

Ciência de Dados para Todos (Data Science For All) - 2018.2 - Análise da Produção Científica e Acadêmica da Universidade de Brasília - Engenharia Biomédica, engenharia eletrônicos e automação, engenharia elétrica p7 e engenharia elétrica p8.

Marcos Vinicius Prescendo Tonin, Lucas Nascimento, Carlos Aragão

20/09/2018

Introdução

O presente trabalho tem por finalidade entender e correlacionar dados que possam algum significado mais profundo do que aparentam ter, assim com os dados da pós-graduação em engenharia Biomédica, engenharia elétrica em seus dois ramos (53001010059P8 e 53001010080P7) e engenharia de sistemas eletrônicos e de automação em mãos busca-se encontrar principais fatores, professores mais envolvidos, temas mais relevantes e afins.

Metodologia

Para um melhor resultado do trabalho buscou-se seguir e adaptar-se a metodologia CRISP-DM, para isso baseou-se no ciclo de projeto usado pelo CRISP-DM.

O ciclo é basicamente definido pelas seguintes fases:

- **Entendimento do negócio** : primeiramente deve-se entender o que se busca encontrar, haja visto que não faz sentido fazer uma análise de dados sem saber o que se busca.
- **Entendimento dos dados** : Busca-se entender os dados de forma mais superficial primeiramente.
- **Preparação de dados** : Então limpa-se o dado e prepara o dado para que se possa facilitar seu processamento.
- **Modelagem** : Faz a modelagem do dado.
- **Avaliação** : Avalia-se o resultado da modelagem.
- **Implatação** : Faz uso dos dados.

As fases não são independentes entre si, mas possuem certa comunicação, dependendo das fases, além de não ser estritamente sequencial, sendo melhor visualizado na figura abaixo.

Por esta imagem consegue-se perceber que o entendimento do negócio e do dado pode muitas vezes ser alternada, indo e voltando.

Delimitações iniciais

domínio de Aplicação do projeto

O domínio de Aplicação do projeto é Produção Científica ou Produção Acadêmica de um subgrupo de pesquisadores vinculados é Universidade de Brasília, sendo vinculado ao subtema **engenharia Biomédica**, tais arquivos foram pegos na plataforma elattes.

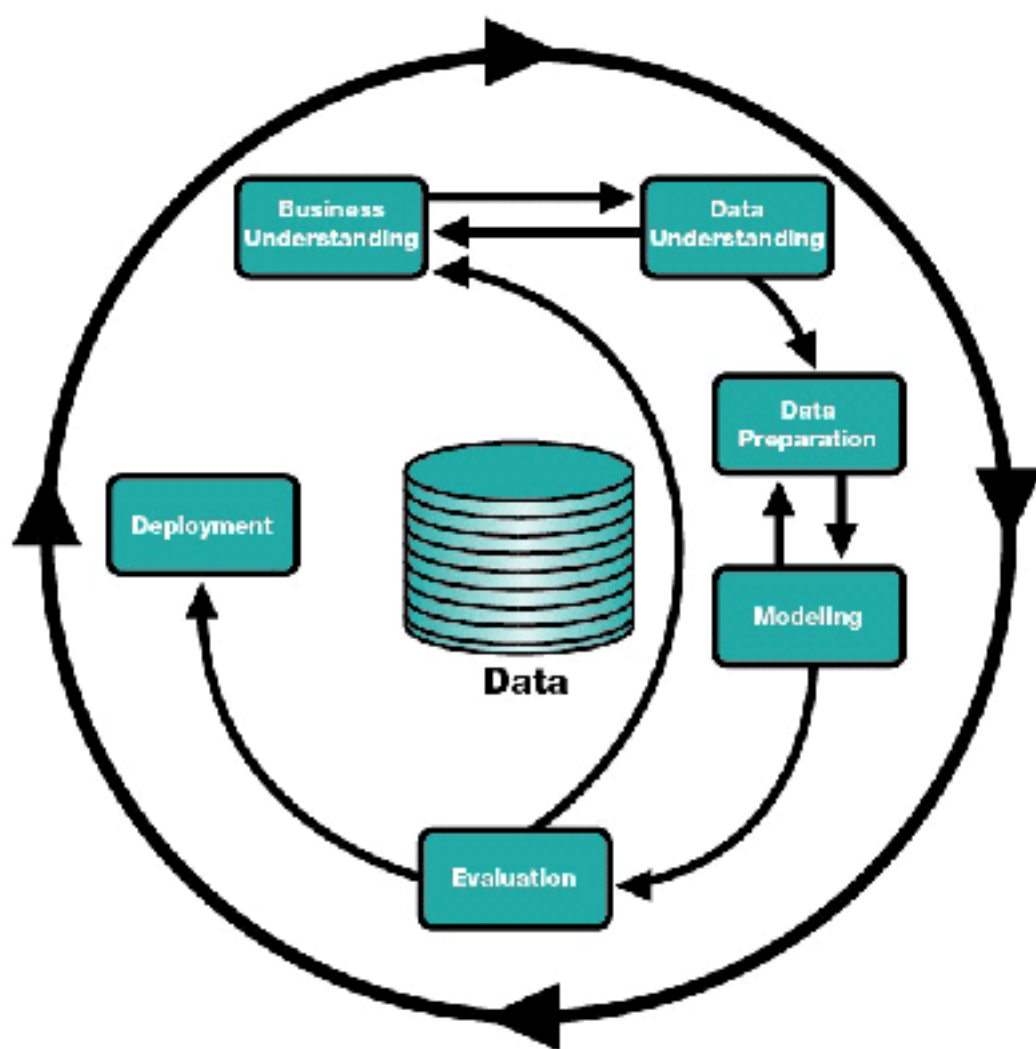


Figure 1: Ciclo do CRISP-DM que será usado como base

Problema abordado

Problema abordado tem por finalidade obter dados de forma descritiva, quantitativa e de modelagem computacional ou estatística, que permitam caracterizar como, porque e também para que ocorre a Produção Científica e Acadêmica na área de engenharia Biomédica.

CRISP-DM Fase 1 - Entendimento do negócio

A Universidade de Brasília é uma das grandes universidades federais, sendo responsável pela Produção de artigos visando artigos acadêmicos, este presente trabalho de como programas :

Programas de pós-graduação Grupo 15

Ids	Área de Avaliação	Nome do Programa	Link Capes	ME	DO	MP
53001010059P8	ENGENHARIA elétrica IV	ENGENHARIA elétrica	ENGENHARIA elétrica (53001010059P8)	-	-	3
53001010080P7	ENGENHARIA elétrica IV	ENGENHARIA elétrica	ENGENHARIA elétrica (53001010080P7)	4	4	-
53001010081P3	ENGENHARIA DE SISTEMAS eletrônicos E DE automação IV	ENGENHARIA DE SISTEMAS eletrônicos E DE automação	ENGENHARIA DE SISTEMAS eletrônicos E DE automação (53001010081P3)	4	4	-
53001010083P6	ENGENHARIA Biomédica IV	ENGENHARIA Biomédica	ENGENHARIA Biomédica (53001010083P6)	3	-	-

Sendo o enfoque principal os dados referente a engenharia Biomédica, que por sua vez tem sua pesquisa muitas voltada a novas soluções na área de Biomédica e demais áreas representadas acima.

Avaliação das Circunstâncias

Este trabalho terá o limite que os próprios dados gerado pela plataforma elattes tem, já que se os dados fornecidos delimitariam o escopo do projeto e até pode-se ir, ou melhor, até onde os dados permitem uma análise fidedigna.

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, através da plataforma elattes. Os dados foram gerados no mês de setembro de 2018, data em que o presente grupo começou a realizar o trabalho e compilam informações entre os anos de 2010 e 2017, das áreas expostas na seção anterior. Os arquivos estão no formato JSON, sendo todos fornecidos pelos docentes responsável por esta disciplina.

Perfil profissional dos docentes vinculados às pós-graduações

```
json.perfil.biomed <- "dados-2018-2/engenharia-biomedica/279.profile.json"
file.info(json.perfil.biomed)
```

```
##                                     size isdir mode
## dados-2018-2/engenharia-biomedica/279.profile.json 770014 FALSE 666
##                                                         mtime
```

```
## dados-2018-2/engenharia-biomedica/279.profile.json 2018-09-20 19:26:09
##                                     ctime
## dados-2018-2/engenharia-biomedica/279.profile.json 2018-09-20 19:26:09
##                                     atime exe
## dados-2018-2/engenharia-biomedica/279.profile.json 2018-11-16 19:03:04 no

json.perfil.auto <- "dados-2018-2/engenharia-eletronicos-automacao/280.profile.json"
file.info(json.perfil.auto)
```

```
##                                     size
## dados-2018-2/engenharia-eletronicos-automacao/280.profile.json 828607
##                                     isdir mode
## dados-2018-2/engenharia-eletronicos-automacao/280.profile.json FALSE 666
##                                     mtime
## dados-2018-2/engenharia-eletronicos-automacao/280.profile.json 2018-09-20 19:26:09
##                                     ctime
## dados-2018-2/engenharia-eletronicos-automacao/280.profile.json 2018-09-20 19:26:09
##                                     atime
## dados-2018-2/engenharia-eletronicos-automacao/280.profile.json 2018-11-16 19:03:02
##                                     exe
## dados-2018-2/engenharia-eletronicos-automacao/280.profile.json no

json.perfil.ele_p8 <- "dados-2018-2/engenharia-eletrica-p8/281.profile.json"
file.info(json.perfil.ele_p8)
```

```
##                                     size isdir mode
## dados-2018-2/engenharia-eletrica-p8/281.profile.json 559453 FALSE 666
##                                     mtime
## dados-2018-2/engenharia-eletrica-p8/281.profile.json 2018-09-20 19:26:09
##                                     ctime
## dados-2018-2/engenharia-eletrica-p8/281.profile.json 2018-09-20 19:26:09
##                                     atime
## dados-2018-2/engenharia-eletrica-p8/281.profile.json 2018-11-16 19:03:02
##                                     exe
## dados-2018-2/engenharia-eletrica-p8/281.profile.json no

json.perfil.ele_p7 <- "dados-2018-2/engenharia-eletrica-p7/282.profile.json"
file.info(json.perfil.ele_p7)
```

```
##                                     size isdir mode
## dados-2018-2/engenharia-eletrica-p7/282.profile.json 711039 FALSE 666
##                                     mtime
## dados-2018-2/engenharia-eletrica-p7/282.profile.json 2018-09-20 19:26:09
##                                     ctime
## dados-2018-2/engenharia-eletrica-p7/282.profile.json 2018-09-20 19:26:09
##                                     atime
## dados-2018-2/engenharia-eletrica-p7/282.profile.json 2018-11-16 19:03:02
##                                     exe
## dados-2018-2/engenharia-eletrica-p7/282.profile.json no
```

Os arquivos dados-2018-2/engenharia-biomedica/279.profile.json, dados-2018-2/engenharia-eletronicos-automacao/280.profile.json, dados-2018-2/engenharia-eletrica-p7/282.profile.json, dados-2018-2/engenharia-eletrica-p8/281.profile.json apresentam dados sobre o perfil de todos os docentes vinculados a programas de pós-graduação, em engenharia biomédica, automação, engenharia elétrica (p7 e p8), da UnB, entre 2010 e 2017.

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

```
json.advise.biomed <- "dados-2018-2/engenharia-biomedica/279.advise.json"
file.info(json.advise.biomed)
```

```
##                                size isdir mode
## dados-2018-2/engenharia-biomedica/279.advise.json 361237 FALSE 666
##                                                    mtime
## dados-2018-2/engenharia-biomedica/279.advise.json 2018-09-20 19:26:09
##                                                    ctime
## dados-2018-2/engenharia-biomedica/279.advise.json 2018-09-20 19:26:09
##                                                    atime exe
## dados-2018-2/engenharia-biomedica/279.advise.json 2018-11-16 19:03:03 no
```

```
json.advise.auto <- "dados-2018-2/engenharia-eletronicos-automacao/280.advise.json"
file.info(json.advise.auto)
```

```
##                                size isdir
## dados-2018-2/engenharia-eletronicos-automacao/280.advise.json 345359 FALSE
##                                                    mode
## dados-2018-2/engenharia-eletronicos-automacao/280.advise.json 666
##                                                    mtime
## dados-2018-2/engenharia-eletronicos-automacao/280.advise.json 2018-09-20 19:26:09
##                                                    ctime
## dados-2018-2/engenharia-eletronicos-automacao/280.advise.json 2018-09-20 19:26:09
##                                                    atime
## dados-2018-2/engenharia-eletronicos-automacao/280.advise.json 2018-11-16 19:03:00
##                                                    exe
## dados-2018-2/engenharia-eletronicos-automacao/280.advise.json no
```

```
json.advise.ele_p8 <- "dados-2018-2/engenharia-eletrica-p8/281.advise.json"
file.info(json.advise.ele_p8)
```

```
##                                size isdir mode
## dados-2018-2/engenharia-eletrica-p8/281.advise.json 181900 FALSE 666
##                                                    mtime
## dados-2018-2/engenharia-eletrica-p8/281.advise.json 2018-09-20 19:26:09
##                                                    ctime
## dados-2018-2/engenharia-eletrica-p8/281.advise.json 2018-09-20 19:26:09
##                                                    atime
## dados-2018-2/engenharia-eletrica-p8/281.advise.json 2018-11-16 19:03:00
##                                                    exe
## dados-2018-2/engenharia-eletrica-p8/281.advise.json no
```

```
json.advise.ele_p7 <- "dados-2018-2/engenharia-eletrica-p7/282.advise.json"
file.info(json.advise.ele_p7)
```

```
##                                size isdir mode
## dados-2018-2/engenharia-eletrica-p7/282.advise.json 274263 FALSE 666
##                                                    mtime
## dados-2018-2/engenharia-eletrica-p7/282.advise.json 2018-09-20 19:26:09
##                                                    ctime
## dados-2018-2/engenharia-eletrica-p7/282.advise.json 2018-09-20 19:26:09
##                                                    atime
## dados-2018-2/engenharia-eletrica-p7/282.advise.json 2018-11-16 19:03:00
##                                                    exe
```

```
## dados-2018-2/engenharia-eletrica-p7/282.advise.json no
```

Os arquivos dados-2018-2/engenharia-biomedica/279.advise.json, dados-2018-2/engenharia-eletronicos-automacao/280.advise.json, dados-2018-2/engenharia-eletrica-p7/282.advise.json, dados-2018-2/engenharia-eletrica-p8/281.advise.json apresenta dados sobre o orientações de mestrado e doutorado feitas por todos os docentes vinculados a programas de pós-graduação em engenharia Biomédica e demais, da UnB, entre 2010 e 2017.

Produção bibliográfica gerada pelos docentes vinculados às pós-graduações

```
json.producao.bibliografica.biomed <- "dados-2018-2/engenharia-biomedica/279.publication.json"
file.info(json.producao.bibliografica.biomed)
```

```
##                                     size isdir mode
## dados-2018-2/engenharia-biomedica/279.publication.json 294169 FALSE 666
##                                                         mtime
## dados-2018-2/engenharia-biomedica/279.publication.json 2018-09-20 19:26:09
##                                                         ctime
## dados-2018-2/engenharia-biomedica/279.publication.json 2018-09-20 19:26:09
##                                                         atime
## dados-2018-2/engenharia-biomedica/279.publication.json 2018-11-16 19:03:03
##                                                         exe
## dados-2018-2/engenharia-biomedica/279.publication.json no
```

```
json.producao.bibliografica.auto <- "dados-2018-2/engenharia-eletronicos-automacao/280.publication.json"
file.info(json.producao.bibliografica.auto)
```

```
##                                     size
## dados-2018-2/engenharia-eletronicos-automacao/280.publication.json 368732
##                                                         isdir
## dados-2018-2/engenharia-eletronicos-automacao/280.publication.json FALSE
##                                                         mode
## dados-2018-2/engenharia-eletronicos-automacao/280.publication.json 666
##                                                         mtime
## dados-2018-2/engenharia-eletronicos-automacao/280.publication.json 2018-09-20 19:26:09
##                                                         ctime
## dados-2018-2/engenharia-eletronicos-automacao/280.publication.json 2018-09-20 19:26:09
##                                                         atime
## dados-2018-2/engenharia-eletronicos-automacao/280.publication.json 2018-11-16 19:03:00
##                                                         exe
## dados-2018-2/engenharia-eletronicos-automacao/280.publication.json no
```

```
json.producao.bibliografica.ele_p8 <- "dados-2018-2/engenharia-eletrica-p8/281.publication.json"
file.info(json.producao.bibliografica.ele_p8)
```

```
##                                     size isdir mode
## dados-2018-2/engenharia-eletrica-p8/281.publication.json 322640 FALSE 666
##                                                         mtime
## dados-2018-2/engenharia-eletrica-p8/281.publication.json 2018-09-20 19:26:09
##                                                         ctime
## dados-2018-2/engenharia-eletrica-p8/281.publication.json 2018-09-20 19:26:09
##                                                         atime
## dados-2018-2/engenharia-eletrica-p8/281.publication.json 2018-11-16 19:03:00
##                                                         exe
## dados-2018-2/engenharia-eletrica-p8/281.publication.json no
```

```
json.producao.bibliografica.ele_p7 <- "dados-2018-2/engenharia-eletrica-p7/282.publication.json"
file.info(json.producao.bibliografica.ele_p7)
```

```
##                                     size isdir mode
## dados-2018-2/engenharia-eletrica-p7/282.publication.json 362839 FALSE 666
##                                                                mtime
## dados-2018-2/engenharia-eletrica-p7/282.publication.json 2018-09-20 19:26:09
##                                                                ctime
## dados-2018-2/engenharia-eletrica-p7/282.publication.json 2018-09-20 19:26:09
##                                                                atime
## dados-2018-2/engenharia-eletrica-p7/282.publication.json 2018-11-16 19:03:00
##                                                                exe
## dados-2018-2/engenharia-eletrica-p7/282.publication.json no
```

Os arquivos dados-2018-2/engenharia-biomedica/279.publication.json, dados-2018-2/engenharia-eletronicos-automacao/280.publication.json, dados-2018-2/engenharia-eletrica-p7/282.publication.json e dados-2018-2/engenharia-eletrica-p8/281.publication.json apresenta dados sobre a Produção bibliográfica gerada por todos os docentes vinculados a programas de pós-graduação, em engenharia elétrica p7 e também as demais, da UnB, entre 2010 e 2017.

ID's dos docentes participantes e o que contém o arquivo list.json

```
json.list.biomed <- "dados-2018-2/engenharia-biomedica/279.list.json"
file.info(json.list.biomed)
```

```
##                                     size isdir mode
## dados-2018-2/engenharia-biomedica/279.list.json 943 FALSE 666
##                                                                mtime
## dados-2018-2/engenharia-biomedica/279.list.json 2018-09-20 19:26:09
##                                                                ctime
## dados-2018-2/engenharia-biomedica/279.list.json 2018-09-20 19:26:09
##                                                                atime exe
## dados-2018-2/engenharia-biomedica/279.list.json 2018-11-16 19:03:03 no
```

```
json.list.auto <- "dados-2018-2/engenharia-eletronicos-automacao/280.list.json"
file.info(json.list.auto)
```

```
##                                     size isdir
## dados-2018-2/engenharia-eletronicos-automacao/280.list.json 1315 FALSE
##                                                                mode
## dados-2018-2/engenharia-eletronicos-automacao/280.list.json 666
##                                                                mtime
## dados-2018-2/engenharia-eletronicos-automacao/280.list.json 2018-09-20 19:26:09
##                                                                ctime
## dados-2018-2/engenharia-eletronicos-automacao/280.list.json 2018-09-20 19:26:09
##                                                                atime
## dados-2018-2/engenharia-eletronicos-automacao/280.list.json 2018-11-16 19:03:03
##                                                                exe
## dados-2018-2/engenharia-eletronicos-automacao/280.list.json no
```

```
json.list.ele_p8 <- "dados-2018-2/engenharia-eletrica-p8/281.list.json"
file.info(json.list.ele_p8)
```

```
##                                     size isdir mode
## dados-2018-2/engenharia-eletrica-p8/281.list.json 695 FALSE 666
##                                                                mtime
```

```
## dados-2018-2/engenharia-eletrica-p8/281.list.json 2018-09-20 19:26:09
##                                     ctime
## dados-2018-2/engenharia-eletrica-p8/281.list.json 2018-09-20 19:26:09
##                                     atime exe
## dados-2018-2/engenharia-eletrica-p8/281.list.json 2018-11-16 19:03:03 no
```

```
json.list.ele_p7 <- "dados-2018-2/engenharia-eletrica-p7/282.list.json"
file.info(json.list.ele_p7)
```

```
##                                     size isdir mode
## dados-2018-2/engenharia-eletrica-p7/282.list.json 881 FALSE 666
##                                     mtime
## dados-2018-2/engenharia-eletrica-p7/282.list.json 2018-09-20 19:26:09
##                                     ctime
## dados-2018-2/engenharia-eletrica-p7/282.list.json 2018-09-20 19:26:09
##                                     atime exe
## dados-2018-2/engenharia-eletrica-p7/282.list.json 2018-11-16 19:03:03 no
```

Os arquivos dados-2018-2/engenharia-biomedica/279.list.json apresenta o id de todos docentes vinculados a programas de pós-graduação descritos acima da UnB, entre 2010 e 2017. porém tal arquivo se mostra inútil, haja visto que só contém o id e que as outras variáveis como “nome” sempre tem seu valor igual a “”.

