

# Ciência de Dados para Todos (Data Science For All) - 2018.2 - Análise da Produção Científica e Acadêmica da Universidade de Brasília - Relatório Final da Disciplina - Departamento de Ciência da Computação da UnB

*Mateus H. Nagata - 17/0041930, João P. Tavares - 13/0029335, Victor Landim -  
15/0023031*

*08/07/2018*

## Introdução

O presente trabalho é o relatório final da disciplina Tópicos Avançados em Computadores - Turma D - 2018.1, do Departamento de Ciência da Computação da Universidade de Brasília, que trata da análise da produção científica e acadêmica na mesma universidade, em uma ou mais áreas de conhecimento.

O objetivo desse trabalho é compreender a situação dos seguintes cursos de pós-graduação: Antropologia, Ciência do Comportamento, Política Social e Sociologia. Cursos esses que têm o mesmo objetivo comum, que é desvendar os mistérios do ser humano e então sistematizar a forma como as pessoas agem, embora de perspectivas muito diferentes.

O antropólogo deriva a essência do ser humano no seu contexto cultural e biológico. O cientista do comportamento busca seu entendimento em questões neurais e cognitivas para o mesmo objetivo. Já o político social, por outro lado, tendo uma visão mais marxista do mundo, intenta corrigir os males da desigualdade, pobreza e classe por meio de estudos teóricos aprofundados. Por fim, a sociologia tem como objetivo final compreender o ser humano mergulhado na interação complexa que existe entre o indivíduo e a sociedade, a fim de definir o que é a sociedade que molda e o que é próprio do indivíduo.

É fácil notar que essas disciplinas, embora possuam suas particularidades, tem o mesmo objeto de estudo, que é o ser humano. Dessa forma, muito do que é pesquisado em uma das áreas pode ser de extrema valia para a outra. Assim, é comum haver certo sincretismo nos conhecimentos, que é naturalmente benéfico. Esse trabalho analisa as quatro disciplinas por essa proximidade.

A metodologia para desenvolvimento do estudo é baseada no modelo de mineração de dados denominado CRISP-DM (Chapman et al., 2000, Mariscal et al., 2010).

## CRISP-DM (Corresponderia à seção de Metodologia)

Para a condução de nosso estudo no programa de pós graduação, empregamos a metodologia CRISP-DM, que é desenhada especificamente para processos de mineração de dados. CRISP-DM (Cross Industry Standard Process for Data Mining) é formada por 6 etapas que descrevem como o estudo deve ser conduzido. Inicia-se com o Entendimento do Negócio, em que o cientista de dados busca entender detalhadamente o problema que irá ser resolvido. Em seguida, vem a Compreensão dos Dados, em que se organiza e se descreve todos os dados disponíveis, de forma a propiciar-se o início do estudo. Logo depois, o profissional executa a Preparação dos Dados, processo em que se escolhe o formato dos dados, seus atributos e padroniza-se eles. Por fim, vêm os processos de Modelagem, Avaliação e Desenvolvimento, em que o profissional, com os dados em mãos, emprega a estratégia de mineração de dados mais adequada, realiza a análise propriamente dita e recolhe os resultados obtidos.

# **CRISP-DM Fase 1 - Entendimento do Negócio**

## **Sistema Nacional de Pós-Graduação (Contextualização)**

O conhecimento científico produzido por uma universidade tem um grande papel no desenvolvimento da sociedade. As universidades de um país possuem diferentes áreas do conhecimento, mas no caso do Brasil, todas seguem o Sistema Nacional de Pós-Graduação que tem como objetivo lançar diretrizes na tentativa de padronizar os cursos de pós-graduação oferecidos por todo o país. Professores, alunos e os cursos oferecidos nas diferentes universidades brasileiras usam o Sistema Nacional de Pós-Graduação como base para os assuntos a serem discutidos em sala de aula e também em suas pesquisas, um bom exemplo disso são os assuntos mais relevantes para o país, dessa forma garante-se uma relação próxima entre o dinheiro investido nos programas de pós-graduação e o benefício colhido pela população que vem justamente como resultado dos estudos e pesquisas nas diferentes áreas. O Sistema Nacional de Pós-Graduação avalia de tempos em tempos os cursos oferecidos pelas universidades e aloca os recursos disponíveis baseando-se nessas avaliações, que levam em conta diversos fatores numa infinidade de áreas de pesquisa diferentes. Essas diversas áreas de pós-graduação são organizadas em forma de colégios, com áreas e grandes áreas, dessa forma uma determinada avaliação leva em conta cada área e suas particularidades.

## **A UnB dentro do Sistema Nacional de Pós-Graduação (Contextualização)**

### **O que é a UnB?**

A Universidade de Brasília tem uma grande história nas ciências humanas, teve sua fundação e seu início feito por grandes nomes das ciências sociais que são estudados hoje em sala de aula em todo o país. Desde sua fundação, a UnB possui programas de ajuda que buscam atrair alunos e pesquisadores para desenvolverem as diferentes áreas do conhecimento. Bolsas para mestrado e doutorado atraem diversos estudantes para que continuem na universidade após sua graduação, melhorando cada vez mais a própria universidade. (Talvez inserir um gráfico da produção acadêmica com o passar dos anos ? REL 3 JPTC)

Descrição da Universidade de Brasília, com foco na sua produção científica e acadêmica.

### **Descrição das pós-graduações da UnB**

Temos como foco três áreas do conhecimento, são elas a antropologia, ciências do comportamento e políticas sociais, cada uma dessas áreas possuem suas particularidades, mas são todas interligadas e altamente relacionadas com o estudo do indivíduo e da sociedade.

### **Antropologia**

O Departamento de Antropologia (DAN) na UnB é o departamento que abriga o Programa de Pós-Graduação em Antropologia Social. Curso que se destaca por receber a mais alta nota na avaliação CAPES de qualidade de curso universitário (DAN, 2018).

A expressividade na produção antropológica da UnB é tradição que vem da época em que Darcy Ribeiro, o primeiro reitor da universidade e doutor honoris causa dessa, produzia seus estudos aqui voltados ao entendimento da cultura indígena. A importância de Darcy se expressa na frase “A construção do campus brotou do cruzamento de mentes geniais. O inquieto antropólogo Darcy Ribeiro definiu as bases da instituição. Juntamente com ele, o educador Anísio Teixeira planejou o modelo pedagógico e o arquiteto Oscar Niemeyer transformou as ideias em prédios.” (UnB, 2018)

A qualidade do ensino é justamente a importância dada a esta área do conhecimento. Como está escrito na sua descrição (site): “Praticamente a totalidade dos nossos estudantes de pós-graduação recebe financiamento

para sua pesquisa sob a forma de bolsas de estudo e/ou verbas para realização de trabalho de campo e participação em eventos científicos.” Dessa forma, os recursos são o potencial dessa área de estudo na UnB. “O resultado dessa conjunção especial a unir qualidade de ensino e pesquisa, corpo docente altamente qualificado com elevada produção intelectual e infraestrutura física de boa qualidade só poderia ser um corpo discente bem formado, que tem recebido os principais prêmios científicos do país.”

