

Disciplina Ciência de Dados Aplicada e Ciência de Dados para Todos

Relatório 3 - Disciplina Ciência de Dados para Todos

Autor: Luis Gustavo Avelino de Lima Jacinto Data: 27/05/2018

1. Introdução e contextualização

O programa de pesquisa utilizado neste relatório é o de Geografia.

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 início 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/>.

É possível ver o histórico do programa, tendo este atualmente 26 integrantes participantes no corpo docente, sendo possível observar também seus artigos por triênio(13-14-15-16). Também é possível observar as informações do corpo discente, que dividido em mestrado e doutorado tem, respectivamente, 70 e 126 integrantes.

Um importante congresso dessa área é o *XIV Congreso de La Asociación Latinoamericana de Escuelas de Urbanismo Y Planificación*, que será realizado no Chile durante o mês de agosto.

As principais revistas e locais de publicação utilizados no Brasil e no mundo podem ser encontrados pela URL <http://www.posgea.unb.br/revistas> . Além disso, os principais parceiros relacionados a este programa de pesquisa e os orientandos e

orientados são: CsF(Ciência sem Fronteiras), PETROBRAS, CNPq(Conselho Nacional de Desenvolvimento Científico e Tecnológico), CAPES(Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), FINEP(Financiadora de Estudos e Projetos), FAP-DF(Fundação de Apoio a Pesquisa do Distrito Federal) e FINATEC(Fundação de Empreendimentos Científicos e Tecnológicos).

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.

Consultando pela base de dados perfis, provenientes da plataforma elattes, os envolvidos no projeto, temos:

```
> #Perfis Geografia
> perfisGeografia <- read_json('geografia.profile.json')
> length(perfisGeografia)
[1] 23
```

```
> sapply(perfisGeografia, function(x) x$name)
      0975018553829295      1175332440156178      1886939214378140
      "Helen da Costa Gurgel" "Neio Lucio de Oliveira Campos" "Roberto Arnaldo Trancoso Gomes"
      3466441462870689      3634018305250466      3758831342495343
      "Rafael Sanzio Araújo dos Anjos" "Shadia Hussein de Araújo" "Ercília Torres Steinke"
      5127536119098016      5334844470201838      5572269831914055
      "Cristina Maria Costa Leite" "Valdir Adilson Steinke" "Osmar Abílio de Carvalho Junior"
      5709330601484442      6222241339082462      6223294904658578
      "Juscelino Eudámidas Bezerra" "Rogério Elias Soares Uagoda" "Everaldo Batista da Costa"
      6563820088149165      6895446047430585      7063856452054362
      "Lucia Cony Faria Cidade" "Marília Steinberger" "Renato Fontes Guimarães"
      7762550938015741      7810411653917702      8160265101709215
      "Nelba Azevedo Penna" "Fernando Luiz Araújo Sobrinho" "Éder de Souza Martins"
      9066611120459129      9114817474946995      9524636602671805
      "Waleska Valença Manyari" "Mario Diniz de Araujo Neto" "Ruth Elias de Paula Laranja"
      9563937823266641      9980267316711819
      "Marília Luiza Peluso" "Dante Flávio da Costa Reis Júnior"
```

2. Referencial

Essa seção tem como objetivo examinar a literatura relacionada ao desenvolvimento dessa pesquisa. Abordando assuntos como Ciência da ciência e Estudos de Ciência e Tecnologia.

Ciência refere-se a qualquer conhecimento ou prática sistemáticos. Em sentido estrito, ciência refere-se ao sistema de adquirir conhecimento baseado no método científico bem como ao corpo organizado de conhecimento conseguido através de tais pesquisas.

Os estudos sociais da ciência e da tecnologia, se dedicam a investigar as relações entre ciência, tecnologia e sociedade. Estes estudos partem do entendimento de que ciência e tecnologia são atividades sociais, feitas por pessoas, situadas no tempo, no espaço, na cultura e também na sociedade, sendo, assim, impossível compreendê-las separadamente.