Redes de colaboração entre docentes

```
json.graph.biomed<- 'dados-2018-2/engenharia-biomedica/279.graph.json'
file.info(json.graph.biomed)
```

```
##                                     size isdir mode
## dados-2018-2/engenharia-biomedica/279.graph.json 3901 FALSE 666
##                                     mtime
## dados-2018-2/engenharia-biomedica/279.graph.json 2018-09-20 19:26:09
##                                     ctime
## dados-2018-2/engenharia-biomedica/279.graph.json 2018-09-20 19:26:09
##                                     atime exe
## dados-2018-2/engenharia-biomedica/279.graph.json 2018-11-16 19:03:03 no
```

```
json.graph.auto <- "dados-2018-2/engenharia-eletronicos-automacao/280.graph.json"
file.info(json.graph.auto)
```

```
##                                     size isdir
## dados-2018-2/engenharia-eletronicos-automacao/280.graph.json 3901 FALSE
##                                     mode
## dados-2018-2/engenharia-eletronicos-automacao/280.graph.json 666
##                                     mtime
## dados-2018-2/engenharia-eletronicos-automacao/280.graph.json 2018-09-20 19:26:09
##                                     ctime
## dados-2018-2/engenharia-eletronicos-automacao/280.graph.json 2018-09-20 19:26:09
##                                     atime
## dados-2018-2/engenharia-eletronicos-automacao/280.graph.json 2018-11-16 19:03:00
##                                     exe
## dados-2018-2/engenharia-eletronicos-automacao/280.graph.json no
```

```
json.graph.ele_p8 <- "dados-2018-2/engenharia-eletrica-p8/281.graph.json"
file.info(json.graph.ele_p8)
```

```
##                                     size isdir mode
## dados-2018-2/engenharia-eletrica-p8/281.graph.json 1676 FALSE 666
```



```
##                                     mtime
## dados-2018-2/engenharia-eletrica-p8/281.graph.json 2018-09-20 19:26:09
##                                     ctime
## dados-2018-2/engenharia-eletrica-p8/281.graph.json 2018-09-20 19:26:09
##                                     atime exe
## dados-2018-2/engenharia-eletrica-p8/281.graph.json 2018-11-16 19:03:00 no
```

```
json.graph.ele_p7 <- "dados-2018-2/engenharia-eletrica-p7/282.graph.json"
file.info(json.graph.ele_p7)
```

```
##                                     size isdir mode
## dados-2018-2/engenharia-eletrica-p7/282.graph.json 1927 FALSE 666
##                                     mtime
## dados-2018-2/engenharia-eletrica-p7/282.graph.json 2018-09-20 19:26:09
##                                     ctime
## dados-2018-2/engenharia-eletrica-p7/282.graph.json 2018-09-20 19:26:09
##                                     atime exe
## dados-2018-2/engenharia-eletrica-p7/282.graph.json 2018-11-16 19:03:00 no
```

Os arquivos dados-2018-2/engenharia-biomedica/279.graph.json, dados-2018-2/engenharia-eletrica-p7/282.graph.json, dados-2018-2/engenharia-eletrica-p8/281.graph.json e dados-2018-2/engenharia-eletronicos-automacao/280.graph.json apresenta redes de colaboração na co-autoria de artigos científicos, feitas entre os docentes vinculados a programas de pós-graduação da UnB, entre 2010 e 2017.

CRISP-DM Fase.Atividade 2.2 - Descrição dos Dados

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

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

```
## -- Attaching packages ----- tid
## v tibble 1.4.2      v dplyr 0.7.6
## v tidyr 0.8.1      v stringr 1.3.1
## v purrr 0.2.5      v forcats 0.3.0
```

```
## -- Conflicts ----- tid
## x dplyr::filter() masks stats::filter()
## x purrr::flatten() masks jsonlite::flatten()
## x dplyr::lag() masks stats::lag()
```

```
library(stringr)
library(igraph)
```

```
##
## Attaching package: 'igraph'
## The following objects are masked from 'package:dplyr':
##
## as_data_frame, groups, union
## The following objects are masked from 'package:purrr':
##
```

```
##      compose, simplify
## The following object is masked from 'package:tidyr':
##
##      crossing
## The following object is masked from 'package:tibble':
##
##      as_data_frame
## The following objects are masked from 'package:stats':
##
##      decompose, spectrum
## The following object is masked from 'package:base':
##
##      union
```

Com estas bibliotecas seremos capazes de responder e determinar qual o volume de dados, a estrutura dos dados (tipos), codificações usadas, etc..

Descrição dos dados do perfil

TODO LIST : VERIFICAR

- Engenharia p7
- Engenharia p8
- automação

Verificando engenharia Biomédica

```
unb.prof.biomed <- fromJSON("dados-2018-2/engenharia-biomedica/279.profile.json")
```

A quantidade de docentes sob análise é apresentada a seguir.

```
length(unb.prof.biomed)
```

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

Para um melhor entendimento é importante saber como os dados estão dispostos para algum número de ocorrência.

```
#### Usando glimpse
```

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

```
## List of 7
## $ nome : chr "Leandro Xavier Cardoso"
## $ resumo_cv : chr "Possui graduação em Física Bacharelado e Licenciatura pela Universid
## $ areas_de_atuacao : 'data.frame': 1 obs. of 4 variables:
## ..$ grande_area : chr ""
## ..$ area : chr ""
## ..$ sub_area : chr ""
## ..$ especialidade: chr ""
## $ endereco_profissional :List of 8
## ..$ instituicao: chr "UnB - UNIVERSIDADE DE BRASÍLIA - FGA - FACULDADE GAMA"
## ..$ orgao : chr ""
## ..$ unidade : chr ""
## ..$ DDD : chr "61"
## ..$ telefone : chr "31078903"
## ..$ bairro : chr "Setor Leste (Gama)"
```

```

## ..$ cep : chr "72444240"
## ..$ cidade : chr "Brasília"
## $ producao_bibliografica :List of 2
## ..$ EVENTO : 'data.frame': 3 obs. of 11 variables:
## .. ..$ natureza : chr [1:3] "COMPLETO" "COMPLETO" "COMPLETO"
## .. ..$ titulo : chr [1:3] "As Tic's na Educação: Mudança ou Modernização" "A Educação no C
## .. ..$ nome_do_evento : chr [1:3] "IV Colóquio Internacional Educação e Contemporaneidade" "IV Co
## .. ..$ ano_do_trabalho : chr [1:3] "2010" "2010" "2014"
## .. ..$ pais_do_evento : chr [1:3] "Brasil" "Brasil" "Brasil"
## .. ..$ cidade_do_evento: chr [1:3] "São Cristovão" "São Cristovão" "Uberlândia"
## .. ..$ doi : chr [1:3] "" "" ""
## .. ..$ classificacao : chr [1:3] "INTERNACIONAL" "INTERNACIONAL" "NACIONAL"
## .. ..$ paginas : chr [1:3] " - " " - " "2636 - 2638"
## .. ..$ autores :List of 3
## .. ..$ autores-endogeno:List of 3
## ..$ PERIODICO: 'data.frame': 10 obs. of 10 variables:
## .. ..$ natureza : chr [1:10] "COMPLETO" "COMPLETO" "COMPLETO" "COMPLETO" ...
## .. ..$ titulo : chr [1:10] "Thermoluminescent dose reconstruction using quartz extracted
## .. ..$ periodico : chr [1:10] "Journal of Physics. Conference Series (Online)" "Scientia Plen
## .. ..$ ano : chr [1:10] "2010" "2011" "2011" "2013" ...
## .. ..$ volume : chr [1:10] "249" "7" "80" "477" ...
## .. ..$ issn : chr [1:10] "17426596" "18082793" "20103778" "17426588" ...
## .. ..$ paginas : chr [1:10] "012031 - " "014101 - " "285 - 290" "012011 - " ...
## .. ..$ doi : chr [1:10] "10.1088/1742-6596/249/1/012031" "" "" "10.1088/1742-6596/477/
## .. ..$ autores :List of 10
## .. ..$ autores-endogeno:List of 10
## $ orientacoes_academicas:List of 2
## ..$ ORIENTACAO_EM_ANDAMENTO_MESTRADO: 'data.frame': 5 obs. of 13 variables:
## .. ..$ natureza : chr [1:5] "Dissertação de mestrado" "Dissertação de mestrado"
## .. ..$ titulo : chr [1:5] "Utilização de dosímetros termoluminescentes comercia
## .. ..$ ano : chr [1:5] "2014" "2014" "2015" "2015" ...
## .. ..$ id_lattes_aluno : chr [1:5] "" "" "" "" ...
## .. ..$ nome_aluno : chr [1:5] "Rafael Assunção Gomes de Souza" "Marcelo Oppermann"
## .. ..$ instituicao : chr [1:5] "Universidade de Brasília" "Universidade de Brasília
## .. ..$ curso : chr [1:5] "Engenharia Biomédica" "Engenharia Biomédica" "Engen
## .. ..$ codigo_do_curso : chr [1:5] "60059672" "90000006" "90000006" "60059672" ...
## .. ..$ 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
## ..$ OUTRAS_ORIENTACOES_CONCLUIDAS : 'data.frame': 5 obs. of 13 variables:
## .. ..$ natureza : chr [1:5] "MONOGRAFIA_DE_CONCLUSAO_DE_CURSO_APERFEICOAMENTO_E_I
## .. ..$ titulo : chr [1:5] "Insuficiência na Aprendizagem de Matemática do 9º an
## .. ..$ ano : chr [1:5] "2011" "2011" "2011" "2011" ...
## .. ..$ id_lattes_aluno : chr [1:5] "" "" "" "" ...
## .. ..$ nome_aluno : chr [1:5] "Décio Luiz Alves Barreto e outros" "Izaque dos Santo
## .. ..$ instituicao : chr [1:5] "Faculdade Serigy" "Faculdade Serigy" "Faculdade Ser
## .. ..$ curso : chr [1:5] "Fundamentos e Métodos do ensino da Matemática" "Funda
## .. ..$ codigo_do_curso : chr [1:5] "90000005" "90000005" "90000005" "90000005" ...
## .. ..$ 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
## $ senioridade : chr "9"
```

Podemos inferir que:

- Que o professor não é da área da engenharia elétrica, por formação, mas acabou por aderir a subárea engenharia Biomédica.
- não é nativo da UnB, sendo formado no UFG.
- Atualmente trabalha na UnB do Gama.
- Sua senioridade é de 9.

Verificando engenharia de sistemas eletrônicos e de automação

Fazendo processo similar para as outras áreas.

```
unb.prof.auto <- fromJSON("dados-2018-2/engenharia-eletronicos-automacao/280.profile.json")
```

A quantidade de docentes sob análise é apresentada a seguir.

```
length(unb.prof.auto)
```

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

Verificando terceiro elemento.

```
### Usando glimpse
```

```
glimpse(unb.prof.auto[[3]], width = 30)
```

```
## List of 7
```

```
## $ nome : chr "Adson Ferreira da Rocha"
```

```
## $ resumo_cv : chr "Engenheiro Eletricista pela Universidade de Brasília (1988), Mestre e
```

```
## $ areas_de_atuacao : 'data.frame': 2 obs. of 4 variables:
```

```
## ..$ grande_area : chr [1:2] "ENGENHARIAS" "ENGENHARIAS"
```

```
## ..$ area : chr [1:2] "Engenharia Biomédica" "Engenharia Elétrica"
```

```
## ..$ sub_area : chr [1:2] "" ""
```

```
## ..$ especialidade: chr [1:2] "" ""
```

```
## $ endereco_profissional :List of 8
```

```
## ..$ instituicao: chr "Universidade de Brasília"
```

```
## ..$ orgao : chr "Faculdade UnB Gama"
```

```
## ..$ unidade : chr ""
```

```
## ..$ DDD : chr "61"
```

```
## ..$ telefone : chr "34843443"
```

```
## ..$ bairro : chr "ASA NORTE"
```

```
## ..$ cep : chr "72405-610"
```

```
## ..$ cidade : chr "Brasilia"
```

```
## $ producao_bibliografica :List of 4
```

```
## ..$ ARTIGO_ACEITO : 'data.frame': 1 obs. of 10 variables:
```

```
## .. ..$ natureza : chr "NAO_INFORMADO"
```

```
## .. ..$ titulo : chr "TISSUE FORMATION INDUCER SYSTEM FOR DIABETIC FOOT WITH LED LIGHT EMIS
```

```
## .. ..$ periodico : chr "Revista Brasileira de Engenharia Biomédica (Impresso)"
```

```
## .. ..$ ano : chr "2015"
```

```
## .. ..$ volume : chr ""
```

```
## .. ..$ issn : chr "15173151"
```

```
## .. ..$ paginas : chr " - "
```

```
## .. ..$ doi : chr ""
```

```
## .. ..$ autores :List of 1
```

```
## .. ..$ autores-endogeno:List of 1
```

```
## ..$ CAPITULO_DE_LIVRO: 'data.frame': 9 obs. of 13 variables:
```

```

## ..$ tipo : chr [1:9] "Capítulo de livro publicado" "Capítulo de livro publica
## ..$ titulo_do_capitulo : chr [1:9] "Knee Angle Estimation Algorithm for Myoelectric Contro
## ..$ titulo_do_livro : chr [1:9] "Biomedical Engineering Systems and Technologies" "Self
## ..$ ano : chr [1:9] "2010" "2011" "2012" "2012" ...
## ..$ doi : chr [1:9] "" "" "" "" ...
## ..$ pais_de_publicacao : chr [1:9] "Brasil" "Croácia" "Brasil" "Brasil" ...
## ..$ isbn : chr [1:9] "9783642117" "9789533075464" "9788563262028" "978888947
## ..$ nome_da_editora : chr [1:9] "Springer-Verlag" "Intech" "São Miguel" "Editore Marett
## ..$ numero_da_edicao_revisao: chr [1:9] "" "" "00" "oo" ...
## ..$ organizadores : chr [1:9] "Ana Fred; Joaquim Filipe; Hugo Gamboa" "Josphat Igadwa
## ..$ paginas : chr [1:9] "124 - 135" "401 - 424" "50 - 55" "180 - 181" ...
## ..$ autores :List of 9
## ..$ autores-endogeno :List of 9
## ..$ EVENTO : 'data.frame': 49 obs. of 11 variables:
## ..$ natureza : chr [1:49] "COMPLETO" "COMPLETO" "COMPLETO" "COMPLETO" ...
## ..$ titulo : chr [1:49] "Automatic Tracking of Innervation Zones Using Image Processing
## ..$ nome_do_evento : chr [1:49] "ISSNIP Biosignals and Biorobotics Conference 2010" "31st Annu
## ..$ ano_do_trabalho : chr [1:49] "2010" "2010" "2010" "2010" ...
## ..$ pais_do_evento : chr [1:49] "Brasil" "Canadá" "Estados Unidos" "Brasil" ...
## ..$ cidade_do_evento: chr [1:49] "Vitória" "Minneapolis" "Minneapolis" "Fortaleza" ...
## ..$ doi : chr [1:49] "" "" "" "" ...
## ..$ classificacao : chr [1:49] "INTERNACIONAL" "INTERNACIONAL" "INTERNACIONAL" "NACIONAL" ...
## ..$ paginas : chr [1:49] "214 - 218" "428 - 431" "225 - 229" "1 - 10" ...
## ..$ autores :List of 49
## ..$ autores-endogeno:List of 49
## ..$ PERIODICO : 'data.frame': 22 obs. of 10 variables:
## ..$ natureza : chr [1:22] "COMPLETO" "COMPLETO" "COMPLETO" "COMPLETO" ...
## ..$ titulo : chr [1:22] "Prótese para controle de fluxo esofagiano como nova técnica p
## ..$ periodico : chr [1:22] "Revista Brasileira de Engenharia Biomédica (Impresso)" "BioMed
## ..$ ano : chr [1:22] "2010" "2010" "2011" "2011" ...
## ..$ volume : chr [1:22] "26" "1" "3" "32" ...
## ..$ issn : chr [1:22] "15173151" "1475925X" "21507988" "09673334" ...
## ..$ paginas : chr [1:22] "49 - 54" "1 - 28" "202 - 209" "543 - 557" ...
## ..$ doi : chr [1:22] "" "10.1186/1475-925X-9-5" "" "10.1088/0967-3334/32/5/004" ...
## ..$ autores :List of 22
## ..$ autores-endogeno:List of 22
## $ orientacoes_academicas:List of 6
## ..$ ORIENTACAO_CONCLUIDA_DOUTORADO : 'data.frame': 4 obs. of 13 variables:
## ..$ natureza : chr [1:4] "Tese de doutorado" "Tese de doutorado" "Tese de dou
## ..$ titulo : chr [1:4] "Influência da Eletroestimulação Neuromuscular de Ba
## ..$ ano : chr [1:4] "2011" "2013" "2013" "2014"
## ..$ id_lattes_aluno : chr [1:4] "7129464687368571" "5330755818114960" "" "0441397018
## ..$ nome_aluno : chr [1:4] "Kenia Fonseca Pires" "Fabiano Araujo Soares" "Maria
## ..$ instituicao : chr [1:4] "Universidade de Brasília" "Universidade de Brasília
## ..$ curso : chr [1:4] "Ciências Médicas" "ENGENHARIA DE SISTEMAS ELETRÔNICO
## ..$ codigo_do_curso : chr [1:4] "60021152" "60057840" "51500140" "60021152"
## ..$ bolsa : chr [1:4] "SIM" "SIM" "SIM" "SIM"
## ..$ agencia_financiadora : chr [1:4] "Conselho Nacional de Desenvolvimento Científico e T
## ..$ codigo_agencia_financiadora: chr [1:4] "002200000000" "045000000000" "045000000000" "002200
## ..$ nome_orientadores :List of 4
## ..$ id_lattes_orientadores :List of 4
## ..$ ORIENTACAO_CONCLUIDA_MESTRADO : 'data.frame': 9 obs. of 13 variables:
## ..$ natureza : chr [1:9] "Dissertação de mestrado" "Dissertação de mestrado"
## ..$ titulo : chr [1:9] "O Efeito da Estimulação Elétrica Neuromuscular no Pa

```

```

## ..$ ano : chr [1:9] "2010" "2010" "2011" "2011" ...
## ..$ id_lattes_aluno : chr [1:9] "3215508254397316" "" "5003191281650601" "" ...
## ..$ nome_aluno : chr [1:9] "Leina Adriana Barbosa Pimenta" "Patrícia Mendonça C
## ..$ instituicao : chr [1:9] "Universidade de Brasília" "Universidade de Brasília
## ..$ curso : chr [1:9] "Ciências Médicas" "ENGENHARIA ELÉTRICA" "Ciências M
## ..$ codigo_do_curso : chr [1:9] "60021152" "60057831" "60021152" "60059672" ...
## ..$ bolsa : chr [1:9] "SIM" "NAO" "NAO" "NAO" ...
## ..$ agencia_financiadora : chr [1:9] "Coordenação de Aperfeiçoamento de Pessoal de Nível S
## ..$ codigo_agencia_financiadora: chr [1:9] "045000000000" "" "" "" ...
## ..$ nome_orientadores :List of 9
## ..$ id_lattes_orientadores :List of 9
## ..$ ORIENTACAO_EM_ANDAMENTO_DOUTORADO :'data.frame': 3 obs. of 13 variables:
## ..$ natureza : chr [1:3] "Tese de doutorado" "Tese de doutorado" "Tese de dou
## ..$ titulo : chr [1:3] "Desenvolvimento de Sistema de Assistência à Tosse en
## ..$ ano : chr [1:3] "2015" "2015" "2015"
## ..$ id_lattes_aluno : chr [1:3] "0180507585649457" "1289651463334045" "5003191281650
## ..$ nome_aluno : chr [1:3] "Felipe Soares Macedo" "Cristina Akemi Shimoda Uechi
## ..$ instituicao : chr [1:3] "Universidade de Brasília" "Universidade de Brasília
## ..$ curso : chr [1:3] "Ciências Médicas" "ENGENHARIA ELÉTRICA" "Ciências M
## ..$ codigo_do_curso : chr [1:3] "60021152" "60057831" "60021152"
## ..$ bolsa : chr [1:3] "SIM" "NAO" "NAO"
## ..$ agencia_financiadora : chr [1:3] "Coordenação de Aperfeiçoamento de Pessoal de Nível S
## ..$ codigo_agencia_financiadora: chr [1:3] "045000000000" "" ""
## ..$ nome_orientadores :List of 3
## ..$ id_lattes_orientadores :List of 3
## ..$ ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA:'data.frame': 2 obs. of 13 variables:
## ..$ natureza : chr [1:2] "Iniciação Científica" "Iniciação Científica"
## ..$ titulo : chr [1:2] "Desenvolvimento de Metodologia para Controle de Sis
## ..$ ano : chr [1:2] "2017" "2017"
## ..$ id_lattes_aluno : chr [1:2] "" ""
## ..$ nome_aluno : chr [1:2] "Rafael Zamarion Campagnoli" "Luiza Irina Lima dos S
## ..$ instituicao : chr [1:2] "Universidade de Brasília" "Universidade de Brasília
## ..$ curso : chr [1:2] "Engenharia Elétrica" "Engenharia Eletrônica"
## ..$ codigo_do_curso : chr [1:2] "60176172" "60339870"
## ..$ bolsa : chr [1:2] "NAO" "SIM"
## ..$ agencia_financiadora : chr [1:2] "" "Conselho Nacional de Desenvolvimento Científico
## ..$ codigo_agencia_financiadora: chr [1:2] "" "002200000000"
## ..$ nome_orientadores :List of 2
## ..$ id_lattes_orientadores :List of 2
## ..$ ORIENTACAO_EM_ANDAMENTO_MESTRADO :'data.frame': 4 obs. of 13 variables:
## ..$ natureza : chr [1:4] "Dissertação de mestrado" "Dissertação de mestrado"
## ..$ titulo : chr [1:4] "Sistema para proteção contra formação de fístulas á
## ..$ ano : chr [1:4] "2015" "2017" "2017" "2017"
## ..$ id_lattes_aluno : chr [1:4] "5235360070351853" "0603371215993939" "7212628088041
## ..$ nome_aluno : chr [1:4] "Sylvia de Sousa Faria" "Pedro Henrique Gonçalves Ina
## ..$ instituicao : chr [1:4] "Universidade de Brasília" "Universidade de Brasília
## ..$ curso : chr [1:4] "Engenharia Biomédica" "Engenharia Biomédica" "Engen
## ..$ codigo_do_curso : chr [1:4] "60059672" "60059672" "60059672" "60059672"
## ..$ bolsa : chr [1:4] "SIM" "SIM" "SIM" "SIM"
## ..$ agencia_financiadora : chr [1:4] "Fundação de Apoio à Pesquisa do Distrito Federal" "
## ..$ codigo_agencia_financiadora: chr [1:4] "786500000001" "001200000991" "045000000000" "045000
## ..$ nome_orientadores :List of 4
## ..$ id_lattes_orientadores :List of 4
## ..$ OUTRAS_ORIENTACOES_CONCLUIDAS :'data.frame': 6 obs. of 13 variables:

