Análise da Produção Científica e Acadêmica da Universidade Federal do Ceará - Ciência de Dados para Todos - Departamento de Ciência da Computação da Universidade de Brasília

Professores: Jorge Henrique Cabral Fernandes e Ricardo Barros Sampaio

Grupo 13
Amanda Bezerra da Silva 15/0057113
Edgar Sampaio de Barros 16/0005213
George Geonardo de Pontes da Silva 12/0012197
Vinícius Costa e Silva 15/0052138

30/06/2019

Contents

1 Introdução

Este trabalho foi realizado pelo Grupo 13 da disciplina Ciência de Dados para Todos (Data Science For All), Turma D, do 1º semestre de 2019, do Departamento de Ciência da Computação da Universidade de Brasília (UnB) e visa fornecer análises sobre 3 (três) programas de pós-graduação da Universidade Federal do Ceará (UFC), sendo eles: Ciências da Computação, Computação e Matemática.

O desenvolvimento deste trabalho segue o modelo metodológico de mineração de dados CRISP-DM em conjunto com tecnologias e técnicas estudadas ao longo do semestre na realização da disciplina. Como resultado do trabalho, espera-se produzir relatórios da avaliação da qualidade dos programas de mestrado e doutorado dos cursos selecionados da UFC, com base em dados recolhidos da plataforma Lattes (um currículo de pesquisadores do Brasil).

2 Contextualização

2.1 O que é ciência?

Uma das definições de ciência concebida por Fernandes, 2017 é a de que ciência é um estudo metódico e organizado sobre qualquer fenômeno que pode acontecer no universo com o objetivo de explicar e prever esses fenômenos. A produção da ciência é uma atividade fundamentalmente social.

A ciência reúne conhecimentos de diversas áreas, onde cada uma possui seu conjunto de princípios e métodos de verificação. Na ciência e em suas áreas é importante que o conhecimento seja certificado, ou seja, que se apresentem provas que comprovem sua veracidade. Assim surge a necessidade de se ter metodologia científica, que é o estudo dos métodos e instrumentos necessários para produção de um trabalho científico.

A ciência é dividida principalmente em ciências exatas e ciências humandas, mas há diversas outras formas de subdividi-la, podendo ser de acordo com os alvos, da finalidade, do grau de precisão, da área e dos métodos empregados no estudo.

A ciência e suas várias áreas vem evoluindo cada vez mais graças às tecnologias e novas formas de pesquisa e verificação.

2.2 O que é ciência no Brasil?

Segundo o rank feito pelo Fórum Econômico Mundial, em 2010 - 2011 o Brasil ficou em 56º no aspecto de desenvolvimento mundial de tecnologia da informação.

A produção científica brasileira começou de fato nas primeiras décadas do século XIX. Mas as primeiras universidades brasileiras surgiram no início do século XX e mais ou menos nessa mesma época surgiu a Academia Brasileira de Ciências. Em 1951 foi criado o Conselho Nacional de Pesquisa (CNPq) e então surge a Coordenação de Pessoal de Ensino Superior (Capes). De lá pra cá o Brasil começou a ganhar destaque na produção científica mundial.

No Brasil, a ciência é produzida em sua maior parte por universidades públicas e institutos próprios de pesquisa. As universidades e institutos recebem financiamento para que possam realizar suas produções científicas, tais financiamentos podem vim por parte do governo, de companhias públicas, da indústria, de associações e até mesmo de outros países.

3 Metodologia

Proposto ainda em 1996, O modelo CRISP-DM (*Cross Industry Standard Process for Data Mining*) utilizado neste trabalho é um modelo de referência que fornece um processo bem estruturado para realização de projetos de mineração de dados e é um padrão para mineração de dados que é o mais comumente utilizado na indústria. Trata-se de um modelo amplamente utilizado no mercado e que base de um princípio de entendimento do negócio cujos dados querem ser analisados, e então dos dados e sua posterior preparação, modelagem e análise.

O CRISP-DM quebra o processo de mineração de dados em seis fases distintas, embora não estritamente sequenciais. É comum, durante o desenvolvimento da solução para o problema, que se alterne entre as fases do CRISP-DM conforme seja necessário.

O ciclo de vida do CRISP-DM apresenta 6 (seis) fases, ilustradas na figura 1.

3.1 Delimitações iniciais

Algumas delimitações são apresentadas a seguir usando a estrutura do CRISP-DM.

3.1.1 Domínio de aplicação do projeto

O domínio de aplicação deste projeto é o da produção científica e acadêmica de um subgrupo de pesquisadores vinculados à Universidade Federal do Ceará.

3.1.2 Tipo de problema abordado

O problema abordado neste trabalho é o da produção de análises descritivas, quantitativas e de modelagem computacional ou estatística, que permitam caracterizar como e porque ocorre a produção científica e acadêmica de um subgrupo de pesquisadores da Universidade Federal do Ceará. Mais especificamente, consiste no subgrupo formado pelos pesquisadores dos cursos de Ciências da Computação, Computação e Matemática da UFC.

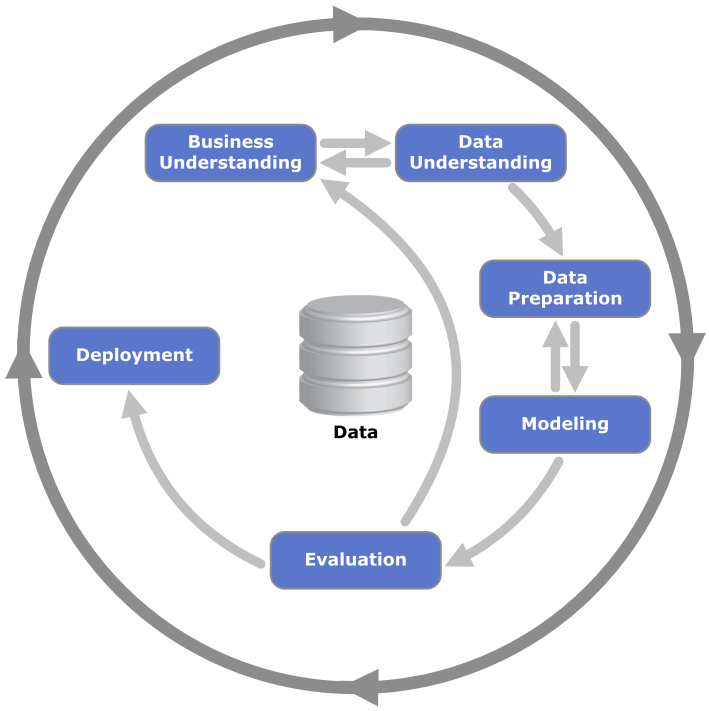


Figure 1: Fases CRISP-DM

3.2 Modelo de Referência CRISP-DM

Cada fase do ciclo de vida do CRISP-DM é composto por um conjunto de atividades, as quais podem ser subdivididas em atividades ainda mais específicas. Dessa maneira, busca-se compreender o todo antes de se aprofundar em análises mais específicas do problema, de modo a ter certeza que os modelos e análises posteriormente desenvolvidas irão refletir corretamente o domínio do problema que se deseja tratar.

A seguir é apresentado uma descrição das fases de forma resumida:

- Fase 1 - Entendimento do negócio: definição dos objetivos e necessidades do projeto sob a perspectiva do negócio para que se possa definir o problema ou tema de data mining a ser abordado.
- Fase 2 - Entendimento dos dados: realização de coleta, organização e descrição dos dados disponíveis, além da análise qualitativa e quantitativa dos mesmos a fim de determinar a viabilidade do projeto.
- Fase 3 - Preparação dos dados: estruturação e limpeza dos dados disponíveis. Nesta fase os dados são preparados para que se possa iniciar as análises e produções de gráficos.
- Fase 4 - Modelagem: aplicação de técnicas de data mining e escolha de modelo estatísticos mais apropriado para o projeto.
- Fase 5 - Avaliação: avaliação dos resultados obtidos e do processo como um todo, analisando se todas as questões relevantes para o projeto foram abordadas de forma adequada e se existem fatores que tenham sido negligenciados.
- Fase 6 - Implementação: definição de estratégia de implantação dos produtos desenvolvidos pelo projeto para o ambiente operacional, além de monitoramento dos resultados para possíveis adaptações do modelo aplicado.

4 CRISP-DM Fase 1: Entendimento do Negócio

4.1 O que é o Sistema Nacional de Pós-Graduação?

O Sistema Nacional de Pós-Graduação (SNPG) é composto por cursos e programas de pós-graduação avaliados e reconhecidos pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e é o principal responsável pela produção de conhecimento científico no Brasil.

Para garantir a qualidade dos programas de pós-graduação que compõem o Sistema Nacional de Pós-Graduação, é realizada a Avaliação do Sistema Nacional de Pós-Graduação, sob orientação da Diretoria de Avaliação/Capes e participação da comunidade acadêmico-científica.

A avaliação é realizada em 49 áreas de avaliação, número vigente em 2017, e segue sistema e quesitos estabelecidos no Conselho Técnico Científico da Educação Superior (CTC-ES).

4.2 Os Colégios, Grandes Áreas e Áreas da Pós-Graduação Brasileira

Abaixo encontra-se a distribuição das áreas de avaliação, organizadas em 3 Colégios e 9 Grandes Áreas.

4.2.1 Colégio de Ciências da Vida

CIÊNCIAS AGRÁRIAS	CIÊNCIAS BIOLÓGICAS	CIÊNCIAS DA SAÚDE
Ciência de Alimentos	Biodiversidade	Educação Física
Ciências Agrárias I	Ciências Biológicas I	Enfermagem
Medicina Veterinária	Ciências Biológicas II	Farmácia
Zootecnia / Recursos Pesqueiros	Ciências Biológicas III	Medicina I
-	-	Medicina II
-	-	Medicina III
-	-	Nutrição
-	-	Odontologia
-	-	Saúde Coletiva

4.2.2 Colégio de Ciências Exatas, Tecnológicas e Multidisciplinar

CIÊNCIAS EXATAS E DA TERRA	ENGENHARIAS	MULTIDISCIPLINAR
Astronomia / Física	Engenharias I	Biotecnologia
Ciência da Computação	Engenharias II	Ciências Ambientais
Geociências	Engenharias III	Ensino
Matemática / Probabilidade e Estatística	Engenharias IV	Interdisciplinar
Química	-	Materiais

4.2.3 Colégio de Humanidades

CIÊNCIAS HUMANAS	CIÊNCIAS SOCIAIS APLICADAS	LINGUÍSTICA, LETRAS E ARTES
Antropologia / Arqueologia	Administração Pública e de Empresas, Ciências Contábeis e Turismo	Artes

CIÊNCIAS HUMANAS	CIÊNCIAS SOCIAIS APLICADAS	LINGUÍSTICA, LETRAS E ARTES
Ciência Política e Relações Internacionais	Arquitetura, Urbanismo e Design	Linguística e Literatura
Ciências da Religião e Teologia	Comunicação e Informação	-
Educação	Direito	-
Filosofia	Economia	-
Geografia	Planejamento Urbano e Regional / Demografia	-
História	Serviço Social	-
Psicologia	-	-
Sociologia	-	-

5 A Universidade Federal do Ceará dentro do Sistema Nacional de Pós-Graduação

De acordo com os dados disponíveis na Plataforma Sucupira em 2019, na Universidade Federal do Ceará existem 79 programas de pós-graduação em funcionamento, 3 em projeto, 9 desativados e 4 em desativação, totalizando 95 programas. Dos programas existentes, 76 são da modalidade acadêmico e 19 são da modalidade profissional.

5.1 Contextualização dos Programas

Os programas de Ciências da Computação, Computação e Matemática da Universidade Federal do Ceará são o foco deste trabalho. A seções a seguir apresentam informações a respeito destes programas, retiradas da Plataforma Sucupira em 2019.

5.1.1 Ciências da Computação

Os cursos de computação se iniciaram na UFC em 1975, inicialmente com o curso de Tecnologia de Processamento de Dados. Atualmente, reúne ao todo 237 discentes matriculados e 31 docentes. Além disso, conta com 79 disciplinas de Mestrado e Doutorado em vigência. O mestrado em Ciência da Computação da UFC começou no ano de 1995, recebendo aprovação da CAPES em 1997 e recebendo nota 4 em 2004. O doutorado foi consolidado em 2004, também com nota 4, e acabou por iniciar suas atividades em 2005.

Atualmente há 5 grupos de pesquisa atuando dentro do programa de Mestrado e Doutorado de Ciências da Computação da UFC: o ARIDA (Advanced Research in Database), o CRAb (Computer Graphics, Virtual Reality and Animation), o GREaT (grupo de Redes de Computadores, Engenharia de Software e Sistemas), o ParGO (Paralelismo, Grafos e Otimização) e o LOGIA (Lógica e Inteligência Artificial).

Em 2013, o programa de mestrado e doutorado recebeu conceito 5 na avaliação trienal da CAPES (referente ao período dos anos 2010-2011-2012).

Abaixo encontra-se a distribuição das áreas de concentração, linhas de pesquisa e projetos de pesquisa deste programa.

5.1.1.1 Área de Concentração Ciência da Computação

Organiza-se nas seguintes linhas de pesquisa:

- Ciência Computacional, que possui 3 projetos de pesquisas em andamento

- Sistemas de Informação, que possui 27 projetos de pesquisas em andamento
- Teoria da Computação, que possui 10 projetos de pesquisas em andamento

Totalizando 40 projetos de pesquisas.

Os projetos de pesquisa atualmente em andamento na pós-graduação em Ciências da Computação da UFC são:

Projeto	Linha de pesquisa
AGREGAÇÃO DE CRENÇAS E PREFERÊNCIAS EM PROGRAMAÇÃO EM LÓGICA E EM FRAMEWORKS DE ARGUMENTAÇÃO	TEORIA DA COMPUTAÇÃO
Analyzing big data with time-dependent graphs and machine learning: application to urban traffic analysis and protein function annotation	SISTEMAS DE INFORMAÇÃO
Atualização da Monitoração Aeronáutica e Auto-Sustentabilidade	SISTEMAS DE INFORMAÇÃO
Coloração e Infecção em Grafos	TEORIA DA COMPUTAÇÃO
DIAG_Cloud: Diagnóstico de componentes computacionais e análise de dados em nuvens computacionais	SISTEMAS DE INFORMAÇÃO
Failure Analysis and Prediction - FAP	SISTEMAS DE INFORMAÇÃO
Geração de Malhas Através de Particionamento Implícito por Decomposição Espacial Recursiva	CIÊNCIA COMPUTACIONAL
GERAÇÃO E ADAPTAÇÃO DE MALHAS PARA SIMULAÇÃO DE RESERVATÓRIOS	CIÊNCIA COMPUTACIONAL
Gerência de Qualidade de Serviço para Redes de Acesso via Rádio de Quinta Geração com Múltiplas Tecnologias de Acesso e Multiconectividade	SISTEMAS DE INFORMAÇÃO
GERENCIAMENTO DA SAÚDE DE VEÍCULOS AÉREOS NÃO TRIPULADOS	TEORIA DA COMPUTAÇÃO
INCT-Ines: Instituto Nacional de Ciência e Tecnologia para Engenharia de Software (INES)	SISTEMAS DE INFORMAÇÃO
Inteligência Científica e Tecnológica na Segurança Pública	SISTEMAS DE INFORMAÇÃO
Internet das Coisas e dos Dados	SISTEMAS DE INFORMAÇÃO
IoT Data and Personal Privacy in Smart Cities	SISTEMAS DE INFORMAÇÃO
Large?Scale Time Dependent Graphs	SISTEMAS DE INFORMAÇÃO
MAximUM2IoT - Abordagem para Avaliação da Qualidade de Aplicações Ubíquas no cenário de VANETs e IoT	SISTEMAS DE INFORMAÇÃO
Medição e Avaliação de Desempenho em Redes de Computadores	SISTEMAS DE INFORMAÇÃO
METIS: Scientific Investigation on Self-driving Database Systems	SISTEMAS DE INFORMAÇÃO
Modelagem de comportamento contextual de aplicações para Internet das Coisas utilizando programação baseada em gatilhos	SISTEMAS DE INFORMAÇÃO
Modelos, Expressividade, Complexidade e Aplicações	TEORIA DA COMPUTAÇÃO
O patrimônio cultural e artístico de Fortaleza na cultura digital	SISTEMAS DE INFORMAÇÃO

Projeto	Linha de pesquisa
PIBIC 2018/2019 - AdApT - Abordagem para Avaliação de Aplicações no Ambiente de Internet das Coisas	SISTEMAS DE INFORMAÇÃO
Plataforma para diagnóstico rápido utilizando biossensores nanotecnológicos de baixo custo baseado em ondas acústicas superficiais e aptâmeros	SISTEMAS DE INFORMAÇÃO
Privacidade de Dados: Modelos, Mecanismos e Aplicações	SISTEMAS DE INFORMAÇÃO
Problemas de Coloração e Infecção em Grafos	TEORIA DA COMPUTAÇÃO
Problemas de Otimização em Grafos acrescidos de restrições de conflito, imposição e dependência	TEORIA DA COMPUTAÇÃO
Problemas em grafos: complexidade e métodos	TEORIA DA COMPUTAÇÃO
Projeto MC-MemES: Máquina de Consultas para Banco de Dados em Memórias de Estado Sólido	SISTEMAS DE INFORMAÇÃO
Projeto PROBRAL - Graph Coloring: Extremal Combinatorics, Graph Theory and Algorithms	TEORIA DA COMPUTAÇÃO
Replicação Elástica para Banco de Dados em Nuvem	SISTEMAS DE INFORMAÇÃO
Residência em Segurança da Informação - RSI	SISTEMAS DE INFORMAÇÃO
Simulação e Animação Baseadas em Física para Computação Gráfica e Realidade Virtual	CIÊNCIA COMPUTACIONAL
SISTEMA DE RECONHECIMENTO FACIAL DE BAIXO CUSTO	TEORIA DA COMPUTAÇÃO
Subprojeto do INCT-Ines: Descoberta de Padrões de Mobilidade a partir de Trajetórias de Objetos Móveis e Dados Contextuais	SISTEMAS DE INFORMAÇÃO
Tecnologias e soluções para habilitar o paradigma de nuvens de coisas	SISTEMAS DE INFORMAÇÃO
Tools & CAS: Desenvolvimento e evolução de Aplicativos Móveis e Sistemas Web	SISTEMAS DE INFORMAÇÃO
UFC Smart Campus: O Uso de Internet das Coisas no Campus do Pici da Universidade Federal do Ceará	SISTEMAS DE INFORMAÇÃO
Uma Arquitetura descentralizada de suporte à computação em névoa com ênfase no offloading de processamento e dados contextuais	SISTEMAS DE INFORMAÇÃO
Uso de técnicas de aprendizado de máquina para aumento da eficiência do setor de varejo	TEORIA DA COMPUTAÇÃO
UTILIZAÇÃO DA TECNOLOGIA DA INFORMAÇÃO NA ADAPTAÇÃO E CAPACITAÇÃO DE ESTUDANTES AFRICANOS ? TI2EA	SISTEMAS DE INFORMAÇÃO

5.1.2 Computação

Este programa não possui discentes matriculados e reúne 17 docentes. Além disso, conta com 16 disciplinas de Mestrado e Doutorado em vigência.

