

# 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

*Fernanda Amorim, Gustavo Monteiro e Daniel Lima*

11/06/2018

## Introdução

Este documento representa o Relatório Final da Disciplina de Ciência de Dados para Todos (Data Science for All). A matéria tem por objetivo o ensino de técnicas de análise de dados utilizando a plataforma R. Tendo em vista o grande contingente de dados que tem sido alimentado pela sociedade, a construção de conhecimento nesta área é de suma importância para a formação acadêmica dos alunos.

No Relatório Final da Disciplina será feita uma análise descritiva de 3 programas de Pós-Graduação da UnB: Agronomia, Agronegócios e Ensino em Ciências Ambientais em Rede Nacional (Programa em Rede). Primeiramente será feita uma breve descrição dos programas, mostrando os objetivos e o históricos deles na Universidade. Posteriormente, serão feitas análises descritivas utilizando dados dos professores, das publicações, das orientações e grupos de pesquisas. Para a análise, será usado os conhecimentos adquiridos ao longo da disciplina e a metodologia CRISP-DM, que consiste em modelo estruturado de processos para análise de dados.

## Metodologia

Para desenvolvimento do trabalho devem ser seguidos, da forma mais simplificada e coerente possível, as fases e atividades genéricas do ciclo de vida de um projeto executado em aderência ao CRISP-DM, conforme ilustra de forma geral a figura 1. Em outras palavras, a produção do relatório deve seguir a metodologia CRISP-DM.

Perceba que a Figura 1 sugere haver grande flexibilidade na execução das fases, de modo que se pode retornar a fases anteriores em muitos pontos.

“A widely used methodology for data mining is the Cross-Industry Standard Process for Data Mining (CRISP-DM) which was initiated in 1996 (...) with the intent of providing a process that is **reliable and repeatable** by people with little data-mining background, with a framework within which experience can be recorded, to support the replication of projects, to support planning and management, as well as to demonstrate data mining as a mature discipline (...)” [Sullivan, Rob. Introduction to Data Mining for the Life Sciences. Springer Science & Business Media. 2012]

Para fazer as análises descritivas nas bases de Pós-Graduação, serão usados os seguintes pacotes do Software Livre R: *jsonlite*, *readxl*, *dplyr*, *ggplot2*, *tidyverse*. Os pacotes *jsonlite* e *readxl* são usados para fazer a leitura das bases. Os pacotes *dplyr* e *tidyverse* são utilizados para manipulação dos dados e o pacote *ggplot2* possui as funções que geram os gráficos para auxiliar na visualização dos dados.

## Delimitações iniciais

Em aderência à estrutura do CRISP-DM, algumas delimitações de contexto para o trabalho são apresentadas a seguir:

## Domínio de Aplicação do projeto

O domínio de aplicação do projeto é o da produção científica e acadêmica brasileira, mais especificamente a produção científica ou produção acadêmica de um subgrupo de pesquisadores vinculados à Universidade de Brasília. Neste relatório, em específico, será realizado um estudos com as bases de Pós-Graduação de Agronomia, Agronegócios e Ensino em Ciências Ambientais em Rede Nacional. Todas as base foram retiradas do portal <http://unb.elattes.com.br> em formato JSON, ao todo são 5 bases com informações acerca dos programas de Pós-Graduação.

## Tipo de Problema abordado

O tipo de problema abordado é o da produção de análises descritivas e quantitativas, que permitam caracterizar a produção científica e acadêmica de um grupo de pesquisadores. Essa caracterização visa subsidiar a tomada de decisão por membros dos três programas de Pós-Graduação estudados neste relatório, isto é, de acordo com as análises feitas é possível avaliar os programas de uma melhor forma porque os dados conseguem descrever as necessidades pesquisadores.

## Modelo CRISP-DM

Miner (2012), aprofunda: "(...) In CRISP-DM, the complete life cycle of a data mining project is represented with **six phases**: business understanding (determining the purpose of the study), data understanding (data exploration and understanding), data preparation, modeling, evaluation, and deployment.(...). [Miner, Gary. Practical Text Mining and Statistical Analysis for Non-structure Text Data Applications. Academic Press, 2012.]

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

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

A produção do conhecimento científico, no Brasil, é predominantemente efetuada por meio do Sistema Nacional de Pós-Graduação - SNPG, e mais fortemente relacionada com a formação de doutores nesse sistema (Pátaro e Mezzomo, 2013), por meio de cursos de pós-graduação *strictu sensu*.

Fernandes e Sampaio (2017) já indicaram que a ciência é reconhecidamente um elemento essencial para o desenvolvimento social e econômico de qualquer nação. Assim sendo, faz-se mister aprimorar o SNPG como forma de promoção desse crescimento, visando maximizar o retorno decorrente do emprego dos recursos nele aplicados. A promoção do crescimento do SNPG se dá predominantemente por meio de avaliações regulares de seus programas de pós-graduação, sob responsabilidade da CAPES, que realiza a cada quatro anos um complexo (Leite, 2018, p. 13) e custoso processo de coleta de dados, análise e deliberação sobre as pós-graduações *strictu sensu*, em coerência com o estabelecido no Plano Nacional de Pós-Graduação (PNPG) 2012-2020 (CAPES, 2010) e nos diversos documentos que definem os critérios de organização da pós-graduação em cada área do conhecimento (CAPES, 2018). Leite (2018) faz uma apresentação geral de como se organizam e são avaliadas as pós-graduações no Brasil.

O Plano Nacional de Pós-Graduação (PNPG), por outro lado, define diretrizes estratégicas para desenvolvimento da pós-graduação brasileira, que deve abordar prioritariamente grandes temas de interesse nacional, tais como a redução das assimetrias de desenvolvimento entre as regiões do Brasil, a formação de professores para a educação básica, a formação de recursos humanos para as empresas, a resposta aos grandes desafios brasileiros sobre Água, Energia, Transporte, Controle de Fronteiras, Agronegócio, Amazônia, Amazônia Azul (Mar), Saúde,

Defesa, Programa Espacial, além de Justiça, Segurança Pública, Criminologia e Desequilíbrio Regional. O PNPG também traça as diretrizes para financiamento da pós-graduação e sua internacionalização, apresentando conclusões e recomendações.

As avaliações do SNPG, ao atribuírem mensurações de desempenho às diversas pós-graduações que dele fazem parte, geram incentivos e penalidades aos programas, tendo em vista a limitada disponibilidade de recursos para investimento em bolsas, taxas de bancada etc. Embora o sistema seja altamente sofisticado ele é também altamente criticado (Azevedo et al., 2016), sobretudo porque há percalços na busca por um equilíbrio entre as diferentes concepções de finalidade da ciência. Se de um lado a promoção do conhecimento gerado predominantemente nas ditas ciências *hard* contribui para criar fluxos econômicos mais intensos, isso não significa que essa promoção possa ocorrer em detrimento da menor promoção na geração de conhecimento sobre problemas sociais, predominantemente gerado nas ditas ciências *soft*, especialmente das áreas de humanidades, sob pena de ampliação de desigualdades (Azevedo et al., 2016).

Não há solução simples, mas postula-se, nesta disciplina, que uma maior agilidade na avaliação e a utilização de critérios mais objetivos, poderá facilitar a melhoria do sistema.

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

A partir de 2018, as diversas áreas da pós-graduação brasileira foram organizadas na forma de colégios, grandes áreas e áreas, conforme apresentam as tabelas a seguir.

### 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

### 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

## Colégio de Humanidades

CIÊNCIAS HUMANAS	CIÊNCIAS SOCIAIS APLICADAS	LINGÜÍSTICA, LETRAS E ARTES
Antropol/Arqueol	Admin.Púb./Empr.,C.Contáb. e Tur.	Artes
Ciência Pol. e Rel. Int.	Arquit., Urban. e Design	Linguística e Literatura
Ciências da Religião e Teol.	Comunicação e Informação	-
Educação	Direito	-
Filosofia	Economia	-
Geografia	Planej. Urbano e Reg. / Demografia	-
História	Serviço Social	-
Psicologia	-	-
Sociologia	-	-

Cada um desses colégios, grandes áreas e áreas de conhecimento possuem dinâmicas próprias, e, portanto, não há um modelo universal que se aplique a todas. Existem aspectos comuns, mas também grandes peculiaridades, descritas parcialmente nos correspondentes documentos de área disponíveis em CAPES (2018).

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

### O que é a UnB?

A Universidade de Brasília foi criada em 1962, idealizada pelos Educadores Darcy Ribeiro e Anísio Texeira com missão de ser uma Universidade transformadora que produz, integra e divulga o conhecimento. Auxiliando na formação de cidadãos comprometidos com a ética, a responsabilidade social e o desenvolvimento sustentável.

### Descrição das pós-graduações da UnB

DESCREVER OS PROGRAMAS AQUI!

O programa de Pós-Graduação em Agronomia é alocado na Faculdade de Agronomia e Medicina Veterinária e possui foco principal na Área de concentração na Produção Sustentável. As três linhas de pesquisa principais presentes no programa são: 1. Solo, água e qualidade ambiental: o objetivo desta linha de pesquisa diagnosticar e monitorar os recursos de solo e água da região dos cerrados, a partir das bacias hidrográficas, avaliar o uso e ocupação atual das terras, com ênfase nos principais usos agrícolas da região dos cerrados, estudar as alterações que ocorrem nos solos, através de seus diferentes usos e manejos, no sentido de se buscar índices de qualidade do solo, por meio de análises físicas, químicas e biológicas; estudar os diferentes métodos de irrigação, suas eficiências e o manejo da água de irrigação, cultivo hidropônico em hortaliças, além das diversas interações entre a água e os nutrientes, que afetam a produtividade dos cultivos; estudar adubações adequadas de macro e micronutrientes e suas combinações e estudar as variáveis climáticas e seus efeitos sobre a produção vegetação vegetal.

2. Recursos genéticos e melhoramento vegetal: Esta linha de pesquisa tem como objetivo efetuar a caracterização de recursos genéticos vegetais com base em características morfológicas, moleculares e agronômicas como subsídio para sua utilização per si e em programas

de melhoramento genético. Diferentes espécies de fruteiras (maracujá, mamão, etc) e hortaliças (tomate, cenoura, etc) e culturas de grãos (soja, trigo, etc) de importância atual e potencial para o agronegócio brasileiro são contempladas, incluindo aquelas da biodiversidade brasileira, como as plantas nativas do Cerrado. A realização de avaliações genômicas para o descobrimento de genes úteis ao melhoramento através da compreensão do controle genético de características de importância econômica como resistência a patógenos, tolerância à seca e ao alumínio e aumento da qualidade nutricional também estão contempladas.

3. Sistemas de produção agrícola sustentáveis: Esta linha de pesquisa visa desenvolver conhecimentos e tecnologias para as principais culturas da região, principalmente dentro das áreas de cultivos anuais (soja, trigo, et), horticultura (cenoura, tomate, etc) e fruticultura (maracujá, mamão, etc), incluindo o manejo integrado de pragas, controle biológico, solarização e manejo do ambiente, especialmente água de irrigação e adubação, visando o fortalecimento da planta e suas estruturas de defesa. Aspectos qualitativos e quantitativos da produção e dos recursos naturais serão avaliados em diferentes sistemas de cultivo, enfatizando os sistemas conservacionistas: plantio direto, orgânico, integração lavoura-pecuária e agroflorestas.. As culturas estudadas e as práticas conservacionistas estudadas a obtenção de produtos de qualidade com sustentabilidade econômica e respeito aos recursos naturais, enfatizando os sistemas conservacionistas.

O Programa é coordenado pelo Prof.Dr. Cícero Célio de Figueiredo e é composto pelos seguintes docentes:

Alessandra Monteiro de Paula Ana Maria Resende Junqueira Carlos Roberto Spehar Cícero Célio de Figueiredo Cícero Lopes da Silva Cristina Schetino Bastos Delvio Sandri Ernandes Rodrigues de Alencar Fabiana Carmanini Ribeiro Fabio Gelape Faleiro (EMBRAPA) Jader Galba Busato Jean Kleber de Abreu Mattos João José da Silva Junior José Ricardo Peixoto Luiz Eduardo Bassay Blum Marcelo Fagioli Maria Lucrecia Gerosa Ramos Marilusa Pinto Coelho Lacerda Michelle Souza Vilela Nara Oliveira Silva Souza Nilton Tadeu Vilela Junqueira (EMBRAPA) Osvaldo Kiyoshi Yamanishi Ricardo Carmona Rose Gomes Monnerat Sólón de Pontes (EMBRAPA) Tairone Paiva Leão Tiago Pereira da Silva Correia Warley Marcos Nascimento (EMBRAPA)

O programa conta com 10 laboratórios especializados em áreas , que são:

Agrometeorologia Biologia do Solo Bromatologia Estação Biológica Fazenda Água Limpa Física do Solo Fitopatologia Fruticultura Geoprocessamento Hidrologia

Já o Programa de pós-graduação em Rede Nacional para Ensino das Ciências Ambientais (PROFCIAMB) é um Programa incentivado pela CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) que teve início em 2016. O programa tem parceria com a Agência Nacional de águas, que aportou recursos para a estruturação da Rede e possui o objetivo de formar pelo menos 150 mestres na temática do ensino em água e ciências ambientais no período de 2016 - 2020.

A rede é composta por diversos polos por todo o Brasil. Os centros estão distribuídos da seguinte forma:

- *Norte:* UFAM, UFPA
- *Nordeste:* UFPE, UFS, UEFS
- *Centro-Oeste:* UNB
- *Sudeste:* USP
- *Sul:* UFPR, UEM

Com o programa espera-se qualificar uma grande quantidade de professores do ensino básico e também profissionais de áreas relacionadas à divulgação da ciência, como museus, por exemplo.

Na UnB, o programa é coordenado pelo professor Mauricio Amazonas, o qual possui atuação nas áreas de Economia do Meio Ambiente, Economia dos Recursos Naturais, Economia Agrária, dentre outras. Isso ressalta o caráter multidisciplinar do programa.

Os docentes do programa são:

Cristiane Gomes Barreto Izabel Cristina Bruno Bacellar Zaneti Mauricio de Carvalho Amazonas Pedro Henrique Zuchi da Conceição Thomas Ludewigs

A quantidade de docentes indica o quanto o programa é novo.

Neste Relatório, serão feitas análises de dados dos professores, de publicações e de orientações para que se possa avaliar o andamento dos programas de pós-graduação e gerar novas perspectivas e soluções para estes.

