Ciência de Dados para Todos (Data Science For All) - 2018.1 - Análise da Produção Científica e Acadêmica da Universidade de Brasília - Relatório Final da Disciplina - Áreas Geografia, Geologia, Geociências Aplicadas e Geodinâmica

02/07/2018

Introdução

A análise da Produção Científica e Acadêmica da Universidade de Brasília é relevante para o entendimento de como a Universidade está gerindo e investindo seus recursos, para as pessoas do meio acadêmico obterem informações que lhe sejam úteis e contribui na divulgação do conhecimento científico através de uma abordagem simplificada do mesmo para as pessoas não inseridas nesse ambiente. Assim, esse trabalho tem como objetivo o uso da disciplina de Ciência dos Dados para analisar, entender e gerar novas informações sobre os dados acadêmicos da Universidade de Brasília.

Esse trabalho terá o enfoque nas áreas de Geografia, Geologia, Geociências Aplicadas e Geodinâmica, devido a gama de áreas existentes na universidade, dessa forma permitindo uma análise mais detalhada e robusta dos dados a serem trabalhos. Para a análise, será utilizado *datasets* referentes aos professores, publicações e orientações dos mesmos, os quais foram disponibilizados na plataforma E-Lattes.

Por fim, ao longo desse trabalho será demonstrado como é realizado um trabalho de Ciência dos Dados, incluindo o conhecimento do contexto dos dados que serão trabalhados, seleção dos dados mais relevantes, limpeza e análise para gerar gráficos e informações de utilidade para aqueles interessados nas produções científicas e acadêmicas, além da possibilidade de obter informações desconhecidas até então. Com isso, esperamos que ao fim desse trabalho o entendimento de como as áreas de Geografia e Geociências em geral da Universidade de Brasília estão trabalhando e pesquisando.

Metodologia

Ferramentas e Técnicas

Foi utilizada a linguagem R e plataforma RStudio para a manipulação e análise dos dados, e este documento foi gerado no formato R Markdown. Foram utilizadas as seguintes bibliotecas no RStudio:

```
library(jsonlite)
library(listviewer)
library(readxl)
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
library(readr)
library(purrr)
```

```
##
## Attaching package: 'purrr'
```

```
## The following object is masked from 'package:jsonlite':
##
##     flatten
```

```
library(tibble)
library(ggpubr)
```

```
## Loading required package: magrittr
```

```
##
## Attaching package: 'magrittr'
```

```
## The following object is masked from 'package:purrr':
##
##     set_names
```

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

CRISP-DM

A metodologia aplicada para produção desse relatório é a *Cross Industry Standard Process for Data Mining* (CRISP-DM). É ideal para Big Data e demais cenários que envolvam processos relacionados à análise de grandes volumes de dados. O CRISP-DM divide o processo de mineração de dados em seis fases principais. A sequência dessas fases não é rigorosa e se move iterativamente entre as fases, como for necessário. Dentro de cada fase no CRISP-DM existe uma estrutura hierárquica de atividades genéricas para serem realizadas, que podem determinar a execução de atividades específicas. As etapas são definidas a seguir:

Business Understanding: Entendimento do negócio

Etapa em que busca-se uma compreensão adequada do problema que necessita ser resolvido. São definidos os principais objetivos e expectativas em relação ao trabalho como um todo, e é feita a Avaliação das Circunstâncias para definir recursos ou dificuldades que podem influenciar o projeto.

Data Understanding: Compreensão dos dados

Tem o objetivo de inspecionar, organizar e descrever todos os dados disponíveis. Busca-se saber quais dados podem ser relevantes para decifrar o problema. É nesta etapa em que ocorrem: a Coleta inicial dos dados; a Descrição dos dados; a Análise exploratória dos dados; a Verificação de sua qualidade.

Data Preparation: Preparação dos dados

Nesta etapa o profissional deverá realizar a manipulação técnica dos dados, realizando a sua "filtragem" para torná-los bem definidos, organizados e bem inspecionados. É preciso preparar todas as databases, definir o formato e os atributos dos dados que serão trabalhados, e ajustar demais questões técnicas. É dividida nas atividades de: Seleção dos dados; Limpeza dos dados: Construção dos dados, ou seja, a criação de novas variáveis que auxiliam a análise; Integração dos dados; Formatação dos dados, para facilitar a análise.

Modeling: Modelagem

São selecionadas e aplicadas as técnicas de mineração de dados mais apropriadas, dependendo dos objetivos identificados na etapa de entendimento do negócio. É dividida em quatro atividades: Seleção das técnicas de modelagem; Testes de modelagem; Construção de modelo com base nos testes; Avaliação do modelo.

Evaluation: Avaliação do modelo

Trata-se do acompanhamento dos resultados objetivos e a avaliação da aplicabilidade confiável das informações e conhecimentos obtidos. Possui três tarefas: Avaliação dos resultados; Revisão do processo; Determinação dos etapas seguintes, se será necessário realizar a revisão de passos tomados.

Deployment: Desenvolvimento

É a aplicação prática de todo o conhecimento que foi obtido por meio do trabalho de mineração e modelagem a partir das análises dos dados realizadas. O conhecimento obtido deve ser apresentado de forma palpável e aplicável ao cliente.

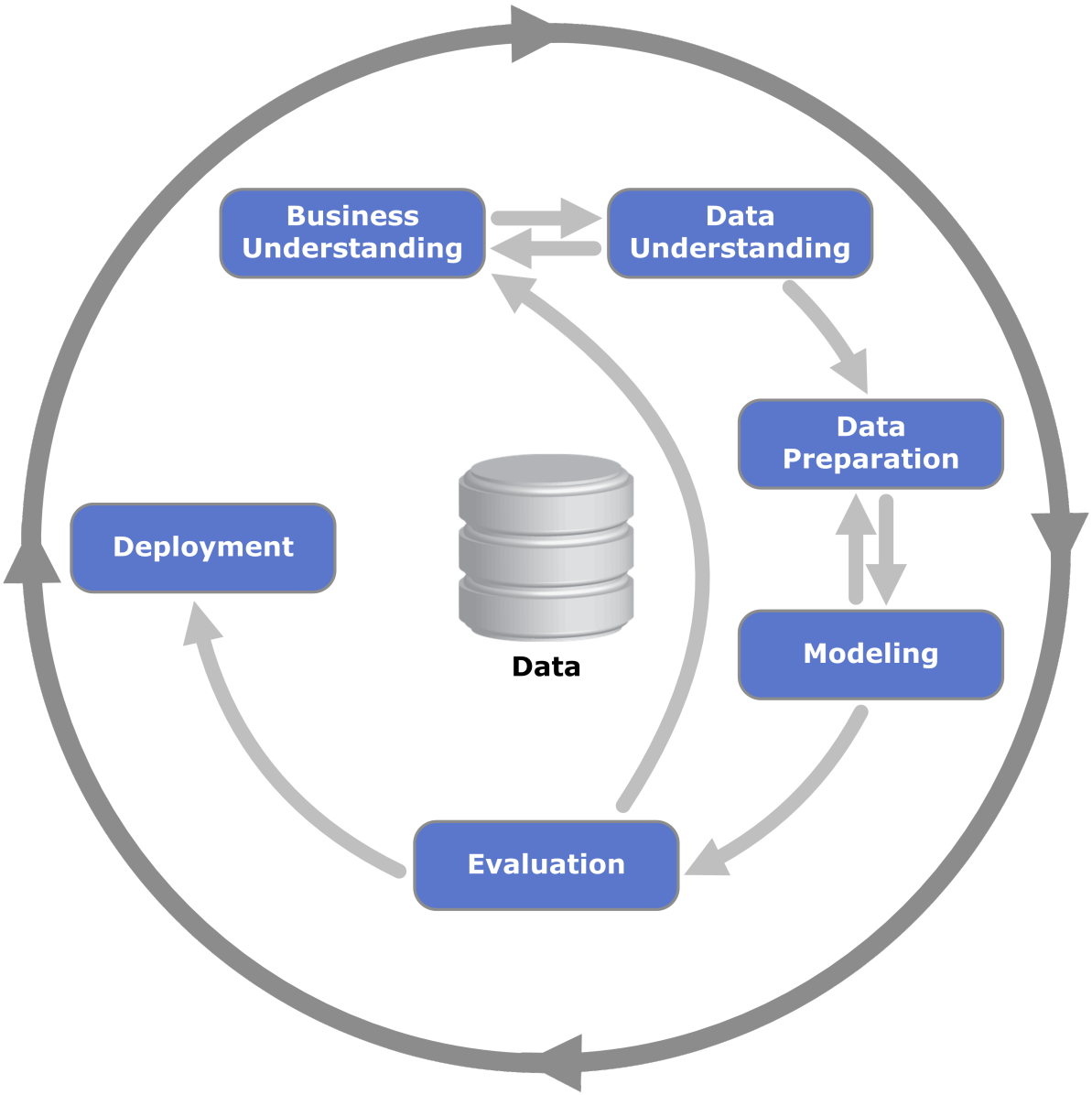


Figura I

Na *figura I* as setas mostradas na imagem do processo indicam as dependências mais importantes e ocorrentes entre as fases do processo. O círculo externo no diagrama simboliza a natureza cíclica da própria mineração de dados. Um processo de mineração de dados continua mesmo depois que uma solução foi implementada.

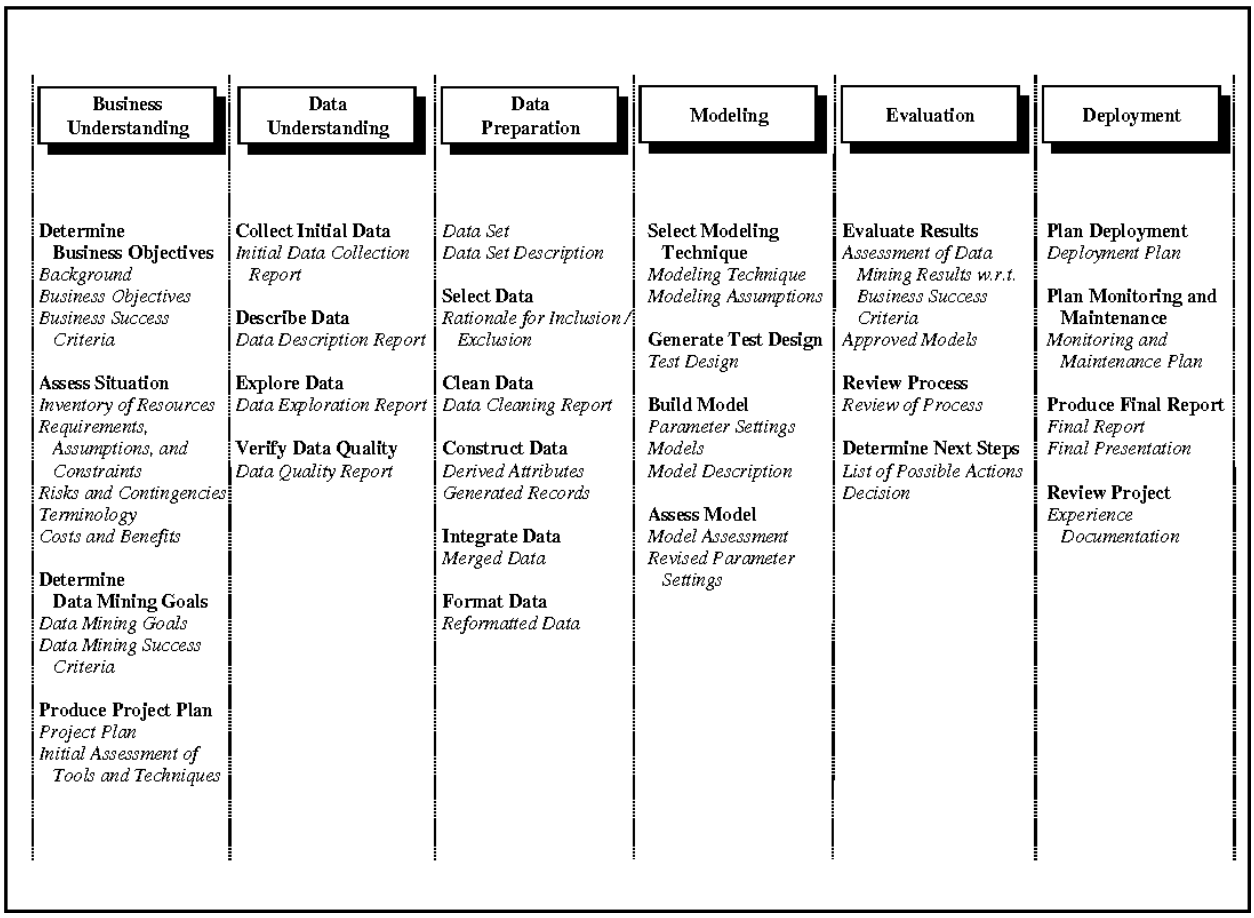


Figura II

Na *figura II* é esboçado as fases acompanhados das tarefas e saídas geradas.

Como surgiu o CRISP-DM?

O CRISP-DM foi idealizado no ano de 1996 e se tornou um projeto da União Europeia sob a iniciativa de financiamento *European Strategic Program on Research in Information Technology* (ESPRIT) no ano de 1997. O projeto foi liderado por cinco empresas: SPSS, Teradata, Daimler AG, NCR Corporation e OHRA. O modelo de trabalho nasceu a partir da iniciativa de profissionais que trabalhavam com data mining.

CRISP-DM Fase 1 - Entendimento do Negócio

O que é o Sistema Nacional de Pós-Graduação? (Contextualização)

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 atribuirem 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 constribui 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

CIÊNCIAS EXATAS E DA TERRA	ENGENHARIAS	MULTIDISCIPLINAR
Matemática / Probabilidade e Estatística	Engenharias IV	Interdisciplinar
Química	-	Materiais

Colégio de Humanidades

CIÊNCIAS HUMANAS	CIÊNCIAS SOCIAIS APLICADAS	LINGUÍSTICA, 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	-	-

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

O que é a UnB?