```

```
## ..$ natureza : chr [1:6] "MONOGRAFIA_DE_CONCLUSAO_DE_CURSO_APERFEICOAMENTO_E_
## ..$ titulo : chr [1:6] "Ricardo Minotto" "Eletrocardiograma Portátil Basead
## ..$ ano : chr [1:6] "2011" "2012" "2012" "2015" ...
## ..$ id_lattes_aluno : chr [1:6] "" "" "" "" ...
## ..$ nome_aluno : chr [1:6] "Ricardo Minotto" "Vinicius do Couto Pinheiro" "Rafa
## ..$ instituicao : chr [1:6] "Universidade de Brasília" "Universidade de Brasília
## ..$ curso : chr [1:6] "Especialização em Engenharia Clínica" "Engenharia E
## ..$ codigo_do_curso : chr [1:6] "90000004" "60176172" "60176172" "60339870" ...
## ..$ bolsa : chr [1:6] "NAO" "NAO" "NAO" "NAO" ...
## ..$ agencia_financiadora : chr [1:6] "" "" "" "" ...
## ..$ codigo_agencia_financiadora: chr [1:6] "" "" "" "" ...
## ..$ nome_orientadores :List of 6
## ..$ id_lattes_orientadores :List of 6
## $ senioridade : chr "9"
```

Verificando engenharia elétrica p8

```
unb.prof.ele_p8 <- fromJSON("dados-2018-2/engenharia-eletrica-p8/281.profile.json")
```

A quantidade de docentes sob análise é apresentada a seguir.

```
length(unb.prof.ele_p8)
```

```
## [1] 11
```

Verificando oitavo elemento.

```
### Usando glimpse
```

```
glimpse(unb.prof.ele_p8[[8]], width = 30)
```

```
## List of 7
## $ nome : chr "Anderson Clayton Alves Nascimento"
## $ resumo_cv : chr "Possui graduação em Engenharia Elétrica pela Universidade de Brasília
## $ areas_de_atuacao : 'data.frame': 3 obs. of 4 variables:
## ..$ grande_area : chr [1:3] "ENGENHARIAS" "ENGENHARIAS" "ENGENHARIAS"
## ..$ area : chr [1:3] "Engenharia Elétrica" "Engenharia Elétrica" "Engenharia Elétrica"
## ..$ sub_area : chr [1:3] "Criptografia" "Criptografia" "Criptografia"
## ..$ especialidade: chr [1:3] "Sistemas Incondicionalmente Seguros e Criptografia Quântica" "Comput
## $ endereco_profissional :List of 8
## ..$ instituicao: chr "University of Washington Tacoma"
## ..$ orgao : chr "Institute of Technology"
## ..$ unidade : chr ""
## ..$ DDD : chr "1"
## ..$ telefone : chr "25369259"
## ..$ bairro : chr "NA"
## ..$ cep : chr ""
## ..$ cidade : chr "Tacoma"
## $ producao_bibliografica :List of 5
## ..$ ARTIGO_ACEITO : 'data.frame': 1 obs. of 10 variables:
## ..$ natureza : chr "NAO_INFORMADO"
## ..$ titulo : chr "On the Oblivious Transfer Capacity of Generalized Erasure Channels a
## ..$ periodico : chr "IEEE TRANSACTIONS ON INFORMATION THEORY"
## ..$ ano : chr "2017"
## ..$ volume : chr ""
## ..$ issn : chr "00189448"
## ..$ paginas : chr " - "
## ..$ doi : chr ""
```

```

## ..$ autores :List of 1
## ..$ autores-endogeno:List of 1
## ..$ CAPITULO_DE_LIVRO:'data.frame': 1 obs. of 13 variables:
## ..$ tipo : chr "Capítulo de livro publicado"
## ..$ titulo_do_capitulo : chr "Análise de Malware: Investigação de Códigos Maliciosos através de..."
## ..$ titulo_do_livro : chr "Minicursos do XI Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais"
## ..$ ano : chr "2011"
## ..$ doi : chr ""
## ..$ pais_de_publicacao : chr "Brasil"
## ..$ isbn : chr "9788576692591"
## ..$ nome_da_editora : chr "Sociedade Brasileira de Computação"
## ..$ numero_da_edicao_revisao: chr "1"
## ..$ organizadores : chr "Nascimento, A.C. A.; De Sousa Jr., R. T.; Faleiros, A. C.; R..."
## ..$ paginas : chr "9 - 52"
## ..$ autores :List of 1
## ..$ autores-endogeno :List of 1
## ..$ EVENTO : 'data.frame': 22 obs. of 11 variables:
## ..$ natureza : chr [1:22] "COMPLETO" "COMPLETO" "COMPLETO" "COMPLETO" ...
## ..$ titulo : chr [1:22] "Oblivious Transfer Based on the McEliece Assumptions with Unc..."
## ..$ nome_do_evento : chr [1:22] "O Simpósio Brasileiro em Segurança da Informação e de Sistemas..."
## ..$ ano_do_trabalho : chr [1:22] "2010" "2010" "2010" "2010" ...
## ..$ pais_do_evento : chr [1:22] "Brasil" "Brasil" "Alemanha" "Alemanha" ...
## ..$ cidade_do_evento: chr [1:22] "Fortaleza" "Fortaleza" "Cheju Island, Korea" "Cheju Island, Korea" ...
## ..$ doi : chr [1:22] "" "" "" "" ...
## ..$ classificacao : chr [1:22] "NACIONAL" "INTERNACIONAL" "INTERNACIONAL" "INTERNACIONAL" ...
## ..$ paginas : chr [1:22] "1 - 12" " " "337 - 350" "307 - 321" ...
## ..$ autores :List of 22
## ..$ autores-endogeno:List of 22
## ..$ LIVRO : 'data.frame': 3 obs. of 13 variables:
## ..$ titulo : chr [1:3] "ANAIS DO 11 SIMPÓSIO BRASILEIRO DE SEGURANÇA DA INFORMAÇÃO E DE SISTEMAS COMPUTACIONAIS"
## ..$ ano : chr [1:3] "2011" "2016" "2016"
## ..$ tipo : chr [1:3] "LIVRO_ORGANIZADO_OU_EDICAO" "LIVRO_ORGANIZADO_OU_EDICAO" "LIVRO_ORGANIZADO_OU_EDICAO"
## ..$ natureza : chr [1:3] "ANAIS" "ANAIS" "ANAIS"
## ..$ pais_de_publicacao : chr [1:3] "Brasil" "Alemanha" "Alemanha"
## ..$ isbn : chr [1:3] "978-85-7669-2" "9783319491745" "9783319458700"
## ..$ doi : chr [1:3] "" "10.1007/978-3-319-49175-2" "10.1007/978-3-319-45871-2"
## ..$ nome_da_editora : chr [1:3] "Sociedade Brasileira de Computação" "Springer International Publishing" "Springer International Publishing"
## ..$ numero_da_edicao_revisao: chr [1:3] "11" "1" "1"
## ..$ numero_de_paginas : chr [1:3] "414" "" "342"
## ..$ numero_de_volumes : chr [1:3] "1" "1" "1"
## ..$ autores :List of 3
## ..$ autores-endogeno :List of 3
## ..$ PERIODICO : 'data.frame': 16 obs. of 10 variables:
## ..$ natureza : chr [1:16] "COMPLETO" "COMPLETO" "COMPLETO" "COMPLETO" ...
## ..$ titulo : chr [1:16] "A Fault Analytic Method against HB+" "Universally Composable and..."
## ..$ periodico : chr [1:16] "IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences"
## ..$ ano : chr [1:16] "2011" "2011" "2011" "2011" ...
## ..$ volume : chr [1:16] "E94A" "E94A" "57" "6" ...
## ..$ issn : chr [1:16] "09168508" "09168508" "00189448" "18099807" ...
## ..$ paginas : chr [1:16] "885 - 859" "725 - 734" "5566 - 5571" "8 - 27" ...
## ..$ doi : chr [1:16] "" "" "" "10.5769/J201101001" ...
## ..$ autores :List of 16
## ..$ autores-endogeno:List of 16
## $ orientacoes_academicas:List of 1

```



```
## ..$ ORIENTACAO_CONCLUIDA_MESTRADO:'data.frame': 13 obs. of 13 variables:
## ..$ natureza : chr [1:13] "Dissertação de mestrado" "Dissertação de mestrado"
## ..$ titulo : chr [1:13] "ELABORAC&#807;A&#771;O DE UM PROTOCOLO EFICIENTE D
## ..$ ano : chr [1:13] "2010" "2010" "2010" "2011" ...
## ..$ id_lattes_aluno : chr [1:13] "4682614800956953" "0166522145556051" "" "754565644
## ..$ nome_aluno : chr [1:13] "Vinicius de Moraes Alves" "Rafael Tonicelli de Mel
## ..$ instituicao : chr [1:13] "Universidade de Brasília" "Universidade de Brasília
## ..$ curso : chr [1:13] "Engenharia Elétrica" "Engenharia Elétrica" "Engenh
## ..$ codigo_do_curso : chr [1:13] "51500140" "51500140" "51500140" "51500140" ...
## ..$ bolsa : chr [1:13] "NAO" "NAO" "NAO" "NAO" ...
## ..$ agencia_financiadora : chr [1:13] "" "" "" "" ...
## ..$ codigo_agencia_financiadora: chr [1:13] "" "" "" "" ...
## ..$ nome_orientadores :List of 13
## ..$ id_lattes_orientadores :List of 13
## $ senioridade : chr "8"
```

Verificando engenharia elétrica p7

```
unb.prof.ele_p7 <- fromJSON("dados-2018-2/engenharia-eletrica-p7/282.profile.json")
```

A quantidade de docentes sob análise é apresentada a seguir.

```
length(unb.prof.ele_p7)
```

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

Verificando sétimo elemento.

```
### Usando glimpse
```

```
glimpse(unb.prof.ele_p7[[7]], width = 30)
```

```
## List of 7
```

```
## $ nome : chr "Sébastien Roland Marie Joseph RONDINEAU"
## $ resumo_cv : chr "- IEEE Senior Member. - Possui Mestrado em Ciencias da Comput., Proces
## $ areas_de_atuacao : 'data.frame': 5 obs. of 4 variables:
## ..$ grande_area : chr [1:5] "ENGENHARIAS" "ENGENHARIAS" "CIENCIAS_EXATAS_E_DA_TERRA" "CIENCIAS_EX
## ..$ area : chr [1:5] "Engenharia Elétrica" "Engenharia Elétrica" "Matemática" "Ciência da C
## ..$ sub_area : chr [1:5] "Telecomunicações" "Telecomunicações" "Matemática Aplicada" "Matemáti
## ..$ especialidade: chr [1:5] "Teoria Eletromagnetica, Microondas, Propagação de Ondas, Antenas" "S
## $ endereco_profissional :List of 8
## ..$ instituicao: chr "Universidade de Brasília"
## ..$ orgao : chr "Campus Gama"
## ..$ unidade : chr ""
## ..$ DDD : chr "51"
## ..$ telefone : chr "91109009"
## ..$ bairro : chr "Setor Leste (Gama)"
## ..$ cep : chr "72444240"
## ..$ cidade : chr "Brasília"
## $ producao_bibliografica :List of 2
## ..$ EVENTO : 'data.frame': 1 obs. of 11 variables:
## ..$ natureza : chr "COMPLETO"
## ..$ titulo : chr "The NSPR, a Brazilian Radio with Capabilities that Fully Realize Mob
## ..$ nome_do_evento : chr "Simposio Internacional de Guerra Eletronica, XIII"
## ..$ ano_do_trabalho : chr "2012"
## ..$ pais_do_evento : chr "Brasil"
## ..$ cidade_do_evento: chr "Sao Jose dos Campos, sp"
## ..$ doi : chr ""
```

```

## .. .$ classificacao : chr "INTERNACIONAL"
## .. .$ paginas : chr " - "
## .. .$ autores :List of 1
## .. .$ autores-endogeno:List of 1
## .. $ PERIODICO:'data.frame': 3 obs. of 10 variables:
## .. .$ natureza : chr [1:3] "COMPLETO" "COMPLETO" "COMPLETO"
## .. .$ titulo : chr [1:3] "Micro-Coaxial Impedance Transformers" "Array Pattern Synthesis
## .. .$ periodico : chr [1:3] "IEEE Transactions on Microwave Theory and Techniques" "IEEE Tr
## .. .$ ano : chr [1:3] "2010" "2016" "2017"
## .. .$ volume : chr [1:3] "58" "64" "65"
## .. .$ issn : chr [1:3] "00189480" "0018926X" "0018926X"
## .. .$ paginas : chr [1:3] "2908 - 2914" "4228 - 4234" "1 - 1"
## .. .$ doi : chr [1:3] "10.1109/TMTT.2010.2078410" "10.1109/TAP.2016.2594300" "10.1109,
## .. .$ autores :List of 3
## .. .$ autores-endogeno:List of 3
## $ orientacoes_academicas:List of 3
## .. $ ORIENTACAO_EM_ANDAMENTO_GRADUACAO :'data.frame': 3 obs. of 13 variables:
## .. .$ natureza : chr [1:3] "Trabalho de conclusão de curso de graduação" "Traba
## .. .$ titulo : chr [1:3] "não reciprocidade amagnetica e isolamento baseadas em
## .. .$ ano : chr [1:3] "2016" "2016" "2017"
## .. .$ id_lattes_aluno : chr [1:3] "" "" ""
## .. .$ nome_aluno : chr [1:3] "Eduardo Gonçalves Sousa" "Jeann Feitosa" "Marconi C
## .. .$ instituicao : chr [1:3] "Universidade de Brasília" "Universidade de Brasília
## .. .$ curso : chr [1:3] "Abi - Engenharia" "Abi - Engenharia" "Engenharia Me
## .. .$ codigo_do_curso : chr [1:3] "60192089" "60192089" "60072377"
## .. .$ bolsa : chr [1:3] "NAO" "SIM" "NAO"
## .. .$ agencia_financiadora : chr [1:3] "" "Fundação de Apoio à Pesquisa do Distrito Federal
## .. .$ codigo_agencia_financiadora: chr [1:3] "" "786500000001" ""
## .. .$ nome_orientadores :List of 3
## .. .$ id_lattes_orientadores :List of 3
## .. $ ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA:'data.frame': 3 obs. of 13 variables:
## .. .$ natureza : chr [1:3] "Iniciação Científica" "Iniciação Científica" "Inici
## .. .$ titulo : chr [1:3] "Iniciação ao desenvolvimento do sistema irradiante
## .. .$ ano : chr [1:3] "2016" "2016" "2016"
## .. .$ id_lattes_aluno : chr [1:3] "" "" ""
## .. .$ nome_aluno : chr [1:3] "Jalusa Maria da Silva Ferrari" "Paulo Fillipe Rodri
## .. .$ instituicao : chr [1:3] "Universidade de Brasília" "Universidade de Brasília
## .. .$ curso : chr [1:3] "Abi - Engenharia" "Engenharia Automotiva" "Abi - Eng
## .. .$ codigo_do_curso : chr [1:3] "60192089" "60095938" "60192089"
## .. .$ bolsa : chr [1:3] "SIM" "SIM" "SIM"
## .. .$ agencia_financiadora : chr [1:3] "Fundação de Apoio à Pesquisa do Distrito Federal" "
## .. .$ codigo_agencia_financiadora: chr [1:3] "786500000001" "786500000001" "786500000001"
## .. .$ nome_orientadores :List of 3
## .. .$ id_lattes_orientadores :List of 3
## .. $ 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] "Circulateurs RF construits avec des boucles résonnan
## .. .$ ano : chr [1:2] "2017" "2017"
## .. .$ id_lattes_aluno : chr [1:2] "" ""
## .. .$ nome_aluno : chr [1:2] "Valentin Favard" "Eduardo Gonçalves"
## .. .$ instituicao : chr [1:2] "Université de Limoges" "Universidade de Brasília"
## .. .$ curso : chr [1:2] "Faculté des Sciences" "ENGENHARIA ELÉTRICA"
## .. .$ codigo_do_curso : chr [1:2] "90000015" "60057831"
## .. .$ bolsa : chr [1:2] "SIM" "NAO"

```

```

##   .. ..$ agencia_financiadora      : chr [1:2] "Ministère de La Recherche" ""
##   .. ..$ codigo_agencia_financiadora: chr [1:2] "000400000997" ""
##   .. ..$ nome_orientadores         :List of 2
##   .. ..$ id_lattes_orientadores    :List of 2
##   $ senioridade                    : chr "9"

####Potencial de utilização dos dados do perfil dos docentes

####Descrição dos dados de orientações

unb.adv.biomed <- fromJSON("dados-2018-2/engenharia-biomedica/279.advise.json")
# Mostrando as listas presentes neste arquivo.
names(unb.adv.biomed)

## [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"

# Explorando um nível de detalhe de Orientações de doutorados concluídas
names(unb.adv.biomed$ORIENTACAO_CONCLUIDA_DOUTORADO)

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

##### DOUTORADO 2017
#Buscando cursos que mais produziram doutorados.
head(sort(table(unb.adv.biomed$ORIENTACAO_CONCLUIDA_DOUTORADO$`2017`$curso), decreasing = TRUE), 10)

##
## ENGENHARIA DE SISTEMAS ELETRÔNICOS E DE AUTOMAÇÃO
##                                     1
##           Pós-graduação em Ciências Médicas
##                                     1

#Sabendo suas instituições
head(sort(table(unb.adv.biomed$ORIENTACAO_CONCLUIDA_DOUTORADO$`2017`$instituicao), decreasing = TRUE), 10)

##
## Universidade de Brasília - Campus Darcy Ribeiro
##                                     1
##           Universidade de Brasília
##                                     1

#Sabendo seus orientadores
data_orienM <- capture.output(str(unb.adv.biomed$ORIENTACAO_CONCLUIDA_DOUTORADO$`2017`$nome_orientadores))
unique(data_orienM)

## [1] "List of 2"
## [2] " $ : chr \"Marcus Vinícius Chaffim Costa\""
## [3] " $ : chr \"Marilia Miranda Forte Gomes\""

##### MESTRADO 2017
#Buscando cursos que mais produziram mestrados.

```

```

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

##
##          Engenharia Biomédica Mestrado em Engenharia Biomédica
##                      9                      4
##          ENGENHARIA ELÉTRICA                      Física
##                      1                      1
##          PPDSCI/CEAM
##                      1

head(sort(table(unb.adv.biomed$ORIENTACAO_CONCLUIDA_MESTRADO$`2017`$instituicao), decreasing = TRUE), 10)

##
##          Faculdade do Gama da UnB
##                      6
##          Universidade de Brasília
##                      5
##          Universidade de Brasília - Faculdade UnB-Gama
##                      3
##          Faculdade UNB Gama - FGA
##                      1
##          Universidade de Brasília - Campus Darcy Ribeiro
##                      1

#Sabendo suas instituições
head(sort(table(unb.adv.biomed$ORIENTACAO_CONCLUIDA_MESTRADO$`2017`$instituicao), decreasing = TRUE), 10)

##
##          Faculdade do Gama da UnB
##                      6
##          Universidade de Brasília
##                      5
##          Universidade de Brasília - Faculdade UnB-Gama
##                      3
##          Faculdade UNB Gama - FGA
##                      1
##          Universidade de Brasília - Campus Darcy Ribeiro
##                      1

#Sabendo seus orientadores
data_orien <- capture.output(str(unb.adv.biomed$ORIENTACAO_CONCLUIDA_MESTRADO$`2017`$nome_orientadores))
unique(data_orien)

## [1] "List of 16"
## [2] " $ : chr \"Georges Daniel Amvame Nze\""
## [3] " $ : chr \"Jose Felicio da Silva\""
## [4] " $ : chr \"Lourdes Mattos Brasil\""
## [5] " $ : chr \"Marcelino Monteiro de Andrade\""
## [6] " $ : chr \"Marilia Miranda Forte Gomes\""
## [7] " $ : chr \"Ronni Geraldo Gomes de Amorim\""

```

Como se pode perceber apenas u, professor que orientou doutorado concluído em 2017 (“Marilia Miranda Forte Gomes”) também fez parte dos professores que orientaram no mestrado:

- “Georges Daniel Amvame Nze”
- “Jose Felicio da Silva”
- “Lourdes Mattos Brasil”

- “Marcelino Monteiro de Andrade”
- “Marilia Miranda Forte Gomes”
- “Ronni Geraldo Gomes de Amorim”

Sendo o professor “Marcus Vinícius Chaffim Costa” responsável apenas por orientações de doutorado em 2017.