Enquanto o programa de pós-graduação em Agronegócios, possui apenas uma área de concentração e integradora de duas linhas de pesquisa com objetivos focalizados. Inicialmente possui um aspecto multidisciplinar, agregando docentes de diversas formações em torno da mesmo tema, de tal forma que preserve o campo de conhecimento de cada um. As linhas de pesquisa do programa são:

1. Agricultura Familiar e Agronegócios: Desenvolvimento de estudos relacionados ao universo da agricultura familiar, buscando compreender seu processo histórico de transformação, seus desafios atuais, tendências e perspectivas futuras. Investigar os impactos das políticas públicas direcionadas ao setor com prioridade para políticas de reforma agrária, extensão rural, crédito, educação do campo e de geração de tecnologia. Priorizar nos estudos a construção de conhecimentos que contribuam para a promoção do desenvolvimento sustentável, com ênfase em aspectos como a viabilidade econômica da agricultura familiar, a reversão do processo de exclusão social e o uso sustentável dos recursos naturais. Também são objetivos desta linha de pesquisa conhecer os diferentes atores, marcados por especificidades de ciclo de vida, classe, étnico-raciais, geração e gênero.
2. Competitividade e sustentabilidade no Agronegócio: Desenvolvimento de estudos e pesquisas voltados para a dinâmica competitiva do agronegócio nos âmbitos regional, nacional e internacional, levando em consideração as principais abordagens teóricas relacionadas ao estudo dos sistemas agroindustriais de forma a apoiar a formulação de políticas públicas e a tomada de decisão no âmbito das organizações do agronegócio.

O programa é coordenado pelo Prof.Dr. Karim Marini Thomé e é composto pelos seguintes docentes:

Ana Maria Resende Junqueira Antônio Maria Gomes de Castro Armando Fornazier Carlos Rosano Peña Flávio Borges Botelho Filho João Paulo Guimarães Soares José Eustáquio Ribeiro Vieira Filho José Márcio Carvalho Karim Marini Thomé Maria Júlia Pantoja Marlon Vinícius Brisola Mauro Eduardo Del Grossi Patricia Guarnieri Suzana Maria Valle Lima Vânia Ferreira Roque-Specht

## 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.

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 sítio da CAPES (2018).

## Avaliação das Circunstâncias

### Avaliação preliminar das pós-graduações na UnB

O Programa de Pós-Graduação em Agronomia possui 22 membros do corpo docente 403 orientações acadêmicas concluídas entre os anos de 2010 e 2017.

Já o PROFCIAMB, um programa bastante novo, tem 5 membros, com 44 orientações acadêmicas concluídas no mesmo período anterior. É importante ressaltar que apesar do programa existir desde 2016, os dados também incluem informações dos docentes antes desse ano.

## Avaliação preliminar da produção científica e acadêmica da UnB

Entre os anos de 2010 e 2017, o programa de agronomia teve 466 artigos, publicados ou aceitos, e 127 livros e capítulos de livros publicados pelos membros do corpo docente e seus orientandos. O PROFCIAMB teve 36 artigos publicados ou aceitos e 21 livros e capítulos de livros publicados.

## CRISP-DM Fase 2 - Entendimento dos Dados

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

Todos os arquivos com dados apresentados foram fornecidos pela plataforma <http://unb.elattes.com.br>, disponibilizada para os alunos pelos professores responsáveis. Os arquivos estão no formato JSON, e seus atributos iniciais e conteúdos são apresentados a seguir.

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

```
#setwd("C:\\Users\\amori\\Documents\\UnB\\DS4A\\233.files")
#json.perfil <- "233.profile.json"
#file.info(json.perfil)

json.perfil.ambientais <- "dados/ambientais.profile.json"
json.perfil.agronomia <- "dados/233.profile.json"
json.perfil.agronegocio <- "dados/232.profile.json"
df <- rbind(file.info(json.perfil.ambientais),
            file.info(json.perfil.agronomia),
            file.info(json.perfil.agronegocio))
rownames(df) <- gsub("dados/", "", rownames(df))
df
```

	size <dbl>	isdir <lgl>	mode <S3: oclmode>	mtime <S3: POSIXct>	ctime <S3: POSIXct>
ambientais.profile.json	112838	FALSE	664	2018-07-08 15:24:32	2018-07-08 15:24:32
233.profile.json	1646833	FALSE	664	2018-07-08 15:24:32	2018-07-08 15:24:32
232.profile.json	720238	FALSE	664	2018-06-21 01:01:30	2018-07-08 15:24:32
3 rows   1-6 of 11 columns					

Os arquivos ambientais.profile.json, 233.profile.json, 232.profile.json apresentam dados sobre o perfil de todos os docentes vinculados a programas de pós-graduação 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

```
json.advise.ambientais <- "dados/ambientais.advise.json"
json.advise.agronomia <- "dados/233.advise.json"
json.advise.agronegocio <- "dados/232.advise.json"
df.advise <- rbind(file.info(json.advise.ambientais),
  file.info(json.advise.agronomia),
  file.info(json.advise.agronegocio))
rownames(df.advise) <- gsub("dados/", "", rownames(df.advise))
df.advise
```

	size <dbl>	isdir <lgl>	mode <S3: octmode>	mtime <S3: POSIXct>	ctime <S3: POSIXct>
ambientais.advise.json	54901	FALSE	664	2018-07-08 15:24:32	2018-07-08 15:24:32
233.advise.json	611536	FALSE	664	2018-07-08 15:24:32	2018-07-08 15:24:32
232.advise.json	248400	FALSE	664	2018-06-21 01:01:30	2018-07-08 15:27:11

3 rows | 1-6 of 11 columns

Os arquivos ambientais.advise.json, 233.advise.json, 232.advise.json apresentam dados sobre o orientações de mestrado e doutorado feitas por todos os docentes vinculados a programas de pós-graduação da UnB, entre 2010 e 2017. Esse arquivo foi fornecido pelos docentes responsáveis pela disciplina.

## Produção bibliográfica gerada pelos docentes vinculados às pós-graduações

```
json.producao.bibliografica.ambientais <- "dados/ambientais.publication.json"
json.producao.bibliografica.agronomia <- "dados/233.publication.json"
json.producao.bibliografica.agronegocio <- "dados/232.publication.json"
df.bib <- rbind(file.info(json.producao.bibliografica.ambientais),
  file.info(json.producao.bibliografica.agronomia),
  file.info(json.producao.bibliografica.agronegocio))
rownames(df.bib) <- gsub("dados/", "", x = rownames(df.bib))
df.bib
```

	size <dbl>	isdir <lgl>	mode <S3: octmode>	mtime <S3: POSIXct>	ctime <S3: POSIXct>
ambientais.publication.json	49668	FALSE	664	2018-07-08 15:24:32	2018-07-08 15:24:32
233.publication.json	824605	FALSE	664	2018-07-08 15:24:32	2018-07-08 15:24:32
232.publication.json	394437	FALSE	664	2018-06-21 01:01:30	2018-07-08 15:27:11

3 rows | 1-6 of 11 columns



Os arquivos ambientais.publication.json, 233.publication.json, 232.publication.json apresentam dados sobre a produção bibliográfica gerada por todos os docentes vinculados a programas de pós-graduação da UnB, entre 2010 e 2017.

## Redes de colaboração entre docentes

```
df.graph <- rbind(file.info('dados/ambientais.graph.json'),
file.info('dados/ambientais.graph.json'),
file.info('dados/ambientais.graph.json'))
rownames(df.graph) <- gsub('dados/', '', x = rownames(df.graph))
df.graph
```

	size <dbl>	isdir <lgl>	mode <S3: octmode>	mtime <S3: POSIXct>	ctime <S3: POSIXct>
ambientais.graph.json	577	FALSE	664	2018-07-08 15:24:32	2018-07-08 15:24:32
ambientais.graph.json1	577	FALSE	664	2018-07-08 15:24:32	2018-07-08 15:24:32
ambientais.graph.json2	577	FALSE	664	2018-07-08 15:24:32	2018-07-08 15:24:32

3 rows | 1-6 of 11 columns

Os arquivos ambientais.graph.json, ambientais.graph.json1, ambientais.graph.json2 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

Para ler e manipular inicialmente esses dados, serão usadas primordialmente as bibliotecas seguintes

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

```
## — Attaching packages — tidyverse 1.2.1 —
```

```
## ✓ tibble 1.4.2   ✓ dplyr 0.7.4
## ✓ tidyr 0.8.0   ✓ stringr 1.3.0
## ✓ purrr 0.2.4   ✓ forcats 0.2.0
```

```
## — Conflicts — tidyverse_conflicts() —
## * dplyr::filter() masks stats::filter()
## * purrr::flatten() masks jsonlite::flatten()
## * dplyr::lag() masks stats::lag()
```

## Descrição dos dados do perfil

O arquivo 233.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.

```
agro.prof <- fromJSON("dados/233.profile.json")
ambientais.prof <- fromJSON("dados/ambientais.profile.json")
agronegocio.prof <- fromJSON("dados/232.profile.json")
```

A quantidade de docentes nos programas sob análise é apresentada a seguir.

```
c("Agronomia" = length(agro.prof),
  "Ambientais" = length(ambientais.prof),
  "Agronegócio" = length(agronegocio.prof))
```

```
##   Agronomia   Ambientais   Agronegócio
##         22         5         16
```

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(agro.prof[[1]]$areas_de_atuacao$area)
```

```
##   chr [1:6] "Engenharia Agrícola" "Engenharia Agrícola" ...
```

```
glimpse(agro.prof[[1]]$producao_bibliografica$PERIODICO$natureza)
```

```
##   chr [1:15] "COMPLETO" "COMPLETO" "COMPLETO" "COMPLETO" "COMPLETO" ...
```

```
glimpse(agro.prof[[1]]$orientacoes_academicas$OUTRAS_ORIENTACOES_CONCLUIDAS)
```

```
## Observations: 50
## Variables: 13
## $ natureza      <chr> "INICIACAO_CIENTIFICA", "INICIACAO...
## $ titulo        <chr> "Qualidade Físico-Química e Microb...
## $ ano           <chr> "2010", "2010", "2010", "2011", "2...
## $ id_lattes_aluno <chr> "", "", "", "", "", "", "", "", "", "...
## $ nome_aluno    <chr> "Andressa Harumi Koyama", "Angelin...
## $ instituicao    <chr> "Universidade de Brasília", "Unive...
## $ curso         <chr> "Agronomia", "Agronomia", "Agronom...
## $ codigo_do_curso <chr> "900000010", "900000010", "900000010"...
## $ bolsa        <chr> "SIM", "NAO", "SIM", "SIM", "NAO",...
## $ agencia_financiadora <chr> "Conselho Nacional de Desenvolve...
```

```
## $ codigo_agencia_financiadora <chr> "002200000000", "", "002200000000"...  
## $ nome_orientadores      <list> ["Ernandes Rodrigues de Alencar",...  
## $ id_lattes_orientadores  <list> ["0646231743976574", "06462317439...
```

```
glimpse(agro.prof[[1]]$senioridade)
```

```
## chr "9"
```

```
glimpse(ambientais.prof[[1]]$areas_de_atuacao$area)
```

```
## chr [1:6] "Economia" "Economia" "Economia" "Economia" "Economia" ...
```

```
glimpse(ambientais.prof[[1]]$producao_bibliografica$PERIODICO$natureza)
```

```
## chr [1:5] "COMPLETO" "COMPLETO" "COMPLETO" "COMPLETO" "COMPLETO"
```

```
glimpse(ambientais.prof[[1]]$orientacoes_academicas$OUTRAS_ORIENTACOES_CONCLUIDAS)
```

```
## Observations: 5  
## Variables: 13  
## $ natureza      <chr> "MONOGRAFIA_DE_CONCLUSAO_DE_CURSO_...  
## $ titulo        <chr> "Serviços Ambientais: evolução rec...  
## $ ano           <chr> "2010", "2011", "2011", "2011", "2...  
## $ id_lattes_aluno <chr> "", "", "", "", ""  
## $ nome_aluno     <chr> "Fabiano Faria de Carvalho Hecht",...  
## $ instituicao     <chr> "Universidade de Brasília", "Unive...  
## $ curso          <chr> "Desenvolvimento Sustentável e Dir...  
## $ codigo_do_curso <chr> "90000017", "90000016", "90000016"...  
## $ bolsa          <chr> "NAO", "NAO", "NAO", "NAO", "NAO"  
## $ agencia_financiadora <chr> "", "", "", "", ""  
## $ codigo_agencia_financiadora <chr> "", "", "", "", ""  
## $ nome_orientadores <list> ["Mauricio de Carvalho Amazonas",...  
## $ id_lattes_orientadores <list> ["0974214976685536", "09742149766...
```

```
glimpse(ambientais.prof[[1]]$senioridade)
```

```
## chr "9"
```

```
glimpse(agronegocio.prof[[1]]$areas_de_atuacao$area)
```

```
## chr [1:4] "Psicologia" "Administração" "Psicologia" "Administração"
```

```
glimpse(agronegocio.prof[[1]]$producao_bibliografica$PERIODICO$natureza)
```

```
## chr [1:6] "COMPLETO" "COMPLETO" "COMPLETO" "COMPLETO" "COMPLETO" ...
```

```
glimpse(agronegocio.prof[[1]]$orientacoes_academicas$OUTRAS_ORIENTACOES_CONCLUIDAS)
```

```
## Observations: 7
## Variables: 13
## $ natureza      <chr> "MONOGRAFIA_DE_CONCLUSAO_DE_CURSO_...
## $ titulo        <chr> "Diagnóstico de Necessidades de Ca...
## $ ano           <chr> "2012", "2012", "2012", "2013", "2...
## $ id_lattes_aluno <chr> "", "", "", "", "", "", ""
## $ nome_aluno     <chr> "Antônio Inácio Soares de Sousa Ne...
## $ instituicao     <chr> "Escola Nacional de Administração ...
## $ curso          <chr> "Curso de Especialização em Gestão...
## $ codigo_do_curso <chr> "90000017", "90000017", "90000017"...
## $ bolsa         <chr> "NAO", "NAO", "NAO", "NAO", "NAO",...
## $ agencia_financiadora <chr> "", "", "", "", "", "", ""
## $ codigo_agencia_financiadora <chr> "", "", "", "", "", "", ""
## $ nome_orientadores <list> ["Maria Júlia Pantoja", "Maria Jú...
## $ id_lattes_orientadores <list> ["0011938955607677", "00119389556...
```