Abaixo encontra-se a distribuição das áreas de concentração, linhas de pesquisa e projetos de pesquisa deste programa.

5.1.2.1 Área de Concentração Computação

Organiza-se nas seguintes linhas de pesquisa:

- Algoritmos e Teoria da Computação
- Engenharia de Software e Metodologias *Sistemas de Computação

Não existem projetos de pesquisas em andamento para nenhuma das linhas de pesquisa desta área.

5.1.3 Matemática

A programa de Pós-Graduação em Matemática da UFC foi criado em 1965, tendo seu formato atual delineado em 1995 com a criação do Doutorado em Matemática, tendo este programa ao todo 67 discentes matriculados e 31 docentes. Além disso, conta com 85 disciplinas de Mestrado e Doutorado em vigência.

Abaixo encontra-se a distribuição das áreas de concentração, linhas de pesquisa e projetos de pesquisa deste programa.

5.1.3.1 Área de Concentração Análise

Possui apenas a linha de pesquisa Análise que conta com 3 projetos de pesquisa em andamento.

5.1.3.2 Área de Concentração Combinatória

Possui apenas a linha de pesquisa Combinatória que conta com 2 projetos de pesquisa em andamento.

5.1.3.3 Área de Concentração Geometria Diferencial

Organiza-se nas seguintes linhas de pesquisa:

- Análise Geométrica, que possui 3 projetos de pesquisa em andamento
- Geometria Diferencial, que possui 4 projetos de pesquisa em andamento

5.1.3.4 Área de Concentração Matemática.

Não possui linhas de pesquisa e projetos de pesquisa.

5.1.3.5 Área de Concentração Topologia e Singularidades

Organiza-se nas seguintes linhas de pesquisa:

- Singularidades, que possui 2 projetos de pesquisa em andamento
- Sistemas Dinâmicos e Teoria Ergódica, que possui 2 projetos de pesquisa em andamento
- Topologia, que possui 1 projeto de pesquisa em andamento

Projeto	Linha de pesquisa

5.1.3.6 Área de Concentração Álgebra

Possui apenas a linha de pesquisa Álgebra que conta com 2 projetos de pesquisa em andamento.

Ao todo são 20 programas de pesquisa para Matemática. São esses:

Projeto	Linha de pesquisa
Análise de Equações Diferencias Parciais	ANÁLISE
Análise Qualitativa de Equações Diferenciais Parciais Não-Lineares	ANÁLISE
Análse Geométrica e Teoria de Singularidade em Espaços Estrafica	ANÁLISE GEOMÉTRICA
Aspectos Analíticos e Geométricos de EDPs Não Lineares e Aplicações	ANÁLISE
Coloração Backbone, Coloração Gulosa e Convexidade Cíclica	COMBINATÓRIA
Decay of correlations and statistical properties of high-dimensional dynamical systems	SISTEMAS DIN MICOS E TEORIA ERGÓDICA
Discriminante de corpos abelianos e reticulados.	ÁLGEBRA
Dualidade e localização em geometria diferencial e na teoria de índice	GEOMETRIA DIFERENCIAL
Fluxos geométricos e área renormalizada em espaços estratificados	ANÁLISE GEOMÉTRICA
Geometria Diferencial de Subvariedades	GEOMETRIA DIFERENCIAL
Geometria e análise não linear em espaços singulares e aplicações	PROJETO ISOLADO
Massa em Relatividade Geral	GEOMETRIA DIFERENCIAL
Núcleo de Análise Geométrica, Singularidades e Combinatória da PGMAT-UFC.	ANÁLISE GEOMÉTRICA
Pesquisa em Singularidades e Topologia	SINGULARIDADES
Singularidades	SINGULARIDADES
Solitons de Ricci, métrica CPE e Variedades Quase-Einstein	GEOMETRIA DIFERENCIAL
Symbolic dynamics and non-uniform hyperbolicity	SISTEMAS DIN MICOS E TEORIA ERGÓDICA
Teoria dos Grafos e Combinatória Extremal.	COMBINATÓRIA
Tópicos de Álgebras Não-Associativas	ÁLGEBRA
Topologia e Singularidades	TOPOLOGIA

6 Entendimento dos Dados

Seguindo-se a metodologia CRISP-DM, após o entendimento do universo que se deseja estudar, o trabalho agora se propõe a fazer as análises de dados retirados da plataforma Unb eLattes, com posterior interpretação das análises e discussão dos resultados.

6.1 Coleta Inicial dos Dados

Os dados a serem analisados foram coletados através da plataforma UnB eLattes, refletindo o período de 2014 a 2018 (correspondente ao período da última avaliação da CAPES). Os dados são referentes aos pesquisadores identificadas durante a fase de entendimento do negócio desse trabalho.

Através da plataforma eLattes, foi feito o download dos arquivos json gerados, contendo os dados brutos que serão analisados. Serão analisados os arquivos *advise.json*, *profile.json* e *publication.json*. Esses arquivos serão estudados para cada um dos três programas de pós-graduação que estão sendo estudados nesse trabalho.

Uma parte vital no CRISP-DM é o correto entendimento dos dados que estão sendo trabalhados, processo que deve ser realizado antes de uma análise mais aprofundada dos mesmos.

De início, serão utilizadas as seguintes bibliotecas para entendimento e análise dos dados da plataforma eLattes:

```
library(jsonlite)
library(readr)
library(ggplot2)
library(stringr)
library(tidyr)
library(readxl)
library(listviewer)
library(dplyr)
library(scales)
library(tm)
library(wordcloud)
```

Primeiramente, será carregado o script (fornecido pelos professores) para transformar listas em data frames:

```
source("elattes.ls2df.R")
```

6.1.1 Dados de perfil

A análise e entendimento dos dados será inicializado pelo estudo do arquivo JSON correspondente aos perfis individuais de cada pesquisador, em cada um dos três programas de pós-graduação em estudo. Os arquivos serão importados para as seguintes variáveis:

```
cic_profile <- jsonlite::fromJSON("cic_profile.json")
comp_profile <- jsonlite::fromJSON("comp_profile.json")
mat_profile <- jsonlite::fromJSON("mat_profile.json")
```

A quantidade de docentes que serão analisados em cada um dos três programas de pós-graduação são:

```
  length(cic_profile)
```

```
## [1] 30
```

```
  length(comp_profile)
```

```
## [1] 15
```

```
  length(mat_profile)
```

```
## [1] 27
```

6.1.2 Dados de orientações

```
cic_advise <- jsonlite::fromJSON("cic_advise.json")
comp_advise <- jsonlite::fromJSON("comp_advise.json")
mat_advise <- jsonlite::fromJSON("mat_advise.json")
```

Pode-se entender melhor os dados das orientações dentro desse conjunto de dados utilizando-se a função names, que possibilita investigar quais são as colunas de cada uma das variáveis criadas, como por exemplo:

```
names(cic_advise)
```

```
## [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"
```

6.1.3 Dados das publicações

```
cic_publication <- jsonlite::fromJSON("cic_publication.json")
comp_publication <- jsonlite::fromJSON("comp_publication.json")
mat_publication <- jsonlite::fromJSON("mat_publication.json")
```

7 Análise dos Dados

De posse dos dados e com um entendimento melhor da maneira que são estruturados, é possível prosseguir para sua análise. A seguir serão analisados separadamente os três programas de Pós-Graduação sob análise.

7.1 Pós-Graduação em Ciência da Computação

7.1.1 Orientações

7.1.1.1 Números de orientações completas por ano

```
for (i in 6:9){
  print(names(cic_advise[i]))
  print(cic_advise[[i]] %>%
          sapply(function(x)
            length(x$ano)))
}
```

```
## [1] "ORIENTACAO_CONCLUIDA_POS_DOUTORADO"
## 2014 2015 2016 2017 2018
##    1    0    2    0    0
## [1] "ORIENTACAO_CONCLUIDA_DOUTORADO"
## 2014 2015 2016 2017 2018
##    7    6   13   10   11
## [1] "ORIENTACAO_CONCLUIDA_MESTRADO"
## 2014 2015 2016 2017 2018
##   22   18   28   22   22
## [1] "OUTRAS_ORIENTACOES_CONCLUIDAS"
## 2014 2015 2016 2017 2018
##   23   24   39   47   29
```

A partir de algumas análises preliminares na base de dados, é possível perceber que o número de orientações completas por ano é sempre maior para outros tipos de orientações concluídas (que não mestrado, doutorado e pós-doutorado), e que dessas três, o número de orientações concluídas de mestrado é sempre maior.

7.1.1.2 Disposição de orientações por situação e ano

```
for (i in 1:length(cic_advise)){
  print(names(cic_advise[i]))
  print(cic_advise[[i]] %>%
          sapply(function(x)
            length(x$ano)) )
}
```

```
## [1] "ORIENTACAO_EM_ANDAMENTO_DE_POS_DOUTORADO"
## 2014 2015 2016 2017 2018
##    0    0    1    2    0
## [1] "ORIENTACAO_EM_ANDAMENTO_DOUTORADO"
## 2014 2015 2016 2017 2018
##   18   11   13   18   14
## [1] "ORIENTACAO_EM_ANDAMENTO_MESTRADO"
## 2014 2015 2016 2017 2018
##    1    3   11   27   36
## [1] "ORIENTACAO_EM_ANDAMENTO_GRADUACAO"
## 2014 2015 2016 2017 2018
##    0    0    0    0    0
## [1] "ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA"
## 2014 2015 2016 2017 2018
##    0    0    0    1   13
## [1] "ORIENTACAO_CONCLUIDA_POS_DOUTORADO"
## 2014 2015 2016 2017 2018
##    1    0    2    0    0
## [1] "ORIENTACAO_CONCLUIDA_DOUTORADO"
## 2014 2015 2016 2017 2018
##    7    6   13   10   11
## [1] "ORIENTACAO_CONCLUIDA_MESTRADO"
## 2014 2015 2016 2017 2018
##   22   18   28   22   22
## [1] "OUTRAS_ORIENTACOES_CONCLUIDAS"
## 2014 2015 2016 2017 2018
##   23   24   39   47   29
```

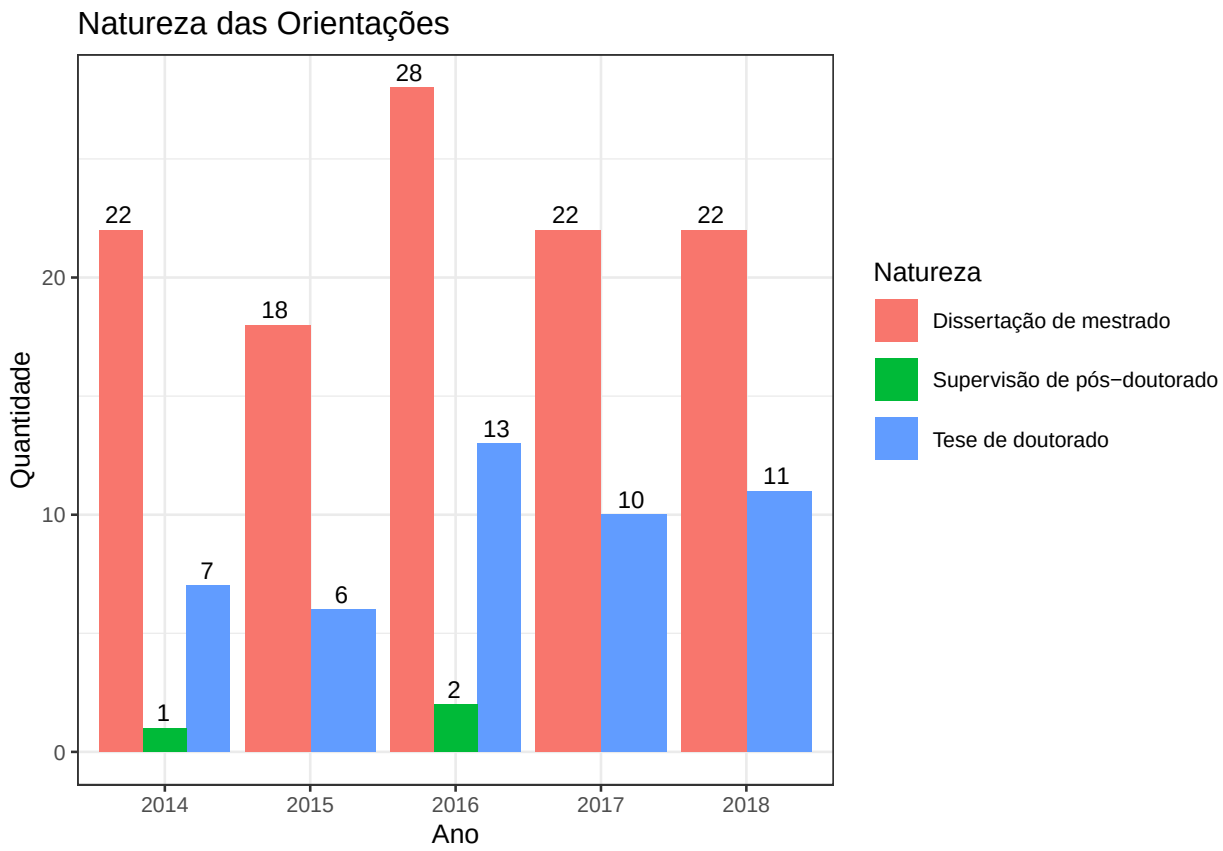
7.1.1.3 Natureza das orientações

```
cic_public.periodico.df <- pub.ls2df(cic_publication, 1) #artigos
cic_public.livros.df <- pub.ls2df(cic_publication, 2) #livros
cic_public.textojornais.df <- pub.ls2df(cic_publication, 4) #textos em jornais
cic_public.eventos.df <- pub.ls2df(cic_publication, 5) #eventos
```

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

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

```
ggplot(cic_orient.df,aes(ano, fill=factor(natureza))) +
  geom_bar(stat = "count",position='dodge') +
  ggtitle("Natureza das Orientações") +
  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")+labs(fill="Natureza")+theme_bw(base_size = 10)+
  geom_text(hjust=0.6,
            vjust=-0.4,
            size=3,color='black',
            position = position_dodge(width=0.9),
            stat = "count", aes(group=factor(natureza),
                                label=formatC(..count.., big.mark=",")),
            check_overlap = TRUE)
```



O gráfico acima permite perceber que a quantidade de dissertações de mestrado é sempre maior ano a ano. A quantidade de supervisões de pós-doutorado é bem menor e não há resultados registrados para os anos de 2017 e 2018.