Descrição dos dados list.json

Descrição de biomedicina

```
unb.list.biomed <- fromJSON("dados-2018-2/engenharia-biomedica/279.list.json")
#analizando a quantidade de elementos presente em list.json
length(unb.list.biomed$fiocruz$id)
```

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

```
# Mostrando alguns id, verificar que eles são diferentes
d_listID <- unb.list.biomed$fiocruz$id
unique(d_listID)
```

```
## [1] "1141716826787805" "0535100751136568" "5330755818114960"
## [4] "5810353896294133" "2957228356035337" "0201204222182378"
## [7] "9190489069187153" "4739013535126469" "4839052902231824"
## [10] "1524924375222848" "9169095482512290" "4086384842130773"
## [13] "5928104758017036" "1154673226500318" "7294738832905991"
```

```
#Mostrando nome sempre igual ""
d_listNO <- unb.list.biomed$fiocruz$nome
unique(d_listNO)
```

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

```
#Mostrando periodo sempre igual 2010-2017
d_listDT <- unb.list.biomed$fiocruz$periodo
unique(d_listDT)
```

```
## [[1]]
## [1] "2010" "2017"
```

Feito a análise, percebe-se que apenas o campo id que muda, tanto “**nome**” (obtendo sempre o valor “”) e “**periodo**”(obtendo sempre o valor[“2010”,“2017”]), por isso considera-se este arquivo JSON como inútil, haja visto que não matemática para obter quaisquer dado plausível, somente os id.

Descrição de engenharia de automação

```
unb.list.auto <- fromJSON("dados-2018-2/engenharia-eletronicos-automacao/280.list.json")
#analizando a quantidade de elementos presente em list.json
length(unb.list.auto$fiocruz$id)
```

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

```
# Mostrando alguns id, verificar que eles são diferentes
d_listID <- unb.list.auto$fiocruz$id
unique(d_listID)
```

```
## [1] "0301021863146083" "1141716826787805" "5202540363896199"
## [4] "6959718976106778" "2065104260659969" "2443108673822686"
## [7] "4442238042952234" "3269842674142738" "6103264336506171"
## [10] "8850603018835560" "8032871065166582" "6509594616230305"
```

```
## [13] "5224996059295555" "1641340120936086" "0335058635088347"
## [16] "4465619366143200" "9986885788660516" "7284320756124645"
## [19] "9271885452744905" "7445354319753187" "3740757562716147"
```

```
#Mostrando nome sempre igual ""
d_listNO <- unb.list.auto$fiocruz$nome
unique(d_listNO)
```

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

```
#Mostrando periodo sempre igual 2010-2017
d_listDT <- unb.list.auto$fiocruz$periodo
unique(d_listDT)
```

```
## [[1]]
## [1] "2010" "2017"
```

Novamente, considera-se este arquivo JSON como inútil, haja visto que não matemática para obter quaisquer dado plausível, somente os id. ##### Descrição de engenharia elétrica p8

```
unb.list.ele_p8 <- fromJSON("dados-2018-2/engenharia-eletrica-p8/281.list.json")
#analizando a quantidade de elementos presente em list.json
length(unb.list.ele_p8$fiocruz$id)
```

```
## [1] 11
```

```
# Mostrando alguns id, verificar que eles são diferentes
d_listID <- unb.list.ele_p8$fiocruz$id
unique(d_listID)
```

```
## [1] "4438697154081039" "1521815731111261" "5453814076903614"
## [4] "3636638127781726" "1808368824068028" "9788199690491510"
## [7] "2217050272348357" "5810353896294133" "1786889674911887"
## [10] "3196088341529197" "3408305339297459"
```

```
#Mostrando nome sempre igual ""
d_listNO <- unb.list.ele_p8$fiocruz$nome
unique(d_listNO)
```

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

```
#Mostrando periodo sempre igual 2010-2017
d_listDT <- unb.list.ele_p8$fiocruz$periodo
unique(d_listDT)
```

```
## [[1]]
## [1] "2010" "2017"
```

Sendo este arquivo JSON sem qualquer significado, haja visto que não matemática para obter quaisquer dado plausível, somente os id.

Descrição de engenharia elétrica p7

```
unb.list.ele_p7 <- fromJSON("dados-2018-2/engenharia-eletrica-p7/282.list.json")
#analizando a quantidade de elementos presente em list.json
length(unb.list.ele_p7$fiocruz$id)
```

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

```
# Mostrando alguns id, verificar que eles são diferentes
d_listID <- unb.list.ele_p7$fiocruz$id
unique(d_listID)

## [1] "8999676470821349" "1573568521807396" "3636638127781726"
## [4] "3155646870666658" "7997341055520362" "1786889674911887"
## [7] "8795500242651581" "9279262545449206" "0997629063334821"
## [10] "8709856465061922" "9868233765380165" "3196088341529197"
## [13] "7293346445799114" "8701122785988842"
```

```
#Mostrando nome sempre igual ""
d_listNO <- unb.list.ele_p7$fiocruz$nome
unique(d_listNO)
```

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

```
#Mostrando periodo sempre igual 2010-2017
d_listDT <- unb.list.ele_p7$fiocruz$periodo
unique(d_listDT)
```

```
## [[1]]
## [1] "2010" "2017"
```

Novamente, arquivo JSON sem qualquer significado, haja visto que não matemática para obter quaisquer dado plausível, somente os id.

Descrição dos dados de Produção bibliográfica

TODO LIST : VERIFICAR

- Engenharia p7
- Engenharia p8
- automação

Verificando engenharia biomédica

```
unb.pub.biomed <- fromJSON("dados-2018-2/engenharia-biomedica/279.publication.json")
# Verificando os tipos de produções que existe.
names(unb.pub.biomed)
```

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

```
#Analisando o qual tipo de informação se tem em periódicos no ano 2012.
names(unb.pub.biomed$PERIODICO$`2012`)
```

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

```
#Nomes dos periódicos em que tiveram mais publicações na área de engenharia Biomédica.
head(sort(table(unb.pub.biomed$PERIODICO$`2017`$periodico), decreasing = TRUE), 10)
```

```
##
## International Journal of Biosensors & Bioelectronics
## 2
## Revista Brasileira de Ensino de Física (Online)
## 2
## ACTA PAUL DE ENFERM
## 1
## ADOLESCÊNCIA & SAÚDE
## 1
## Advanced Materials Letters
## 1
## Advances in High Energy Physics
## 1
## AFRICAN JOURNAL OF BIOTECHNOLOGY
## 1
## ANNALS OF PHYSICS
## 1
## ARTEFACTUM (RIO DE JANEIRO)
## 1
## CADERNO DE FÍSICA DA UEFS
## 1
```

```
#Nomes dos periódicos em que tiveram mais publicações de artigos aceitos na área de engenharia Biomédica
head(sort(table(unb.pub.biomed$ARTIGO_ACEITO$`2017`$periodico), decreasing = TRUE), 10)
```

```
##
## Acta Paulista de Enfermagem
## 1
## Advanced Materials Letters
## 1
## Ciência e Saúde Coletiva (Impresso)
## 1
## IEEE Journal of Biomedical and Health Informatics
## 1
## IEEE Latin America Transactions
## 1
## JOURNAL OF NANOSCIENCE AND NANOTECHNOLOGY
## 1
```

```
#Nomes dos autores que produziram um tipo de Produção que não estava contemplada .
head(sort(table(unb.pub.biomed$DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA$`2011`$autores), decreasing = TRUE), 10)
```

```
##
##
##
## GOMES, Marília Miranda Forte;GOMES, MARÍLIA MIRANDA FORTE;GOMES, MARÍLIA M F;GOMES, Marília Miranda Forte
##
##
##
##
##
##
##
```


Verificando engenharia automação

Verificando engenharia elétrica p8

Verificando engenharia elétrica p7

Descrição dos dados de redes de colaboração

TODO LIST : VERIFICAR

- biomed - explicar algo
- Engenharia p7
- Engenharia p8
- automação

Verificando engenharia biomédica

```
unb.graph.biomed <- fromJSON("dados-2018-2/engenharia-biomedica/279.graph.json")  
# Suas variáveis  
names(unb.graph.biomed)
```

```
## [1] "label" "nodes" "links"
```

```
# Quantidade de nós  
length(unb.graph.biomed$nodes$id)
```

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

```
# Quantidade de links de fonte  
length(unb.graph.biomed$links$source)
```

```
## [1] 28
```

```
# Quantidade de links de chegada  
length(unb.graph.biomed$links$target)
```

```
## [1] 28
```

```
#exemplos de pesos da aresta  
str(unb.graph.biomed$links$weight)
```

```
## chr [1:28] "1" "3" "9" "1" "2" "3" "1" "1" "5" "15" "1" "3" "9" "1" ...
```

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

Arquivo Profile

TODO LIST : VERIFICAR

- Engenharia p7
- Engenharia p8
- automação

Verificando engenharia biomédica

```
# Total de áreas de atuação de todos profissionais  
sum(sapply(unb.prof.biomed, function(x) nrow(x$areas_de_atuacao)))
```

```
## [1] 57
```

```
# número de pessoas por grande area
table(unlist(sapply(unb.prof.biomed, function(x) (x$areas_de_atuacao$grande_area))))
```

```
##
##                                CIENCIAS_BIOLOGICAS
##                                1                      4
##          CIENCIAS_DA_SAUDE  CIENCIAS_EXATAS_E_DA_TERRA
##                                7                      9
## CIENCIAS_SOCIAIS_APLICADAS                                ENGENHARIAS
##                                1                      34
##                                OUTROS
##                                1
```

```
# número de pessoas que produziram os tipos de Produção específico
table(unlist(sapply(unb.prof.biomed, function(x) names(x$producao_bibliografica))))
```

```
##
##                                ARTIGO_ACEITO
##                                7
##          CAPITULO_DE_LIVRO
##                                9
##  DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA
##                                1
##                                EVENTO
##                                14
##                                LIVRO
##                                3
##                                PERIODICO
##                                15
##          TEXTO_EM_JORNAIS
##                                1
```

```
# número de publicações por tipo
```

```
#####ARTIGO ACEITO#####
```

```
sum(sapply(unb.prof.biomed, function(x) length(x$producao_bibliografica$ARTIGO_ACEITO$ano)))
```

```
## [1] 7
```

```
# número de pessoas por quantitativo de produções por pessoa 0 = 1; 1 = 2...
```

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

```
##
## 0 1
## 8 7
```

```
#####CAPITULO DE LIVRO#####
```

```
sum(sapply(unb.prof.biomed, function(x) length(x$producao_bibliografica$CAPITULO_DE_LIVRO$ano)))
```

```
## [1] 47
```

```
# número de pessoas por quantitativo de produções por pessoa 0 = 1; 1 = 2...
```

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

```
##
## 0 1 3 5 9 17
## 6 4 1 1 2 1
```

```
#####LIVRO#####
sum(sapply(unb.prof.biomed, function(x) length(x$producao_bibliografica$LIVRO$ano)))

## [1] 3

# número de pessoas por quantitativo de produções por pessoa 0 = 1; 1 = 2...
table(unlist(sapply(unb.prof.biomed, function(x) length(x$producao_bibliografica$LIVRO$ano))))

##
## 0 1
## 12 3

#####periódico#####
sum(sapply(unb.prof.biomed, function(x) length(x$producao_bibliografica$PERIODICO$ano)))

## [1] 172

# número de pessoas por quantitativo de produções por pessoa 0 = 1; 1 = 2...
table(unlist(sapply(unb.prof.biomed, function(x) length(x$producao_bibliografica$PERIODICO$ano))))

##
## 1 2 3 4 6 8 9 10 11 13 16 22 29 36
## 1 2 1 1 1 1 1 1 1 1 1 1 1 1

#####TEXTO EM JORNAIS#####
sum(sapply(unb.prof.biomed, function(x) length(x$producao_bibliografica$TEXTO_EM_JORNAIS$ano)))

## [1] 1

# número de pessoas por quantitativo de produções por pessoa 0 = 1; 1 = 2...
table(unlist(sapply(unb.prof.biomed, function(x) length(x$producao_bibliografica$TEXTO_EM_JORNAIS$ano))))

##
## 0 1
## 14 1

#####DEMAIS TIPOS#####
sum(sapply(unb.prof.biomed, function(x) length(x$producao_bibliografica$DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRFICA$ano)))

## [1] 1

# número de pessoas por quantitativo de produções por pessoa 0 = 1; 1 = 2...
table(unlist(sapply(unb.prof.biomed, function(x) length(x$producao_bibliografica$DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRFICA$ano))))

##
## 0 1
## 14 1

# número de produções por ano
table(unlist(sapply(unb.prof.biomed, function(x) (x$producao_bibliografica$ARTIGO_ACEITO$ano))))

##
## 2015 2017
## 1 6

table(unlist(sapply(unb.prof.biomed, function(x) (x$producao_bibliografica$CAPITULO_DE_LIVRO$ano))))

##
## 2010 2011 2012 2013 2014 2015 2016 2017
## 1 3 8 2 2 21 5 5
```

```

table(unlist(sapply(unb.prof.biomed, function(x) (x$producao_bibliografica$LIVRO$ano))))

##
## 2015 2016 2017
##    1    1    1
table(unlist(sapply(unb.prof.biomed, function(x) (x$producao_bibliografica$PERIODICO$ano))))

##
## 2010 2011 2012 2013 2014 2015 2016 2017
##    14    9   10   29   18   39   21   32
table(unlist(sapply(unb.prof.biomed, function(x) (x$producao_bibliografica$TEXTOS$ano))))

##
## 2010
##    1
# número de pessoas que realizaram diferentes tipos de orientações
length(unlist(sapply(unb.prof.biomed, function(x) names(x$orientacoes_academicas))))

## [1] 60
# número de pessoas por tipo de orientação
table(unlist(sapply(unb.prof.biomed, function(x) names(x$orientacoes_academicas))))

##
##          ORIENTACAO_CONCLUIDA_DOUTORADO
##                                6
##          ORIENTACAO_CONCLUIDA_MESTRADO
##                                13
##          ORIENTACAO_CONCLUIDA_POS_DOUTORADO
##                                1
##          ORIENTACAO_EM_ANDAMENTO_DOUTORADO
##                                2
##          ORIENTACAO_EM_ANDAMENTO_GRADUACAO
##                                3
## ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA
##                                7
##          ORIENTACAO_EM_ANDAMENTO_MESTRADO
##                                13
##          OUTRAS_ORIENTACOES_CONCLUIDAS
##                                15
#número de orientações concluídas
sum(sapply(unb.prof.biomed, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTRADO$ar

## [1] 109
sum(sapply(unb.prof.biomed, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTORADO$ar

## [1] 14
sum(sapply(unb.prof.biomed, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_POS_DOUTORADO$ar

## [1] 2
# número de pessoas por quantitativo de orientações por pessoa 0 = 1; 1 = 2...
table(unlist(sapply(unb.prof.biomed, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_M

```

```
##
## 0 1 3 6 7 9 11 18 19
## 2 1 2 1 3 3 1 1 1

table(unlist(sapply(unb.prof.biomed, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTORADO))))

##
## 0 1 2 3 4
## 9 2 1 2 1

table(unlist(sapply(unb.prof.biomed, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_POS_DOUTORADO))))

##
## 0 2
## 14 1

# número de orientações por ano
table(unlist(sapply(unb.prof.biomed, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTRADO))))

##
## 2010 2011 2012 2013 2014 2015 2016 2017
## 7 4 10 12 23 24 13 16

table(unlist(sapply(unb.prof.biomed, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTORADO))))

##
## 2011 2012 2013 2014 2015 2017
## 2 2 6 1 1 2

table(unlist(sapply(unb.prof.biomed, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_POS_DOUTORADO))))

##
## 2014 2017
## 1 1
```

Verificando engenharia automação

Verificando engenharia elétrica p8

Verificando engenharia elétrica p7

Arquivo publicação

TODO LIST : VERIFICAR

- Engenharia p7
- Engenharia p8
- automação

Verificando engenharia biomédica

```
#Criando um data-frame com todos os anos
unb.pub.biomed.df <- data.frame()
for (i in 1:length(unb.pub.biomed[[1]]))
  unb.pub.biomed.df <- rbind(unb.pub.biomed.df, unb.pub.biomed$PERIODICO[[i]])
glimpse(unb.pub.biomed.df)
```

```
## Observations: 147
## Variables: 10
```

```
## $ natureza      <chr> "COMPLETO", "COMPLETO", "COMPLETO", "COMPLE...
## $ titulo        <chr> "Prótese para controle de fluxo esofagiano ...
## $ periodico     <chr> "Revista Brasileira de Engenharia Biomédica...
## $ ano           <chr> "2010", "2010", "2010", "2010", "2010", "20...
## $ volume        <chr> "26", "249", "20", "11", "14", "9", "53", "...
## $ issn          <chr> "15173151", "17426596", "10546618", "151942...
## $ paginas       <chr> "49 - 54", "012031 - ", "192 - 200", "23 - ...
## $ doi           <chr> "", "10.1088/1742-6596/249/1/012031", "10.1...
## $ autores       <list> [<"Rosa, S. S. R. F.", "da Rocha, A F", "B...
## $ `autores-endogeno` <list> ["1141716826787805", "0201204222182378", "...

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

## Observations: 147
## Variables: 10
## $ natureza      <chr> "COMPLETO", "COMPLETO", "COMPLETO", "COMPLE...
## $ titulo        <chr> "Prótese para controle de fluxo esofagiano ...
## $ periodico     <chr> "Revista Brasileira de Engenharia Biomédica...
## $ ano           <chr> "2010", "2010", "2010", "2010", "2010", "20...
## $ volume        <chr> "26", "249", "20", "11", "14", "9", "53", "...
## $ issn          <chr> "15173151", "17426596", "10546618", "151942...
## $ paginas       <chr> "49 - 54", "012031 - ", "192 - 200", "23 - ...
## $ doi           <chr> "", "10.1088/1742-6596/249/1/012031", "10.1...
## $ autores       <chr> "Rosa, S. S. R. F.; da Rocha, A F; Brasil, ...
## $ `autores-endogeno` <chr> "1141716826787805", "0201204222182378", "91...

### ARQUIVO PROCESSADO E LIMPO
## publicações por natureza - todas foram completas
table(unb.pub.biomed.df$natureza)

##
## COMPLETO
##      147

## publicações 2010 até 2017
table(unb.pub.biomed.df$ano)

##
## 2010 2011 2012 2013 2014 2015 2016 2017
##    12    8    9   27   14   31   20   26

## publicações por periódico, mostrando os top-5 na área de Biomédica
d <- table(unb.pub.biomed.df$periodico)
head(sort(d,decreasing = TRUE),n=5)

##
##      Revista Brasileira de Engenharia Biomédica (Impresso)
##                                                    7
## Global Journal of Engineering Science and Research Management
##                                                    6
##      Biomedical Engineering Online (Online)
##                                                    4
```

```
##          Revista Brasileira de Ensino de Física (Online)
##                                     4
##          Revista Brasileira de Inovação Tecnológica em Saúde
##                                     4

## Mostrando os autores que tiveram mais publicação
### Neste caso é importante mostra que como a não uniformidade de como se escreve acaba por tornar este
head(sort(table(toupper(unlist(strsplit(unb.pub.biomed.df$autores,";")))),decreasing = TRUE),n=10)

##
##          AMORIM, R. G. G.
##          17
##          AMORIM, R. G. G.
##          16
##          MARÃES, V. R. F. S.
##          14
##          BRASIL, L. M.
##          7
##          MARÃES, V. R. F. S.
##          7
##          SILVA, W. B.
##          7
##          GOMES, MARÍLIA MIRANDA FORTE
##          7
## ROSA, SUÉLIA DE SIQUEIRA RODRIGUES FLEURY
##          7
##          DA ROCHA, A. F.
##          6
##          GOMES, MARÍLIA MIRANDA FORTE
##          6

## Mostrando os autores-endogeno que tiveram mais publicação
head(sort(table(unlist(unb.pub.biomed.df$`autores-endogeno`)),decreasing = TRUE),n=10)

##
## 1154673226500318 4086384842130773 1141716826787805 7294738832905991
##          31          29          13          12
## 9169095482512290 4839052902231824 9190489069187153 0535100751136568
##          10          9          9          4
## 4739013535126469 0201204222182378
##          4          3
```