Outra razão para sua qualidade é o volume e atualização de sua biblioteca. Quantitativamente, “Além da aquisição regular de livros traduzidos e publicados originalmente no Brasil, entre 1976 e 1990 foram comprados um total de 1503 livros no exterior, sendo que no biênio 89/91 foram adquiridos 512 exemplares, portanto 33,2% do total” (DAN, 2018)

Um destaque dentre as produções do departamento é o “Anuário Antropológico”. “Seu corpo editorial e organizador é integrado por docentes do Departamento, que tem dele participado substantivamente, com artigos, ensaios e resenhas. O Anuário foi iniciado em 1976 tendo alcançado até o presente um total de 42 edições” (DAN, 2018)

## **Ciência do Comportamento**

O programa de ensino sobre ciências do comportamento analisado possui cursos de mestrado e doutorado que têm como finalidade promover a base científica e gerar docentes e pesquisadores nesse campo da psicologia. Os cursos possuem duas grandes áreas de concentração: a análise do comportamento e a cognição e neurociências do comportamento.

A área da análise do comportamento tem como objetivo conhecer melhor os processos básicos do comportamento não só humano, mas também de animais e investigar os fenômenos associados a esses comportamentos, além de interpretá-los e saber como são influenciados por relações diversas estabelecidas por inúmeras razões naturais.

A área de concentração referente à cognição e neurociências do comportamento tem o seu foco voltado para produzir conhecimento sobre os processos que envolvem a parte cognitiva e da psicobiologia do comportamento humano e animal, mas diferente da análise do comportamento, deseja-se obter dados a respeito dos processos relacionados a memória, percepção, emoção e como o ambiente em que vive afeta esses processos.

Dentro dessas áreas de concentração ainda possuímos duas linhas de pesquisas distintas para cada uma. Em análise do comportamento temos as linhas processos comportamentais básicos e análise comportamental aplicada. Já na área da cognição e neurociência do comportamento temos pesquisas sobre cognição, percepção e sensação e também cérebro, drogas e comportamento.

A linha de pesquisas sobre processos comportamentais básicos busca compreender como as relações funcionais estabelecidas por contingências ambientais diversas afetam os processos básicos do comportamento, tanto em um contexto humano quanto animal.

A segunda linha de pesquisa dentro da mesma área de concentração refere-se a análise comportamental aplicada, que tem como objetivo principal investigar e interpretar fenômenos comportamentais que ocorrem em seus ambientes naturais usando metodologias obtidas de pesquisas experimentais no campo de aprendizagem. Na mesma linha, também procura-se saber como desenvolver e avaliar procedimentos que promovam o estabelecimento e manutenção de repertórios comportamentais eficazes na resolução de problemas em diferentes contextos.

Na outra área de concentração do programa de ensino da pós-graduação temos a primeira linha de pesquisa sendo a cognição, percepção e sensação. “Os objetivos dessa linha de pesquisa consistem em investigar, por meio de métodos experimentais ou observacionais, sob quais condições ou fatores cognitivos, sensoriais e perceptuais influenciam o desempenho humano em sua interação com o ambiente, visando desenvolver modelos teóricos e estratégias de investigação em psicologia e neurociências do comportamento”.

A outra linha de pesquisa dessa área de concentração tem como tema cérebro, drogas e comportamento. Seu objetivo é avaliar o comportamento humano e como este afeta e é afetado pelo cérebro. Também tem

como foco avaliar a dependência de drogas psicotrópicas no comportamento e gerar pesquisas relacionadas a diversos transtornos de memória, ansiedade e estresse.

Os assuntos abordados nesse programa de ensino são de interesse mundial, linhas de pesquisas sobre comportamento podem resolver problemas diversos que há muito estão presentes no cotidiano de uma sociedade como a nossa. Dada a importância do tema, outras universidades do Brasil e também em outros países provavelmente possuem linhas de pesquisa semelhantes em seus programas de ensino contribuindo para a descoberta de melhores soluções para problemas semelhantes em todo o mundo.

## **Política Social**

O Programa de Pós-Graduação em Política Social (PPGPS) está vinculado ao Departamento de Serviço Social (SER), do Instituto de Ciências Humanas (IH), da Universidade de Brasília (UnB) e constitui importante espaço de formação, pesquisa e produção de conhecimento na área da Política Social. Desde 2013 está avaliado com o Conceito 6 pela CAPES. Forma anualmente mestres e doutores em Política Social e muitas teses e dissertações foram publicadas em livros e/ou artigos, e adotadas como referências para a formação em âmbito de graduação e pós-graduação. O PPGPS recebeu por três vezes o Prêmio Capes de Melhor Tese da Área de Serviço Social (2008, 2011 e 2014).

Desde seu início, o PPGPS mantém como área de concentração “Estado, Política Social e Cidadania”, e tem como eixo analítico a relação dialética e contraditória entre Estado, Classes Sociais e Direitos, mediada pelas lutas que decorrem do conflito entre capital e trabalho.

O corpo docente do Programa de Pós-Graduação em Política Social é composto atualmente por 21 professores/as orientadores/as, sendo 18 permanentes e 3 colaboradores/as.

## **O que a Organização precisa realmente alcançar?**

Vários stakeholders estão envolvidos no projeto em curso, e poderíamos considerar cada um deles como distintas organizações que possuem interesses distintos e complementares. Elas são: \* A Disciplina Ciência de Dados para Todos 2018.1, que quer comprovar que seus alunos dominam ferramentas e técnicas de ciência de dados, para fins de avaliação de rendimento da disciplina. \* A UnB, representada pelos decanatos de pós-graduação (DPG) e de pesquisa e inovação (DPI), que querem dispor de instrumentos para realização de avaliações contínuas de suas pós-graduações. \* O SNPG, que assim com o DPG e DPI, também pode se beneficiar do uso de instrumentos para realização de avaliações contínuas de suas pós-graduações. \* Os interessados em melhor conhecer o que é produzido pelo Sistema Nacional de Pós-graduação, como empresas privadas, que querem desfrutar dos benefícios gerados pela ciência brasileira.

A fim de dar maior fidelidade e homogeneidade ao exercício realizado na disciplina, focaremos em atendimento aos interesses comuns das organizações DPI, DPG e CAPES, que desejam dispor de instrumentos ágeis para avaliação contínua da pós-graduação brasileira.

Com base no exposto, o objetivo do trabalho final a ser alcançado pelos produtos de mineração de dados desenvolvido pelos alunos da disciplina Ciência de Dados para Todos é produzir, tomando por base inicial os dados fornecidos pelos professores responsáveis pela disciplina, ferramentas para análise e avaliação contínuas e de baixo custo, do desempenho de um conjunto de pós-graduações que estão vinculadas a uma mesma subárea ou grupo de conhecimento. Cada área de pós-graduação apresenta suas características peculiares, assim como cada um dos programas vinculados a essas áreas. Como já informado, características peculiares de cada programa podem ser obtidas a partir de visita ao site da CAPES (2018).

## CRISP-DM Fase 2 - Entendimento dos Dados

Doravante, a fim de facilitar aos alunos seguirem a metodologia CRISP-DM, os nomes das seções e subseções de texto serão prefixadas com o número e nome da fase e atividade genérica do CRISP-DM. Fica facultado aos grupos seguir ou não a sequência prevista, tendo em vista que se pode retornar às fases anteriores, bem como podem haver atividades que não foram adequadas às características do problema específico sob análise.

### CRISP-DM Fase.Atividade 2.1 - Coleta inicial dos dados

Todos os arquivos com dados iniciais a seguir apresentados foram fornecidos pelos professores responsáveis pela disciplina. Os dados foram gerados no mês de maio de 2018, e compilam informações entre os anos de 2010 e 2017. Os arquivos estão no formato JSON, e seus atributos iniciais e conteúdos são apresentados a seguir.