A Universidade de Brasília (UnB) é uma universidade pública federal brasileira. É a maior instituição de ensino superior do Centro-Oeste do Brasil e uma das mais importantes do país.

"Resultado do sonho e do trabalho de educadores como Darcy Ribeiro e Anísio Teixeira, a UnB é, desde 1962, ano de sua criação, uma das principais referências acadêmicas nacionais. A diversidade cultural presente em seus quatro campi é uma de suas características marcantes. A pluralidade, aliada à busca permanente por soluções inovadoras, move a produção científica e o cotidiano da instituição. A UnB segue atuante em todas as áreas do conhecimento, aberta às principais demandas do Brasil e do mundo." (www.unb.br/a-unb)

Geografia

A Geografia tem como objeto de estudo a superfície terrestre e seus fenômenos, bem como a relação humana com essa superfície. Dessa maneira, o geógrafo analisa a relação da população com a região que ocupa e os efeitos dessa ocupação. O Geógrafo pode seguir duas áreas com focos diferentes: a carreira acadêmica, como professores e pesquisadores; ou, uma carreira profissional, ligada a indústria.

O Programa de Pós-graduação em Geografia(PPGGEA) da Universidade de Brasília(UnB) foi criado em 1996 já com o curso de mestrado, e a partir de 2011 inicio o curso de Doutorado.

De acordo com o programa, seu principal central é o desenvolvimento de inovações científicas na área do conhecimento da Geografia e na formação de docentes, pesquisadores e recursos humanos especializados. O portal do programa pode ser acessado pela URL http://www.posgea.unb.br/.

O PPGGEA conta com diversos laboratórios nas subáreas de Geografia. São esses:

- Centro de Cartografia Aplicada e Informação Geográfica (CIGA): Conta estrutura física organizada com equipamentos básicos para o desenvolvimento de atividades direcionadas ao manuseio de ferramentas geográficas voltadas para a educação geográfica e o planejamento e gestão do território.
- Centro de Documentação Geográfica – Milton Santos (CDG): Espaço é destinado para os alunos de pós-graduação do GEA.

- Laboratório de Cartografia e Fotointerpretação: Laboratório equipado com mapoteca, fotografias aéreas em diversas escalas, retroprojetor, estereoscópicos de bolso e de espelho e mesas para trabalhos cartográficos. Consta ainda de um acervo de cartas sistemáticas de suporte aos projetos desenvolvidos no GEA por professores e estudantes (Graduação e Pós-Graduação)
- Laboratório de Geografia Física(LAGEF): LAGEF desenvolve pesquisas em Geomorfologia, pedologia, educação ambiental, ensino de Geografia, biogeografia e gestão ambiental. Apóia atividades de pesquisa com a participação dos alunos de graduação de pós-graduação. O laboratório abriga os projetos de pesquisa desenvolvidos pelos alunos do PIBIC/ Cnpq (Programa Institucional de Bolsas de Iniciação Científica/ DPP/UnB).
- Laboratório de Geoiconografia e Multimídias (LAGIM): Um dos focos prioritários do LAGIM é o de aproximar-se dos demais laboratórios e lhes propor ações conjuntas.
- Laboratório de Sistemas de Informações Espaciais (LSIE): O LSIE é um espaço programado para desenvolver atividades de pesquisa com alunos da Graduação e Pós-Graduação do Departamento de Geografia, assim como outros departamentos de áreas afins da Universidade. O LSIE desenvolve atividades de pesquisa e extensão no campo da geomática, com o propósito de integrar, adquirir e gerenciar dados e/ou informações espaciais. O laboratório trabalha em projetos com os diferentes órgãos e instituições que visam o desenvolvimento sustentado do meio ambiente.
- Laboratório de Análises Territoriais (LATER): Os principais objetivos do LATER são os seguintes: Organizar grupos de pesquisa temáticos para potencializar esforços e estreitar os laços entre docentes, alunos e participantes diversos; Fortalecer um espaço para debates em interação com a graduação; Organizar os projetos em grupos para discussões, trabalhos, pesquisas e orientações e torná-los mais produtivos. Os trabalhos desenvolvidos no laboratório congregam alunos de graduação, pós-graduação e professores, proporcionando suporte teórico e metodológico para a elaboração dos trabalhos científicos no âmbito territorial, cujos produtos foram publicações em revistas, seminários e eventos de vários tipos.
- Laboratório de Climatologia Geográfica (LCGea): O objetivo deste laboratório é ampliar e difundir os conhecimentos teóricos e práticos da Climatologia Geográfica Brasileira; ampliar os estudos de clima do Distrito Federal; estreitar laços com a sociedade civil e ampliar as relações profissionais visando a solução de problemas da comunidade; e- fornecer subsídios técnicos para o planejamento e a gestão do território.
- Laboratório de Geografia, Ambiente e Saúde (LAGAS): O LAGAS tem como objetivo principal proporcionar um espaço para discussões e apoio para elaboração e desenvolvimento de projetos de pesquisa e extensão que englobam uma abordagem integrada entre a análise geográfica das questões de saúde, meio ambiente e novas metodologias baseadas em geotecnologias, a fim de ampliar o conhecimento científico sobre esses temas e apoiar a formação de profissionais nessa área.

Segundo dados do portal do PPGGEA, constam 359 registros de teses e dissertações vinculados a esse programa de pós graduação.

Geociências Aplicadas e Geodinâmica

O Curso de Pós-graduação em GEOCIÊNCIAS APLICADAS, criado em 2008, tem como objetivo principal impulsionar o ensino e a pesquisa das geotecnologias que nas últimas décadas consagraram-se como ferramentas de coleta, organização, análise e processamento de dados e de informações nas diferentes aéreas das ciências da terra, a fim de demonstrar e desenvolver o potencial de suas aplicações às práticas e ao conhecimento geocientífico. O uso das geotecnologias na abordagem da informação geoespacial ou geográfica foi o processo que viabilizou a integração de um maior número de elementos e dados de diferentes origens, possibilitando maior materialidade na compreensão das relações entre o espaço geográfico e os componentes e constituintes terrestres. O Curso de Pós-graduação em GEOCIÊNCIAS APLICADAS oferece titulação nos níveis de Mestrado e Doutorado. O Curso está estruturado em 3 (três) áreas de concentração:

- Geofísica Aplicada

- Geoprocessamento e Análise Ambiental
- Hidrogeologia e Meio Ambiente

A duração do curso de mestrado é de no mínimo 2 e no máximo 4 semestres letivos, enquanto o de doutorado abrange o mínimo de 4 e o máximo de 8 semestres regulares.

Geologia

O Programa de Geologia da UnB - Universidade de Brasília é vinculado ao Instituto de Geociências da UnB (IG).

O IG ocupa área de mais de 6.550 m 2 no prédio do Instituto Central de Ciências (ICC). O IG possui ainda dois prédios próprios, onde estão instalados o Observatório Sismológico e o Laboratório de Geocronologia, com área adicional de 2.500 m 2 . No ICC funcionam a Direção e Secretaria, salas de aula, laboratórios, salas individuais de professores, salas para alunos de pós-graduação, o Museu de Geociências, os Centros Acadêmicos de Geologia Jorge Gushiken (CAGEO) e de Geofísica (CAGEF) e o Grupo Espeleológico da Geologia (GREGEO).

IG apresenta caráter multidisciplinar e procura integrar seus pesquisadores em trabalho de equipe, em torno de um objetivo comum, que é o de realizar pesquisa de alto nível e formar recursos humanos de qualidade.

O Instituto de Geociências da UnB conta com laboratórios em todas as subáreas das Geociências, aos quais os estudantes têm amplo acesso. Quase todos os laboratórios possuem equipamentos de última geração, a fim de atenderem adequadamente às atividades de ensino e pesquisa desenvolvidas no âmbito do Instituto, bem como ao intercâmbio com outras universidades, instituições de pesquisa e empresas do País e do exterior. Encontram-se em pleno funcionamento os laboratórios de laminação, preparação de amostras, separação de minerais pesados, computação, microscopia, inclusões fluidas, difratometria de raios X, microssonda eletrônica, geoquímica, isótopos estáveis, geocronologia, micropaleontologia, microscopia eletrônica de varredura, sensoriamento remoto e análise espacial, geofísica aplicada, laboratórios do Observatório Sismológico e de estudos da litosfera. O Laboratório de Geocronologia, por exemplo, possui excelente estrutura para análises geocronológicos e de geoquímica isotópica, que têm sido realizadas para a comunidade científica nacional e internacional, além de empresas públicas e privadas. Veículos apropriados para trabalhos de campo estão à disposição dos estudantes e professores.

Em decorrência da qualificação e dedicação de seu corpo docente, de seus técnicos e de seus estudantes e da qualidade de sua infraestrutura, todos os cursos oferecidos pelo Instituto de Geociências são bem avaliados pelo MEC e outros institutos de avaliação:

- Programa de Pós-Graduação em Geologia - 6;
- Programa de Pós-Graduação em Geociências Aplicadas - 4;
- Curso de Graduação em Geologia - 5;
- Curso de Graduação em Geofísica - 4;
- Curso de Graduação em Ciências Ambientais - 4.

O Museu de Geociências é um dos principais museus da UnB. Possui acervo expressivo e diversificado, que atende à visitação diária da comunidade da UnB, de grupos de estudantes do ensino fundamental e médio e da sociedade em geral.

O Observatório Sismológico (SIS) é um Centro do Instituto de Geociências (IG) da Universidade de Brasília (UnB). Suas responsabilidades envolvem três áreas, que são: o ensino (níveis de graduação e pós-graduação), a Extensão e a Pesquisa relacionada à sismicidade e à estrutura do interior da Terra. Sua principal atividade é o monitoramento sismográfico da sismicidade brasileira, natural e induzida por reservatórios. Dispõe de sede própria com boa infraestrutura física (laboratórios de Análise de Dados e de Computação; Oficina Eletrônica, Mostra de Sismologia, Biblioteca, entre outros), instrumental (sistemas sismográficos digitais completos, analisadores de espectro, gravímetro, geradores e medidores de rádio-freqüências, geradores de função, GPS's etc.), computacional (uma rede local com seis estações de trabalho, PC's e periféricos), e recur-

sos humanos (professores, técnicos-administrativos, prestadores de serviços e alunos bolsistas de graduação e pós-graduação).

Outros aspectos que caracterizam a produção científica e acadêmica da UnB

O que a Organização precisa realmente alcançar?

Vários stakeholders estão envolvidos neste projeto, e poderíamos considerar cada um deles como distintas organizações que possuem interesses distintos e complementares. Elas são:

- A Disciplina Ciência de Dados para Todos 2018.1, que quer comprovar que seus alunos dominam ferramentas e técnicas de ciência de dados, para fins de avaliação de rendimento da disciplina.
- A UnB, representada pelos decanatos de pós-graduação (DPG) e de pesquisa e inovação (DPI), que querem dispor de instrumentos para realização de avaliações contínuas de suas pós-graduações.
- O SNPG, que assim com o DPG e DPI, também pode se beneficiar do uso de instrumentos para realização de avaliações contínuas de suas pós-graduações.
- Os interessados em melhor conhecer o que é produzido pelo Sistema Nacional de Pós-graduação, como empresas privadas, que querem desfrutar dos benefícios gerados pela ciência brasileira.

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

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

Avaliação das Circunstâncias

Este trabalho irá demonstrar conhecimentos de Ciência de Dados aplicados aos datasets com informações dos cursos das áreas de Geografia, Geologia, e Geociências, de acordo com o que foi definido pela disciplina e seus professores.

Avaliação preliminar das pós-graduações na UnB

Geografia

Esse programa histórico do programa, tem atualmente 26 integrantes participantes no corpo docente. O corpo discente, que divido em mestrado e doutorado tem, respectivamente, 70 e 126 integrantes.

Esse programa é formado por uma área de concentração: Gestão Ambiental e territorial e duas linhas de pesquisa reestruturadas da seguinte forma:

1. Produção do Espaço e Território Nacional
2. Representação Espacial da Dinâmica Territorial e Ambiental

As linhas de pesquisa do programa estão de acordo com a formação acadêmica do corpo docente e visam atender aos objetivos estipulados, abrigando alunos que possuem interesse na qualificação em Geografia e outras áreas relacionadas como Arquitetura e Urbanismo, Sociologia, Planejamento Urbano e Regional, Ecologia, Geoprocessamento, Ensino de Geografia, entre outras.

As linhas desse programa pesquisa contemplam objetivos dos projetos desenvolvidos e professores e também os interesses de qualificação profissional dos discentes. De acordo com o portal, a proposta do curso possui como elemento estruturador fundamental os projetos de pesquisa que se articulam diretamente com a temática das disciplinas oferecidas e com os projetos de dissertação desenvolvidos pelos alunos. Isso permite que grupos de professores articulados por uma temática participem de projetos comuns, juntamente com seus grupos de orientandos, reunidos nos laboratórios.

Geociências Aplicadas e Geodinâmica

O programa de Geociências Aplicadas e Geodinâmica tem atualmente 11 pessoas em seu corpo docente, tendo iniciado em 2008. É dividido em 3 áreas de aplicação:

- Geofísica Aplicada
- Geoprocessamento e Análise Ambiental
- Hidrogeologia e Meio Ambiente

Tem um total de 28 Teses de Doutorado já publicadas e avaliadas, sendo a primeira datada de 30 de julho de 2012 e a mais recente de 8 de maio de 2017. Já as dissertações de Mestrado tem um total de 101 publicadas e avaliadas, iniciando em março de 2010 e indo até 18 de abril de 2017. Essas teses e dissertações são em diversas áreas dentro das 3 áreas de concentração do programa.