```
glimpse(agronegocio.prof[[1]]$senioridade)
```

```
## chr "8"
```

Uma breve inspeção visual dos atributos anteriormente apresentados permite inferir que o pesquisador de agronomia (1o caso) sob análise:

- Atua predominantemente na área de engenharia agrícola e tecnologia de alimentos.
- Trabalha no Faculdade de Agronomia e Medicina Veterinária.
- Possui quinze artigos recentes publicados entre os anos de 2010 e 2017.
- Faz orientações de mestrado, doutorado, as duas em andamento, além de ter orientações de mestrado, doutorado e iniciação científica concluídas
- Foi classificado com senioridade 9.

Já o pesquisador do PROFCIAMB tem as seguintes características:

- Atua predominantemente na área de economia (potencialmente avaliação econômica de projetos ambientais).
- Tem senioridade 9.
- Publicou 5 artigos entre os anos de 2010 e 2017.
- Realizou 5 orientações de monografia e trabalho de conclusão de curso de graduação

Enquanto a pesquisadora do agronegócio(3º caso):

- Atua predominantemente na área da psicologia;
- Possui senioridade 8;
- Realizou 7 orientações de monografias, 5 orientações de mestrado finalizadas e 6 em andamento.

## Potencial de utilização dos dados do perfil dos docentes

Esses dados terão potencial para responder às questões de *data mining*? O que é possível gerar a partir desses dados? A fim de compreender a relevância dos dados para a avaliação da produção acadêmica nas pós-graduações brasileiras pode-se recorrer a trabalhos como os seguintes:

- Leite (2018) apresenta, em suas “Considerações básicas sobre a Avaliação do Sistema Nacional de Pós-Graduação”, o conjunto dos itens que são tópicos de avaliação das pós-graduações pela CAPES, e que envolvem, entre outros:
  - Avaliação do corpo docente, com 20% a 30% de peso na avaliação total do programa, a depender do seu tipo. Analisando-se de forma mais detalhada os critérios de avaliação do corpo docente, indicados por Leite, o que é possível gerar com base nos dados disponíveis em unb.prof? Há dados que permitam identificar o perfil do docente, como proposto pela CAPES, inclusive no documento de área específica na qual atua o pesquisador? Que outros aspectos relevantes para a CAPES podem ser levantados com base nos dados dessa fonte?
  - Avaliação do corpo discente, Teses e dissertações, com 30% a 20% de peso na avaliação total do programa, a depender de seu tipo. Os dados sobre orientação permitem fazer quais tipos de avaliações do corpo discente?
  - Avaliação da produção intelectual, com 40% de peso na avaliação total. Qual a relevância dos dados em unb.prof para essa avaliação? Que outros arquivos podem melhor subsidiar essa avaliação?
- Em busca de considerar outros fatores relevantes para a avaliação da pós-graduação, não considerados no modelo da CAPES, pode-se recorrer ao trabalho de Kalpazidou Schmidt e Graversen (2018), que apresentam um conjunto de fatores persistentes que facilitam a existência de ambientes de pesquisa inovadores e dinâmicos, dentre os quais se destaca:
  - Atividade em pesquisas com elevado grau de impacto social;
  - Promoção de elevado grau de autonomia individual, tanto do ponto de vista teórico quanto metodológico;
  - Possuem um bom clima de trabalho, baseado no trabalho em times;
  - São internacionalmente bem conhecidas etc.

Estariam esses fatores contemplados, de alguma forma, mesmo que parcialmente, nos dados presentes em unb.prof? Ou em qualquer outros dos arquivos? Cabe explorar.

## Descrição dos dados de orientações

```
agro.adv <- fromJSON("dados/233.advise.json")
names(agro.adv)
```

```
## [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(agro.adv$ORIENTACAO_CONCLUIDA_DOUTORADO)
```

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

```
length(agro.adv$ORIENTACAO_CONCLUIDA_DOUTORADO$`2016`$natureza)
```

```
## [1] 18
```

```
head(sort(table(agro.adv$ORIENTACAO_CONCLUIDA_DOUTORADO$`2017`$curso), decreasing = TRUE), 10)
```

```
##
##              Agronomia      Pós-graduação em Agronomia
##              3              2
## Mestrado e Doutorado em Agronomia
##              1
```

```
head(sort(table(agro.adv$ORIENTACAO_CONCLUIDA_MESTRADO$`2017`$curso), decreasing = TRUE), 10)
```

```
##
##              Agronomia      Agronegócios
##              14              2
## Pós graduação em Agronomia Mestrado e Doutorado em Agronomia
##              2              1
## Produção Vegetal
##              1
```

```
ambientais.adv <- fromJSON("dados/ambientais.advise.json")
names(ambientais.adv)
```

```
## [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(ambientais.adv$ORIENTACAO_CONCLUIDA_DOUTORADO)
```

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

```
length(ambientais.adv$ORIENTACAO_CONCLUIDA_DOUTORADO$`2016`$natureza)
```

```
## [1] 0
```

```
head(sort(table(ambientais.adv$ORIENTACAO_CONCLUIDA_DOUTORADO$`2017`$curso), decreasing = TRUE), 10)
```

```
## Desenvolvimento Sustentável  
## 1
```

```
head(sort(table(ambientais.adv$ORIENTACAO_CONCLUIDA_MESTRADO$`2017`$curso), decreasing = TRUE), 10)
```

```
##  
## Centro de Desenvolvimento Sustentável  
## 2  
## Mestrado em Desenvolvimento Sustentável  
## 1
```

```
agronegocio.adv <- fromJSON("dados/232.advise.json")  
names(agronegocio.adv)
```

```
## [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(agronegocio.adv$ORIENTACAO_CONCLUIDA_DOUTORADO)
```

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

```
length(agronegocio.adv$ORIENTACAO_CONCLUIDA_DOUTORADO$`2016`$natureza)
```

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

```
head(sort(table(agronegocio.adv$ORIENTACAO_CONCLUIDA_DOUTORADO$`2017`$curso), decreasing = TRUE), 10)
```

```
## Agronomia  
## 1
```

```
head(sort(table(agronegocio.adv$ORIENTACAO_CONCLUIDA_MESTRADO$`2017`$curso), decreasing = TRUE), 10)
```

```
##
##               Agronegócios
##               6
## Programa de Pós-Graduação em Agronegócios
##               4
##               Economia
##               3
##           Mestrado em Agronegócio
##               1
##       Pós Graduação em Gestão Pública
##               1
```

## Descrição dos dados de produção bibliográfica

```
agro.pub <- fromJSON("dados/233.publication.json")
names(agro.pub)
```

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

```
names(agro.pub$PERIODICO$`2012`)
```

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

```
head(sort(table(agro.pub$PERIODICO$`2016`$periodico), decreasing = TRUE), 10)
```

```
##
## Pesquisa Agropecuária Brasileira (Online)
##               7
##               Agrotrópica (Itabuna)
##               2
## Australian Journal of Crop Science
##               2
## Bioscience Journal (UFU. Impresso)
##               2
## Horticultura Brasileira (Impresso)
##               2
```



```
##          MAGISTRA CRUZ DAS ALMAS-BA
##                                     2
## Revista Brasileira de Ciência do Solo (Online)
##                                     2
##          Acta Agronomica
##                                     1
##          Acta Horticulturae
##                                     1
## Agriculture, Ecosystems & Environment (Print)
##                                     1
```

```
head(sort(table(agro.pub$LIVRO$`2017`$nome_da_editora), decreasing = TRUE), 10)
```

```
##
##          UnB - CVT AAO          Embrapa Novas Edições Acadêmicas
##                                     1                                     1
```

```
ambientais.pub <- fromJSON("dados/ambientais.publication.json")
names(ambientais.pub)
```

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

```
names(ambientais.pub$PERIODICO$`2012`)
```

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

```
head(sort(table(ambientais.pub$PERIODICO$`2016`$periodico), decreasing = TRUE), 10)
```

```
##
## Fronteiras: Journal of Social, Technological and Environmental Science
##                                     3
##          EXPEDIÇÕES: TEORIA DA HISTÓRIA E HISTORIOGRAFIA
##                                     1
##          SUSTENTABILIDADE EM DEBATE
##                                     1
```

```
head(sort(table(ambientais.pub$LIVRO$`2017`$nome_da_editora), decreasing = TRUE), 10)
```

```
## integer(0)
```

```
agronegocio.pub <- fromJSON("dados/232.publication.json")  
names(agronegocio.pub)
```

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

```
names(agronegocio.pub$PERIODICO$`2012`)
```

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

```
head(sort(table(agronegocio.pub$PERIODICO$`2016`$periodico), decreasing = TRUE), 10)
```

```
##  
##          Revista Gestão Industrial  
##                                4  
##      Texto para Discussão (IPEA. Brasília)  
##                                3  
##      Revista Brasileira de Agroecologia  
##                                2  
##      Revista de economia e sociologia rural  
##                                2  
##      Revista em Agronegócios e Meio Ambiente  
##                                2  
## Revista em Gestão, Inovação e Sustentabilidade  
##                                2  
##      Brazilian Journal of Biology (Online)  
##                                1  
##      British Food Journal (1966)  
##                                1  
##                                Cepal Review  
##                                1  
##      Economia e Sociedade (UNICAMP. Impresso)  
##                                1  
##
```

```
head(sort(table(agronegocio.pub$LIVRO$`2017`$nome_da_editora), decreasing = TRUE), 10)
```

```
##
## Art Letras Editora      UnB - CVT AA0      IPEA
##                          2                2      1
```

## 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.

### Arquivo Profile

```
# jsonedit(agro.prof)
# Número de áreas de atuação cumulativo
c("Agronomia" = sum(sapply(agro.prof, function(x) nrow(x$sareas_de_atuacao))),
  "Ambientais" = sum(sapply(ambientais.prof, function(x) nrow(x$sareas_de_atuacao))),
  "Agronegócio" = sum(sapply(agronegocio.prof, function(x) nrow(x$sareas_de_atuacao)))) %>%
  knitr::kable(caption = "Número de áreas de atuação por programa")
```

Número de áreas de atuação por programa

	x
Agronomia	100
Ambientais	24
Agronegócio	74

```
# Número de áreas de atuação por pessoa
table(unlist(sapply(agro.prof, function(x) nrow(x$sareas_de_atuacao)))) %>%
  knitr::kable(caption = "Número de áreas de atuação por pessoa para agronomia")
```

Número de áreas de atuação por pessoa para agronomia

Var1	Freq
2	1
3	5
4	5
5	3
6	8

```
table(unlist(sapply(ambientais.prof, function(x) nrow(x$areas_de_atuacao)))) %>%
  knitr::kable(caption = "Número de áreas de atuação por pessoa para ambientais")
```

Número de áreas de atuação por pessoa para ambientais

Var1	Freq
1	1
5	1
6	3

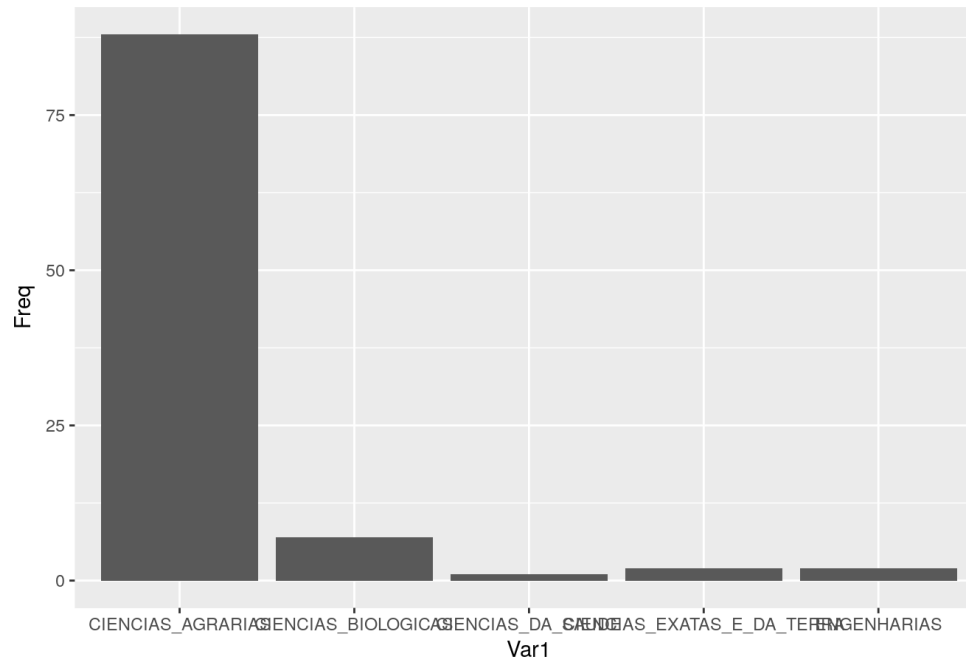
```
table(unlist(sapply(agronegocio.prof, function(x) nrow(x$areas_de_atuacao)))) %>%
  knitr::kable(caption = "Número de áreas de atuação por pessoa para agronegócio")
```

Número de áreas de atuação por pessoa para agronegócio

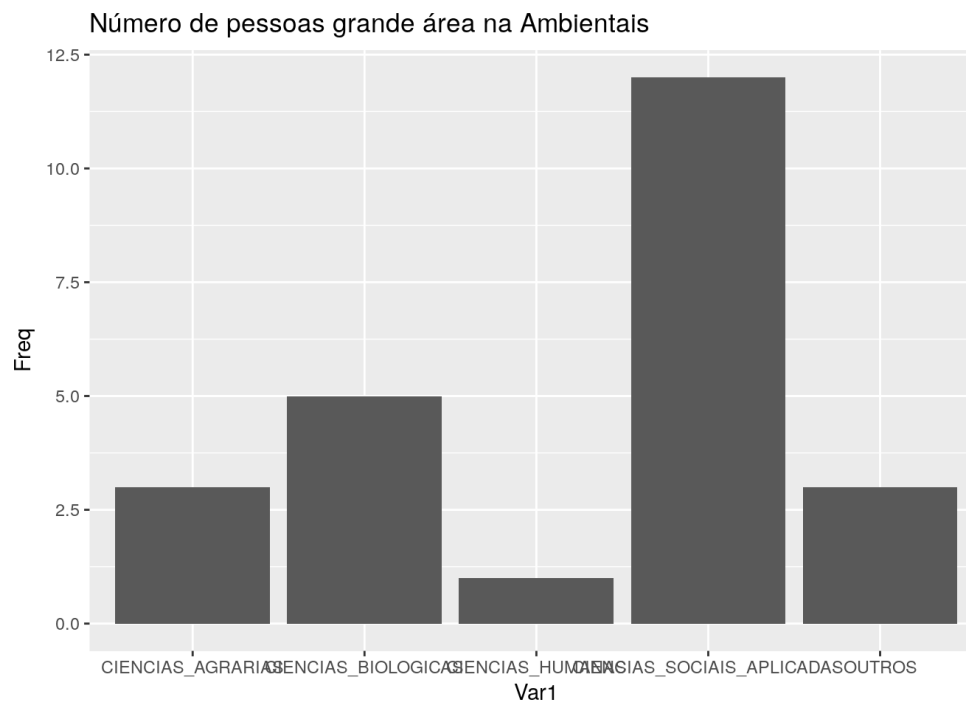
Var1	Freq
1	1
2	1
3	2
4	2
5	3
6	7