#### Perfil profissional dos docentes vinculados às pós-graduações

```
PER_ant <- fromJSON("236.profile.json")
PER_soc <- fromJSON("328.profile.json")
PER_pol <- fromJSON("314.profile.json")
PER_cie <- fromJSON("255.profile.json")
```

Os arquivos acima apresentam dados sobre o perfil dos docentes vinculados a programas de pós-graduação de antropologia, sociologia, política social e ciência do comportamento da UnB, entre 2010 e 2017. Esse arquivo foi fornecido pelos docentes responsáveis pela disciplina.

#### Orientações de mestrado e doutorado realizadas pelos docentes vinculados às pós-graduações

```
ORI_ant <- fromJSON("236.advise.json")
ORI_soc <- fromJSON("328.advise.json")
ORI_pol <- fromJSON("314.advise.json")
ORI_cie <- fromJSON("255.advise.json")
```

Os arquivos com sufixo “ORI” apresenta dados sobre o orientações de mestrado e doutorado feitas por todos os docentes vinculados aos respectivos programas de pós-graduação da UnB, entre 2010 e 2017. Esse arquivo foi fornecido pelos docentes responsáveis pela disciplina.

#### Agrupamento dos docentes conforme áreas de atuação

```
GRA_ant <- fromJSON("236.graph.json")
GRA_soc <- fromJSON("328.graph.json")
GRA_pol <- fromJSON("314.graph.json")
GRA_cie <- fromJSON("255.graph.json")
```

#### Redes de colaboração entre docentes

Os arquivos com sufixo GRA apresentam redes de colaboração na co-autoria de artigos científicos, feitas entre os docentes vinculados a programas de pós-graduação da UnB, entre 2010 e 2017.

## CRISP-DM Fase.Atividade 2.2 - Descrição dos Dados

A leitura e manipulação dos dados é uma das principais responsáveis por uma boa análise de dados, para que possamos analisar de forma eficiente, as seguintes bibliotecas foram utilizadas:

```
library(jsonlite)
library(listviewer)
library(readxl)
library(readr)
library(readtext)
library(ggplot2)
library(tidyverse)
```

### Descrição dos dados do perfil

O arquivo unb.perfis.json, que contém dados que caracterizam o perfil profissional de todos os docentes do grupo sob análise, podem ser lido por meio do comando seguinte.

```
PER_ant <- fromJSON("236.profile.json")
PER_soc <- fromJSON("328.profile.json")
PER_pol <- fromJSON("314.profile.json")
PUB_cie <- fromJSON("255.publication.json")
```

A quantidade de docentes sob análise é apresentada a seguir.

```
length(PER_ant)
```

```
## [1] 26
```

```
length(PER_soc)
```

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

```
length(PER_pol)
```

```
## [1] 21
```

```
length(PER_cie)
```

```
## [1] 20
```

Para gerar uma apresentação inicial dos dados que estão contido nos dados de perfil dos docentes, pode-se usar a função glimpse, da biblioteca dplyr, como ilustra o código seguinte, que apresenta os atributos típicos que podem ser obtidos relativamente a um pesquisador específico, o mais antigo docente ainda em exercício na UnB a ter criado seu registro na plataforma Lattes.

```
library(dplyr)
glimpse(PER_ant[[1]], width = 30)
```

```
## List of 7
## $ nome                : chr "Sílvia Maria Ferreira Guimarães"
## $ resumo_cv            : chr "Possui graduação em Ciências Sociais (1998), mestrado em Antropologia"
## $ areas_de_atuacao     : 'data.frame': 4 obs. of 4 variables:
## ..$ grande_area       : chr [1:4] "CIENCIAS_HUMANAS" "CIENCIAS_HUMANAS" "CIENCIAS_HUMANAS" "CIENCIAS_HUMANAS"
## ..$ area              : chr [1:4] "Antropologia" "Antropologia" "Antropologia" "Antropologia"
## ..$ sub_area          : chr [1:4] "Teoria Antropológica" "Etnologia indígena" "Antropologia da Saúde" "Antropologia da Saúde"
## ..$ especialidade     : chr [1:4] "" "" "" ""
## $ endereco_profissional :List of 8
```

```

## ..$ instituicao: chr "Universidade de Brasília"
## ..$ orgao      : chr "Departamento de Antropologia"
## ..$ unidade    : chr ""
## ..$ DDD        : chr "61"
## ..$ telefone   : chr "31071550"
## ..$ bairro     : chr "Asa Norte"
## ..$ cep        : chr "70910900"
## ..$ cidade     : chr "Brasília"
## $ producao_bibliografica :List of 5
## ..$ CAPITULO_DE_LIVRO      :'data.frame':  13 obs. of  13 variables:
## .. ..$ tipo                : chr [1:13] "Capítulo de livro publicado" "Capítulo de livro publi
## .. ..$ titulo_do_capitulo   : chr [1:13] "Caminhos e Territórios Guarani-Mbyá" "Metamorfoses Sa
## .. ..$ titulo_do_livro      : chr [1:13] "Línguas e Culturas Tupí" "Variações interétnicas: etn
## .. ..$ ano                  : chr [1:13] "2010" "2012" "2012" "2015" ...
## .. ..$ doi                  : chr [1:13] "" "" "" "" ...
## .. ..$ pais_de_publicacao   : chr [1:13] "Brasil" "Brasil" "Brasil" "Brasil" ...
## .. ..$ isbn                 : chr [1:13] "9788599944202" "9788573003628" "9788580424751" "97885
## .. ..$ nome_da_editora      : chr [1:13] "Curt Nimuendajú" "IBAMA/CEPPAC/IEB" "Editora CRV" "Ed
## .. ..$ numero_da_edicao_revisao: chr [1:13] "" "" "1" "1" ...
## .. ..$ organizadores       : chr [1:13] "CABRAL, A. S. A; RODRIGUES, A; DUARTE, F." "BAINES, S
## .. ..$ paginas              : chr [1:13] "149 - 159" "497 - 510" "133 - 150" "363 - " ...
## .. ..$ autores              :List of 13
## .. ..$ autores-endogeno     :List of 13
## ..$ DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA:'data.frame':  1 obs. of  9 variables:
## .. ..$ natureza            : chr "Mesa de debate"
## .. ..$ titulo               : chr "Cine diálogos socioambientais e quartas sustentáveis-sagrada terra
## .. ..$ ano                  : chr "2011"
## .. ..$ pais_de_publicacao   : chr "Brasil"
## .. ..$ editora              : chr ""
## .. ..$ doi                  : chr ""
## .. ..$ numero_de_paginas    : chr ""
## .. ..$ autores              :List of 1
## .. ..$ autores-endogeno     :List of 1
## ..$ EVENTO                  :'data.frame':  50 obs. of  11 variables:
## .. ..$ natureza            : chr [1:50] "RESUMO" "RESUMO" "RESUMO" "RESUMO" ...
## .. ..$ titulo               : chr [1:50] "O sistema médico Sanumá-Yanomami no contexto de atenção à saú
## .. ..$ nome_do_evento       : chr [1:50] "V Congresso Brasileiro de Ciências Sociais e Humanas em Saúde
## .. ..$ ano_do_trabalho      : chr [1:50] "2011" "2011" "2011" "2011" ...
## .. ..$ pais_do_evento       : chr [1:50] "Brasil" "Brasil" "Brasil" "Brasil" ...
## .. ..$ cidade_do_evento     : chr [1:50] "São Paulo" "Belém" "Boa Vista" "Fortaleza" ...
## .. ..$ doi                  : chr [1:50] "" "" "" "" ...
## .. ..$ classificacao        : chr [1:50] "NACIONAL" "NAO_INFORMADO" "NACIONAL" "INTERNACIONAL" ...
## .. ..$ paginas              : chr [1:50] " - " "28 - 28" " - " "39 - 39" ...
## .. ..$ autores              :List of 50
## .. ..$ autores-endogeno     :List of 50
## ..$ LIVRO                    :'data.frame':  1 obs. of  13 variables:
## .. ..$ titulo               : chr "Antropologia e saúde: diálogos indisciplinados"
## .. ..$ ano                  : chr "2017"
## .. ..$ tipo                 : chr "LIVRO_PUBLICADO"
## .. ..$ natureza             : chr "COLETANEA"
## .. ..$ pais_de_publicacao   : chr "Brasil"
## .. ..$ isbn                 : chr "9788593128196"
## .. ..$ doi                  : chr ""
## .. ..$ nome_da_editora      : chr "Editora da UFJF"