Geologia

O Programa de Geologia da UnB iniciou no ano de 2012, com 18 linhas de pesquisa, e atualmente possui sete áreas de concentração:

- Bioestratigrafia e Paleoecologia;
- Geofísica Aplicada;
- Geologia Regional;
- Geoquímica;
- Mineralogia e Petrologia;
- Processamento de Dados em Geologia e Análise Ambiental;
- Prospecção e Geologia Econômica.

De acordo com a Avaliação Trienal do programa, o corpo docente do Programa é bem qualificado, mesclando docentes experientes com mais de 10 anos de titulação máxima e jovens doutores com formação diversificada em vários centros nacionais e internacionais, o que permite cobrir com qualidade todas as áreas de pesquisa e formação em desenvolvimento.

O corpo discente que participa do Programa atinge, em média, 104 alunos por ano durante o triênio, sendo em sua maioria estudantes de mestrado. As teses e dissertações possuem qualidade inequívoca, o que é atestado pela publicação de seus resultados em periódicos de ampla circulação, inclusive em periódicos internacionais de nível elevado.

O Programa de Geologia da UnB é dos mais tradicionais e qualificados da área, constituindo um pólo de conhecimentos com forte inserção regional e nacional. A infra-estrutura analítica, em particular o Laboratório de Geologia Isotópica e o Laboratório de Microssonda Eletrônica vem servindo a várias instituições e centros de pesquisa nacionais. Vários docentes são colaboradores em outros programas da área, entre outros UFMT e UFC, atuando em cooperações do tipo Procad e Casadinho. Adicionalmente, o Programa coordena um projeto Pronex e um projeto MCT/CNPq/Instituto Nacional de Ciência e Tecnologia.

CRISP-DM Fase 2 - Entendimento dos Dados

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

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

Perfil profissional dos docentes vinculados às pós-graduações

```
json.geografia.perfil <- 'data/Geografia.profile.json'
file.info(json.geografia.perfil)
```

```
##                                size isdir mode               mtime
## data/Geografia.profile.json 1164871 FALSE  664 2018-07-08 10:06:21
##                                             ctime               atime  uid
## data/Geografia.profile.json 2018-07-08 14:01:27 2018-07-08 14:09:38 1000
##                               gid uname grname
## data/Geografia.profile.json 1000  luis   luis
```

O arquivo data/Geografia.profile.json apresenta dados sobre o perfil de todos os docentes vinculados a programas de pós-graduação de Geografia da UnB, entre 2010 e 2017.

```
json.geologia.perfil <- 'data/geologia.profile.json'
file.info(json.geologia.perfil)
```

```
##                                size isdir mode               mtime
## data/geologia.profile.json 1701579 FALSE  664 2018-07-08 19:09:09
##                                             ctime               atime  uid
## data/geologia.profile.json 2018-07-08 19:09:09 2018-07-08 19:09:11 1000
##                               gid uname grname
## data/geologia.profile.json 1000  luis   luis
```

O arquivo data/geologia.profile.json apresenta dados sobre o perfil de todos os docentes vinculados a programas de pós-graduação de Geologia da UnB, entre 2010 e 2017.

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

```
json.geografia.advise <- "data/Geografia.advise.json"
file.info(json.geografia.advise)
```

```
##                                size isdir mode               mtime
## data/Geografia.advise.json 541062 FALSE  664 2018-07-08 10:06:21
##                                             ctime               atime  uid
## data/Geografia.advise.json 2018-07-08 14:01:27 2018-07-08 14:09:38 1000
##                               gid uname grname
## data/Geografia.advise.json 1000  luis   luis
```

O arquivo data/Geografia.advise.json apresenta dados sobre o orientações de mestrado e doutorado feitas por todos os docentes vinculados a programas de pós-graduação de Geografia da UnB, entre 2010 e 2017.

```
json.geologia.advise <- "data/geologia.advise.json"
file.info(json.geologia.advise)
```

```
##                              size isdir mode               mtime
## data/geologia.advise.json 541431 FALSE  664 2018-07-08 19:09:09
##                                          ctime               atime  uid
## data/geologia.advise.json 2018-07-08 19:09:09 2018-07-08 19:09:11 1000
##                             gid uname grname
## data/geologia.advise.json 1000  luis   luis
```

O arquivo data/geologia.advise.json apresenta dados sobre o orientações de mestrado e doutorado feitas por todos os docentes vinculados a programas de pós-graduação de Geologia da UnB, entre 2010 e 2017.

Produção bibliográfica gerada pelos docentes vinculados às pós-graduações

```
json.geografia.producao.bibliografica <- "data/Geografia.publication.json"
file.info(json.geografia.producao.bibliografica)
```

```
##                                  size isdir mode               mtime
## data/Geografia.publication.json 413174 FALSE  664 2018-07-08 10:06:21
##                                                 ctime               atime
## data/Geografia.publication.json 2018-07-08 14:01:27 2018-07-08 14:09:38
##                                uid  gid uname grname
## data/Geografia.publication.json 1000 1000  luis   luis
```

O arquivo data/Geografia.publication.json apresenta dados sobre a produção bibliográfica gerada por todos os docentes vinculados a programas de pós-graduação de Geografia da UnB, entre 2010 e 2017.

```
json.geologia.producao.bibliografica <- "data/geologia.publication.json"
file.info(json.geologia.producao.bibliografica)
```

```
##                                 size isdir mode               mtime
## data/geologia.publication.json 866922 FALSE  664 2018-07-08 19:09:09
##                                                ctime               atime
## data/geologia.publication.json 2018-07-08 19:09:09 2018-07-08 19:09:11
##                               uid  gid uname grname
## data/geologia.publication.json 1000 1000  luis   luis
```

O arquivo data/geologia.publication.json apresenta dados sobre a produção bibliográfica gerada por todos os docentes vinculados a programas de pós-graduação de Geologia da UnB, entre 2010 e 2017.

Agrupamento dos docentes conforme áreas de atuação

```
json.geografia.researchers_by_area <- "data/Geografia.researchers_by_area.json"
file.info(json.geografia.researchers_by_area)
```

```
##                                       size isdir mode
## data/Geografia.researchers_by_area.json 1164 FALSE  664
##                                                       mtime
## data/Geografia.researchers_by_area.json 2018-07-08 10:06:21
##                                                       ctime
## data/Geografia.researchers_by_area.json 2018-07-08 14:01:27
##                                                       atime  uid  gid
## data/Geografia.researchers_by_area.json 2018-07-08 14:09:38 1000 1000
```

```
##                                         uname grname
## data/Geografia.researchers_by_area.json  luis   luis
```

O arquivo data/Geografia.researchers_by_area.json apresenta as vinculações de todos os docentes de Geografia que declararam atuar em cada uma das áreas de pós-graduação do Sistema Nacional de Pós-Graduação da CAPES, conforme apresenta-se registrada essa informação no currículo Lattes de cada um, em data recente

```
json.geologia.researchers_by_area <- "data/geologia.researchers_by_area.json"
file.info(json.geologia.researchers_by_area)
```

```
##                                        size isdir mode mtime ctime atime
## data/geologia.researchers_by_area.json   NA    NA <NA>  <NA>  <NA>  <NA>
##                                        uid gid uname grname
## data/geologia.researchers_by_area.json  NA  NA  <NA>   <NA>
```

O arquivo data/geologia.researchers_by_area.json apresenta as vinculações de todos os docentes de Geologia que declararam atuar em cada uma das áreas de pós-graduação do Sistema Nacional de Pós-Graduação da CAPES, conforme apresenta-se registrada essa informação no currículo Lattes de cada um, em data recente

Redes de colaboração entre docentes

```
file.info('data/Geografia.graph.json')
```

```
##                            size isdir mode               mtime
## data/Geografia.graph.json 4360 FALSE  664 2018-07-08 10:06:21
##                                     ctime               atime  uid
## data/Geografia.graph.json 2018-07-08 14:01:27 2018-07-08 14:09:38 1000
##                            gid uname grname
## data/Geografia.graph.json 1000  luis   luis
```

O arquivo `rdata/Geografia.graph.json` apresenta redes de colaboração na co-autoria de artigos científicos, feitas entre os docentes vinculados a programas de pós-graduação de Geografia da UnB, entre 2010 e 2017.

```
file.info('data/geologia.graph.json')
```

```
##                            size isdir mode               mtime
## data/geologia.graph.json 12391 FALSE  664 2018-07-08 19:09:09
##                                     ctime               atime  uid  gid
## data/geologia.graph.json 2018-07-08 19:09:09 2018-07-08 19:09:11 1000 1000
##                           uname grname
## data/geologia.graph.json  luis   luis
```

O arquivo `rdata/geologia.graph.json` apresenta redes de colaboração na co-autoria de artigos científicos, feitas entre os docentes vinculados a programas de pós-graduação de Geologia da UnB, entre 2010 e 2017.

CRISP-DM Fase.Atividade 2.2 - Descrição dos Dados

Descrição dos dados do perfil

O arquivo Geografia.profile.json, contém dados que caracterizam o perfil profissional de cada um dos docentes do grupo sob análise. Esse arquivo pode ser lido através do seguinte comando, da biblioteca *jsonlite*.

```
unb.geografia.prof <- fromJSON("data/Geografia.profile.json")
```

A quantidade de docentes de Geografia sob análise é apresentada a seguir.

```
length(unb.geografia.prof)
```

```
## [1] 23
```

Através do comando *names()* pode ser obtido uma ideia sobre as principais informações que podem ser obtidas de cada docente. Como pode-se observar abaixo essas informações são divididas em sete áreas principais.

```
names(unb.geografia.prof[[1]])
```