3. Metodologia

CRISP-DM

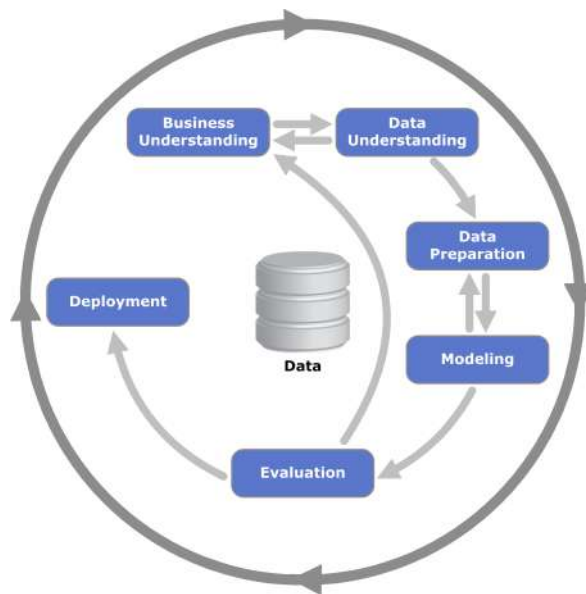
Este relatório tem como objetivo alcançar maior compreensão e entendimento do foco do PPGGEA seguindo o método CRISP-DM, que é definido em seis etapas

- *Business Understanding*
- *Data Understanding*
- *Data preparativos*
- *Modeling*
- *Evaluation*
- *Deployment*

No arquivo de Perfis temos 7 elementos principais:

Cada elemento contém informações sobre os integrantes desse programa.

```
"nome" "resumo_cv" "areas_de_atuacao" "endereco_profissional"
"producao_bibliografica" "orientacoes_academicas" "senioridade"
```



Em seguida, foram coletados esses dados dessa base de dados para uma melhor análise do conjunto como um todo utilizando funções como `glimpse()`, `summary()`, `names()` e também os gráficos da biblioteca `ggplot2`, para conseguir melhor visualização dos dados de forma a ter uma visão mais clara de alguns dados especificamente. Também foram utilizadas as bibliotecas `jsonlite`, `tidyr`, `ggplot2`, `dplyr`. E para leitura dos dados foi utilizado o método `read_json()`.

4. Resultados

Análise dos perfis dos integrantes do programa de pós-graduação

O arquivo “*geografia.profile.json*” pois dados referentes ao programa de pós graduação de geografia.

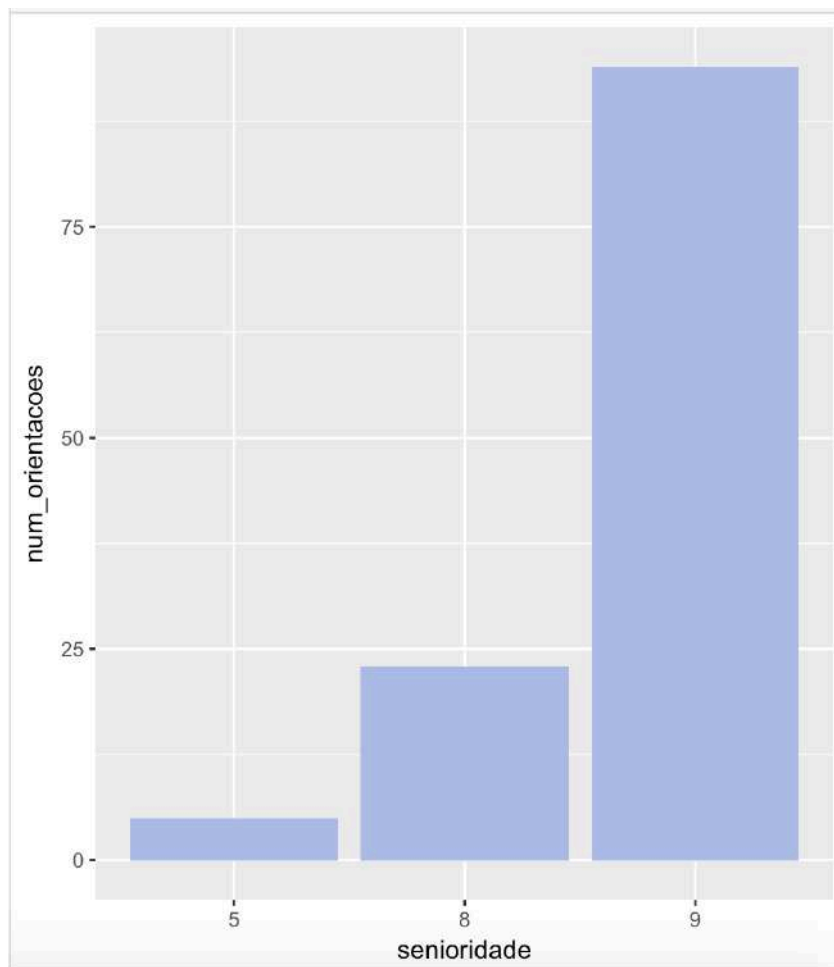
Dos professores presentes nesse arquivos pode-se analisar as seguintes informações de cada um, através da função `names()`:

