

# Ciência de Dados para Todos (Data Science For All) - 2018.2 - Análise da Produção Científica e Acadêmica da Universidade de Brasília

Engenharia Biomédica, Engenharia de Sistemas Eletrônicos e de Automação, Engenharia Elétrica P7 e Engenharia Elétrica P8.

*Marcos Vinicius Prescendo Tonin, Lucas Nascimento Santos Souza e Carlos Aragão*

*02/12/2018*

## Introdução

O presente trabalho tem por finalidade entender, correlacionar e analisar os dados da produção acadêmica e científica dos seguintes cursos de pós-graduação: Engenharia Biomédica, Engenharia Elétrica em seus dois ramos (53001010059P8 e 53001010080P7) e Engenharia de Sistemas Eletrônicos e de Automação.

## 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 da metodologia CRISP-DM.

O ciclo é 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 inicialmente.
- **Preparação de dados:** Então os dados são tratados para que se possa facilitar o processamento dos mesmos.
- **Modelagem:** Realiza a modelagem do dado.
- **Avaliação:** Avalia-se o resultado da modelagem.
- **Implantação:** Utilização dos dados.

Para melhor visualização das fases e a comunicação entre elas, deve-se observar a Figura 1. Por esta figura consegue-se perceber que, por exemplo, o entendimento do negócio e do dado pode muitas vezes ser se alternar no ciclo da metodologia adotada.

## Delimitações iniciais

### Domínio de Aplicação do projeto

O domínio de Aplicação do projeto é a produção científica e acadêmica dos docentes e discentes dos seguintes cursos de pós-graduação da Universidade de Brasília: **Engenharia Biomédica, Engenharia Elétrica P7 e P8 e Engenharia de Sistemas Eletrônicos e de Automação**. Os dados para análise foram obtidos a partir da plataforma elattes e da plataforma sucupira.

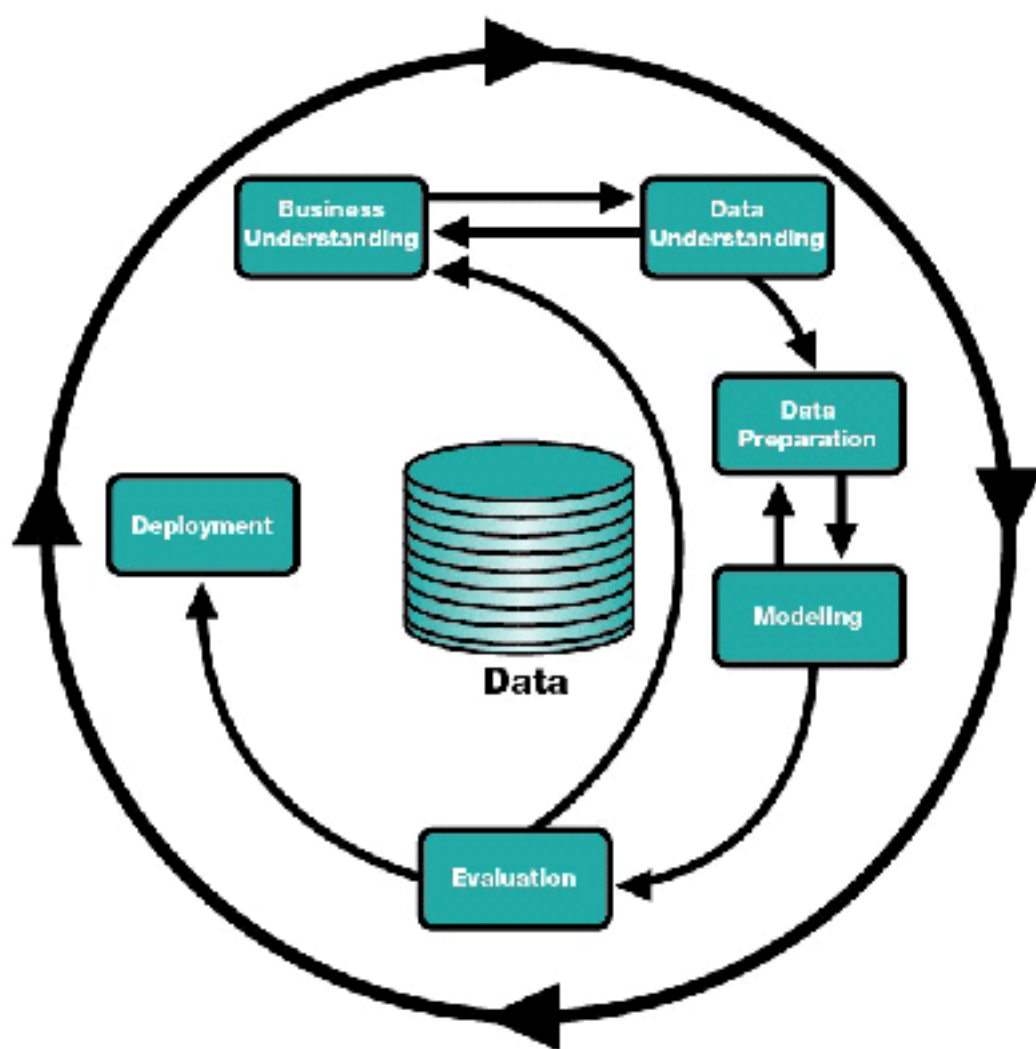


Figure 1: Ciclo do CRISP-DM que será usado como base

## Problema abordado

O problema 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 nos cursos de pós-graduação descritos anteriormente.

## Engenharia Biomédica

O programa de pós-graduação da engenharia biomédica - possui as informações: Código: 53001010083P6; Área-base: 31300006.

O corpo discente é composto de 15 membros, já o docente é de 673, conforme scripts.

Numero de disciplina retirado do site <https://fga.unb.br/pgengbiomedica/lista-de-oferta> : 9 disciplinas.

Na área de concetração engenharia biomédica, há 4 áreas(linhas) de pesquisa:

- **Análise e Desenvolvimento de Sistemas Inteligentes e de Saúde :** " Relaciona os aspectos ligados à análise, desenvolvimento, aquisição, representação e descoberta do conhecimento para a produção de sistemas inteligentes e de saúde. É uma área multidisciplinar, que apoia a tomada de decisão, diagnóstico, tratamento e desenvolvimento de técnicas para prática na saúde. Aborda os principais avanços teóricos e práticos de metodologias e técnicas utilizadas no desenvolvimento de Sistemas Inteligentes e de Saúde. As principais áreas de investigação são: Análise e modelagem de dados epidemiológicos; Análise de sistemas de saúde; Gestão e avaliação de tecnologias em saúde; Engenharia Clínica; Inteligência Artificial; Aprendizado de Máquina; Sistemas Baseados em Conhecimento; Paradigmas Simbólicos; Conexionistas, Evolutivos e Híbridos; Inteligência Computacional; Nanotecnologia; entre outras."
- **Física Médica e Bioengenharia.** " Relaciona a aplicação e as consequências das radiações ionizantes e não ionizantes na área da saúde, que a cada dia se tornam mais complexas e desafiadoras. Isso exige que os profissionais da área possuam uma formação específica de grande qualidade em diferentes campos, tais como física, biologia e química. Ademais, estes profissionais devem dominar os fundamentos de proteção radiológica e conhecer normas nacionais e recomendações internacionais a ela relacionadas, visando a sua própria proteção, bem como à proteção do paciente, do público geral e do meio ambiente. "
- **Sistemas Eletrônicos e Instrumentação Biomédica :** " Abrange a pesquisa e o desenvolvimento de transdutores, equipamentos, aparelhos ou sistemas microprocessados de uso biológico ou aplicação médica, odontológica ou laboratorial. A linha aborda o desenvolvimento e análise tanto de software como de hardware, abrangendo componentes e técnicas de microeletrônica, circuitos integrados e dispositivos programáveis. A ênfase diz respeito a aplicações em prevenção, diagnóstico, tratamento ou reabilitação. Em comparação com outros produtos de engenharia, a instrumentação biomédica se distingue por obter seus sinais de entrada a partir de seres vivos ou seus extratos, e ainda por aplicar energia (eletromagnética, mecânica, térmica) diretamente a tecidos vivos. As principais áreas de investigação são: Desenvolvimento de eletrodos ativos; Aquisição de sinais biológicos e biomédicos; Estimulação de tecidos ou sistemas biológicos e biomédicos; Desenvolvimento e análise de dispositivos microeletrônicos; Projetos e configuração de dispositivos programáveis; Projeto de circuitos integrados. "
- **Processamento de Sinais e Imagens :** " Sinais e imagens são representações da realidade, na forma de funções matemáticas a serem submetidas a métodos rigorosos e eficientes de extração e tratamento de informação. Pode-se definir um sinal como um conjunto físico, abstrato ou simbólico de dados de interesse específico (i.e., informação) que muda ao longo do tempo ou do espaço. Uma imagem, especificamente, é conceituada como uma representação visual, real ou pictórica, parcial ou total, de uma estrutura, um ser ou até mesmo de ideias. É possível agrupar sinais e imagens como um mesmo tipo de objeto matemático, sendo que as imagens constituem sinais em domínios de duas ou mais dimensões. Processamento de sinais e imagens, portanto, consiste em utilizar métodos científicos e técnicas de engenharia para trabalhar com toda a informação contida nestas representações, conforme definido pela Sociedade de Processamento de Sinais do IEEE em 2009. Isto significa o uso de representações, formalismos e técnicas matemáticas, estatísticas, computacionais, heurísticas e/ou linguísticas

na informação existente para fins de representação, modelagem, análise, síntese, revelação, recuperação, percepção, identificação, aquisição, extração, aprendizagem, segurança e/ou análise forense. Na engenharia biomédica, a linha de processamento de sinais é fundamental para aplicações como desenvolvimento de sistemas de auxílio a diagnóstico, formação de imagens médicas, controle de próteses e outros dispositivos médicos, desenvolvimento de métodos eficientes de tratamento e condicionamento físico, análise científica de dados relacionados à saúde etc.”

Informações retiradas do próprio site, disponível em :[https://fga.unb.br/articles/0001/4233/Linhas\\_de\\_Pesquisa\\_2016.pdf](https://fga.unb.br/articles/0001/4233/Linhas_de_Pesquisa_2016.pdf) .

Há muitas projetos de pesquisas, no total 41 projetos. Sendo descrito no site <https://fga.unb.br/pgengbiomedica/projetos-de-pesquisas>:

Projetos de Pesquisa	Labs	Período	Financiador Linha	Situação
AUTOBAC - Plataforma Embarcada para a Identificação de Bactérias por Processamento de Imagens	LIPIS	2011- Atual	Financiadora de Estudos e Projetos	Andamento
Biomaterial Látex	BioEng	2011- 2016	Capes	Andamento
Acompanhamento do crescimento e desenvolvimento de recém-nascidos pré-termo		2009- Atual	Instrumentação e Modelagem Biomédica	Andamento
Arte e Tecnociência: interações expandidas e condição cívica em software art	LART	2010- 2014	Capes/CNPq	Processamento de Sinais Biomédicos e Imagens Médicas
Atlas Anatômico 3D Aplicado à Mama	LIS	2009- 2010	FINATEC	Sistemas Inteligentes Aplicados à Eng. Biomédica
Biosensores Aplicados no Controle da Obesidade e Diabete Mellitus	BioEng	2010- 2015	CNPq	Instrumentação e Modelagem Biomédica
Bolsa de Produtividade em Pesquisa - Construção de um Ambiente Web de Ensino e Aprendizagem para a Área de Saúde	LIS	2010- 2012	CNPq	Sistemas Inteligentes Aplicados à Eng. Biomédica
CIBERCOMUNICAÇÃO: interatividade, imersão, autonomia e mobilidade em software art - PQ 1B	LART	2009- 2010	CNPq	Projeto Isolado
Colchão Inteligente Derivado de Latex anti-escara	BioEng	2009- 2014		Instrumentação e Modelagem Biomédica
Computação Massivamente Paralela Utilizando Hardware Gráfico		2008- Atual	FAP/DF	Instrumentação e Modelagem Biomédica
Consultório Médico Virtual (CMV)	LIS	2009- Atual	UnB/Gama	Sistemas Inteligentes Aplicados à Eng. Biomédica

Projetos de Pesquisa	Labs	Período	Financiador	Linha	Situação
Desenvolvimento de Dispositivo de Infusão Contínua de Insulina (Bomba de Insulina)		2010-2011		Instrumentação e Modelagem Biomédica	Andamento
Desenvolvimento de Modelos e Busca de Parâmetros para Minimização da Energia usada na Reversão Elétrica da Fibrilação Cardíaca		2009-Atual	FINATEC	Processamento de Sinais Biomédicos e Imagens Médicas	Andamento
Desenvolvimento de um oclutor infantil, no formato de lente de contato, feito à base de látex natural aplicado a pacientes com tendência a desenvolver o olho ambliope.	BioEng	2009-2015		Instrumentação e Modelagem Biomédica	Andamento
Desenvolvimento de um sistema físico de controle de fluxo esofágico para o tratamento da diabetes mellitus	BioEng	2010-2012	FAP/DF	Instrumentação e Modelagem Biomédica	Andamento
Desenvolvimento de uma palmilha para pé diabético derivada de látex natural com controle de pressão plantar e com indução de neoformação tecidual.	BioEng	2009-2012	CAPEF	Instrumentação e Modelagem Biomédica	Andamento
Dinâmica de Sistemas Complexos com Aplicações em Biologia e Meio Ambiente		2007-2010		Instrumentação e Modelagem Biomédica	Andamento
Estudo da Exequibilidade, Tolerabilidade e Eficácia do Módulo Controlador de Fluxo Esofágico	BioEng	2009-2010		Instrumentação e Modelagem Biomédica	Andamento
Estudo de sistemas embarcáveis para controle de módulo de fluxo esofágico em aplicação para tratamento de obesidade em humanos	BioEng	2010-2013	INBD – Engenharia Eletrônica	Instrumentação e Modelagem Biomédica	Andamento
Estudo experimental para desenvolvimento de sistema de monitoração clínico e informatizado de fatores de risco para distúrbios cardiovasculares, por meio de avaliação de diversas variáveis do Sistema Nervoso Autônomo.		2010-2011		Instrumentação e Modelagem Biomédica	Andamento
IEA - Instrumentação Embarcada em Aeronave	LIPIS	2009-2010		Projeto Isolado	Andamento
MINIVANT - Desenvolvimento de um sistema de Mini-VANT para imageamento giro-estabilizado	LIPIS	2010-Atual	FINEP	Projeto Isolado	Andamento
Modelagem e caracterização de motores de ciclo Otto alimentados com etanol e Hidrogênio		2009-2011	FAP/DF	Projeto Isolado	Andamento
Modelo Matematico da Força de Resistencia do Osso Humano a Perfurações em Cirurgias Ortopédicas - Aplicação Osteotomia de Tibia		2010-2015		Instrumentação e Modelagem Biomédica	Andamento
Obesidade: Busca de Novo Método Tratamento, novo Índice de Diagnostico e Controle	BioEng	2010-2020	Capes	Instrumentação e Modelagem Biomédica	Andamento
Preditores de morbi/mortalidade em portadores de doenças crônico-degenerativas em regime de internação domiciliar: coorte prospectiva		2010-Atual		Sistemas Inteligentes Aplicados à Eng. Biomédica	Andamento

Projetos de Pesquisa	Labs	Período	Financiador	Linha	Situação
Prevalência de tabagismo nos adolescentes e jovens do ensino médio da escola pública em Ceilândia		2010- Atual		Sistemas Inteligentes Aplicados à Eng. Biomédica	Andamento
ProDC: Promover o Desempenho Cognitivo	LIS	2010- Atual		Sistemas Inteligentes Aplicados à Eng. Biomédica	Andamento
Propriedades físicas de nanopartículas fora do equilíbrio, modelagem e projeto automatizado de nanocircuitos eletrônicos		2010- 2012	CNPq	Instrumentação e Modelagem Biomédica	Andamento
Prótese de pé ativa para reabilitação e assistência a locomoção humana	BioEng	2009- 2015		Instrumentação e Modelagem Biomédica	Andamento
Pêndulo Invertido	BioEng	2009- 2011		Instrumentação e Modelagem Biomédica	Andamento
Re-estruturação de um dispositivo para UTI neonatal e Home Care - CPAP	BioEng	2010- 2012	Brakko - Cooperação	Instrumentação e Modelagem Biomédica	Andamento
SAPIO - Sistema de Aquisição e Processamento de Imagens Ovitrapas	LIPIS/	2008- 2010	FINEP	Processamento de Sinais Biomédicos e Imagens Médicas	Andamento
Sistema de aquisição de sinais de temperatura e vibração em cirurgia de acesso ao nervo facial		2010- 2012		Instrumentação e Modelagem Biomédica	Andamento
Sistema de Controle de Velocidade, Temperatura e Força em Cirurgias Ortopédicas		2009- 2014		Instrumentação e Modelagem Biomédica	Andamento
Transtorno de Déficit de Atenção e Hiperatividade (TDAH)	LIS	2010- 2010		Sistemas Inteligentes Aplicados à Eng. Biomédica	Andamento
Monitoramento eletrônico da variação de temperatura durante a cirurgia descompressiva do nervo facial estudo experimental em gatos	BioEng	2011- 2012	FPDF		Andamento
Biblioteca digital de imagens e sinais biomédicos	LabGS/DI/LIPIS	2011- 2012	UnB		Andamento
Determinação do fator Gama para Cálculo de Taxa de Exposição em Medicina Nuclear	FCE	2010- atual	UnB	Física Médica	Andamento

A organização original em uma única área de concentração (Engenharia Biomédica) com quatro linhas de pesquisa (Física Médica; Instrumentação e Modelagem Biomédica; Processamento de Sinais Biomédicos e Imagens Médicas; Sistemas Inteligentes Aplicados à Engenharia Biomédica) é coerente, consistente e em consonância com a proposta aprovada. O número de projetos (39 no total) é elevado para a dimensão do corpo docente, parecendo corresponder aos temas das dissertações dos alunos. Recomenda-se concentrar os estudos em projetos consistentes. A proposta de programa aponta para possíveis mudanças no futuro próximo, em

função de processo de credenciamento e credenciamento de novos docentes, visando aumentar a produção e melhorar sua distribuição. A proposta curricular é adequada: 24 créditos, sendo 16 gerais, obrigatórias e 8 optativas, das áreas. Admite dispensa de cursos com teste de proficiência e há iniciativa de ensino à distância. A proposta valoriza o estabelecimento de cooperações com grupos de pesquisa nacionais (12), internacionais (13) e empresas (4). Aponta ainda ação de interesse social/sanitário (controle da dengue) e nucleação com o IFMT (Cuiabá). A autoavaliação reflete um esforço de busca de amadurecimento e consolidação. O Programa conta com 10 laboratórios com aparato instrumental adequado para a realização das atividades de ensino e pesquisa. Ao longo do período, parte das atividades e Laboratórios funcionaram em instalações provisórias. Mais ao final, houve mudança para um novo edifício, o que interferiu no próprio desempenho acadêmico/científico do programa. Há disponibilidade de recursos para um edifício adicional, previsto para o próximo quadriênio.