```

```

## .. .$ numero_da_edicao_revisao: chr "1"
## .. .$ numero_de_paginas      : chr ""
## .. .$ numero_de_volumes      : chr "1"
## .. .$ autores                :List of 1
## .. .$ autores-endogeno       :List of 1
## ..$ PERIODICO                  :'data.frame':  9 obs. of  10 variables:
## .. .$ natureza                : chr [1:9] "COMPLETO" "COMPLETO" "COMPLETO" "COMPLETO" ...
## .. .$ titulo                  : chr [1:9] "Corpos e ciclos da vida sanumá-yanomami" "O drama ritual da mo
## .. .$ periodico               : chr [1:9] "Horizontes Antropológicos (UFRGS. Impresso)" "Tellus (UCDB)" "
## .. .$ ano                     : chr [1:9] "2010" "2010" "2011" "2013" ...
## .. .$ volume                  : chr [1:9] "16" "19" "5" "56" ...
## .. .$ issn                    : chr [1:9] "01047183" "15199452" "19828829" "00347701" ...
## .. .$ paginas                 : chr [1:9] "261 - 286" "111 - 128" "57 - 72" "181 - 205" ...
## .. .$ doi                     : chr [1:9] "10.1590/S0104-71832010000200012" "" "" "" ...
## .. .$ autores                 :List of 9
## .. .$ autores-endogeno:List of 9
## $ orientacoes_academicas:List of 6
## ..$ ORIENTACAO_CONCLUIDA_MESTRADO                  :'data.frame': 10 obs. of  13 variables:
## .. .$ natureza                : chr [1:10] "Dissertação de mestrado" "Dissertação de mestrado"
## .. .$ titulo                  : chr [1:10] "Dilemas na construção e efetivação da política naci
## .. .$ ano                     : chr [1:10] "2015" "2016" "2016" "2017" ...
## .. .$ id_lattes_aluno         : chr [1:10] "1986784071425235" "3082949565387056" "940442666772
## .. .$ nome_aluno              : chr [1:10] "Alane Andreino Ribeiro" "Jéssica Camila de Sousa
## .. .$ instituicao              : chr [1:10] "Universidade de Brasília" "Universidade de Brasília
## .. .$ curso                   : chr [1:10] "Pós-Graduação. em Ciências e Tecnologias em Saúde"
## .. .$ codigo_do_curso         : chr [1:10] "90000017" "90000017" "90000017" "90000017" ...
## .. .$ bolsa                   : chr [1:10] "NAO" "NAO" "NAO" "NAO" ...
## .. .$ agencia_financiadora    : chr [1:10] "" "" "" "" ...
## .. .$ codigo_agencia_financiadora: chr [1:10] "" "" "" "" ...
## .. .$ nome_orientadores       :List of 10
## .. .$ id_lattes_orientadores  :List of 10
## ..$ ORIENTACAO_EM_ANDAMENTO_DOUTORADO              :'data.frame':  3 obs. of  13 variables:
## .. .$ natureza                : chr [1:3] "Tese de doutorado" "Tese de doutorado" "Tese de dou
## .. .$ titulo                  : chr [1:3] "Interdisciplinaridade e produção científica e tecno
## .. .$ ano                     : chr [1:3] "2015" "2016" "2016"
## .. .$ id_lattes_aluno         : chr [1:3] "1734397644031826" "6795640994074965" ""
## .. .$ nome_aluno              : chr [1:3] "Mario Fabrício Fleury Rosa" "Caroline de Oliveira A
## .. .$ instituicao              : chr [1:3] "Universidade de Brasília" "Universidade de Brasília
## .. .$ curso                   : chr [1:3] "Pós-Graduação. em Ciências e Tecnologias em Saúde"
## .. .$ codigo_do_curso         : chr [1:3] "90000017" "90000017" "90000017"
## .. .$ bolsa                   : chr [1:3] "SIM" "SIM" "NAO"
## .. .$ agencia_financiadora    : chr [1:3] "" "Conselho Nacional de Desenvolvimento Científico
## .. .$ codigo_agencia_financiadora: chr [1:3] "" "002200000000" ""
## .. .$ nome_orientadores       :List of 3
## .. .$ id_lattes_orientadores  :List of 3
## ..$ ORIENTACAO_EM_ANDAMENTO_GRADUACAO              :'data.frame':  4 obs. of  13 variables:
## .. .$ natureza                : chr [1:4] "Trabalho de conclusão de curso de graduação" "Traba
## .. .$ titulo                  : chr [1:4] "Suínos e domesticidade tardia" "ETNOBIOGRAFIAS DE V
## .. .$ ano                     : chr [1:4] "2017" "2017" "2017" "2017"
## .. .$ id_lattes_aluno         : chr [1:4] "" "" "" ""
## .. .$ nome_aluno              : chr [1:4] "André Macedo Ferreira" "Rosânia de Oliveira Nascime
## .. .$ instituicao              : chr [1:4] "Universidade de Brasília" "Universidade de Brasília
## .. .$ curso                   : chr [1:4] "Ciências Sociais" "Ciências Sociais" "Ciências Soci
## .. .$ codigo_do_curso         : chr [1:4] "60252650" "60252650" "60235306" "60252650"