7.1.2 Perfil

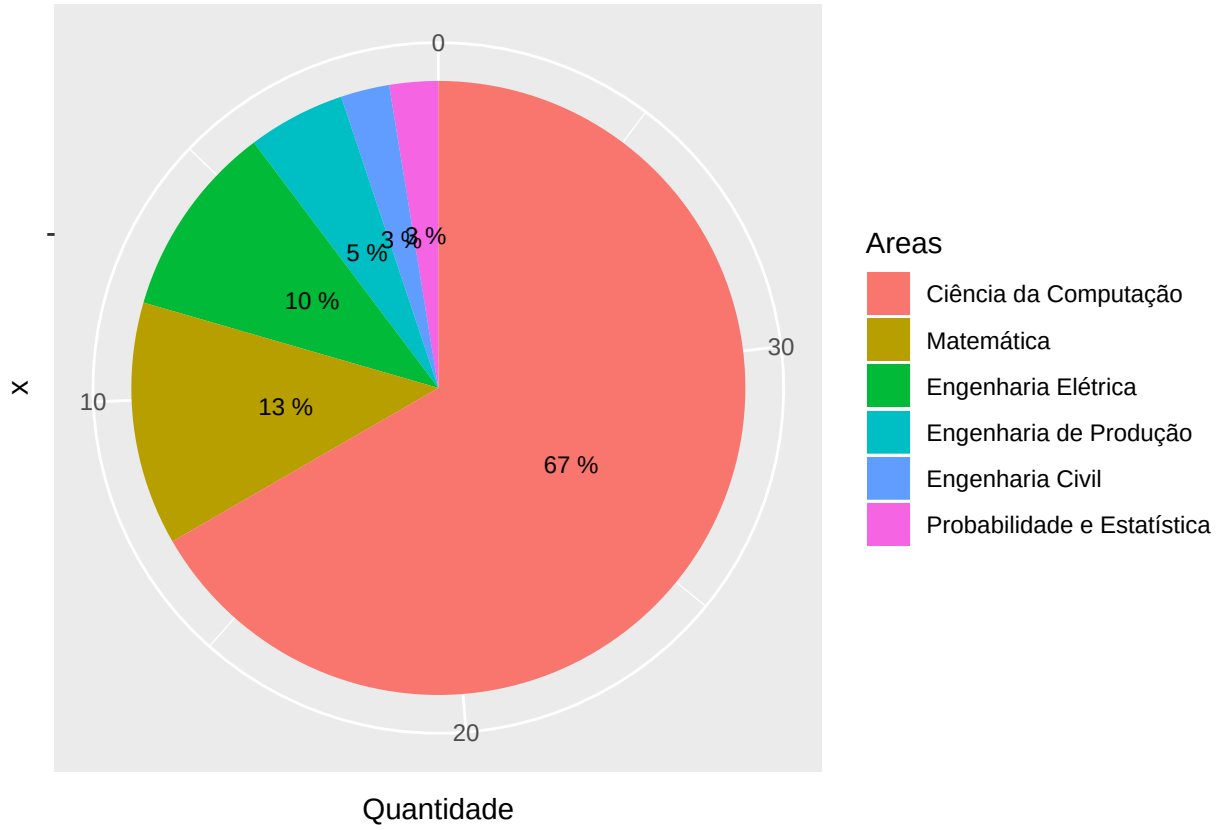
7.1.2.1 Distribuição de áreas de atuação dos pesquisadores

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

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

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

ggplot(areas_atuacao_cic, aes(x="", y=Quantidade, fill=Areas))+
  geom_bar(width = 1, stat = "identity") +
  coord_polar("y", start = 0, direction = -1) +
  geom_text(data = areas_atuacao_cic,
            aes(x ="", y=Quantidade, label = paste(Porcentagem, "%")),
            position = position_stack(vjust = 0.5), size=3)
```



Percebe-se que a área de atuação com maior recorrência é a *Ciência da Computação* o que faz todo sentido já que se trata do programa de Ciência da Computação. Percebe-se também uma grande porcentagem em relação a área *Matemática* o que pode indicar que no programa de Matemática existem grandes chances de a área de Ciência da Computação ter uma porcentagem expressiva também.

7.1.2.2 Distribuição de subáreas de atuação mais frequentes dos pesquisadores

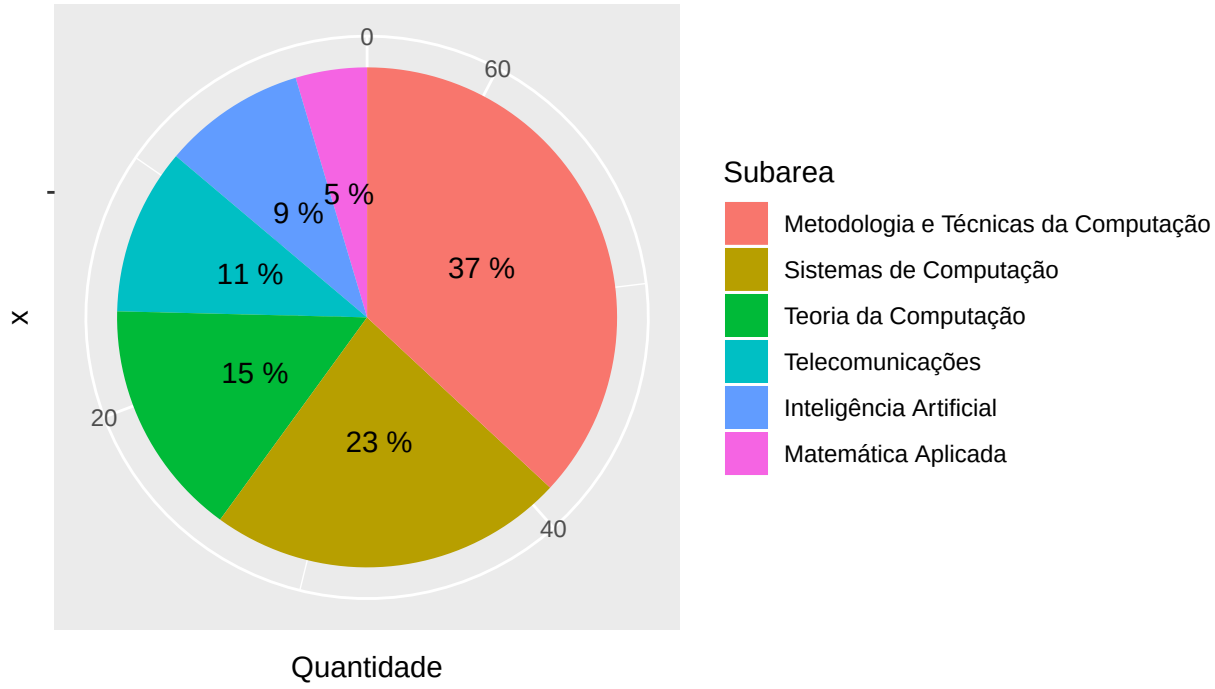
```
subarea <- cic_profile %>%
  sapply(function(x) (x$areas_de_atuacao$sub_area)) %>%
  unlist() %>% table() %>%
  sort(decreasing = TRUE) %>%   as.data.frame() %>%
  filter(!. == "") %>% head(6)

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))
```



A partir desses dados, é fácil perceber que a sub-área de *Metodologia e Técnicas de Computação* é que possui a maior frequência de atuação, seguida por *Teoria da Computação* e *Sistemas de Computação*.

7.1.2.3 Distribuição de especialidades mais frequentes dos pesquisadores

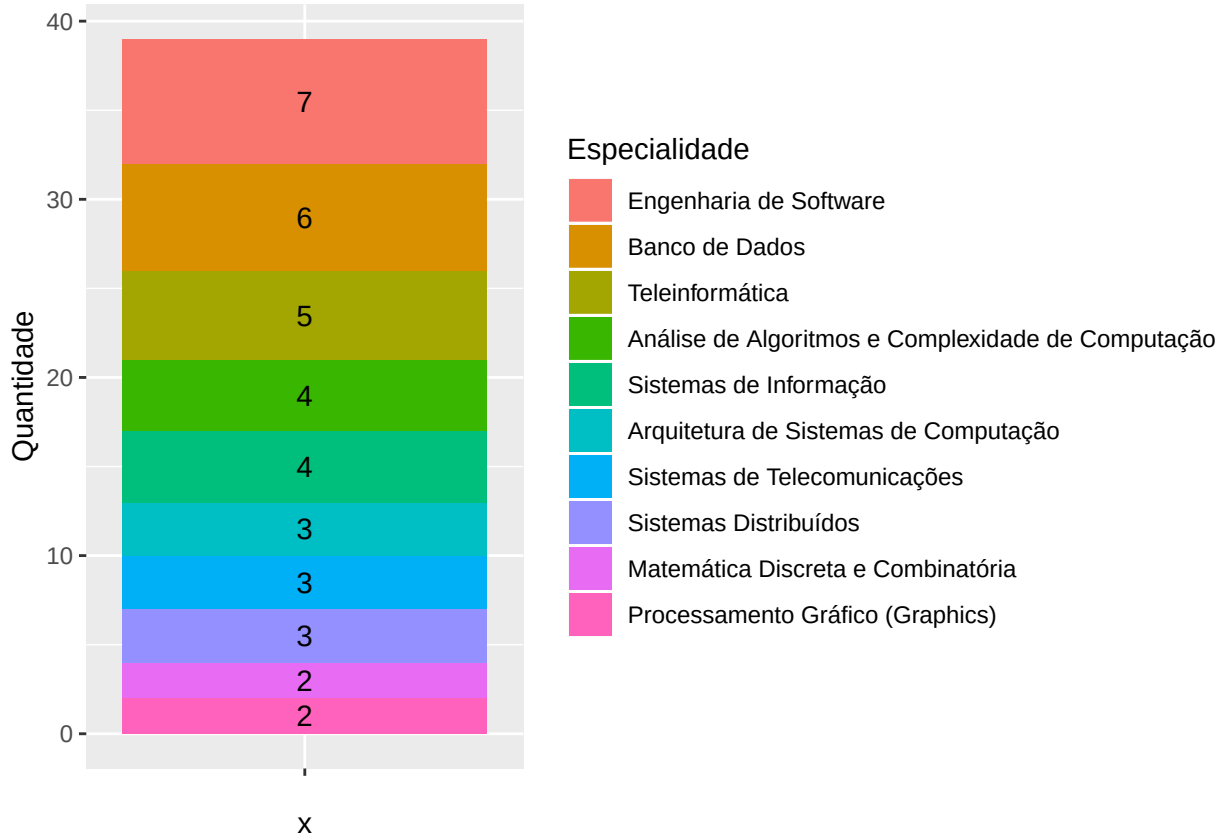
```
especialidades_frequentes <- cic_profile %>%
  sapply(function(x) unique(x$areas_de_atuacao$especialidade)) %>%
  unlist() %>% table() %>% sort(decreasing = TRUE) %>%
  as.data.frame() %>% filter(!. == "") %>% head(10)

quantidade <- sum(especialidades_frequentes$Freq)

especialidades_frequentes <- mutate(especialidades_frequentes, percent = round(especialidades_frequentes

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))
```



Aqui vê-se as 10 especialidades mais recorrentes e percebe-se uma boa distribuição da quantidade entre elas.

7.1.3 Publicações

7.1.3.1 Quantidade de Publicações por tipo

```
for (i in 1:length(cic_publication)){
  print(names(cic_publication[i]))
  print(cic_publication[[i]] %>%
          sapply(function(x)
            length(x$ano)) %>% sum())
}
```

```
## [1] "PERIODICO"
## [1] 175
## [1] "LIVRO"
## [1] 15
## [1] "CAPITULO_DE_LIVRO"
## [1] 29
## [1] "TEXTO_EM_JORNAIS"
## [1] 1
## [1] "EVENTO"
## [1] 556
## [1] "ARTIGO_ACEITO"
## [1] 6
## [1] "DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA"
## [1] 7
```

Percebe-se que o tipo mais recorrente de publicações é do tipo Evento, possui bem mais ocorrências do que o segundo colocado, Periódico.

7.1.3.2 Quantidade de publicações por tipo e por ano

```
publication_tipo_cic <- cic_publication %>%
  sapply(function(x)
    sapply(x, function(x)
      length(x$autores)))
publication_tipo_cic
```

```
##      PERIODICO LIVRO CAPITULO_DE_LIVRO TEXTO_EM_JORNAIS EVENTO
## 2014        22     0                 3                0    101
## 2015        31     4                 5                1    103
## 2016        40     2                 7                0    122
## 2017        49     5                 6                0    126
## 2018        33     4                 8                0    104
##      ARTIGO_ACEITO DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA
## 2014             1                                      3
## 2015             0                                      2
## 2016             0                                      0
## 2017             1                                      2
## 2018             4                                      0
```

Destaca-se que nos anos de 2016 e 2017 houveram mais publicações de Periódicos e Eventos, havendo uma queda dessas publicações no ano de 2018, porém este ano apresentou um aumento de Artigos Aceitos e Capítulos de Livro.

7.1.3.3 Participação em eventos por país

```
cic_publication$EVENTO %>%
  sapply(function(x)
    (x$pais_do_evento)) %>%
  unlist() %>% table() %>%
  sort(decreasing = TRUE)
```

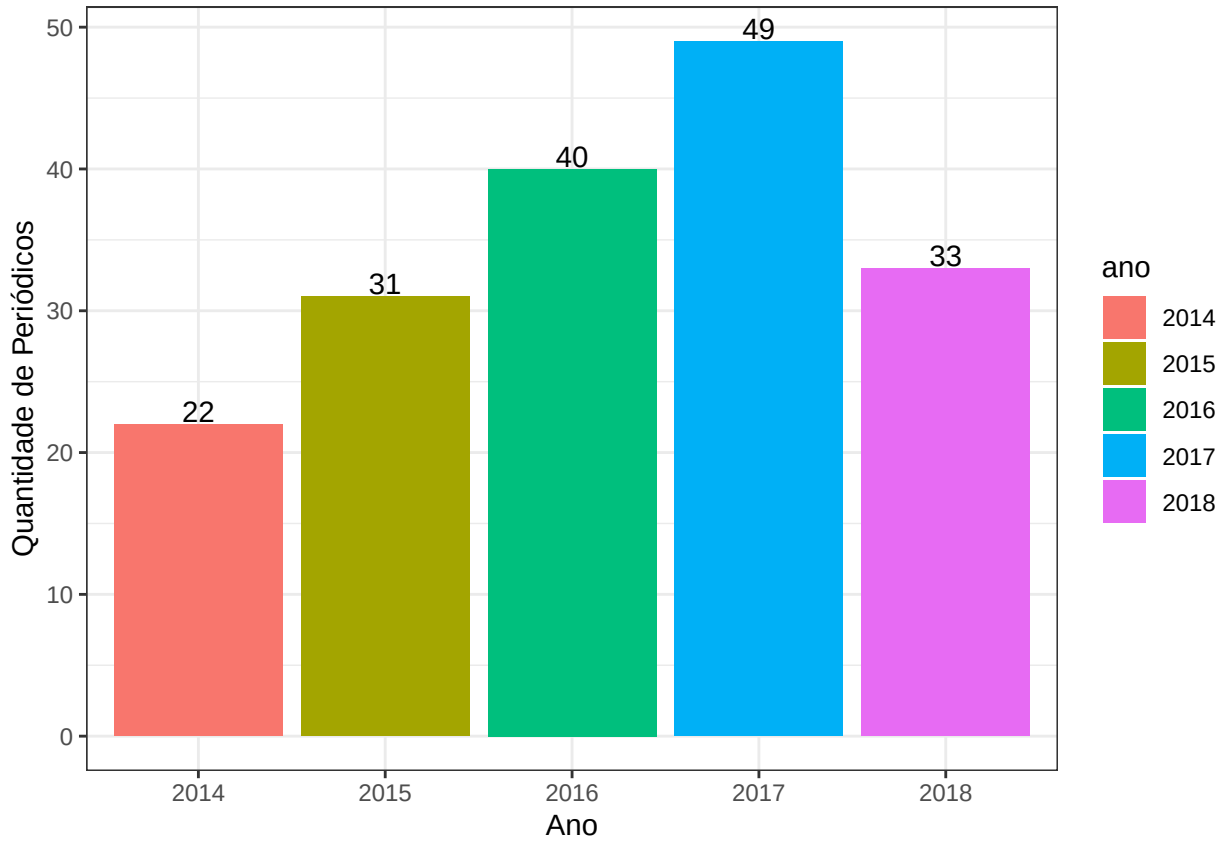
```
## .
##           Brasil   Estados Unidos           Itália          Espanha
##              427               43               14               11
##         Portugal           Canadá           França         Alemanha
##               10                7                5                4
##          Turquia          Uruguai        Argentina          Bélgica
##                4                3                2                2
##            Chile    Coréia do Sul          Holanda            Japão
##                2                2                2                2
##         Marrocos República Tcheca        Austrália          Austria
##                2                2                1                1
##            China         Colômbia  Emirados Árabes     Grã-Bretanha
##                1                1                1                1
##           Grécia       Iugoslávia            Malta             Peru
##                1                1                1                1
##          Polônia           Suécia
##                1                1
```

Dentro da Pós-Graduação de Ciência da Computação, a presença em eventos é maior no Brasil, embora exista um grande número de participações em eventos também nos Estados Unidos (com 43 participações), e também na Itália, Espanha e Portugal (embora em menor quantidade).

É interessante observar que este programa possui uma grande participação em eventos em outros países e em vários continentes.

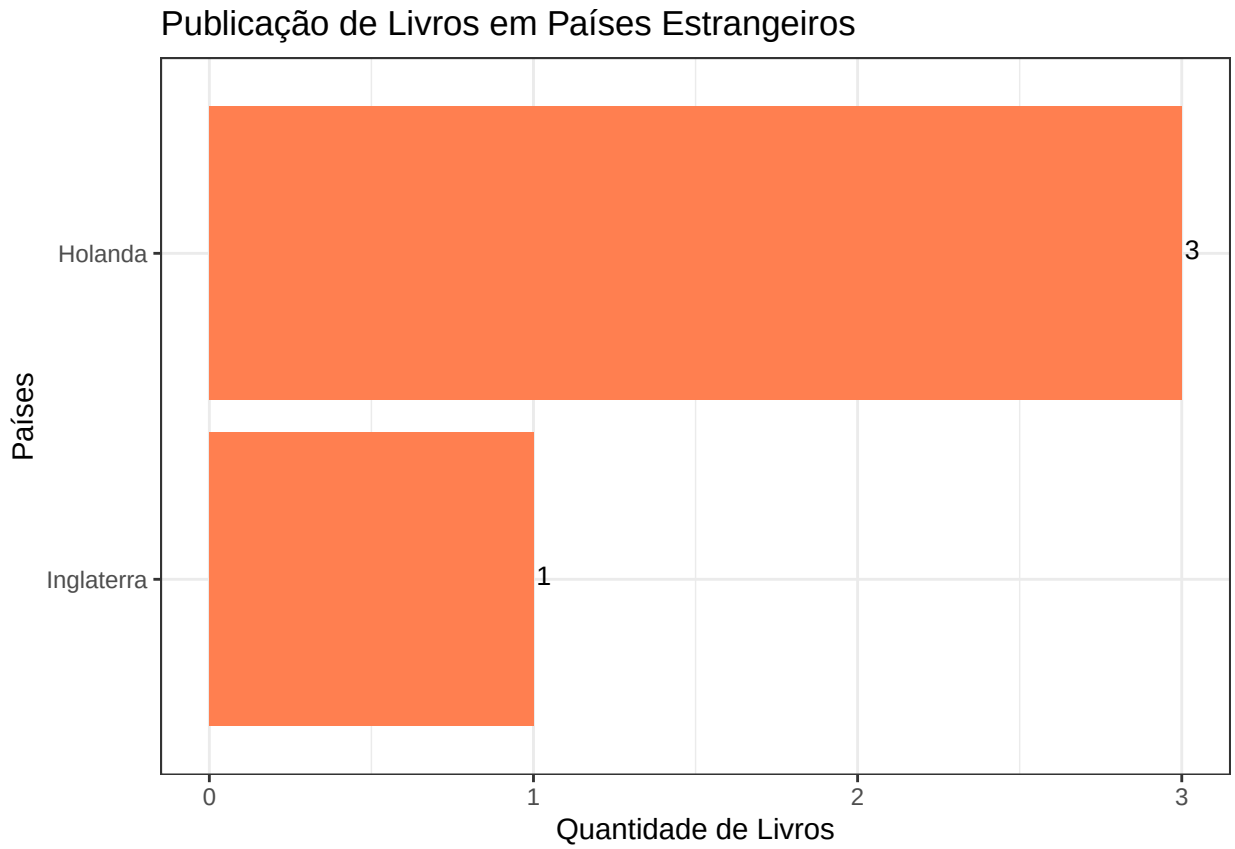
7.1.3.4 Publicações por ano