```
# Número de pessoas por grande area
grande.area.agro <- table(unlist(sapply(agro.prof, function(x) (x$areas_de_atuacao$grande_area))))
grande.area.ambientais <- table(unlist(sapply(ambientais.prof, function(x) (x$areas_de_atuacao$grande_area))))
grande.area.agronegocio <- table(unlist(sapply(agronegocio.prof, function(x) (x$areas_de_atuacao$grande_area))))
grande.area.agro <- data.frame(grande.area.agro)
grande.area.ambientais <- data.frame(grande.area.ambientais)
grande.area.agronegocio <- data.frame(grande.area.agronegocio)
ggplot(grande.area.agro, aes(x=Var1, y=Freq)) + geom_col()+
  ggtitle("Número de pessoas grande área na Agronomia")
```

Número de pessoas grande área na Agronomia

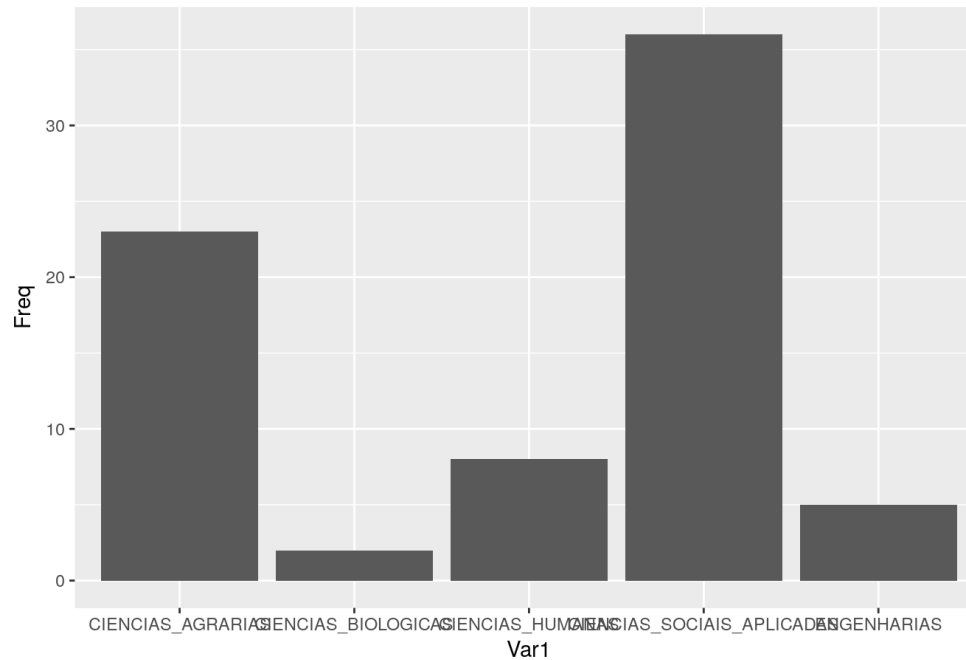


```
ggplot(grande.area.ambientais, aes(x=Var1, y=Freq)) + geom_col() +  
ggtitle("Número de pessoas grande área na Ambientais")
```



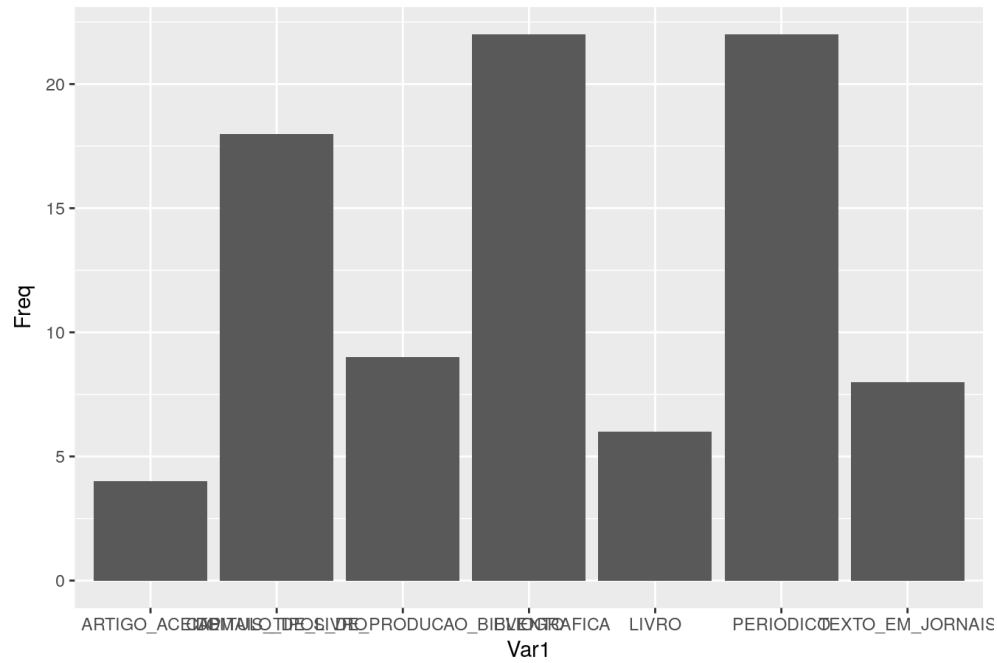
```
ggplot(grande.area.agronegocio, aes(x=Var1, y=Freq)) + geom_col() +  
ggtitle("Número de pessoas grande área na Agronegócio")
```

Número de pessoas grande área na Agronegócio



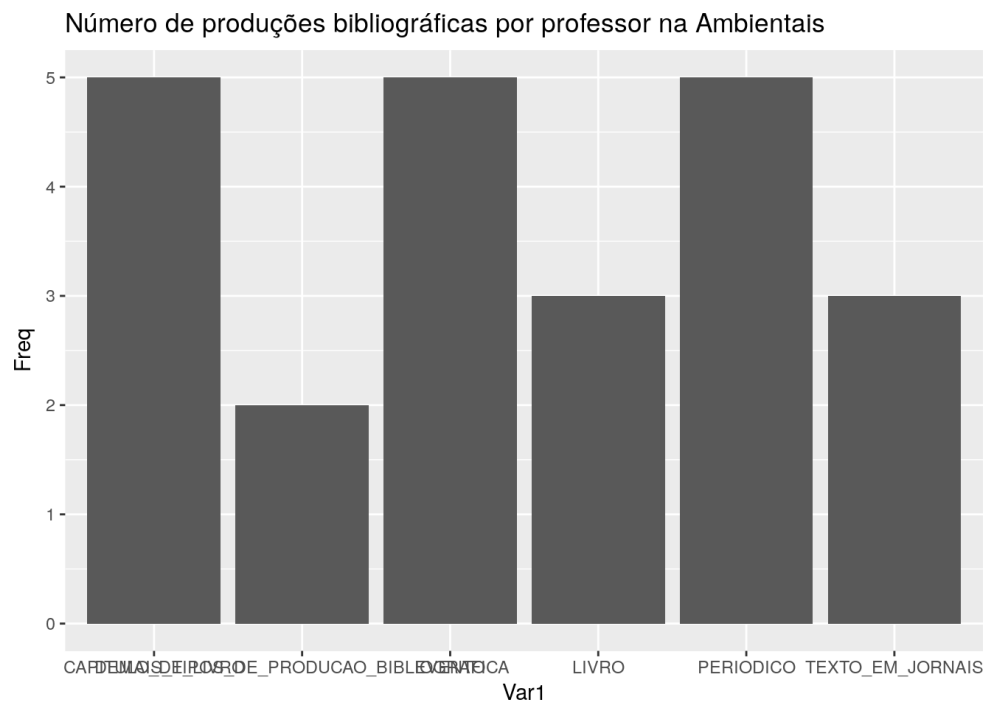
```
# Número de pessoas que produziram os específicos tipos de produção
agro.prod <- table(unlist(sapply(agro.prof, function(x) names(x$producao_bibliografica))))
ambientais.prod <- table(unlist(sapply(ambientais.prof, function(x) names(x$producao_bibliografica))))
agronegocio.prod <- table(unlist(sapply(agronegocio.prof, function(x) names(x$producao_bibliografica))))
agro.prod <- data.frame(agro.prod)
ambientais.prod <- data.frame(ambientais.prod)
agronegocio.prod <- data.frame(agronegocio.prod)
ggplot(agro.prod, aes(x= Var1, y=Freq)) + geom_col()+
  ggtitle("Número de produções bibliográficas por professor na Agronomia")
```

Número de produções bibliográficas por professor na Agronomia



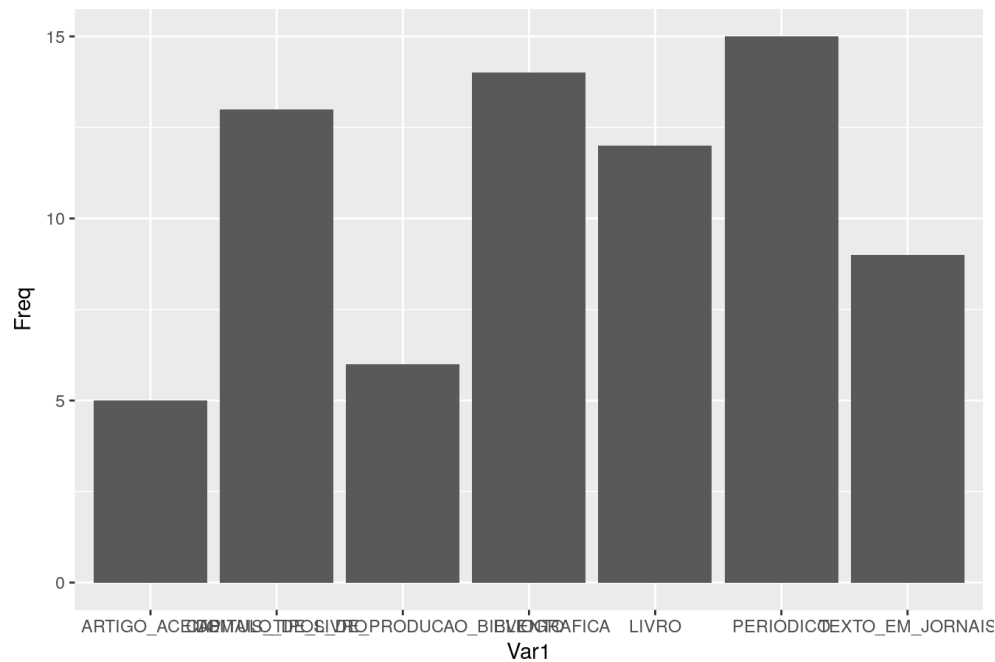
```
ggplot(ambientais.prod, aes(x= Var1, y=Freq)) + geom_col()+  
ggtitle("Número de produções bibliográficas por professor na Ambientais")
```





```
ggplot(agronegocio.prod, aes(x= Var1, y=Freq)) + geom_col()+  
ggtitle("Número de produções bibliográficas por professor na Agronegócio")
```

Número de produções bibliográficas por professor na Agronegócio



```
# Número de publicações por tipo
c("Agronomia" = sum(sapply(agro.prof,
                           function(x) length(x$producao_bibliografica$ARTIGO_ACEITO$ano))),
  "Ambientais" = sum(sapply(ambientais.prof,
                           function(x) length(x$producao_bibliografica$ARTIGO_ACEITO$ano))),
  "Agronegocio" = sum(sapply(agronegocio.prof,
                             function(x) length(x$producao_bibliografica$ARTIGO_ACEITO$ano)))) %>%
  knitr::kable(caption="Número de artigos aceitos por programa")
```

Número de artigos aceitos por programa

	x
Agronomia	9
Ambientais	0
Agronegocio	7

```
c("Agronomia Livro" = sum(sapply(agro.prof, function(x) length(x$producao_bibliografica$LIVRO$ano))),
  "Agronomia Periodico" = sum(sapply(agro.prof, function(x) length(x$producao_bibliografica$PERIODICO$ano))),
  "Agronomia Jornais" = sum(sapply(agro.prof, function(x) length(x$producao_bibliografica$TEXTO_EM_JORNAIS$ano))),
```

```
"Ambientais Livro" = sum(sapply(ambientais.prof, function(x) length(x$producao_bibliografica$LIVRO$ano))),
"Ambientais Periodico" = sum(sapply(ambientais.prof, function(x) length(x$producao_bibliografica$PERIODICO$ano))),
"Ambientais Jornais" = sum(sapply(ambientais.prof, function(x) length(x$producao_bibliografica$TEXT0_EM_JORNAIS$ano))),
"Agronegocio Livro" = sum(sapply(agronegocio.prof, function(x) length(x$producao_bibliografica$LIVRO$ano))),
"Agronegocio Periodico" = sum(sapply(agronegocio.prof, function(x) length(x$producao_bibliografica$PERIODICO$ano))),
"Agronegócio Jornais" = sum(sapply(agronegocio.prof, function(x) length(x$producao_bibliografica$TEXT0_EM_JORNAIS$ano)))) %>%
  knitr::kable(caption = "Número de Livros, Periódicos e Texto em Jonais por programa")
```

Número de Livros, Periódicos e Texto em Jonais por programa

	x
Agronomia Livro	27
Agronomia Periodico	505
Agronomia Jornais	19
Ambientais Livro	5
Ambientais Periodico	31
Ambientais Jornais	5
Agronegocio Livro	46
Agronegocio Periodico	267
Agronegócio Jornais	30

```
# Número de pessoas por quantitativo de produções por pessoa 0 = 1; 1 = 2...
table(unlist(sapply(agro.prof, function(x) length(x$producao_bibliografica$ARTIGO_ACEITO$ano)))) %>%
  knitr::kable(caption = "Número de pessoas por quantitativo de produções de artigo na Agronomia")
```

Número de pessoas por quantitativo de produções de artigo na Agronomia

Var1	Freq
0	18
1	2
3	1
4	1

```
table(unlist(sapply(ambientais.prof, function(x) length(x$producao_bibliografica$ARTIGO_ACEITO$ano)))) %>%
  knitr::kable(caption = "Número de pessoas por quantitativo de produções de artigo em Ambientais")
```

Número de pessoas por quantitativo de produções de artigo em Ambientais

Var1	Freq
0	5

```
table(unlist(sapply(agronegocio.prof, function(x) length(x$producao_bibliografica$ARTIGO_ACEITO$ano)))) %>%  
  knitr::kable(caption = "Número de pessoas por quantitativo de produções de artigo em Agronegócio")
```

Número de pessoas por quantitativo de produções de artigo em Agronegócio

Var1	Freq
0	11
1	3
2	2

```
table(unlist(sapply(agro.prof, function(x) length(x$producao_bibliografica$CAPITULO_DE_LIVRO$ano)))) %>%  
  knitr::kable(caption = "Número de pessoas por quantitativo de produções de Capítulos de Livro na Agronomia")
```

Número de pessoas por quantitativo de produções de Capítulos de Livro na Agronomia

Var1	Freq
0	4
1	9
2	3
3	2
Var1	Freq
32	1
47	1