Sobre a última avaliação do curso engenharia biomédica, serve: <https://sucupira.capes.gov.br/sucupira/public/consultas/avaliacao/viewPreenchimentoFicha.jsf?idFicha=5739&popup=true>. Verifica-se que as notas desde de 2010 até 2017 é 3 e que ainda continua em funcionamento.

Sobre a última avaliação de 2017 temos que :

Nos quesitos de avaliação :

- Proposta do programa : Bom
- Corpo Docente : Regular
- Corpo Discente, Teses e Dissertações : Bom
- Produção Intelectual : Regular
- Inserção Social : Muito Bom

Especificamente sobre a proposta do programa em 2017, tem-se que

Nos quesitos de avaliação :

- Proposta do programa : Bom
- Corpo Docente : Regular
- Corpo Discente, Teses e Dissertações : Bom
- Produção Intelectual : Regular
- Inserção Social : Muito Bom

“A organização original em uma única área de concentração (Engenharia Biomédica) com quatro linhas de pesquisa (Física Médica; Instrumentação e Modelagem Biomédica; Processamento de Sinais Biomédicos e Imagens Médicas; Sistemas Inteligentes Aplicados à Engenharia Biomédica) é coerente, consistente e em consonância com a proposta aprovada. O número de projetos (39 no total) é elevado para a dimensão do corpo docente, parecendo corresponder aos temas das dissertações dos alunos. Recomenda-se concentrar os estudos em projetos consistentes. A proposta de programa aponta para possíveis mudanças no futuro próximo, em função de processo de credenciamento e credenciamento de novos docentes, visando aumentar a produção e melhorar sua distribuição. A proposta curricular é adequada: 24 créditos, sendo 16 gerais, obrigatórias e 8 optativas, das áreas. Admite dispensa de cursos com teste de proficiência e há iniciativa de ensino à distância. A proposta valoriza o estabelecimento de cooperações com grupos de pesquisa nacionais (12), internacionais (13) e empresas (4). Aponta ainda ação de interesse social/sanitário (controle da dengue) e nucleação com o IFMT (Cuiabá). A autoavaliação reflete um esforço de busca de amadurecimento e consolidação. O Programa conta com 10 laboratórios com aparato instrumental adequado para a realização das atividades de ensino e pesquisa. Ao longo do período, parte das atividades e Laboratórios funcionaram em instalações provisórias. Mais ao final, houve mudança para um novo edifício, o que interferiu no próprio desempenho acadêmico/científico do programa. Há disponibilidade de recursos para um edifício adicional, previsto para o próximo quadriênio.” Informação retirada de : <https://sucupira.capes.gov.br/sucupira/public/consultas/avaliacao/viewPreenchimentoFicha.jsf?idFicha=5739&popup=true>

Sobre o corpo docente: “O número informado de docentes do Programa é estável (8 a 9 no período), mas compreende várias entradas e saídas de docentes no quadriênio, e alternâncias de classificação de docentes entre permanentes e colaboradores, sendo constituído por professores em tempo integral e dedicação exclusiva. O Programa conta com apenas um bolsista de Produtividade em Pesquisa do CNPq. A Proposta aponta para a necessidade de credenciamentos e descredenciamentos de docentes, deixando claro que o núcleo de docentes permanentes do Programa ainda não está consolidado. A carga letiva dos docentes permanentes encontra-se um pouco abaixo da média, mas está com boa distribuição. No caso da graduação, alcança “muito bom” em termos de carga didática, e o número de alunos de iniciação científica é de pouco mais de um por docente.” Informação retirada de : <https://sucupira.capes.gov.br/sucupira/public/consultas/avaliacao/viewPreenchimentoFicha.jsf?idFicha=5739&popup=true>

O corpo discente é descrito por “O número de dissertações concluídas por docente e o tempo médio de titulação são muito bons, enquanto o percentual de docentes responsável pelas titulações ano a ano é bom. Entretanto, resultou regular o indicador de qualidade das dissertações, que admite como produção intelectual mínima, relacionada à dissertação, um artigo completo em congresso ou periódico.” Informação retirada de : <https://sucupira.capes.gov.br/sucupira/public/consultas/avaliacao/viewPreenchimentoFicha.jsf?idFicha=5739&popup=true>

Já sobre produção intelectual :

Itens de Avaliação	Peso	Avaliação
4.1. Publicações qualificadas do Programa por docente permanente.	50.0	Fraco
4.2. Distribuição de publicações qualificadas em relação ao corpo docente permanente do Programa.	30.0	Regular
4.3. Produção técnica, patentes e outras produções consideradas relevantes.	20.0	Bom

“O programa apresentou como produção intelectual um total de 63 artigos completos no quadriênio. Entretanto, após a distribuição nos diferentes estratos do QUALIS, apenas 16 foram qualificados como produção relevante (extratos A1, A2 e B1). Tal montante corresponde a meio artigo relevante por docente por ano, aquém da média dos programas da área. Naturalmente, a autoria desses artigos concentra-se em uma fração inadequada do corpo docente. Cabe salientar que alguns docentes permanentes têm investido em solicitação de patentes. A lista de produção intelectual tem um alto número de trabalhos listados como capítulos de livros, e mesmo livros, mas não há justificativa da importância das obras conforme orientação do documento de área das Engenharias IV. Em vários casos, é claro tratar-se de comunicações em congressos (p/ex da série IFMBE Proceedings).” Informação retirada de : <https://sucupira.capes.gov.br/sucupira/public/consultas/avaliacao/viewPreenchimentoFicha.jsf?idFicha=5739&popup=true>

Por fim, sobre inserção social :

Itens de Avaliação	Peso	Avaliação
5.1. Inserção e impacto regional e (ou) nacional do programa.	40.0	Muito Bom
5.2. Integração e cooperação com outros programas e centros de pesquisa e desenvolvimento profissional relacionados à área de conhecimento do programa, com vistas ao desenvolvimento da pesquisa e da pós-graduação.	40.0	Muito Bom
5.3 - Visibilidade ou transparência dada pelo programa a sua atuação.	20.0	Muito Bom



Engenharia Elétrica P7

Engenharia Elétrica P8

Engenharia de Sistemas Eletrônicos e de Automação

## CRISP-DM Fase 1 - Entendimento do negócio

A Universidade de Brasília é uma das grandes universidades federais, tendo enfoque na produção de artigos acadêmicos. Os programas a seguir são todos da área de avaliação ENGENHARIAS IV e serão analisados pelo trabalho como já exposto anteriormente:

### Programas de pós-graduação Grupo 15

Nome do Programa	ME	DO	MP
ENGENHARIA ELÉTRICA (53001010059P8)	-	-	3
ENGENHARIA ELÉTRICA (53001010080P7)	4	4	-
ENGENHARIA DE SISTEMAS ELETRÔNICOS E DE AUTOMAÇÃO (53001010081P3)	4	4	-
ENGENHARIA BIOMÉDICA (53001010083P6)	3	-	-

Sendo o enfoque principal os dados referente a Engenharia Biomédica, que por sua vez possui muitas de suas pesquisas voltadas a novas soluções na área de Biomédica e demais áreas representadas na tabela 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)
json.perfil.auto <- "dados-2018-2/engenharia-eletronicos-automacao/280.profile.json"
file.info(json.perfil.auto)
json.perfil.ele_p8 <- "dados-2018-2/engenharia-eletrica-p8/281.profile.json"
file.info(json.perfil.ele_p8)
json.perfil.ele_p7 <- "dados-2018-2/engenharia-eletrica-p7/282.profile.json"
file.info(json.perfil.ele_p7)
```

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)
json.advise.auto <- "dados-2018-2/engenharia-eletronicos-automacao/280.advise.json"
file.info(json.advise.auto)
json.advise.ele_p8 <- "dados-2018-2/engenharia-eletrica-p8/281.advise.json"
file.info(json.advise.ele_p8)
json.advise.ele_p7 <- "dados-2018-2/engenharia-eletrica-p7/282.advise.json"
file.info(json.advise.ele_p7)
```

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 2018-12-02 15:20:10
##                                                                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 2018-12-02 15:20:08
##                                                                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 2018-12-02 15:20:10
##                                                                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 2018-12-02 15:20:10
##                                                                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 2018-12-02 15:20:10 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 2018-12-02 15:20:10
##                                     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 2018-12-02 15:20:10 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 2018-12-02 15:20:10 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 2018-12-02 15:20:10 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 2018-12-02 15:20:08
##                                     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 2018-12-02 15:20:10 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 2018-12-02 15:20:10 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 -----
## 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 -----
## 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

library(tm)
```

```
## Loading required package: NLP

##
## Attaching package: 'NLP'

## The following object is masked from 'package:ggplot2':
##
##   annotate
```

```
library(wordcloud)
```

```
## Loading required package: RColorBrewer
```

```
library(RColorBrewer)
library(SnowballC)
```

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

### 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.



```
## ..$ 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" "Engenharia Biomédica"
## ..$ 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_LATENTES"
## ..$ titulo : chr [1:5] "Insuficiência na Aprendizagem de Matemática do 9º ano"
## ..$ 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 Santos"
## ..$ instituicao : chr [1:5] "Faculdade Serigy" "Faculdade Serigy" "Faculdade Serigy"
## ..$ curso : chr [1:5] "Fundamentos e Métodos do ensino da Matemática" "Fundamentos e Métodos do ensino da Matemática"
## ..$ 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]])
```

```
## List of 7
```

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

```
## $ resumo_cv : chr "Engenheiro Eletricista pela Universidade de Brasília (1988), Mestre em Engenharia de Sistemas Eletrônicos e de Automação"
```

```
## $ 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 P
## .. ..$ 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 e
## ..$ 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 In
## ..$ 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_I
## ..$ titulo : chr [1:6] "Ricardo Minotto" "Eletrocardiograma Portátil Baseado
## ..$ 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" "Rafael
## ..$ 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"
```

Professor, senioridade 9, Adson Rocha formado na UnB, em 1988, em engenharia elétrica.

### 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, para poder ter uma noção.

```
### Usando glimpse
glimpse(unb.prof.ele_p8[[8]])
```

```
## 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" "Computa
## $ 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 ap
## .. ..$ 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 atrav
## .. ..$ titulo_do_livro : chr "Minicursos do XI Simpósio Brasileiro em Segurança da Informa
## .. ..$ 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.; Ra
## .. ..$ 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 Unco
## .. ..$ nome_do_evento : chr [1:22] "O Simpósio Brasileiro em Segurança da Informação e de Sistema
## .. ..$ 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, K
## .. ..$ 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 INFORM
## .. ..$ ano          : chr [1:3] "2011" "2016" "2016"
## .. ..$ tipo         : chr [1:3] "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
## .. ..$ nome_da_editora : chr [1:3] "Sociedade Brasileira de Computação" "Springer Internat
## .. ..$ 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 a
## .. ..$ periodico    : chr [1:16] "IEICE Transactions on Fundamentals of Electronics, Communicat
## .. ..$ 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 DI
## .. ..$ 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"
```

Trata-se do professor Anderson Clayton Alves Nascimento, formado na UnB em 1998, em engenharia elétrica, sendo sua especialidade em criptografia e sua senioridade é 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]])
```

```
## List of 7
## $ nome : chr "Sébastien Roland Marie Joseph RONDINEAU"
## $ resumo_cv : chr "- IEEE Senior Member. - Possui Mestrado em Ciencias da Comput., Proce
## $ 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 c
## .. .$ 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"
```

É o professor Sébastien, membro senior do IEEE, tem interesse em matemática aplicada na computação e em telecomunicações, atualmente está na UnB do Gama.

## Descrevendo dados de orientação

### Engenharia biomédica

```
###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 um, 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.

## Engenharia eletrônicos e automação

### Engenharia Elétrica p8

###Descrição dos dados de orientações

```
unb.adv.ele_p8 <- fromJSON("dados-2018-2/engenharia-eletrica-p8/281.advise.json")
##### DOUTORADO 2017
#Buscando cursos que mais produziram doutorados.
head(sort(table(unb.adv.ele_p8$ORIENTACAO_CONCLUIDA_DOUTORADO$`2017`$curso), decreasing = TRUE), 10)
```

```
##
## ENGENHARIA ELÉTRICA          Doutorado
##                2                1
```

*#Sabendo suas instituições*

```
head(sort(table(unb.adv.ele_p8$ORIENTACAO_CONCLUIDA_DOUTORADO$`2017`$instituicao), decreasing = TRUE), 10)
```

```

##
##
##
## Instituto de Computação - Universidade Estadual de Campinas
##
##
#Sabendo seus orientadores
data_orienM <- capture.output(str(unb.adv.ele_p8$ORIENTACAO_CONCLUIDA_DOUTORADO$`2017`$nome_orientadores))
unique(data_orienM)

## [1] "List of 3"
## [2] " $ : chr \"Anderson de Rezende Rocha\""
## [3] " $ : chr \"Joao Paulo Carvalho Lustosa da Costa\""
## [4] " $ : chr \"Rafael Timóteo de Sousa Jr\""

#Sabendo os nomes dos trabalhos
data_orienM <- capture.output(str(unb.adv.ele_p8$ORIENTACAO_CONCLUIDA_DOUTORADO$`2017`$titulo))
unique(data_orienM)

## [1] " chr [1:3] \"iSuper-resolution in low-quality videos for forensics, surveillance, and mobile ap
##### MESTRADO 2017
#Buscando cursos que mais produziram mestrados.
head(sort(table(unb.adv.ele_p8$ORIENTACAO_CONCLUIDA_MESTRADO$`2017`$curso), decreasing = TRUE), 10)

##
##
## Engenharia Elétrica
## 6
## ENGENHARIA ELÉTRICA
## 3
## Electrical Engineering and Information Technology
## 1
## Engenharia Biomédica
## 1
## Mestrado
## 1
## Sistemas Mecatrônicos
## 1
head(sort(table(unb.adv.ele_p8$ORIENTACAO_CONCLUIDA_MESTRADO$`2017`$instituicao), decreasing = TRUE), 10)

##
##
## Universidade de Brasília
## 11
## Instituto de Computação - Universidade Estadual de Campinas
## 1
## Technische Universität Ilmenau
## 1
#Sabendo seus orientadores
data_orien3 <- capture.output(str(unb.adv.ele_p8$ORIENTACAO_CONCLUIDA_MESTRADO$`2017`$nome_orientadores))
unique(data_orien3)

## [1] "List of 13"
## [2] " $ : chr \"Anderson de Rezende Rocha\""
## [3] " $ : chr \"Daniel Guerreiro e Silva\""
## [4] " $ : chr \"Díbio Leandro Borges\""
## [5] " $ : chr \"Georges Daniel Amvame Nze\""

```

```
## [6] " $ : chr \"Joao Paulo Carvalho Lustosa da Costa\""
## [7] " $ : chr \"Rafael Timóteo de Sousa Jr\""
```

*#Sabendo seus orientados*

```
data_orien2 <- capture.output(str(unb.adv.ele_p8$ORIENTACAO_CONCLUIDA_MESTRADO$`2017`$nome_aluno))
unique(data_orien2)
```

```
## [1] " chr [1:13] \"Geise Kelly da Silva Santos\" ..."
```

Mais uma vez, os dados mais atualizados são de 2017. Há apenas 3 doutorados concluídos em 2017, sendo duas na UnB e uma em conjunto com o Instituto de Computação. Os orientadores são:

- “Anderson de Rezende Rocha”
- “Joao Paulo Carvalho Lustosa da Costa”
- “Rafael Timóteo de Sousa Jr”

Um dos trabalhos tem o título : “iSuper-resolution in low-quality videos for forensics, surveillance, and mobile applications”.

Já no mestrado, podemos ver um grande problema da ciência de dados, o tratamento dos mesmos.

## Engenharia Elétrica p7

###Descrição dos dados de orientações

```
unb.adv.ele_p7 <- fromJSON("dados-2018-2/engenharia-eletrica-p7/282.advise.json")
##### DOUTORADO 2017
#Buscando cursos que mais produziram doutorados.
head(sort(table(unb.adv.ele_p7$ORIENTACAO_CONCLUIDA_DOUTORADO$`2017`$curso), decreasing = TRUE), 10)
```

```
## ENGENHARIA ELÉTRICA
##                2
```

*#Sabendo suas instituições*

```
head(sort(table(unb.adv.ele_p7$ORIENTACAO_CONCLUIDA_DOUTORADO$`2017`$instituicao), decreasing = TRUE), 10)
```

```
## Universidade de Brasília
##                2
```

*#Sabendo seus orientadores*

```
data_orienM <- capture.output(str(unb.adv.ele_p7$ORIENTACAO_CONCLUIDA_DOUTORADO$`2017`$nome_orientadores))
unique(data_orienM)
```

```
## [1] "List of 2"
## [2] " $ : chr \"Joao Paulo Carvalho Lustosa da Costa\""
## [3] " $ : chr \"Rafael Timóteo de Sousa Jr\""
```

*#Sabendo os nomes dos trabalhos*