```
cic_public.periodico.df %>%
  ggplot(aes(x = ano)) + geom_bar(aes(fill=ano)) +
  geom_text(stat = "count", aes(label=formatC(..count.., big.mark=",")),vjust=-0.1) +
  theme_bw()+labs(x="Ano",y="Quantidade de Periódicos") +
  scale_y_continuous(labels = comma)
```



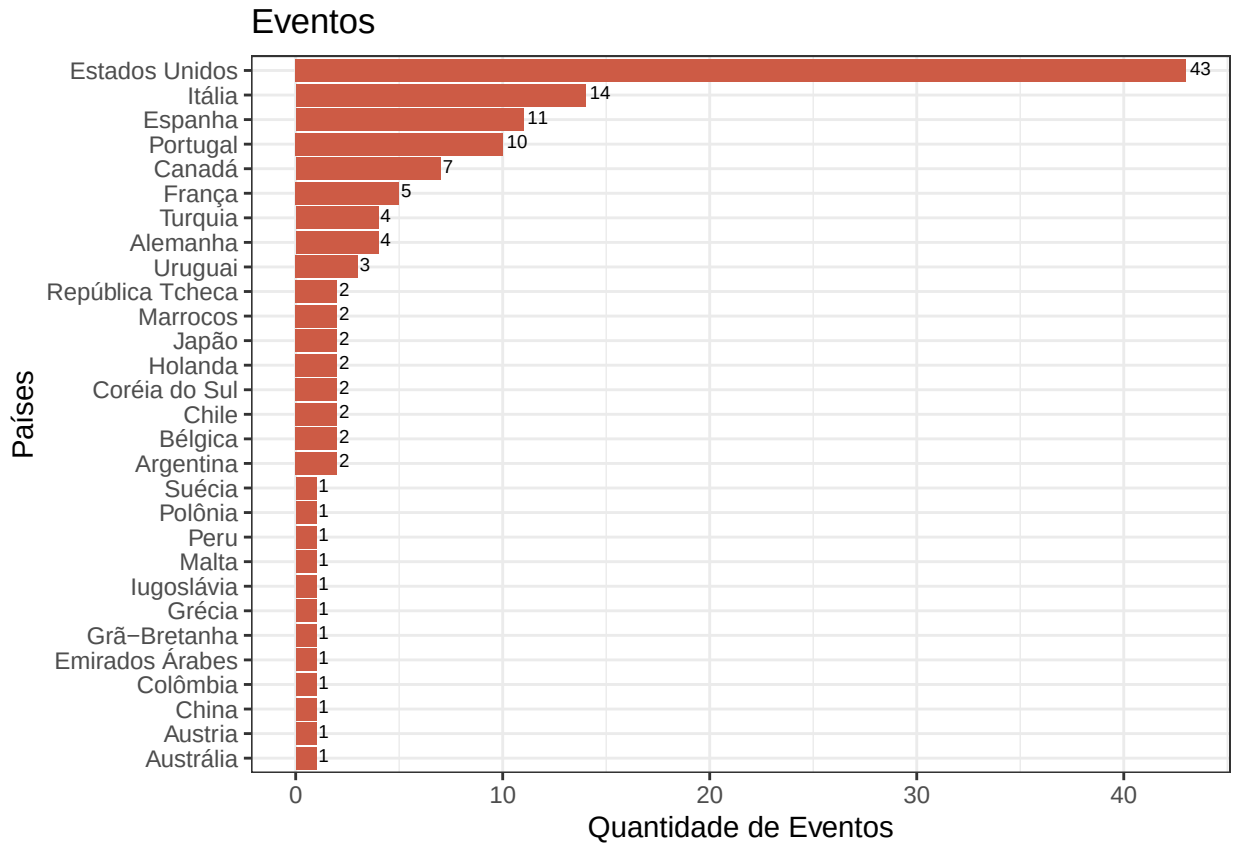
7.1.3.5 Publicações de livros fora do Brasil

```
cic_public.livros.df %>%
  group_by(pais_de_publicacao) %>%
  summarise(Quantidade = n()) %>%
  filter(pais_de_publicacao != "Brasil") %>%
  ggplot(aes(x = reorder(pais_de_publicacao, (Quantidade)), y = Quantidade)) +
  geom_col(fill = "coral")+
  geom_text(aes(label=comma(Quantidade)),
            hjust=-0.2,
            vjust=0.3,
            size=3.5) + coord_flip() +
  labs(title = "Publicação de Livros em Países Estrangeiros", x = "Países", y = "Quantidade de Livros")+
  theme_bw()
```



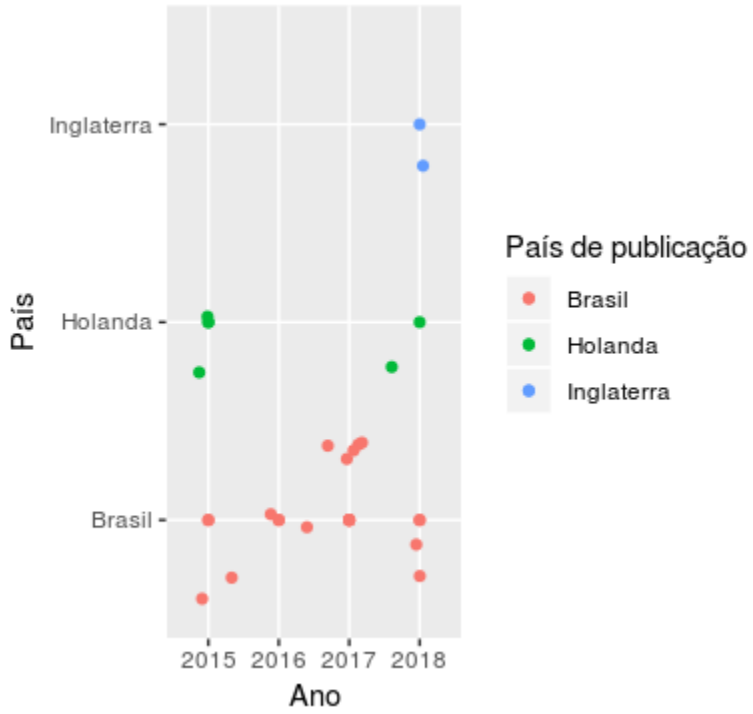
7.1.3.6 Eventos por País

```
cic_public.eventos.df %>%
  group_by(pais_do_evento) %>%
  summarise(Quantidade = n()) %>%
  filter(pais_do_evento != "Brasil") %>%
  ggplot(aes(x = reorder(pais_do_evento, (Quantidade)), y = Quantidade)) +
  geom_col(fill = "coral3")+
  geom_text(aes(label=comma(Quantidade)),
            hjust=-0.2,
            vjust=0.3,
            size=2.5)+ coord_flip() +
  labs(title = "Eventos", x = "Países", y = "Quantidade de Eventos") + theme_bw()
```



Algumas análises básicas também permitem perceber que os Estados Unidos são o país em que os docentes da Pós-Graduação em Ciência da Computação mais participam de eventos, seguido (em quantidade bem menore), pela Itália e Espanha. Nesse gráfico foram mostrados apenas os eventos acontecidos fora do país.

7.1.3.7 Publicações de livros por ano agrupados por país



O maior número de publicações de livros acontece, obviamente, no Brasil. Mas percebe-se, nos anos de 2015 e 2018, que os docentes da Pós-Graduação em Ciência da Computação publicaram livros na Holanda e, especificamente em 2018, também na Inglaterra.

7.2 Pós-Graduação em Computação

7.2.1 Orientações

7.2.1.1 Números de orientações completas por ano

```
for (i in 6:9){
  print(names(comp_advise[i]))
  print(comp_advise[[i]] %>%
          sapply(function(x)
            length(x$ano)))
}
```

```
## [1] "ORIENTACAO_CONCLUIDA_POS_DOUTORADO"
## 2014 2015 2016 2017 2018
##    0    0    0    0    0
## [1] "ORIENTACAO_CONCLUIDA_DOUTORADO"
## 2014 2015 2016 2017 2018
##    0    1    0    0    2
## [1] "ORIENTACAO_CONCLUIDA_MESTRADO"
## 2014 2015 2016 2017 2018
##    0    4    4    2    7
## [1] "OUTRAS_ORIENTACOES_CONCLUIDAS"
## 2014 2015 2016 2017 2018
##   35   34   48   33   27
```

Percebe-se que existem bem mais orientações concluídas do tipo Outras Orientações, havendo poucas orientações de Mestrado, Doutorado e nenuma de Pós Doutorado nos anos analisados.

É possível também notar que cresce o número de orientações completas por ano para o período estudado.

7.2.1.2 Disposição de orientações por situação e ano

```
for (i in 1:length(comp_advise)){
  print(names(comp_advise[i]))
  print(comp_advise[[i]] %>%
          sapply(function(x)
            length(x$ano)) )
}
```

```
## [1] "ORIENTACAO_EM_ANDAMENTO_DE_POS_DOUTORADO"
## 2014 2015 2016 2017 2018
##    0    0    0    0    0
## [1] "ORIENTACAO_EM_ANDAMENTO_DOUTORADO"
## 2014 2015 2016 2017 2018
##    0    0    2    0    3
## [1] "ORIENTACAO_EM_ANDAMENTO_MESTRADO"
## 2014 2015 2016 2017 2018
##    0    2    3    8    6
## [1] "ORIENTACAO_EM_ANDAMENTO_GRADUACAO"
## 2014 2015 2016 2017 2018
##    0    1    0    3    5
## [1] "ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA"
## 2014 2015 2016 2017 2018
##    0    0    0    0    2
## [1] "ORIENTACAO_CONCLUIDA_POS_DOUTORADO"
## 2014 2015 2016 2017 2018
##    0    0    0    0    0
## [1] "ORIENTACAO_CONCLUIDA_DOUTORADO"
## 2014 2015 2016 2017 2018
##    0    1    0    0    2
## [1] "ORIENTACAO_CONCLUIDA_MESTRADO"
## 2014 2015 2016 2017 2018
##    0    4    4    2    7
## [1] "OUTRAS_ORIENTACOES_CONCLUIDAS"
## 2014 2015 2016 2017 2018
##   35   34   48   33   27
```

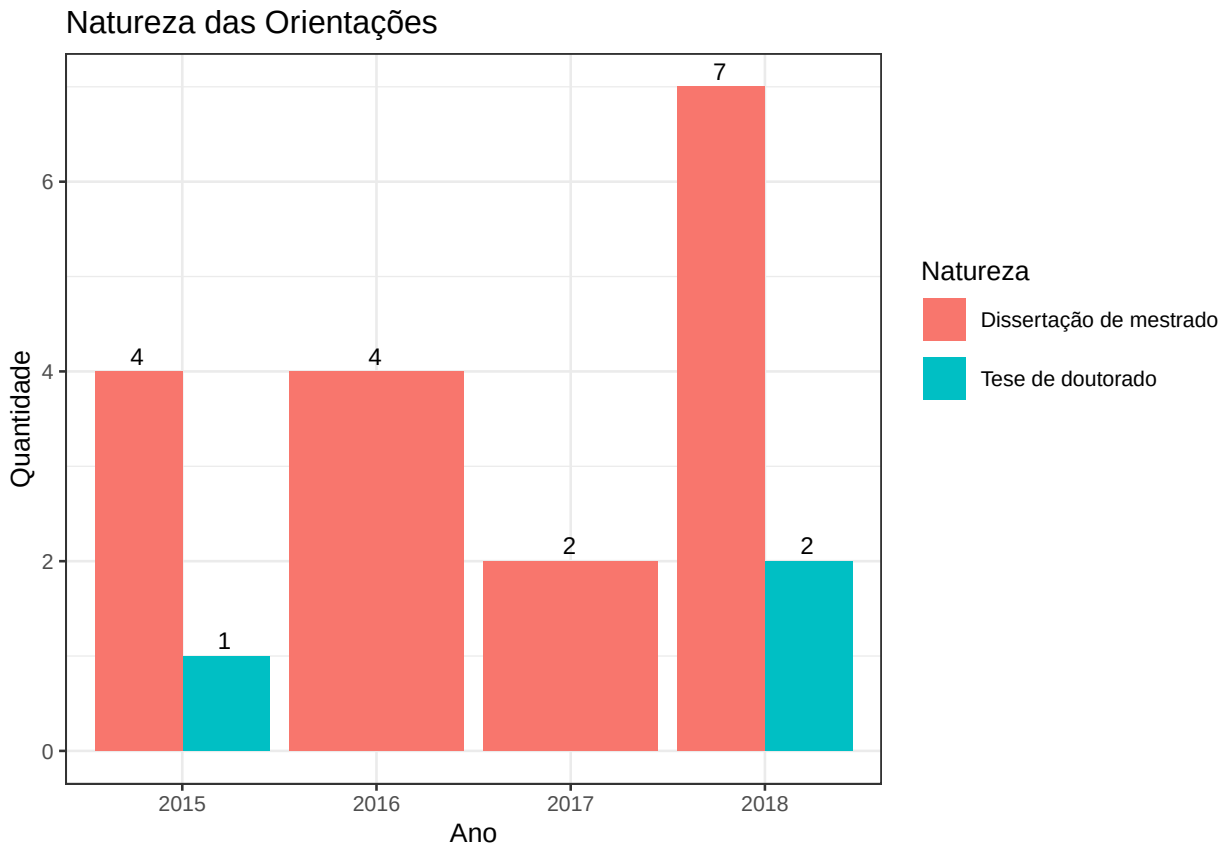
7.2.1.3 Natureza das orientações

```
comp_public.periodico.df <- pub.ls2df(comp_publication, 1) #artigos
comp_public.livros.df <- pub.ls2df(comp_publication, 2) #livros
comp_public.textojornais.df <- pub.ls2df(comp_publication, 4) #textos em jornais
comp_public.eventos.df <- pub.ls2df(comp_publication, 5) #eventos
```

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

```
comp_orient.df <- rbind(rbind(comp_orient.posdoutorado.df, comp_orient.doutorado.df), comp_orient.mestra
```

```
ggplot(comp_orient.df,aes(ano, fill=factor(natureza))) +
  geom_bar(stat = "count",position='dodge') +
  ggtitle("Natureza das Orientações") +
  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")+labs(fill="Natureza")+theme_bw(base_size = 10)+
  geom_text(hjust=0.6,
            vjust=-0.4,
            size=3,
            color='black',
            position = position_dodge(width=0.9),
            stat = "count",
            aes(group=factor(natureza),
                label=formatC(..count.., big.mark=",")),
            check_overlap = TRUE)
```



7.2.2 Perfil

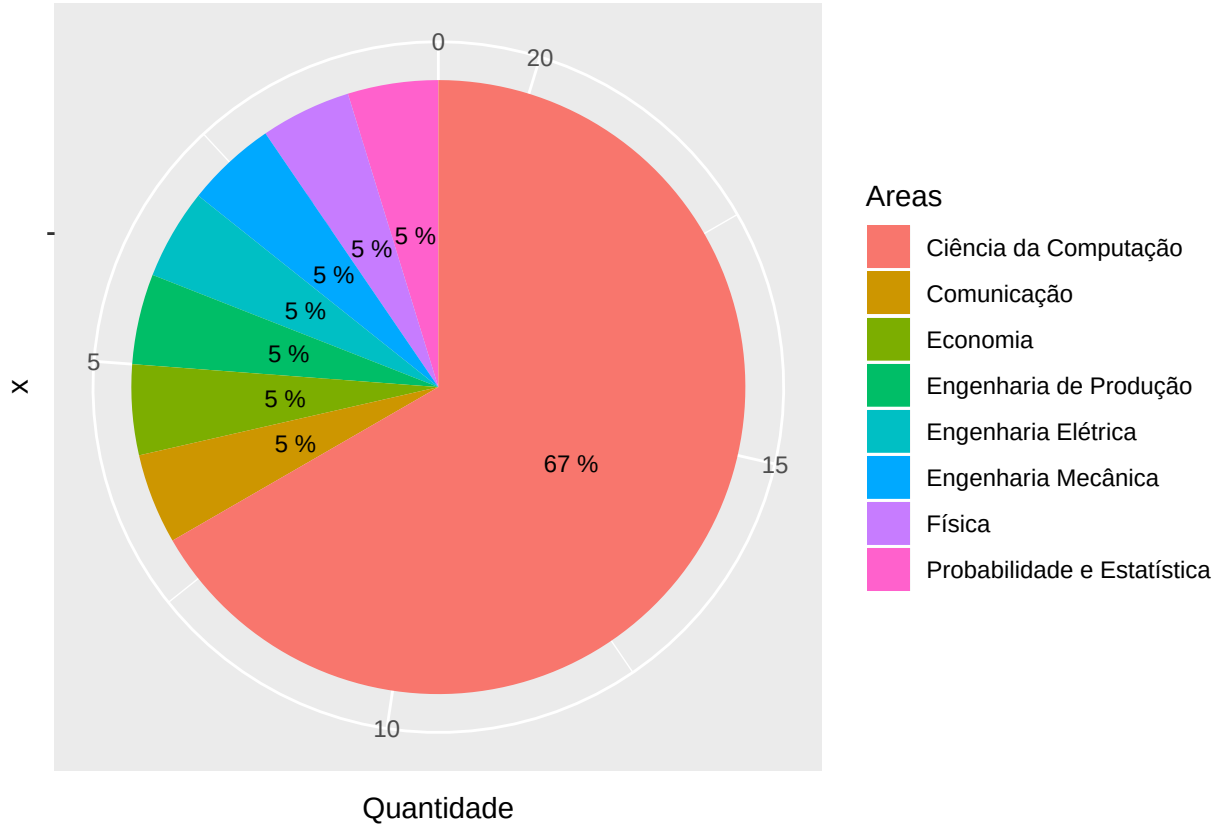
7.2.2.1 Distribuição de áreas de atuação dos pesquisadores

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

quantidade <- sum(areas_atuacao_comp$Freq)
areas_atuacao_comp <- mutate(areas_atuacao_comp, percent = round(areas_atuacao_comp$Freq/quantidade * 1

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

ggplot(areas_atuacao_comp, aes(x="", y=Quantidade, fill=Areas))+
  geom_bar(width = 1, stat = "identity") +
  coord_polar("y", start = 0, direction = -1) +
  geom_text(data = areas_atuacao_comp,
            aes(x ="", y=Quantidade, label = paste(Porcentagem, "%")),
            position = position_stack(vjust = 0.5), size=3)
```



Percebe-se que a área de atuação com maior recorrência é a *Ciência da Computação* o que faz todo sentido já que se trata de um programa de Computação. Percebe-se também uma grande porcentagem em relação às demais áreas existe uma distribuição bem uniforme, demonstrando que o ponto forte desse programa de fato são assuntos relacionados à Computação e Ciência da Computação.