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

Ainda com o intuito de obter uma análise inicial dos dados contidos nos perfis dos docentes, pode-se usar a função *glimpse()*, que apresenta os atributos típicos que podem ser obtidos relativamente a um pesquisador específico.

```
glimpse(unb.geografia.prof[[1]], width = 30)
```

```
## List of 7
##  $ nome                    : chr "Helen da Costa Gurgel"
##  $ resumo_cv          : chr "Possui graduação em Geografia pela Universidade Federal Fluminense (1996), mes
##  $ areas_de_atuacao       :'data.frame':  6 obs. of  4 variables:
##   ..$ grande_area  : chr [1:6] "CIENCIAS_EXATAS_E_DA_TERRA" "CIENCIAS_EXATAS_E_DA_TERRA" "CIENCIAS_HUMAN
##   ..$ area         : chr [1:6] "Geociências" "Geociências" "Geografia" "Saúde Coletiva" ...
##   ..$ sub_area     : chr [1:6] "Geofísica" "Geografia Física" "Geografia Humana" "Saúde Pública" ...
##   ..$ especialidade: chr [1:6] "Sensoriamento Remoto" "Geoprocessamento" "Geografia da Saúde" "" ...
##  $ endereco_profissional :List of 8
##   ..$ instituicao: chr "Universidade de Brasília"
##   ..$ orgao      : chr "Departamento de Geografia"
##   ..$ unidade    : chr ""
##   ..$ DDD        : chr "61"
##   ..$ telefone   : chr "33072373"
##   ..$ bairro     : chr "Asa Norte"
##   ..$ cep        : chr "70910900"
##   ..$ cidade     : chr "Brasília"
##  $ producao_bibiografica :List of 4
##   ..$ CAPITULO_DE_LIVRO:'data.frame':    11 obs. of  13 variables:
##   .. ..$ tipo              : chr [1:11] "Capítulo de livro publicado" "Capítulo de livro publicado" "Capítu
##   .. ..$ titulo_do_capitulo    : chr [1:11] "Unidades de Conservação e Desenvolvimento: A Contribuição do
##   .. ..$ titulo_do_livro      : chr [1:11] "Dez anos do Sistema Nacional de Unidades de Conservação da Natu
##   .. ..$ ano                      : chr [1:11] "2011" "2011" "2011" "2011" ...
##   .. ..$ doi                      : chr [1:11] "" "" "" "" ...
##   .. ..$ pais_de_publicacao       : chr [1:11] "Brasil" "Brasil" "Brasil" "Brasil" ...
##   .. ..$ isbn              : chr [1:11] "9788577381456" "9788533417779" "9788533417779" "9788533417779" .
##   .. ..$ nome_da_editora      : chr [1:11] "MMA" "Ministério da Saúde" "Ministério da Saúde" "Ministério d
##   .. ..$ numero_da_edicao_revisao: chr [1:11] "1" "1" "1" "1" ...
##   .. ..$ organizadores        : chr [1:11] "Rodrigo Medeiros; Fábio França Silva Araújo" "Carlos Machado d
##   .. ..$ paginas                  : chr [1:11] "55 - 88" "25 - 52" "53 - 71" "73 - 86" ...
##   .. ..$ autores                  :List of 11
##   .. ..$ autores-endogeno         :List of 11
##   ..$ EVENTO           :'data.frame':    52 obs. of  11 variables:
##   .. ..$ natureza         : chr [1:52] "RESUMO" "COMPLETO" "COMPLETO" "RESUMO" ...
##   .. ..$ titulo       : chr [1:52] "Presença de Flebotomíneos (Diptera: Psychodidae) transmissores da Leis
```

```
##  . . ..$ nome_do_evento  : chr [1:52] "XXVIII Congresso Brasileiro de Zoologia" "XVI Simpósio Brasileiro d
##   .. ..$ ano_do_trabalho : chr [1:52] "2010" "2013" "2013" "2013" ...
##   .. ..$ pais_do_evento  : chr [1:52] "Brasil" "Brasil" "Brasil" "Brasil" ...
##  . . ..$ cidade_do_evento: chr [1:52] "Belém - PA" "Foz do Iguaçu ? PR" "São Luiz - MA" "Alto Paraiso de Goi
##   .. ..$ doi             : chr [1:52] "" "" "" "" ...
##   .. ..$ classificacao   : chr [1:52] "NACIONAL" "NACIONAL" "NACIONAL" "LOCAL" ...
##   .. ..$ paginas         : chr [1:52] " - " "6238 - 6245" " - " " - " ...
##   .. ..$ autores         :List of 52
##   .. ..$ autores-endogeno:List of 52
##   ..$ LIVRO            :'data.frame':    1 obs. of  13 variables:
##   .. ..$ titulo               : chr "Anais do 7º Simpósio Nacional de Geografia da Saúde"
##   .. ..$ ano                  : chr "2015"
##   .. ..$ tipo                 : chr "LIVRO_ORGANIZADO_OU_EDICAO"
##   .. ..$ natureza             : chr "ANAIS"
##   .. ..$ pais_de_publicacao   : chr "Brasil"
##   .. ..$ isbn                 : chr "19805829"
##   .. ..$ doi                  : chr ""
##   .. ..$ nome_da_editora      : chr "Universidade de Brasília"
##   .. ..$ numero_da_edicao_revisao: chr "1"
##   .. ..$ numero_de_paginas    : chr "1312"
##   .. ..$ numero_de_volumes    : chr "1"
##   .. ..$ autores              :List of 1
##   .. ..$ autores-endogeno     :List of 1
##   ..$ PERIODICO        :'data.frame':    12 obs. of  10 variables:
##   .. ..$ natureza        : chr [1:12] "COMPLETO" "COMPLETO" "COMPLETO" "COMPLETO" ...
##  . . ..$ titulo       : chr [1:12] "Spatial clustering and longitudinal variation of Anopheles darlingi (D
##  . . ..$ periodico     : chr [1:12] "Bulletin of Entomological Research" "Confins (Paris)" "Malaria Journa
##   .. ..$ ano             : chr [1:12] "2011" "2011" "2013" "2013" ...
##   .. ..$ volume          : chr [1:12] "-" "13" "12" "" ...
##   .. ..$ issn            : chr [1:12] "00074853" "19589212" "14752875" "16790944" ...
##   .. ..$ paginas         : chr [1:12] "1 - 16" "1 - " "192 - " "15 - " ...
##  . . ..$ doi         : chr [1:12] "10.1017/S0007485311000265" "10.4000/confins.7348" "10.1186/1475-2875-
##   .. ..$ autores         :List of 12
##   .. ..$ autores-endogeno:List of 12
##  $ orientacoes_academicas:List of 5
##   ..$ ORIENTACAO_CONCLUIDA_DOUTORADO   :'data.frame':    2 obs. of  13 variables:
##   .. ..$ natureza                  : chr [1:2] "Tese de doutorado" "Tese de doutorado"
##  . . ..$ titulo              : chr [1:2] "Estudo dos padrões espaço-temporais de ocorrência da diarreia no
##   .. ..$ ano                       : chr [1:2] "2015" "2016"
##   .. ..$ id_lattes_aluno           : chr [1:2] "2856454630796379" "8261731624731487"
##   .. ..$ nome_aluno             : chr [1:2] "Marcus Andre Fuckner" "MIssifany Silveira"
##  . . ..$ instituicao          : chr [1:2] "Universidade de Brasília" "Universidade de Brasília"
##   .. ..$ curso                     : chr [1:2] "Geografia" "Geografia"
##   .. ..$ codigo_do_curso           : chr [1:2] "51500434" "51500434"
##   .. ..$ bolsa                     : chr [1:2] "NAO" "NAO"
##   .. ..$ agencia_financiadora      : chr [1:2] "" ""
##   .. ..$ codigo_agencia_financiadora: chr [1:2] "" ""
##   .. ..$ nome_orientadores         :List of 2
##   .. ..$ id_lattes_orientadores    :List of 2
##   ..$ ORIENTACAO_CONCLUIDA_MESTRADO    :'data.frame':    4 obs. of  13 variables:
##  . . ..$ natureza             : chr [1:4] "Dissertação de mestrado" "Dissertação de mestrado" "Dissertaçã
##  . . ..$ titulo              : chr [1:4] "Análise Temporal do Uso e Cobertura da Terra na Bacia Hidrográfi
##   .. ..$ ano                       : chr [1:4] "2016" "2016" "2016" "2016"
##  . . ..$ id_lattes_aluno        : chr [1:4] "5045240633932615" "3413533328523209" "4048463184776211" "09
```

```
##  .. ..$ nome_aluno            : chr [1:4] "Lucas Garcia Magalhães Peres" "Alexandre Sauma da Silva" "Naya
##  .. ..$ instituicao           : chr [1:4] "Universidade de Brasília" "Universidade de Brasília" "Univers
##  .. ..$ curso                 : chr [1:4] "Geografia" "Geografia" "Geografia" "Geografia"
##  .. ..$ codigo_do_curso          : chr [1:4] "51500434" "51500434" "51500434" "51500434"
##  .. ..$ bolsa                       : chr [1:4] "SIM" "NAO" "SIM" "NAO"
##  .. ..$ agencia_financiadora     : chr [1:4] "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
##  .. ..$ codigo_agencia_financiadora: chr [1:4] "045000000000" "" "045000000000" ""
##  .. ..$ nome_orientadores          :List of 4
##  .. ..$ id_lattes_orientadores     :List of 4
##  ..$ ORIENTACAO_EM_ANDAMENTO_DOUTORADO:'data.frame':    5 obs. of  13 variables:
##  .. ..$ natureza              : chr [1:5] "Tese de doutorado" "Tese de doutorado" "Tese de doutorado" "Tes
##  .. ..$ titulo               : chr [1:5] "O desempenho do sistema único de saúde: uma análise geográfica d
##  .. ..$ ano                         : chr [1:5] "2014" "2014" "2016" "2016" ...
##  .. ..$ id_lattes_aluno        : chr [1:5] "" "3852290153832016" "4424334504086302" "2241554336609585"
##  .. ..$ nome_aluno            : chr [1:5] "Daniel Alvão de Carvalho Júnior" "Leandro da Silva Gregório" "Y
##  .. ..$ instituicao           : chr [1:5] "Universidade de Brasília" "Universidade de Brasília" "Univers
##  .. ..$ curso                : chr [1:5] "Geografia" "Geografia" "Geografia" "Geografia" ...
##  .. ..$ codigo_do_curso         : chr [1:5] "51500434" "51500434" "51500434" "51500434" ...
##  .. ..$ bolsa                       : chr [1:5] "NAO" "NAO" "NAO" "NAO" ...
##  .. ..$ agencia_financiadora        : chr [1:5] "" "" "" "" ...
##  .. ..$ codigo_agencia_financiadora: chr [1:5] "" "" "" "" ...
##  .. ..$ nome_orientadores          :List of 5
##  .. ..$ id_lattes_orientadores     :List of 5
##  ..$ ORIENTACAO_EM_ANDAMENTO_MESTRADO :'data.frame':    2 obs. of  13 variables:
##  .. ..$ natureza              : chr [1:2] "Dissertação de mestrado" "Dissertação de mestrado"
##  .. ..$ titulo               : chr [1:2] "Análise dos aspectos geoambientais e da distribuição espacial da
##  .. ..$ ano                         : chr [1:2] "2017" "2017"
##  .. ..$ id_lattes_aluno             : chr [1:2] "" "5691395606608035"
##  .. ..$ nome_aluno            : chr [1:2] "Eraldo Jair Gonçalves Dias" "Amarílis Bahia Bezerra"
##  .. ..$ instituicao           : chr [1:2] "Universidade de Brasília" "Universidade de Brasília"
##  .. ..$ curso                       : chr [1:2] "Geografia" "Geografia"
##  .. ..$ codigo_do_curso             : chr [1:2] "51500434" "51500434"
##  .. ..$ bolsa                       : chr [1:2] "NAO" "NAO"
##  .. ..$ agencia_financiadora        : chr [1:2] "" ""
##  .. ..$ codigo_agencia_financiadora: chr [1:2] "" ""
##  .. ..$ nome_orientadores          :List of 2
##  .. ..$ id_lattes_orientadores     :List of 2
##  ..$ OUTRAS_ORIENTACOES_CONCLUIDAS    :'data.frame':    46 obs. of  13 variables:
##  .. ..$ natureza              : chr [1:46] "TRABALHO_DE_CONCLUSAO_DE_CURSO_GRADUACAO" "TRABALHO_DE_CONCL
##  .. ..$ titulo               : chr [1:46] "Lixo Urbano e Qualidade de Vida: Um Problema em Araguapaz" "Impa
##  .. ..$ ano                         : chr [1:46] "2012" "2012" "2012" "2012" ...
##  .. ..$ id_lattes_aluno             : chr [1:46] "" "" "" "" ...
##  .. ..$ nome_aluno            : chr [1:46] "Silvânia Borges de Oliveira da Mata" "Wellen Cintia B. dos San
##  .. ..$ instituicao           : chr [1:46] "Universidade de Brasília" "Universidade de Brasília" "Univers
##  .. ..$ curso                : chr [1:46] "Geografia" "Geografia" "Geografia" "Geografia" ...
##  .. ..$ codigo_do_curso         : chr [1:46] "60108886" "60108886" "60108886" "60108886" ...
##  .. ..$ bolsa                       : chr [1:46] "NAO" "NAO" "NAO" "NAO" ...
##  .. ..$ agencia_financiadora        : chr [1:46] "" "" "" "" ...
##  .. ..$ codigo_agencia_financiadora: chr [1:46] "" "" "" "" ...
##  .. ..$ nome_orientadores          :List of 46
##  .. ..$ id_lattes_orientadores     :List of 46
##  $ senioridade           : chr "8"
```

Observando os atributos do pesquisador em análise, pode-se perceber que:

- É uma pesquisadora que formada em Geografia.
- Trabalha no departamento de Geografia.
- Possui um capítulo de livro publicado no ano de 2011.
- Possui cinco orientação de doutorado em andamento, uma destas iniciada em 2014.
- Possui senioridade 8.

Potencial de utilização dos dados do perfil dos docentes

Descrição dos dados de orientações

Assim como no *dataset* de perfis, a leitura será feita através da biblioteca jsonlite.

```
unb.geografia.adv <- fromJSON("data/Geografia.advise.json")
```

Para maior entendimento dos dados contidos nesse *dataset* serão utilizados os seguintes métodos.

```
names(unb.geografia.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(unb.geografia.adv$ORIENTACAO_CONCLUIDA_DOUTORADO)
```

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

Percebe-se que as orientações se subdividem em quatro principais áreas e, dentre cada ma delas, variam entre os anos de 2010 até 2017.

Explorando os dados de orientações concluídas de mestrado e doutorados, no ano de 2017, percebe-se que existem 14 e 17 orientações, respectivamente. Dentre as de mestrado apenas uma não era do curso de geografia. E dentre as de doutorado, apenas duas não são do curso de geografia.

```
length(unb.geografia.adv$ORIENTACAO_CONCLUIDA_DOUTORADO$`2017`$natureza)
```

```
## [1] 17
```

```
sort(table(unb.geografia.adv$ORIENTACAO_CONCLUIDA_DOUTORADO$`2017`$curso), decreasing = TRUE)
```

```
##
##                                          Geografia
##                                                 15
## Desenvolvimento Sociedade e Cooperação Internacional
##                                                  1
##                               DOUTORADO EM GEOGRAFIA
##                                                  1
```

```
length(unb.geografia.adv$ORIENTACAO_CONCLUIDA_MESTRADO$`2017`$natureza)
```

```
## [1] 14
```

```
sort(table(unb.geografia.adv$ORIENTACAO_CONCLUIDA_MESTRADO$`2017`$curso), decreasing = TRUE)
```

```
##
##                                                          Geografia
##                                                                 13
## Pós-Graduação Desenvolvimento Sociedade e Cooperação Internacional
##                                                                  1
```

Descrição dos dados de produção bibliográfica

Fazendo a leitura do arquivos Geografia.publication.json, e para entender sobre o conteúdo dos dados presentes dentro dele, utiliza-se a função *names* e, a partir dela , observa-se que as publicações bibliográficas estão divididas em sete áreas principais. Representadas abaixo:

```
unb.geografia.pub <- fromJSON("data/Geografia.publication.json")
names(unb.geografia.pub)
```

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

Afim de entender mais sobre os dados presentes em alguns das áreas principais, foi explorados os dados de periodicos e livros no ano de 2017. Pode-se observar abaixo a quantidade de registros, os atributos que contém em cada área, os periodicos e o país de publicação dos livros.

```
length(unb.geografia.pub$PERIODICO$`2017`)
```

```
## [1] 10
```

```
names(unb.geografia.pub$PERIODICO$`2017`)
```

