realizando limpeza
titulos <- public.periodico.df['titulo'] %>%
  tolower() %>%
  removePunctuation() %>%
  stripWhitespace %>%
  removeNumbers() %>%
  stemDocument(language = "english") %>%
  removeWords(palavras_vazias)%>%
  VectorSource() %>%
  VCorpus()
# Exibindo gráfico
titulos_tdm <- TermDocumentMatrix(titulos)
titulos_m <- as.matrix(titulos_tdm)
term_frequency <- rowSums(titulos_m)
term_frequency <- sort(term_frequency, decreasing = TRUE)
barplot(term_frequency[1:25], col = "tan", las = 2, main = "Palavras mais
frequentes nos títulos das publicações")
```

Palavras mais frequentes nos títulos das publicações



```
# Exibindo wordcloud
term_vec <- names(term_frequency)
wordcloud(term_vec, term_frequency, max.words = 60, scale=c(4,.1))
title(main="Nuvem de palavras dos títulos dos artigos")
```



Avaliando os resultados dos gráficos de Palavras mais frequentes nos títulos das publicações e nos wordclouds gerados através dessas palavras, é possível apontar os termos e assuntos mais discutidos nessas publicações. Nos três programas é evidente a predominância do idioma inglês nos títulos das publicações. Os termos “group” e “algebra” são fortemente utilizados pelos pesquisadores de Matemática e Matemática aplicada, em ambos, estes são os termos mais comuns. Por outro lado, os termos mais comuns em Ensino de Matemática são “optim” (resultado da stemização, significa qualquer palavra derivada de “optimization”) e “problem”. Através dessas observações é possível apontar que o programa de Ensino de Matemática possui um enfoque muito diferente dos outros dois programas, o que condiz com o que pôde ser observado nas outras análises realizadas até aqui.

3.4.12 Análise de Rede

A partir da análise de rede é possível aprender como os professores como os professores se relacionam e aprender se eles formam grupos dentro da rede em que se encontram. Aqui podemos ver dados armazenados no graph.json de cada programa. A coluna label serve para identificar o pesquisador na visualização da rede mais abaixo. A largura do vértice indica que os pesquisadores trabalharam mais vezes. Em seguida é possível ver os agrupamentos de cada programa.

- Ensino de Matemática

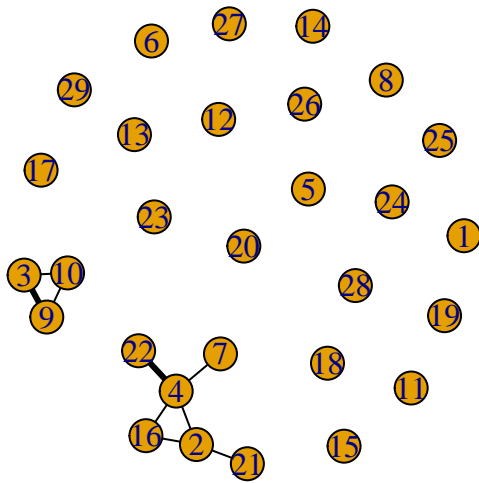
```
# Lê os dados de ensino de matemática
graph_ens_mat <- fromJSON("./Ensino de Matemática (33002010223P9)/graph.json")
graph_ens_mat$nodes
```

##		id	label	name
## 1	0348490713417429	1		Leliane Nunes de Barros
## 2	1464247269026445	2		David Pires Dias
## 3	1601481697363454	3		Maria Cristina Bonomi
## 4	1727582332230890	4		Barbara Corominas Valerio
## 5	1915297691969734	5		Elvia Mureb Sallum
## 6	2153527263061692	6		Vera Helena Giusti de Souza
## 7	2422103751979129	7		Antonio Carlos Brolezzi
## 8	2628621250028497	8		Orlando Stanley Juriaans
## 9	3612359023677691	9		Ana Paula Jahn
## 10	3963038169664451	10		Iole de Freitas Druck
## 11	4159733067004447	11		Ricardo Bianconi
## 12	4654251951434427	12		Eduardo do Nascimento Marcos
## 13	5502172167494560	13	Elisete da Conceição Quintaneiro Aubin	
## 14	5618622435626525	14	Leônidas de Oliveira Brandão	
## 15	5798035148953676	15	Helena Maria Avila de Castro	
## 16	6261615324975968	16	Cláudia Cueva Candido	
## 17	6884084734880165	17	Oscar João Abdounur	
## 18	6912823669386029	18	Francisco Cesar Polcino Milies	
## 19	7037569509778870	19	Lígia Carla Pinto Henriques Jorge Rodrigues	
## 20	7191150286119863	20	Martha Salerno Monteiro	
## 21	7193492880677720	21	Cristina Cerri	
## 22	7507242119255330	22	Rosa Maria dos Santos Barreiro Chaves	
## 23	7522850820895210	23	Antonio Luiz Pereira	
## 24	7810711686517284	24	Circe Mary Silva da Silva Dynnikov	
## 25	7902197542498193	25	Rogério Augusto dos Santos Fajardo	
## 26	8165895654064360	26	Viviana Giampaoli	
## 27	8280771165693528	27	Marcos Nascimento Magalhaes	
## 28	8883044509372931	28	Lisbeth Kaiserlian Cordani	
## 29	9799994124940260	29	Silvia Nagib Elian	