7.2.2.2 Distribuição de subáreas de atuação mais frequentes dos pesquisadores

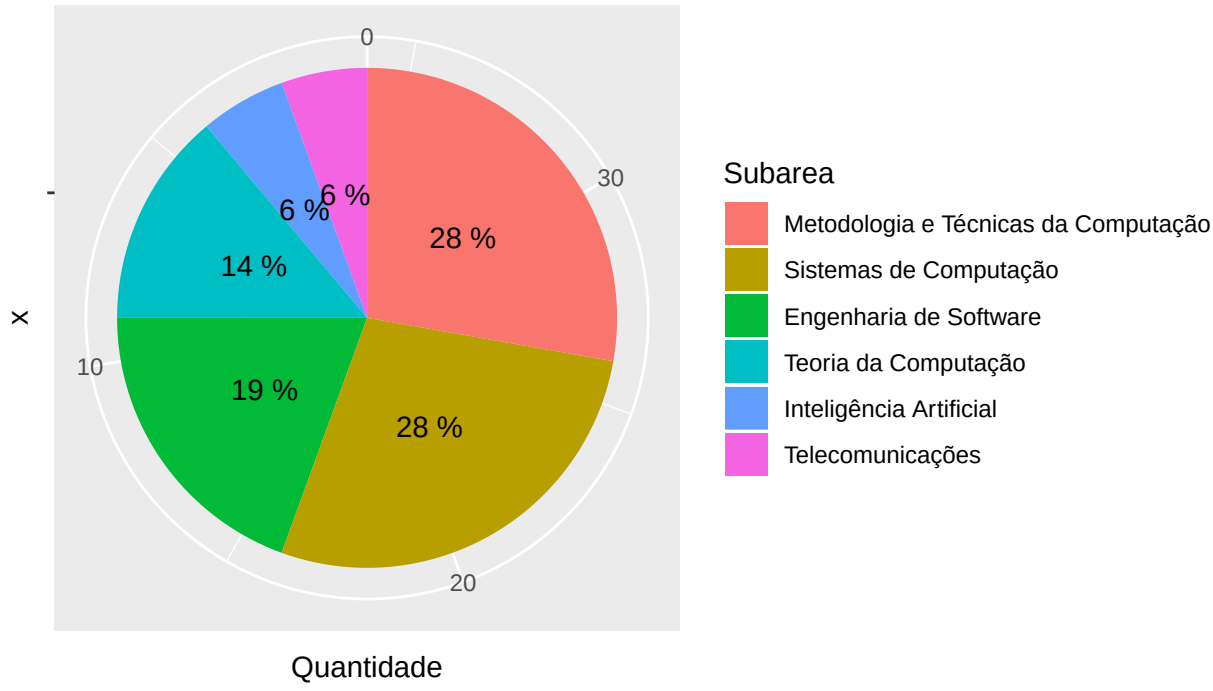
```r
subarea <- comp_profile %>%
  sapply(function(x) (x$areas_de_atuacao$sub_area)) %>%
  unlist() %>% table() %>%
  sort(decreasing = TRUE) %>%   as.data.frame() %>%
  filter(!. == "") %>% head(6)

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))
```



Através da análise acima, pode-se ver que a área com maior número de docentes atuantes é em *Sistemas de Computação*, seguida de perto por *Metodologia e Técnicas de Computação* e *Teoria da Computação*.

7.2.2.3 Distribuição de especialidades mais frequentes dos pesquisadores

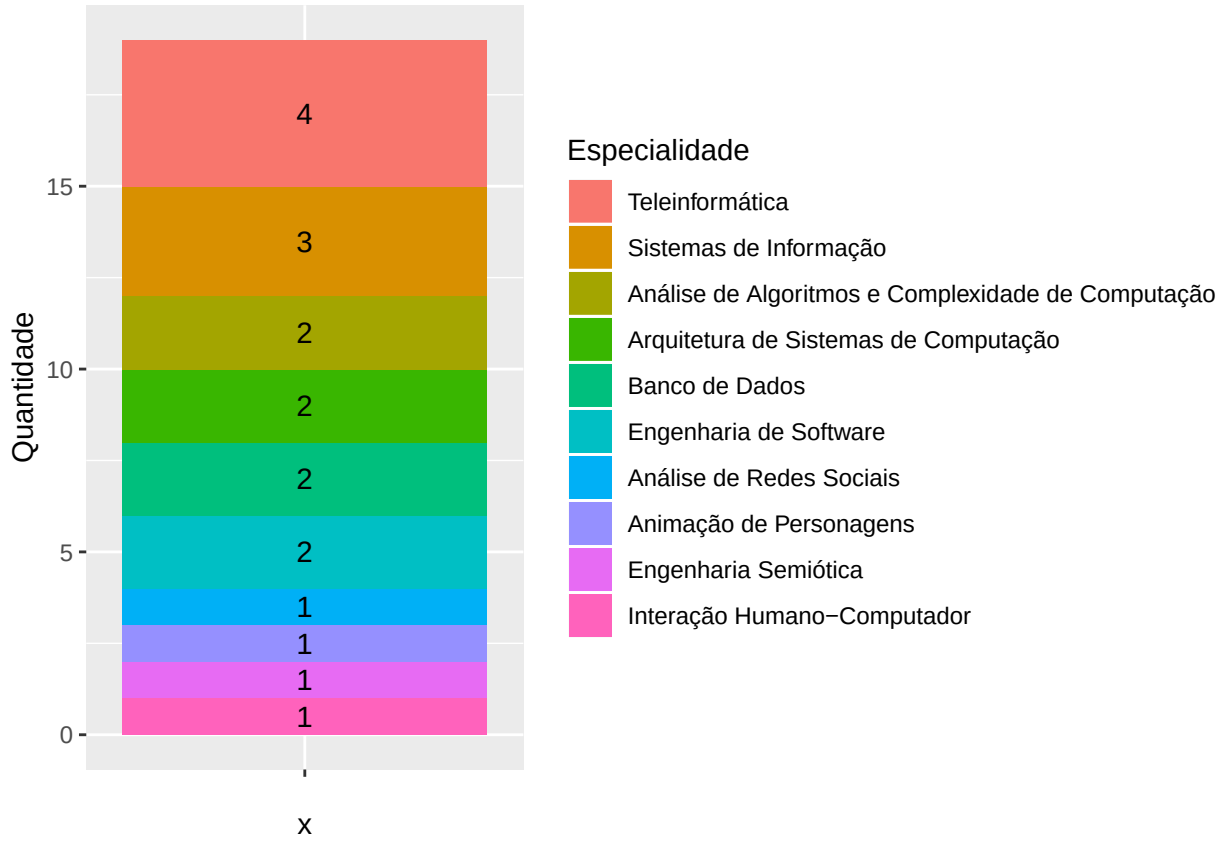
```r
especialidades_frequentes <- comp_profile %>%
  sapply(function(x) unique(x$areas_de_atuacao$especialidade)) %>%
  unlist() %>% table() %>% sort(decreasing = TRUE) %>%
  as.data.frame() %>% filter(!. == "") %>% head(10)

quantidade <- sum(especialidades_frequentes$Freq)

especialidades_frequentes <- mutate(especialidades_frequentes, percent = round(especialidades_frequentes

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))
```



Aqui vê-se as 10 especialidades mais recorrentes e percebe-se uma boa distribuição da quantidade entre elas.

7.2.3 Publicações

7.2.3.1 Quantidade de Publicações por tipo

```
for (i in 1:length(comp_publication)){
  print(names(comp_publication[i]))
  print(comp_publication[[i]] %>%
          sapply(function(x)
            length(x$ano)) %>% sum())
}
```

```
## [1] "PERIODICO"
## [1] 42
## [1] "LIVRO"
## [1] 2
## [1] "CAPITULO_DE_LIVRO"
## [1] 10
## [1] "TEXTO_EM_JORNAIS"
## [1] 0
## [1] "EVENTO"
## [1] 298
## [1] "ARTIGO_ACEITO"
## [1] 0
## [1] "DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA"
## [1] 7
```

Percebe-se que o tipo mais recorrente de publicações é do tipo Evento, possui bem mais ocorrências do que o segundo colocado, Periódico.

7.2.3.2 Quantidade de publicações por tipo e por ano

```
publication_tipo_comp <- comp_publication %>%
  sapply(function(x)
    sapply(x, function(x)
      length(x$autores)))
publication_tipo_comp
```

```
##      PERIODICO LIVRO CAPITULO_DE_LIVRO TEXTO_EM_JORNAIS EVENTO
## 2014         7     0                 1                0     63
## 2015         6     1                 2                0     54
## 2016        10     0                 1                0     76
## 2017         8     0                 2                0     63
## 2018        11     1                 4                0     42
##      ARTIGO_ACEITO DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA
## 2014             0                                      4
## 2015             0                                      3
## 2016             0                                      0
## 2017             0                                      0
## 2018             0                                      0
```

Destaca-se que nos ano de 2016 houveram mais publicações de Periódicos e Eventos. Outro ponto interessante é a ausência de periódicos do tipo Artigo Aceito e Texto em Jornais em todos os anos.

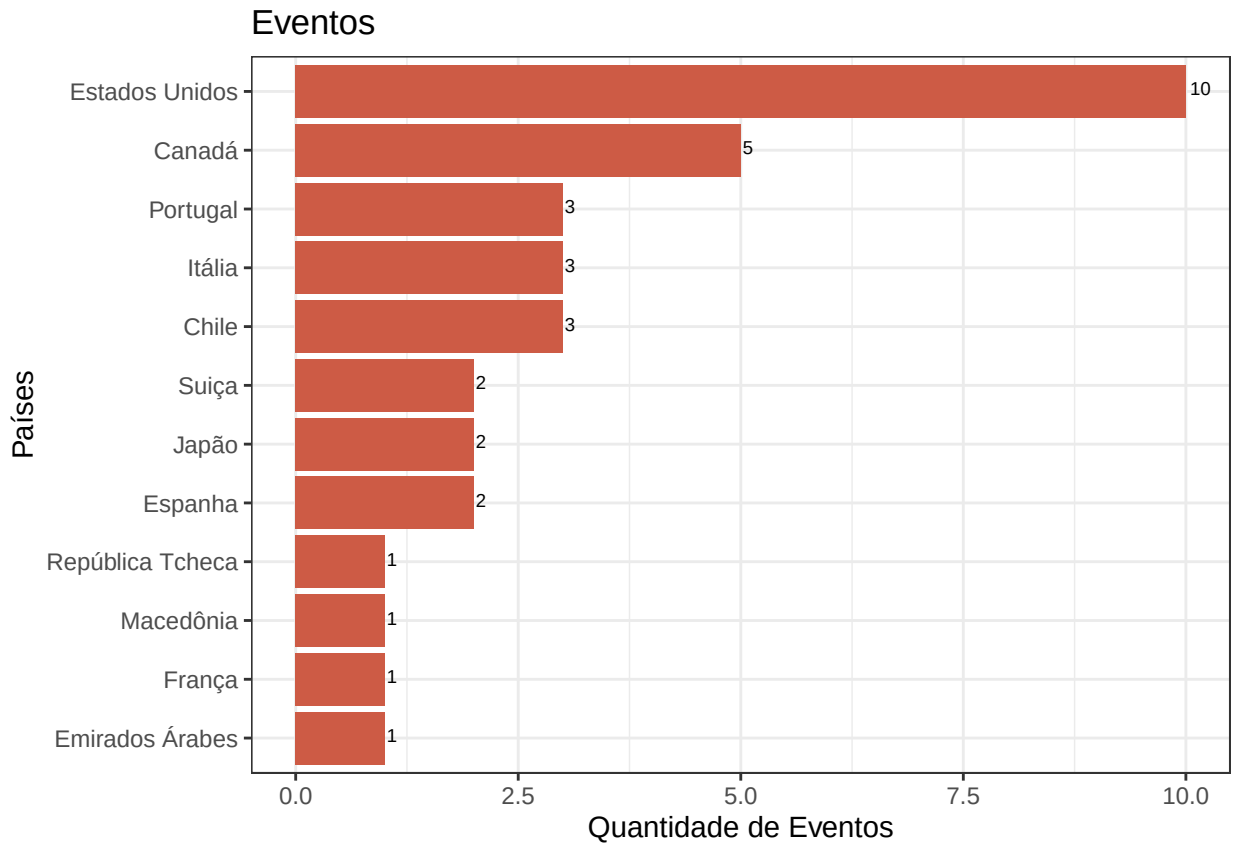
7.2.3.3 Participação em eventos por país

```
comp_publication$EVENTO %>%
  sapply(function(x)
    (x$pais_do_evento)) %>%
  unlist() %>% table() %>%
  sort(decreasing = TRUE)
```

```
## .
##           Brasil   Estados Unidos           Canadá            Chile
##              264               10                5                3
##           Itália         Portugal          Espanha            Japão
##                3                3                2                2
##            Suiça  Emirados Árabes           França        Macedônia
##                2                1                1                1
## República Tcheca
##                1
```

Este programa possui um bom número de participações em eventos em outros países.

7.2.3.4 Eventos por país



7.3 Pós-Graduação em Matemática

7.3.1 Orientações

7.3.1.1 Números de orientações completas por ano

```
for (i in 6:9){
  print(names(mat_advise[i]))
  print(mat_advise[[i]] %>%
          sapply(function(x)
            length(x$ano)))
}
```

```
## [1] "ORIENTACAO_CONCLUIDA_POS_DOUTORADO"
## 2014 2015 2016 2017 2018
##    0    0    1    0    1
## [1] "ORIENTACAO_CONCLUIDA_DOUTORADO"
## 2014 2015 2016 2017 2018
##    7    4    9    5    5
## [1] "ORIENTACAO_CONCLUIDA_MESTRADO"
## 2014 2015 2016 2017 2018
##   24   18   20   17   15
## [1] "OUTRAS_ORIENTACOES_CONCLUIDAS"
## 2014 2015 2016 2017 2018
##   19   11   18   10    6
```

No geral percebe-se que existem mais orientações concluídas do tipo Mestrado e Outras Orientações, havendo bem poucas orientações de Pós Doutorado. Pode-se supor que isto aconteça devido ao nível de exigência das orientações de Pós Doutorado.