```



```

## ..$ bolsa : chr [1:4] "NAO" "NAO" "NAO" "NAO"
## ..$ agencia_financiadora : chr [1:4] "" "" "" ""
## ..$ codigo_agencia_financiadora: chr [1:4] "" "" "" ""
## ..$ nome_orientadores :List of 4
## ..$ id_lattes_orientadores :List of 4
## ..$ ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA:'data.frame': 4 obs. of 13 variables:
## ..$ natureza : chr [1:4] "Iniciação Científica" "Iniciação Científica" "Inici
## ..$ titulo : chr [1:4] "Mapeando a Saúde entre os Tikuna" "Práticas de cuid
## ..$ ano : chr [1:4] "2017" "2017" "2017" "2017"
## ..$ id_lattes_aluno : chr [1:4] "" "" "" ""
## ..$ nome_aluno : chr [1:4] "Iury da Costa Felipe" "Vitor Lucio Ferrras de Arauj
## ..$ instituicao : chr [1:4] "Universidade de Brasília" "Universidade de Brasília
## ..$ curso : chr [1:4] "Abi - Ciências Sociais" "Enfermagem" "Medicina" "Abi
## ..$ codigo_do_curso : chr [1:4] "60662107" "60284269" "60224770" "60662107"
## ..$ bolsa : chr [1:4] "NAO" "SIM" "NAO" "SIM"
## ..$ agencia_financiadora : chr [1:4] "" "Decanato de Pós-Graduação" "" "Decanato de Pós-G
## ..$ codigo_agencia_financiadora: chr [1:4] "" "001800000992" "" "001800000992"
## ..$ nome_orientadores :List of 4
## ..$ id_lattes_orientadores :List of 4
## ..$ ORIENTACAO_EM_ANDAMENTO_MESTRADO :'data.frame': 3 obs. of 13 variables:
## ..$ natureza : chr [1:3] "Dissertação de mestrado" "Dissertação de mestrado"
## ..$ titulo : chr [1:3] "FARMÁCIA VIVA: FORTALECIMENTO DA MEDICINA ANCESTRAL
## ..$ ano : chr [1:3] "2016" "2016" "2016"
## ..$ id_lattes_aluno : chr [1:3] "" "9830970369575104" "9764033878314492"
## ..$ nome_aluno : chr [1:3] "Alan Miguel Alves" "Melina Soares Rodrigues" "ANA C
## ..$ instituicao : chr [1:3] "Universidade de Brasília" "Universidade de Brasília
## ..$ curso : chr [1:3] "Sustentabilidade junto a Povos e Terras Tradiciona
## ..$ codigo_do_curso : chr [1:3] "90000018" "90000017" "90000017"
## ..$ bolsa : chr [1:3] "NAO" "SIM" "NAO"
## ..$ agencia_financiadora : chr [1:3] "" "Conselho Nacional de Desenvolvimento Científico
## ..$ codigo_agencia_financiadora: chr [1:3] "" "002200000000" ""
## ..$ nome_orientadores :List of 3
## ..$ id_lattes_orientadores :List of 3
## ..$ OUTRAS_ORIENTACOES_CONCLUIDAS :'data.frame': 66 obs. of 13 variables:
## ..$ natureza : chr [1:66] "ORIENTACAO-DE-OUTRA-NATUREZA" "ORIENTACAO-DE-OUTRA
## ..$ titulo : chr [1:66] "Práticas e noções do corpo entre terapeutas popula
## ..$ ano : chr [1:66] "2010" "2010" "2011" "2011" ...
## ..$ id_lattes_aluno : chr [1:66] "" "" "" "" ...
## ..$ nome_aluno : chr [1:66] "Grasiela dos Santos da Silva" "Elias Rodrigues Barl
## ..$ instituicao : chr [1:66] "Universidade de Brasília" "Universidade de Brasília
## ..$ curso : chr [1:66] "Saúde Coletiva" "Saúde Coletiva" "Saúde Coletiva"
## ..$ codigo_do_curso : chr [1:66] "90000009" "90000009" "90000009" "90000009" ...
## ..$ bolsa : chr [1:66] "SIM" "SIM" "NAO" "SIM" ...
## ..$ agencia_financiadora : chr [1:66] "Decanato de Assuntos Comunitários" "Decanato de Ass
## ..$ codigo_agencia_financiadora: chr [1:66] "001400000995" "001400000995" "" "002200000000" ...
## ..$ nome_orientadores :List of 66
## ..$ id_lattes_orientadores :List of 66
## $ senioridade : chr "9"

```

## Descrição dos dados de orientações

```

ORI_ant <- fromJSON("236.advise.json")
names(ORI_ant)

```

```

## [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"

names(ORI_ant$ORIENTACAO_CONCLUIDA_DOUTORADO)

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

length(ORI_ant$ORIENTACAO_CONCLUIDA_DOUTORADO$`2016`$natureza)

## [1] 8

head(sort(table(ORI_ant$ORIENTACAO_CONCLUIDA_DOUTORADO$`2017`$curso), decreasing = TRUE), 10)

##
##                               Antropologia
##                               3
##                               Antropologia Social
##                               3
##                               Direito
##                               3
##                               Doutorado em Antropologia - PPGAS/UnB
##                               1
##                               Literatura
##                               1
## Programa de Pós-Graduação em Antropologia Social
##                               1
## Programa de Pós Graduação Em Antropologia Social
##                               1

head(sort(table(ORI_ant$ORIENTACAO_CONCLUIDA_MESTRADO$`2017`$curso), decreasing = TRUE), 10)

##
##                               Antropologia
##                               9
##                               Pós-Graduação. em Ciências e Tecnologias em Saúde
##                               4
##                               Sustentabilidade junto a Povos e Terras Tradiciona
##                               3
##                               Antropologia Social
##                               2
##                               Desenvolvimento Sustentável
##                               2
## Mestrado Profissional em Sustentabilidade junto a Povos e Terras Tradiciona
##                               2

```

## Descrição dos dados de produção bibliográfica

```

PUB_ant <- fromJSON("236.publication.json")
names(PUB_ant)

```

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

```
names(PUB_ant$PERIODICO$`2012`)
```

```
## [1] "natureza"      "titulo"         "periodico"
## [4] "ano"           "volume"         "issn"
## [7] "paginas"       "doi"            "autores"
## [10] "autores-endogeno"
```

```
head(sort(table(PUB_ant$PERIODICO$`2017`$periodico), decreasing = TRUE), 10)
```

```
##
##      INTERETHNIC@ - REVISTA DE ESTUDOS EM RELAÇÕES INTERÉTNICAS
##                                     3
##      ANUÁRIO ANTROPOLÓGICO
##                                     2
##      REVISTA DE ANTROPOLOGIA
##                                     2
##      Abya-Yala: Revista sobre Acesso à Justiça e Direitos nas Américas
##                                     1
##      Anales de Antropología
##                                     1
##      Análise Social (Lisboa)
##                                     1
##      ANTROPOLÍTICA: REVISTA CONTEMPORÂNEA DE ANTROPOLOGIA
##                                     1
##      Cadernos de Gênero e Diversidade
##                                     1
##      CADERNOS PAGU
##                                     1
##      DÉJÀ LU
##                                     1
```

```
head(sort(table(PUB_ant$LIVRO$`2015`$nome_da_editora), decreasing = TRUE), 10)
```

```
##
##
##                                     1
##      Editora Universidade de Brasília
##                                     1
##      Fondo de Cultura Económica/Colegio de México
##                                     1
##      Langaa
##                                     1
```

## Descrição dos dados de redes de colaboração

### CRISP-DM Fase.Atividade 2.3 - Análise exploratória dos dados

Como já informado, a análise exploratória dos dados possibilita um entendimento mais profundo da relação estatística existente entre os dados dos *datasets* para um melhor entendimento da qualidade daqueles dados para os objetivos do projeto.

!!!Como já informado, a análise exploratória dos dados é responsabilidade parcial dos alunos, tendo em vista que este relatório apresenta uma análise exploratória preliminar. O relatório final deve conter análises exploratórias dos dados que sejam significativas e aprofundadas.

### CRISP-DM Fase.Atividade 2.4 - Verificação da qualidade dos dados.

Como já informado, a verificação da qualidade dos dados envolve responder se os dados disponíveis estão realmente completos.

As informações disponíveis são suficientes para o trabalho proposto? Neste projeto, a verificação da qualidade dos dados é responsabilidade dos alunos.

## CRISP-DM Fase 3 - Preparação dos Dados

Como já informado, na fase de **Preparação dos Dados** os *datasets* que serão utilizados em todo o trabalho são construídos a partir dos dados brutos. Aqui os dados são “filtrados” retirando-se partes que não interessam e selecionando-se os “campos” necessários para o trabalho de mineração.

São 5 as atividades genéricas nesta fase de preparação dos dados, a seguir divididas em subseções

### CRISP-DM Fase.Atividade 3.1 - Seleção dos dados.

Como já informado, a seleção dos dados envolve identificar quais dados, da nossa “montanha de dados”, serão realmente utilizados.

Quais variáveis dos dados brutos serão convertidas para o *dataset*?