```
table(unlist(sapply(ambientais.prof, function(x) length(x$producao_bibliografica$CAPITULO_DE_LIVRO$ano)))) %>%  
  knitr::kable(caption = "Número de pessoas por quantitativo de produções de Capítulos de Livros em Ambientais")
```

Número de pessoas por quantitativo de produções de Capítulos de Livros em Ambientais

Var1	Freq
1	2

Var1	Freq
2	2
10	1

```
table(unlist(sapply(agronegocio.prof, function(x) length(x$producao_bibliografica$CAPITULO_DE_LIVRO$ano)))) %>%
  knitr::kable(caption = "Número de pessoas por quantitativo de produções de Capítulos de Livros em Agronegócio")
```

Número de pessoas por quantitativo de produções de Capítulos de Livros em Agronegócio

Var1	Freq
0	3
1	3
2	1
3	2
5	1
6	3
8	1
9	1
10	1

```
table(unlist(sapply(agro.prof, function(x) length(x$producao_bibliografica$LIVRO$ano)))) %>%
  knitr::kable(caption = "Livros por ano na Agronomia")
```

Livros por ano na Agronomia

Var1	Freq
0	16
1	2
4	2
8	1
9	1

```
table(unlist(sapply(ambientais.prof, function(x) length(x$producao_bibliografica$LIVRO$ano)))) %>%
  knitr::kable(caption = "Livros por ano em Ambientais")
```

Livros por ano em Ambientais

Var1	Freq
0	2
1	1
2	2

```
table(unlist(sapply(agronegocio.prof, function(x) length(x$producao_bibliografica$LIVRO$ano))))%>%  
  knitr::kable(caption = "Livros por ano em Agronegocio")
```

Livros por ano em Agronegocio

Var1	Freq
0	4
1	3
2	2
3	2
4	1
5	1
6	1
8	1
10	1

```
table(unlist(sapply(agro.prof, function(x) length(x$producao_bibliografica$PERIODICO$ano))))%>%  
  knitr::kable(caption = "Periódicos por ano na Agronomia")
```

Periódicos por ano na Agronomia

Var1	Freq
7	1
8	1
9	3
12	2
15	2

Var1	Freq
16	1
19	3
22	1
23	1
34	1
35	1
36	1
37	1
38	1
40	1
71	1

```
table(unlist(sapply(ambientais.prof, function(x) length(x$producao_bibliografica$PERIODICO$ano))))%>%
  knitr::kable(caption = "Periódicos por ano em Ambientais")
```

Periódicos por ano em Ambientais

Var1	Freq
1	1
5	2
8	1
Var1	Freq

```
table(unlist(sapply(agronegocio.prof, function(x) length(x$producao_bibliografica$PERIODICO$ano))))%>%
  knitr::kable(caption = "Periódicos por ano em Agronegocio")
```

Periódicos por ano em Agronegocio

Var1	Freq
0	1
1	1
5	1

Var1	Freq
6	1
8	1
12	1
14	2
15	1
18	2
19	1
27	1
35	1
36	1
39	1

```
table(unlist(sapply(agro.prof, function(x) length(x$producao_bibliografica$TEXT0_EM_JORNAIS$ano))))%>%
  knitr::kable(caption = "Textos em jornais por ano na Agronomia")
```

Textos em jornais por ano na Agronomia

Var1	Freq
0	14
1	5
2	2
10	1

```
table(unlist(sapply(ambientais.prof, function(x) length(x$producao_bibliografica$TEXT0_EM_JORNAIS$ano))))%>%
  knitr::kable(caption = "Textos em jornais por ano em ambientais")
```

Textos em jornais por ano em ambientais

Var1	Freq
0	2
1	2
3	1



```
table(unlist(sapply(agronegocio.prof, function(x) length(x$producao_bibliografica$TEXT0_EM_JORNAIS$ano))))%>%
knitr::kable(caption = "Textos em jornais por ano em Agronegócio")
```

Textos em jornais por ano em Agronegócio

Var1	Freq
0	7
1	4
2	2
3	1
4	1
15	1

```
# Número de pessoas que realizaram diferentes tipos de orientações
length(unlist(sapply(agro.prof, function(x) names(x$orientacoes_academicas))))
```

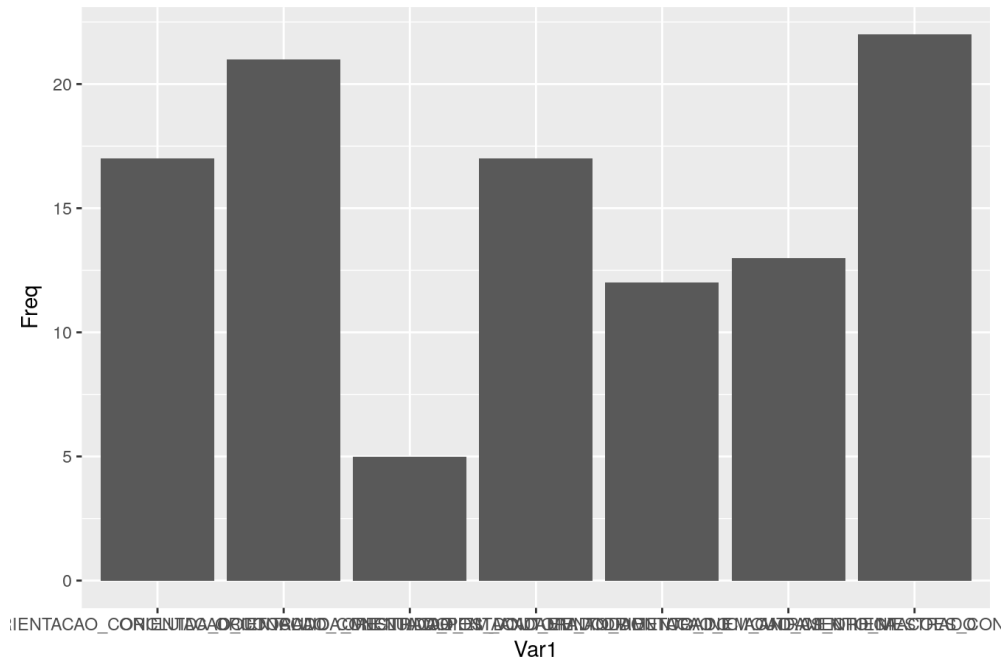
```
## [1] 107
```

```
length(unlist(sapply(ambientais.prof, function(x) names(x$orientacoes_academicas))))
```

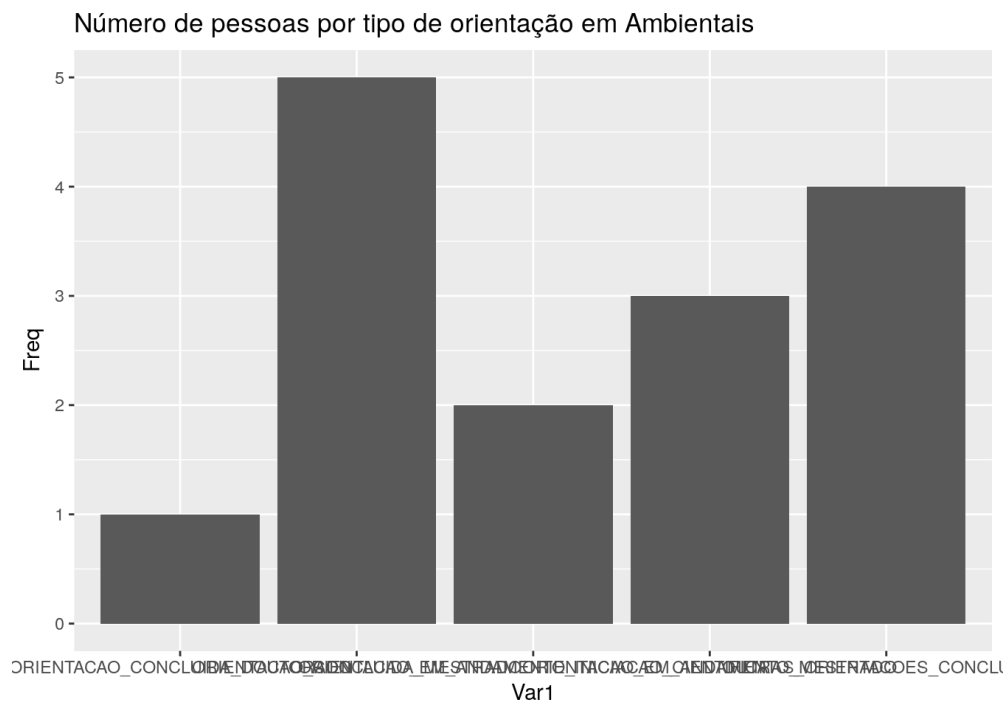
```
## [1] 15
```

```
# Número de pessoas por tipo de orientação
orienta.agro <- table(unlist(sapply(agro.prof, function(x) names(x$orientacoes_academicas))))
orienta.agro <- data.frame(orienta.agro)
ggplot(orienta.agro, aes(x=Var1, y = Freq)) + geom_col()+
  ggtitle("Número de pessoas por tipo de orientação na Agronomia")
```

Número de pessoas por tipo de orientação na Agronomia

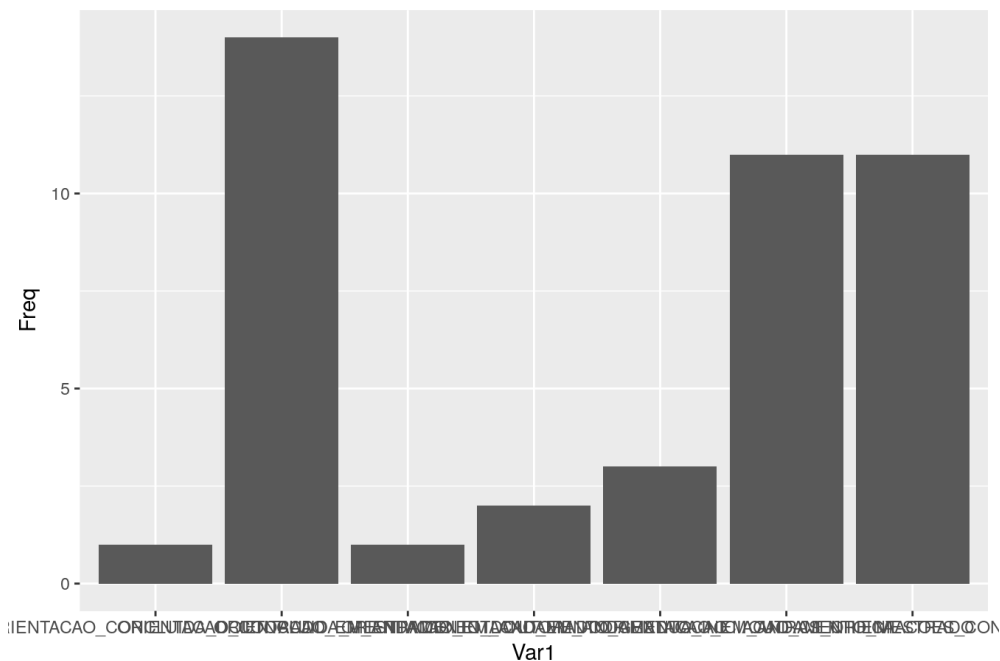


```
orienta.ambientais <- table(unlist(sapply(ambientais.prof, function(x) names(x$orientacoes_academicas))))
orienta.ambientais <- data.frame(orienta.ambientais)
ggplot(orienta.ambientais, aes(x=Var1, y = Freq)) + geom_col()+
  ggtitle("Número de pessoas por tipo de orientação em Ambientais")
```



```
orienta.agronegocio <- table(unlist(sapply(agronegocio.prof, function(x) names(x$orientacoes_academicas))))
orienta.agronegocio <- data.frame(orienta.agronegocio)
ggplot(orienta.agronegocio, aes(x=Var1, y = Freq)) + geom_col()+
  ggtitle("Número de pessoas por tipo de orientação em Agronegócio")
```

Número de pessoas por tipo de orientação em Agronegócio



```
#Número de orientações concluídas
c("Mestrado" = sum(sapply(agro.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTRADO$ano))),
  "Doutorado" = sum(sapply(agro.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTORADO$ano))),
  "Pos Doutorado" = sum(sapply(agro.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_POS_DOUTORADO$ano)))) %>%
  knitr::kable(caption = "Número de orientações concluídas na Agronomia")
```

Número de orientações concluídas na Agronomia

	x
Mestrado	177
Doutorado	87
Pos Doutorado	7

```
c("Mestrado" = sum(sapply(ambientais.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTRADO$ano))),
  "Doutorado" = sum(sapply(ambientais.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTORADO$ano))))
```

```
ADO$ano))),
"Pos Doutorado" = sum(sapply(ambientais.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_PO
S_DOUTORADO$ano)))) %>%
knitr::kable(caption = "Número de Orientações Concluídas em Ambientais")
```

Número de Orientações Concluídas em Ambientais

	x
Mestrado	29
Doutorado	8
Pos Doutorado	0

```
c("Mestrado" = sum(sapply(agronegocio.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MEST
RADO$ano))),
"Doutorado" = sum(sapply(agronegocio.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTO
RADO$ano))),
"Pos-Doutorado" = sum(sapply(agronegocio.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_P
OS_DOUTORADO$ano)))) %>%
knitr::kable(caption = "Número de Orientações Concluídas em Agronegócio")
```

Número de Orientações Concluídas em Agronegócio

	x
Mestrado	123
Doutorado	8
Pos-Doutorado	0

```
# Número de pessoas por quantitativo de orientações por pessoa 0 = 1; 1 = 2...
table(unlist(sapply(agro.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTRADO$ano))))
%>%
knitr::kable(caption = "Número de pessoas por quantitativo de orientações de mestrados na Agronomia")
```

Número de pessoas por quantitativo de orientações de mestrados na Agronomia

Var1	Freq
0	1
1	2
2	1
4	2

Var1	Freq
5	2
6	1
7	2
8	4
9	2
11	1
12	1
13	1
15	1
34	1

```
table(unlist(sapply(agro.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTORADO$ano))))
%>%
  knitr::kable(caption = "Número de pessoas por quantitativo de orientações de doutorados na Agronomia")
```

Número de pessoas por quantitativo de orientações de doutorados na Agronomia

Var1	Freq
0	5
1	2
2	2
3	4
4	3
5	1
7	1
8	1
10	1
13	1
14	1

```
table(unlist(sapply(agro.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_POS_DOUTORADO$ano)))) %>%
```

```
knitr::kable(caption = "Número de pessoas por quantitativo de orientações de Pós-Doutorados na Agronomia")
```

Número de pessoas por quantitativo de orientações de Pós-Doutorados na Agronomia

Var1	Freq
0	17
1	3
2	2
Var1	Freq