```
# Gera visualização da rede
g_ens_mat <- graph_from_data_frame(graph_ens_mat$links,
                                   directed=FALSE,
                                   vertices=graph_ens_mat$nodes)

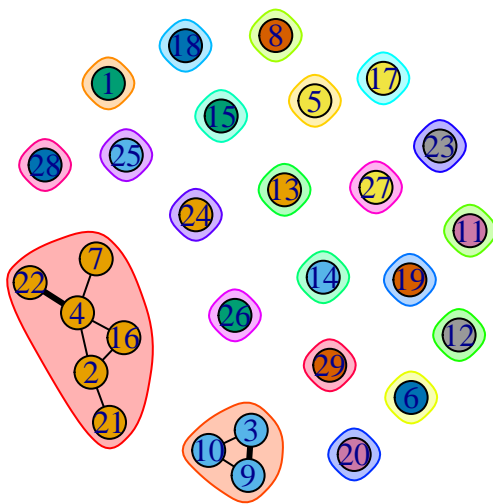
plot(g_ens_mat,
     layout=layout_nicely(g_ens_mat),
     edge.width = graph_ens_mat$links$weight,
     edge.color = "black",
     main="Rede de Ensino de Matemática")
```

Rede de Ensino de Matemática



```
# Gera visualização das comunidades
ens_mat_community = fastgreedy_community(g_ens_mat)
plot(ens_mat_community,
     g_ens_mat,
     edge.width = graph_ens_mat$links$weight,
     main="Agrupamentos de Ensino de Matemática")
```


Agrupamentos de Ensino de Matemática



- Matemática

Lê os dados de Matemática