7.3.1.2 Disposição de orientações por situação e ano

```
for (i in 1:length(mat_advise)){
  print(names(mat_advise[i]))
  print(mat_advise[[i]] %>%
          sapply(function(x)
            length(x$ano)) )
}
```

```
## [1] "ORIENTACAO_EM_ANDAMENTO_DE_POS_DOUTORADO"
## 2014 2015 2016 2017 2018
##    0    0    0    1    3
## [1] "ORIENTACAO_EM_ANDAMENTO_DOUTORADO"
## 2014 2015 2016 2017 2018
##    2    2    5    4    7
## [1] "ORIENTACAO_EM_ANDAMENTO_MESTRADO"
## 2014 2015 2016 2017 2018
##    0    0    3    3   13
## [1] "ORIENTACAO_EM_ANDAMENTO_GRADUACAO"
## 2014 2015 2016 2017 2018
##    0    0    0    0    0
## [1] "ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA"
## 2014 2015 2016 2017 2018
##    0    1    5    6    9
## [1] "ORIENTACAO_CONCLUIDA_POS_DOUTORADO"
## 2014 2015 2016 2017 2018
##    0    0    1    0    1
## [1] "ORIENTACAO_CONCLUIDA_DOUTORADO"
```

```
## 2014 2015 2016 2017 2018
##    7    4    9    5    5
## [1] "ORIENTACAO_CONCLUIDA_MESTRADO"
## 2014 2015 2016 2017 2018
##   24   18   20   17   15
## [1] "OUTRAS_ORIENTACOES_CONCLUIDAS"
## 2014 2015 2016 2017 2018
##   19   11   18   10    6
```

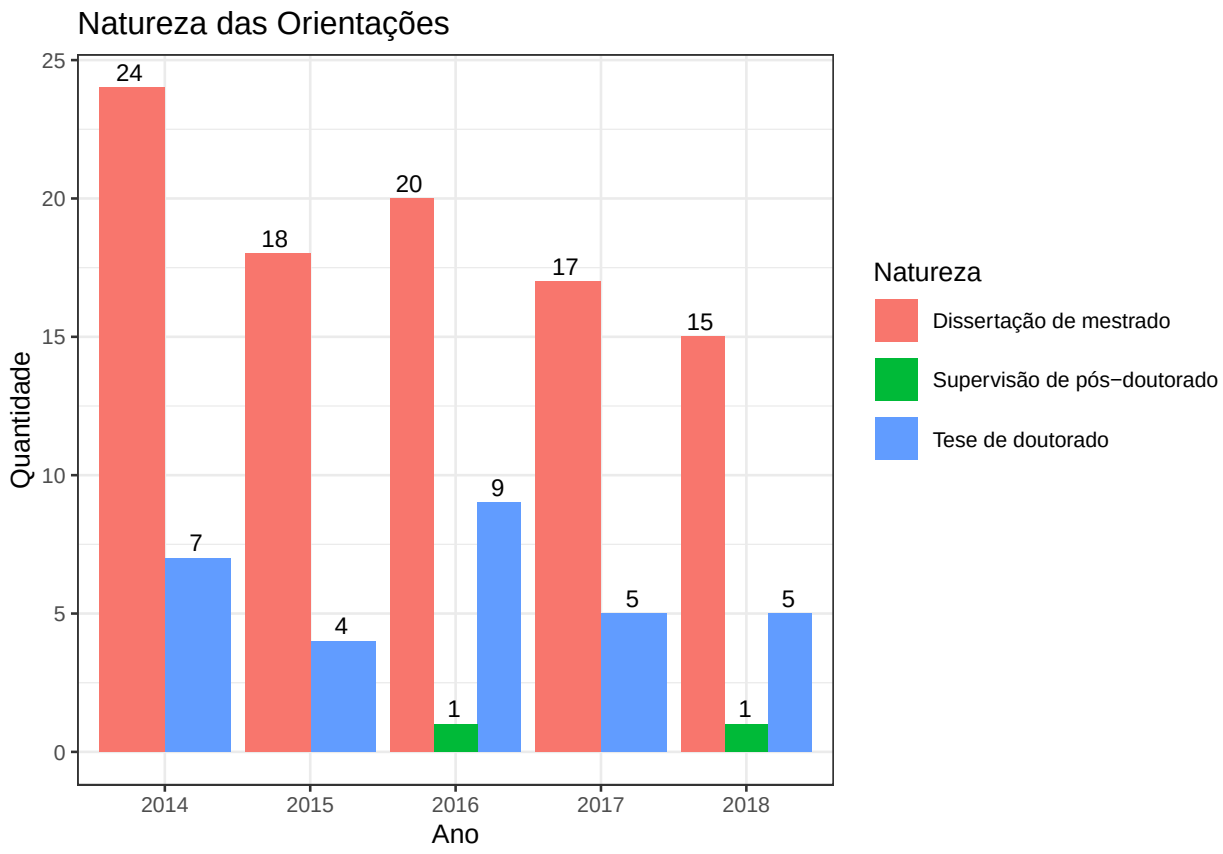
7.3.1.3 Natureza das orientações

```
mat_public.periodico.df <- pub.ls2df(mat_publication, 1) #artigos
mat_public.livros.df <- pub.ls2df(mat_publication, 2) #livros
mat_public.textojornais.df <- pub.ls2df(mat_publication, 4) #textos em jornais
mat_public.eventos.df <- pub.ls2df(mat_publication, 5) #eventos
```

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

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

```
ggplot(mat_orient.df,aes(ano, fill=factor(natureza))) +
  geom_bar(stat = "count",position='dodge') +
  ggtitle("Natureza das Orientações") +
  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")+labs(fill="Natureza")+theme_bw(base_size = 10)+
  geom_text(hjust=0.6,
            vjust=-0.4,
            size=3,color='black',
            position = position_dodge(width=0.9),
            stat = "count",
            aes(group=factor(natureza),
                label=formatC(..count.., big.mark=",")),
            check_overlap = TRUE)
```



7.3.2 Perfil

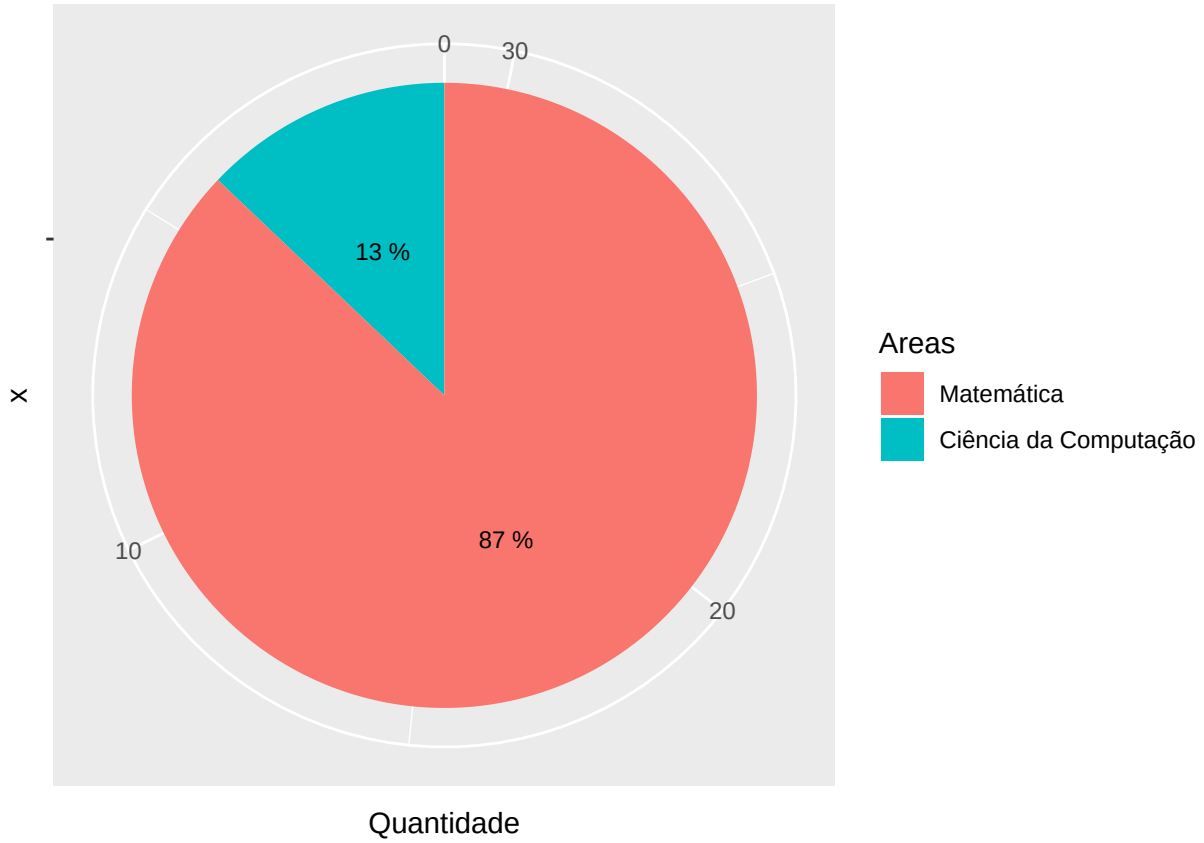
7.3.2.1 Distribuição de áreas de atuação dos pesquisadores

```
areas_atuacao_mat <- mat_profile %>%
  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), size=3)
```



Percebe-se que a área de atuação com maior recorrência é a *Ciência da Computação* e apenas uma parcela reduzida equivale a área *Matemática*, o que se analisado isoladamente pode causar estranheza, pois seria de se esperar que no programa de Matemática a área *Matemática* tivesse uma maior porcentagem.

7.3.2.2 Distribuição de subáreas de atuação mais frequentes dos pesquisadores

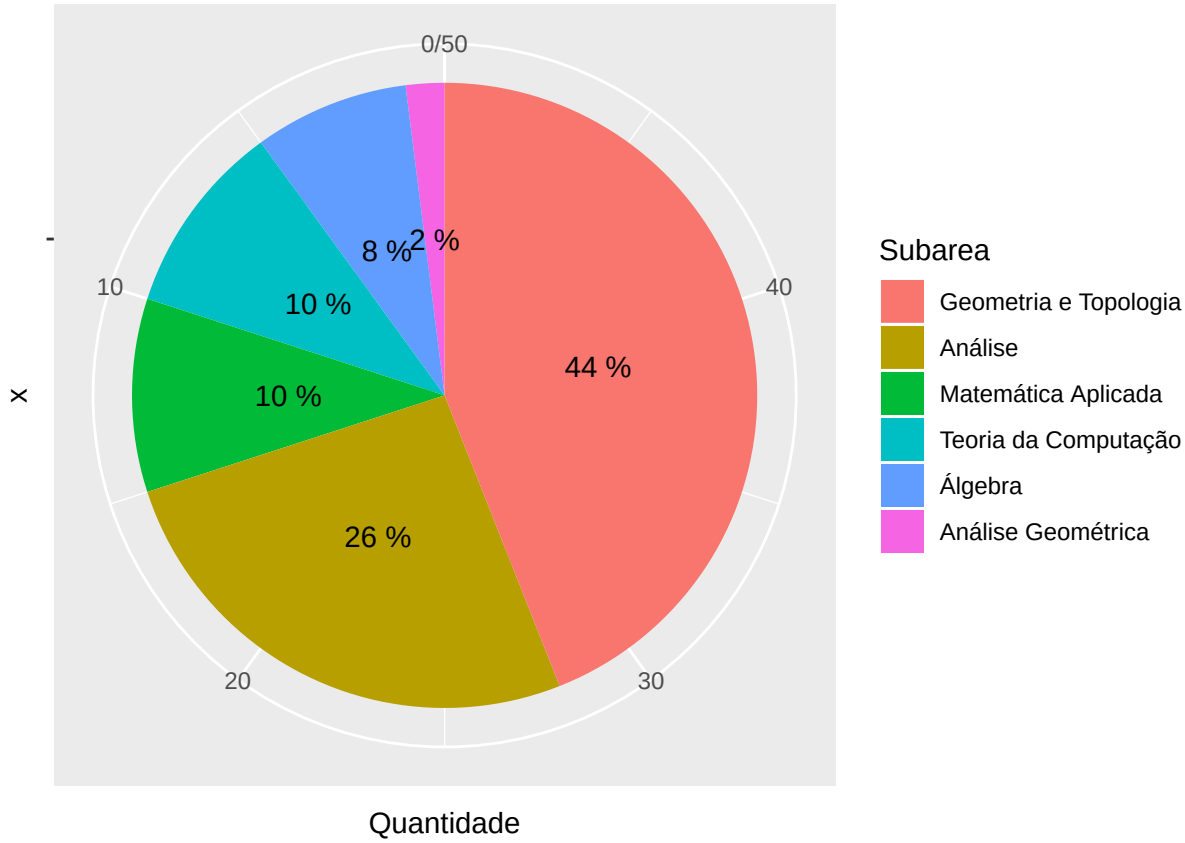
```
subarea <- mat_profile %>%
  sapply(function(x) (x$areas_de_atuacao$sub_area)) %>%
  unlist() %>% table() %>%
  sort(decreasing = TRUE) %>%   as.data.frame() %>%
  filter(!. == "") %>% head(6)

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))
```



Através da análise acima, pode-se ver que a área com maior número de docentes atuantes é em *Geometria e Topologia*, seguida por *Análise*, *Matemática Aplicada* e *Teoria da Computação*.

7.3.2.3 Distribuição de especialidades mais frequentes dos pesquisadores

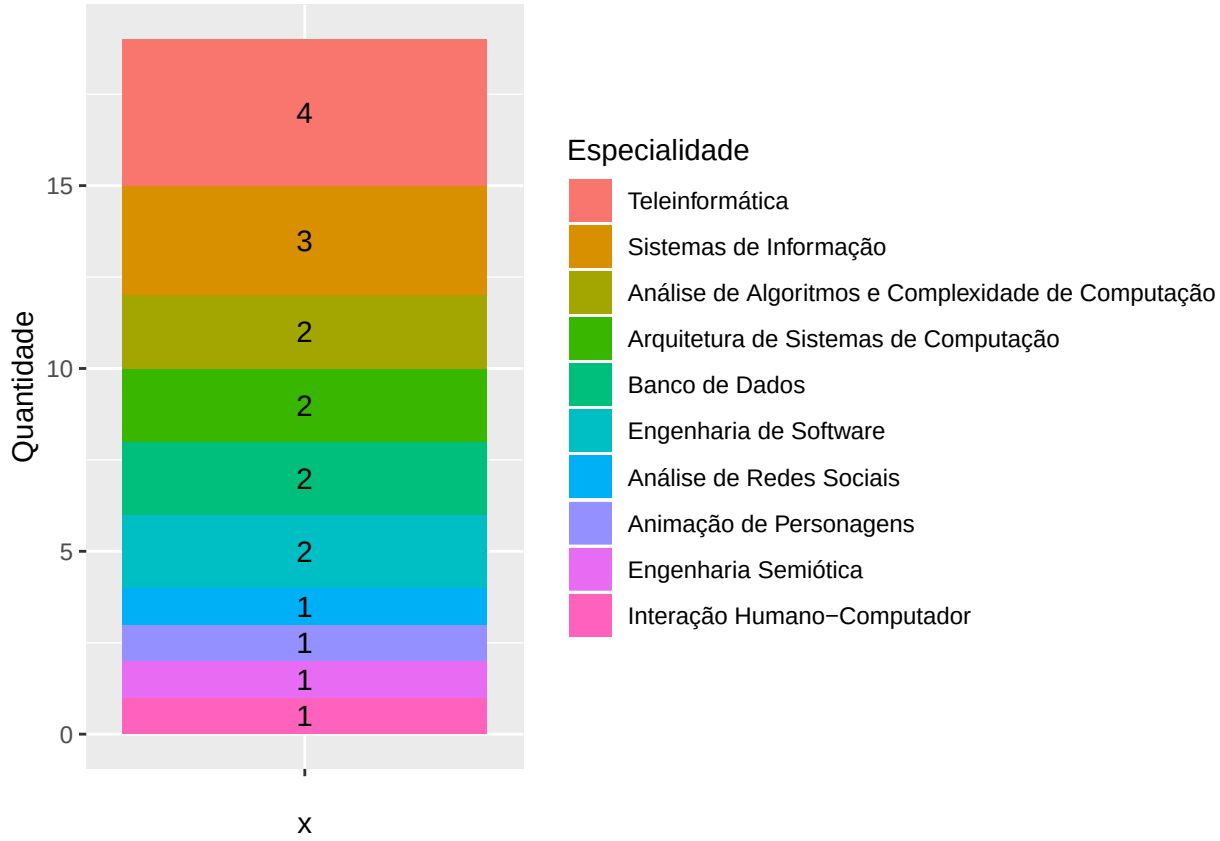
```
especialidades_frequentes <- comp_profile %>%
  sapply(function(x) unique(x$areas_de_atuacao$especialidade)) %>%
  unlist() %>% table() %>% sort(decreasing = TRUE) %>%
  as.data.frame() %>% filter(!. == "") %>% head(10)

quantidade <- sum(especialidades_frequentes$Freq)

especialidades_frequentes <- mutate(especialidades_frequentes, percent = round(especialidades_frequentes

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))
```



Aqui vê-se as 10 especialidades mais recorrentes e percebe-se uma boa distribuição da quantidade entre elas.
Publicações

7.3.2.4 Quantidade de Publicações por tipo

```
for (i in 1:length(mat_publication)){
  print(names(mat_publication[i]))
  print(mat_publication[[i]] %>%
          sapply(function(x)
            length(x$ano)) %>% sum())
}
```

```
## [1] "PERIODICO"
## [1] 151
## [1] "LIVRO"
## [1] 7
## [1] "CAPITULO_DE_LIVRO"
## [1] 4
## [1] "TEXTO_EM_JORNAIS"
## [1] 0
## [1] "EVENTO"
## [1] 28
## [1] "ARTIGO_ACEITO"
## [1] 14
## [1] "DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA"
## [1] 35
```

Percebe-se que o tipo mais recorrente de publicações é do tipo Periódico.

7.3.2.5 Quantidade de publicações por tipo e por ano

```
publication_tipo_mat <- mat_publication %>%
  sapply(function(x)
    sapply(x, function(x)
      length(x$autores)))
publication_tipo_mat
```

```
##      PERIODICO LIVRO CAPITULO_DE_LIVRO TEXTO_EM_JORNAIS EVENTO
## 2014        28     1                 0                0     10
## 2015        30     1                 1                0      3
## 2016        32     2                 1                0      4
## 2017        32     1                 0                0      4
## 2018        29     2                 2                0      7
##      ARTIGO_ACEITO DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA
## 2014             0                                     2
## 2015             0                                     4
## 2016             0                                     2
## 2017             1                                     9
## 2018            13                                    18
```

Destaca-se que nos anos de 2016 e 2017 houveram mais publicações de Periódicos e em 2018 houve uma queda, mas no ano de 2018 houveram mais publicações do tipo Artigo Aceito e Demais Tipos de Produção Bibliografica. Pode-se imaginar que o programa focou mais nesses outros tipos de publicação no ano de 2018 em relação a Periódicos.