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

```
sort(table(unb.geografia.pub$PERIODICO$`2017`$periodico), decreasing = TRUE)
```

```
##
##                                                 Confins (Paris)
##                                                               2
##                                                 CONFINS (PARIS)
##                                                               2
##       GEOINGÁ: REVISTA DO PROGRAMA DE PÓS-GRADUAÇÃO EM GEOGRAFIA
##                                                               2
##         Hygeia.Revista Brasileira de Geografia Médica e da Saúde
##                                                               2
##                   ISPRS International Journal of Geo-Information
##                                                               2
##                           REVISTA BRASILEIRA DE GEOGRAFIA FÍSICA
##                                                               2
##                                      Agri-environmental Sciences
```

```
##                                                                1
##                                                BRASÍLIA EM DEBATE
##                                                                1
##              BUILDING THE WAY - REVISTA DO CURSO DE LETRAS DA UEG
##                                                                1
##                                                          Climate
##                                                                1
##                                             CUADERNOS DE GEOGRAFÌA
##                                                                1
##                                                   CYBERGEO (PARIS)
##                                                                1
##                                                             Data
##                                                                1
##                                         Estudos Geográficos (UNESP)
##                                                                1
##                                           Finisterra (Lisboa. 1966)
##                                                                1
##                                                         GEODERMA
##                                                                1
##                                                   GEOGRAPHIA (UFF)
##                                                                1
##                      GEOSABERES REVISTA DE ESTUDOS GEOEDUCACIONAIS
##                                                                1
##                                                GEOTEXTOS (ONLINE)
##                                                                1
##                                             GEOUSP: espaço e tempo
##                                                                1
##                             INTERNATIONAL JOURNAL OF BIOMETEOROLOGY
##                                                                1
##                                                 LAND USE POLICY
##                                                                1
## PESQUISAR - REVISTA DE ESTUDOS E PESQUISAS EM ENSINO DE GEOGRAFIA
##                                                                1
##                                       Regional Environmental Change
##                                                                1
##                                    Revista brasileira de climatologia
##                                                                1
##                                  REVISTA BRASILEIRA DE GEOMORFOLOGIA
##                                                                1
##                      REVISTA BRASILEIRA DE PLANEJAMENTO E ORÇAMENTO
##                                                                1
##                                                 REVISTA CENÁRIO
##                                                                1
##                                       REVISTA CERRADOS (UNIMONTES)
##                                                                1
##                                     Revista de Ensino de Geografia
##                                                                1
##                                   Revista Espaço e Geografia (UnB)
##                                                                1
##                                               Revista Geoaraguaia
##                                                                1
##                                     Revista Querubim (Online)
##                                                                1
##                        Revista Ra'e Ga Espaço Geográfico em Análise
```

```
##                                                          1
##                                          SER SOCIAL (UNB)
##                                                          1
##                  SOJ Microbiology & Infectious Diseases
##                                                          1
##                               TERR@ PLURAL (UEPG. ONLINE)
##                                                          1
```

```
length(unb.geografia.pub$LIVRO$`2017`)
```

```
## [1] 13
```

```
names(unb.geografia.pub$LIVRO$`2017`)
```

```
##  [1] "titulo"                   "ano"
##  [3] "tipo"                     "natureza"
##  [5] "pais_de_publicacao"       "isbn"
##  [7] "doi"                      "nome_da_editora"
##  [9] "numero_da_edicao_revisao" "numero_de_paginas"
## [11] "numero_de_volumes"        "autores"
## [13] "autores-endogeno"
```

```
sort(table(unb.geografia.pub$LIVRO$`2017`$pais_de_publicacao), decreasing = TRUE)
```

```
## Brasil
##      4
```

Descrição dos dados de agregação de docentes por área

Fazendo a leitura do arquivo Geografia.researchers_by_area.json, que é referente a vinculação dos docentes que atuam nas áreas de pós-graduação do Sistema Nacional de Pós-Graduação da CAPES, e entendendo os atributos existentes dentro desse *dataset* através do método *names()* podemos criar uma tabela relacionando a área de atuação e o número de docentes vinculados. O resultado pode ser observado abaixo:

```
unb.geografia.area <- fromJSON("data/Geografia.researchers_by_area.json")
names(unb.geografia.area)
```

```
## [1] "Areas dos pesquisadores"
```

```
unb.geografia.area.df <- cbind(names(unb.geografia.area$`Areas dos pesquisadores`),
          (sapply(unb.geografia.area$`Areas dos pesquisadores`, function(x) length(x))))
rownames(unb.geografia.area.df) <- c(1:nrow(unb.geografia.area.df)); colnames(unb.geografia.area.df) <-
glimpse(unb.geografia.area.df)
```

```
##  chr [1:12, 1:2] "Agronomia" "Arquitetura e Urbanismo" ...
##  - attr(*, "dimnames")=List of 2
##   ..$ : chr [1:12] "1" "2" "3" "4" ...
##   ..$ : chr [1:2] "Area" "Professores"
```

```
unb.geografia.area.df[]
```

```
##    Area                          Professores
## 1  "Agronomia"                   "1"
## 2  "Arquitetura e Urbanismo"     "2"
## 3  "Ciências Ambientais"         "2"
## 4  "Ecologia"                    "1"
## 5  "Educação"                    "4"
## 6  "Filosofia"                   "1"
```

```
## 7  "Geociências"                    "9"
## 8  "Geografia"                      "15"
## 9  "História"                       "1"
## 10 "Planejamento Urbano e Regional" "6"
## 11 "Saúde Coletiva"                 "1"
## 12 "Turismo"                        "2"
```

Descrição dos dados de redes de colaboração

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

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. Abaixo serão feitas as análises relacionadas a cada *dataset* utilizado nesse projeto.

Arquivo Profile

Agora trabalharemos novamente com o *dataset* de profile para conseguir compreensão mais a fundo desses dados. Primeiramente explorando a quantidade de áreas de atuação da geografia.

```
sum(sapply(unb.geografia.prof, function(x) nrow(x$areas_de_atuacao)))
```

```
## [1] 98
```

Em seguida, explorando a quantidade de pessoas por área de atuação, grande área e pessoas que produziram os específicos tipos de produção

```
table(unlist(sapply(unb.geografia.prof, function(x) nrow(x$areas_de_atuacao))))
```

```
##
## 1 2 3 4 5 6
## 1 1 7 3 4 7
```

```
table(unlist(sapply(unb.geografia.prof, function(x) (x$areas_de_atuacao$grande_area))))
```

```
##
##          CIENCIAS_AGRARIAS        CIENCIAS_BIOLOGICAS
##                          2                          1
##          CIENCIAS_DA_SAUDE CIENCIAS_EXATAS_E_DA_TERRA
##                          2                         29
##           CIENCIAS_HUMANAS CIENCIAS_SOCIAIS_APLICADAS
##                         50                         11
##                     OUTROS
##                          3
```

```
table(unlist(sapply(unb.geografia.prof, function(x) names(x$producao_bibiografica))))
```

```
##
##                          ARTIGO_ACEITO
##                                      3
##                      CAPITULO_DE_LIVRO
##                                     21
## DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA
##                                      7
##                                 EVENTO
##                                     23
```

```
##                                   LIVRO
##                                      14
##                                PERIODICO
##                                      23
##                          TEXTO_EM_JORNAIS
##                                       8
```

Percebe-se que a maior parte dos docentes atuam na área de Ciência Humanas, que é o contexto da Geografia, e maior parte das suas produções bibliográficas estão nas áreas de **Periódico**, **Evento** e **Capítulo de Livro**, respectivamente.

Em seguida, exploramos o número de publicações de cada tipo seguinte:

- Artigo aceito
- Capítulo de livro
- Periodico
- Texto em jornais

```
sum(sapply(unb.geografia.prof, function(x) length(x$producao_bibiografica$ARTIGO_ACEITO$ano)))
```

```
## [1] 4
```

```
sum(sapply(unb.geografia.prof, function(x) length(x$producao_bibiografica$CAPITULO_DE_LIVRO$ano)))
```

```
## [1] 132
```

```
sum(sapply(unb.geografia.prof, function(x) length(x$producao_bibiografica$LIVRO$ano)))
```

```
## [1] 44
```

```
sum(sapply(unb.geografia.prof, function(x) length(x$producao_bibiografica$PERIODICO$ano)))
```

```
## [1] 501
```

```
sum(sapply(unb.geografia.prof, function(x) length(x$producao_bibiografica$TEXTO_EM_JORNAIS$ano)))
```

```
## [1] 41
```

E também o número de pessoas por quantitativo de produções por pessoa:

```
table(unlist(sapply(unb.geografia.prof, function(x) length(x$producao_bibiografica$ARTIGO_ACEITO$ano))))
```

```
##
##  0  1  2
## 20  2  1
```

```
table(unlist(sapply(unb.geografia.prof, function(x) length(x$producao_bibiografica$CAPITULO_DE_LIVRO$ano
```

```
##
##  0  1  2  3  4  5  6  7  8 11 13 16 20
##  2  3  3  1  1  3  2  3  1  1  1  1  1
```

```
table(unlist(sapply(unb.geografia.prof, function(x) length(x$producao_bibiografica$LIVRO$ano))))
```

```
##
##  0  1  2  3  4  6 10
##  9  5  3  1  2  2  1
```

```
table(unlist(sapply(unb.geografia.prof, function(x) length(x$producao_bibiografica$PERIODICO$ano))))
```

```
##
##  1  5  6  7  8  9 11 12 20 22 23 24 28 47 63 66 86
```

```
##  1  2  2  2  1  2  1  2  1  1  1  2  1  1  1  1  1
```

```
table(unlist(sapply(unb.geografia.prof, function(x) length(x$producao_bibiografica$TEXTO_EM_JORNAIS$ano)
```

```
##
##  0  1  2  3  5 12 15
## 15  2  2  1  1  1  1
```

Novamente é possível observar a predominância nas quantidades referentes a Perídodicos e Capítulo de Livros quando comparado aos demais.

Agora, exploraremos a quantidade referente ao número de produções feitas por ano de cada um dos tipos supracitados

```
table(unlist(sapply(unb.geografia.prof, function(x) (x$producao_bibiografica$ARTIGO_ACEITO$ano))))
```

```
##
## 2012 2014 2017
##    2    1    1
```

```
table(unlist(sapply(unb.geografia.prof, function(x) (x$producao_bibiografica$CAPITULO_DE_LIVRO$ano))))
```

```
##
## 2010 2011 2012 2013 2014 2015 2016 2017
##   14   22   19   23   16   10    9   19
```

```
table(unlist(sapply(unb.geografia.prof, function(x) (x$producao_bibiografica$LIVRO$ano))))
```

```
##
## 2010 2011 2012 2013 2014 2015 2016 2017
##    7    5    4    5    8    5    5    5
```

```
table(unlist(sapply(unb.geografia.prof, function(x) (x$producao_bibiografica$PERIODICO$ano))))
```

```
##
## 2010 2011 2012 2013 2014 2015 2016 2017
##   81   51   83   78   66   43   47   52
```

```
table(unlist(sapply(unb.geografia.prof, function(x) (x$producao_bibiografica$TEXTO_EM_JORNAIS$ano))))
```

```
##
## 2010 2011 2012 2013 2014 2015 2016 2017
##   19   12    1    1    3    3    1    1
```

De maneira geral, percebe-se que o número de produções diminuiu drasticamente de 2010 até 2017, principalmente de Textos em jornais e Periódicos.

Agora, vamos abordar o número de pessoas que realizaram diferentes tipo de orientações, para os seguintes tipos:

- Orientação concluída de mestrado
- Orientação concluída de doutorado
- Orientação concluída de pós-doutorado

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

```
## [1] 122
```

```
table(unlist(sapply(unb.geografia.prof, function(x) names(x$orientacoes_academicas))))
```

```
##
##               ORIENTACAO_CONCLUIDA_DOUTORADO
```

```
##                                         19
##                 ORIENTACAO_CONCLUIDA_MESTRADO
##                                         22
##            ORIENTACAO_CONCLUIDA_POS_DOUTORADO
##                                          2
##             ORIENTACAO_EM_ANDAMENTO_DOUTORADO
##                                         22
##             ORIENTACAO_EM_ANDAMENTO_GRADUACAO
##                                          7
## ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA
##                                          9
##              ORIENTACAO_EM_ANDAMENTO_MESTRADO
##                                         19
##                 OUTRAS_ORIENTACOES_CONCLUIDAS
##                                         22
```

```
sum(sapply(unb.geografia.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTRADO
```

```
## [1] 181
```

```
sum(sapply(unb.geografia.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTORAD
```

```
## [1] 58
```

```
sum(sapply(unb.geografia.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_POS_DOUT
```

```
## [1] 3
```

Em seguida, a número de pessoas por quantitativo de orientações por pessoa:

```
table(unlist(sapply(unb.geografia.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA
```

```
##
##  0  1  2  3  4  5  7  8  9 10 11 12 13 14 31
##  1  2  1  3  2  1  2  1  1  2  2  2  1  1  1
```

```
table(unlist(sapply(unb.geografia.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA
```

```
##
## 0 1 2 3 4 5 6
## 4 5 4 3 1 4 2
```

```
table(unlist(sapply(unb.geografia.prof, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA
```

```
##
##  0  1  2
## 21  1  1
```

Por fim, a quantidade de orientações a cada ano:

```
table(unlist(sapply(unb.geografia.prof, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTR
```

```
##
## 2010 2011 2012 2013 2014 2015 2016 2017
##   16   20   23   18   23   41   26   14
```

```
table(unlist(sapply(unb.geografia.prof, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTO
```

```
##
## 2011 2012 2013 2014 2015 2016 2017
##    1    1    7    2   11   19   17
```

```
table(unlist(sapply(unb.geografia.prof, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_POS_D
```

```
##
## 2013 2014
##    1    2
```

Em contrapartida a quantidade de produção bibliográfica, o número de orientações a cada ano, de modo geral, cresceu consideravelmente de 2010 até 2017.

Arquivo Publicação

Agora criamos um data frame com a junção dos dados das publicações ao longo dos anos. Podemos observar que o *data frame* final contém 10 variáveis e um total de 388 observações, o que indica que temos 10 campos possíveis para cada publicação, desde natureza e título até número de páginas e autores, e que temos um total de 388 publicações ao longo dos anos presentes no *dataset*.