```
graph_mat <- fromJSON("./Matemática (33002010005P1)/graph.json")
graph_mat$nodes
```

##	id	label	name
## 1	0368187000548549	1	Raul Antonio Ferraz
## 2	0814550345087037	2	Cristian Andres Ortiz Gonzalez
## 3	0964053090112695	3	Leonardo Pellegrini Rodrigues
## 4	1244442414351450	4	Daniela Mariz Silva Vieira
## 5	1318171263288733	5	Marcos Martins Alexandrino da Silva
## 6	1334333415248806	6	Alexandre Lymberopoulos
## 7	1344385349084972	7	Martha Patricia Dussan Angulo
## 8	1504083676579535	8	Mikhajolo Dokuchaev
## 9	1520212806910322	9	Daniel Victor Tausk
## 10	1597520020384471	10	Salvador Addas Zanata
## 11	1606389825259677	11	Edson Vargas
## 12	1703939123866491	12	Roberto Mossa
## 13	2115528633747994	13	Severino Toscano do Rego Melo
## 14	2202693274986226	14	Jorge Manuel Sotomayor Tello
## 15	2281591530807032	15	Juan Carlos Gutiérrez Fernández
## 16	2382469130424035	16	Ivan Chestakov
## 17	2628621250028497	17	Orlando Stanley Juriaans
## 18	2869491387764118	18	Paolo Piccione
## 19	3066990520731287	19	Ofelia Teresa Alas
## 20	3487250524564847	20	Ricardo dos Santos Freire Júnior
## 21	3762450205090582	21	Hugo Luiz Mariano
## 22	4031361680529891	22	Wilson Albeiro Cuellar Carrera
## 23	4110545446460055	23	Rodrigo Bissacot Proença
## 24	4159733067004447	24	Ricardo Bianconi
## 25	4183069998276255	25	Lucia Satie Ikemoto Murakami
## 26	4595265210121698	26	Christina Brech
## 27	4627360041573918	27	Yoshiharu Kohayakawa
## 28	4654251951434427	28	Eduardo do Nascimento Marcos

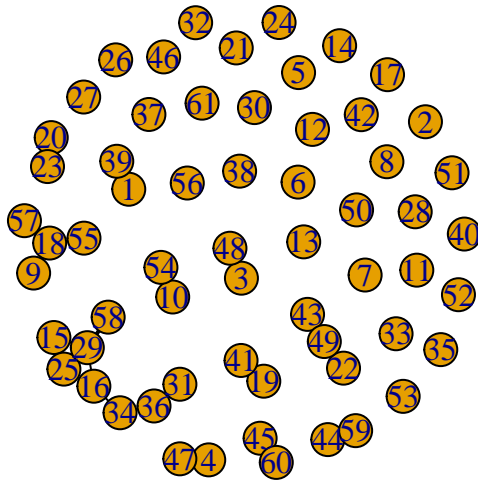
## 29	4822164054567444	29	Alexandre Grichkov
## 30	4924243158075998	30	Flavio Ulhoa Coelho
## 31	5220668843910824	31	Kostiantyn Iusenko
## 32	5563844615917020	32	Henrique Guzzo Junior
## 33	6227659115867830	33	Francisco Miraglia Neto
## 34	6279103178873169	34	Iryna Kashuba
## 35	6414101223818111	35	Jaime Angulo Pava
## 36	6765895362060524	36	Vyacheslav Futorny
## 37	6837855526877302	37	Fabiano Gustavo Braga Brito
## 38	6876840066785728	38	Ivan Struchiner
## 39	6912823669386029	39	Francisco Cesar Polcino Milies
## 40	6924559566010266	40	Claudio Gorodski
## 41	7034382452298591	41	Lucia Renato Junqueira
## 42	7457186791833218	42	Paulo Domingos Cordaro
## 43	7500070840563045	43	Eloi Medina Galego
## 44	7522850820895210	44	Antonio Luiz Pereira
## 45	7541987739314374	45	Jairo Zacarias Goncalves
## 46	7650178604363178	46	Artur Hideyuki Tomita
## 47	7886856633475996	47	Mary Lilian Lourenco
## 48	7902197542498193	48	Rogério Augusto dos Santos Fajardo
## 49	8035634297294419	49	Valentin Raphael Henri Ferenczi
## 50	8367930445383308	50	Sylvain Philippe Pierre Bonnot
## 51	8455495596164984	51	Sinai Robins
## 52	8580052002817169	52	Daciberg Lima Gonçalves
## 53	8905672608345856	53	Edson de Faria Francisco
## 54	9060045633534022	54	Pedro Antonio Santoro Salomão
## 55	9074528435754216	55	Pierluigi Benevieri
## 56	9134120309868145	56	Albert Meads Fisher
## 57	9171102073588628	57	Gaetano Siciliano
## 58	9358236443006461	58	Maria de Lourdes Merlini Giuliani
## 59	9681052469456395	59	Marcone Corrêa Pereira
## 60	9711211873693776	60	Javier Sánchez Serdà
## 61	9764244146941023	61	Glaucio Terra

```
# Gera visualização da rede
```

```
g_mat <- graph_from_data_frame(graph_mat$links,
                               directed=FALSE,
                               vertices=graph_mat$nodes)

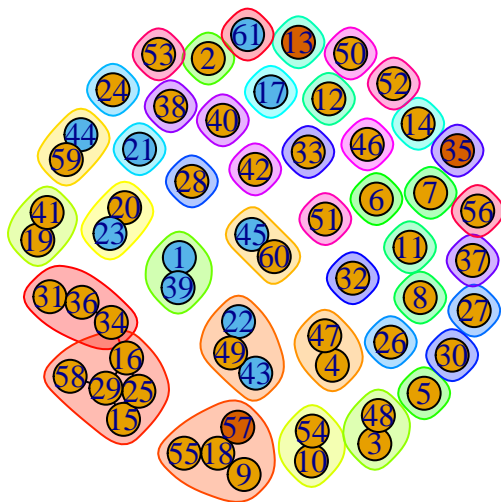
plot(g_mat,
     layout=layout_nicely(g_mat),
     edge.width = graph_mat$links$weight,
     edge.color = "black",
     main="Rede de Matemática")
```

Rede de Matemática



```
# Gera visualização das comunidades
mat_community = fastgreedy.community(g_mat)
plot(mat_community,
     g_mat,
     graph_mat$links$weight,
     main="Grupos de Matemática")
```

Grupos de Matemática



- Matemática Aplicada