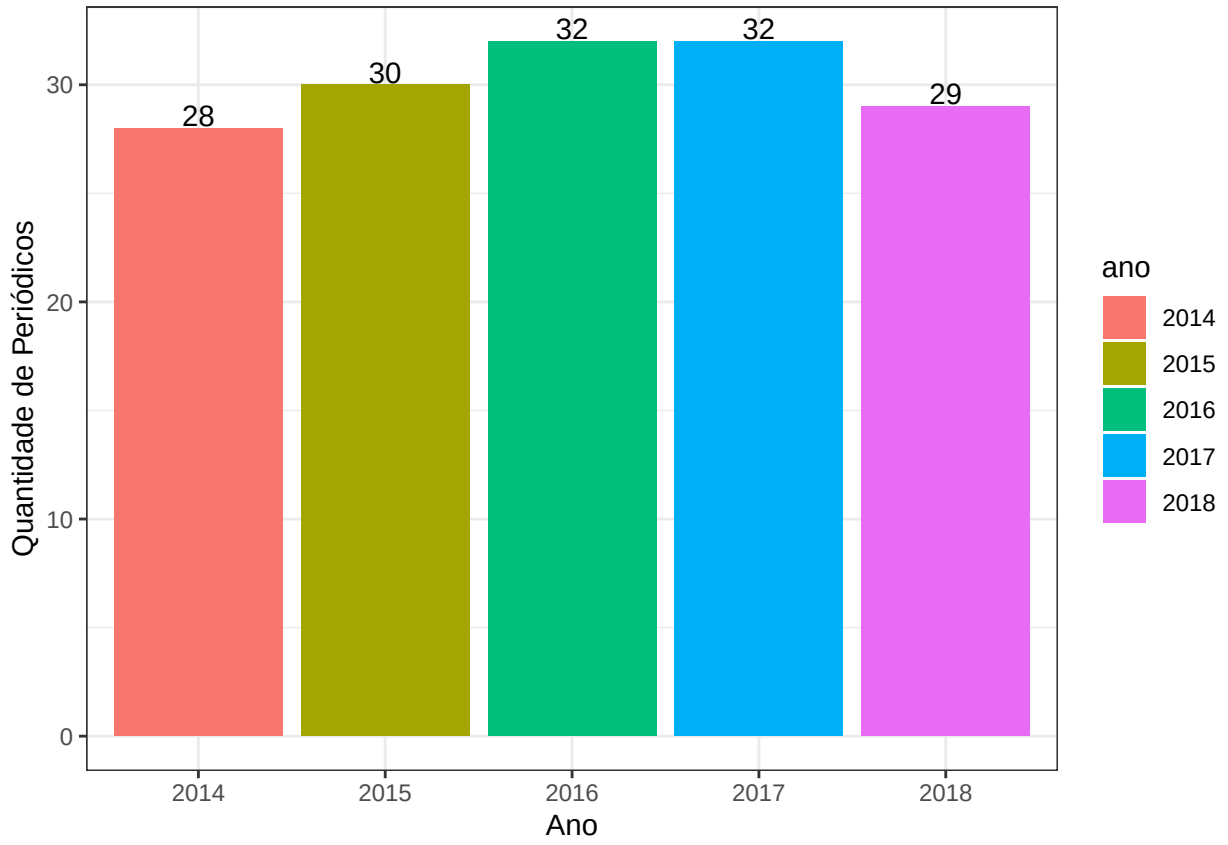
7.3.2.6 Participação em eventos por país

```
mat_publication$EVENTO %>%
  sapply(function(x)
    (x$pais_do_evento)) %>%
  unlist() %>% table() %>%
  sort(decreasing = TRUE)
```

```
## .
##  Brasil  França   Chile Espanha  Itália
##      21       4       1       1       1
```

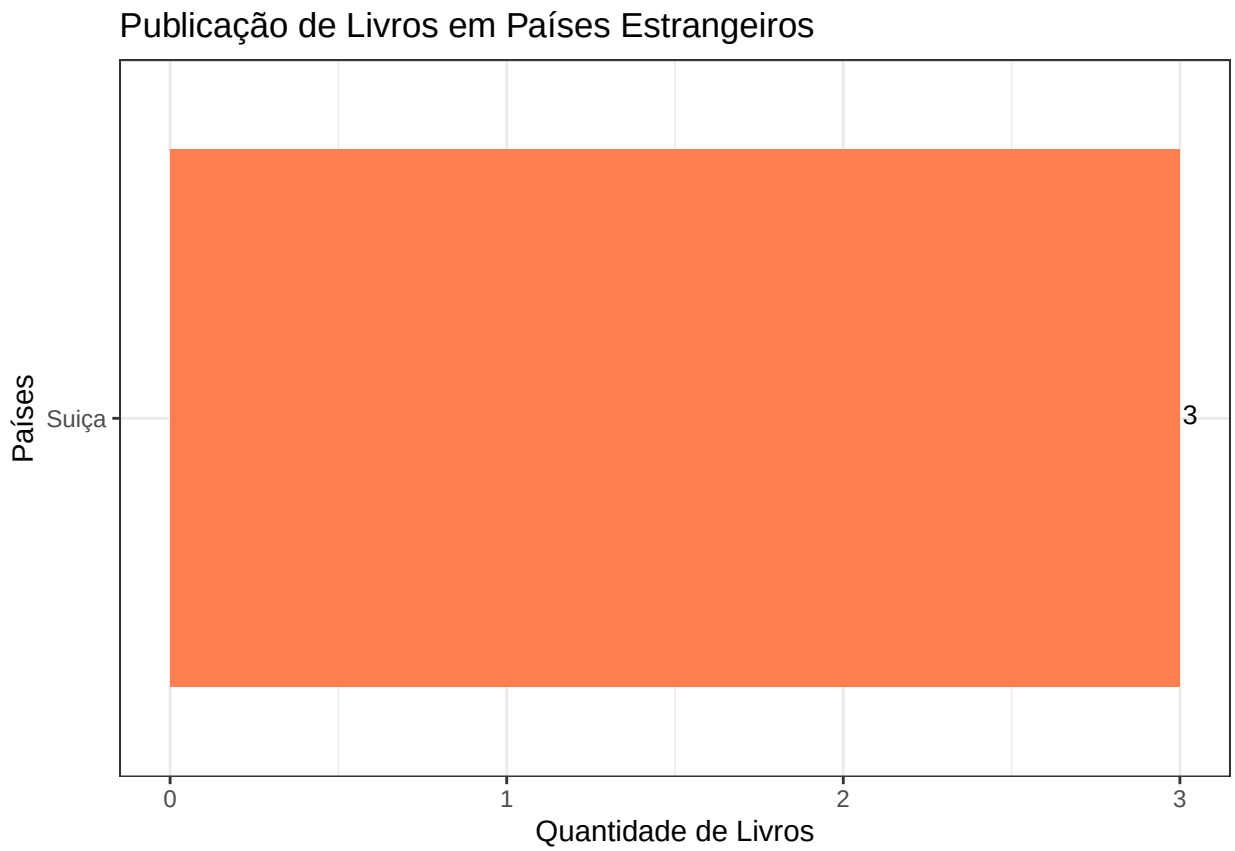
É interessante que o país tenha participado de mais eventos na França do que Chile que é um país mais próximo do Brasil.

7.3.2.7 Publicações por ano

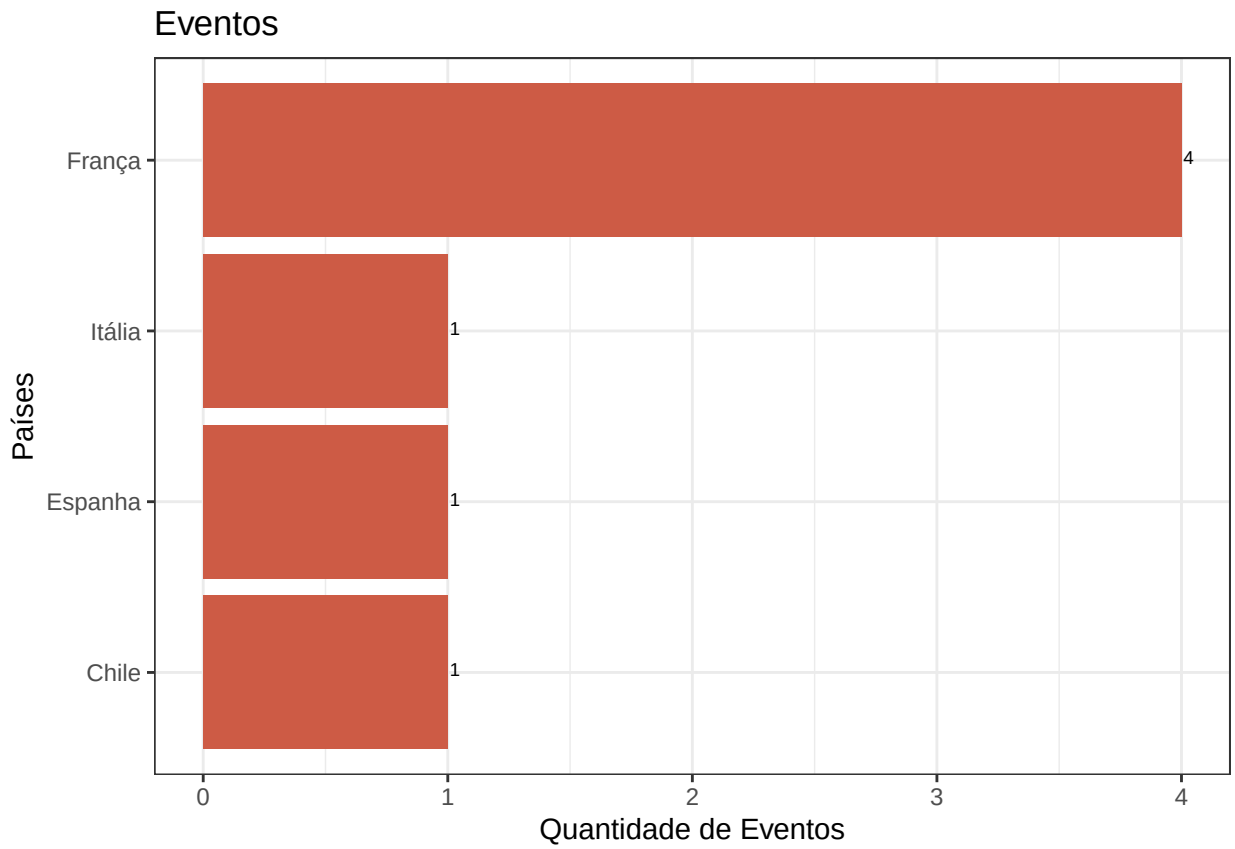


7.3.2.8 Publicações de livros fora do Brasil

```
mat_public.livros.df %>%
  group_by(pais_de_publicacao) %>%
  summarise(Quantidade = n()) %>%
  filter(pais_de_publicacao != "Brasil") %>%
  ggplot(aes(x = reorder(pais_de_publicacao, (Quantidade)), y = Quantidade)) +
  geom_col(fill = "coral")+
  geom_text(aes(label=comma(Quantidade)),
            hjust=-0.2,vjust=0.3,size=3.5) + coord_flip() +
  labs(title = "Publicação de Livros em Países Estrangeiros",
       x = "Países",
       y = "Quantidade de Livros")+
  theme_bw()
```



7.3.2.9 Eventos por país



8 Modelos de Análise

8.1 Análise de Redes

Uma rede (ou um grafo de redes) representa uma conexão entre indivíduos. Na representação por redes, a presença de uma aresta entre um ou mais nós representa a existência de uma relação entre os indivíduos conectados, formando assim a rede. Há vários tipos de padrões e relacionamentos que podem ser modelos e representados como uma rede dessa maneira.

A partir dos arquivos graph.json podemos criar modelos de relacionamento entre os pesquisadores de diversos departamentos. Para tal será utilizada a biblioteca **igraph**, que permite a visualização gráfica desses relacionamentos na forma de uma estrutura de dados conhecida como *grafo*.

Começamos nossa análise carregando a biblioteca **igraph**:

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

## The following objects are masked from 'package:dplyr':
##
##     as_data_frame, groups, union

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

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

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

A análise de redes detalhada será feita para cada um dos programas de pós-graduação a seguir:

8.1.1 Ciência da Computação

Começamos carregando o arquivo json para a memória e criando variáveis para representar os nós e os relacionamentos entre os diversos autores do programa de Pós-Graduação em Ciência da Computação.

```
  cic_graph <- jsonlite::fromJSON("cic_graph.json")
  cic_nodes <- cic_graph$nodes
  cic_nodes$properties <- NULL
  cic_relations <- cic_graph$links
```

Podemos observar as primeiras linhas dos nós (ou seja, os pesquisadores em si):

```
head(cic_nodes)
```

```
##                 id label
## 1 0597956911969596     1
## 2 0656977742590515     2
## 3 0771942436828005     3
## 4 0802023762311924     4
## 5 1744732999336375     5
## 6 2807107019851322     6
```

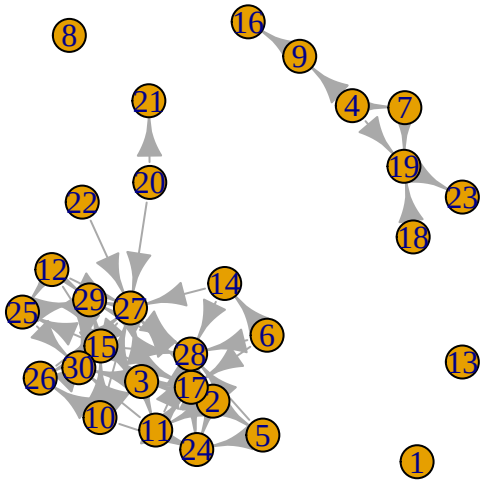
O mesmo pode ser feita para ser observar os relacionamentos (as arestas no grafo):

```
head(cic_relations)
```

```
##             source           target weigth
## 1 0656977742590515 0771942436828005      1
## 2 0656977742590515 1744732999336375      1
## 3 0656977742590515 3614256141054800      3
## 4 0656977742590515 6631267110894080      2
## 5 0656977742590515 8908026219336623      3
## 6 0656977742590515 9553770402705512      1
```

A biblioteca **igraph** facilita bastante o processo de geração das redes a partir dos dados fornecidos.

```
  graph_cic <- graph_from_data_frame(cic_relations, directed=TRUE, vertices=cic_nodes)
  plot(graph_cic)
```



Pode-se observar facilmente que há dois *clusters* distintos de pesquisadores que já trabalharam em conjunto, além de três outros pesquisadores que trabalham apenas sozinhos.

8.1.2 Computação

Fazendo-se a análise anterior para a Pós-Graduação em Computação, obtemos os seguintes resultados:

```
comp_graph <- jsonlite::fromJSON("comp_graph.json")
comp_nodes <- comp_graph$nodes
comp_nodes$properties <- NULL
comp_relations <- comp_graph$links
```

Observando-se os primeiros dados dos conjuntos de nós e das arestas:

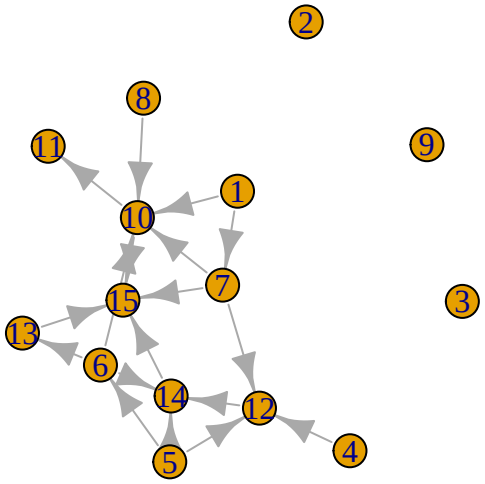
```
head(comp_nodes)
```

```
##                 id label
## 1 1556476759915826     1
## 2 1866537425954359     2
## 3 2028071357762491     3
## 4 2149928021922564     4
## 5 2807107019851322     5
## 6 2812695592756155     6
```

```
  head(comp_relations)
```

```
##             source           target weigth
## 1 1556476759915826 3125027229507836      1
## 2 1556476759915826 4277471687235814      1
## 3 2149928021922564 6631267110894080      1
## 4 2807107019851322 2812695592756155      7
## 5 2807107019851322 6631267110894080      1
## 6 2807107019851322 7332276159500899      1
```

```
  graph_comp <- graph_from_data_frame(comp_relations, directed=TRUE, vertices=comp_nodes)
  plot(graph_comp)
```



Na análise para a Pós-Graduação em Computação fica bem claro que há um número menor de pesquisadores, e diferentemente da Pós-Graduação em Ciência da Computação, aqui fica bem claro que não há uma separação tão grandes entre diferentes *clusters* de pesquisa.

8.1.3 Matemática

Por último, a análise de redes será feita para a Pós-Graduação em Matemática.

```
  mat_graph <- jsonlite::fromJSON("mat_graph.json")
  mat_nodes <- mat_graph$nodes
  mat_nodes$properties <- NULL
  mat_relations <- mat_graph$links
```

Observando-se os dados:

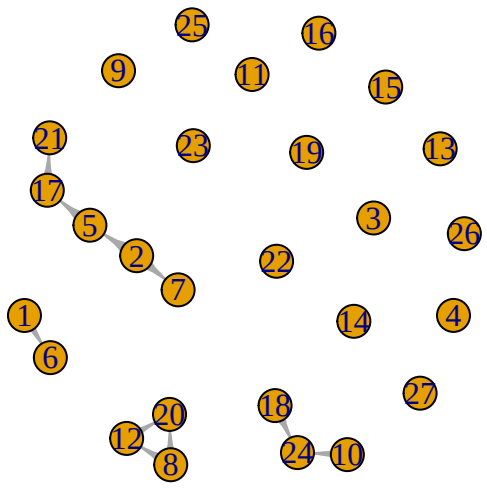
```
  head(mat_nodes)
```

```
##                  id label
## 1 0499138914431785     1
## 2 1113531859811863     2
## 3 1270947513791152     3
## 4 1299070184304880     4
## 5 1873757687453531     5
## 6 1912187013775217     6
```

```
  head(mat_relations)
```

```
##             source           target weigth
## 1 0499138914431785 1912187013775217      1
## 2 1113531859811863 1873757687453531      2
## 3 1113531859811863 1999855370737290      1
## 4 1873757687453531 5686283061097960      1
## 5 2132614695901416 4695081445531168      3
## 6 2132614695901416 7659965567201224      4
```

```
  graph_mat <- graph_from_data_frame(mat_relations, directed=TRUE, vertices=mat_nodes)
  plot(graph_mat)
```



A análise das redes no programa de pós-graduação em Matemática é interessante por ser bastante diferente dos outros dois programas sendo estudados. Em primeiro lugar, percebe-se facilmente que há muito menos colaboração entre os diversos pesquisadores desse departamento do que nos outros dois em estudo, observando-se apenas quatro agrupamentos de pesquisadores que já trabalharam em conjunto na parte inferior do grafo.