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

```
## Observations: 338
## Variables: 10
## $ natureza           <chr> "COMPLETO", "COMPLETO", "COMPLETO", "COMPLE...
## $ titulo             <chr> "Caracterização de uma topossequência no Pa...
## $ periodico          <chr> "Boletim de Pesquisa e Desenvolvimento (Emb...
## $ ano                <chr> "2010", "2010", "2010", "2010", "2010", "20...
## $ volume             <chr> "277", "01", "01", "01", "01", "01", "01", ...
## $ issn               <chr> "1676918X", "21774366", "21774366", "217743...
## $ paginas            <chr> "5 - 68", "01 - 25", "36 - ", "01 - 11", "0...
## $ doi                <chr> "", "", "", "", "", "", "", "", "", "", "",...
## $ autores            <list> [<"SOUZA, V. V.", "PASSO, D. P.", "GOMES, ...
## $ `autores-endogeno` <list> ["1886939214378140", "3466441462870689", "...
```

Por fim, limpamos o *data frame* de quaisquer listas que possa a ter, para assim facilitar a análise a ser feita posteriormente.

```
unb.geografia.pub.df$autores <- gsub("\",\"|\", \"", "; ", unb.geografia.pub.df$autores)
unb.geografia.pub.df$autores <- gsub("\"|c\\(|\\)", "", unb.geografia.pub.df$autores)
unb.geografia.pub.df$`autores-endogeno` <- gsub(",", ";", unb.geografia.pub.df$`autores-endogeno`)
unb.geografia.pub.df$`autores-endogeno` <- gsub("\"|c\\(|\\)", "", unb.geografia.pub.df$`autores-endoge
glimpse(unb.geografia.pub.df)
```

```
## Observations: 338
## Variables: 10
## $ natureza           <chr> "COMPLETO", "COMPLETO", "COMPLETO", "COMPLE...
## $ titulo             <chr> "Caracterização de uma topossequência no Pa...
## $ periodico          <chr> "Boletim de Pesquisa e Desenvolvimento (Emb...
## $ ano                <chr> "2010", "2010", "2010", "2010", "2010", "20...
## $ volume             <chr> "277", "01", "01", "01", "01", "01", "01", ...
## $ issn               <chr> "1676918X", "21774366", "21774366", "217743...
## $ paginas            <chr> "5 - 68", "01 - 25", "36 - ", "01 - 11", "0...
## $ doi                <chr> "", "", "", "", "", "", "", "", "", "", "",...
## $ autores            <chr> "SOUZA, V. V.; PASSO, D. P.; GOMES, R. A. T...
## $ `autores-endogeno` <chr> "1886939214378140", "3466441462870689", "34...
```

Arquivo Orientação

Agora para o arquivo de Orientações, primeiro agrupamos os dados a partir das orientações de Pós-Doutorado, Doutorado e Mestrado, assim focando apenas no programa de Pós-Graduação. assim, criamos um *data frame* para os tipos de orientações que desejamos e outro para as orientações em si. Assim, temos o *data frame* *unb.geografia.adv.df* cujo tem 13 campos para descrever as orientações e um total de 226. Já para o *data frame* *unb.geografia.adv.tipo.df* temos apenas um total de 168 orientações, ou seja, das 226 orientações, apenas 168 são do programa de Pós-Graduação.

```
unb.geografia.adv.tipo.df <- data.frame()
unb.geografia.adv.df <- data.frame()
for (i in 1:length(unb.geografia.adv[[1]]))
    unb.geografia.adv.tipo.df <- rbind(unb.geografia.adv.tipo.df,unb.geografia.adv$ORIENTACAO_CONCLUIDA
unb.geografia.adv.df <- rbind(unb.geografia.adv.df, unb.geografia.adv.tipo.df); unb.geografia.adv.tipo.
for (i in 1:length(unb.geografia.adv[[1]]))
    unb.geografia.adv.tipo.df <- rbind(unb.geografia.adv.tipo.df,unb.geografia.adv$ORIENTACAO_CONCLUIDA
unb.geografia.adv.df <- rbind(unb.geografia.adv.df, unb.geografia.adv.tipo.df); unb.geografia.adv.tipo.
for (i in 1:length(unb.geografia.adv[[1]]))
    unb.geografia.adv.tipo.df <- rbind(unb.geografia.adv.tipo.df,unb.geografia.adv$ORIENTACAO_CONCLUIDA
unb.geografia.adv.df <- rbind(unb.geografia.adv.df, unb.geografia.adv.tipo.df)
glimpse(unb.geografia.adv.df)
```

```
## Observations: 226
## Variables: 13
## $ natureza                    <chr> "Supervisão de pós-doutorado", "Su...
## $ titulo                      <chr> "", "", "Genética de paisagem de R...
## $ ano                         <chr> "2013", "2014", "2014", "2011", "2...
## $ id_lattes_aluno             <chr> "", "", "", "5330376452257907", "1...
## $ nome_aluno                  <chr> "Potira Meirelles Hermuche", "Fagn...
## $ instituicao                 <chr> "Universidade de Brasília", "Unive...
## $ curso                       <chr> "", "", "", "Geografia", "Geociênc...
## $ codigo_do_curso             <chr> "", "", "", "90000022", "60050870"...
## $ bolsa                       <chr> "SIM", "SIM", "NAO", "NAO", "NAO",...
## $ agencia_financiadora        <chr> "Coordenação de Aperfeiçoamento de...
## $ codigo_agencia_financiadora <chr> "045000000000", "002200000000", ""...
## $ nome_orientadores           <list> ["Renato Fontes Guimarães", "Neio...
## $ id_lattes_orientadores      <list> ["7063856452054362", "11753324401...
```

```
glimpse(unb.geografia.adv.tipo.df)
```

```
## Observations: 168
## Variables: 13
## $ natureza                    <chr> "Dissertação de mestrado", "Disser...
## $ titulo                      <chr> "Inclusão Digital e Exclusão Socia...
## $ ano                         <chr> "2010", "2010", "2010", "2010", "2...
## $ id_lattes_aluno             <chr> "5666083832593501", "8595273216773...
## $ nome_aluno                  <chr> "Vevila Rezende Costa", "Frederico...
## $ instituicao                 <chr> "Universidade de Brasília", "Unive...
## $ curso                       <chr> "Geografia", "Geografia", "Geograf...
## $ codigo_do_curso             <chr> "51500434", "51500434", "51500434"...
## $ bolsa                       <chr> "NAO", "SIM", "SIM", "SIM", "SIM",...
## $ agencia_financiadora        <chr> "", "Coordenação de Aperfeiçoament...
## $ codigo_agencia_financiadora <chr> "", "045000000000", "045000000000"...
## $ nome_orientadores           <list> ["Neio Lucio de Oliveira Campos",...
## $ id_lattes_orientadores      <list> ["1175332440156178", "11753324401...
```

Em seguida removemos as possíveis listas nos dados e em seguida separamos os orientadores pelo seu ID Lattes:

```
unb.geografia.adv.df$nome_orientadores <- gsub("\"|c\\(|\\)", "", unb.geografia.adv.df$nome_orientadores
unb.geografia.adv.df$id_lattes_orientadores <- gsub("\"|c\\(|\\)", "", unb.geografia.adv.df$id_lattes_or
unb.geografia.adv.df <- separate(unb.geografia.adv.df, nome_orientadores, into = c("ori1", "ori2"), sep
```

```
## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 210 rows [1,
## 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 18, 19, 20, 21, 22, ...].
```

```
unb.geografia.adv.df <- separate(unb.geografia.adv.df, id_lattes_orientadores, into = c("idLattes1", "id
```

```
## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 210 rows [1,
## 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 18, 19, 20, 21, 22, ...].
```

Depois, vemos quantas orientações ocorreram por ano no formato de uma tabela:

```
table(unb.geografia.adv.df$ano)
```

```
##
## 2010 2011 2012 2013 2014 2015 2016 2017
##   15   18   21   26   27   46   42   31
```

Para no fim observarmos os 20 docentes que mais orientaram ao longo dos anos em ordem decrescente:

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

```
##
##   Osmar Abílio de Carvalho Junior     Neio Lucio de Oliveira Campos
##                                29                                18
##     Fernando Luiz Araújo Sobrinho              Marilia Steinberger
##                                17                                16
##         Everaldo Batista da Costa    Roberto Arnaldo Trancoso Gomes
##                                15                                13
##            Valdir Adilson Steinke           Lucia Cony Faria Cidade
##                                13                                12
##              Marilia Luiza Peluso           Renato Fontes Guimarães
##                                11                                11
##       Ruth Elias de Paula Laranja            Ercília Torres Steinke
##                                11                                 9
##        Mario Diniz de Araujo Neto               Nelba Azevedo Penna
##                                 8                                 8
##   Osmar Abílio de Carvalho Junior       Rogério Elias Soares Uagoda
##                                 7                                 7
##        Cristina Maria Costa Leite Dante Flávio da Costa Reis Júnior
##                                 6                                 6
##             Helen da Costa Gurgel    Rafael Sanzio Araújo dos Anjos
##                                 6                                 6
```

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

Conforme visto nas seções anteriores, os dados dos *datasets* após serem trabalhados, permitem análises e entendimento da situação do programa de Pós-Graduação. Portanto, os dados referentes ao Departamento de Geografia e do Instituto de Geociências contém conteúdo de qualidade, ou seja, os dados disponíveis estão completos, ou pelo menos com uma completude suficiente para estudo, possibilitando a realização do trabalho proposto e da geração de novas informações.

Assim, nas próximas seções será demonstrado com mais detalhes o trabalho feito em cima dos dados para obter uma análise com maior detalhe e de melhor qualidade até a apresentada aqui.

CRISP-DM Fase 3 - Preparação dos Dados

CRISP-DM Fase.Atividade 3.1 - Seleção dos dados

Para a preparação dos dados foram selecionados dados para alguns *data frames*. A seguir está os respectivos *data frames* e seus dados:

- Perfil: ID Lattes, resumo do CV, nome, endereço profissional, instituição, orgão, unidade, DDD, telefone, bairro, CEP, cidade, senioridade.
- Produção: ID Lattes, produção bibliográfica, tipo de produção.
- Orientação: ID Lattes, orientações acadêmicas, orientação.
- Áreas de atuação: ID Lattes, áreas de atuação.

Como pode ser visto, é a partir do ID Lattes que os docentes serão identificados através dos *data frames*, assim todos estes contém esse campo. Quanto aos outros dados, no caso do *data frame* Perfil, são os dados mínimo para identificação e perfil de cada professor. Para os *data frames* Produção, Orientação e Áreas de Atuação, são os dados mais relevantes para de cada tipo, para entender como está o funcionamento e desenvolvimento dos programas de Pós-Graduação aqui analisados.

CRISP-DM Fase.Atividade 3.2 - Limpeza dos dados

A limpeza dos dados já está contida em conjunto com a construção dos mesmos na seção abaixo.

CRISP-DM Fase.Atividade 3.3 - Construção dos dados

Criação de funções para facilitar o trabalho de construção de dados, segundo o modelo: "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".

```
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)
}

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_producao2dataframe<- function( lista_producao ) {
  df_saida <- NULL

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

  df_saida <- cv_tplista2tpchar(df_saida)
  return(df_saida)


}


concatenadf <- function( df1, df2) {
  for( coluna in names(df1 ) ) {
    if( !is.element(coluna, names(df2) )) {
      df2[coluna] <- NA
    }
  }

  for( coluna in names(df2 ) ) {

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

  df_final <- rbind(df1 , df2)
  return(df_final)

}

extrai_1perfil <- 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_1perfil <- data.frame( idLattes , nome, resumo_cv ,instituicao ,
                            orgao, unidade , DDD, telefone, bairro,cep,cidade , senioridade,
                            stringsAsFactors = FALSE)

  return(df_1perfil)
}

extrai_perfis <- function(jsonProfessores) {
  df_saida <- data.frame()
  for( i in 1:length(jsonProfessores)) {
    jsonProfessor <- jsonProfessores[i]
    df_professor <- extrai_1perfil(jsonProfessor)
    if( nrow(df_saida) > 0 ) {
      df_saida <- rbind(df_saida , df_professor)
    } else {
      df_saida <- df_professor
    }
  }

  return(df_saida)
}

extrai_1producao <- function(professor) {
  idLattes <- names(professor)
  df_1producao <<- NULL
  producao_bibliografica <- professor[[idLattes]]$producao_bibiografica  #list
  for( tipo_producao in names(producao_bibliografica)) {
    df_temporario <- cv_tplista2tpchar ( producao_bibliografica[[tipo_producao]])
    df_temporario$tipo_producao <-  tipo_producao
    df_temporario$idLattes <-  idLattes
    df_1producao <- concatenadf( df_1producao , df_temporario  )
  }
  return(df_1producao)
}

extrai_producoes <- function( jsonProfessores) {
  df_saida <- data.frame()
  for( i in 1:length(jsonProfessores)) {
    jsonProfessor <- jsonProfessores[i]
    df_producao <- extrai_1producao(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)
}

extrai_1orientacao <- function(professor) {
  idLattes <- names(professor)
  df_1orientacao <- 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_1orientacao <- concatenadf( df_1orientacao , df_temporario  )
  }
  return(df_1orientacao)
}

extrai_orientacoes <- function(jsonProfessores) {
  df_saida <- data.frame()
  for( i in 1:length(jsonProfessores)) {
    jsonProfessor <- jsonProfessores[i]
    df_orientacao <- extrai_1orientacao(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)
}

extrai_1area_de_atuacao <- function(professor){
  idLattes <- names(professor)
  df_1area <-  professor[[idLattes]]$areas_de_atuacao
  df_1area$idLattes <- idLattes
  return(df_1area)
}

extrai_areas_atuacao <- function(jsonProfessores){
  df_saida <- data.frame()
  for( i in 1:length(jsonProfessores)) {
    jsonProfessor <- jsonProfessores[i]
    df_area_atuacao <- extrai_1area_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)
}
```

CRISP-DM Fase.Atividade 3.4 - Integração dos dados

Em seguida é feito a criação do arquivo para análise, ainda tendo como base o modelo.

```
unb.geografia.prof.json <- read_file("data/Geografia.profile.json")
unb.prof.df.capes <- read.csv("data/PesqPosCapes.csv",
                              sep = ";", header = TRUE, colClasses = "character")
unb.geografia.prof.json <- fromJSON(unb.geografia.prof.json)

unb.geografia.prof.df.professores <- extrai_perfis(unb.geografia.prof)

unb.geografia.prof.df.publicacoes <- extrai_producoes(unb.geografia.prof)

unb.geografia.prof.df.orientacoes <- extrai_orientacoes(unb.geografia.prof)

unb.geografia.prof.df.areas.de.atuacao <- extrai_areas_atuacao(unb.geografia.prof)

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

unb.geografia.prof.df <- data.frame()
unb.geografia.prof.df <- unb.geografia.prof.df.professores %>%
  select(idLattes, nome, resumo_cv, senioridade) %>%
  left_join(
    unb.geografia.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(
    unb.geografia.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(
    unb.geografia.prof.df.areas.de.atuacao %>%
      select(area, idLattes) %>%
      group_by(idLattes) %>%
      summarise(n_distinct(area)),
    by = "idLattes") %>%
  left_join(
    unb.prof.df.capes %>%
      select(AreaPos, idLattes) %>%
      group_by(idLattes) %>%
      summarise(n_distinct(AreaPos)),
    by = "idLattes")
```