```
table(unlist(sapply(ambientais.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTRADO$ano)))) %>%
knitr::kable(caption = "Número de pessoas por quantitativo de orientações de mestrado em Ambientais")
```

Número de pessoas por quantitativo de orientações de mestrado em Ambientais

Var1	Freq
3	1
5	1
6	1
7	1
8	1

```
table(unlist(sapply(ambientais.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTORADO$ano)))) %>%
knitr::kable(caption = "Número de pessoas por quantitativo de orientações de doutorado em Ambientais")
```

Número de pessoas por quantitativo de orientações de doutorado em Ambientais

Var1	Freq
0	4
3	1

```
table(unlist(sapply(ambientais.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_POS_DOUTORADO$ano)))) %>%
knitr::kable(caption = "Número de pessoas por quantitativo de orientações de Pós-Doutorado em Ambientais")
```

Número de pessoas por quantitativo de orientações de Pós-Doutorado em Ambientais

Var1	Freq
0	5

```
table(unlist(sapply(agronegocio.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTRADO$ano)))) %>%
  knitr::kable(caption = "Número de pessoas por quantitativo de orientações de mestrado em Agronegócio")
```

Número de pessoas por quantitativo de orientações de mestrado em Agronegócio

Var1	Freq
0	2
2	1
3	1
4	1
5	2
6	1
7	1
8	3
10	1
11	1
12	1
34	1

```
table(unlist(sapply(agronegocio.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTORADO$ano)))) %>%
  knitr::kable(caption = "Número de pessoas por quantitativo de orientações de doutorado em Agronegócio")
```

Número de pessoas por quantitativo de orientações de doutorado em Agronegócio

Var1	Freq
0	15
8	1

```
table(unlist(sapply(agronegocio.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_POS_DOUTORADO$ano)))) %>%
```



```
knitr::kable(caption = "Número de pessoas por quantitativo de orientações de Pós-Doutorado em Agronegócio")
```

Número de pessoas por quantitativo de orientações de Pós-Doutorado em Agronegócio

Var1	Freq
0	16

```
# Número de orientações por ano
table(unlist(sapply(agro.prof, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTRADO$ano)))) %>%
  knitr::kable(caption = "Número orientações de mestrado por ano na Agronomia")
```

Número orientações de mestrado por ano na Agronomia

Var1	Freq
2010	21
2011	18
2012	25
2013	21
2014	18
2015	29
2016	25
2017	20

```
table(unlist(sapply(agro.prof, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTORADO$ano)))) %>%
  knitr::kable(caption = "Número de orientações de doutorado por ano na Agronomia")
```

Número de orientações de doutorado por ano na Agronomia

Var1	Freq
2010	3
2011	3
2012	13
2013	17
2014	13
2015	12

Var1	Freq
2016	19
2017	7

```
table(unlist(sapply(agro.prof, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_POS_DOUTORADO$ano))))%>%
  knitr::kable(caption = "Número orientações de pós-doutorado por ano na Agronomia")
```

Número orientações de pós-doutorado por ano na Agronomia

Var1	Freq
2011	2
2012	1
2013	1
2014	2
2015	1

```
table(unlist(sapply(ambientais.prof, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTRADO$ano))))%>%
  knitr::kable(caption = "Número orientações de mestrado por ano em Ambientais")
```

Número orientações de mestrado por ano em Ambientais

Var1	Freq
2010	4
2011	3
2012	4
2013	6
2014	3
2015	2
2016	4
2017	3

```
table(unlist(sapply(ambientais.prof, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTORADO$ano))))%>%
```

```
knitr::kable(caption = "Número orientações de doutorado por ano em Ambientais")
```

Número orientações de doutorado por ano em Ambientais

Var1	Freq
2012	1
2013	1
2017	1

```
table(unlist(sapply(ambientais.prof, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_POS_DOUTORADO$ano))))%>%
knitr::kable(caption = "Número orientações de pós-doutorado por ano na Ambientais")
```

Número orientações de pós-doutorado por ano na Ambientais

Freq
------

```
table(unlist(sapply(agronegocio.prof, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTRADO$ano))))%>%
knitr::kable(caption = "Número orientações de mestrado por ano em Agronegocio")
```

Número orientações de mestrado por ano em Agronegocio

Var1	Freq
2010	11
2011	9
2012	16
2013	19
2014	15
2015	17
2016	19
2017	17

```
table(unlist(sapply(agronegocio.prof, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTORADO$ano))))%>%
knitr::kable(caption = "Número orientações de doutorado por ano em Agronegocio")
```

Número orientações de doutorado por ano em Agronegocio

Var1	Freq
2012	2
2014	4
2016	1
2017	1

```
table(unlist(sapply(agronegocio.prof, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_POS_DOUTORADO$ano))))>%>%
  knitr::kable(caption = "Número orientações de pós-doutorado por ano na Agronegocio")
```

Número orientações de pós-doutorado por ano na Agronegocio

Freq

Com os resultados apresentados acima, é possível observar que há uma concentração de pessoas na Grande Área das Ciências Agrárias, o que significa que há pouco acadêmicos de outras áreas aplicando os conhecimentos na Agronomia. Além disso, os dados mostram que o maior número de orientações dos últimos anos foram de iniciação científica e graduação. Já o programa PROFCIAMB possui maior concentração na área de ciências sociais aplicada. O programa de Agronegócio também possui mais orientações na área de Ciências Sociais Aplicadas.

## Arquivo Publicação

```
# Visualizar a estrutura do arquivo de Publicacao
#jsonedit(agro.pub)
#Criando um data-frame com todos os anos
agro.pub.df <- data.frame()
for (i in 1:length(agro.pub[[1]]))
  agro.pub.df <- rbind(agro.pub.df, agro.pub$PERIODICO[[i]])
glimpse(agro.pub.df)
```

```
## Observations: 438
## Variables: 10
## $ natureza      <chr> "COMPLETO", "COMPLETO", "COMPLETO", "COMPLE...
## $ titulo        <chr> "Evaluating the potential of forest species...
## $ periodico     <chr> "Water, Air and Soil Pollution (Print)", "E...
## $ ano           <chr> "2010", "2010", "2010", "2010", "2010", "20...
## $ volume        <chr> "208", "", "34", "5", "134", "30", "30", "3...
## $ issn          <chr> "00496979", "00142336", "01000683", "198099...
## $ paginas       <chr> "79 - 89", "1 - 6", "907 - 916", "133 - 143...
## $ doi           <chr> "10.1007/s11270-009-0150-5", "10.1007/s1068...
## $ autores       <list> [<"Mendes Filho, P. F.", "Vasconcellos, R...
## $ `autores-endogeno` <list> ["9293816616786808", "1377231146497584", "...
```

```
ambientais.pub.df <- data.frame()
for (i in 1:length(ambientais.pub[[1]]))
  ambientais.pub.df <- rbind(ambientais.pub.df, ambientais.pub$PERIODICO[[i]])
glimpse(ambientais.pub.df)
```

```
## Observations: 31
## Variables: 10
## $ natureza      <chr> "COMPLETO", "COMPLETO", "COMPLETO", "COMPLE...
## $ titulo        <chr> "Proteção ambiental e conflitos sociais: an...
## $ periodico     <chr> "Economia Política do Desenvolvimento", "Ci...
## $ ano           <chr> "2010", "2010", "2011", "2011", "2011", "20...
## $ volume        <chr> "3", "62", "9", "Ano11", "31", "41", "43", ...
## $ issn          <chr> "19840756", "00096725", "1678765X", "166718...
## $ paginas       <chr> "19 - 37", "42 - 46", "209 - 224", "22 - 28...
## $ doi           <chr> "", "", "", "", "", "10.1590/S0044-59672011...
## $ autores       <list> [<"CEZAR, K. G.", "BARRETO, C. G.", "Nasci...
## $ `autores-endogeno` <list> ["3536396604369119", "0974214976685536", "...]
```

```
agronegocio.pub.df <- data.frame()
for (i in 1:length(agronegocio.pub[[1]]))
  agronegocio.pub.df <- rbind(agronegocio.pub.df, agronegocio.pub$PERIODICO[[i]])
glimpse(agronegocio.pub.df)
```

```
## Observations: 31
## Variables: 10
## $ natureza      <chr> "COMPLETO", "COMPLETO", "COMPLETO", "COMPLE...
## $ titulo        <chr> "Proteção ambiental e conflitos sociais: an...
## $ periodico     <chr> "Economia Política do Desenvolvimento", "Ci...
## $ ano           <chr> "2010", "2010", "2011", "2011", "2011", "20...
## $ volume        <chr> "3", "62", "9", "Ano11", "31", "41", "43", ...
## $ issn          <chr> "19840756", "00096725", "1678765X", "166718...
## $ paginas       <chr> "19 - 37", "42 - 46", "209 - 224", "22 - 28...
## $ doi           <chr> "", "", "", "", "", "10.1590/S0044-59672011...
## $ autores       <list> [<"CEZAR, K. G.", "BARRETO, C. G.", "Nasci...
## $ `autores-endogeno` <list> ["3536396604369119", "0974214976685536", "...]
```

```
# Limpando o data-frame de listas
agro.pub.df$autores <- gsub("\\\\", "\\ ", "\\ ", "; ", agro.pub.df$autores)
agro.pub.df$autores <- gsub("\\[c\\\\(\\\\\\\\)", "", agro.pub.df$autores)
agro.pub.df$`autores-endogeno` <- gsub("\\", " ", "; ", agro.pub.df$`autores-endogeno`)
agro.pub.df$`autores-endogeno` <- gsub("\\[c\\\\(\\\\\\\\)", "", agro.pub.df$`autores-endogeno`)
glimpse(agro.pub.df)
```

```
## Observations: 438
## Variables: 10
## $ natureza      <chr> "COMPLETO", "COMPLETO", "COMPLETO", "COMPLE...
## $ titulo        <chr> "Evaluating the potential of forest species..."
```

```
## $ periodico      <chr> "Water, Air and Soil Pollution (Print)", "E...
## $ ano            <chr> "2010", "2010", "2010", "2010", "2010", "20...
## $ volume         <chr> "208", "", "34", "5", "134", "30", "30", "3...
## $ issn           <chr> "00496979", "00142336", "01000683", "198099...
## $ paginas        <chr> "79 - 89", "1 - 6", "907 - 916", "133 - 143...
## $ doi            <chr> "10.1007/s11270-009-0150-5", "10.1007/s1068...
## $ autores        <chr> "Mendes Filho, P. F.; Vasconcellos, R.L.F.;...
## $ `autores-endogeno` <chr> "9293816616786808", "1377231146497584", "51..."
```

```
ambientais.pub.df$autores <- gsub("\\", "\\\\", "\\\"", "; ", ambientais.pub.df$autores)
ambientais.pub.df$autores <- gsub("\\{c\\\\(\\\\)", "", ambientais.pub.df$autores)
ambientais.pub.df$`autores-endogeno` <- gsub(",", ";", ambientais.pub.df$`autores-endogeno`)
ambientais.pub.df$`autores-endogeno` <- gsub("\\{c\\\\(\\\\)", "", ambientais.pub.df$`autores-endogeno`)
glimpse(ambientais.pub.df)
```

```
## Observations: 31
## Variables: 10
## $ natureza      <chr> "COMPLETO", "COMPLETO", "COMPLETO", "COMPLE...
## $ titulo        <chr> "Proteção ambiental e conflitos sociais: an...
## $ periodico     <chr> "Economia Política do Desenvolvimento", "Ci...
## $ ano           <chr> "2010", "2010", "2011", "2011", "2011", "20...
## $ volume        <chr> "3", "62", "9", "Ano11", "31", "41", "43", ...
## $ issn          <chr> "19840756", "00096725", "1678765X", "166718...
## $ paginas       <chr> "19 - 37", "42 - 46", "209 - 224", "22 - 28...
## $ doi           <chr> "", "", "", "", "", "10.1590/S0044-59672011...
## $ autores       <chr> "CEZAR, K. G.; BARRETO, C. G.; Nascimento, ...
## $ `autores-endogeno` <chr> "3536396604369119", "0974214976685536", "35..."
```

## Arquivo Orientação

```
#Orientação
#Visualizar a estrutura do json no painel Viewer
#jsonedit(agro.adv)
#Reunir todos os anos e orientações concluídas em um mesmo data-frame
agro.adv.tipo.df <- data.frame(); agro.adv.df <- data.frame()
for (i in 1:length(agro.adv[[1]]))
  agro.adv.tipo.df <- rbind(agro.adv.tipo.df, agro.adv$ORIENTACAO_CONCLUIDA_POS_DOUTORADO[[i]])
agro.adv.df <- rbind(agro.adv.df, agro.adv.tipo.df); agro.adv.tipo.df <- data.frame()
for (i in 1:length(agro.adv[[1]]))
  agro.adv.tipo.df <- rbind(agro.adv.tipo.df, agro.adv$ORIENTACAO_CONCLUIDA_DOUTORADO[[i]])
agro.adv.df <- rbind(agro.adv.df, agro.adv.tipo.df); agro.adv.tipo.df <- data.frame()
for (i in 1:length(agro.adv[[1]]))
  agro.adv.tipo.df <- rbind(agro.adv.tipo.df, agro.adv$ORIENTACAO_CONCLUIDA_MESTRADO[[i]])
agro.adv.df <- rbind(agro.adv.df, agro.adv.tipo.df)
glimpse(agro.adv.df)
```

```
## Observations: 263
## Variables: 13
## $ natureza          <chr> "Supervisão de pós-doutorado", "Su...
## $ titulo            <chr> "", "", "", "", "", "", "", "Fator...
## $ ano               <chr> "2011", "2011", "2012", "2013", "2...
## $ id_lattes_aluno   <chr> "", "", "", "", "", "", "", "", ""...
## $ nome_aluno        <chr> "Sebastião Avelino Neto", "Patrici...
## $ instituicao        <chr> "Universidade Estadual de Goiás", ...
## $ curso             <chr> "", "", "", "", "", "", "", "Entom...
## $ codigo_do_curso   <chr> "", "", "", "", "", "", "", "32010...
## $ bolsa            <chr> "SIM", "SIM", "SIM", "SIM", "SIM",...
## $ agencia_financiadora <chr> "Coordenação de Aperfeiçoamento de...
## $ codigo_agencia_financiadora <chr> "045000000000", "045000000000", "0...
## $ nome_orientadores <list> ["Delvio Sandri", "Jose Ricardo P...
## $ id_lattes_orientadores <list> ["0720020819590155", "59902627103...
```