Não é raro cometer o erro de selecionar dados para um modelo preditivo com base em uma falsa ideia de que aqueles dados contém a resposta para o modelo que se quer construir. Surge o cuidado de se separar o sinal do ruído (Silver, Nate. The Signal and the Noise: Why so many predictions fail - but some don't. USA: The Penguin Press HC, 2012.).

### CRISP-DM Fase.Atividade 3.2 - Limpeza dos dados

### CRISP-DM Fase.Atividade 3.3 - Construção dos dados

Como já informado, a construção dos dados envolve a criação de novas variáveis a partir de outras presentes nos *datasets*.

### CRISP-DM Fase.Atividade 3.4 - Integração dos dados

Nessa fase damos ênfase na criação de datasets com diferentes tipos de dados, que foram unidos a partir dos arquivos iniciais.

## **CRISP-DM Fase.Atividade 3.5 - Formatação dos dados**

A formatação de dados aborda a parte de modificação e adequação dos arquivos iniciais para facilitar a análise a depender do método utilizado de mineração de dados.

## **CRISP-DM Fase 4 - Modelagem**

A modelagem de dados é referente a forma como a análise dos dados é estruturada, seja por meio de uma abordagem estatística ou computacional, é a lógica por trás de como funciona a análise.

### **CRISP-DM Fase.Atividade 4.1 - Seleção das técnicas de modelagem**

Diferentes técnicas de modelagem podem ser utilizadas para analisar o conjunto de dados trabalhado, podem ser basicamente modelos que puxam mais para a análise estatística ou computacional, embora os dois sempre são necessários nessa fase.

### **CRISP-DM Fase.Atividade 4.2 - Realização de testes de modelagem**

Na realização de testes de modelagem diferentes modelos estatísticos ou computacionais são previamente testados e avaliados. Pode-se dividir o conjunto de dados criado anteriormente para se ter uma base de treino na construção de modelos.

## **CRISP-DM Fase 5 - Avaliação**

A avaliação do modelo e da análise de dados feita nos diz se os resultados obtidos realmente possuem sentido e se são condizentes com o esperado para o que foi feito.

### **CRISP-DM Fase.Atividade 5.1 - Avaliação dos resultados**

### **CRISP-DM Fase.Atividade 5.2 - Revisão do processo**

Caso o modelo tenha sido adequado para solucionar o problema proposto a revisão serve para garantir que o processo foi bem sucedido usando esse modelo. Bom revisar e levar em consideração também o caso de poder utilizar o modelo para outras análises.

## **CRISP-DM Fase 6 - Implantação**

A ultima fase do CRISP-DM se refere à entrega final do produto a um usuário, que passaria a utilizar esse produto regularmente. Nessa etapa o produto já está feito e pronto para ser entregue, faltando no máximo pequenos ajustes para se adequar ao que o usuário precisa. A repetição da análise deve ser garantida, visto que os dados sempre estão em constante mudança, logo a análise precisa funcionar sem que ocorra problemas com outros dados além dos disponíveis inicialmente.

## CRISP-DM Fase.Atividade 6.1 - Planejamento da transição

A parte de transição diz respeito a aplicabilidade das análises feitas aqui para o usuário final, aquele que realmente precisa dos resultados dessa análise. Para que seja garantido o bom planejamento da transição o cliente deve conseguir reproduzir a mesma análise e obter resultados satisfatórios e dentro do que havia sido pedido.

## CRISP-DM Fase.Atividade 6.2 - Planejamento do monitoramento dos produtos

O monitoramento busca a realização de atividades voltadas a garantir o correto funcionamento da análise, nem sempre o profissional responsável pela análise de dados estará disponível para ajudar o cliente, logo o monitoramento tem como fim evitar qualquer tipo de problema que o usuário final venha a ter na hora da realização da análise.

## CRISP-DM Fase.Atividade 6.3 - Planejamento de manutenção

A manutenção do que foi desenvolvido dá ao usuário final a segurança que ele não ficará sem qualquer tipo de suporte, mesmo se seus dados mudarem um pouco ou algum outro pequeno problema acontecer, a análise não pode ficar inutilizada totalmente.

## CRISP-DM Fase.Atividade 6.4 - Produção do relatório final

A entrega do relatório do grupo, tomando como base este aqui, reflete a execução desta etapa.

## Visualizações

```
ORI_ant <- ORI_ant %>%
  lapply(function(x) {
    bind_rows(x, .id = 'ano')
  }) %>%
  bind_rows(.id = 'tipo_orientacao')

ORI_ant$nome_orientadores <- as.character(ORI_ant$nome_orientadores)

ORI_pol <- ORI_pol %>%
  lapply(function(x) {
    bind_rows(x, .id = 'ano')
  }) %>%
  bind_rows(.id = 'tipo_orientacao')

ORI_pol$nome_orientadores <- as.character(ORI_pol$nome_orientadores)

ORI_cie <- ORI_cie %>%
  lapply(function(x) {
    bind_rows(x, .id = 'ano')
  }) %>%
  bind_rows(.id = 'tipo_orientacao')

ORI_cie$nome_orientadores <- as.character(ORI_cie$nome_orientadores)
```