```
# Lê os dados de Matemática Aplicada
graph_mat_apl <- fromJSON("./MATEMÁTICA APLICADA (33002010006P8)/graph.json")
graph_mat_apl$nodes
```

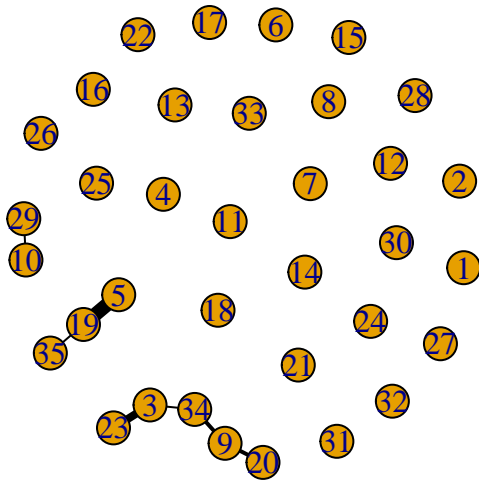
##	id	label	name
## 1	0647513129067496	1	Claudia Monteiro Peixoto
## 2	0694302682822936	2	Eduardo Colli

## 3	0746337016670411	3	Saulo Rabello Maciel de Barros
## 4	1498618533380124	4	Walter Alberto de Siqueira Pedra
## 5	1597520020384471	5	Salvador Addas Zanata
## 6	1606389825259677	6	Edson Vargas
## 7	1914965078179670	7	Renato Vicente
## 8	2202693274986226	8	Jorge Manuel Sotomayor Tello
## 9	2376991776742062	9	Gleiciane da Silva Aragão
## 10	2443304092293827	10	Walter Figueiredo Mascarenhas
## 11	2506136880977644	11	Antoine Laurain
## 12	2612915149105321	12	Gabriel Haeser
## 13	3469377550928789	13	Orlando Francisco Lopes
## 14	4110545446460055	14	Rodrigo Bissacot Proença
## 15	4361743248742740	15	Ana Cristina de Oliveira Mereu
## 16	4517534884485194	16	Antonio Elias Fabris
## 17	4804300167674691	17	Clodoaldo Grotta Ragazzo
## 18	5181059029789860	18	Luis Carlos de Castro Santos
## 19	5658903073741596	19	Fábio Armando Tal
## 20	6218135906579844	20	Sergio Muniz Oliva Filho
## 21	6419833437574297	21	Joyce da Silva Bevilacqua
## 22	6518215277080266	22	Nelson Mugayar Kuhl
## 23	6547630839036017	23	Pedro da Silva Peixoto
## 24	6721706447042143	24	Luciana Luna Anna Lomonaco
## 25	6748364534679596	25	Frank Michael Forger
## 26	6790064280260422	26	Pedro Tavares Paes Lopes
## 27	7457186791833218	27	Paulo Domingos Cordaro
## 28	7625751421942524	28	Christian Dieter Jäkel
## 29	7750373154186324	29	Ernesto Julián Goldberg Birgin
## 30	7981328728299342	30	Sonia Regina Leite Garcia
## 31	8644214878865621	31	André Salles de Carvalho
## 32	8872656701324553	32	Pedro Aladar Tonelli
## 33	9582404119292455	33	Julio Michael Stern
## 34	9681052469456395	34	Marcone Corrêa Pereira
## 35	9893531212718568	35	Manuel Valentim de Pera Garcia

```
# Gera visualização da rede
g_mat_apl <- graph_from_data_frame(graph_mat_apl$links,
                                   directed=FALSE,
                                   vertices=graph_mat_apl$nodes)

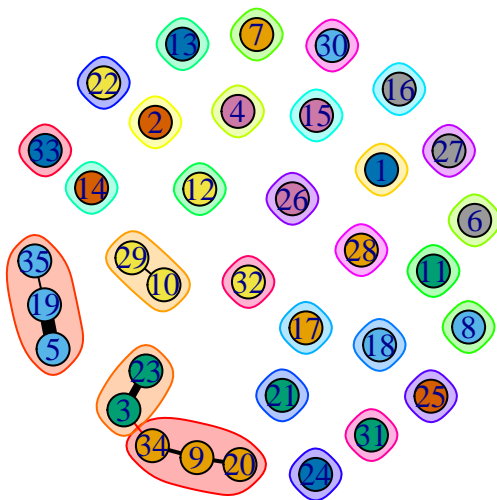
plot(g_mat_apl,
     layout=layout_nicely(g_mat_apl),
     edge.width = graph_mat_apl$links$weight,
     edge.color = "black",
     main="Rede de Matemática Aplicada")
```

Rede de Matemática Aplicada