```
data_orienM <- capture.output(str(unb.adv.ele_p7$ORIENTACAO_CONCLUIDA_DOUTORADO$`2017`$titulo))
unique(data_orienM)
```

```
## [1] " chr [1:2] \"Beamforming Techniques for Next Generation Communication Systems\" ..."
```

##### MESTRADO 2017

*#Buscando cursos que mais produziram mestrados.*

```
head(sort(table(toupper(unb.adv.ele_p7$ORIENTACAO_CONCLUIDA_MESTRADO$`2017`$curso), decreasing = TRUE), 10))
```

```
##
##                ENGENHARIA ELÉTRICA
##                12
```

```
## ELECTRICAL ENGINEERING AND INFORMATION TECHNOLOGY
##                                     1
## PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA
##                                     1

head(sort(table(unb.adv.ele_p7$ORIENTACAO_CONCLUIDA_MESTRADO$`2017`$instituicao), decreasing = TRUE), 1)

##
##      Universidade de Brasília Technische Universität Ilmenau
##              13                      1

#Sabendo seus orientadores
data_orien3 <- capture.output(str(unb.adv.ele_p7$ORIENTACAO_CONCLUIDA_MESTRADO$`2017`$nome_orientadores))
unique(data_orien3)

## [1] "List of 14"
## [2] " $ : chr \"Andre Noll Barreto\""
## [3] " $ : chr \"Daniel Guerreiro e Silva\""
## [4] " $ : chr \"Francisco Damasceno Freitas\""
## [5] " $ : chr \"Joao Paulo Carvalho Lustosa da Costa\""
## [6] " $ : chr \"Kleber Melo e Silva\""
## [7] " $ : chr \"Rafael Timóteo de Sousa Jr\""

#Sabendo seus orientados
data_orien2 <- capture.output(str(unb.adv.ele_p7$ORIENTACAO_CONCLUIDA_MESTRADO$`2017`$nome_aluno))
unique(data_orien2)

## [1] " chr [1:14] \"Sávio Oliveira de Oliveira Neves\" \"Lucas Sousa e Silva\" ..."
```

Há apenas três doutorados concluídos em 2017 e 14 mestados neste mesmo ano. Os professores Rafael Timóteo e João da Costa orientaram mestrados e doutorados concluídos em 2017. Houve um aluno da Universitat Ilmenau - Binghua Guo, “Low Complexity Schemes to Semi-Blind Joint Channel and Symbol Estimation in MIMO-OFDM” .

## 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 há 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 tem 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 há 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 tem matemática para obter quaisquer dado plausível, somente os id.

### Descrição dos dados de Produção bibliográfica

#### TODO LIST : VERIFICAR

- 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] "TEXTO_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))

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

Verificando engenharia automação

Verificando engenharia elétrica p8

```
unb.pub.ele_p8 <- fromJSON("dados-2018-2/engenharia-eletrica-p8/281.publication.json")
# Verificando os tipos de produções que existe.
names(unb.pub.ele_p8)
```

```
## [1] "PERIODICO"
## [2] "LIVRO"
## [3] "CAPITULO_DE_LIVRO"
## [4] "TEXTO_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.ele_p8$PERIODICO$`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.ele_p8$PERIODICO$`2017`$periodico), decreasing = TRUE), 10)
```

```
##
##                                     IEEE Access
##                                     2
##             IEEE Transactions on Information Forensics and Security
##                                     2
##                                     JOURNAL OF MOLECULAR MODELING
```



```
##
##
## Applied Sciences-Basel 2
## 1
## CLEI ELECTRONIC JOURNAL 1
## 1
## Cluster Computing 1
## Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization 1
## Concurrency and Computation 1
## Digital Signal Processing (Print) 1
## IEEE Journal of Biomedical and Health Informatics 1
```

```
#Nomes dos periódicos em que tiveram mais publicações de artigos aceitos na área de engenharia Biomédica
head(sort(table(unb.pub.ele_p8$ARTIGO_ACEITO$`2017`$periodico), decreasing = TRUE), 10)
```

```
## IEEE TRANSACTIONS ON INFORMATION THEORY
## 1
```

```
#Nomes dos autores que produziram um tipo de Produção que não estava contemplada em 2011 .
head(sort(table(unb.pub.ele_p8$DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA$`2011`$autores), decreasing = TRUE), 10)
```

```
## ARANHA, D. F.;ARANHA, DIEGO F.
## 1
```

### Verificando engenharia elétrica p7

```
unb.pub.ele_p7 <- fromJSON("dados-2018-2/engenharia-eletrica-p7/282.publication.json")
# Verificando os tipos de produções que existe.
names(unb.pub.ele_p7)
```

```
## [1] "PERIODICO"
## [2] "LIVRO"
## [3] "CAPITULO_DE_LIVRO"
## [4] "TEXTO_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.ele_p7$PERIODICO$`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.ele_p7$PERIODICO$`2017`$periodico), decreasing = TRUE), 10)
```

```
##
## JOURNAL OF MOLECULAR MODELING 2
## Applied Sciences-Basel 1
##
```

```
##                                CLEI ELECTRONIC JOURNAL
##                                1
##                                Cluster Computing
##                                1
##                                Digital Signal Processing (Print)
##                                1
##                                IEEE Access
##                                1
##                                IEEE ANTENNAS AND PROPAGATION MAGAZINE
##                                1
##                                IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION
##                                1
##                                IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT
##                                1
##                                IEEE Transactions on Power Delivery
##                                1

#Nomes dos periódicos em que tiveram mais publicações de artigos aceitos na área de engenharia Biomédica
head(sort(table(unb.pub.ele_p7$ARTIGO_ACEITO$`2017`$periodico), decreasing = TRUE), 10)

## IEEE TRANSACTIONS ON COMMUNICATIONS
##                                1

#Nomes dos autores que produziram um tipo de Produção que não estava contemplada em 2014.
head(sort(table(unb.pub.ele_p7$DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA$`2014`$autores), decreasing = TRUE), 10)

## integer(0)
```

Descrição dos dados de redes de colaboração

TODO LIST : VERIFICAR

- 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" ...
```

## Engenharia automação e eletrônica

### Engenharia elétrica p8

```
unb.graph.ele_p8 <- fromJSON("dados-2018-2/engenharia-eletrica-p8/281.graph.json")  
# Suas variáveis  
names(unb.graph.ele_p8)
```

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

```
# Quantidade de nós
```

```
length(unb.graph.ele_p8$nodes$id)
```

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

```
# Quantidade de links de fonte
```

```
length(unb.graph.ele_p8$links$source)
```

```
## [1] 6
```

```
# Quantidade de links de chegada
```

```
length(unb.graph.ele_p8$links$target)
```

```
## [1] 6
```

```
#exemplos de pesos da aresta
```

```
str(unb.graph.ele_p8$links$weight)
```

```
## chr [1:6] "27" "2" "2" "3" "4" "1"
```

### Engenharia elétrica p7

```
unb.graph.ele_p7 <- fromJSON("dados-2018-2/engenharia-eletrica-p7/282.graph.json")  
# Suas variáveis  
names(unb.graph.ele_p7)
```

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

```
# Quantidade de nós
```

```
length(unb.graph.ele_p7$nodes$id)
```

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

```
# Quantidade de links de fonte
```

```
length(unb.graph.ele_p7$links$source)
```

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

```
# Quantidade de links de chegada
```

```
length(unb.graph.ele_p7$links$target)
```

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

```
#exemplos de pesos da aresta
```

```
str(unb.graph.ele_p7$links$weight)
```

```
## chr [1:5] "27" "2" "13" "7" "1"
```

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

### Arquivo Profile

#### TODO LIST : VERIFICAR

- 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_BIBLIOG)))

## [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_BIBLIOG))))

##
## 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_EM_JORNAIS$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$ano))))
```

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

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

## [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_D

##
## 0 1 2 3 4
## 9 2 1 2 1
table(unlist(sapply(unb.prof.biomed, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_P

##
## 0 2
## 14 1
# número de orientações por ano
table(unlist(sapply(unb.prof.biomed, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTRAD

##
## 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_DOUTORAI

##
## 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_DOU

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

Pelos dados, a maioria dos profissionais atuam na área de engenharia (34). A maioria da produção científica é dada pela publicações em periódicos. Um fato interessante é que em 2015 houve um grande número de capítulos de livro publicado, fugindo do padrão de 4. Há apenas um texto em jornal publicado em 2010. Houve apeans dois pós-doutorado feito em biomédica na UnB em 2014 e 2017.

## Verificando engenharia automação

### Verificando engenharia elétrica p8

```
# Total de áreas de atuação de todos profissionais
sum(sapply(unb.prof.ele_p8, function(x) nrow(x$areas_de_atuacao)))
```

```
## [1] 45
```

```
# número de pessoas por grande area  
table(unlist(sapply(unb.prof.ele_p8, function(x) (x$areas_de_atuacao$grande_area))))
```

```
##  
## CIENCIAS_EXATAS_E_DA_TERRA          CIENCIAS_HUMANAS  
##                20                      1  
##                ENGENHARIAS  
##                24
```

```
# número de pessoas que produziram os tipos de Produção específico  
table(unlist(sapply(unb.prof.ele_p8, function(x) names(x$producao_bibliografica))))
```

```
##  
##                ARTIGO_ACEITO  
##                1  
##                CAPITULO_DE_LIVRO  
##                7  
## DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA  
##                1  
##                EVENTO  
##                10  
##                LIVRO  
##                6  
##                PERIODICO  
##                10  
##                TEXTO_EM_JORNAIS  
##                3
```

```
# número de publicações por tipo  
#####ARTIGO ACEITO#####  
sum(sapply(unb.prof.ele_p8, function(x) length(x$producao_bibliografica$ARTIGO_ACEITO$ano)))
```

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

```
# número de pessoas por quantitativo de produções por pessoa 0 = 1; 1 = 2...  
table(unlist(sapply(unb.prof.ele_p8, function(x) length(x$producao_bibliografica$ARTIGO_ACEITO$ano))))
```

```
##  
## 0 1  
## 10 1
```

```
#####CAPITULO DE LIVRO#####  
sum(sapply(unb.prof.ele_p8, function(x) length(x$producao_bibliografica$CAPITULO_DE_LIVRO$ano)))
```

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

```
# número de pessoas por quantitativo de produções por pessoa 0 = 1; 1 = 2...  
table(unlist(sapply(unb.prof.ele_p8, function(x) length(x$producao_bibliografica$CAPITULO_DE_LIVRO$ano))))
```

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

```
#####LIVRO#####  
sum(sapply(unb.prof.ele_p8, function(x) length(x$producao_bibliografica$LIVRO$ano)))
```

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



```

# número de pessoas por quantitativo de produções por pessoa 0 = 1; 1 = 2...
table(unlist(sapply(unb.prof.ele_p8, function(x) length(x$producao_bibliografica$LIVRO$ano))))

##
## 0 1 2 3 4 9
## 5 2 1 1 1 1

#####periódico#####
sum(sapply(unb.prof.ele_p8, function(x) length(x$producao_bibliografica$PERIODICO$ano)))

## [1] 179

# número de pessoas por quantitativo de produções por pessoa 0 = 1; 1 = 2...
table(unlist(sapply(unb.prof.ele_p8, function(x) length(x$producao_bibliografica$PERIODICO$ano))))

##
## 0 1 2 3 9 16 28 45 63
## 1 1 1 2 2 1 1 1 1

#####TEXTO EM JORNAIS#####
sum(sapply(unb.prof.ele_p8, function(x) length(x$producao_bibliografica$TEXTO_EM_JORNAIS$ano)))

## [1] 14

# número de pessoas por quantitativo de produções por pessoa 0 = 1; 1 = 2...
table(unlist(sapply(unb.prof.ele_p8, function(x) length(x$producao_bibliografica$TEXTO_EM_JORNAIS$ano))))

##
## 0 2 4 8
## 8 1 1 1

#####DEMAIS TIPOS#####
sum(sapply(unb.prof.ele_p8, function(x) length(x$producao_bibliografica$DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRFICA$ano)))

## [1] 3

# número de pessoas por quantitativo de produções por pessoa 0 = 1; 1 = 2...
table(unlist(sapply(unb.prof.ele_p8, function(x) length(x$producao_bibliografica$DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRFICA$ano))))

##
## 0 3
## 10 1

# número de produções por ano
table(unlist(sapply(unb.prof.ele_p8, function(x) (x$producao_bibliografica$ARTIGO_ACEITO$ano))))

##
## 2017
## 1

table(unlist(sapply(unb.prof.ele_p8, function(x) (x$producao_bibliografica$CAPITULO_DE_LIVRO$ano))))

##
## 2010 2011 2012 2013 2014 2015 2016 2017
## 2 5 4 2 2 5 4 2

table(unlist(sapply(unb.prof.ele_p8, function(x) (x$producao_bibliografica$LIVRO$ano))))

##
## 2010 2011 2013 2014 2015 2016 2017
## 1 3 7 2 2 3 2

```

```

table(unlist(sapply(unb.prof.ele_p8, function(x) (x$producao_bibliografica$PERIODICO$ano))))

##
## 2010 2011 2012 2013 2014 2015 2016 2017
##    8   18   19   18   18   21   39   38

table(unlist(sapply(unb.prof.ele_p8, function(x) (x$producao_bibliografica$TEXTOS_EM_JORNAIS$ano))))

##
## 2011 2012 2013 2015 2017
##    3    6    1    2    2

# número de pessoas que realizaram diferentes tipos de orientações
length(unlist(sapply(unb.prof.ele_p8, function(x) names(x$orientacoes_academicas))))

## [1] 45

# número de pessoas por tipo de orientação
table(unlist(sapply(unb.prof.ele_p8, function(x) names(x$orientacoes_academicas))))

##
##          ORIENTACAO_CONCLUIDA_DOUTORADO
##                                5
##          ORIENTACAO_CONCLUIDA_MESTRADO
##                                10
##          ORIENTACAO_CONCLUIDA_POS_DOUTORADO
##                                3
##          ORIENTACAO_EM_ANDAMENTO_DOUTORADO
##                                5
##          ORIENTACAO_EM_ANDAMENTO_GRADUACAO
##                                2
## ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA
##                                5
##          ORIENTACAO_EM_ANDAMENTO_MESTRADO
##                                7
##          OUTRAS_ORIENTACOES_CONCLUIDAS
##                                8

#número de orientações concluídas
sum(sapply(unb.prof.ele_p8, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTRADO$ano)))

## [1] 108

sum(sapply(unb.prof.ele_p8, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTORADO$ano)))

## [1] 17

sum(sapply(unb.prof.ele_p8, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_POS_DOUTORADO$ano)))

## [1] 12

# número de pessoas por quantitativo de orientações por pessoa 0 = 1; 1 = 2...
table(unlist(sapply(unb.prof.ele_p8, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTRADO$ano))))

##
## 0  4  5  6  7 13 15 20 28
## 1  2  1  2  1  1  1  1  1

```

```

table(unlist(sapply(unb.prof.ele_p8, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTORADO)),
##
## 0 1 2 5 8
## 6 2 1 1 1
table(unlist(sapply(unb.prof.ele_p8, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_POS_DOUTORADO)),
##
## 0 1 10
## 8 2 1
# número de orientações por ano
table(unlist(sapply(unb.prof.ele_p8, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTRADO))),
##
## 2010 2011 2012 2013 2014 2015 2016 2017
## 9 10 26 12 11 5 22 13
table(unlist(sapply(unb.prof.ele_p8, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTORADO))),
##
## 2012 2013 2014 2015 2016 2017
## 3 3 3 2 3 3
table(unlist(sapply(unb.prof.ele_p8, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_POS_DOUTORADO))),
##
## 2012 2013 2014 2015 2016 2017
## 1 1 1 4 3 2
# números de orientações não concluídas por ano
# Trabalho de mestrado em andamento
table(unlist(sapply(unb.prof.ele_p8, function(x) (x$orientacoes_academicas$ORIENTACAO_EM_ANDAMENTO_MESTRADO))),
##
## 2015 2016 2017
## 2 4 12
# DOUTORADO em andamento
table(unlist(sapply(unb.prof.ele_p8, function(x) (x$orientacoes_academicas$ORIENTACAO_EM_ANDAMENTO_DOUTORADO))),
##
## 2013 2014 2015 2016 2017
## 1 5 5 6 15
# Trabalho de Graduação em andamento
table(unlist(sapply(unb.prof.ele_p8, function(x) (x$orientacoes_academicas$ORIENTACAO_EM_ANDAMENTO_GRADUACAO))),
##
## 2017
## 4

```

Em engenharia elétrica p8, tem uma boa divisão das grandes áreas dos orientadores, 20 para ciências exatas e da terra e 24 para engenharias, que é maioria de todas as pesquisas feitas. Houve 108 orientações concluídas de mestrado e apenas 17 doutorado, que não é tão diferente de pós-doutorados: 12. Há um doutorado de 2013 que não foi concluído ainda.

**Verificando engenharia elétrica p7**

```

# Total de áreas de atuação de todos profissionais
sum(sapply(unb.prof.ele_p7, function(x) nrow(x$areas_de_atuacao)))

## [1] 55

# número de pessoas por grande area
table(unlist(sapply(unb.prof.ele_p7, function(x) (x$areas_de_atuacao$grande_area))))

##
##              CIENCIAS_EXATAS_E_DA_TERRA
##              1                      15
##      CIENCIAS_HUMANAS              ENGENHARIAS
##              1                      38

# número de pessoas que produziram os tipos de Produção específico
table(unlist(sapply(unb.prof.ele_p7, function(x) names(x$producao_bibliografica))))

##
##      ARTIGO_ACEITO CAPITULO_DE_LIVRO          EVENTO          LIVRO
##              1              8              14              4
##      PERIODICO  TEXTO_EM_JORNAIS
##              14              4