Arquivo Orientação

TODO LIST : VERIFICAR

- Engenharia p7
- Engenharia p8
- automação

Verificando engenharia biomédica

```
#Orientação
#Visualizar a estrutura do json no painel Viewer
#jsonedit(unb.adv.biomed)
#Reunir todos os anos e orientações concluídas em um mesmo data-frame
unb.adv.biomed.tipo.df <- data.frame(); unb.adv.biomed.df <- data.frame()
```

```

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

## Observations: 124
## Variables: 13
## $ natureza <chr> "Supervisão de pós-doutorado", "Su...
## $ titulo <chr> "", "", "Influência da Eletroestim...
## $ ano <chr> "2014", "2017", "2011", "2011", "2...
## $ id_lattes_aluno <chr> "", "", "7129464687368571", "92327...
## $ nome_aluno <chr> "Leandro Xavier Cardoso", "Glécia ...
## $ instituicao <chr> "Universidade de Brasília - Faculd...
## $ curso <chr> "", "", "Ciências Médicas", "ENGEN...
## $ codigo_do_curso <chr> "", "", "60021152", "60057840", "6...
## $ bolsa <chr> "SIM", "SIM", "SIM", "NAO", "NAO",...
## $ agencia_financiadora <chr> "Coordenação de Aperfeiçoamento de...
## $ codigo_agencia_financiadora <chr> "045000000000", "045000000000", "0...
## $ nome_orientadores <list> ["Lourdes Mattos Brasil", "Lourde...
## $ id_lattes_orientadores <list> ["9190489069187153", "91904890691...

#Transformar as colunas de listas em caracteres eliminando c("")
unb.adv.biomed.df$nome_orientadores <- gsub("\\|c\\(|\\)", "", unb.adv.biomed.df$nome_orientadores)
unb.adv.biomed.df$id_lattes_orientadores <- gsub("\\|c\\(|\\)", "", unb.adv.biomed.df$id_lattes_orientadores)
#Separar as colunas com dois orientadores
unb.adv.biomed.df <- separate(unb.adv.biomed.df, nome_orientadores, into = c("ori1", "ori2"), sep = ",")

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 123 rows [1,
## 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].

# Contando quem tem apenas 1 orientador
sum(is.na(unb.adv.biomed.df$ori2))

## [1] 123

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

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 123 rows [1,
## 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].

# Contando quem tem apenas 1 orientador
sum(is.na(unb.adv.biomed.df$idLattes2))

## [1] 123

#Numero de orientacoes por ano
table(unb.adv.biomed.df$ano)

##
## 2010 2011 2012 2013 2014 2015 2016 2017
##    7    6   12   17   25   25   13   19

```



```
sort(table(unb.adv.biomed.df$instituicao),decreasing = TRUE)
```

#Cursos que estão envolvido na pós-graduação de engenharia Biomédica

##		
##	Engenharia Biomédica	
##		41
##	Mestrado em Engenharia Biomédica	
##		26
##	Programa de Pós-Graduação em Engenharia Biomédica	
##		7
##	Engenharia Elétrica	
##		6
##	Física	
##		6
##	Ciências Médicas	
##		5
##	ENGENHARIA DE SISTEMAS ELETRÔNICOS E DE AUTOMAÇÃO	
##		5
##	Pós-Graduação em Engenharia Biomédica	
##		5
##	ENGENHARIA ELÉTRICA	
##		3
##		
##		2
##	ciências de materiais	
##		2
##	PPDSCI/CEAM	
##		2
##	Sistemas Mecatrônicos	
##		2
##	Ciência da Computação - Uern - Ufersa	

```
##                                     1
##                               Ciências da Saúde
##                                     1
##                               Ciências e Tecnologias em Saúde
##                                     1
##                               Ciências Mecânicas
##                                     1
##                               Demografia
##                                     1
##                               Educação Física
##                                     1
##                               Informática
##                                     1
##                               Medicina (Clínica Médica)
##                                     1
##                               Nanociência e Nanobiotecnologia
##                                     1
##                               Pós-graduação em Ciências Médicas
##                                     1
##                               Psicologia
##                                     1
##                               Psicologia Clínica e Cultura
##                                     1
```

```
cursos_d5<-head(cursos_d,5)
table(unb.adv.biomed.df$codigo_do_curso)
```

```
##
##          51500027 51500132 51500140 51500248 60009322 60018704 60021152
##           2         6         1         6         1         1         2         5
## 60021160 60021179 60027002 60027894 60045850 60057831 60057840 60059672
##           1         1         1         1         1         3         5        33
## 60059753 60471972 90000003 90000013 90000019 90000022 90000023 90000024
##           1         1         5         1         1         1         7         9
## 90000027 90000035 90000036 90000046 90000052
##           1         1         1        19         6
```

```
#Quantidades de natureza dos trabalhos orientado
table(unb.adv.biomed.df$natureza)
```

```
##
##      Dissertação de mestrado Supervisão de pós-doutorado
##                               109                          2
##      Tese de doutorado
##                               13
```

```
#Tabela com nome de professor e numero de orientacoes
```

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

```
##
##          Lourdes Mattos Brasil
##                               24
##      Suélia de Siqueira Rodrigues Fleury Rosa
##                               20
##          Adson Ferreira da Rocha
##                               13
```

```
##           Jose Felicio da Silva
##                               11
##           Marilia Miranda Forte Gomes
##                               11
## Cristiano Jacques Miosso Rodrigues Mendes
##                               10
##           Marcelino Monteiro de Andrade
##                               7
##           Ronni Geraldo Gomes de Amorim
##                               7
## Vera Regina Fernandes da Silva Marães
##                               7
##           Georges Daniel Amvame Nze
##                               6
##           Marcella Lemos Brettas Carneiro
##                               3
##           Sergio Ricardo Menezes Mateus
##                               3
## Suélia de Siqueira Rodrigues Fleury Rosa
##                               1
##           Fabiano Araujo Soares
##                               1
##           Marcus Vinícius Chaffim Costa
##                               1
```

```
# Tabela com nome dos alunos que mais foram orientados
head(sort(table(toupper(unb.adv.biomed.df$nome_aluno)),decreasing = TRUE),n=10)
```

```
##
## AMILTON DOS REIS CAPISTRANO      ANTONIO DOMINGUES NETO
##                               3      2
## CAMILA CADENA DE ALMEIDA  CRISTINA AKEMI SHIMODA UECHI
##                               2      2
## LEINA ADRIANA BARBOSA PIMENTA      LUIZ ALBER LEMOS
##                               2      2
## MARIA DO CARMO DOS REIS      ROBERTO AGUIAR LIMA
##                               2      2
## ROOZBEH TAHMASEBI      SIMONE BEZERRA FRANCO
##                               2      2
```

```
# Quantidade de alunos que não são bolsistas
sum(unb.adv.biomed.df$bolsa == 'NAO')
```

```
## [1] 78
```

```
# Quantidade de alunos que são bolsistas
sum(unb.adv.biomed.df$bolsa == 'SIM')
```

```
## [1] 46
```

```
# Reparar que a maioria não tem agência financiadora por trás e há dois alunos que tem bolsa mas que não
table(unb.adv.biomed.df$agencia_financiadora)
```

```
##
##
##
##
```

```
## 1
## Conselho Nacional de Desenvolvimento Científico e Tecnológico
## 10
## Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
## 32
## Fundação de Apoio à Pesquisa do Distrito Federal
## 1
```

Verificando engenharia automação

Verificando engenharia elétrica p8

Verificando engenharia elétrica p7

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

Para o levantamento de informação os dados em geral são preciso, como por exemplo saber quais os professores mais envolvidos em orientações, quantos artigos foram concluídos, informações por ano. Mas houve dificuldade quando o dado que se queria analisar dependia de como a pessoa que cadastrou escreveu, como por exemplo os autores nos artigos, haja visto que não foi encontrado um padrão, dificultou qualquer análise.

CRISP-DM Fase 3 - Preparação dos Dados

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

Para a principal utilização dos dados foi definidos os **data frames** unb.adv.biomed.df, que neste caso contém alguns informações de qual a natureza da pesquisa produzida, qual aluno produziu, seus respectivos orientadores, o ano e etc. também foi definido os dados importantes do arquivo profile.json, gerando o **data frame** unb.prof.biomed.df, que já está limpo, com as principais colunas definidas.

CRISP-DM Fase.Atividade 3.2 - Limpeza dos dados

Esta etapa se mistura com a próxima pelo fato de limpar e preparar o dado, ou seja, já construir.

CRISP-DM Fase.Atividade 3.3 - Construção dos dados

Construindo e limpando alguns dados de interesse.

```
# Funcoes auxiliares que serão usadas
# estas funções foram disponibilizada pelo professor da disciplina
# converte as colunas de um dataframe tipo lista em tipo character
cv_tplista2tpchar <- function( df ) {
  for( variavel in names(df) ) {
    if (class(df[[variavel]]) == "list" ) {
      df[[variavel]] <- lapply(df[[variavel]] , function(x) lista2texto( x ) )
      df[[variavel]] <- as.character( df[[variavel]] )
    }
  }
  return(df)
}
###
```

```
# converte o conteudo de lista em array de characters
```

```

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

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

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

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

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

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

```

```

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

}

# Extracao dos perfis dos professores

extra1_perfil <- 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)
}

extra1_perfis <- function(jsonProfessores) {
  df_saida <- data.frame()
  for( i in 1:length(jsonProfessores)) {
    jsonProfessor <- jsonProfessores[i]
    df_professor <- extra1_perfil(jsonProfessor)
    if( nrow(df_saida) > 0 ) {
      df_saida <- rbind(df_saida , df_professor)
    } else {
      df_saida <- df_professor
    }
  }

  return(df_saida)
}

# Extracao da producao bibliografica dos professores

extra1_producao <- function(professor) {
  idLattes <- names(professor)
  df_1producao <- NULL
  producao_bibliografica <- professor[[idLattes]]$producao_bibliografica #list
  for( tipo_producao in names(producao_bibliografica)) {
    df_temporario <- cv_tplista2tpchar ( producao_bibliografica[[tipo_producao]])
  }
}

```

```

    df_temporario$tipo_producao <- tipo_producao
    df_temporario$idLattes <- idLattes
    df_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)
}

# Extracao das orientacoes dos professores

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)
}

# Extracao das areas de atuacao dos professores

```

```

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

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

##### Inicio #####
#### Começo da Preparação dos dados referente a profile

unb.prof.biomed.json <- read_file("dados-2018-2/engenharia-biomedica/279.profile.json")
unb.prof.biomed.df.capes <- read_csv("dados-2018-2/PesqPosCapes.csv",
                                     sep = ";", header = TRUE, colClasses = "character")
unb.prof.biomed <- fromJSON(unb.prof.biomed.json)
length(unb.prof.biomed)

## [1] 15

# extrai perfis dos professores
unb.prof.biomed.df.professores <- extrai_perfis(unb.prof.biomed)

# extrai producao bibliografica de todos os professores
unb.prof.biomed.df.publicacoes <- extrai_producoes(unb.prof.biomed)

#extrai orientacoes
unb.prof.biomed.df.orientacoes <- extrai_orientacoes(unb.prof.biomed)

#extrai areas de atuacao
unb.prof.biomed.df.areas.de.atuacao <- extrai_areas_atuacao(unb.prof.biomed)

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

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

glimpse(unb.prof.biomed.df)

## Observations: 15
## Variables: 16
## $ idLattes <chr> "0201204222182378", "05...
## $ nome <chr> "Leandro Xavier Cardoso...
## $ resumo_cv <chr> "Possui graduação em Fí...
## $ senioridade <chr> "9", "6", "9", "9", "5"...
## $ ORIENTACAO_CONCLUIDA_DOUTORADO <int> NA, 1, 4, 3, 1, NA, NA,...
## $ ORIENTACAO_CONCLUIDA_MESTRADO <int> NA, 9, 9, 18, NA, 11, 7...
## $ ORIENTACAO_CONCLUIDA_POS_DOUTORADO <int> NA, NA, NA, NA, NA, NA,...
## $ OUTRAS_ORIENTACOES_CONCLUIDAS <int> 5, 2, 6, 148, 48, 20, 1...
## $ CAPITULO_DE_LIVRO <int> NA, 1, 9, 9, NA, 1, NA,...
## $ DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA <int> NA, NA, NA, NA, NA, NA,...
## $ EVENTO <int> 3, 23, 51, 65, 10, 15, ...
## $ LIVRO <int> NA, NA, NA, 1, NA, NA, ...
## $ PERIODICO <int> 10, 8, 22, 36, 1, 2, 29...
## $ TEXTO_EM_JORNAIS <int> NA, NA, NA, 1, NA, NA, ...
## $ `n_distinct(area)` <int> 1, 3, 2, 3, 2, 3, 2, 1,...
## $ `n_distinct(AreaPos)` <int> 1, 1, 3, 1, 1, 1, 2, 1,...

head(unb.prof.biomed.df,3)

##           idLattes                                nome
## 1 0201204222182378                      Leandro Xavier Cardoso
## 2 0535100751136568 Cristiano Jacques Miosso Rodrigues Mendes
## 3 1141716826787805                      Adson Ferreira da Rocha

```

```
##
## 1
## 2
## 3 Engenheiro Eletricista pela Universidade de Brasília (1988), Mestre em Engenharia Elétrica pela Un
##  senioridade ORIENTACAO_CONCLUIDA_DOUTORADO ORIENTACAO_CONCLUIDA_MESTRADO
## 1          9          NA          NA
## 2          6          1          9
## 3          9          4          9
##  ORIENTACAO_CONCLUIDA_POS_DOUTORADO OUTRAS_ORIENTACOES_CONCLUIDAS
## 1          NA          5
## 2          NA          2
## 3          NA          6
##  CAPITULO_DE_LIVRO DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA EVENTO LIVRO
## 1          NA          NA          3          NA
## 2          1          NA          23          NA
## 3          9          NA          51          NA
##  PERIODICO TEXTO_EM_JORNAIS n_distinct(area) n_distinct(AreaPos)
## 1          10          NA          1          1
## 2          8          NA          3          1
## 3          22          NA          2          3

## Mostrando dados processados referente ao arquivo 279.publication.json
glimpse(unb.pub.biomed.df)

## Observations: 147
## Variables: 10
## $ natureza      <chr> "COMPLETO", "COMPLETO", "COMPLETO", "COMPLE...
## $ titulo         <chr> "Prótese para controle de fluxo esofagiano ...
## $ periodico      <chr> "Revista Brasileira de Engenharia Biomédica...
## $ ano            <chr> "2010", "2010", "2010", "2010", "2010", "20...
## $ volume         <chr> "26", "249", "20", "11", "14", "9", "53", "...
## $ issn           <chr> "15173151", "17426596", "10546618", "151942...
## $ paginas        <chr> "49 - 54", "012031 - ", "192 - 200", "23 - ...
## $ doi            <chr> "", "10.1088/1742-6596/249/1/012031", "10.1...
## $ autores        <chr> "Rosa, S. S. R. F.; da Rocha, A F; Brasil, ...
## $ `autores-endogeno` <chr> "1141716826787805", "0201204222182378", "91...

#Podemos ver que contém natureza, titulo, periodico , autores ....
head(unb.pub.biomed.df,3)

##  natureza
## 1 COMPLETO
## 2 COMPLETO
## 3 COMPLETO
##
## 1 Prótese para controle de fluxo esofagiano como nova técnica para o tratamento da obesidade (QUALIS
## 2 Thermoluminescent dose reconstruction using quartz extracted from unfired build
## 3 Breast cancer image assessment using an adaptative network-based fuzzy inference sys
##                                periodico ano volume
## 1 Revista Brasileira de Engenharia Biomédica (Impresso) 2010 26
## 2 Journal of Physics. Conference Series (Online) 2010 249
## 3 Pattern Recognition and Image Analysis 2010 20
##      issn paginas doi
## 1 15173151 49 - 54
## 2 17426596 012031 - 10.1088/1742-6596/249/1/012031
```

```
## 3 10546618 192 - 200      10.1134/S1054661810020112
##
## 1
## 2 Campos, Simara S; Almeida, Geângela M; CARDOSO, L. X.;Cardoso, Leandro X;CARDOSO, L X;XAVIER CARDOSO
## 3      Fernandes, F. C.; BRASIL, L. M.;Brasil, L. M.;Brasil, L. M.
##      autores-endogeno
## 1 1141716826787805
## 2 0201204222182378
## 3 9190489069187153
```

```
## Mostrando dados processados referente ao arquivo 279.advise.json
glimpse(unb.adv.biomed.df)
```

```
## Observations: 124
## Variables: 15
## $ natureza      <chr> "Supervisão de pós-doutorado", "Su...
## $ titulo        <chr> "", "", "Influência da Eletroestim...
## $ ano           <chr> "2014", "2017", "2011", "2011", "2...
## $ id_lattes_aluno <chr> "", "", "7129464687368571", "92327...
## $ nome_aluno    <chr> "Leandro Xavier Cardoso", "Glécia ...
## $ instituicao    <chr> "Universidade de Brasília - Faculd...
## $ curso         <chr> "", "", "Ciências Médicas", "ENGEN...
## $ codigo_do_curso <chr> "", "", "60021152", "60057840", "6...
## $ bolsa        <chr> "SIM", "SIM", "SIM", "NAO", "NAO",...
## $ agencia_financiadora <chr> "Coordenação de Aperfeiçoamento de...
## $ codigo_agencia_financiadora <chr> "0450000000000", "0450000000000", "0...
## $ ori1         <chr> "Lourdes Mattos Brasil", "Lourdes ...
## $ ori2         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ idLattes1    <chr> "9190489069187153", "9190489069187...
## $ idLattes2    <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA...
```

```
#Podemos ver que contém natureza, titulo, autores, ano, nome aluno, orientadores ....
head(unb.adv.biomed.df,3)
```

```
##      natureza
## 1 Supervisão de pós-doutorado
## 2 Supervisão de pós-doutorado
## 3      Tese de doutorado
##
## 1
## 2
## 3 Influência da Eletroestimulação Neuromuscular de Baixa Frequência nas Variáveis Eletromiográficas
##      ano id_lattes_aluno      nome_aluno
## 1 2014      Leandro Xavier Cardoso
## 2 2017      Glécia Virgolino da Silva Luz
## 3 2011 7129464687368571      Kenia Fonseca Pires
##      instituicao      curso
## 1 Universidade de Brasília - Faculdade UnB-Gama
## 2 Universidade de Brasília - Faculdade UnB-Gama
## 3      Universidade de Brasília Ciências Médicas
##      codigo_do_curso bolsa
## 1      SIM
## 2      SIM
## 3      60021152      SIM
##      agencia_financiadora
## 1      Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
```

```
## 2  Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
## 3  Conselho Nacional de Desenvolvimento Científico e Tecnológico
##  codigo_agencia_financiadora      ori1 ori2
## 1              045000000000    Lourdes Mattos Brasil <NA>
## 2              045000000000    Lourdes Mattos Brasil <NA>
## 3              002200000000 Adson Ferreira da Rocha <NA>
##      idLattes1 idLattes2
## 1 9190489069187153    <NA>
## 2 9190489069187153    <NA>
## 3 1141716826787805    <NA>

##### Os arquivos
##### 279.graph.json - usado para mostrar correlações
#### 279.list.json - Como explicado este arquivo não contém nenhuma informação interessante.
```

Verificando engenharia automação

Verificando engenharia elétrica p8

Verificando engenharia elétrica p7

CRISP-DM Fase.Atividade 3.4 - Integração dos dados

Neste presente trabalho, não se viu a necessidade de fazer merge entre os data frames.

CRISP-DM Fase.Atividade 3.5 - Formatação dos dados

As formatações de dados necessária já foram feitas, como por exemplo “orientadores” virou “orie1” e “orie2”, fazendo com que cada variável contenha apenas um elemento, e não mais uma lista de orientadores.

CRISP-DM Fase 4 - Modelagem

Nesta fase, após ter feito limpeza e Preparação dos dados, tem-se que se decidir qual ferramenta usar : computacionais, matemáticas ou estatística. Esta fase, por enquanto, será omitida.

CRISP-DM Fase 5 - Avaliação

TODO LIST:

- AUTOMAÇÃO
- P7
- P8

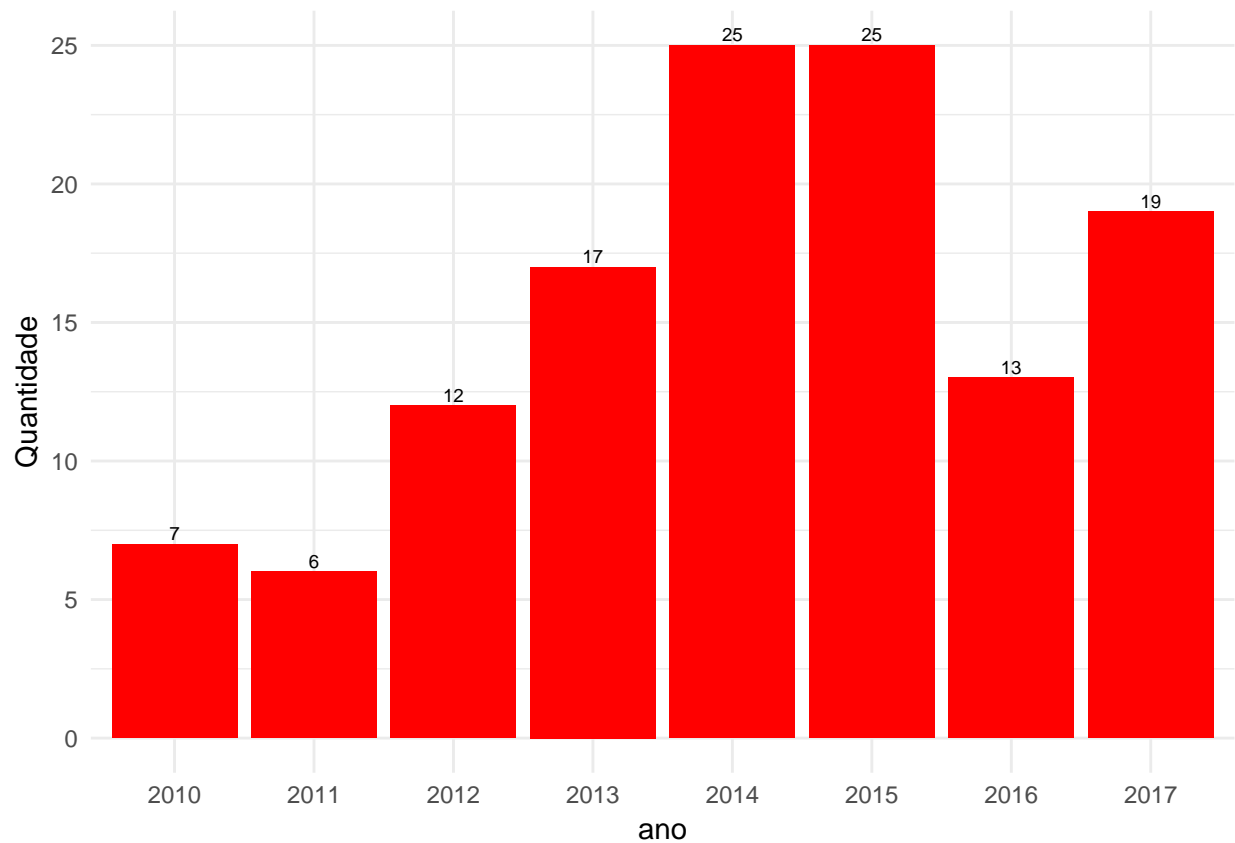
Será exposto em forma de gráficos certos dados que se acharam convenientes e interessantes.