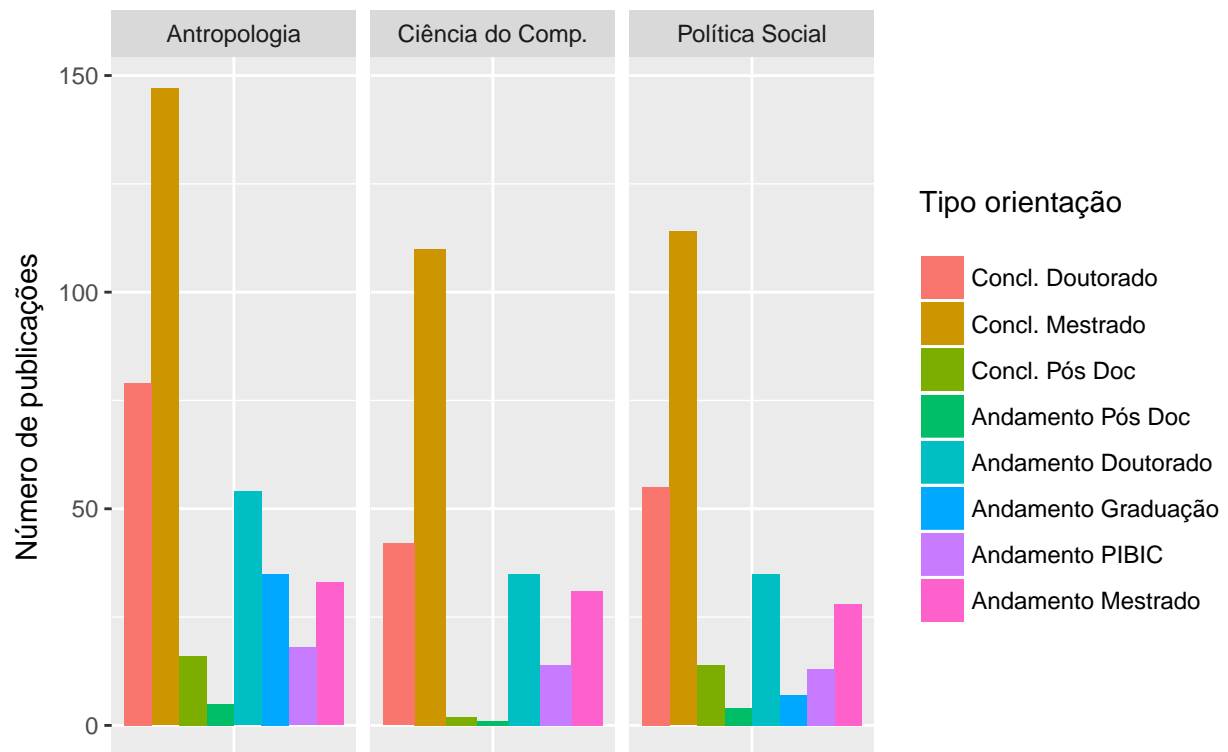
```
ORI_ant <- ORI_ant %>%
  mutate(programa = 'Antropologia')

ORI_pol <- ORI_pol %>%
  mutate(programa = 'Política Social')

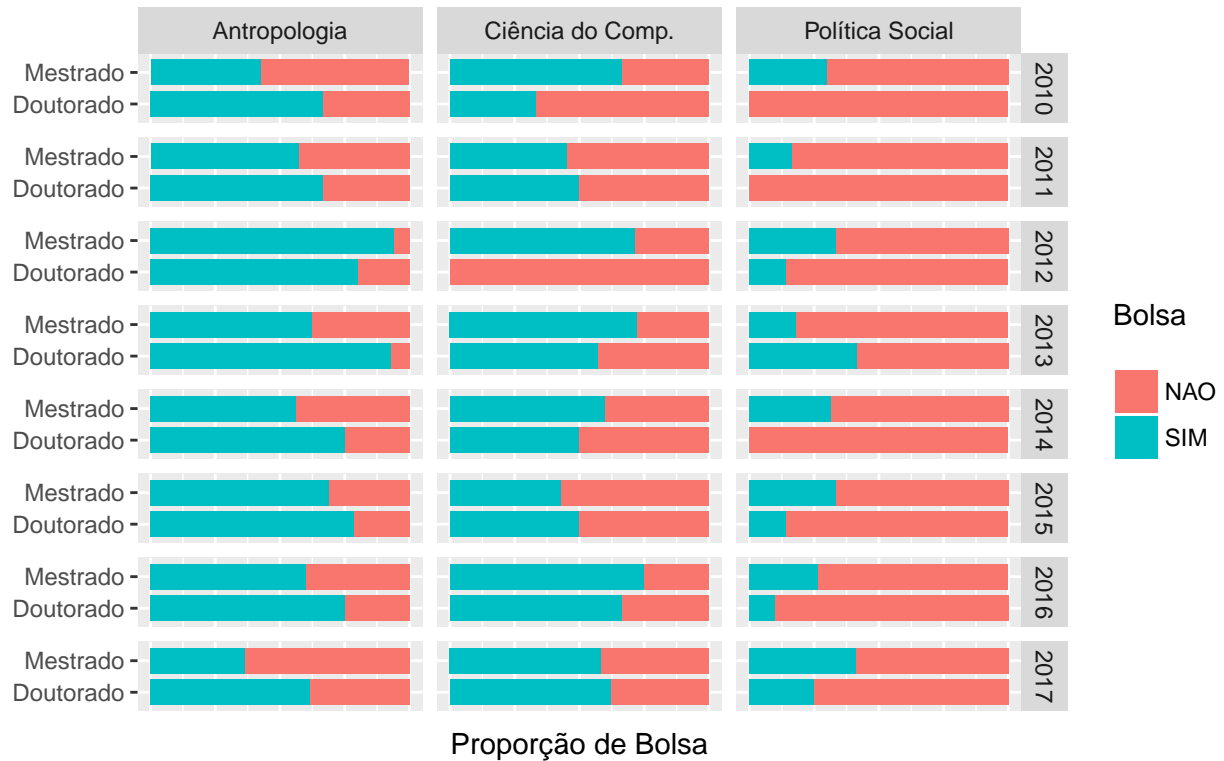
ORI_cie <- ORI_cie %>%
  mutate(programa = 'Ciência do Comp.')

ORI <- rbind(ORI_ant, ORI_pol, ORI_cie)
```