```
ambientais.adv.tipo.df <- data.frame(); ambientais.adv.df <- data.frame()
for (i in 1:length(ambientais.adv[[1]]))
  ambientais.adv.tipo.df <- rbind(ambientais.adv.tipo.df, ambientais.adv$ORIENTACAO_CONCLUIDA_POS_DOUTORADO[[i]])
ambientais.adv.df <- rbind(ambientais.adv.df, ambientais.adv.tipo.df); ambientais.adv.tipo.df <- data.frame()
for (i in 1:length(ambientais.adv[[1]]))
  ambientais.adv.tipo.df <- rbind(ambientais.adv.tipo.df, ambientais.adv$ORIENTACAO_CONCLUIDA_DOUTORADO[[i]])
ambientais.adv.df <- rbind(ambientais.adv.df, ambientais.adv.tipo.df); ambientais.adv.tipo.df <- data.frame()
for (i in 1:length(ambientais.adv[[1]]))
  ambientais.adv.tipo.df <- rbind(ambientais.adv.tipo.df, ambientais.adv$ORIENTACAO_CONCLUIDA_MESTRADO[[i]])
ambientais.adv.df <- rbind(ambientais.adv.df, ambientais.adv.tipo.df)
glimpse(ambientais.adv.df)
```

```
## Observations: 32
## Variables: 13
## $ natureza          <chr> "Tese de doutorado", "Tese de dout...
## $ titulo            <chr> "Educação Ambiental para gestão su...
## $ ano               <chr> "2012", "2013", "2017", "2010", "2...
## $ id_lattes_aluno   <chr> "8651168588446017", "", "", "08329...
## $ nome_aluno        <chr> "Ana Lúcia Soares Machado", "Valér...
## $ instituicao        <chr> "Universidade de Brasília", "Unive...
## $ curso             <chr> "Centro de Desenvolvimento Sustent...
## $ codigo_do_curso   <chr> "900000008", "900000025", "51500442"...
## $ bolsa            <chr> "SIM", "SIM", "SIM", "SIM", "NAO",...
## $ agencia_financiadora <chr> "Coordenação de Aperfeiçoamento de...
## $ codigo_agencia_financiadora <chr> "045000000000", "045000000000", "0...
## $ nome_orientadores <list> ["Izabel Cristina Bruno Bacellar ...
## $ id_lattes_orientadores <list> ["6400809887186849", "64008098871...
```

```
agronegocio.adv.tipo.df <- data.frame(); agronegocio.adv.df <- data.frame()
for (i in 1:length(agronegocio.adv[[1]]))
  agronegocio.adv.tipo.df <- rbind(agronegocio.adv.tipo.df, agronegocio.adv$ORIENTACAO_CONCLUIDA_POS_DOUTORADO[[i]])
```

```

agronegovio.adv.df <- rbind(ambientais.adv.df, ambientais.adv.tipo.df); agronegovio.adv.tipo.df <- data.frame()
for (i in 1:length(agronegovio.adv[[1]]))
  agronegovio.adv.tipo.df <- rbind(agronegovio.adv.tipo.df, agronegovio.adv$ORIENTACAO_CONCLUIDA_DOUTORADO[[i]])
agronegovio.adv.df <- rbind(agronegovio.adv.df, agronegovio.adv.tipo.df); agronegovio.adv.tipo.df <- data.frame()
for (i in 1:length(agronegovio.adv[[1]]))
  agronegovio.adv.tipo.df <- rbind(agronegovio.adv.tipo.df, agronegovio.adv$ORIENTACAO_CONCLUIDA_MESTRADO[[i]])
agronegovio.adv.df <- rbind(agronegovio.adv.df, agronegovio.adv.tipo.df)
glimpse(agronegovio.adv.df)

```

```

## Observations: 129
## Variables: 13
## $ natureza          <chr> "Tese de doutorado", "Tese de dout...
## $ titulo            <chr> "EFEITO DE SILÍCIO EM CARACTERÍSTI...
## $ ano               <chr> "2012", "2012", "2014", "2014", "2...
## $ id_lattes_aluno    <chr> "8986469050322253", "5679819360493...
## $ nome_aluno        <chr> "Marília Cristina dos Santos", "An...
## $ instituicao        <chr> "Universidade de Brasília", "Unive...
## $ curso             <chr> "Agronomia", "Agronomia", "Agronom...
## $ codigo_do_curso    <chr> "90000016", "90000016", "90000021"...
## $ bolsa             <chr> "SIM", "SIM", "NAO", "SIM", "SIM",...
## $ agencia_financiadora <chr> "Coordenação de Aperfeiçoamento de...
## $ codigo_agencia_financiadora <chr> "045000000000", "045000000000", "...
## $ nome_orientadores  <list> ["Ana Maria Resende Junqueira", "...
## $ id_lattes_orientadores <list> ["6823749061158873", "68237490611...

```

```

#Transformar as colunas de listas em caracteres eliminando c("")
agro.adv.df$nome_orientadores <- gsub("\\|c\\(|\\)", "", agro.adv.df$nome_orientadores)
agro.adv.df$id_lattes_orientadores <- gsub("\\|c\\(|\\)", "", agro.adv.df$id_lattes_orientadores)

ambientais.adv.df$nome_orientadores <- gsub("\\|c\\(|\\)", "", ambientais.adv.df$nome_orientadores)
ambientais.adv.df$id_lattes_orientadores <- gsub("\\|c\\(|\\)", "", ambientais.adv.df$id_lattes_orientadores)

agronegovio.adv.df$nome_orientadores <- gsub("\\|c\\(|\\)", "", agronegovio.adv.df$nome_orientadores)
agronegovio.adv.df$id_lattes_orientadores <- gsub("\\|c\\(|\\)", "", agronegovio.adv.df$id_lattes_orientadores)

#Separar as colunas com dois orientadores

agro.adv.df <- separate(agro.adv.df, nome_orientadores, into = c("ori1", "ori2"), sep = ",")

```

```

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 255 rows [1,
## 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].

```

```

agro.adv.df <- separate(agro.adv.df, id_lattes_orientadores, into = c("idLattes1", "idLattes2"), sep = ",")

```

```

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 255 rows [1,
## 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].

```



```
ambientais.adv.df <- separate(ambientais.adv.df, nome_orientadores, into = c("ori1", "ori2"), sep = ",")
```

```
## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 32 rows [1,  
## 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
ambientais.adv.df <- separate(ambientais.adv.df, id_lattes_orientadores, into = c("idLattes1", "idLattes2"), sep  
= ",")
```

```
## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 32 rows [1,  
## 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
agronegocio.adv.df <- separate(agronegocio.adv.df, nome_orientadores, into = c("ori1", "ori2"), sep = ",")
```

```
## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 119 rows  
## [1, 2, 5, 6, 7, 8, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23,  
## 24, ...].
```

```
agronegocio.adv.df <- separate(agronegocio.adv.df, id_lattes_orientadores, into = c("idLattes1", "idLattes2"), se  
p = ",")
```

```
## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 119 rows  
## [1, 2, 5, 6, 7, 8, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23,  
## 24, ...].
```

```
#Número de orientacoes por ano  
table(agro.adv.df$ano) %>%  
knitr::kable(caption = "Número de orientações por ano na Agronomia")
```

Número de orientações por ano na Agronomia

Var1	Freq
2010	24
2011	23
2012	39
2013	36
2014	31
2015	41
2016	43

Var1	Freq
2017	26

```
table(ambientais.adv.df$ano) %>%
  knitr::kable(caption = "Número de orientações por ano em Ambientais")
```

Número de orientações por ano em Ambientais

Var1	Freq
2010	4
2011	3
2012	5
2013	7
2014	3
2015	2
2016	4
2017	4

```
table(agronegocio.adv.df$ano) %>%
  knitr::kable(caption = "Número de orientações por ano em Agronegócio")
```

Número de orientações por ano em Agronegócio

Var1	Freq
2010	11
2011	9
2012	20
2013	18
2014	19
2015	14
2016	21
2017	17

```
#Tabela com nome de professor e numero de orientacoes
head(sort(table(rbind(agro.adv.df$ori1, agro.adv.df$ori2)), decreasing = TRUE), 20)
```

```
##
##      Ana Maria Resende Junqueira      Jose Ricardo Peixoto
##                                40                                26
##      Maria Lucrecia Gerosa Ramos      Fábio Gelape Faleiro
##                                21                                20
##      Jean Kleber de Abreu Mattos      Cícero Célio de Figueiredo
##                                19                                16
##      Marilusa Pinto Coelho Lacerda      Luiz Eduardo Bassay Blum
##                                15                                13
## Rose Gomes Monnerat Solon de Pontes      Delvio Sandri
##                                13                                11
##      Nara Oliveira Silva Souza      Warley Marcos Nascimento
##                                11                                11
##      Carlos Roberto Spehar      Cristina Schetino Bastos
##                                9                                7
##      Ernandes Rodrigues de Alencar      Ricardo Carmona
##                                7                                7
##      Marcelo Fagioli      Alessandra Monteiro de Paula
##                                6                                4
##      Jose Ricardo Peixoto      Osvaldo Kiyoshi Yamanishi
##                                3                                3
```

```
head(sort(table(rbind(ambientais.adv.df$ori1, ambientais.adv.df$ori2)), decreasing = TRUE), 20)
```

```
##
## Izabel Cristina Bruno Bacellar Zaneti
##                                11
##      Mauricio de Carvalho Amazonas
##                                7
##      Thomas Ludewigs
##                                6
##      Pedro Henrique Zuchi da Conceição
##                                5
##      Cristiane Gomes Barreto
##                                3
```

```
head(sort(table(rbind(agronegocio.adv.df$ori1, ambientais.adv.df$ori2)), decreasing = TRUE), 20)
```

```
## Warning in rbind(agronegocio.adv.df$ori1, ambientais.adv.df$ori2): number
## of columns of result is not a multiple of vector length (arg 2)
```

```
##
##      Ana Maria Resende Junqueira      Flávio Borges Botelho Filho
```

##	46	12
##	João Paulo Guimarães Soares	Mauro Eduardo Del Grossi
##	11	10
##	Carlos Rosano Pena	José Márcio Carvalho
##	8	7
##	Suzana Maria Valle Lima	Antônio Maria Gomes de Castro
##	7	6
##	Itibere Saldanha Silva	Maria Júlia Pantoja
##	5	5
##	Marlon Vinícius Brisola	José Eustáquio Ribeiro Vieira Filho
##	4	3
##	Patricia Guarnieri	Vânia Ferreira Roque-Specht
##	3	2

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

Com os dados apresentados, é possível avaliar de acerca do desempenho dos professores, do número de publicações e número de orientações de cada professor. Com essas análises, é possível verificar se o número de publicações é comprometido caso um professor esteja com um grande número de orientações ou se o número de professores é compatível com a quantidade de alunos participantes do programa.

## 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 â€¦ but some donâ€™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*.

```

# Funcoes

# converte as colunas de um dataframe tipo lista em tipo character
cv_tplista2tpchar <- function( df ) {
  for( variavel in names(df)) {
    if (class(df[[variavel]]) == "list" ) {
      df[[variavel]] <- lapply(df[[variavel]] , function(x) lista2texto( x ) )
      df[[variavel]] <- as.character( df[[variavel]] )
    }
  }
  return(df)
}
###

# converte o conteudo de lista em array de characters
lista2texto <- function( lista ) {
  if(is.null(lista)) {
    return ( NULL )
  }
  saida <- ""
  for( j in 1:length(lista)) {
    for( i in 1:length(lista[[j]]) ) {
      elemento <- lista[[j]][i]
      if( !is.null(elemento)) {
        if( i == length(lista[[j]]) & j == length(lista) ) {
          # se for o ultimo elemento nao coloque o ponto e virgula no final
          saida <- paste0( saida , elemento )
        } else {
          # enquanto nao for o ultimo coloque ; separando os elementos concatenados
          saida <- paste0( saida , elemento , sep = " ; " )
        }
      }
    }
  }
  return( saida )
}

# Converte producao elattes separada por anos em um unico dataframe
converte_producao2dataframe<- function( lista_producao ) {
  df_saida <- NULL

  for( ano in names(lista_producao)) {
    df_saida <- rbind(df_saida , lista_producao[[ano]])
  }

  # converte tipo lista em array de character
  df_saida <- cv_tplista2tpchar(df_saida)
  return(df_saida)
}

```

```

}

#concatena dois dataframes com  colunas diferentes
concatenadf <- function( df1, df2) {
  #cria colunas de df1 que faltam em df2
  for( coluna in names(df1) ) {
    if( !is.element(coluna, names(df2)) ) {
      df2[coluna] <- NA
    }
  }

  #cria colunas de df2 que faltam em df1
  for( coluna in names(df2) ) {

    if( !is.element(coluna, names(df1)) ) {
      df1[coluna] <- NA
    }
  }

  #faz o rbind dos dois dataframes
  df_final <- rbind(df1 , df2)
  return(df_final)
}

# Extracao dos perfis dos professores

extrai_lperfil <- function( professor ) {
  idLattes <- names(professor)
  nome <- professor[[idLattes]]$nome
  resumo_cv <- professor[[idLattes]]$resumo_cv
  endereco_profissional <- professor[[idLattes]]$endereco_profissional #list
  instituicao <- endereco_profissional$instituicao
  orgao <- endereco_profissional$orgao
  unidade <- endereco_profissional$unidade
  DDD <- endereco_profissional$DDD
  telefone <- endereco_profissional$telefone
  bairro <- endereco_profissional$bairro
  cep <- endereco_profissional$cep
  cidade <- endereco_profissional$cidade
  senioridade <- professor[[idLattes]]$senioridade
  df_lperfil <- data.frame( idLattes , nome, resumo_cv ,instituicao ,
                           orgao, unidade , DDD, telefone, bairro,cep,cidade , senioridade,
                           stringsAsFactors = FALSE)

  return(df_lperfil)
}

extrai_perfis <- function(jsonProfessores) {

```

```

df_saida <- data.frame()
for( i in 1:length(jsonProfessores)) {
  jsonProfessor <- jsonProfessores[i]
  df_professor <- extrai_lperfil(jsonProfessor)
  if( nrow(df_saida) > 0 ) {
    df_saida <- rbind(df_saida , df_professor)
  } else {
    df_saida <- df_professor
  }
}

return(df_saida)
}

# Extracao da producao bibliografica dos professores

extrai_lproducao <- function(professor) {
  idLattes <- names(professor)
  df_lproducao <- NULL
  producao_bibliografica <- professor[[idLattes]]$producao_bibliografica #list
  for( tipo_producao in names(producao_bibliografica)) {
    df_temporario <- cv_tplista2tpchar ( producao_bibliografica[[tipo_producao]])
    df_temporario$tipo_producao <- tipo_producao
    df_temporario$idLattes <- idLattes
    df_lproducao <- concatenadf( df_lproducao , df_temporario )
  }
  return(df_lproducao)
}

extrai_producoes <- function( jsonProfessores) {
  df_saida <- data.frame()
  for( i in 1:length(jsonProfessores)) {
    jsonProfessor <- jsonProfessores[i]
    df_producao <- extrai_lproducao(jsonProfessor)
    if( nrow(df_saida) > 0 ) {
      df_saida <- concatenadf(df_saida , df_producao)
    } else {
      df_saida <- df_producao
    }
  }
  df_saida <- df_saida %>% filter( !is.na(tipo_producao))
  return(df_saida)
}