Expondo engenharia biomédica

Aqui visa-se ver as quantidades de publicações feitas por anos na área de engenharia de Biomédica.

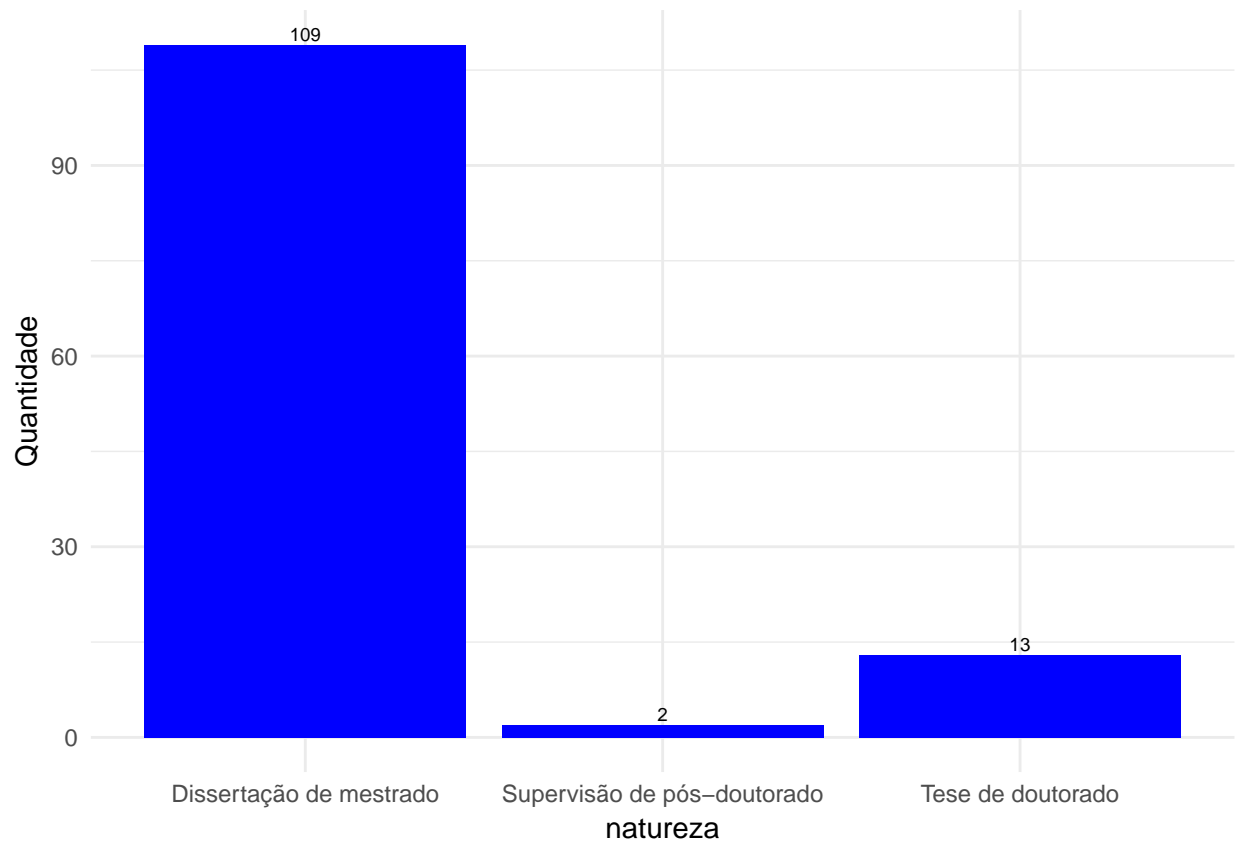
```
unb.adv.biomed.df %>%
group_by(ano) %>%
summarise(Quantidade = n()) %>%
ggplot(aes(x = ano, y = Quantidade)) +
geom_bar(position = "stack", stat = "identity", fill = "red") +
```

```
geom_text(aes(label=Quantidade), vjust=-0.3, size=2.5)+
theme_minimal()
```



Agora verificando a quantidade separada por tipo de publicação.

```
unb.adv.biomed.df %>%
group_by(natureza) %>%
summarise(Quantidade = n()) %>%
ggplot(aes(x = natureza, y = Quantidade)) +
geom_bar(position = "stack", stat = "identity", fill = "blue")+
geom_text(aes(label=Quantidade), vjust=-0.3, size=2.5)+
theme_minimal()
```



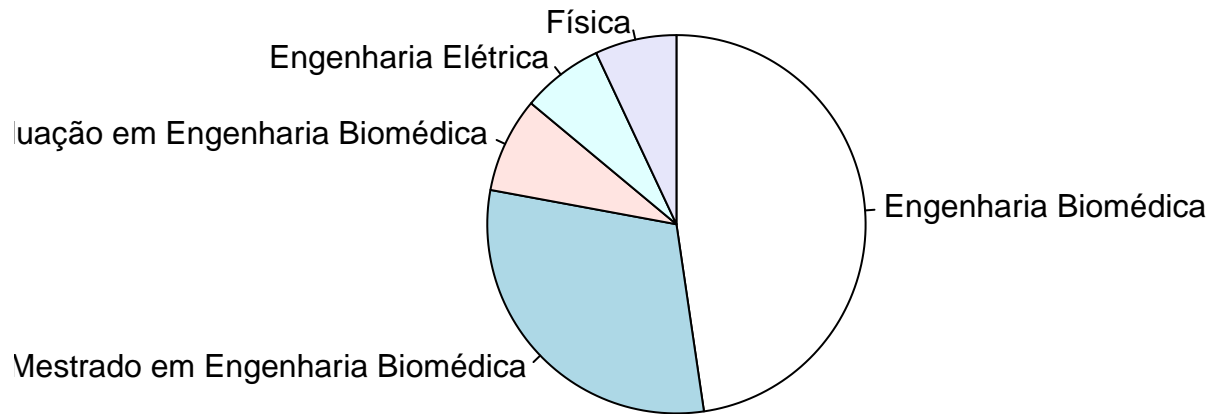
Verificando os cursos mais presentes.

```
cursos_d5
```

```
##
##                               Engenharia Biomédica
##                               41
##           Mestrado em Engenharia Biomédica
##                               26
## Programa de Pós-Graduação em Engenharia Biomédica
##                               7
##                               Engenharia Elétrica
##                               6
##                               Física
##                               6
```

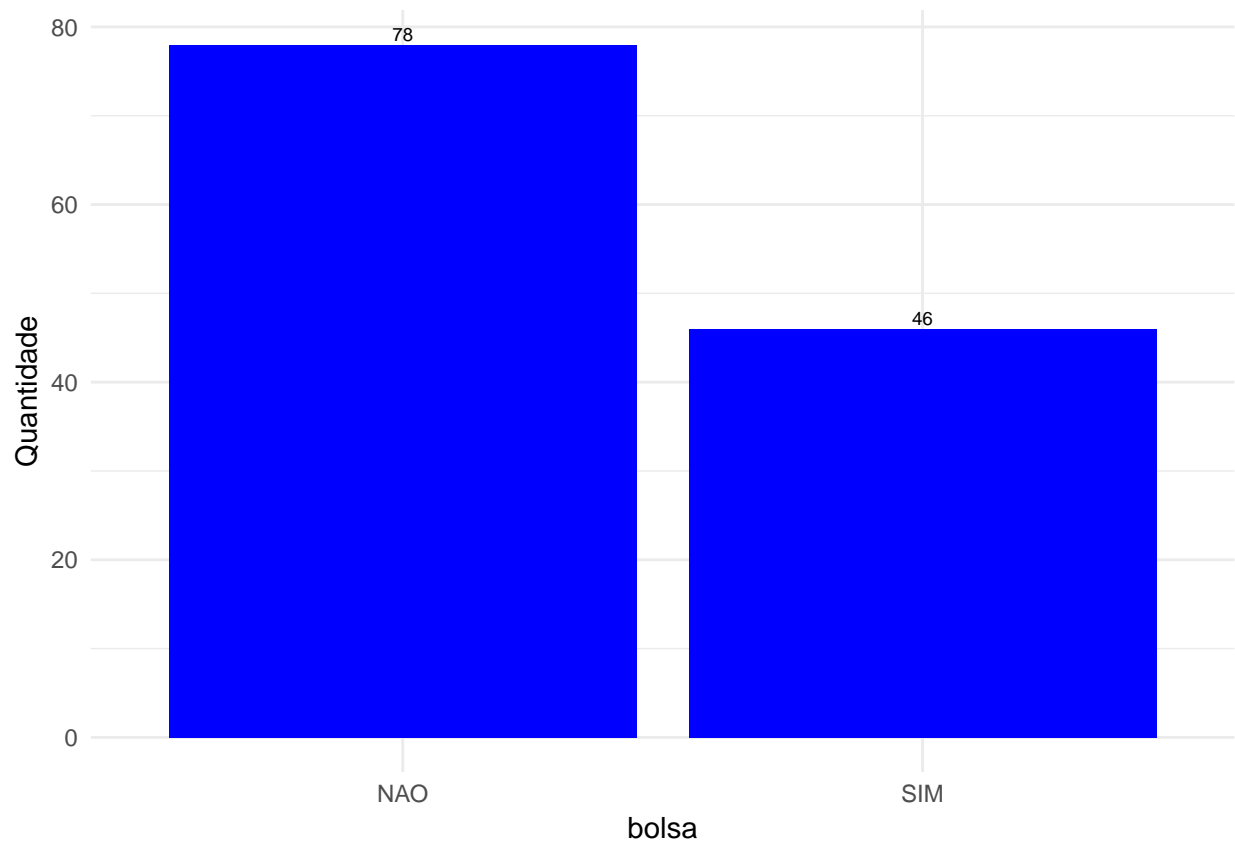
```
pie(cursos_d5,clockwise=TRUE,main="Cursos mais presentes")
```

Cursos mais presentes



Quanto bolsistas tem ?

```
unb.adv.biomed.df %>%  
  group_by(bolsa) %>%  
  summarise(Quantidade = n()) %>%  
  ggplot(aes(x = bolsa, y = Quantidade)) +  
  geom_bar(position = "stack", stat = "identity", fill = "blue") +  
  geom_text(aes(label=Quantidade), vjust=-0.3, size=2.5) +  
  theme_minimal()
```

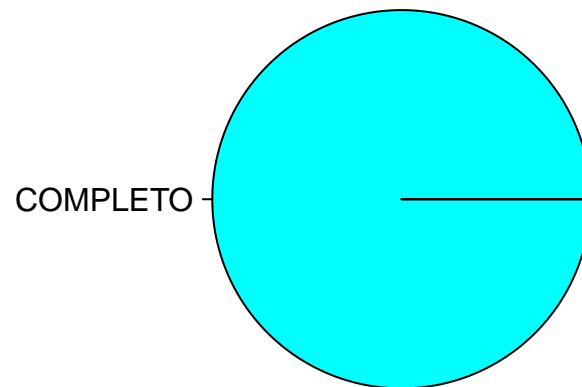


Assim, estes gráficos permitem verificar alguns aspectos importantes sobre as orientações.

Verificando, agora, o arquivo 279.publication.json, contendo informações sobre as publicações.

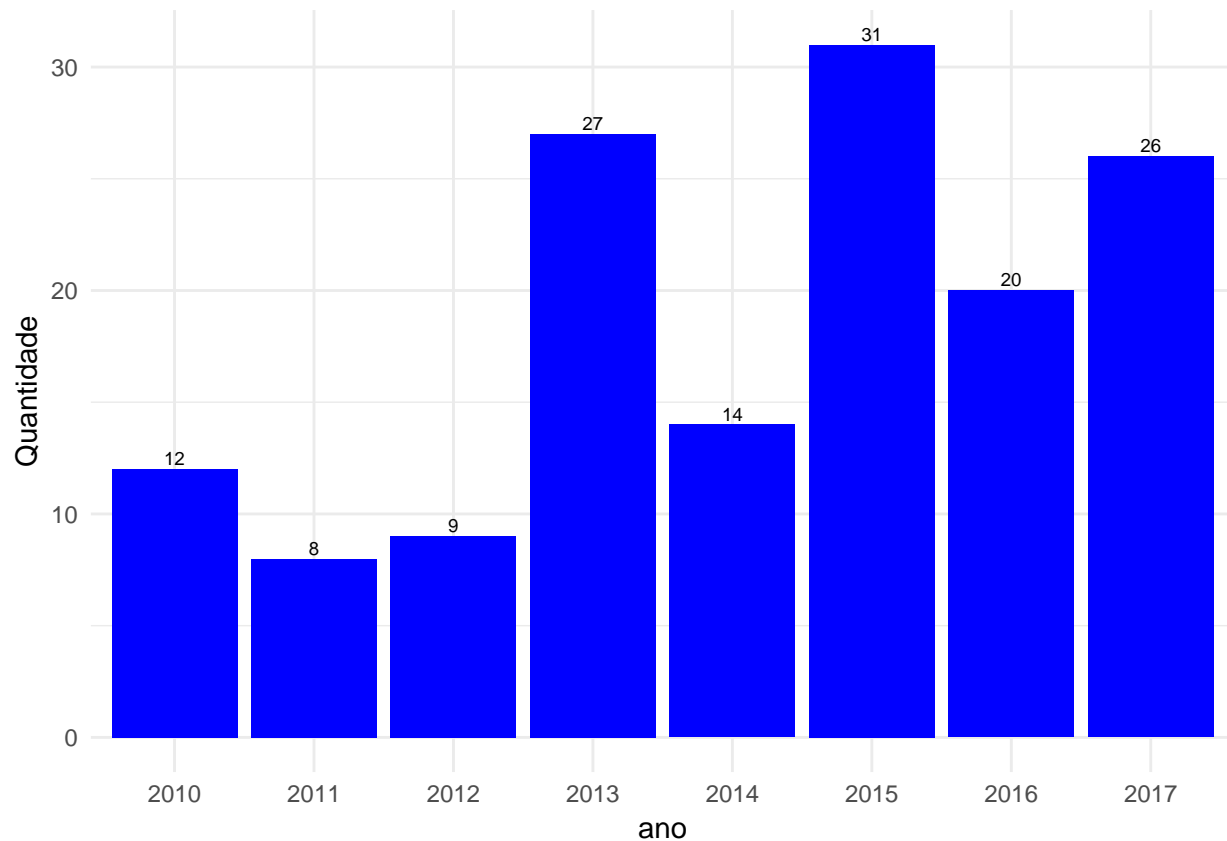
```
pie(table(unb.pub.biomed.df$natureza),main="Natureza das publicações",col = "cyan")
```


Natureza das publicações



Separando por ano e depois por periódico.

```
unb.pub.biomed.df %>%  
group_by(ano) %>%  
summarise(Quantidade = n()) %>%  
ggplot(aes(x = ano, y = Quantidade)) +  
geom_bar(position = "stack", stat = "identity", fill = "blue") +  
geom_text(aes(label=Quantidade), vjust=-0.3, size=2.5) +  
theme_minimal()
```



```
d <- table(unb.pub.biomed.df$periodico)
head(sort(d,decreasing = TRUE),n=5)
```

```
##
##          Revista Brasileira de Engenharia Biomédica (Impresso)
##                                     7
## Global Journal of Engineering Science and Research Management
##                                     6
##          Biomedical Engineering Online (Online)
##                                     4
##          Revista Brasileira de Ensino de Física (Online)
##                                     4
##          Revista Brasileira de Inovação Tecnológica em Saúde
##                                     4
```

Mostrando os autores que mais participaram de publicações.

```
head(sort(table(toupper(unlist(strsplit(unb.pub.biomed.df$autores,";")))),decreasing = TRUE),n=10)
```

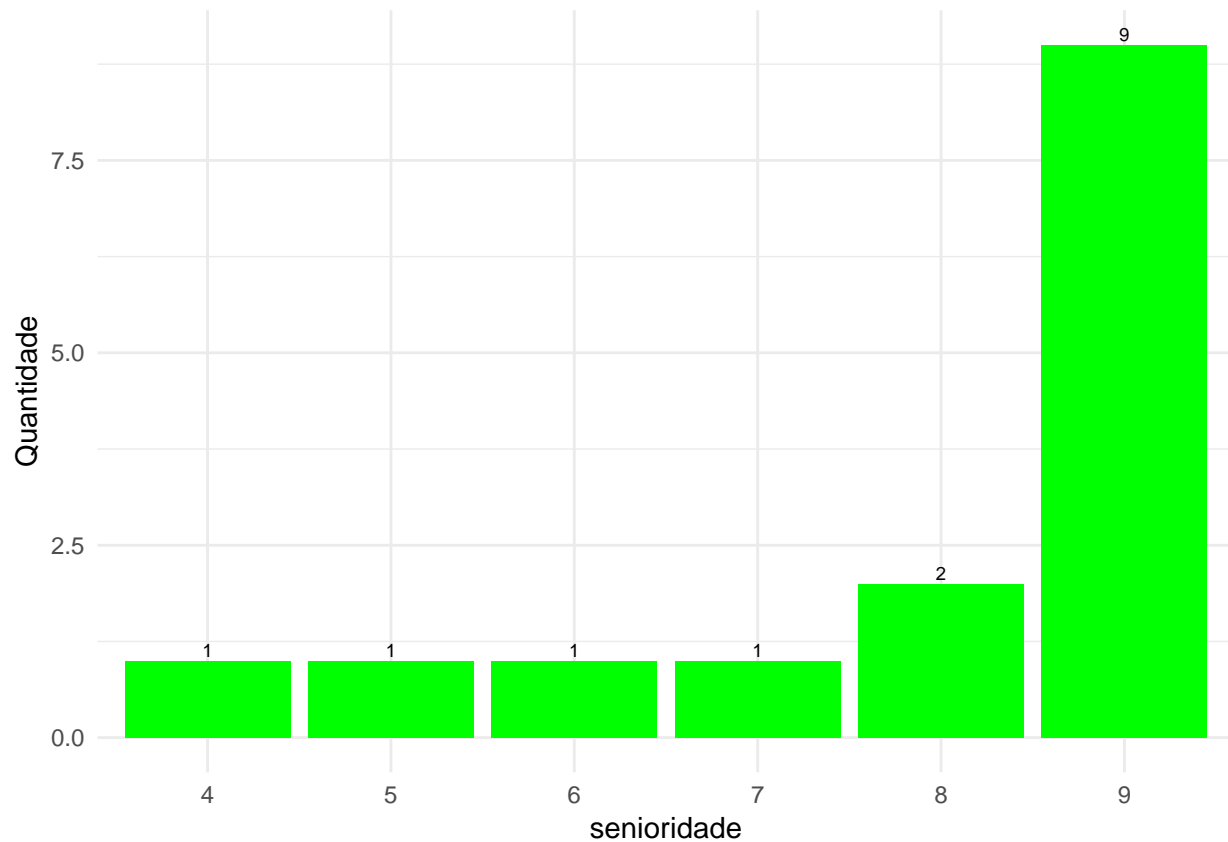
```
##
##          AMORIM, R. G. G.
##                                     17
##          AMORIM, R. G. G.
##                                     16
##          MARÃES, V. R. F. S.
##                                     14
##          BRASIL, L. M.
##                                     7
```

```
##           MARÃES, V. R. F. S.
##           7
##           SILVA, W. B.
##           7
##           GOMES, MARÍLIA MIRANDA FORTE
##           7
## ROSA, SUÉLIA DE SIQUEIRA RODRIGUES FLEURY
##           7
##           DA ROCHA, A. F.
##           6
##           GOMES, MARÍLIA MIRANDA FORTE
##           6
```

Assim, vemos que há algumas incoerência neste dado acima, isto se dá pela falta de uniformidade.