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

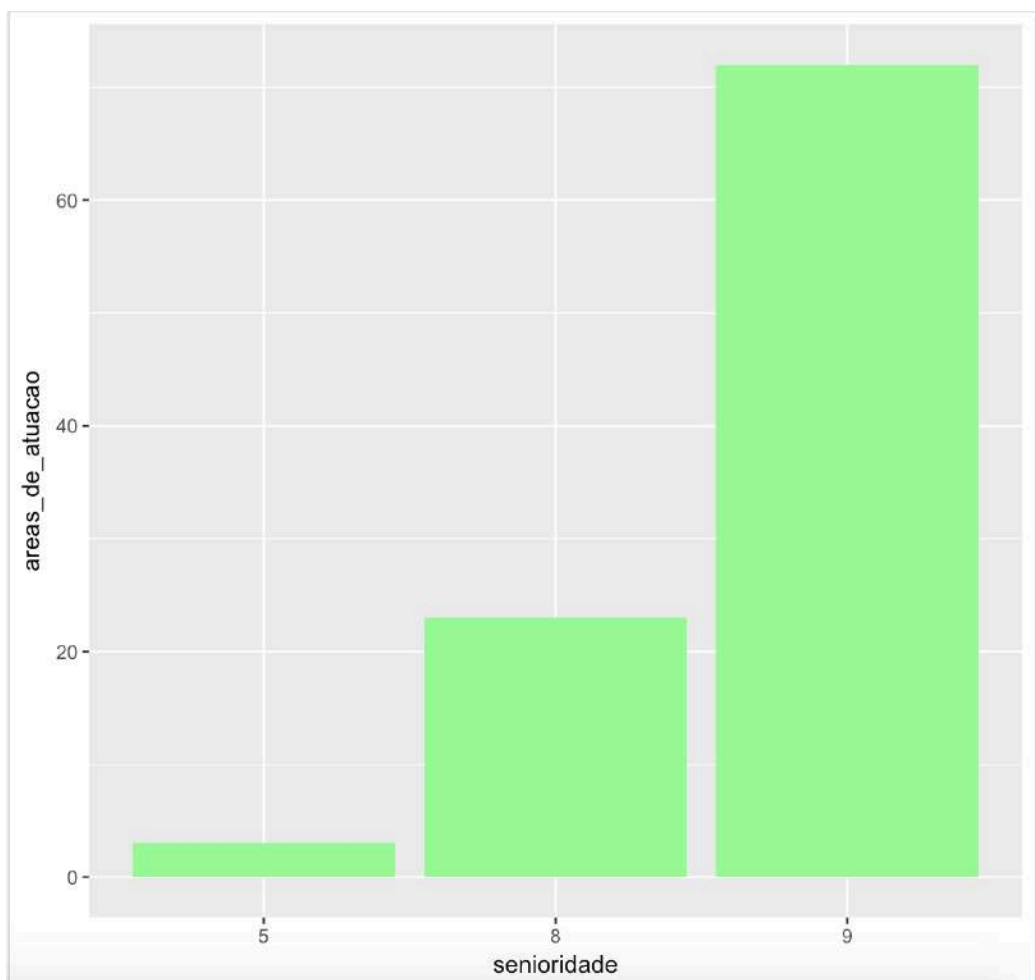
Analisando alguns dos dados presentes nesse arquivos, pode-se observar que existe um alto rendimento de professores a partir da sua senioridade, como mostrado abaixo