Por fim, pode-ser observar a construção dos dados trabalhados, por meio das funções *head* e *glimpse*:

```
glimpse(unb.geografia.prof.df)
```

```
## Observations: 23
## Variables: 16
## $ idLattes                                <chr> "0975018553829295", "11...
## $ nome                                    <chr> "Helen da Costa Gurgel"...
## $ resumo_cv                               <chr> "Possui graduação em Ge...
## $ senioridade                             <chr> "8", "9", "9", "9", "9"...
## $ ORIENTACAO_CONCLUIDA_DOUTORADO          <int> 2, 3, 1, 3, NA, 5, 1, 2...
## $ ORIENTACAO_CONCLUIDA_MESTRADO           <int> 4, 14, 12, 3, 1, 4, 5, ...
## $ ORIENTACAO_CONCLUIDA_POS_DOUTORADO      <int> NA, 1, NA, NA, NA, NA, ...
## $ OUTRAS_ORIENTACOES_CONCLUIDAS           <int> 46, 5, 19, 4, 8, 50, 33...
## $ CAPITULO_DE_LIVRO                       <int> 11, 2, 1, 3, 13, 5, 4, ...
## $ DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA  <int> NA, 3, NA, NA, 6, NA, N...
## $ EVENTO                                  <int> 52, 3, 8, 1, 1, 32, 9, ...
## $ LIVRO                                   <int> 1, NA, NA, 6, 2, 4, 1, ...
## $ PERIODICO                               <int> 12, 6, 63, 24, 12, 24, ...
## $ TEXTO_EM_JORNAIS                        <int> NA, 2, NA, 15, NA, NA, ...
## $ `n_distinct(area)`                      <int> 4, 3, 1, 3, 1, 2, 3, 1,...
## $ `n_distinct(AreaPos)`                   <int> 1, 2, 1, 1, 1, 1, 1, 1,...
```

```
head(unb.geografia.prof.df, 1)
```

```
##           idLattes                  nome
## 1 0975018553829295 Helen da Costa Gurgel
##
## 1 Possui graduação em Geografia pela Universidade Federal Fluminense (1996), mestrado em Sensoriamento Re
##   senioridade ORIENTACAO_CONCLUIDA_DOUTORADO ORIENTACAO_CONCLUIDA_MESTRADO
## 1           8                              2                             4
##   ORIENTACAO_CONCLUIDA_POS_DOUTORADO OUTRAS_ORIENTACOES_CONCLUIDAS
## 1                                 NA                           46
##   CAPITULO_DE_LIVRO DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA EVENTO LIVRO
## 1                11                                     NA     52     1
##   PERIODICO TEXTO_EM_JORNAIS n_distinct(area) n_distinct(AreaPos)
## 1        12              NA                4                   1
```

Análise dos dados

Senioridade

Para análise da senioridade dos docentes, será feita uma correlação entre senioridade e o número de orientações concluídas.

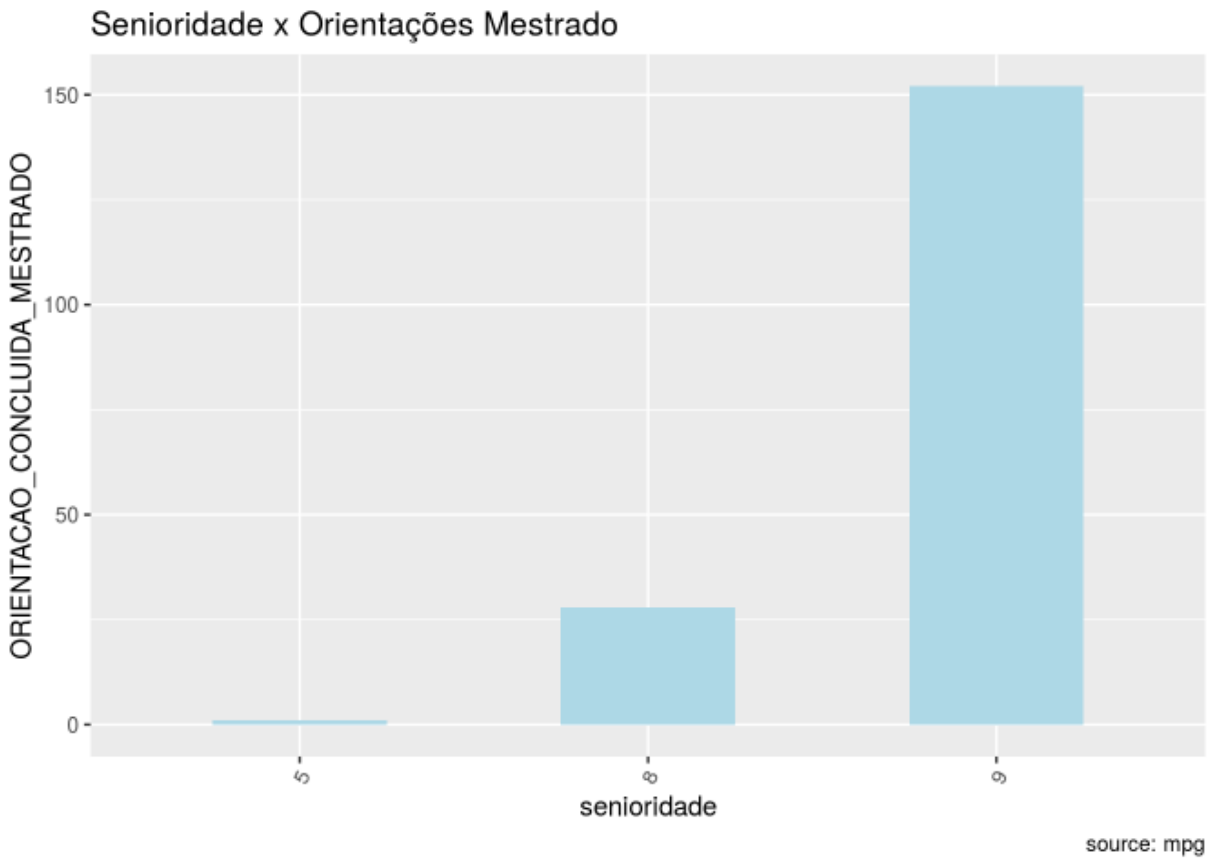
```
plot1 <- ggplot(unb.geografia.prof.df, aes(x=senioridade, y=ORIENTACAO_CONCLUIDA_MESTRADO)) +
  geom_bar(stat="identity", width=.5, fill="#ADD8E6") +
  labs(title="Senioridade x Orientações Mestrado",
       caption="source: mpg") +
  theme(axis.text.x = element_text(angle=65, vjust=0.6))

plot2 <- ggplot(unb.geografia.prof.df, aes(x=senioridade, y=ORIENTACAO_CONCLUIDA_DOUTORADO)) +
  geom_bar(stat="identity", width=.5, fill="#abf069") +
  labs(title="Senioridade x Orientações Doutorado",
       caption="source: mpg") +
  theme(axis.text.x = element_text(angle=65, vjust=0.6))
```

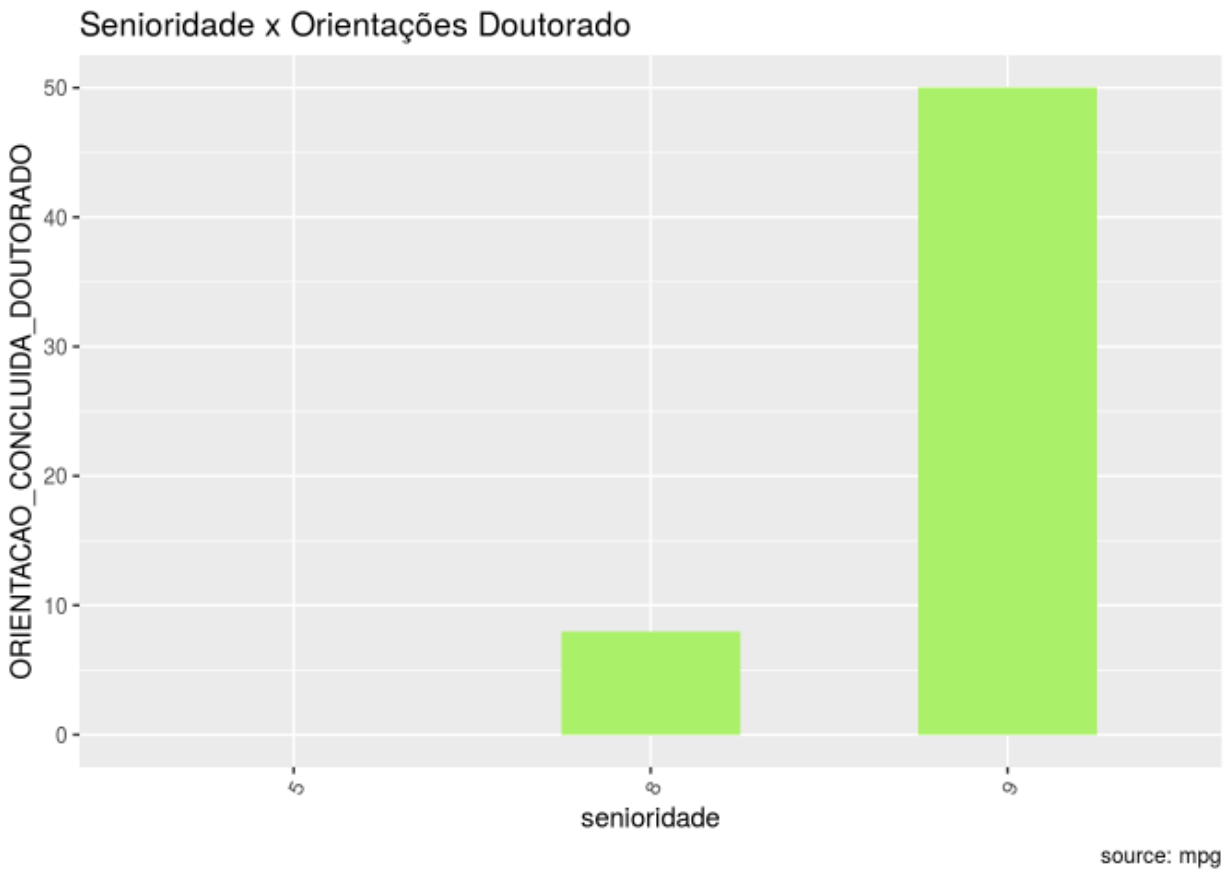
```
plot3 <- ggplot(unb.geografia.prof.df, aes(x=senioridade, y=ORIENTACAO_CONCLUIDA_POS_DOUTORADO)) +
  geom_bar(stat="identity", width=.5, fill="#ba0e42") +
  labs(title="Senioridade x Orientações Pós-Doutorado",
       caption="source: mpg") +
  theme(axis.text.x = element_text(angle=65, vjust=0.6))

ggarrange(plot1, plot2, plot3, ncol=1)
```

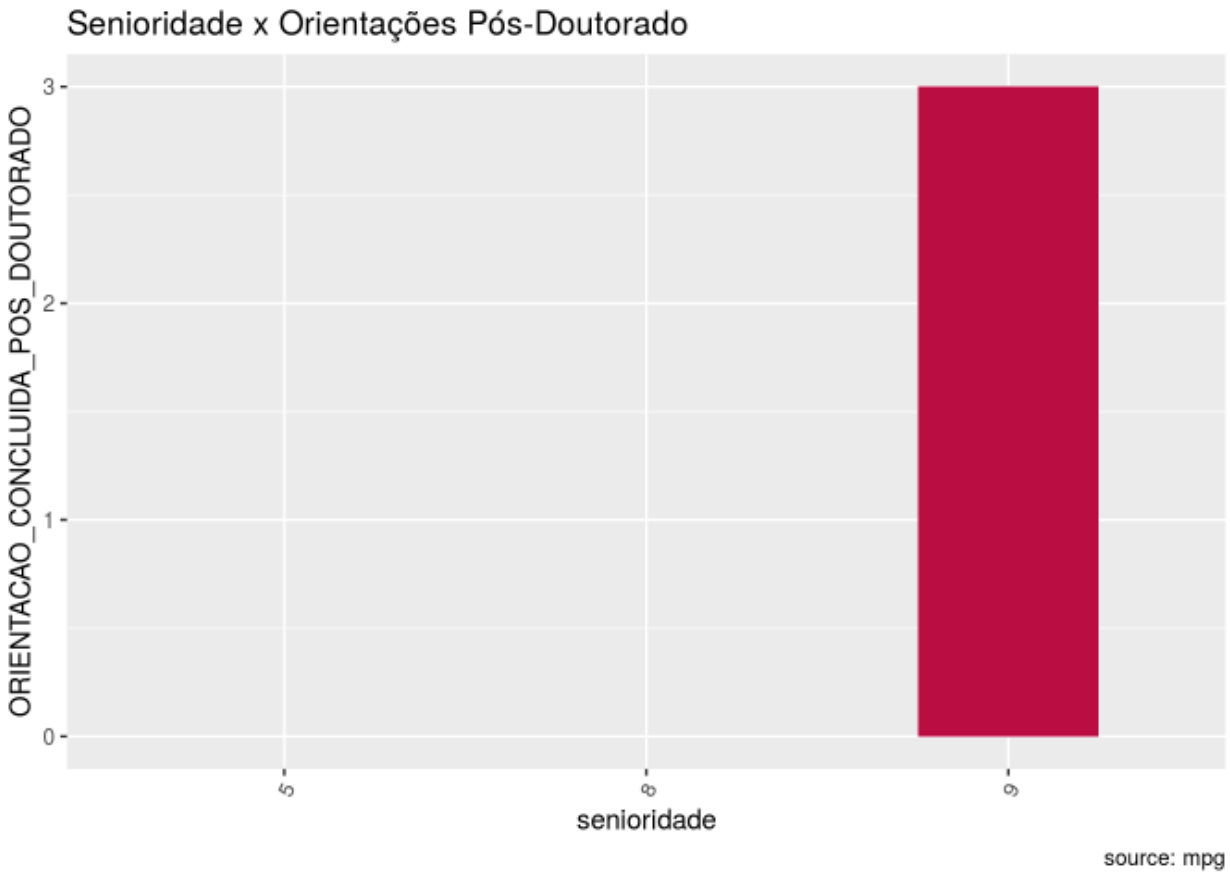
```
## $`1`
```




```
##
## $`2`
```


Senioridade x Orientações Doutorado
ORIENTACAO_CONCLUIDA_DOUTORADO
50
40
30
20
10
0
5
8
9
senioridade
source: mpg

```
##
## $`3`
```




```
##
## attr(,"class")
## [1] "list"      "ggarrange"
```

Percebe-se que existe uma relação diretamente proporcional entre a senioridade do docente e o número de orientações concluídas, visto que em todos os casos quanto maior a senioridade maior o número de orientações.