```
# Gera visualização das comunidades
mat_apl_community = fastgreedy.community(g_mat_apl)
plot(mat_apl_community,
     g_mat_apl,
     edge.width = graph_mat_apl$links$weight,
     main="Grupos de Matemática Aplicada")
```

Grupos de Matemática Aplicada



De acordo com as redes é possível notar que um grande número de pesquisadores de um mesmo programa está isolado. Não há uma colaboração muito grande. Dos poucos que trabalham juntos, poucos colaboraram mais de uma vez. Isso indica uma rede bem fragmentada e poucos agrupamentos. As redes dos programas são muito semelhantes, entretanto é possível encontrar um destaque no programa de Ensino de Matemática onde o pesquisador 4, Barbara Corominas Valerio, apresenta uma influência no grupo onde se encontra. Além ter colaborado mais de uma vez com outro pesquisador, ela serve como ponto de conexão entre outros quatro pesquisadores. Na rede de Matemática, Lucia Renato Junqueira, label 41, se destaca como a que pesquisadora que tem mais relações em seu grupo. Já no grupo de Matemática Aplicada, o destaque é Fábio Armando Tal, label 19, que tem mais colaboração do que qualquer outro pesquisador entre todos os programas.

3.4.13 Comparação entre os programas

Analisando o número de pessoas em cada grande área dos programas avaliados, fica claro a natureza da pesquisa, pois, a grande maioria se concentra em ciências da natureza. Podemos citar como um fator de diferença na presença da grande área humanas em um dos programas avaliados, mas faz todo o sentido por se tratar de um programa voltado ao ensino, que faz-se uso de estudo dessa natureza.

Em todos os programas a maior área de atuação dos pesquisadores foi matemática, o programa matemática foi ainda mais dominante o tema, e ensino da matemática e matemática aplicada ainda foi possível observar que uma grande porcentagem dos pesquisadores possuem temas além da própria matemática.

As especialidades mais frequentes dos pesquisadores dos 3 programas foram similares, tendo ocorrências que apareceram nos 3 programas. De maneira geral os programas matemática e ensino da matemática apresentam um grande número de coincidências e os programas matemática e matemática aplicada tiveram grande similaridade também.

Estatística representa uma grande porcentagem das subáreas frequentes do programa de ensino de matemática, e não está presente no top 7 subáreas dos demais programas. Destaque para a subárea análise que representa uma grande porcentagem em todos os programas.

O programa matemática tem um número bem maior de publicações em revistas, a revista mais publicada recebeu 32 publicações. Enquanto ensino de matemática e matemática aplicada possuem, 17 e 8 respectivamente nas revistas mais públicas.

Uma grande discrepância ocorreu em relação ao número de eventos por ano, ensino de matemática em média possui o dobro dos demais programas.

O programa matemática é disparado o que mais fez publicações nos anos avaliados, ensino de matemática e matemática aplicada não costuma publicar mais que 45 artigos, enquanto a média de matemática é de 100 artigos.

3.4.14 Comparação com os dados da CAPES

Como é possível, observar, o programa de Matemática possui um número bem maior de pesquisadores. Esses dados são refletidos em outros índices como número de trabalhos publicados e quantidade de especializações, onde o programa de Matemática também tem um valor acima dos outros. Esse valor também reflete a quantidade de discentes no programa segundo porta do CAPES: Ensino de Matemática 78, Matemática 132 e Matemática aplicada 110.

```
print(paste('Ensino de Matemática =', length(perfil_ens)))
```

```
## [1] "Ensino de Matemática = 29"
```

```
print(paste('Matemática =', length(perfil_mat)))
```

```
## [1] "Matemática = 61"
```

```
print(paste('Matemática Aplicada =', length(perfil_apl)))
```

```
## [1] "Matemática Aplicada = 35"
```

O programa ensino de matemática por sua natureza de mestrado profissional apresenta algumas diferenças para os demais programas, um gráfico que deixa isso em bastante evidência é o gráfico de eventos por ano.

O programa ensino de matemática recebeu avaliação 4 do mec, e os demais programas receberam nota 5. Isso pode ser compreendido através do presente trabalho, o número de publicações foi menor desse programa e as revistas publicadas com quais menor impactando na sua nota.