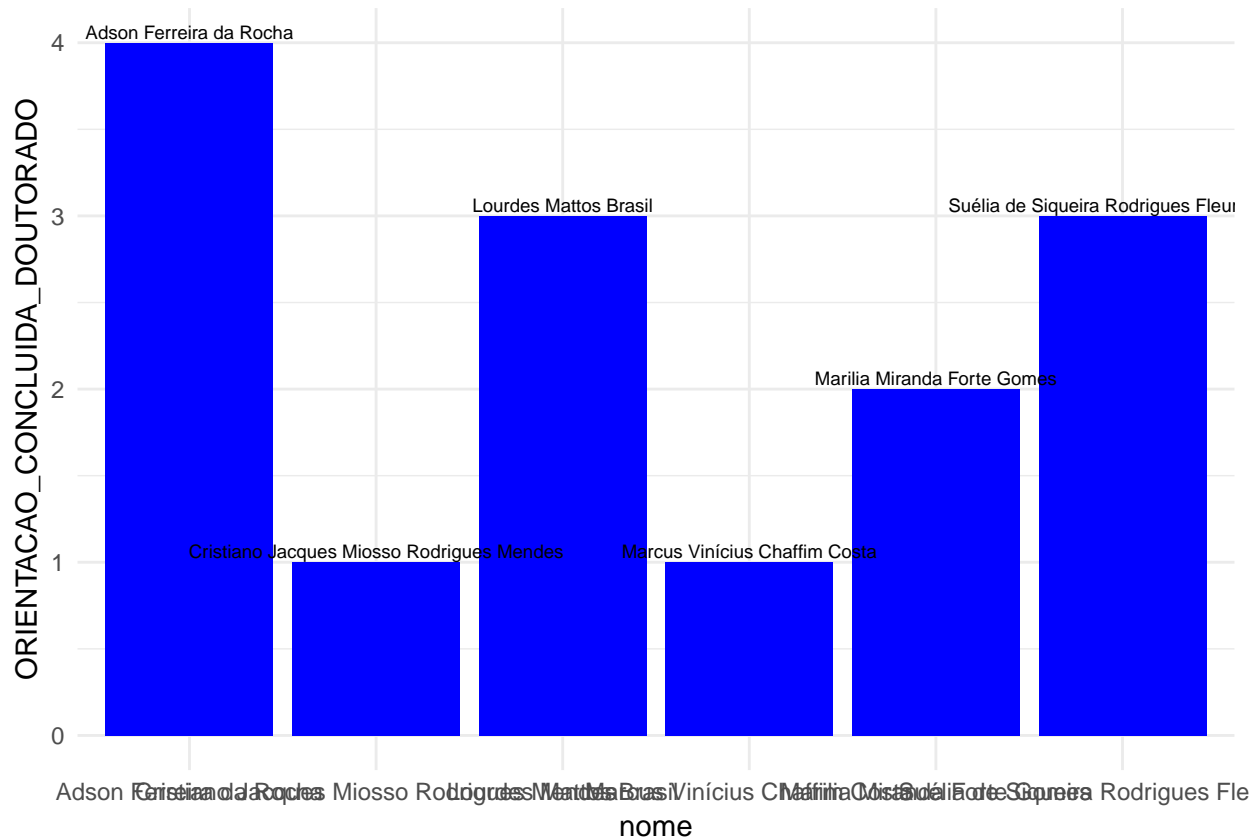
Analisando o arquivo 279.profile.json.

```
unb.prof.biomed.df %>%
  group_by(senioridade) %>%
  summarise(Quantidade = n()) %>%
  ggplot(aes(x = senioridade, y = Quantidade)) +
  geom_bar(position = "stack", stat = "identity", fill = "green") +
  geom_text(aes(label=Quantidade), vjust=-0.3, size=2.5) +
  theme_minimal()
```



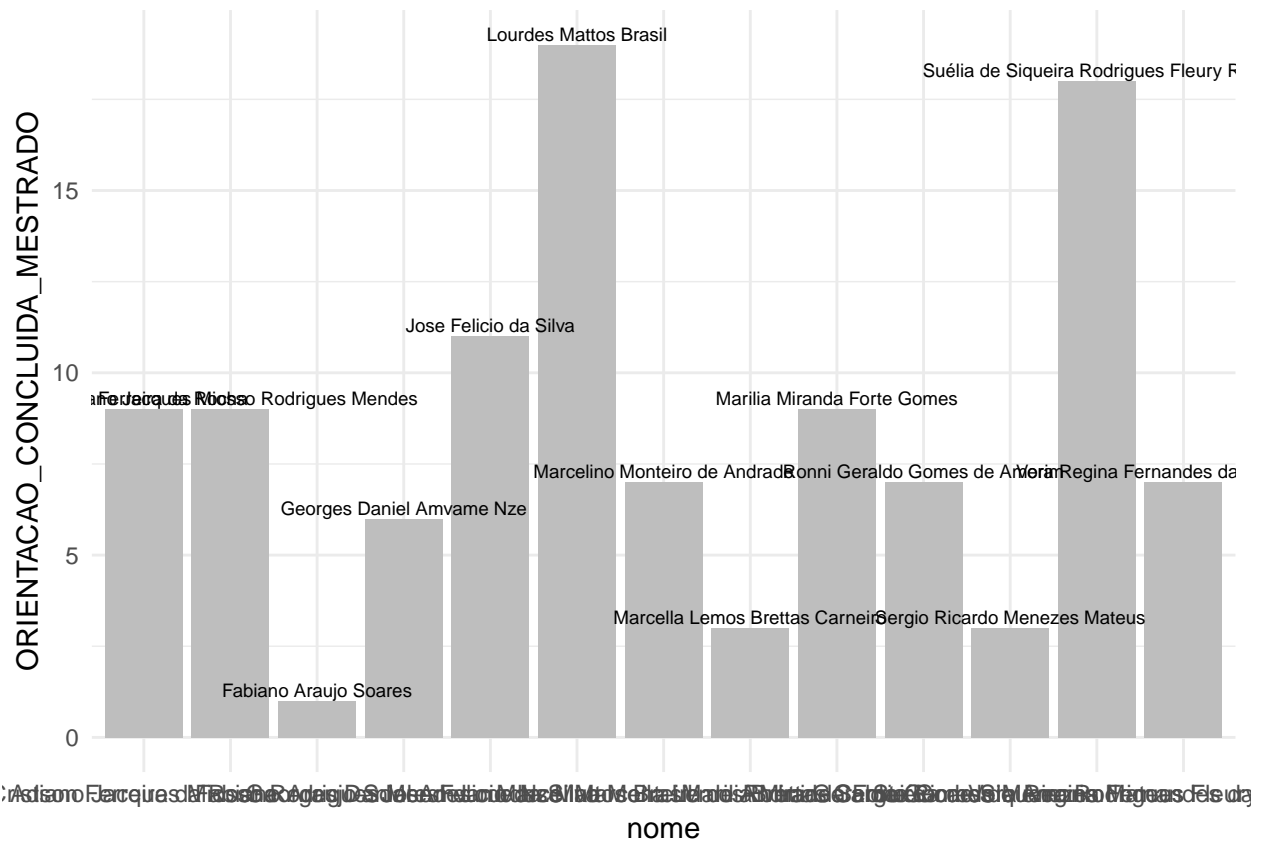
Analisando quantidade de orientações concluídas no doutorado por professor.

```
unb.prof.biomed.df %>%
group_by(nome) %>%
filter(!is.na(ORIENTACAO_CONCLUIDA_DOUTORADO))%>%
ggplot(aes(x = nome, y = ORIENTACAO_CONCLUIDA_DOUTORADO)) +
geom_bar(position = "stack", stat = "identity", fill = "blue")+
geom_text(aes(label=nome), vjust=-0.3, size=2.5)+
theme_minimal()
```



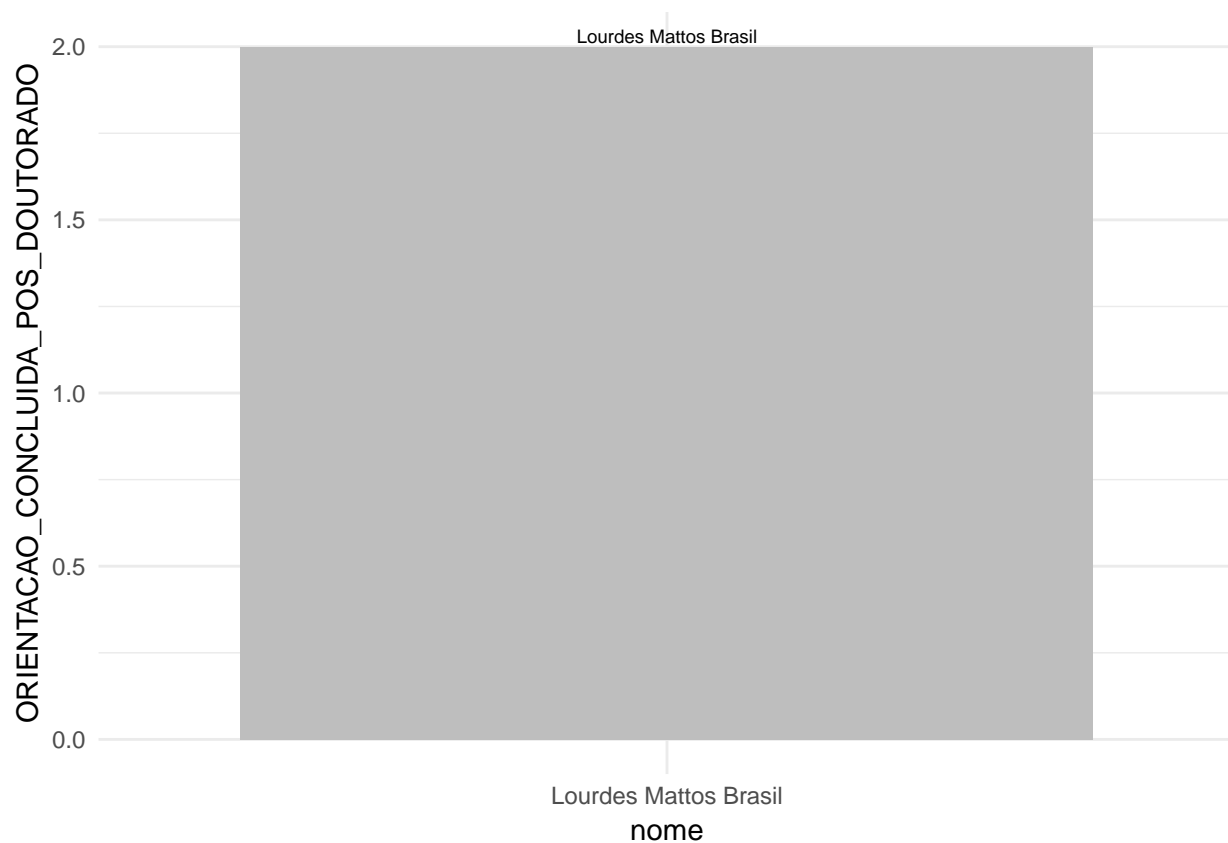
Analisando quantidade de orientações concluídas no mestrado por professor.

```
unb.prof.biomed.df %>%
group_by(nome) %>%
filter(!is.na(ORIENTACAO_CONCLUIDA_MESTRADO))%>%
ggplot(aes(x = nome, y = ORIENTACAO_CONCLUIDA_MESTRADO)) +
geom_bar(position = "stack", stat = "identity", fill = "gray")+
geom_text(aes(label=nome), vjust=-0.3, size=2.5)+
theme_minimal()
```



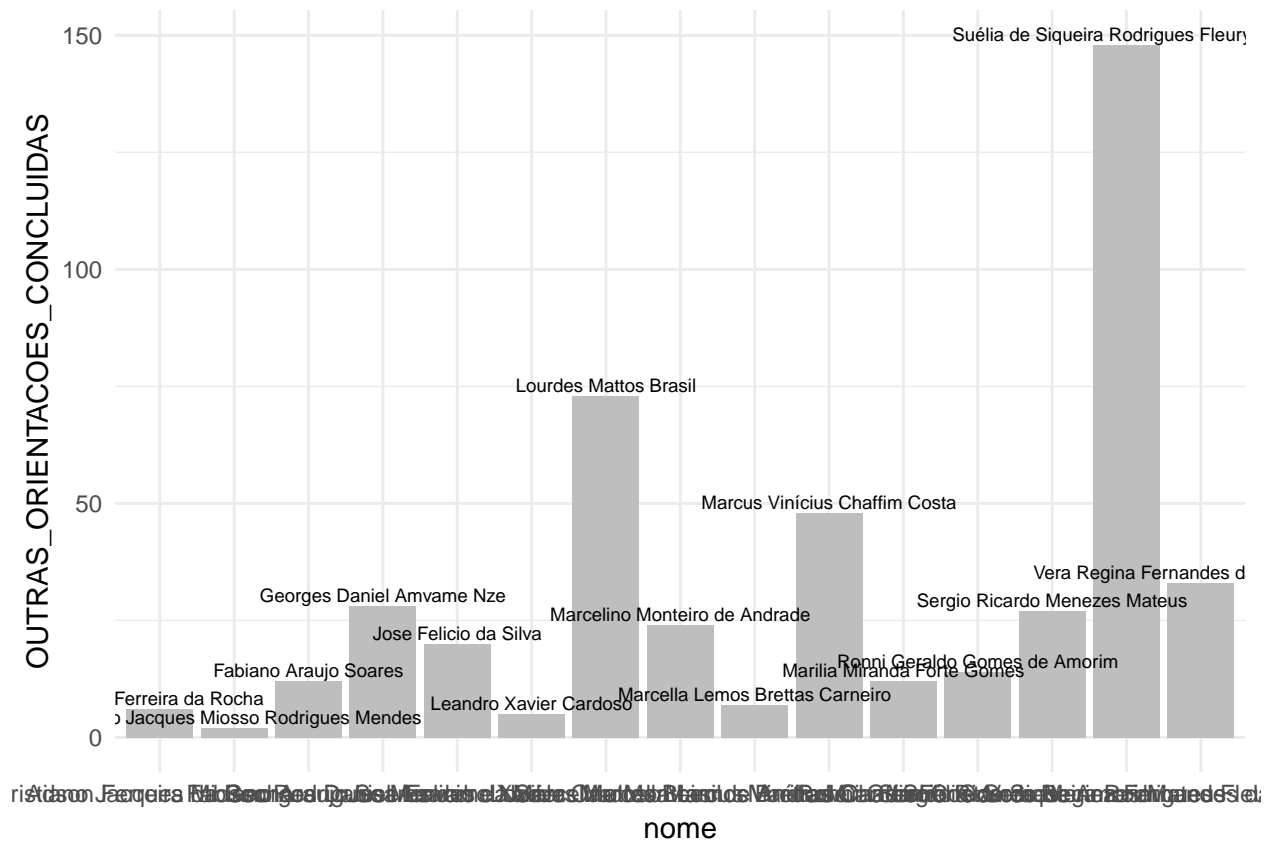
Analisando quantidade de orientações concluídas no pós-doutorado por professor.

```
unb.prof.biomed.df %>%
  group_by(nome) %>%
  filter(!is.na(ORIENTACAO_CONCLUIDA_POS_DOUTORADO))%>%
  ggplot(aes(x = nome, y = ORIENTACAO_CONCLUIDA_POS_DOUTORADO)) +
  geom_bar(position = "stack", stat = "identity", fill = "gray")+
  geom_text(aes(label=nome), vjust=-0.3, size=2.5)+
  theme_minimal()
```



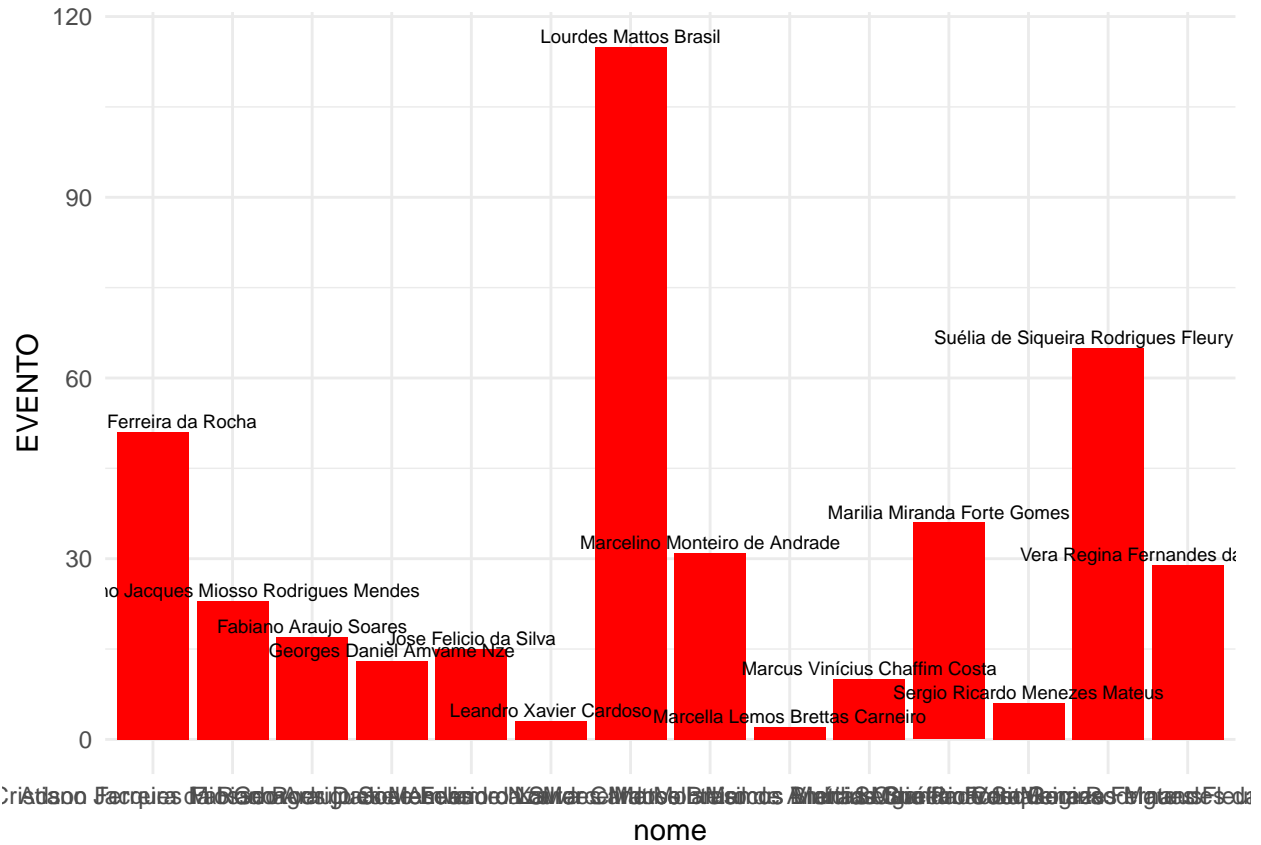
Analisando quantidade de orientações concluídas em outros tipos por professor.

```
unb.prof.biomed.df %>%
  group_by(nome) %>%
  filter(!is.na(OUTRAS_ORIENTACOES_CONCLUIDAS))%>%
  ggplot(aes(x = nome, y = OUTRAS_ORIENTACOES_CONCLUIDAS)) +
  geom_bar(position = "stack", stat = "identity", fill = "gray")+
  geom_text(aes(label=nome), vjust=-0.3, size=2.5)+
  theme_minimal()
```



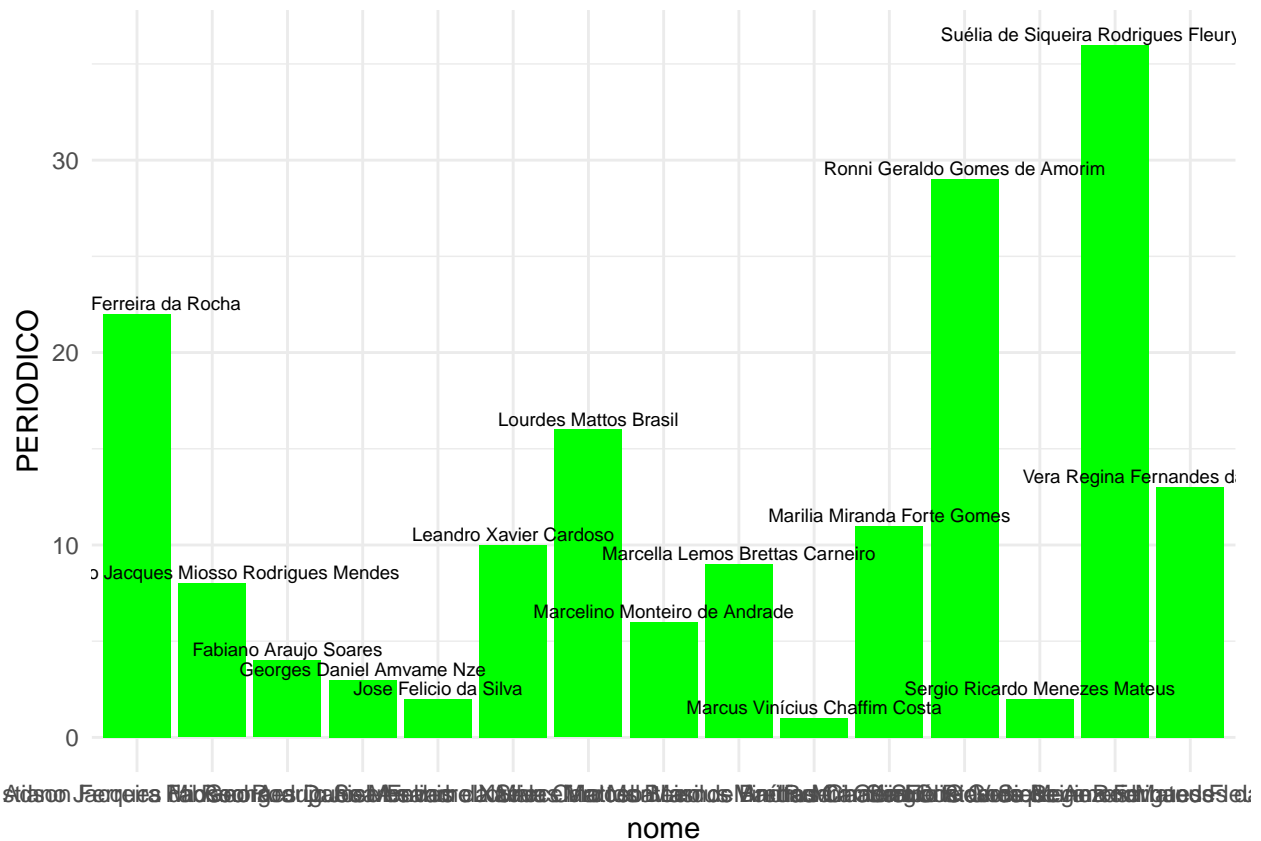
Analisando quantidade de eventos ido por professor.

```
unb.prof.biomed.df %>%
  group_by(nome) %>%
  filter(!is.na(EVENTO))%>%
  ggplot(aes(x = nome, y = EVENTO)) +
  geom_bar(position = "stack", stat = "identity", fill = "red")+
  geom_text(aes(label=nome), vjust=-0.3, size=2.5)+
  theme_minimal()
```