# Extracao das orientacoes dos professores

extrai_lorientacao <- function(professor) {
  idLattes <- names(professor)
  df_lorientacao <- NULL
  orientacoes_academicas <- professor[[idLattes]]$orientacoes_academicas #list
  for( orientacao in names(orientacoes_academicas) ) {

```

```

    df_temporario <- cv_tplista2tpchar ( orientacoes_academicas[[orientacao]])
    df_temporario$orientacao <- orientacao
    df_temporario$idLattes <- idLattes
    df_lorientacao <- concatenadf( df_lorientacao , df_temporario )
  }
  return(df_lorientacao)
}

extraire_orientacoes <- function(jsonProfessores) {
  df_saida <- data.frame()
  for( i in 1:length(jsonProfessores)) {
    jsonProfessor <- jsonProfessores[i]
    df_orientacao <- extraire_lorientacao(jsonProfessor)
    if( nrow(df_saida) > 0 ) {
      df_saida <- concatenadf(df_saida , df_orientacao)
    } else {
      df_saida <- df_orientacao
    }
  }
  df_saida <- df_saida %>% filter(!is.na(idLattes))
  return(df_saida)
}

# Extracao das areas de atuacao dos professores

extraire_larea_de_atuacao <- function(professor){
  idLattes <- names(professor)
  df_larea <- professor[[idLattes]]$areas_de_atuacao
  df_larea$idLattes <- idLattes
  return(df_larea)
}

extraire_areas_atuacao <- function(jsonProfessores){
  df_saida <- data.frame()
  for( i in 1:length(jsonProfessores)) {
    jsonProfessor <- jsonProfessores[i]
    df_area_atuacao <- extraire_larea_de_atuacao(jsonProfessor)
    if( nrow(df_saida) > 0 ) {
      df_saida <- concatenadf(df_saida , df_area_atuacao)
    } else {
      df_saida <- df_area_atuacao
    }
  }
  df_saida <- df_saida %>% filter( !is.na(idLattes))
  return(df_saida)
}

##### Inicio

# colocar o directorio onde está o arquivo json de perfis a serem lidos
agro.prof.json <- read_file("dados/233.profile.json")
ambientais.prof.json <- read_file("dados/ambientais.profile.json")

```



```
agronegocio.prof.json <- read_file("dados/232.profile.json")
# agro.prof.df.capes <- read.csv("data/PesqPosCapes.csv",
#                               sep = ";", header = TRUE, colClasses = "character")

agro.prof <- fromJSON(agro.prof.json)
ambientais.prof <- fromJSON(ambientais.prof.json)
agronegocio.prof <- fromJSON(agronegocio.prof.json)
length(agro.prof)
```

```
## [1] 22
```

```
length(ambientais.prof)
```

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

```
length(agronegocio.prof)
```

```
## [1] 16
```

```
# extrai perfis dos professores
agro.prof.df.professores <- extrai_perfis(agro.prof)
ambientais.prof.df.professores <- extrai_perfis(ambientais.prof)
agronegocio.prof.df.professores <- extrai_perfis(agronegocio.prof)

# extrai producao bibliografica de todos os professores
agro.prof.df.publicacoes <- extrai_producoes(agro.prof)
ambientais.prof.df.publicacoes <- extrai_producoes(ambientais.prof)
agronegocio.prof.df.publicacoes <- extrai_producoes(agronegocio.prof)

#extrai orientacoes
agro.prof.df.orientacoes <- extrai_orientacoes(agro.prof)
ambientais.prof.df.orientacoes <- extrai_orientacoes(ambientais.prof)
agronegocio.prof.df.orientacoes <- extrai_orientacoes(agronegocio.prof)

#extrai areas de atuacao
agro.prof.df.areas.de.atuacao <- extrai_areas_atuacao(agro.prof)
ambientais.prof.df.areas.de.atuacao <- extrai_areas_atuacao(ambientais.prof)
agronegocio.prof.df.areas.de.atuacao <- extrai_areas_atuacao(agronegocio.prof)

#salva os daframes
save(agro.prof.df.professores, agro.prof.df.publicacoes,
     agro.prof.df.orientacoes, agro.prof.df.areas.de.atuacao, file = "dataframes.Rda")

save(ambientais.prof.df.professores, ambientais.prof.df.publicacoes,
     ambientais.prof.df.orientacoes, ambientais.prof.df.areas.de.atuacao, file = "dataframes.Rda")

save(agronegocio.prof.df.professores, agronegocio.prof.df.publicacoes,
```

```

    agronegocio.prof.df.orientacoes, agronegocio.prof.df.areas.de.atuacao, file = "dataframes.Rda")

#cria arquivo para análise
agro.prof.df <- data.frame()
agro.prof.df <- agro.prof.df.professores %>%
  select(idLattes, nome, resumo_cv, senioridade) %>%
  left_join(
    agro.prof.df.orientacoes %>%
      select(orientacao, idLattes) %>%
      filter(!grepl("EM_ANDAMENTO", orientacao)) %>%
      group_by(idLattes) %>%
      count(orientacao) %>%
      spread(key = orientacao, value = n),
    by = "idLattes") %>%
  left_join(
    agro.prof.df.publicacoes %>%
      select(tipo_producao, idLattes) %>%
      filter(!grepl("ARTIGO_ACEITO", tipo_producao)) %>%
      group_by(idLattes) %>%
      count(tipo_producao) %>%
      spread(key = tipo_producao, value = n),
    by = "idLattes") %>%
  left_join(
    agro.prof.df.areas.de.atuacao %>%
      select(area, idLattes) %>%
      group_by(idLattes) %>%
      summarise(n_distinct(area)),
    by = "idLattes")

glimpse(agro.prof.df)

```

```

## Observations: 22
## Variables: 15
## $ idLattes          <chr> "0646231743976574", "07...
## $ nome              <chr> "Ernandes Rodrigues de ...
## $ resumo_cv          <chr> "Possui graduação em En...
## $ senioridade        <chr> "9", "9", "9", "9", "9"...
## $ ORIENTACAO_CONCLUIDA_DOUTORADO <int> NA, 1, 4, 2, 7, 3, 3, 1...
## $ ORIENTACAO_CONCLUIDA_MESTRADO <int> 7, 8, 5, 5, 8, 4, 9, NA...
## $ ORIENTACAO_CONCLUIDA_POS_DOUTORADO <int> NA, 2, NA, NA, NA, NA, ...
## $ OUTRAS_ORIENTACOES_CONCLUIDAS <int> 50, 29, 8, 80, 65, 1, 2...
## $ CAPITULO_DE_LIVRO <int> 1, 1, 2, 1, NA, 1, 4, N...
## $ DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA <int> NA, NA, NA, NA, NA, 1, ...
## $ EVENTO            <int> 17, 44, 8, 41, 35, 1, 2...
## $ LIVRO             <int> NA, NA, 1, NA, NA, NA, ...
## $ PERIODICO         <int> 15, 22, 12, 9, 19, 8, 3...
## $ TEXTO_EM_JORNAIS  <int> NA, NA, 1, 1, 1, NA, NA...
## $ `n_distinct(area)` <int> 2, 3, 3, 1, 2, 1, 1, 1...

```

```

ambientais.prof.df <- data.frame()
ambientais.prof.df <- ambientais.prof.df.professores %>%
  select(idLattes, nome, resumo_cv, senioridade) %>%
  left_join(
    ambientais.prof.df.orientacoes %>%
      select(orientacao, idLattes) %>%
      filter(!grepl("EM_ANDAMENTO", orientacao)) %>%
      group_by(idLattes) %>%
      count(orientacao) %>%
      spread(key = orientacao, value = n),
    by = "idLattes") %>%
  left_join(
    ambientais.prof.df.publicacoes %>%
      select(tipo_producao, idLattes) %>%
      filter(!grepl("ARTIGO_ACEITO", tipo_producao)) %>%
      group_by(idLattes) %>%
      count(tipo_producao) %>%
      spread(key = tipo_producao, value = n),
    by = "idLattes") %>%
  left_join(
    ambientais.prof.df.areas.de.atuacao %>%
      select(area, idLattes) %>%
      group_by(idLattes) %>%
      summarise(n_distinct(area)),
    by = "idLattes")

glimpse(ambientais.prof.df)

```

```

## Observations: 5
## Variables: 14
## $ idLattes          <chr> "0974214976685536", "35...
## $ nome              <chr> "Mauricio de Carvalho A...
## $ resumo_cv          <chr> "Possui graduação em En...
## $ senioridade        <chr> "9", "9", "8", "8", "7"
## $ ORIENTACAO_CONCLUIDA_DOUTORADO <int> NA, NA, NA, 3, NA
## $ ORIENTACAO_CONCLUIDA_MESTRADO <int> 7, 3, 6, 8, 5
## $ OUTRAS_ORIENTACOES_CONCLUIDAS <int> 5, NA, 12, 17, 15
## $ CAPITULO_DE_LIVRO <int> 1, 2, 1, 10, 2
## $ DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA <int> NA, 1, NA, 2, NA
## $ EVENTO            <int> 1, 14, 1, 14, 10
## $ LIVRO              <int> NA, 2, 1, 2, NA
## $ PERIODICO          <int> 5, 12, 5, 8, 1
## $ TEXTO_EM_JORNAIS   <int> NA, 3, 1, NA, 1
## $ `n_distinct(area)` <int> 1, 4, 4, 1, 1

```

```

agronegocio.prof.df <- data.frame()
agronegocio.prof.df <- agronegocio.prof.df.professores %>%
  select(idLattes, nome, resumo_cv, senioridade) %>%
  left_join(

```

```

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

glimpse(agronegocio.prof.df)

```

```

## Observations: 16
## Variables: 14
## $ idLattes          <chr> "0011938955607677", "03...
## $ nome              <chr> "Maria Júlia Pantoja", ...
## $ resumo_cv         <chr> "possui graduação em Ps...
## $ senioridade       <chr> "8", "9", "7", "9", "9"...
## $ ORIENTACAO_CONCLUIDA_DOUTORADO <int> NA, NA, NA, NA, NA, NA,...
## $ ORIENTACAO_CONCLUIDA_MESTRADO  <int> 5, 12, 8, 11, NA, 8, 7,...
## $ OUTRAS_ORIENTACOES_CONCLUIDAS  <int> 7, NA, NA, 7, 10, NA, N...
## $ CAPITULO_DE_LIVRO  <int> 5, NA, 3, 6, 1, NA, 9, ...
## $ DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA <int> NA, NA, NA, 12, NA, NA,...
## $ EVENTO            <int> 17, 13, NA, 70, 40, NA,...
## $ LIVRO              <int> 2, 1, NA, 2, 1, 10, 6, ...
## $ PERIODICO          <int> 6, 8, 18, 15, 39, NA, 1...
## $ TEXTO_EM_JORNAIS   <int> 1, NA, NA, 2, NA, NA, N...
## $ `n_distinct(area)` <int> 2, 2, 1, 2, 1, 5, 3, 3,...

```

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

Como já informado, a integração dos dados envolve a união (merge) de diferentes tabelas para criar um único *dataset* para ser utilizado no R, por exemplo.

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

Como já informado, a formatação de dados envolve a realização de pequenas alterações na estrutura dos dados, como a ordem das variáveis, para permitir a execução de determinado método de data mining.

## Referências

- Fernandes, Jorge H C, Ricardo Barros Sampaio, e João Ribas de Moura. *â€œCiência 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 UnBâ€œ. 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.*
- Azevedo, Mário Luiz Neves de, João Ferreira de Oliveira, e Afrânio Mendes Catani. *â€œO Sistema Nacional de Pós-Graduação (SNPG) e o Plano Nacional de Educação (PNE 2014-2024): regulação, avaliação e financiamentoâ€œ. Revista Brasileira de Política e Administração da Educação 32, n.º 3 (2016). <http://dx.doi.org/10.21573/vol32n32016.68576>.*
- Can, Fazli, Tansel Årzyer, e Faruk Polat, orgs. *State of the Art Applications of Social Network Analysis. Lecture Notes in Social Networks. Switzerland: Springer International Publishing, 2014.*
- CAPES. *â€œDocumentos de Áreaâ€œ. CAPES.gov.br. Acessado 12 de junho de 2018. <http://avaliacaoquadrienal.capes.gov.br/documentos-de-area>.*
- *â€œâ€œâ€œâ€œ. â€œPlano Nacional de Pós-Graduação - PNPG 2011/2020 Vol. 1â€œ. Brasília - DF, dezembro de 2010. <http://www.capes.gov.br/images/stories/download/Livros-PNPG-Volume-I-Mont.pdf>.*
- *â€œâ€œâ€œâ€œ. â€œPlano Nacional de Pós-Graduação - PNPG 2011/2020 Vol. 2â€œ. 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).*
- *â€œâ€œâ€œâ€œ. â€œSucupira: coleta de dados, docentes de pós-graduação stricto sensu no Brasil 2015â€œ. 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 RÃ¼diger Wirth. *â€œCRISP-DM 1.0: Step-by-Step Data Mining Guideâ€œ. USA: CRISP-DM Consortium, 2000. <https://www.the-modeling-agency.com/crisp-dm.pdf>.*
- Datacamp. *â€œMachine Learning with R (Skill Track)â€œ. Datacamp, 2018. <https://www.datacamp.com/tracks/machine-learning>.*
- Fernandes, Jorge H C, e Ricardo Barros Sampaio. *â€œDataScienceForAllâ€œ. Zotero, 13 de junho de 2018. <https://www.zotero.org/groups/2197167/datascienceforall>.*
- *â€œâ€œâ€œâ€œ. â€œEspecificação 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)â€œ. 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. *â€œCiência 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 UnBâ€œ. 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. *â€œJob profile: Higher education lecturerâ€œ, 2018. <https://www.prospects.ac.uk/job-profiles/higher-education-lecturer>.*
- Kalpazidou Schmidt, Evanthia, e Ebbe Krogh Graversen. *â€œPersistent factors facilitating excellence in research environmentsâ€œ. Higher Education 75, n.º 2 (1.º de fevereiro de 2018): 341â€œ63. <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. *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](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)
- 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, Áscar Marbán, e Covadonga Fernández. *A survey of data mining and knowledge discovery process models and methodologies*. The Knowledge Engineering Review 25, n.º 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.º 3 (julho de 2013): 11-17.
- Schwartzman, Simon. *A Ciência da Ciência*. Ciência Hoje 2, n.º 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.
- 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 Grolemund. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. USA: O'Reilly, 2016.