# número de publicações por tipo
#####ARTIGO ACEITO#####
sum(sapply(unb.prof.ele_p7, function(x) length(x$producao_bibliografica$ARTIGO_ACEITO$ano)))

## [1] 1

# número de pessoas por quantitativo de produções por pessoa 0 = 1; 1 = 2...
table(unlist(sapply(unb.prof.ele_p7, function(x) length(x$producao_bibliografica$ARTIGO_ACEITO$ano))))

##
##  0  1
## 13  1

#####CAPITULO DE LIVRO#####
sum(sapply(unb.prof.ele_p7, function(x) length(x$producao_bibliografica$CAPITULO_DE_LIVRO$ano)))

## [1] 30

# número de pessoas por quantitativo de produções por pessoa 0 = 1; 1 = 2...
table(unlist(sapply(unb.prof.ele_p7, function(x) length(x$producao_bibliografica$CAPITULO_DE_LIVRO$ano))))

##
##  0  1  4  5  6  8
##  6  3  2  1  1  1

#####LIVRO#####
sum(sapply(unb.prof.ele_p7, function(x) length(x$producao_bibliografica$LIVRO$ano)))

## [1] 14

# número de pessoas por quantitativo de produções por pessoa 0 = 1; 1 = 2...
table(unlist(sapply(unb.prof.ele_p7, function(x) length(x$producao_bibliografica$LIVRO$ano))))

##
##  0  1  2  9
## 10  1  2  1

```

```
#####periódico#####
sum(sapply(unb.prof.ele_p7, function(x) length(x$producao_bibliografica$PERIODICO$ano)))

## [1] 178

# número de pessoas por quantitativo de produções por pessoa 0 = 1; 1 = 2...
table(unlist(sapply(unb.prof.ele_p7, function(x) length(x$producao_bibliografica$PERIODICO$ano))))

##
## 3 5 7 8 9 10 13 17 28 45
## 1 1 1 4 2 1 1 1 1 1

#####TEXTO EM JORNAIS#####
sum(sapply(unb.prof.ele_p7, function(x) length(x$producao_bibliografica$TEXTO_EM_JORNAIS$ano)))

## [1] 13

# número de pessoas por quantitativo de produções por pessoa 0 = 1; 1 = 2...
table(unlist(sapply(unb.prof.ele_p7, function(x) length(x$producao_bibliografica$TEXTO_EM_JORNAIS$ano))))

##
## 0 1 3 8
## 10 2 1 1

#####DEMAIS TIPOS#####
sum(sapply(unb.prof.ele_p7, function(x) length(x$producao_bibliografica$DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOG)))

## [1] 0

# número de pessoas por quantitativo de produções por pessoa 0 = 1; 1 = 2...
table(unlist(sapply(unb.prof.ele_p7, function(x) length(x$producao_bibliografica$DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOG))))

##
## 0
## 14

# número de produções por ano
table(unlist(sapply(unb.prof.ele_p7, function(x) (x$producao_bibliografica$ARTIGO_ACEITO$ano))))

##
## 2017
## 1

table(unlist(sapply(unb.prof.ele_p7, function(x) (x$producao_bibliografica$CAPITULO_DE_LIVRO$ano))))

##
## 2010 2011 2012 2013 2014 2015 2016 2017
## 10 2 5 3 3 4 2 1

table(unlist(sapply(unb.prof.ele_p7, function(x) (x$producao_bibliografica$LIVRO$ano))))

##
## 2010 2011 2013 2014 2015 2017
## 4 3 3 1 1 2

table(unlist(sapply(unb.prof.ele_p7, function(x) (x$producao_bibliografica$PERIODICO$ano))))

##
## 2010 2011 2012 2013 2014 2015 2016 2017
## 15 18 15 23 16 17 43 31
```

```

table(unlist(sapply(unb.prof.ele_p7, function(x) (x$producao_bibliografica$TEXTOS_JORNAIS$ano))))

##
## 2010 2011 2012 2013 2014 2015
##    1    2    7    1    1    1
# número de pessoas que realizaram diferentes tipos de orientações
length(unlist(sapply(unb.prof.ele_p7, function(x) names(x$orientacoes_academicas))))

## [1] 66
# número de pessoas por tipo de orientação
table(unlist(sapply(unb.prof.ele_p7, function(x) names(x$orientacoes_academicas))))

##
##          ORIENTACAO_CONCLUIDA_DOUTORADO
##                                8
##          ORIENTACAO_CONCLUIDA_MESTRADO
##                                13
##          ORIENTACAO_EM_ANDAMENTO_DOUTORADO
##                                10
##          ORIENTACAO_EM_ANDAMENTO_GRADUACAO
##                                5
## ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA
##                                7
##          ORIENTACAO_EM_ANDAMENTO_MESTRADO
##                                11
##          OUTRAS_ORIENTACOES_CONCLUIDAS
##                                12
#número de orientações concluídas
sum(sapply(unb.prof.ele_p7, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTRADO$ano)))

## [1] 140
sum(sapply(unb.prof.ele_p7, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTORADO$ano)))

## [1] 22
sum(sapply(unb.prof.ele_p7, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_POS_DOUTORADO$ano)))

## [1] 0
# número de pessoas por quantitativo de orientações por pessoa 0 = 1; 1 = 2...
table(unlist(sapply(unb.prof.ele_p7, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTRADO$ano))))

##
## 0  1  4  5  6  8 10 12 14 15 20 28
## 1  1  1  2  1  1  1  2  1  1  1  1
table(unlist(sapply(unb.prof.ele_p7, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTORADO$ano))))

##
## 0 1 2 3 5 6
## 6 2 3 1 1 1
table(unlist(sapply(unb.prof.ele_p7, function(x) length(x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_POS_DOUTORADO$ano))))

##
## 0

```

```
## 14
```

```
# número de orientações por ano  
table(unlist(sapply(unb.prof.ele_p7, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_MESTRAD
```

```
##
```

```
## 2010 2011 2012 2013 2014 2015 2016 2017
```

```
##    6    16    25    19    18    16    26    14
```

```
table(unlist(sapply(unb.prof.ele_p7, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_DOUTORAD
```

```
##
```

```
## 2010 2012 2013 2014 2015 2016 2017
```

```
##    4     3     5     2     4     2     2
```

```
table(unlist(sapply(unb.prof.ele_p7, function(x) (x$orientacoes_academicas$ORIENTACAO_CONCLUIDA_POS_DOU
```

```
## < table of extent 0 >
```

```
# números de orientações não concluídas por ano
```

```
# Trabalho de mestrado em andamento
```

```
table(unlist(sapply(unb.prof.ele_p7, function(x) (x$orientacoes_academicas$ORIENTACAO_EM_ANDAMENTO_MESTR
```

```
##
```

```
## 2015 2016 2017
```

```
##    4     9    22
```

```
# DOUTORADO em andamento
```

```
table(unlist(sapply(unb.prof.ele_p7, function(x) (x$orientacoes_academicas$ORIENTACAO_EM_ANDAMENTO_DOUT
```

```
##
```

```
## 2013 2014 2015 2016 2017
```

```
##    2     5     7     7    15
```

```
# Trabalho de Graduação em andamento
```

```
table(unlist(sapply(unb.prof.ele_p7, function(x) (x$orientacoes_academicas$ORIENTACAO_EM_ANDAMENTO_GRAD
```

```
##
```

```
## 2016 2017
```

```
##    3    10
```

Por fim, na área de engenharia elétrica - “p7”, vê-se, novamente, um domínio da grande área de engenharias, como era de se esperar. Em 2016, é possível observar que se teve o maior número de publicações em periódicos, 42 no total.

Sobre as orientações, dos temas desta pesquisa, foi a que teve mais equilíbrio entre o número de profissionais de mestrados e os que trabalham com doutorados concluídos, tendo uma diferença de 5. Especificamente, sobre as orientações vê-se um enorme número de mestrados concluídos desde 2011, com valor baixo em 2010 (6 no total, número similar a de doutorados feitos: 4, naquele mesmo ano). Um dado que pode ser alarmante é que ainda existe têm dois doutorados iniciados em 2013 que não estão concluídos ainda, estando em seu último ano, pois o tempo máximo do mestrado é de 5 anos, já do mestrado há 4 trabalhos que estão no último ano possível.

## Arquivo publicação

### TODO LIST : VERIFICAR

- automação

```
#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$PERIODICO[[i]])
#glimpse(unb.pub.biomed.df)
# Limpando o data-frame de listas
unb.pub.biomed.df$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)

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

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

##		
##	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
```





```
## publicações 2010 até 2017
table(unb.pub.ele_p8.df$ano)
```

```
##
## 2010 2011 2012 2013 2014 2015 2016 2017
##    8   15   17   17   18   21   36   36
## publicações por periódico, mostrando os top-5 na área de Elétrica p8
d <- table(unb.pub.ele_p8.df$periodico)
head(sort(d,decreasing = TRUE),n=5)
```

```
##
##          IEEE Transactions on Information Forensics and Security
##                                     14
##          Digital Signal Processing (Print)
##                                     7
##          Journal of Communication and Information Systems (Online)
##                                     6
## Journal of Visual Communication and Image Representation (Print)
##                                     6
##          Pattern Recognition Letters
##                                     5
```

```
## 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.ele_p8.df$autores,";")))),decreasing = TRUE),n=10)
```

```
##
##          ROCHA, ANDERSON          ROCHA, ANDERSON
##                52                21
##          GOLDENSTEIN, SIOME      DA COSTA, J. P. C. L.
##                17                15
##          SOUSA JR, R. T. DE REZENDE ROCHA, ANDERSON
##                14                13
##          ROCHA, A                ROCHA, A.
##                13                13
## ROCHA, ANDERSON DE REZENDE      WAINER, JACQUES
##                13                9
```

```
## Mostrando os autores-endogeno que tiveram mais publicação
head(sort(table(unlist(unb.pub.ele_p8.df$`autores-endogeno`)),decreasing = TRUE),n=10)
```

```
##
##          1521815731111261          3196088341529197
##                63                35
##          1786889674911887          4438697154081039
##                19                14
##          3636638127781726          9788199690491510
##                9                9
## 1786889674911887; 3196088341529197          1808368824068028
##                8                3
##          5810353896294133          3408305339297459
##                3                2
```

Processando os dados, pode-se concluir que todas as publicações foram completas, a maioria foi publicada na revista “IEEE Transactions on Information Forensics and Security”, 14 no total. O autor que teve mais publicação foi Anderson de Rezendo Rocha.

```
#Criando um data-frame com todos os anos
unb.pub.ele_p7.df <- data.frame()
for (i in 1:length(unb.pub.ele_p7[[1]]))
  unb.pub.ele_p7.df <- rbind(unb.pub.ele_p7.df, unb.pub.ele_p7$PERIODICO[[i]])
#glimpse(unb.pub.ele_p7.df)
# Limpando o data-frame de listas
unb.pub.ele_p7.df$autores <- gsub("\\", "\\|\\", "\\\"", "; ", unb.pub.ele_p7.df$autores)
unb.pub.ele_p7.df$autores <- gsub("\\\"|c\\|(\\|\\|)", "", unb.pub.ele_p7.df$autores)
unb.pub.ele_p7.df$`autores-endogeno` <- gsub(",", ";", unb.pub.ele_p7.df$`autores-endogeno`)
unb.pub.ele_p7.df$`autores-endogeno` <- gsub("\\\"|c\\|(\\|\\|)", "", unb.pub.ele_p7.df$`autores-endogeno`)
#glimpse(unb.pub.ele_p7.df)

### ARQUIVO PROCESSADO E LIMPO
## publicações por natureza - todas foram completas
table(unb.pub.ele_p7.df$natureza)
```

```
## publicações 2010 até 2017
table(unb.pub.ele_p7.df$ano)
```

```
## publicações por periódico, mostrando os top-5 na área de Elétrica p8
d <- table(unb.pub.ele_p7.df$periodico)
head(sort(d,decreasing = TRUE),n=5)
```

```
## Journal of Communication and Information Systems (Online)
##                                     10
##                               Revista IEEE América Latina
##                                     8
##                               Digital Signal Processing (Print)
##                                     7
##                               IEEE Transactions on Power Delivery
##                                     6
##                               Electric Power Systems Research (Print)
##                                     5
```

```
## Mostrando os autores que tiveram mais publicação
### Neste caso é importante mostra que como a não uniformidade de como se escreve acaba po
head(sort(table(toupper(unlist(strsplit(unb.pub.ele_p7.df$autores, ";")))), decreasing = TRUE))
```

##		
##	DA COSTA, J. P. C. L.	SOUSA JR, R. T.
##	14	14
##	MOLINARO, L. F. R.	ATTUX, ROMIS
##	13	6
##	DE SOUSA JÚNIOR, RAFAEL TIMÓTEO	DA COSTA, JOÃO PAULO C.L.
##	6	5

##	DIAS, U. S.	FRAZAO, A.
##	5	5
##	GALPERIN, H.	SILVA, K. M.
##	5	5

Há um total de 161 publicações em engenharia elétrica p7, espaçado desde 2010 até 2017. Duas revistas 4 revistas se destacam, quando analisamos as revistas que foram enviadas:

- Journal of Communication and Information Systems (Online)
- Revista IEEE América Latina
- Digital Signal Processing (Print)
- IEEE Transactions on Power Delivery

O professor com mais publicação é o professor Da Costa (professor Lustosa), seguido por Molina. E mais uma vez é verificado a dificuldade que se encontra ao analisar nomes de autores que são os mesmo, mas que foram preenchido de foram diferente. ### Arquivo Orientação

## TODO LIST : VERIFICAR

- 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$ORIENTACAO_CONCLUIDA_POS_DOUTORADO[i])
unb.adv.biomed.df <- rbind(unb.adv.biomed.df, unb.adv.biomed.tipo.df); unb.adv.biomed.tipo.df <- data.frame()
for (i in 1:length(unb.adv.biomed[[1]]))
  unb.adv.biomed.tipo.df <- rbind(unb.adv.biomed.tipo.df, unb.adv.biomed$ORIENTACAO_CONCLUIDA_DOUTORADO[i])
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$ORIENTACAO_CONCLUIDA_MESTRADO[i])
unb.adv.biomed.df <- rbind(unb.adv.biomed.df, unb.adv.biomed.tipo.df)
#glimpse(unb.adv.biomed.df)
#Transformar as colunas de listas em caracteres eliminando c("")
unb.adv.biomed.df$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, ...].

#Os que tem 1 orientador pelo menos
length(unb.adv.biomed.df$ori1)

## [1] 124

# 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

#Numero de instituições
sort(table(unb.adv.biomed.df$instituicao),decreasing = TRUE)

##
##                                Universidade de Brasília
##                                                                82
##                                Universidade de Brasília - Faculdade UnB-Gama
##                                                                21
##                                Faculdade do Gama da UnB
##                                                                7
##                                Faculdade UNB Gama - FGA
##                                                                7
##                                Universidade de Brasília - Campus Darcy Ribeiro
##                                                                3
## Centro de Desenvolvimento e Planejamento Regional - Cedeplar/FACE/UFGM
##                                                                1
##                                Faculdade do Gama/ UnB
##                                                                1
##                                Universidade do Estado do Rio Grande do Norte
##                                                                1
##                                Universidade Nacional de Brasília
##                                                                1

#Cursos que estão envolvido na pós-graduação de engenharia Biomédica
cursos_d <- sort(table(unb.adv.biomed.df$curso), decreasing = TRUE)
cursos_d

##
##                                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
```

```
#Quantidas 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)

##
##
##
##                                80
##          Centro de Apoio ao Desenvolvimento Tecnológico
##                                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

```

Das 123 orientações, apenas uma delas contém um segundo orientador. A maioria das orientações ocorrem na UnB - Darcy Ribeiro depois é unidade do Gama, na UnB. Das 123, a maioria é dissertação de mestrado 109, Lourdes Matos Brasil é a professora com mais orientações. 46 alunos são bolsistas.

## Verificando engenharia automação

### Verificando engenharia elétrica p8

```

#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.ele_p8.tipo.df <- data.frame(); unb.adv.ele_p8.df <- data.frame()
for (i in 1:length(unb.adv.ele_p8[[1]]))
  unb.adv.ele_p8.tipo.df <- rbind(unb.adv.ele_p8.tipo.df, unb.adv.ele_p8$ORIENTACAO_CONCLUIDA_POS_DOUTORADO)
unb.adv.ele_p8.df <- rbind(unb.adv.ele_p8.df, unb.adv.ele_p8.tipo.df); unb.adv.ele_p8.tipo.df <- data.frame()
for (i in 1:length(unb.adv.ele_p8[[1]]))
  unb.adv.ele_p8.tipo.df <- rbind(unb.adv.ele_p8.tipo.df, unb.adv.ele_p8$ORIENTACAO_CONCLUIDA_DOUTORADO)
unb.adv.ele_p8.df <- rbind(unb.adv.ele_p8.df, unb.adv.ele_p8.tipo.df); unb.adv.ele_p8.tipo.df <- data.frame()
for (i in 1:length(unb.adv.ele_p8[[1]]))
  unb.adv.ele_p8.tipo.df <- rbind(unb.adv.ele_p8.tipo.df, unb.adv.ele_p8$ORIENTACAO_CONCLUIDA_MESTRADO)
unb.adv.ele_p8.df <- rbind(unb.adv.ele_p8.df, unb.adv.ele_p8.tipo.df)
#glimpse(unb.adv.ele_p8.df)
#Transformar as colunas de listas em caracteres eliminando c("")
unb.adv.ele_p8.df$nome_orientadores <- gsub("\|c\\(|\\)", "", unb.adv.ele_p8.df$nome_orientadores)
unb.adv.ele_p8.df$id_lattes_orientadores <- gsub("\|c\\(|\\)", "", unb.adv.ele_p8.df$id_lattes_orientadores)
#Separar as colunas com dois orientadores
unb.adv.ele_p8.df <- separate(unb.adv.ele_p8.df, nome_orientadores, into = c("ori1", "ori2"), sep = ",")

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 133 rows [1,
## 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].