## Número de publicações por tipo

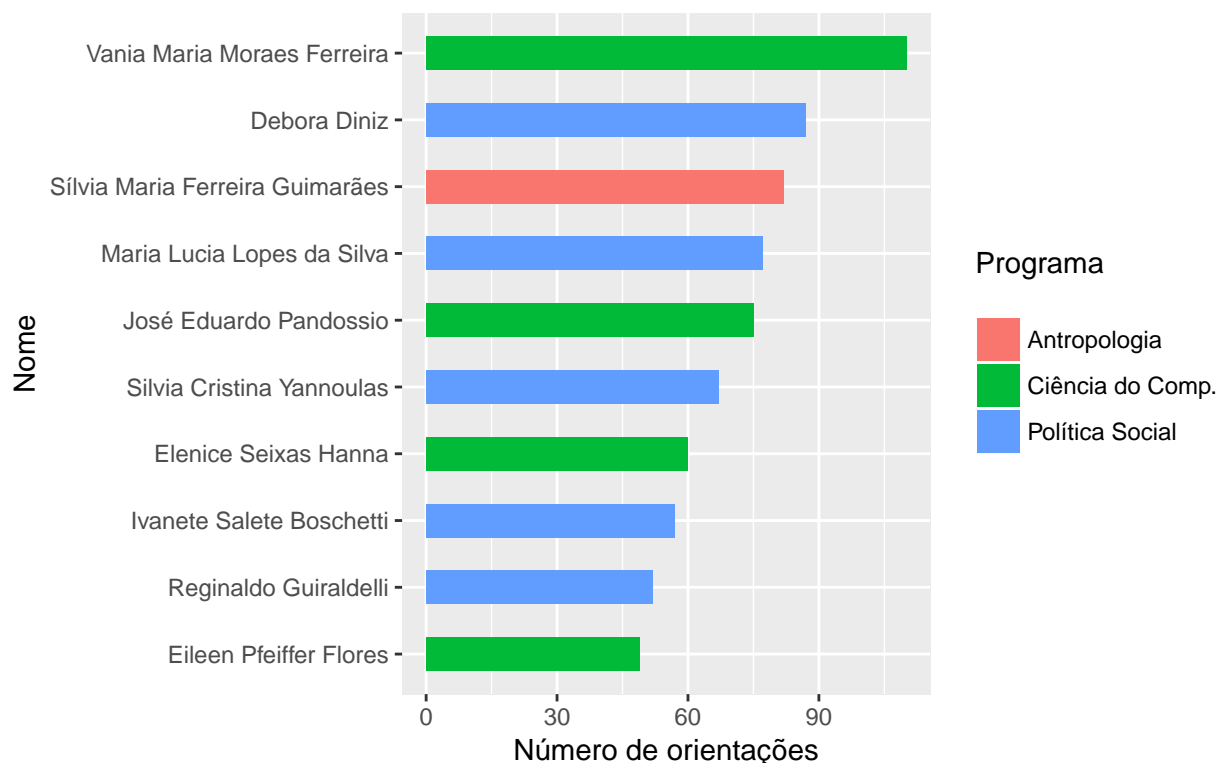


## Número de publicações por tipo por bolsa





## Número de orientações por professor



## Referências

- Azevedo, Mário Luiz Neves de, João Ferreira de Oliveira, e Afrânio Mendes Catani. “O Sistema Nacional de Pós-Graduação (SNPG) e o Plano Nacional de Educação (PNE 2014-2024): regulação, avaliação e financiamento”. *Revista Brasileira de Política e Administração da Educação* 32, nº 3 (2016). <http://dx.doi.org/10.21573/vol32n32016.68576>.
- Can, Fazli, Tansel Özyer, e Faruk Polat, orgs. *State of the Art Applications of Social Network Analysis. Lecture Notes in Social Networks*. Switzerland: Springer International Publishing, 2014.
- CAPES. “Documentos de Área”. [CAPES.gov.br](http://capes.gov.br). Acessado 12 de junho de 2018. <http://avaliacaoquadrienal.capes.gov.br/documentos-de-area>.
- . “Plano Nacional de Pós-Graduação - PNPG 2011/2020 Vol. 1”. Brasília - DF, dezembro de 2010. <http://www.capes.gov.br/images/stories/download/Livros-PNPG-Volume-I-Mont.pdf>.
- . “Plano Nacional de Pós-Graduação - PNPG 2011/2020 Vol. 2”. Brasília - DF, dezembro de 2010. [http://www.capes.gov.br/images/stories/download/PNPG\\_Miolo\\_V2.pdf](http://www.capes.gov.br/images/stories/download/PNPG_Miolo_V2.pdf).
- . “Sucupira: coleta de dados, docentes de pós-graduação stricto sensu no Brasil 2015”. CAPES - Banco de Metadados, 16 de março de 2016. <http://metadados.capes.gov.br/index.php/catalog/63>.
- Chapman, Pete, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, e Rüdiger Wirth. “CRISP-DM 1.0: Step-by-Step Data Mining Guide”. USA: CRISP-DM Consortium, 2000. <https://www.the-modeling-agency.com/crisp-dm.pdf>. DAN. *Corpo Docente*. Departamento de Antropologia da UnB. Disponível em: <http://www.dan.unb.br/dan-corporodocente>. Acesso em: 31 mai. 2018 DAN. História do DAN/UnB. Departamento de Antropologia da UnB. Disponível em: <http://www.dan.unb.br/dan-historia>. Acesso em: 31 mai. 2018 UnB.
- Datacamp. “Machine Learning with R (Skill Track)”. Datacamp, 2018. <https://www.datacamp.com/tracks/machine-learning>.

- Fernandes, Jorge H C, e Ricardo Barros Sampaio. “DataScienceForAll”. Zotero, 13 de junho de 2018. <https://www.zotero.org/groups/2197167/datascienceforall>.
- —. “Especificação do Trabalho Final da Disciplina de Ciência de Dados para Todos 2017.2: Estudo sobre a visibilidade internacional da produção científica das pós-graduações vinculadas às áreas de conhecimento da CAPES, na Universidade de Brasília (Comunicação Interna)”. Disciplina 116297 - Tópicos Avançados em Computadores, turma D, do semestre 2017.2, do Departamento de Ciência da Computação do Instituto de Ciências Exatas da Universidade de Brasília, 28 de novembro de 2017. [https://aprender.ead.unb.br/pluginfile.php/474549/mod\\_resource/content/1/Estudo%20da%20Cie%CC%82ncia.pdf](https://aprender.ead.unb.br/pluginfile.php/474549/mod_resource/content/1/Estudo%20da%20Cie%CC%82ncia.pdf).
- Fernandes, Jorge H C, Ricardo Barros Sampaio, e João Ribas de Moura. “Ciência de Dados para Todos (Data Science For All) - 2018.1 - Análise da Produção Científica e Acadêmica da Universidade de Brasília - Modelo de Relatório Final da Disciplina - Departamento de Ciência da Computação da UnB”. Disciplina 116297 - Tópicos Avançados em Computadores, turma D, do semestre 2018.1, do Departamento de Ciência da Computação do Instituto de Ciências Exatas da Universidade de Brasília, 13 de junho de 2018.
- Frickel, Scott, e Kelly Moore. *The New Political Sociology of Science: Institutions, Networks, and Power*. Science and technology in society. USA: The University of Wisconsin Press, 2006.
- Graduate Prospects Ltd. “Job profile: Higher education lecturer”, 2018. <https://www.prospects.ac.uk/job-profiles/higher-education-lecturer>.
- Kalpazidou Schmidt, Evanthia, e Ebbe Krogh Graversen. “Persistent factors facilitating excellence in research environments”. *Higher Education* 75, n° 2 (1º de fevereiro de 2018): 341-63. <https://doi.org/10.1007/s10734-017-0142-0>.
- Kilduff, Martin, e Wenpin Tsai. *Social Networks and Organizations*. UK: Sage Publications, 2003.
- Kolaczyk, Eric D., e Gábor Csárdi. *Statistical Analysis of Network Data with R*. USA: Springer, 2014.
- Kuhn, Max, Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, et al. “Package ‘Caret’ - Classification and Regression Training”, 27 de maio de 2018. <https://cran.r-project.org/web/packages/caret/caret.pdf>.
- Leite, Fernando César Lima. “Considerações básicas sobre a Avaliação do Sistema Nacional de Pós-Graduação”. Comunicação Pessoal (slides). Universidade de Brasília, abril de 2018. [https://aprender.ead.unb.br/pluginfile.php/502250/mod\\_resource/content/1/Considera%C3%A7%C3%B5es%20b%C3%A1sicas%20sobre%20a%20Avalia%C3%A7%C3%A3o%20do%20Sistema%20Nacional.pdf](https://aprender.ead.unb.br/pluginfile.php/502250/mod_resource/content/1/Considera%C3%A7%C3%B5es%20b%C3%A1sicas%20sobre%20a%20Avalia%C3%A7%C3%A3o%20do%20Sistema%20Nacional.pdf).
- Lusher, Dean, Johan Koskinen, e Garry Robins, orgs. *Exponential Random Graph Models for Social Networks: Theory, methods, and applications*. Structural Analysis in the Social Sciences. USA: Cambridge University Press, 2013.
- Mariscal, Gonzalo, Óscar Marbán, e Covadonga Fernández. “A survey of data mining and knowledge discovery process models and methodologies”. *The Knowledge Engineering Review* 25, n° 2 (2010): 137-66. <https://doi.org/10.1017/S0269888910000032>.
- Nery, Guilherme, Ana Paula Bragaglia, Flávia Clemente, e Suzana Barbosa. “Nem tudo parece o que é: Entenda o que é plágio”. Instituto de Arte e Comunicação Social da UFF, 2009. <http://www.noticias.uff.br/arquivos/cartilha-sobre-plagio-academico.pdf>.
- Nooy, Wouter de, Andrej Mrvar, e Vladimir Batagelj. *Exploratory Social Network Analysis with Pajek*. Structural Analysis in the Social Sciences. USA: Routledge, 2005.
- Pátaro, Cristina Saitê de Oliveira, e Frank Antonio Mezzomo. “Sistema Nacional de Pós-Graduação no Brasil: estrutura, resultados e desafios para política de Estado - Lívio Amaral”. *Revista Educação e Linguagens* 2, n° 3 (julho de 2013): 11-17.
- Schwartzman, Simon. “A Ciência da Ciência”. *Ciência Hoje* 2, n° 11 (março de 1984): 54-59.
- Silver, Nate. *The Signal and the Noise: Why so many predictions fail - but some don't*. USA: The Penguin Press HC, 2012. \*UNB. História. UnB. Disponível em: <https://www.unb.br/a-unb/historia>. Acesso em: 31 mai. 2018
- Vicari, Donatella, Akinori Okada, Giancarlo Ragozini, e Claus Wiehs. *Analysis and Modeling of Complex Data in Behavioral and Social Sciences*. Studies in Classification, Data Analysis, and Knowledge Organization. Switzerland: Springer, 2014.
- Wickham, Hadley, e Garrett Golemund. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. USA: O'Reilly, 2016.