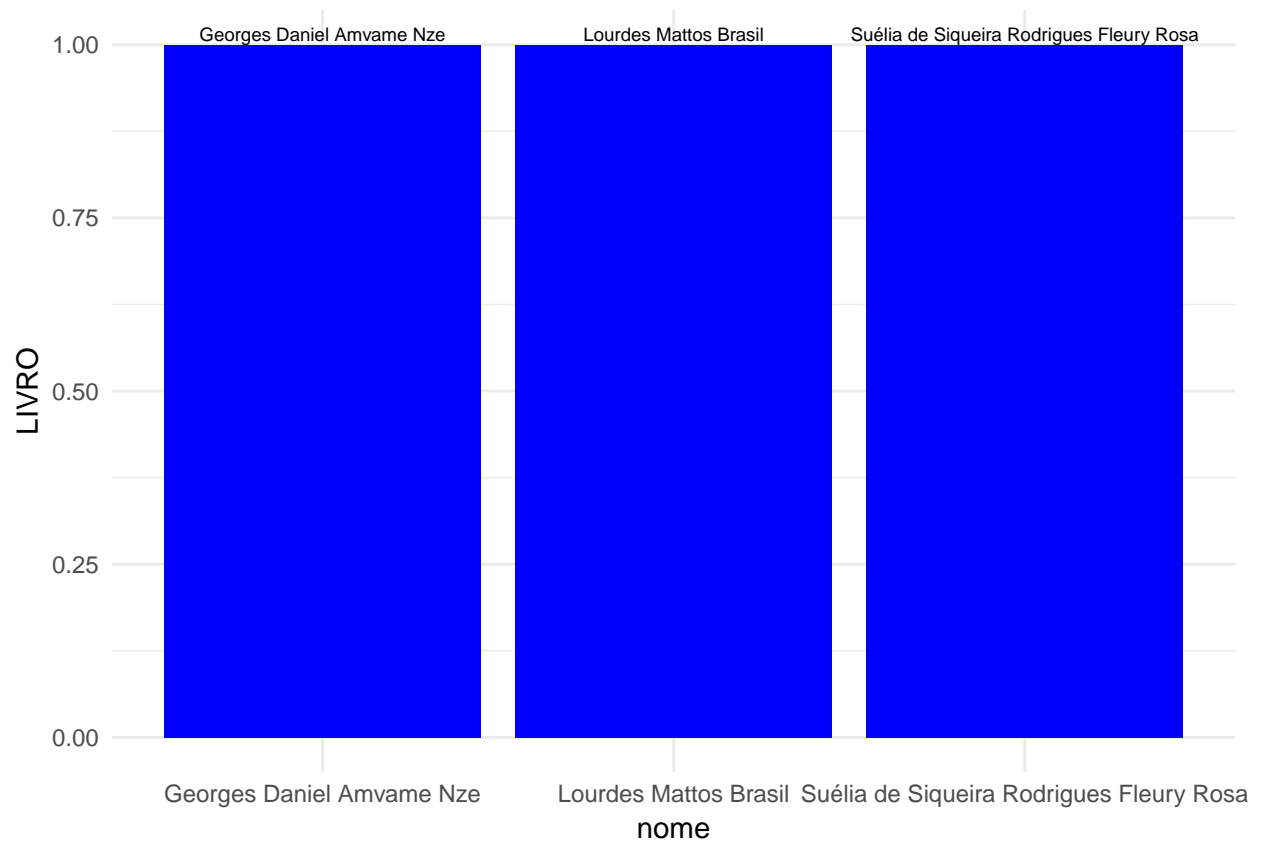
Analisando quantidade de periódicos por professor.

```
unb.prof.biomed.df %>%
  group_by(nome) %>%
  filter(!is.na(PERIODICO))%>%
  ggplot(aes(x = nome, y = PERIODICO)) +
  geom_bar(position = "stack", stat = "identity", fill = "green")+
  geom_text(aes(label=nome), vjust=-0.3, size=2.5)+
  theme_minimal()
```

Analizando quantidade de livro por professor.

```
unb.prof.biomed.df %>%
  group_by(nome) %>%
  filter(!is.na(LIVRO))%>%
  ggplot(aes(x = nome, y = LIVRO)) +
  geom_bar(position = "stack", stat = "identity", fill = "blue")+
  geom_text(aes(label=nome), vjust=-0.3, size=2.5)+
  theme_minimal()
```



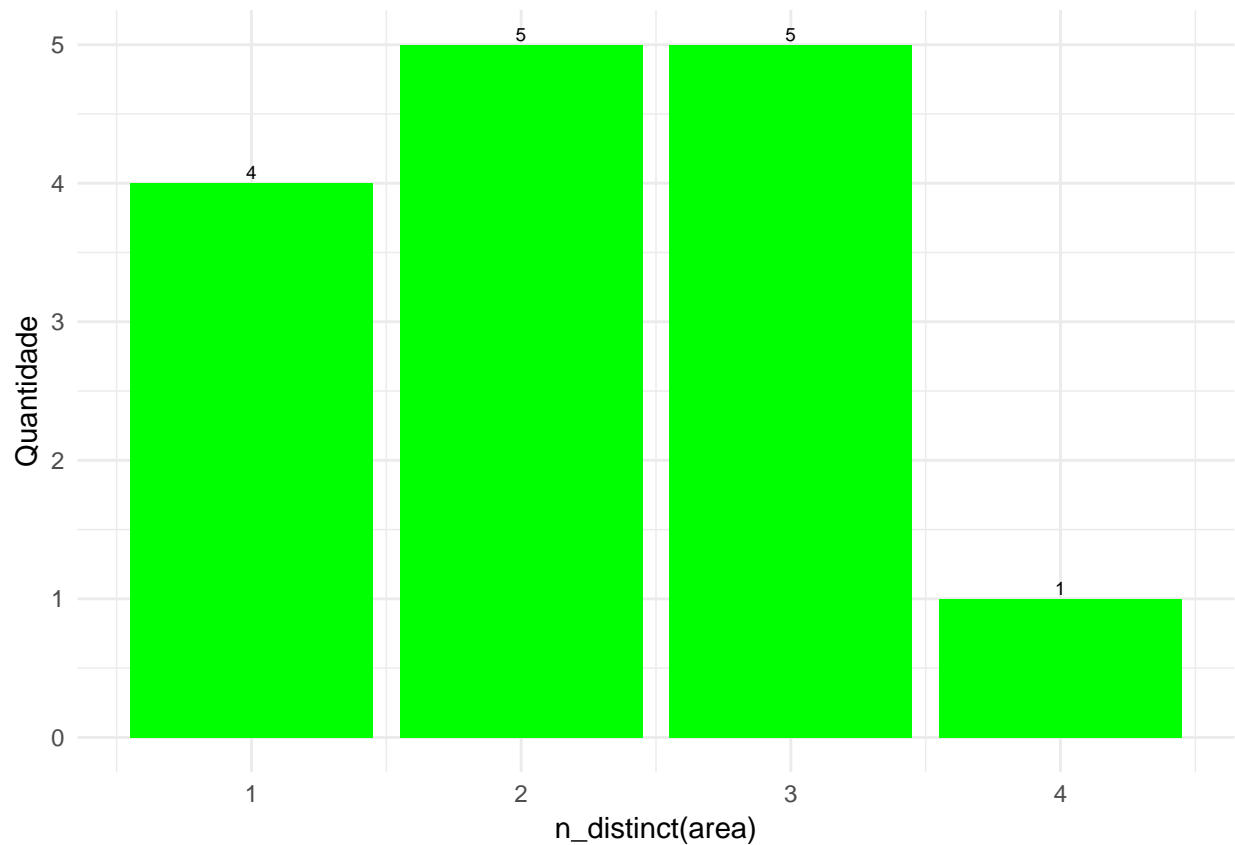
Analisando quantidade de texto em jornais por professor.

```
unb.prof.biomed.df %>%
group_by(nome) %>%
filter(!is.na(TEXTO_EM_JORNAIS))%>%
ggplot(aes(x = nome, y = TEXTO_EM_JORNAIS)) +
geom_bar(position = "stack",stat = "identity", fill = "black")+
geom_text(aes(label=nome), vjust=-0.3, size=2.5)+
theme_minimal()
```



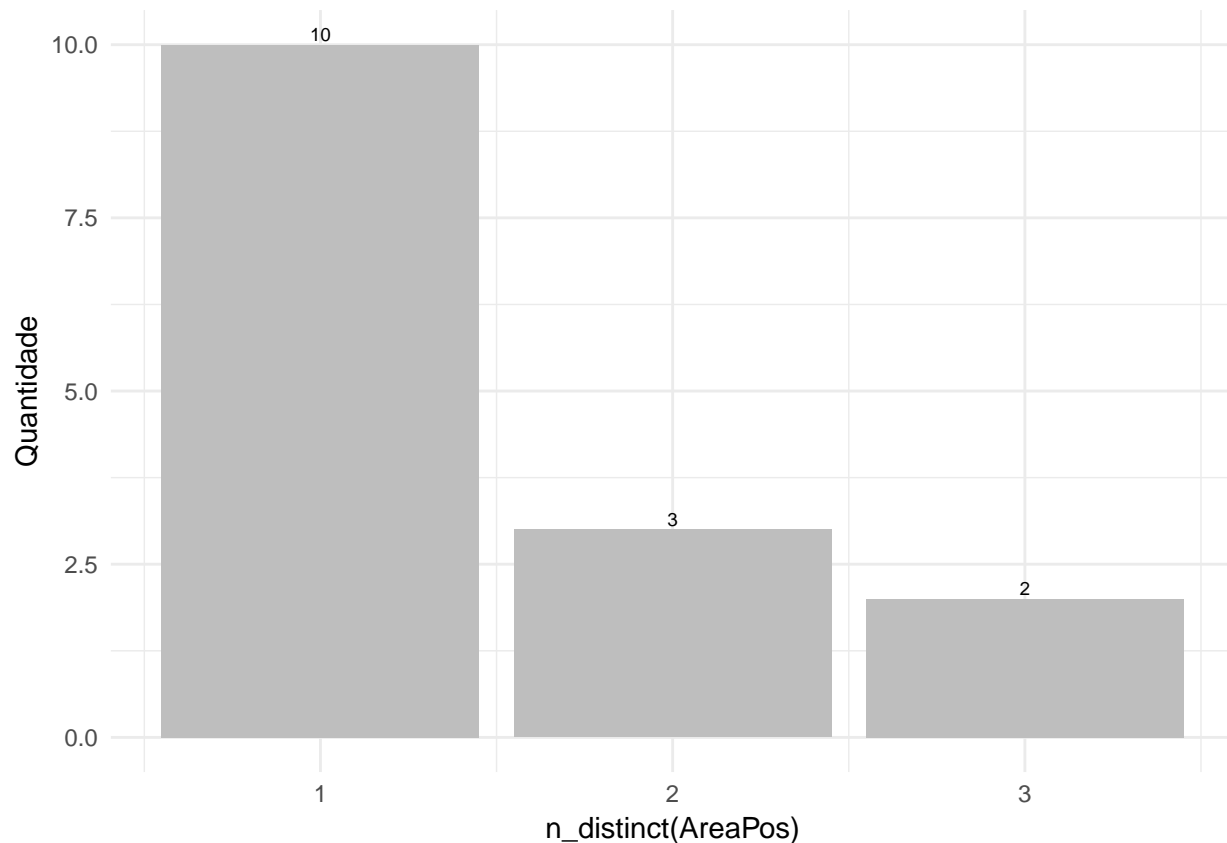
Analisando quantidade de professor por maior área e posteriormente o área pós-graduação.

```
unb.prof.biomed.df %>%
  group_by(`n_distinct(area)`) %>%
  summarise(Quantidade = n()) %>%
  ggplot(aes(x = `n_distinct(area)`, y = Quantidade)) +
  geom_bar(position = "stack", stat = "identity", fill = "green") +
  geom_text(aes(label=Quantidade), vjust=-0.3, size=2.5) +
  theme_minimal()
```



Analisando quantidade de professor por maior área e posteriormente o área pós-graduação.

```
unb.prof.biomed.df %>%
group_by(`n_distinct(AreaPos)`) %>%
summarise(Quantidade = n()) %>%
ggplot(aes(x = `n_distinct(AreaPos)`, y = Quantidade)) +
geom_bar(position = "stack", stat = "identity", fill = "grey") +
geom_text(aes(label=Quantidade), vjust=-0.3, size=2.5) +
theme_minimal()
```



Grafos de engenharia biomédica

```
unb.graph.biomed$links$source
```

```
## [1] "0201204222182378" "0201204222182378" "0535100751136568"
## [4] "0535100751136568" "0535100751136568" "0535100751136568"
## [7] "0535100751136568" "0535100751136568" "0535100751136568"
## [10] "1141716826787805" "1141716826787805" "1141716826787805"
## [13] "1141716826787805" "1141716826787805" "1141716826787805"
## [16] "1154673226500318" "1154673226500318" "1154673226500318"
## [19] "1524924375222848" "2957228356035337" "4739013535126469"
## [22] "4739013535126469" "4739013535126469" "5330755818114960"
## [25] "5810353896294133" "5810353896294133" "7294738832905991"
## [28] "9169095482512290"
```

```
unb.graph.biomed$links$target
```

```
## [1] "5810353896294133" "9190489069187153" "1141716826787805"
## [4] "1154673226500318" "4739013535126469" "5330755818114960"
## [7] "5810353896294133" "9169095482512290" "9190489069187153"
## [10] "1154673226500318" "2957228356035337" "4739013535126469"
## [13] "5330755818114960" "5928104758017036" "9190489069187153"
## [16] "2957228356035337" "5330755818114960" "9190489069187153"
## [19] "5810353896294133" "9190489069187153" "5330755818114960"
## [22] "5810353896294133" "9190489069187153" "9190489069187153"
## [25] "9169095482512290" "9190489069187153" "9190489069187153"
## [28] "9190489069187153"
```

```
list_biomed <- c(rbind(unb.graph.biomed$links$source, unb.graph.biomed$links$target))
list_biomed
```

```
## [1] "0201204222182378" "5810353896294133" "0201204222182378"
## [4] "9190489069187153" "0535100751136568" "1141716826787805"
## [7] "0535100751136568" "1154673226500318" "0535100751136568"
## [10] "4739013535126469" "0535100751136568" "5330755818114960"
## [13] "0535100751136568" "5810353896294133" "0535100751136568"
## [16] "9169095482512290" "0535100751136568" "9190489069187153"
## [19] "1141716826787805" "1154673226500318" "1141716826787805"
## [22] "2957228356035337" "1141716826787805" "4739013535126469"
## [25] "1141716826787805" "5330755818114960" "1141716826787805"
## [28] "5928104758017036" "1141716826787805" "9190489069187153"
## [31] "1154673226500318" "2957228356035337" "1154673226500318"
## [34] "5330755818114960" "1154673226500318" "9190489069187153"
## [37] "1524924375222848" "5810353896294133" "2957228356035337"
## [40] "9190489069187153" "4739013535126469" "5330755818114960"
## [43] "4739013535126469" "5810353896294133" "4739013535126469"
## [46] "9190489069187153" "5330755818114960" "9190489069187153"
## [49] "5810353896294133" "9169095482512290" "5810353896294133"
## [52] "9190489069187153" "7294738832905991" "9190489069187153"
## [55] "9169095482512290" "9190489069187153"
```

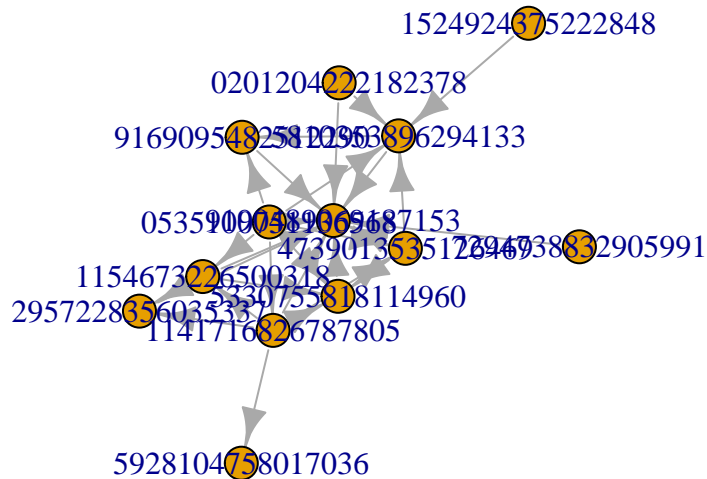
```
g_biomed<-graph(list_biomed, directed=TRUE)
g_biomed[ 1:length(g_biomed) ]
```

```
## 10 x 13 sparse Matrix of class "dgCMatrix"
```

```
## [[ suppressing 13 column names '0201204222182378', '5810353896294133', '9190489069187153' ... ]]
```

```
##
## 0201204222182378 . 1 1 . . . . .
## 5810353896294133 . . 1 . . . . 1 . . .
## 9190489069187153 . . . . . . . . .
## 0535100751136568 . 1 1 . 1 1 1 1 1 . . .
## 1141716826787805 . . 1 . . 1 1 1 . 1 1 .
## 1154673226500318 . . 1 . . . . 1 . 1 . .
## 4739013535126469 . 1 1 . . . . 1 . . . .
## 5330755818114960 . . 1 . . . . . . . .
## 9169095482512290 . . 1 . . . . . . . .
## 2957228356035337 . . 1 . . . . . . . .
```

```
plot(g_biomed)
```



Assim podemos ver algumas relações entre os professores. Abaixo segue o id do professor e seu nome loga em seguida.

```
c(rbind(unb.graph.biomed$nodes$id,unb.graph.biomed$nodes$properties$name))
```

```
## [1] "0201204222182378"
## [2] "Leandro Xavier Cardoso"
## [3] "0535100751136568"
## [4] "Cristiano Jacques Miosso Rodrigues Mendes"
## [5] "1141716826787805"
## [6] "Adson Ferreira da Rocha"
## [7] "1154673226500318"
## [8] "Suélia de Siqueira Rodrigues Fleury Rosa"
## [9] "1524924375222848"
## [10] "Marcus Vinícius Chaffim Costa"
## [11] "2957228356035337"
## [12] "Jose Felicio da Silva"
## [13] "4086384842130773"
## [14] "Ronni Geraldo Gomes de Amorim"
## [15] "4739013535126469"
## [16] "Marcelino Monteiro de Andrade"
## [17] "4839052902231824"
## [18] "Marcella Lemos Brettas Carneiro"
## [19] "5330755818114960"
## [20] "Fabiano Araujo Soares"
## [21] "5810353896294133"
## [22] "Georges Daniel Amvame Nze"
```

```
## [23] "5928104758017036"  
## [24] "Sergio Ricardo Menezes Mateus"  
## [25] "7294738832905991"  
## [26] "Vera Regina Fernandes da Silva Marães"  
## [27] "9169095482512290"  
## [28] "Marilia Miranda Forte Gomes"  
## [29] "9190489069187153"  
## [30] "Lourdes Mattos Brasil"
```

Expondo engenharia automação

Expondo engenharia elétrica p8

Expondo engenharia elétrica p7

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

No caso deste presente trabalho, é basicamente os scripts desenvolvido ao decorrer do mesmo.

Conclusão

Ao final deste trabalho, podemos chegar a conclusão que com a “ferramenta” CRISP-DM foi possível realizar uma análise de dados apartir de vários arquivos **JSON**, assim, foi possível perceber como funciona - pelo menos de maneira simplória - a vida de cientista de dados, como se deve preparar os dados, como se deve buscar informações úteis, analisar determinadas informações, plotar gráficos para um efeito mais visual, ver a dificuldade que falta de padronização causa, ainda mais quando se tem os mais diversos arquivos para analisar das mais diversas fontes.

O trabalho usou como base o arquivo disponibilizado pelo professor, no qual continha as explicações de todas as fase do CRISP-DM, seguindo o modelo passado pelo professor, tentou-se ir percorrendo sobre o script em R que se vinha fazendo e correlacionando com alguma fase do CRISP-DM, mas tal divisão não pode ser considerada totalmente fidedigna, haja visto que muitas vezes há mistura de fases e não apenas fases totalmente isoladas, sem influência de outras, procurou-se explicar de maneira simples como se encaixava cada fase.

Por fim, podemos chegar a alguns resultados valorosos ,como por exemplo : as revistas que tem mais publicações advindas da pós-graduação em engenharia biomédica, mas nem tudo são flores, algumas dificuldades vieram em decorrência da falta de uniformidade, porém podemos aprender bastante com este trabalho e nos interar de como começar a fazer análise de dados.

Referências