```



```

#Os que tem 1 orientador pelo menos
length(unb.adv.ele_p8.df$ori1)

## [1] 135

# Contando quem tem apenas 1 orientador
sum(is.na(unb.adv.ele_p8.df$ori2))

## [1] 133

unb.adv.ele_p8.df <- separate(unb.adv.ele_p8.df, id_lattes_orientadores, into = c("idLattes1", "idLattes2"))

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 133 rows [1,
## 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].

# Contando quem tem apenas 1 orientador
sum(is.na(unb.adv.ele_p8.df$idLattes2))

## [1] 133

#Numero de orientacoes por ano
table(unb.adv.ele_p8.df$ano)

##
## 2010 2011 2012 2013 2014 2015 2016 2017
##    9   10   29   15   15   11   28   18

#Numero de instituições
sort(table(unb.adv.ele_p8.df$instituicao),decreasing = TRUE)

##
##
## Universidade de Brasília
## 90
## Instituto de Computação - Universidade Estadual de Campinas
## 29
## Universidade Estadual de Campinas
## 7
## Universidade Salvador
## 3
## City University of Hong Kong
## 1
## Faculté des Sciences et Techniques - Universtité de Limoges
## 1
## Technische Universität Ilmenau
## 1
## Universidade Federal da Bahia/Universidade Salvador/UEFS
## 1
## Universidade Federal de Goiás
## 1
## Universitat Politècnica de Catanlunya
## 1

#Cursos que estão envolvido na pós-graduação de engenharia eletrica
cursos_d <- sort(table(unb.adv.ele_p8.df$curso), decreasing = TRUE)
cursos_d

##
## Engenharia Elétrica

```

```

##                                     46
##                               ENGENHARIA ELÉTRICA
##                                     30
##                               Mestrado
##                                     14
##
##                                     12
##                               Doutorado
##                                     8
##       Engenharia Elétrica (53001010059F0) (F)
##                                     5
##                               Ciência da Computação
##                                     4
##                               Engenharia Biomédica
##                                     3
##       Mestrado em Engenharia Elétrica
##                                     3
##                               Sistemas e Computação
##                                     2
##                               Sistemas Mecatrônicos
##                                     2
##       Doutorado em Engenharia Elétrica
##                                     1
## Doutorado Multiinstitucional Ciência da Computação
##                                     1
##   Electrical Engineering and Information Technology
##                                     1
##                               Informática
##                                     1
##                               Mestrado Em Redes
##                                     1
##       Telecommunication Engineering & Management
##                                     1

```

```

cursos_d5<-head(cursos_d,5)
table(unb.adv.ele_p8.df$codigo_do_curso)

```

```

##
##       33070059 51500140 60018704 60021179 60025859 60030232 60057831
##       12       4       42       2       1       2       3       30
## 60059672 90000005 90000008 90000009 90000011 90000012 90000014 90000017
##       3       1       12       7       1       1       1       2
## 90000018 90000019 90000020 90000023 90000026
##       3       1       1       1       5

```

```

#Quantidas de natureza dos trabalhos orientado
table(unb.adv.ele_p8.df$natureza)

```

```

##
##       Dissertação de mestrado Supervisão de pós-doutorado
##                                     106                                     12
##       Tese de doutorado
##                                     17

```

```

#Tabela com nome de professor e numero de orientacoes
head(sort(table(rbind(unb.adv.ele_p8.df$ori1, unb.adv.ele_p8.df$ori2)), decreasing = TRUE), 20)

```

```
##
## Anderson de Rezende Rocha Rafael Timóteo de Sousa Jr
## 33 31
## Joao Paulo Carvalho Lustosa da Costa Anderson Clayton Alves Nascimento
## 22 13
## Díbio Leandro Borges Bruno Werneck Pinto Hoelz
## 8 6
## Diego de Freitas Aranha Georges Daniel Amvame Nze
## 6 6
## Daniel Guerreiro e Silva William Ferreira Giozza
## 5 5
## Rafael Timóteo de Sousa Jr
## 2
```

```
# Tabela com nome dos alunos que mais foram orientados
head(sort(table(toupper(unb.adv.ele_p8.df$nome_aluno)),decreasing = TRUE),n=10)
```

```
##
## ANTONIO MANUEL RUBIO SERRANO DANILO FERNANDES TENÓRIO
## 2 2
## FILIPE DE OLIVEIRA COSTA GILVAN MARTINS DURÃES
## 2 2
## GIOVANI CHIACHIA MARINA ATSUMI OIKAWA
## 2 2
## RICARDO KEHRLE MIRANDA VINÍCIUS COUTINHO GUIMARÃES COELHO
## 2 2
## ADRIANA BASTOS PINTO ADRIANA SILVA NEIVA
## 1 1
```

```
# Quantidade de alunos que não são bolsistas
sum(unb.adv.ele_p8.df$bolsa == 'NAO')
```

```
## [1] 83
```

```
# Quantidade de alunos que são bolsistas
sum(unb.adv.ele_p8.df$bolsa == 'SIM')
```

```
## [1] 52
```

```
# Reparar que a maioria não tem agência financiadora por trás e há dois alunos que tem bolsa mas que não
table(unb.adv.ele_p8.df$agencia_financiadora)
```

```
##
##
## 83
## Conselho Nacional de Desenvolvimento Científico e Tecnológico
## 5
## Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
## 25
## Fundação de Amparo a Pesquisa do Estado de Goiás
## 1
## Fundação de Amparo à Pesquisa do Estado de São Paulo
## 8
## Google Inc.
## 1
## Hong Kong City University
## 1
```

```
##                                Intel Labs
##                                2
##                                LG Electronics
##                                2
##                                Motorola Mobility
##                                1
##                                Samsung Eletrônica da Amazônia
##                                5
##                                União Européia
##                                1
```

Das 135 orientações apenas 2 tem mais de um orientador. 2012 é o ano com mais orientação. A principal instituição é, sem novidades, UnB, seguindo pelo instituto de Computação - Universidade Estadual de Campinas. A maioria é dissertação de mestrado 106. Anderson Rocha e Rafael Sousa são os maiores colaboradores. A maioria dos alunos não são bolsistas 83 de 135.

### Verificando engenharia elétrica p7

```
#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.ele_p7.tipo.df <- data.frame(); unb.adv.ele_p7.df <- data.frame()
for (i in 1:length(unb.adv.ele_p7[[1]]))
  unb.adv.ele_p7.tipo.df <- rbind(unb.adv.ele_p7.tipo.df, unb.adv.ele_p7$ORIENTACAO_CONCLUIDA_POS_DOUTORADO)
unb.adv.ele_p7.df <- rbind(unb.adv.ele_p7.df, unb.adv.ele_p7.tipo.df); unb.adv.ele_p7.tipo.df <- data.frame()
for (i in 1:length(unb.adv.ele_p7[[1]]))
  unb.adv.ele_p7.tipo.df <- rbind(unb.adv.ele_p7.tipo.df, unb.adv.ele_p7$ORIENTACAO_CONCLUIDA_DOUTORADO)
unb.adv.ele_p7.df <- rbind(unb.adv.ele_p7.df, unb.adv.ele_p7.tipo.df); unb.adv.ele_p7.tipo.df <- data.frame()
for (i in 1:length(unb.adv.ele_p7[[1]]))
  unb.adv.ele_p7.tipo.df <- rbind(unb.adv.ele_p7.tipo.df, unb.adv.ele_p7$ORIENTACAO_CONCLUIDA_MESTRADO)
unb.adv.ele_p7.df <- rbind(unb.adv.ele_p7.df, unb.adv.ele_p7.tipo.df)
#glimpse(unb.adv.ele_p7.df)
#Transformar as colunas de listas em caracteres eliminando c("")
unb.adv.ele_p7.df$nome_orientadores <- gsub("\\|c\\(|\\|)", "", unb.adv.ele_p7.df$nome_orientadores)
unb.adv.ele_p7.df$id_lattes_orientadores <- gsub("\\|c\\(|\\|)", "", unb.adv.ele_p7.df$id_lattes_orientadores)
#Separar as colunas com dois orientadores
unb.adv.ele_p7.df <- separate(unb.adv.ele_p7.df, nome_orientadores, into = c("ori1", "ori2"), sep = ",")

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 154 rows [1,
## 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].

#Os que tem 1 orientador pelo menos
length(unb.adv.ele_p7.df$ori1)

## [1] 158

# Contando quem tem apenas 1 orientador
sum(is.na(unb.adv.ele_p7.df$ori2))

## [1] 154

unb.adv.ele_p7.df <- separate(unb.adv.ele_p7.df, id_lattes_orientadores, into = c("idLattes1", "idLattes2"))

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 154 rows [1,
## 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```

# Contando quem tem apenas 1 orientador
sum(is.na(unb.adv.ele_p7.df$idLattes2))

## [1] 154

#Numero de orientacoes por ano
table(unb.adv.ele_p7.df$ano)

##
## 2010 2011 2012 2013 2014 2015 2016 2017
##   10   16   27   23   18   20   28   16

#Numero de instituicoes
sort(table(unb.adv.ele_p7.df$instituicao),decreasing = TRUE)

##
##               Universidade de Brasília
##                               146
##               Universidade de Brasilia
##                               7
##               City University of Hong Kong
##                               1
##               Technische Universität Ilmenau
##                               1
##               Universidade Estadual de Campinas
##                               1
## Universidade Federal de Campina Grande
##                               1
## Universitat Politècnica de Catanlunya
##                               1

#Cursos que estão envolvido na pós-graduação de engenharia eletrica
cursos_d <- sort(table(unb.adv.ele_p7.df$curso), decreasing = TRUE)
cursos_d

##
##                               Engenharia Elétrica
##                               77
##                               ENGENHARIA ELÉTRICA
##                               55
## Programa de Pós-Graduação em Engenharia Elétrica
##                               12
##                               Mestrado em Engenharia Elétrica
##                               7
##                               Doutorado em Engenharia Elétrica
##                               4
## Electrical Engineering and Information Technology
##                               1
##                               Mestrado em Engenharia Elétrica - PPGEE
##                               1
##                               Telecommunication Engineering & Management
##                               1

cursos_d5<-head(cursos_d,5)
table(unb.adv.ele_p7.df$codigo_do_curso)

##

```

```
## 22000097 33070210 51500140 60030232 60057831 90000001 90000003 90000006
##      1      1      66      7      55      4      3      1
## 90000011 90000012 90000017 90000018 90000019 90000021
##      1      1      2      3      12      1

#Quantidas de natureza dos trabalhos orientado
table(unb.adv.ele_p7.df$natureza)

##
## Dissertação de mestrado      Tese de doutorado
##      136      22

#Tabela com nome de professor e numero de orientacoes
head(sort(table(rbind(unb.adv.ele_p7.df$ori1, unb.adv.ele_p7.df$ori2)), decreasing = TRUE), 20)

##
##      Rafael Timóteo de Sousa Jr
##      31
##      Joao Paulo Carvalho Lustosa da Costa
##      22
##      Kleber Melo e Silva
##      16
##      Paulo Roberto de Lira Gondim
##      16
##      Andre Noll Barreto
##      12
##      Marco Aurelio Gonçalves de Oliveira
##      11
##      Anésio de Leles Ferreira Filho
##      10
##      Francisco Damasceno Freitas
##      10
##      Ugo Silva Dias
##      10
##      Luis Fernando Ramos Molinaro
##      9
##      Daniel Guerreiro e Silva
##      5
##      Felipe Vigolvinho Lopes
##      4
##      Anésio de Leles Ferreira Filho
##      2
##      Rafael Timóteo de Sousa Jr
##      2
## Leonardo Rodrigues Araujo Xavier de Menezes
##      2

# Tabela com nome dos alunos que mais foram orientados
head(sort(table(toupper(unb.adv.ele_p7.df$nome_aluno)),decreasing = TRUE),n=10)

##
##      ALFREDO MIGUEL PIANETA ESCUDERO      ANTONIO MANUEL RUBIO SERRANO
##      2      2
##      DANIEL VIEIRA      DANILO FERNANDES TENÓRIO
##      2      2
##      EDUARDO JORGE SILVA LEITE JUNIOR      MARIA LEONOR SILVA DE ALMEIDA
```

```
##                                2                                2
##          RICARDO KEHRLE MIRANDA VINÍCIUS COUTINHO GUIMARÃES COELHO
##                                2                                2
##          ABIEZER AMARILIA FERNANDES          ABRAHAM ELIAS ORTEGA PAREDES
##                                1                                1
```

```
# Quantidade de alunos que não são bolsistas
sum(unb.adv.ele_p7.df$bolsa == 'NAO')
```

```
## [1] 106
```

```
# Quantidade de alunos que são bolsistas
sum(unb.adv.ele_p7.df$bolsa == 'SIM')
```

```
## [1] 52
```

```
# Reparar que a maioria não tem agência financiadora por trás e há dois alunos que tem bolsa mas que não
table(unb.adv.ele_p7.df$agencia_financiadora)
```

```
##
##
##                                108
## Conselho Nacional de Desenvolvimento Científico e Tecnológico
##                                7
##  Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
##                                42
##                                Hong Kong City University
##                                1
```

Dos 158 trabalhos orientados apenas 4 possuem 2 orientadores, sendo 154 possuindo apenas um orientador. Apenas um curso não é especificamente engenharia elétrica - Telecommunication Engineering & Management .Das universidades que não são UnB tem as seguintes :

- City University of Hong Kong
- Technische Universität Ilmenau
- Universidade Estadual de Campinas
- Universidade Federal de Campina Grande
- Universitat Politècnica de Catalunya

Houve 22 doutorados e 136 mestrados. Os professores que mais orientaram são:

- Rafael Timóteo de Sousa Jr - 31
- Joao Paulo Carvalho Lustosa da Costa - 22
- Kleber Melo e Silva - 16

Apenas 52 alunos são bolsistas.

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

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

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

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

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

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

# Extracao dos perfis dos professores

extraia_1perfil <- function( professor ) {
  idLattes <- names(professor)
  nome <- professor[[idLattes]]$nome
  resumo_cv <- professor[[idLattes]]$resumo_cv
  endereco_profissional <- professor[[idLattes]]$endereco_profissional #list
  instituicao <- endereco_profissional$instituicao
  orgao <- endereco_profissional$orgao
  unidade <- endereco_profissional$unidade
  DDD <- endereco_profissional$DDD
  telefone <- endereco_profissional$telefone
  bairro <- endereco_profissional$bairro
  cep <- endereco_profissional$cep

```

```

cidade <- endereco_profissional$cidade
senioridade <- professor[[idLattes]]$senioridade
df_1perfil <- data.frame( idLattes , nome, resumo_cv ,instituicao ,
                        orgao, unidade , DDD, telefone, bairro,cep,cidade , senioridade,
                        stringsAsFactors = FALSE)

return(df_1perfil)
}

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

return(df_saida)
}

# Extracao da producao bibliografica dos professores

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

extrai_producoes <- function( jsonProfessores) {
  df_saida <- data.frame()
  for( i in 1:length(jsonProfessores)) {
    jsonProfessor <- jsonProfessores[i]
    df_producao <- extrai_1producao(jsonProfessor)
    if( nrow(df_saida) > 0 ) {
      df_saida <- concatenadf(df_saida , df_producao)
    } else {
      df_saida <- df_producao
    }
  }
  df_saida <- df_saida %>% filter( !is.na(tipo_producao))
  return(df_saida)
}

```

```

# Extracao das orientacoes dos professores

extra1_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)
}

extra1_orientacoes <- function(jsonProfessores) {
  df_saida <- data.frame()
  for( i in 1:length(jsonProfessores)) {
    jsonProfessor <- jsonProfessores[i]
    df_orientacao <- extra1_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

extra1_larea_de_atuacao <- function(professor){
  idLattes <- names(professor)
  df_larea <- professor[[idLattes]]$areas_de_atuacao
  df_larea$idLattes <- idLattes
  return(df_larea)
}

extra1_areas_atuacao <- function(jsonProfessores){
  df_saida <- data.frame()
  for( i in 1:length(jsonProfessores)) {
    jsonProfessor <- jsonProfessores[i]
    df_area_atuacao <- extra1_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)
}

```

```
##### Início #####
#### 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)
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)
#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 sy
##                      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)
#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
##
##      titulo
## 1
## 2
## 3 Influência da Eletroestimulação Neuromuscular de Baixa Freqüência 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.
#numero de docentes
unique(length(unb.prof.biomed.df.orientacoes$nome_aluno))

## [1] 673

```

## Verificando engenharia automação

### Verificando engenharia elétrica p8

```
##### Início #####
#### Começo da Preparação dos dados referente a profile

unb.prof.ele_p8.json <- read_file("dados-2018-2/engenharia-eletrica-p8/281.profile.json")
unb.prof.ele_p8.df.capes <- read_csv("dados-2018-2/PesqPosCapes.csv",
                                     sep = ";", header = TRUE, colClasses = "character")
unb.prof.ele_p8 <- fromJSON(unb.prof.ele_p8.json)
length(unb.prof.ele_p8)

## [1] 11

# extrai perfis dos professores
unb.prof.ele_p8.df.professores <- extrai_perfis(unb.prof.ele_p8)

# extrai producao bibliografica de todos os professores
unb.prof.ele_p8.df.publicacoes <- extrai_producoes(unb.prof.ele_p8)

#extrai orientacoes
unb.prof.ele_p8.df.orientacoes <- extrai_orientacoes(unb.prof.ele_p8)

#extrai areas de atuacao
unb.prof.ele_p8.df.areas.de.atuacao <- extrai_areas_atuacao(unb.prof.ele_p8)

#salva os dataframes
save(unb.prof.ele_p8.df.professores, unb.prof.ele_p8.df.publicacoes,
     unb.prof.ele_p8.df.orientacoes, unb.prof.ele_p8.df.areas.de.atuacao, file = "dataframes.Rda")

#cria arquivo para análise
unb.prof.ele_p8.df <- data.frame()
unb.prof.ele_p8.df <- unb.prof.ele_p8.df.professores %>%
  select(idLattes, nome, resumo_cv, senioridade) %>%
  left_join(
    unb.prof.ele_p8.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.ele_p8.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.ele_p8.df.areas.de.atuacao %>%
      select(area, idLattes) %>%
      group_by(idLattes) %>%

```