professores com maior senioridade possuem tanto maior número de orientações como maior número de áreas de atuação.

```
senioridade <- sapply(perfisGeografia, function(x) x$senioridade)
num_orientacoes <- sapply(perfisGeografia, function(x) length(x$orientacoes_academicas))
data <- data.frame(num_orientacoes, senioridade)
ggplot(data, aes(x=senioridade, y=num_orientacoes)) + geom_bar(stat="identity", fill="#adbce6")
```



```
areas_de_atuacao <- sapply(perfisGeografia, function(x) length(x$areas_de_atuacao))
data2 <- data.frame(senioridade, areas_de_atuacao)
ggplot(data2, aes(x=senioridade, y=areas_de_atuacao)) + geom_bar(stat="identity", fill="#98FB98")
```

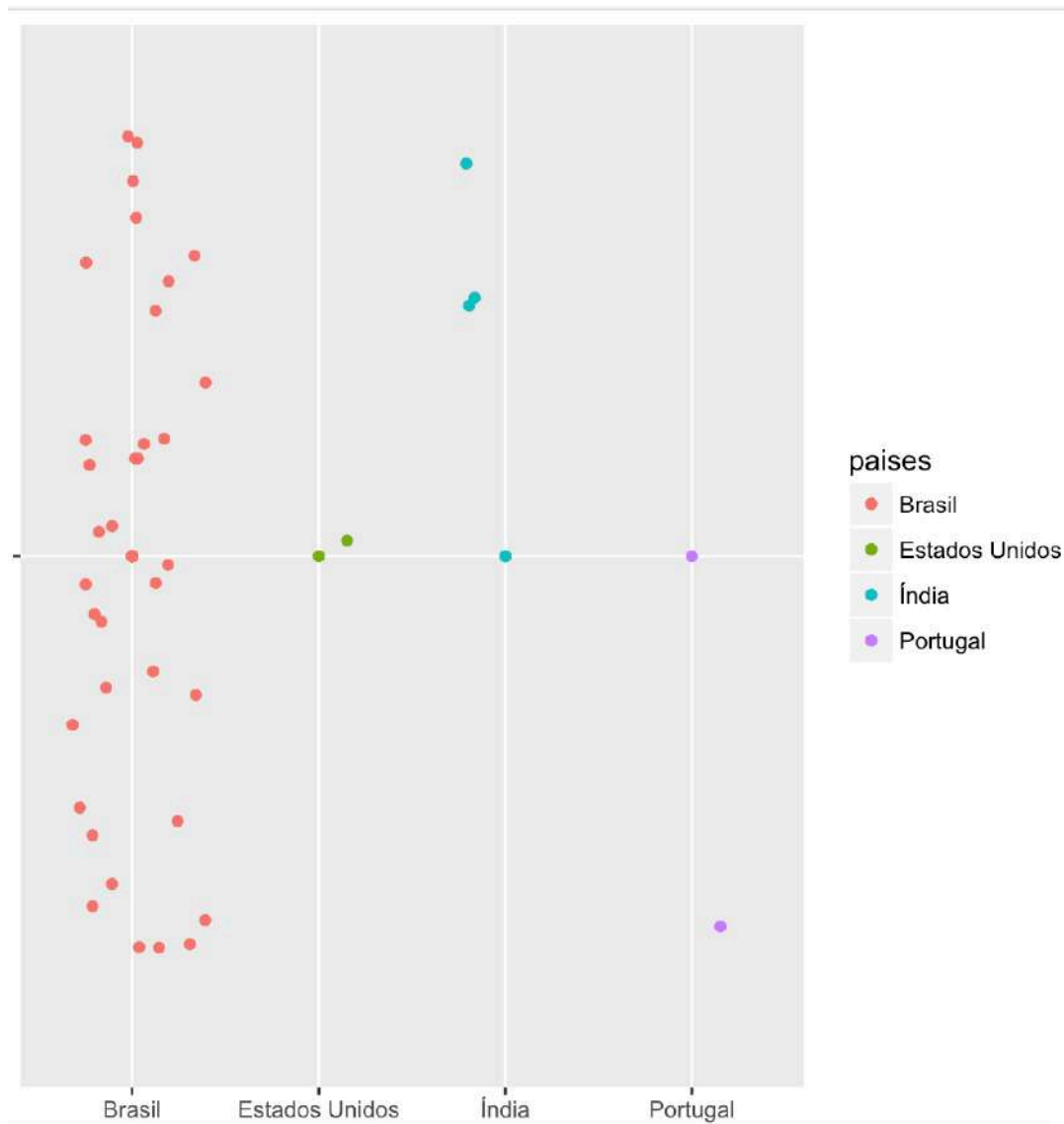


Nível de internacionalização

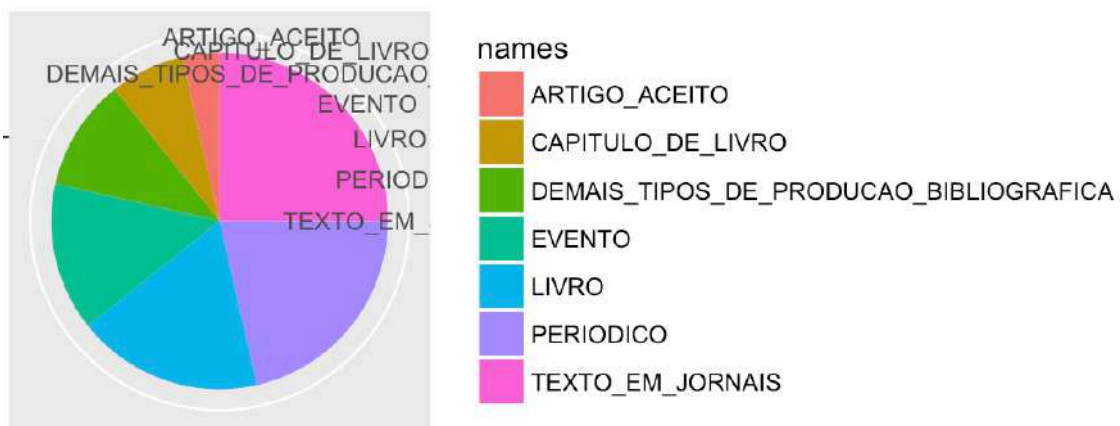
Analisando as publicações referentes a esses professores e os alunos orientados, percebe-se que grande parte das publicações ainda são feitas no Brasil. Das produções bibliográficas existentes, que podem ser classificadas conforme a imagem abaixo:

```
> names(perfisGeografia$"0975018553829295"$producao_bibliografica)
[1] "CAPITULO_DE_LIVRO" "EVENTO"           "LIVRO"            "PERIODICO"
```

A maior parte é nacional, assim o nível de internacionalização não é muito alto, como pode-se conferir com a imagem abaixo:



Das publicações feitas pelos alunos orientados que estão na base de dados, os tipos de publicações subdividem-se nas seguintes categorias abaixo, juntamente com a quantidade de cada categoria:



Percebe-se que maior parte das publicações são Texto em Jornais e Periódicos, e os menos feitos são Artigos Aceitos e Capítulo de Livro.

5. Conclusão

Por meio deste relatório foi possível ter melhor entendimento sobre o programa de pós graduação de geografia da unb, o perfil dos professores e integrantes desse programa e um pouco também sobre as publicações. Também foi possível observar as principais revistas utilizadas pelos integrantes dessa programa e também o reflexo do nível de senioridade dos docentes e a relação com o nível de produções e quantidade de orientados. Além de quantificar tais informações, foi possível ter uma análise mais visual por meio dos gráficos, sendo possível ter uma análise temporal dos dados analisados.