8.2 Nuvem de Palavras

Ainda dentro do escopo dos modelos de análise, o modelo de *Word Cloud*, ou Nuvem de Palavras, é interessante como maneira de representar visualmente a ocorrência de diferentes palavras dentro do conjunto de dados que se quer analisar.

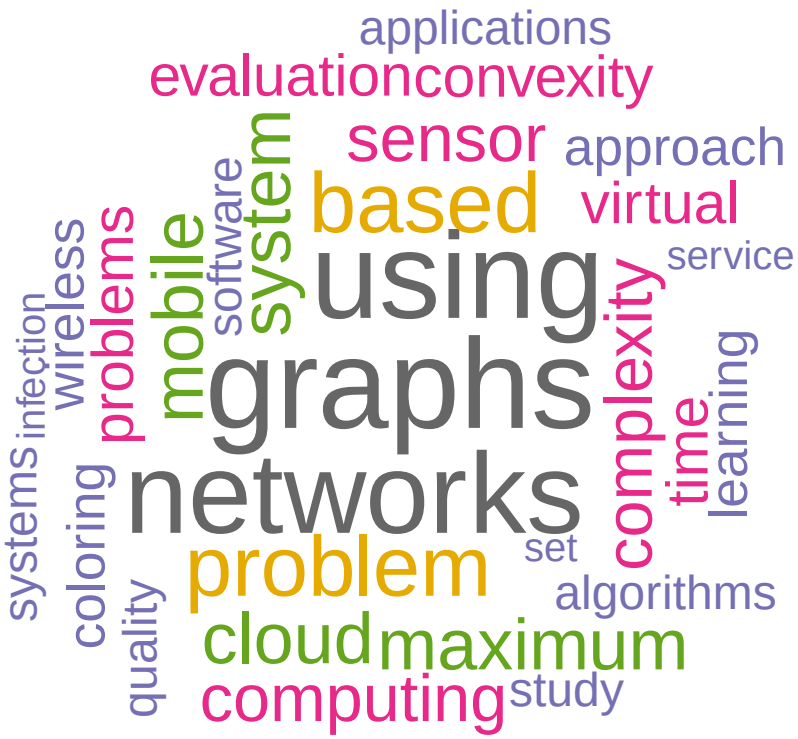
8.2.1 Ciência da Computação

Fazendo-se a limpeza dos dados dos títulos das publicações para criação da nuvem de palavras:

```
  cic_publication.df <- data.frame()
  for (i in 1:length(cic_publication[[1]]))
    cic_publication.df <- rbind(cic_publication.df, cic_publication$PERIODICO[[i]])
  cic_publication.df$autores <- gsub("\",\"|\", \"", "; ", cic_publication.df$autores)
  cic_publication.df$autores <- gsub("\"|c\\(|\\)", "", cic_publication.df$autores)
  cic_publication.df$`autores-endogeno` <- gsub(",", ";", cic_publication.df$`autores-endogeno`)
  cic_publication.df$`autores-endogeno` <- gsub("\"|c\\(|\\)", "", cic_publication.df$`autores-endogeno
```

Com os dados dos títulos das publicações limpos, é possível prosseguir para a realização da análise usando os pacotes **tm** e **wordcloud**.

```
cic_publication.df$titulo <- iconv(cic_publication.df$titulo,from="UTF-8", to = "ASCII//TRANSLIT")
df_wc <- data.frame(doc_id = cic_publication.df$issn, text = cic_publication.df$titulo, stringsAsFactor
docs <- Corpus(DataframeSource(df_wc))
docs <- tm_map(docs, content_transformer(tolower))
docs <- tm_map(docs, removeNumbers)
docs <- tm_map(docs, removeWords, stopwords("pt"))
docs <- tm_map(docs, removeWords, stopwords("english"))
docs <- tm_map(docs, removeWords, c("universidade", "federal","computação","posgraduacao","ciência","ma
docs <- tm_map(docs, removePunctuation)
docs <- tm_map(docs, stripWhitespace)
dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
set.seed(1234)
wordcloud(words =(d$word), freq = d$freq, min.freq = 3,
  max.words=30,scale = c(4, 0.3), random.order=FALSE, rot.per=0.4,
  colors=brewer.pal(8, "Dark2"))
```



A nuvem de palavras acima permite observar que um grande número de publicações na área de ciência da computação está voltada para a área de grafos e redes.

8.2.2 Computação

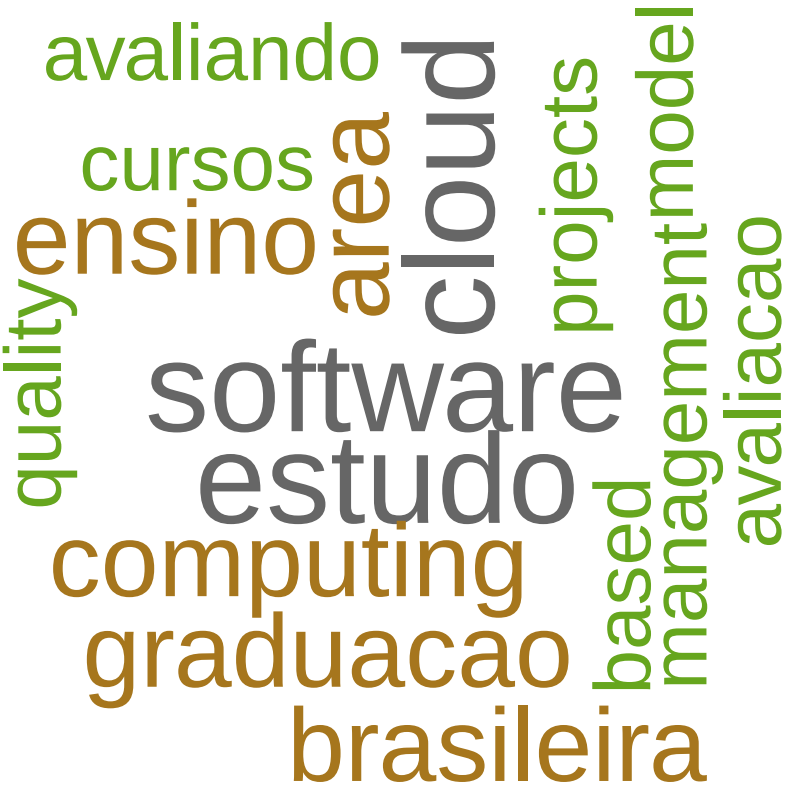
A mesma análise de nuvem de palavras pode ser feita também para o programa de pós-graduação em Computação. Iniciando pela limpeza dos dados:

```
  comp_publication.df <- data.frame()
  for (i in 1:length(comp_publication[[1]]))
    comp_publication.df <- rbind(comp_publication.df, comp_publication$PERIODICO[[i]])
  comp_publication.df$autores <- gsub("\",\"|\", \"", "; ", comp_publication.df$autores)
  comp_publication.df$autores <- gsub("\"|c\\(|\\)", "", comp_publication.df$autores)
  comp_publication.df$`autores-endogeno` <- gsub(",", ";", comp_publication.df$`autores-endogeno`)
  comp_publication.df$`autores-endogeno` <- gsub("\"|c\\(|\\)", "", comp_publication.df$`autores-endoge
```

E fazendo-se a nuvem de palavras:

```
comp_publication.df$titulo <- iconv(comp_publication.df$titulo,from="UTF-8", to = "ASCII//TRANSLIT")
df_wc <- data.frame(doc_id = comp_publication.df$issn, text = comp_publication.df$titulo, stringsAsFacto
docs <- Corpus(DataframeSource(df_wc))
docs <- tm_map(docs, content_transformer(tolower))
docs <- tm_map(docs, removeNumbers)
docs <- tm_map(docs, removeWords, stopwords("pt"))
docs <- tm_map(docs, removeWords, stopwords("english"))
```

```
docs <- tm_map(docs, removeWords, c("universidade", "federal","computação","posgraduacao","ciência","ma
docs <- tm_map(docs, removePunctuation)
docs <- tm_map(docs, stripWhitespace)
dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
set.seed(1234)
wordcloud(words =(d$word), freq = d$freq, min.freq = 3,
  max.words=30,scale = c(4, 0.3), random.order=FALSE, rot.per=0.4,
  colors=brewer.pal(8, "Dark2"))
```



No caso do programa de pós-graduação em Computação, percebe-se uma frequência maior de aparecimento de palavras relacionados a ensino e estudo, explicitando o foco educacional dessa área da computação.

8.2.3 Matemática

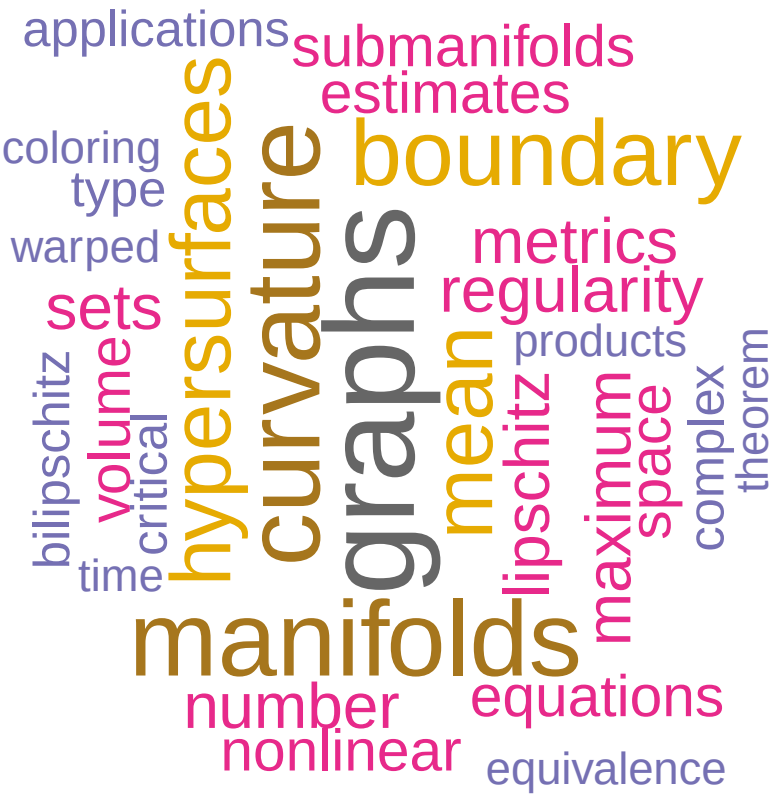
Por último, pode-se realizar a mesma análise para a pós-graduação em matemática:

```
  mat_publication.df <- data.frame()
  for (i in 1:length(mat_publication[[1]]))
    mat_publication.df <- rbind(mat_publication.df, mat_publication$PERIODICO[[i]])
  mat_publication.df$autores <- gsub("\",\"|\", \"", "; ", mat_publication.df$autores)
  mat_publication.df$autores <- gsub("\"|c\\(|\\)", "", mat_publication.df$autores)
```

```
  mat_publication.df$`autores-endogeno` <- gsub(",", ";", mat_publication.df$`autores-endogeno`)
  mat_publication.df$`autores-endogeno` <- gsub("\"|c\\(|\\)", "", mat_publication.df$`autores-endogeno`
```

E a nuvem de palavras:

```
mat_publication.df$titulo <- iconv(mat_publication.df$titulo,from="UTF-8", to = "ASCII//TRANSLIT")
df_wc <- data.frame(doc_id = mat_publication.df$issn, text = mat_publication.df$titulo, stringsAsFactors
docs <- Corpus(DataframeSource(df_wc))
docs <- tm_map(docs, content_transformer(tolower))
docs <- tm_map(docs, removeNumbers)
docs <- tm_map(docs, removeWords, stopwords("pt"))
docs <- tm_map(docs, removeWords, stopwords("english"))
docs <- tm_map(docs, removeWords, c("universidade", "federal","computação","posgraduacao","ciência","ma
docs <- tm_map(docs, removePunctuation)
docs <- tm_map(docs, stripWhitespace)
dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
set.seed(1234)
wordcloud(words =(d$word), freq = d$freq, min.freq = 3,
  max.words=30,scale = c(4, 0.3), random.order=FALSE, rot.per=0.4,
  colors=brewer.pal(8, "Dark2"))
```



No caso do programa de pós-graduação em Matemática, percebe-se também uma grande ocorrência de publicações envolvendo grafos. Como foi observado um grande número de publicações envolvendo grafos

também dentro do programa de Pós-Graduação em Ciência da Computação, uma análise posterior poderia revelar se ocorre colaboração entre departamentos diferentes com foco em pesquisas nessa área.

9 Análises Finais

9.1 Orientações

No geral percebe-se que existem poucas orientações de Pós Doutorado em relação aos demais tipos. Pode-se supor que isto aconteça devido ao nível de exigência das orientações de Pós Doutorado.

Outra questão interessante é o grande número de orientações concluídas do tipo *Outras Orientações* que ocorre em todos os programas analisados.

9.2 Perfil

Ao analisar as informações em relação ao perfil dos pesquisadores, percebe-se que muitos pesquisadores do programa de Ciência da Computação atuam na área de *Matemática* e muitos pesquisadores do programa de Matemática atuam na área de *Ciência da Computação*. Demonstrando como há uma relação forte entre as áreas.

Metodologia e Técnicas da Computação, *Teoria da Computação*, *Sistemas de Computação* e *Telecomunicações* estão entre as 6 sub áreas de atuação mais frequentes dos programas de Computação e Ciência da Computação. Tal constatação demonstra como os dois programas estão bem relacionados e trabalhos futuros poderiam explorar o grau de colaboração entre os participantes de ambos programas.

No geral, ao se analisar as informações a respeito de área de atuação, subárea de atuação e especialidades dos pesquisadores, percebe-se que os 3 programas analisados atuam em sua maioria em conteúdos de Exatas e tem boa relação com subáreas relacionadas à Engenharias.

9.3 Publicações

Pôde-se observar que a Ciência da Computação produz bem mais material do tipo Evento e nos anos analisados apresentou pouca variação na quantidade de Eventos.

A Computação também se destaca na produção de bem mais Eventos do que os outros tipos de publicação e isso pode ser explicado pela similaridade entre os programas de Ciência da Computação e de Computação. A quantidade de Eventos apresentou uma queda nos últimos dois anos analisados.

Diferente da Ciência da Computação e da Computação, os tipos de publicações da Matemática são bem mais de Periódicos, ficando o tipo Evento apenas em segundo lugar e sua quantidade sendo cerca de 5 vezes menor. Nos anos analisados obteve pouca variação em relação a quantidade de Periódicos publicados.

Em relação a participação em Eventos em outros países, percebe-se que os programas de Ciências da Computação e Computação possuem bem mais participações do que o programa de Matemática. Matemática participou em eventos nos anos analisados em apenas 4 países, já Computação participou em 12 países e Ciências da Computação em 29 países.

Nos programas de Computação e Ciências da Computação ocorreram muitas participações em Eventos no Estados Unidos, um indicativo de que o país e o Brasil possuem uma boa rede de interação e troca de conhecimento.

No geral, em todos os programas percebe-se que 2016 reuniu maiores números de publicações e orientações. A partir dessa percepção é possível criar teorias a respeito do por quê esse ano apresentou melhores resultados e com trabalhos seguintes buscar investigar essas teorias.

10 Conclusão

Com este trabalho foi possível trabalhar conceitos de ciências de dados e exercitar o uso de ferramentas relacionados a esta área.

A utilização da metodologia CRISP-DM foi de grande ajuda no seu desenvolvimento e guiou de forma eficiente os autores. Devido à sua complexidade a fase final não foi realizada, mas as demais foram abordadas e exploradas.

Diversas análises pontuais e gerais puderam ser levantadas e como oportunidade de trabalhos futuros os que mais chamam a atenção é o aprofundamento nas relações entre os 3 programas analisados, a exploração de quais seriam as orientação do tipo *Outras Orientações* que teve grande número de ocorrência e a investigação do por que o ano de 2016 reuniu maiores números de publicações e orientações.

11 Referências Bibliográficas

- SOBRE as áreas de avaliação. [S. l.], 20 jan. 2016. Disponível em: https://www.capes.gov.br/avaliacao/sobre-as-areas-de-avaliacao. Acesso em: 10 abr. 2019.
- SOBRE a avaliação. [S. l.], 20 jan. 2016. Disponível em: https://www.capes.gov.br/avaliacao/sobre-a-avaliacao. Acesso em: 10 abr. 2019.
- PLATAFORMA Sucupira. [S.l.], 20 jan. 2016. Disponível em: https://sucupira.capes.gov.br/sucupira/. Acesso em: 10 abr. 2019.
- CRISP-DM – a Standard Methodology to Ensure a Good Outcome. [S. l.], 20 jan. 2016. Disponível em: https://www.datasciencecentral.com/profiles/blogs/crisp-dm-a-standard-methodology-to-ensure-a-good-outcome. Acesso em: 13 abr. 2019.
- CIÊNCIAS da Computação. [S. l.], 13 mar. 2015. Disponível em: http://www.ufc.br/ensino/guia-de-profissoes/576-ciencia-da-computacao. Acesso em: 13 abr. 2019.
- MATEMÁTICA. [S. l.], 13 mar. 2015. Disponível em: http://www.mat.ufc.br/portal/ptbr/laboratorios. Acesso em: 13 abr. 2019.
- Ciência e tecnologia do Brasil. [S. l.]. Disponível em: https://pt.wikipedia.org/wiki/Ci%C3%AAncia_e_tecnologia_do_Brasil. Acesso em: 13 abr. 2019.
- Fernandes, Jorge Henrique Cabral. "Considerações Preliminares sobre a Ciência e sua avaliação". 2017.