```

    summarise(n_distinct(area)),
  by = "idLattes") %>%
left_join(
  unb.prof.ele_p8.df.capes %>%
    select(AreaPos, idLattes) %>%
    group_by(idLattes) %>%
    summarise(n_distinct(AreaPos)),
  by = "idLattes")

#glimpse(unb.prof.ele_p8.df)
head(unb.prof.ele_p8.df,3)

```

```

##           idLattes                                nome
## 1 1521815731111261          Anderson de Rezende Rocha
## 2 1786889674911887 Joao Paulo Carvalho Lustosa da Costa
## 3 1808368824068028          Díbio Leandro Borges
##
## 1 Anderson de Rezende Rocha é professor associado (Livre-Docente) do Instituto de Computação, Univer
## 2
## 3
##   senioridade ORIENTACAO_CONCLUIDA_DOUTORADO ORIENTACAO_CONCLUIDA_MESTRADO
## 1           9                             8                          15
## 2           8                             2                          20
## 3           8                             NA                          7
##   ORIENTACAO_CONCLUIDA_POS_DOUTORADO OUTRAS_ORIENTACOES_CONCLUIDAS
## 1                               10                          23
## 2                               NA                          29
## 3                               1                           9
##   CAPITULO_DE_LIVRO  DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA  EVENTO  LIVRO
## 1                   4                                     NA     66    NA
## 2                   4                                     NA     99     1
## 3                   NA                                     NA     12    NA
##   PERIODICO  TEXTO_EM_JORNAIS  n_distinct(area)  n_distinct(AreaPos)
## 1         63                 2                 1                 1
## 2         28                 8                 1                 2
## 3          3                NA                 2                 2

```

```

## Mostrando dados processados referente ao arquivo 281.publication.json
#glimpse(unb.pub.ele_p8.df)
#Podemos ver que contém natureza, titulo, periodico , autores ....
head(unb.pub.ele_p8.df,3)

```

```

##   natureza
## 1 COMPLETO
## 2 COMPLETO
## 3 COMPLETO
##
##                                     titulo
## 1      Automatic fruit and vegetable classification from images
## 2      Progressive randomization: Seeing the unseen
## 3 Efficient implementation of elliptic curve cryptography in wireless sensors
##                                     periodico  ano volume  issn
## 1      Computers and Electronics in Agriculture 2010    70 01681699
## 2 Computer Vision and Image Understanding (Print) 2010 114 10773142
## 3      Advances in Mathematics of Communications 2010    4 19305346

```



```

##      paginas                                doi
## 1   96 - 104 10.1016/j.compag.2009.09.002
## 2 349 - 362 10.1016/j.cviu.2009.10.002
## 3 169 - 187 10.3934/amc.2010.4.169
##
## 1 ROCHA, Anderson;ROCHA, ANDERSON;ROCHA, A.;DE REZENDE ROCHA, ANDERSON;ROCHA, ANDERSON DE REZENDE;RO
## 2                                ROCHA, Anderson;ROCHA, ANDERSON;ROCHA, A.;DE REZENDE ROCHA, A
## 3                                ARANHA
##      autores-endogeno
## 1 1521815731111261
## 2 1521815731111261
## 3 9788199690491510

## Mostrando dados processados referente ao arquivo 281.advise.json
#glimpse(unb.adv.ele_p8.df)
#Podemos ver que contém natureza, titulo, autores, ano, nome aluno, orientadores ....
head(unb.adv.ele_p8.df,3)

##      natureza
## 1 Supervisão de pós-doutorado
## 2 Supervisão de pós-doutorado
## 3 Supervisão de pós-doutorado
##
## 1 Reconhecimento e Classificação de Padrões por Meio de Técnicas de Engenharia de Características, F
## 2
## 3
##      ano id_lattes_aluno                                nome_aluno
## 1 2012                                Otávio Augusto Bizetto Penatti
## 2 2013                                Marina Atsumi Oikawa
## 3 2014                                Leandro Luís Galdino de Oliveira
##
##      instituicao curso
## 1 Instituto de Computação - Universidade Estadual de Campinas
## 2 Instituto de Computação - Universidade Estadual de Campinas
## 3                                Universidade Federal de Goiás
##      codigo_do_curso bolsa                                agencia_financiadora
## 1                                SIM                                Samsung Eletrônica da Amazônia
## 2                                SIM                                União Européia
## 3                                SIM Fundação de Amparo a Pesquisa do Estado de Goiás
##      codigo_agencia_financiadora                                ori1 ori2
## 1                                002400000993 Anderson de Rezende Rocha <NA>
## 2                                002600000997 Anderson de Rezende Rocha <NA>
## 3                                IXOT00000002 Dúbio Leandro Borges <NA>
##      idLattes1 idLattes2
## 1 1521815731111261 <NA>
## 2 1521815731111261 <NA>
## 3 1808368824068028 <NA>

##### Os arquivos
##### 279.graph.json - usado para mostrar correlações
##### 279.list.json - Como explicado este arquivo não contém nenhuma informação interessante.
#numero de docentes
unique(length(unb.prof.ele_p8.df.orientacoes$nome_aluno))

## [1] 340

```

## Verificando engenharia elétrica p7

```
##### Início #####
#### Começo da Preparação dos dados referente a profile

unb.prof.ele_p7.json <- read_file("dados-2018-2/engenharia-eletrica-p7/282.profile.json")
unb.prof.ele_p7.df.capes <- read_csv("dados-2018-2/PesqPosCapes.csv",
                                     sep = ";", header = TRUE, colClasses = "character")
unb.prof.ele_p7 <- fromJSON(unb.prof.ele_p7.json)
length(unb.prof.ele_p7)

## [1] 14

# extrai perfis dos professores
unb.prof.ele_p7.df.professores <- extrai_perfis(unb.prof.ele_p7)

# extrai producao bibliografica de todos os professores
unb.prof.ele_p7.df.publicacoes <- extrai_producoes(unb.prof.ele_p7)

#extrai orientacoes
unb.prof.ele_p7.df.orientacoes <- extrai_orientacoes(unb.prof.ele_p7)

#extrai areas de atuacao
unb.prof.ele_p7.df.areas.de.atuacao <- extrai_areas_atuacao(unb.prof.ele_p7)

#salva os daframes
save(unb.prof.ele_p7.df.professores, unb.prof.ele_p7.df.publicacoes,
     unb.prof.ele_p7.df.orientacoes, unb.prof.ele_p7.df.areas.de.atuacao, file = "dataframes.Rda")

#cria arquivo para análise
unb.prof.ele_p7.df <- data.frame()
unb.prof.ele_p7.df <- unb.prof.ele_p7.df.professores %>%
  select(idLattes, nome, resumo_cv, senioridade) %>%
  left_join(
    unb.prof.ele_p7.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.ele_p7.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.ele_p7.df.areas.de.atuacao %>%
      select(area, idLattes) %>%
      group_by(idLattes) %>%
      summarise(n_distinct(area)),
    by = "idLattes") %>%
```

```

left_join(
  unb.prof.ele_p7.df.capes %>%
    select(AreaPos, idLattes) %>%
    group_by(idLattes) %>%
    summarise(n_distinct(AreaPos)),
  by = "idLattes")

```

```
glimpse(unb.prof.ele_p7.df)
```

```

## Observations: 14
## Variables: 14
## $ idLattes          <chr> "0997629063334821", "1573568521...
## $ nome              <chr> "Luis Fernando Ramos Molinaro",...
## $ resumo_cv         <chr> "Possui graduação em Engenharia...
## $ senioridade       <chr> "9", "7", "8", "6", "9", "7", "...
## $ ORIENTACAO_CONCLUIDA_DOUTORADO <int> 3, NA, 2, NA, 5, NA, NA, 2, NA,...
## $ ORIENTACAO_CONCLUIDA_MESTRADO <int> 6, 12, 20, 4, 28, 5, NA, 8, 10,...
## $ OUTRAS_ORIENTACOES_CONCLUIDAS <int> NA, 38, 29, 11, 13, 5, NA, 14, ...
## $ CAPITULO_DE_LIVRO <int> 8, 1, 4, NA, 6, 1, NA, NA, NA, ...
## $ EVENTO            <int> 17, 36, 99, 59, 96, 23, 1, 33, ...
## $ LIVRO              <int> 2, NA, 1, NA, 9, NA, NA, NA, NA...
## $ PERIODICO          <int> 17, 9, 28, 7, 45, 9, 3, 8, 8, 1...
## $ TEXTO_EM_JORNAIS   <int> NA, 1, 8, NA, NA, NA, NA, NA, 3...
## $ `n_distinct(area)` <int> 1, 1, 1, 1, 3, 2, 3, 1, 2, 1, 1...
## $ `n_distinct(AreaPos)` <int> 1, 1, 2, 1, 1, 1, 2, 1, 1, 1, 1...

```

```
head(unb.prof.ele_p7.df,3)
```

```

##           idLattes                                nome
## 1 0997629063334821      Luis Fernando Ramos Molinaro
## 2 1573568521807396      Anésio de Leles Ferreira Filho
## 3 1786889674911887 Joao Paulo Carvalho Lustosa da Costa
##
## 1 Possui graduação em Engenharia Elétrica pela Universidade de Brasília (1979), mestrado em Engenharia
## 2
## 3
##   senioridade ORIENTACAO_CONCLUIDA_DOUTORADO ORIENTACAO_CONCLUIDA_MESTRADO
## 1           9                             3                             6
## 2           7                             NA                             12
## 3           8                             2                             20
##   OUTRAS_ORIENTACOES_CONCLUIDAS CAPITULO_DE_LIVRO EVENTO LIVRO PERIODICO
## 1                             NA                8       17       2       17
## 2                             38                1       36      NA       9
## 3                             29                4       99       1       28
##   TEXTO_EM_JORNAIS n_distinct(area) n_distinct(AreaPos)
## 1                NA                1                1
## 2                1                1                1
## 3                8                1                2

```

```
## Mostrando dados processados referente ao arquivo 281.publication.json
```

```
glimpse(unb.pub.ele_p7.df)
```

```

## Observations: 161
## Variables: 10
## $ natureza          <chr> "COMPLETO", "COMPLETO", "COMPLETO", "COMPLE...

```

```
## $ titulo <chr> "Computing Rightmost Eigenvalues for Small-...
## $ periodico <chr> "IEEE Transactions on Power Systems", "Elec...
## $ ano <chr> "2010", "2010", "2010", "2010", "2010", "20...
## $ volume <chr> "25", "80", "2010", "58", "2", "5", "38", "...
## $ issn <chr> "15580679", "03787796", "16875591", "001892...
## $ paginas <chr> "929 - 938", "84 - 90", "1 - 5", "988 - 993...
## $ doi <chr> "10.1109/tpwrs.2009.2036822", "10.1016/j.ep...
## $ autores <chr> "Rommès, Joost; MARTINS, NELSON; FREITAS, F...
## $ `autores-endogeno` <chr> "7997341055520362", "8795500242651581", "92..."
```

```
#Podemos ver que contém natureza, titulo, periodico , autores ....
head(unb.pub.ele_p7.df,3)
```

```
## natureza
## 1 COMPLETO
## 2 COMPLETO
## 3 COMPLETO
##
## 1 Computing Rightmost Eigenvalues for Small-Sig
## 2 Distance
## 3 Modeling and Characterization for Microstrip Filters in the Manufacturing Process through the Unsc
## periodico ano volume issn paginas
## 1 IEEE Transactions on Power Systems 2010 25 15580679 929 - 938
## 2 Electric Power Systems Research (Print) 2010 80 03787796 84 - 90
## 3 Modelling and Simulation in Engineering 2010 2010 16875591 1 - 5
## doi
## 1 10.1109/tpwrs.2009.2036822
## 2 10.1016/j.epsr.2009.08.012
## 3 10.1155/2010/691241
##
## autores
## 1 Rommès, Joost; MARTINS, NELSON; FREITAS, F. D.;Freitas, Francisco Damasceno
## 2 SILVA, Kleber Melo; NEVES, W. L. A.; SOUZA, B. A.
## 3 Ortega Paredes, Abraham E.; de Menezes, Leonardo R. A. X.; Abdalla, Humberto; Romani, Ivan N. A.
## autores-endogeno
## 1 7997341055520362
## 2 8795500242651581
## 3 9279262545449206
```

```
## Mostrando dados processados referente ao arquivo 281.advise.json
glimpse(unb.adv.ele_p7.df)
```

```
## Observations: 158
## Variables: 15
## $ natureza <chr> "Tese de doutorado", "Tese de dout...
## $ titulo <chr> "Aplicação da Transformada da Ince...
## $ ano <chr> "2010", "2010", "2010", "2010", "2...
## $ id_lattes_aluno <chr> "6817085573778215", "9354456664541...
## $ nome_aluno <chr> "Abraham Elias Ortega Paredes", "M...
## $ instituicao <chr> "Universidade de Brasília", "Unive...
## $ curso <chr> "Engenharia Elétrica", "Doutorado ...
## $ codigo_do_curso <chr> "60030232", "900000003", "51500140"...
## $ bolsa <chr> "SIM", "NAO", "NAO", "NAO", "NAO",...
## $ agencia_financiadora <chr> "Conselho Nacional de Desenvolve...
## $ codigo_agencia_financiadora <chr> "002200000000", "", "", "", ""...
## $ ori1 <chr> "Leonardo Rodrigues Araujo Xavier ...
```

```
## $ ori2 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ idLattes1 <chr> "9279262545449206", "0997629063334...
## $ idLattes2 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA...

#Podemos ver que contêm natureza, titulo, autores, ano, nome aluno, orientadores ....
head(unb.adv.ele_p7.df,3)
```

```
##          natureza
## 1 Tese de doutorado
## 2 Tese de doutorado
## 3 Tese de doutorado
##
## 1          Aplicação da Transformada da Incerteza para Projetos de
## 2      Análise do Perfil do Profissional de Tecnologia da Informação e Comunicação a partir de Percepções
## 3 Método para determinação do limite de penetração da geração distribuída fotovoltaica em sistemas rurais
##      ano id_lattes_aluno      nome_aluno
## 1 2010 6817085573778215 Abraham Elias Ortega Paredes
## 2 2010 9354456664541013   Marcelo Stehling de Castro
## 3 2010 6471578140876893      RAFAEL AMARAL SHAYANI
##          instituicao      curso
## 1 Universidade de Brasília      Engenharia Elétrica
## 2 Universidade de Brasília Doutorado em Engenharia Elétrica
## 3 Universidade de Brasília      Engenharia Elétrica
##      codigo_do_curso bolsa
## 1          60030232     SIM
## 2          90000003     NAO
## 3          51500140     NAO
##
##          agencia_financiadora
## 1 Conselho Nacional de Desenvolvimento Científico e Tecnológico
## 2
## 3
##      codigo_agencia_financiadora      ori1
## 1          002200000000 Leonardo Rodrigues Araujo Xavier de Menezes
## 2
## 3
##          Luis Fernando Ramos Molinaro
##          Marco Aurelio Gonçalves de Oliveira
##      ori2      idLattes1 idLattes2
## 1 <NA> 9279262545449206      <NA>
## 2 <NA> 0997629063334821      <NA>
## 3 <NA> 8709856465061922      <NA>
```

```
##### Os arquivos
##### 279.graph.json - usado para mostrar correlações
##### 279.list.json - Como explicado este arquivo não contem nenhuma informação interessante.
#numero de docentes
unique(length(unb.prof.ele_p7.df.orientacoes$nome_aluno))
```