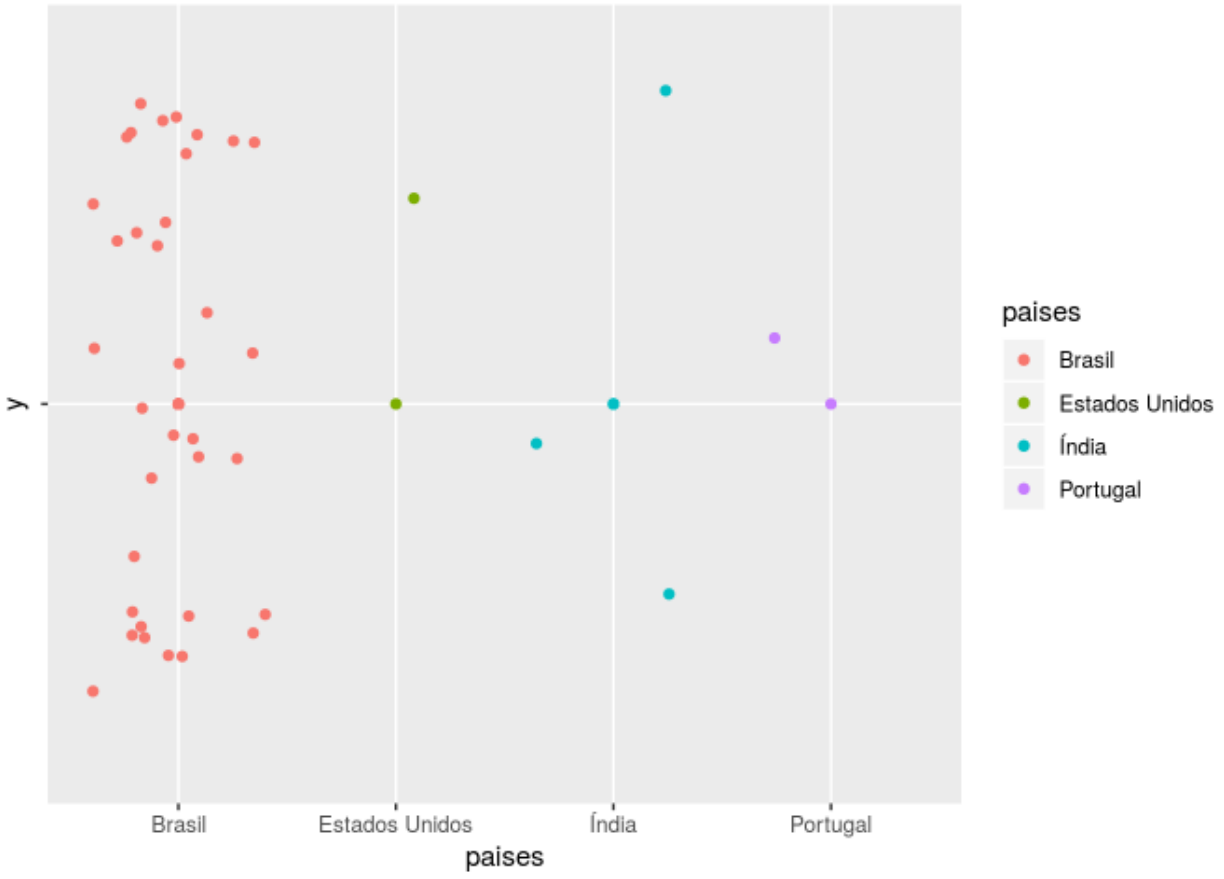
Nível de internacionalização

Para análise do nível de internacionalização das publicações referentes ao programa de pós graduação de geografia da UnB, foi gerado o seguinte gráfico, utilizando a biblioteca ggplot2.

```
eventos2017 <- unb.geografia.pub$EVENTO$`2017`
names(eventos2017)
```

```
##  [1] "natureza"         "titulo"           "nome_do_evento"
##  [4] "ano_do_trabalho"  "pais_do_evento"   "cidade_do_evento"
##  [7] "doi"              "classificacao"    "paginas"
## [10] "autores"          "autores-endogeno"
```

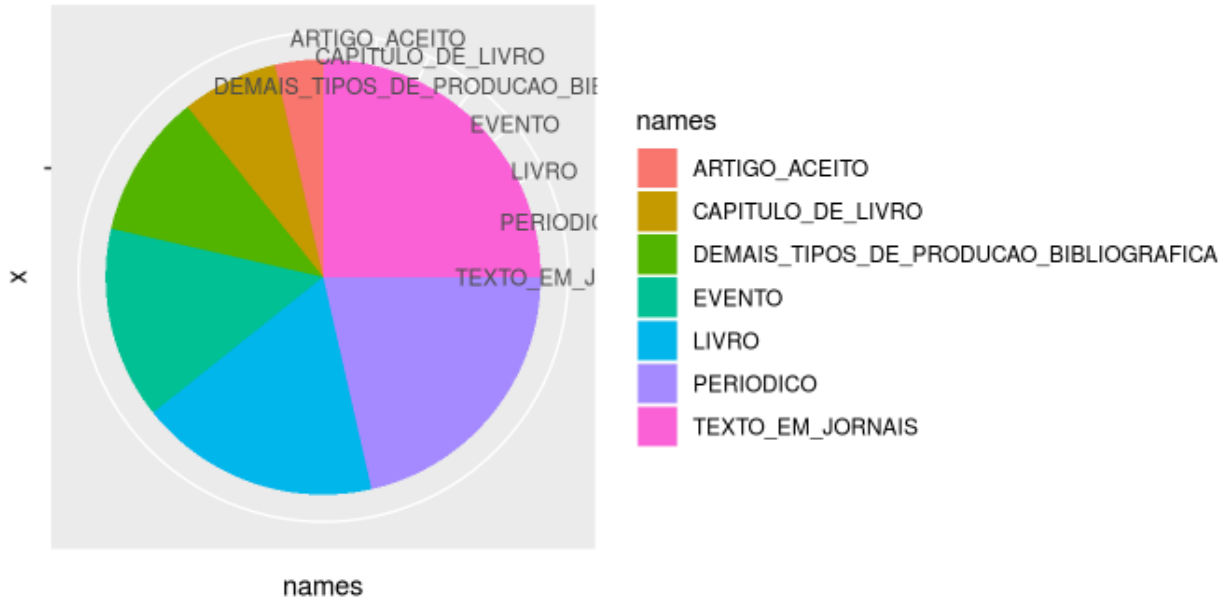
```
paises <- eventos2017$pais_do_evento
natureza <- eventos2017$natureza
data <- data.frame(paises, natureza)
ggplot(data, aes(x=paises, y="", col=paises)) + geom_point() + geom_jitter()
```



Publicações

Quanto a análise das publicações feitas pelos docentes, essas podem ser observadas no gráfico abaixo:

```
names <- names(unb.geografia.pub)
tamanhos <- sapply(unb.geografia.pub, function(x) length(x))
data <- data.frame(names, tamanhos)
ggplot(data, aes(x="", y=names, fill=names))+
  geom_bar(width = 1, stat = "identity") + coord_polar("y", start=0)
```



É evidente que no programa del pós-graduação de geografia da UnB há mais publicações de Periódicos, Texto em Jornais, Livros e em Eventos. Além disso, percebe-se que artigos aceitos é de longe a menor representação, em termos quantitativos, das publicações desse programa.

CRISP-DM Fase.Atividade 3.5 - Formatação dos dados

Quanto a formatação dos dados, não houve necessidade de ser feita pelos integrantes do grupo, pois ao longo da seleção até a integração dos dados foi visado a formatação, de forma que não houvesse necessidade de integração posterior.

CRISP-DM Fase 4 - Modelagem

Essa fase não se aplica nesse relatório, visto que modelos de análise não está mais no escopo desse projeto.

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

Não se aplica.

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

Não se aplica.

CRISP-DM Fase.Atividade 4.3 - Construção do modelo definitivo

Não se aplica.

CRISP-DM Fase.Atividade 4.4 - Avaliação do modelo

Não se aplica.

CRISP-DM Fase 5 - Avaliação

CRISP-DM Fase.Atividade 5.1 - Avaliação dos resultados

Após o tratamento dos dados, as análises demonstraram que houve um crescimento ao longo dos anos da atividade da Pós-Graduação nas áreas de Geografia, Geociências Aplicadas e Geodinâmica, e, Geologia. Isso era esperado considerando que ao decorrer da existência do programa mais pessoas estaram interessadas e haverá uma melhoria na qualidade dos cursos oferecidos.

CRISP-DM Fase.Atividade 5.2 - Revisão do processo

Não se aplica.

CRISP-DM Fase.Atividade 5.3 - Determinação dos etapas seguintes

Para continuação do trabalho, os próximos passos seriam de selecionar uma técnica de modelagem para aplicar aos dados, realizando testes até a obtenção do modelo definitivo, cujo conteria a análise desejada. Ou seja, a seção 4 antes prevista no trabalho. Quanto ao modelo aplicado e como trabalhar nele ainda não foi definido e/ou pensado, portanto é deixado aqui em aberto para uma avaliação futura.

CRISP-DM Fase 6 - Implantação (*deployment*)

CRISP-DM Fase.Atividade 6.1 - Planejamento da transição

Criando uma interface de fácil uso para o cliente, como uma aplicação *Web*, de forma que usuários sem experiência e conhecimento técnico das ferramentas de *Data Science* conseguissem utilizar esta aplicação de forma simples e intuitiva.

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

Para o monitoramento dos produtos é preciso a automatização da coleta dos dados e da execução dos scripts responsáveis pela análise dos dados e geração dos gráficos para assim não haver a necessidade de interferência humana a menos que haja realmente alguma mudança a ser feita na aplicação.

CRISP-DM Fase.Atividade 6.3 - Planejamento de manutenção

Como já dito acima, o ideal seria a automatização da aplicação, para ser funcional independente da execução de pessoas, como também a automatização da coleta dos dados, pois assim não seria necessário que alguém adicionasse manualmente novos arquivos com dados a serem analisados.

Também a generalização da aplicação, de forma que o usuário possa dizer quais dados ele deseja que sejam trabalhados, de que forma e como será a saída destes, se será em um arquivo em formato específico, através de gráficos e etc.

E, a aplicação de modelos para análises mais aprofundadas dos dados, como *Machine Learning*, processamento de texto e outras, possibilitando um entendimento melhor do programa de Pós-Graduação em diversos aspectos.

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

A entrega deste realatório reflete essa etapa.

CRISP-DM Fase.Atividade 6.5 - Apresentação do relatório final

Não se aplica, pois não haverá apresentação do relatório.

CRISP-DM Fase.Atividade 6.6 - Revisão sobre a execução do projeto

Com o decorrer do trabalho aprendemos que trabalhar com dados não é uma tarefa fácil, que exige várias etapas para poder realmente transformar eles para um formato em que seja possível estudá-los e analisá-los, para só assim gerar gráficos e informações com aplicabilidade em problemas do mundo real.

As possíveis melhorias futuras para esse trabalho estão listadas nas seções 6.1 a 6.3, que visam o aprimoramento de forma que tenha uma utilidade real para pessoas que não se encontram na área de tecnologia ou que não tem conhecimento prévio sobre *Data Science*.

Por fim, quanto a disciplina Ciência dos Dados para Todos contribuiu para conhecer como os dados gerados tem grande relevância quando são aplicadas técnicas para tratar e lidar com eles. Assim, esperamos conseguir aplicar o aprendido na disciplina em trabalhos futuros, tanto acadêmicos como profissionais, aproveitando e explorando realmente tudo o que os dados do que estão sendo trabalhos tem a oferecer.

Referências

- Fernandes, Jorge H C, Ricardo Barros Sampaio, e João Ribas de Moura. "Ciência de Dados para Todos (Data Science For All) - 2018.1 - Análise da Produção Científica e Acadêmica da Universidade de Brasília - Modelo de Relatório Final da Disciplina - Departamento de Ciência da Computação da UnB". Disciplina 116297 - Tópicos Avançados em Computadores, turma D, do semestre 2018.1, do Departamento de Ciência da Computação do Instituto de Ciências Exatas da Universidade de Brasília, 13 de junho de 2018.
- POSGEA. Apresentação, 2018. Disponíivel em http://www.posgea.unb.br/site/apresentacao.
- IGD. Geociências Aplicadas - Instituto de Geociências - Universidade de Brasília, 2018. Disponíivel em http://www.igd.unb.br/index.php?option=com_content&view=article&id=23&Itemid=136.