```
## [1] 511
```

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

Neste presente trabalho, não se viu a necessidade de fazer merge entres os datas 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:

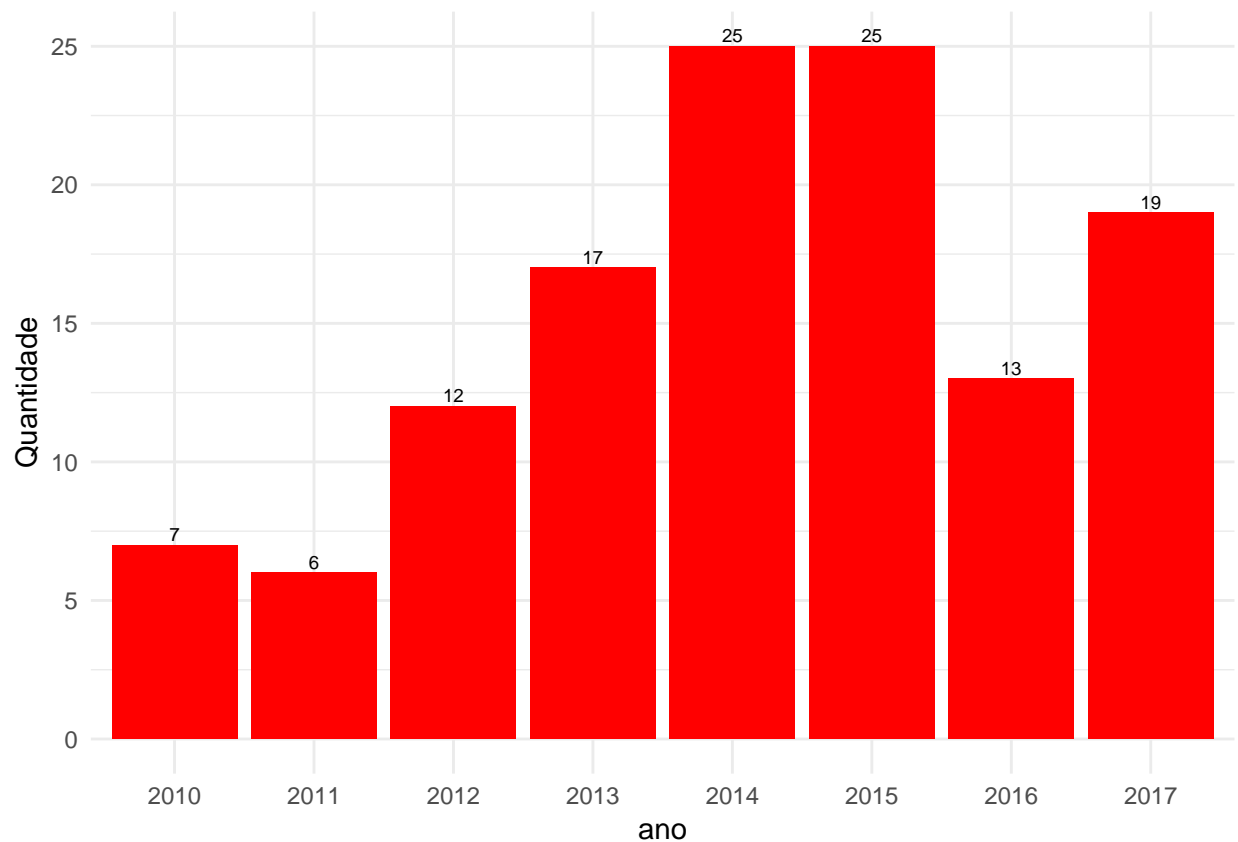
- AUTOMAÇÃO
- P7
- P8

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

### Expondo engenharia biomédica

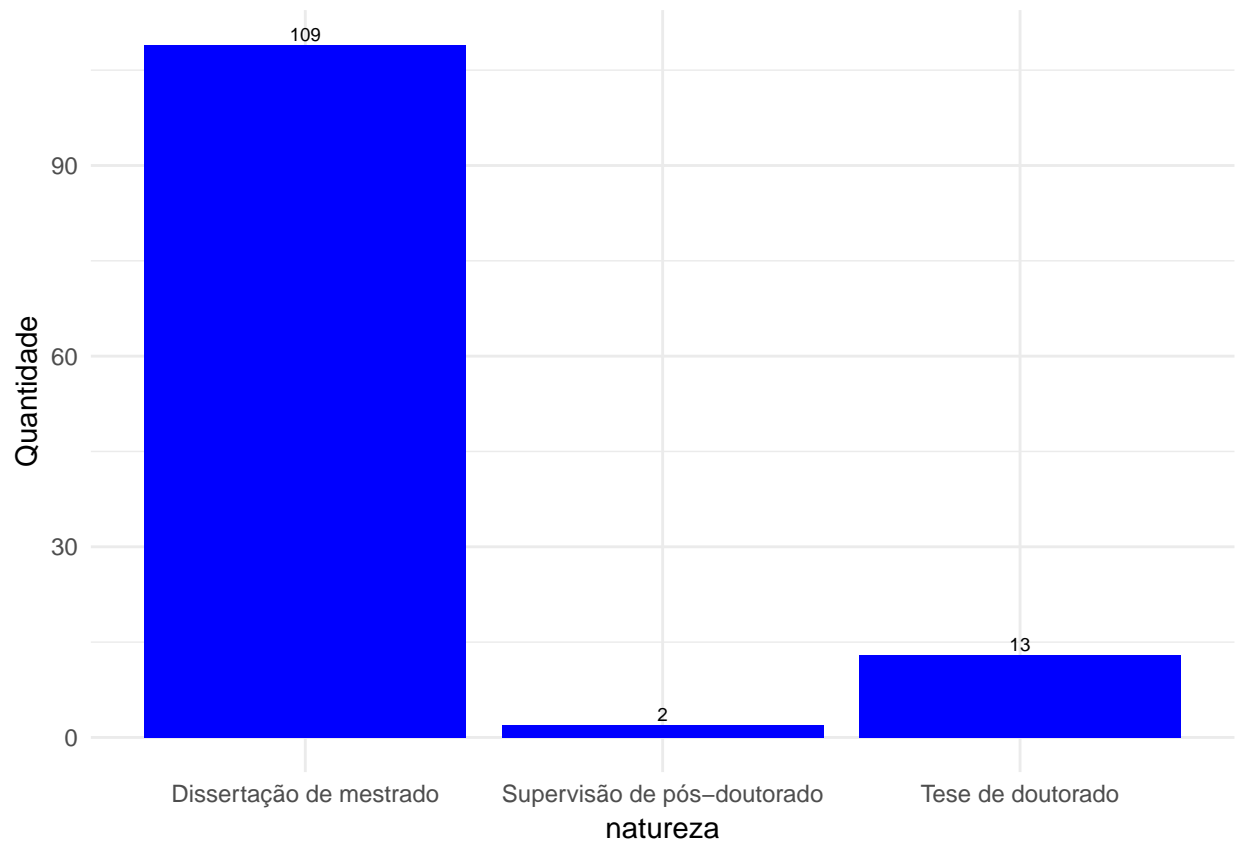
Aqui visa-se ver as quantidades de publicações feitas por anos na área de engenharia de Biomédica.

```
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()
```



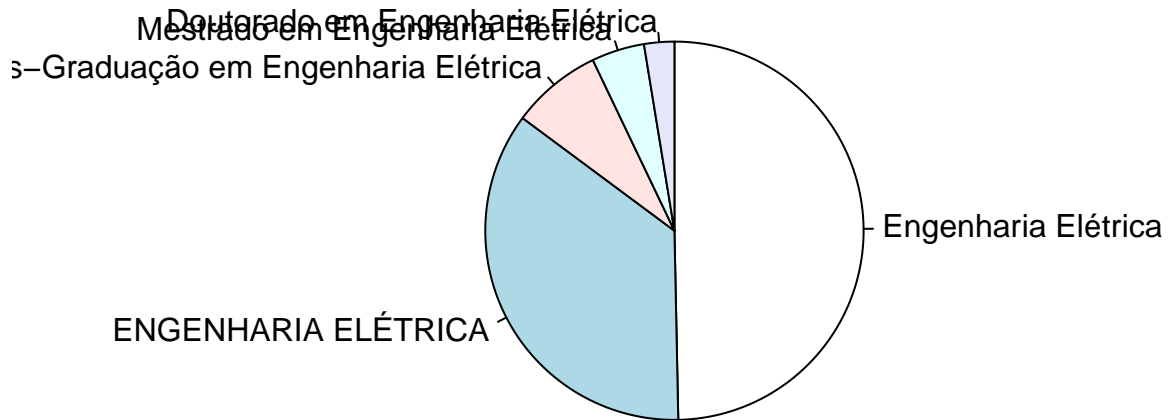
Verificando os cursos mais presentes.

```
cursos_d5

##
##          Engenharia Elétrica
##                      77
##          ENGENHARIA ELÉTRICA
##                      55
## Programa de Pós-Graduação em Engenharia Elétrica
##                      12
##          Mestrado em Engenharia Elétrica
##                      7
##          Doutorado em Engenharia Elétrica
##                      4
```

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

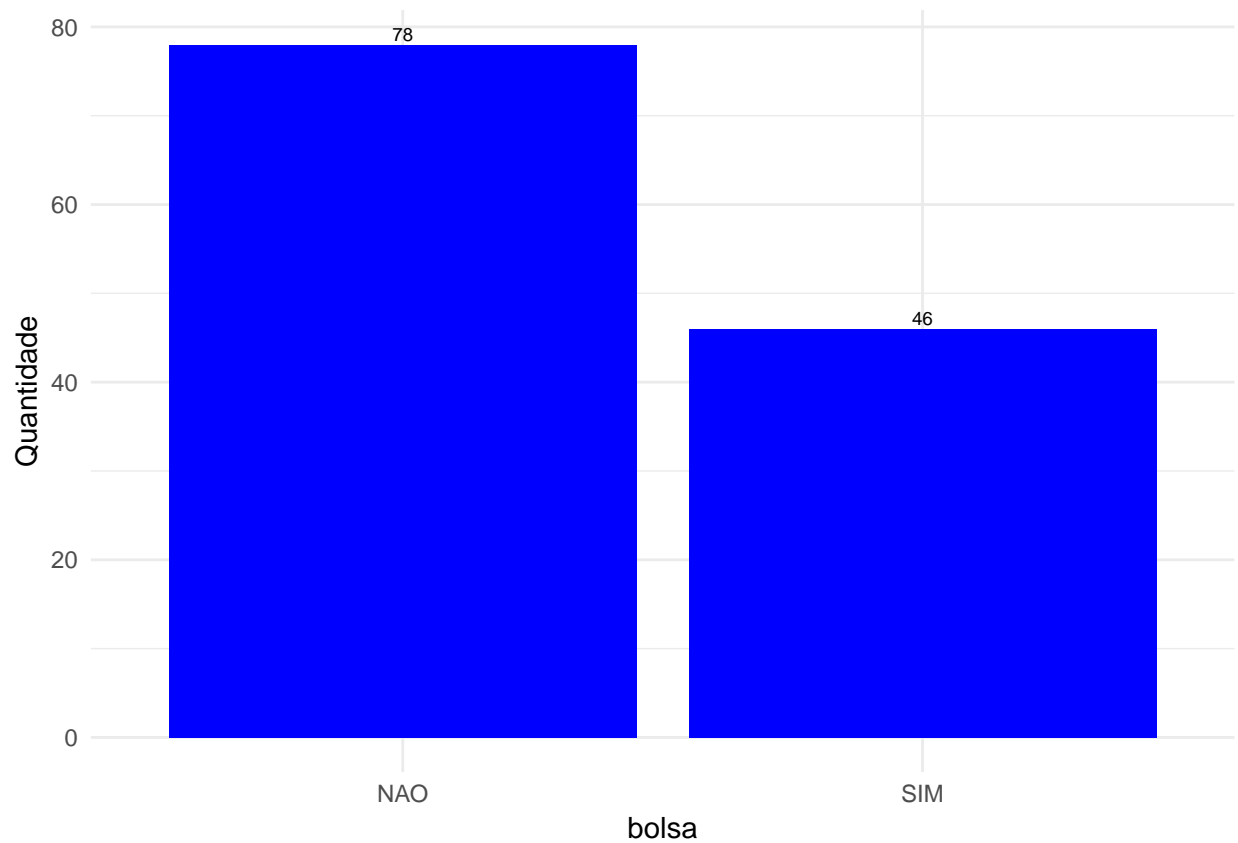
## Cursos mais presentes



Quantos 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()
```



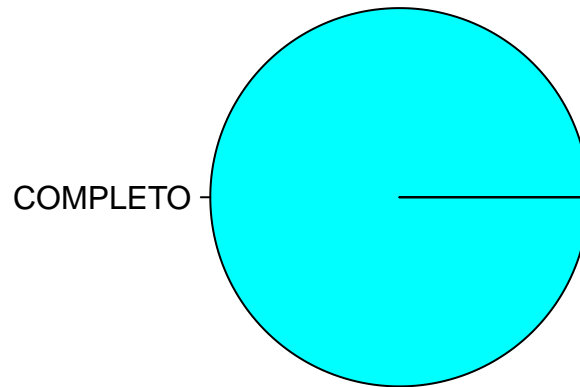


Assim, estes gráficos permitem verificar alguns aspectos importantes sobre as orientações.

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

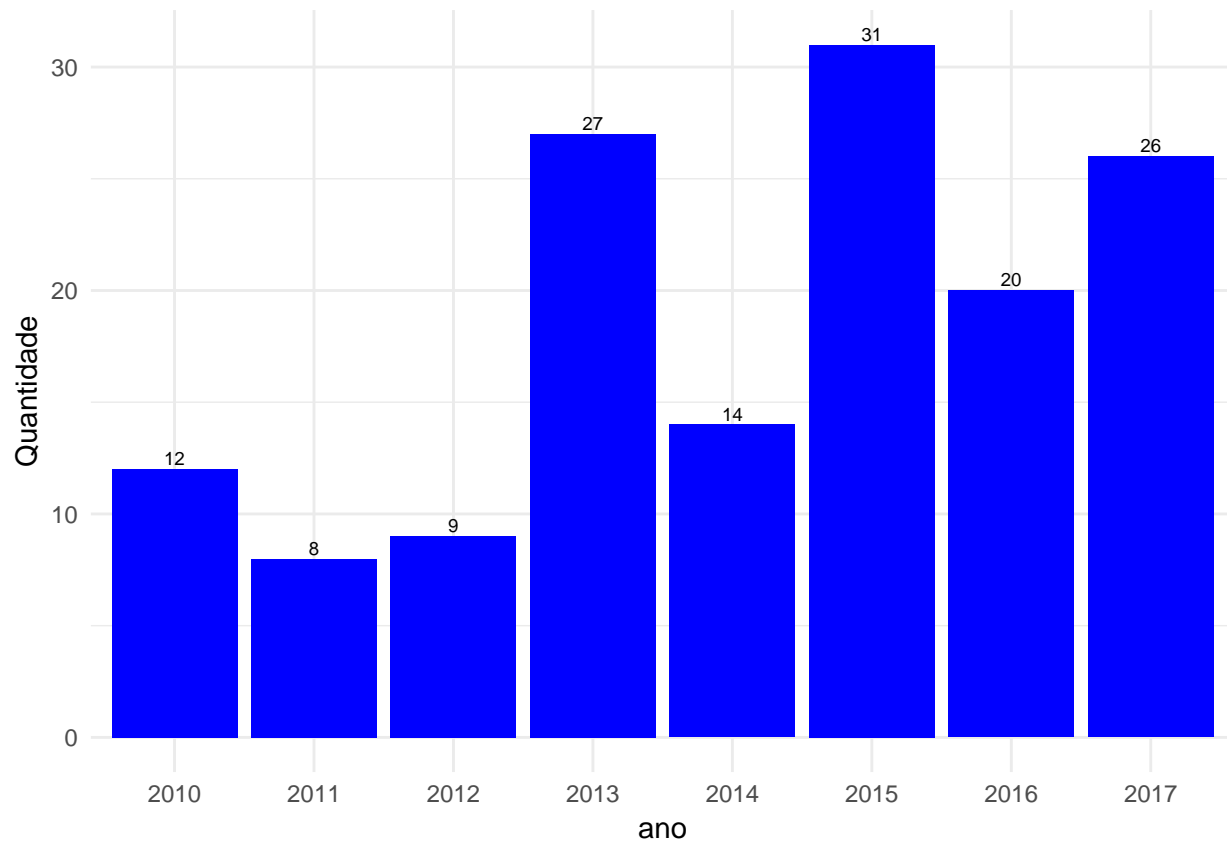
```
pie(table(unb.pub.biomed.df$natureza),main="Natureza das publicações",col = "cyan")
```

## 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)
```

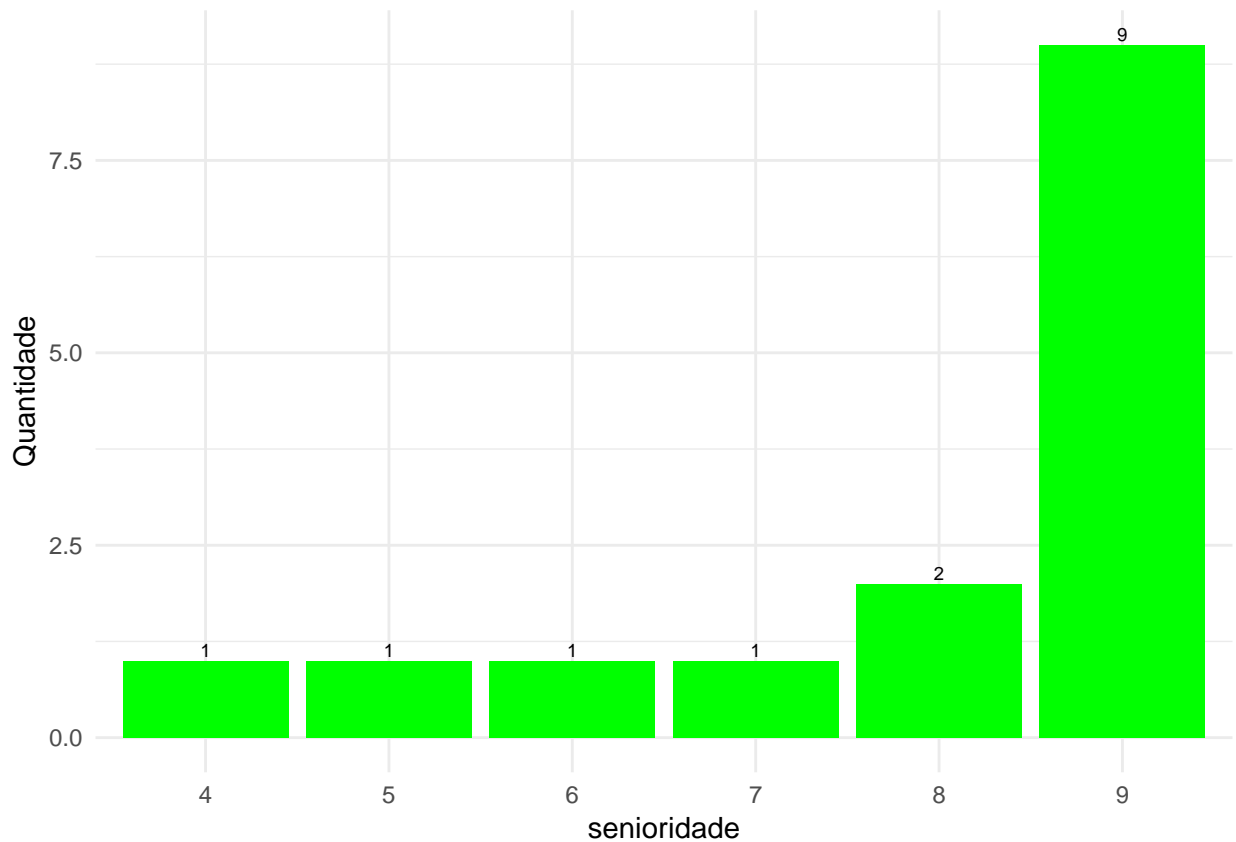
```
##
##          AMORIM, R. G. G.
##                                     17
##          AMORIM, R. G. G.
##                                     16
##          MARÃES, V. R. F. S.
##                                     14
##          BRASIL, L. M.
##                                     7
```

```
##           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.

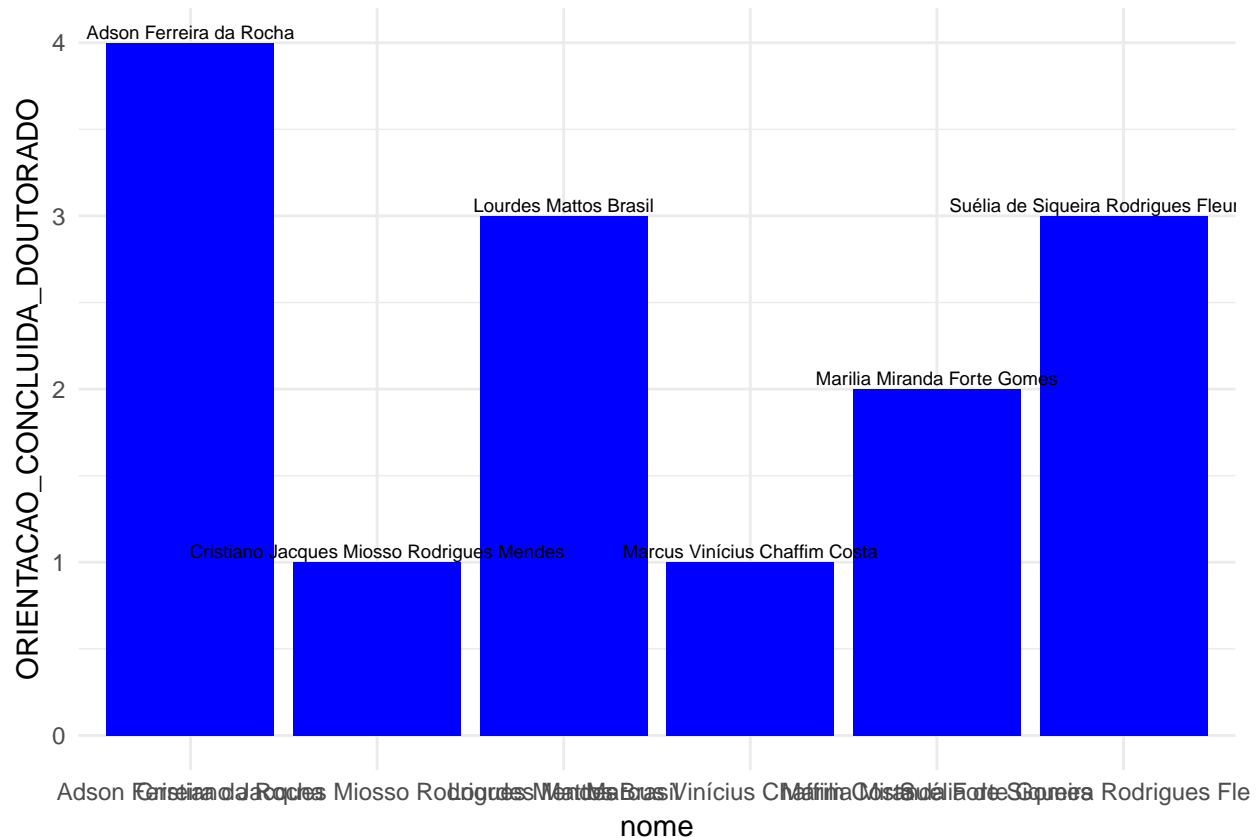
Analisando o arquivo 279.profile.json.

```
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()
```



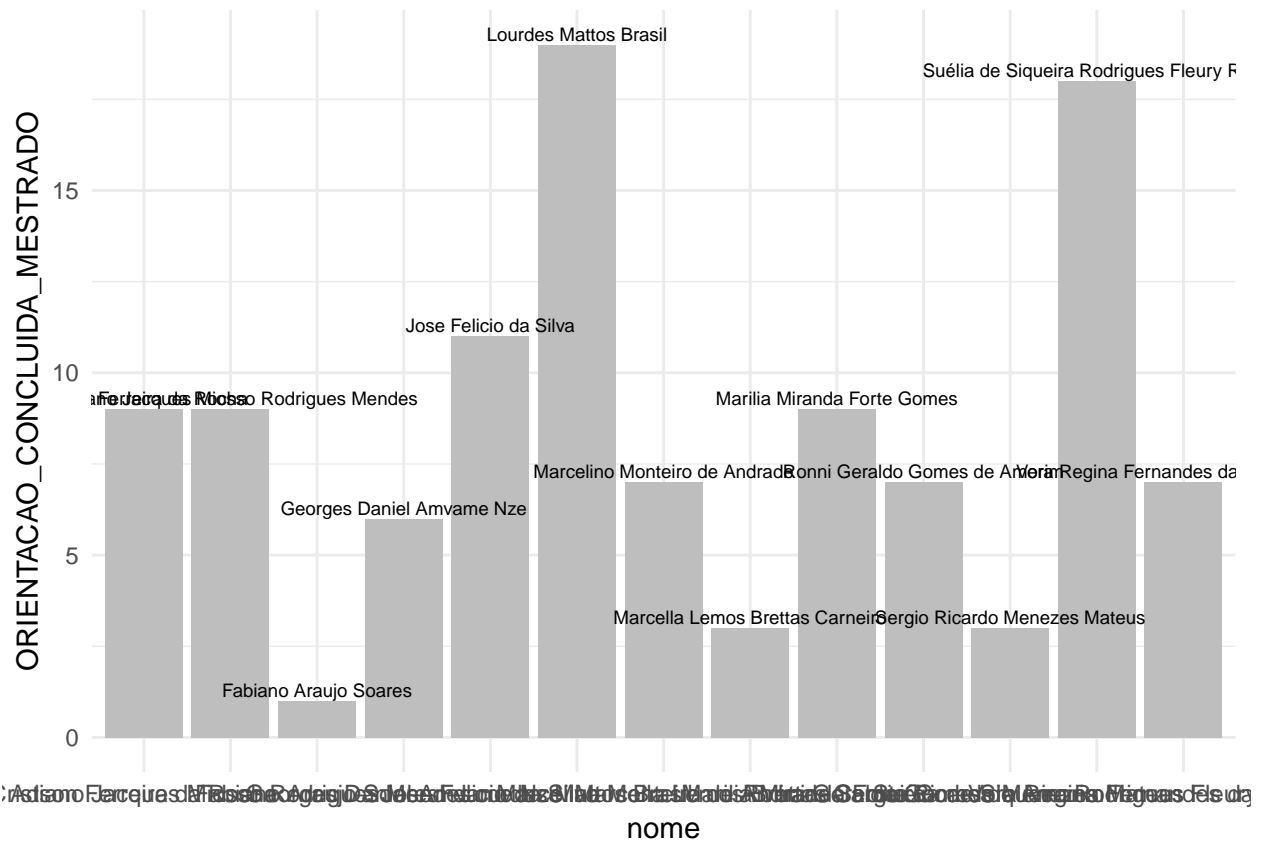
Analisando quantidade de orientações concluídas no doutorado por professor.

```
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()
```



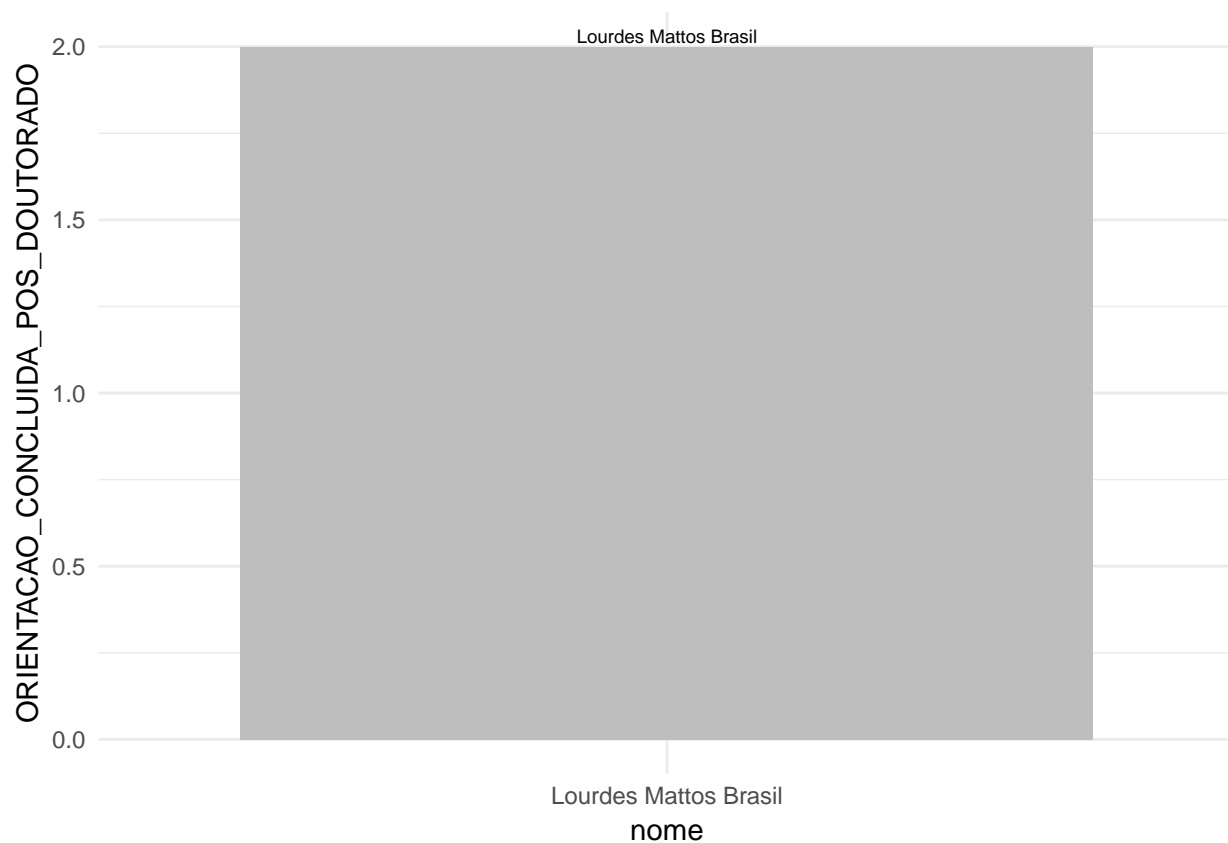
Analisando quantidade de orientações concluídas no mestrado por professor.

```
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()
```



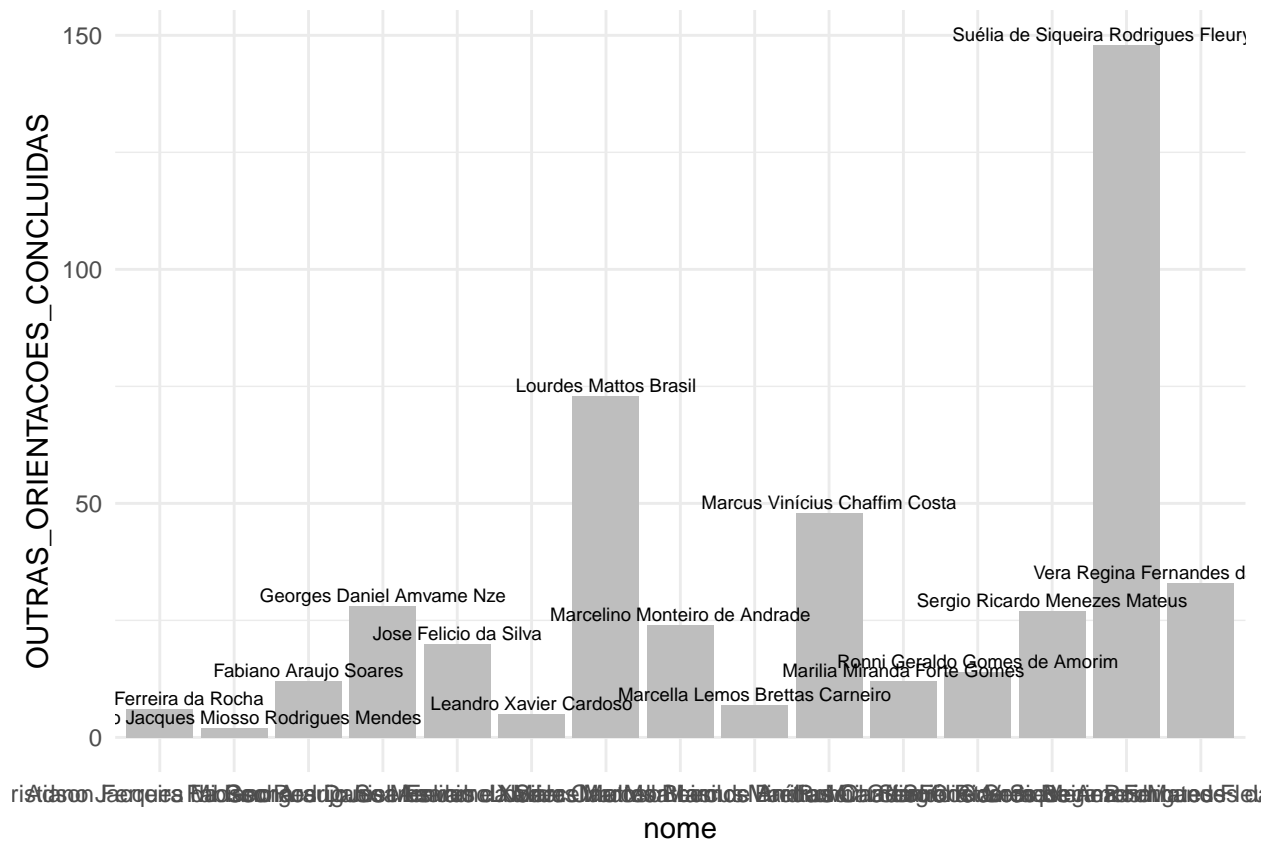
Analisando quantidade de orientações concluídas no pós-doutorado por professor.

```
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.

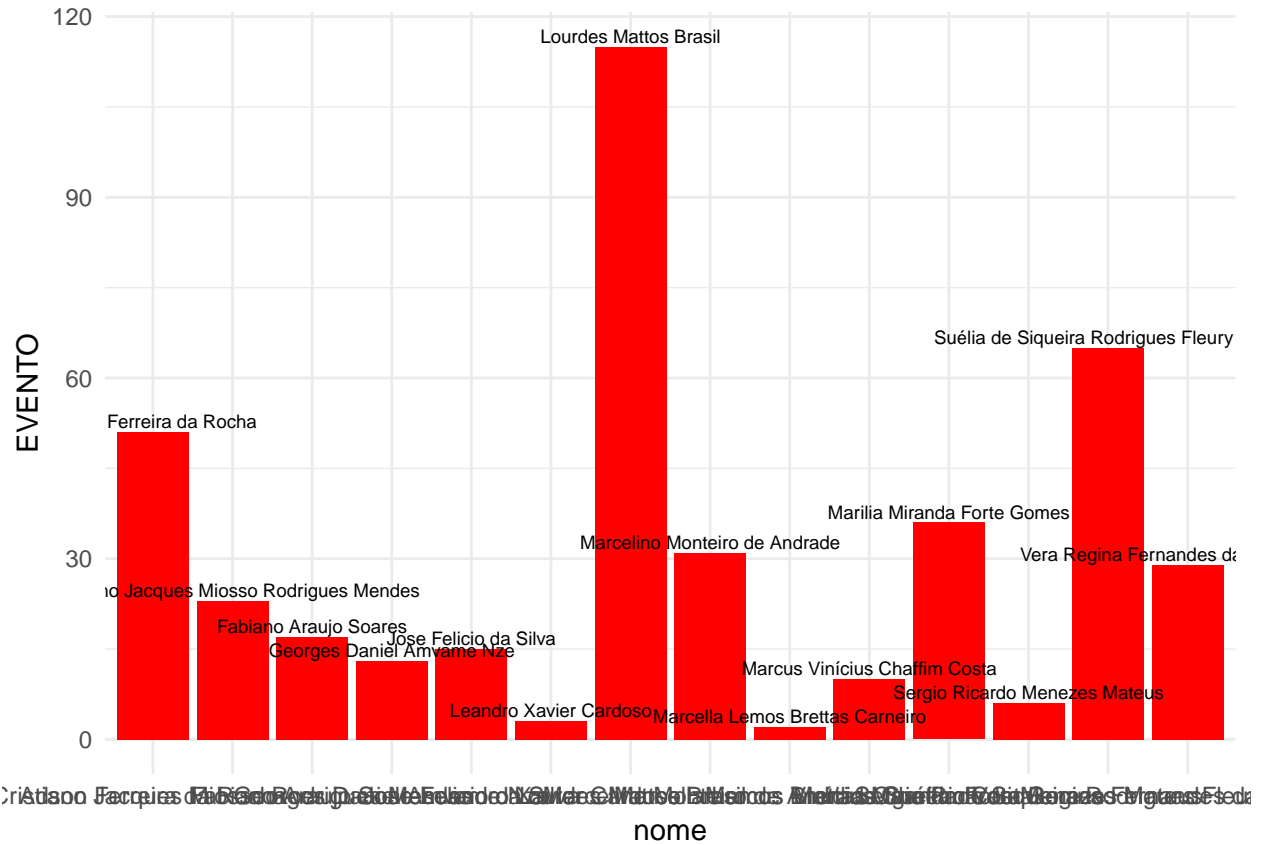
```
unb.prof.biomed.df %>%
  group_by(nome) %>%
  filter(!is.na(OUTRAS_ORIENTACOES_CONCLUIDAS))%>%
  ggplot(aes(x = nome, y = OUTRAS_ORIENTACOES_CONCLUIDAS)) +
  geom_bar(position = "stack", stat = "identity", fill = "gray")+
  geom_text(aes(label=nome), vjust=-0.3, size=2.5)+
  theme_minimal()
```



Analisando quantidade de eventos ido por professor.

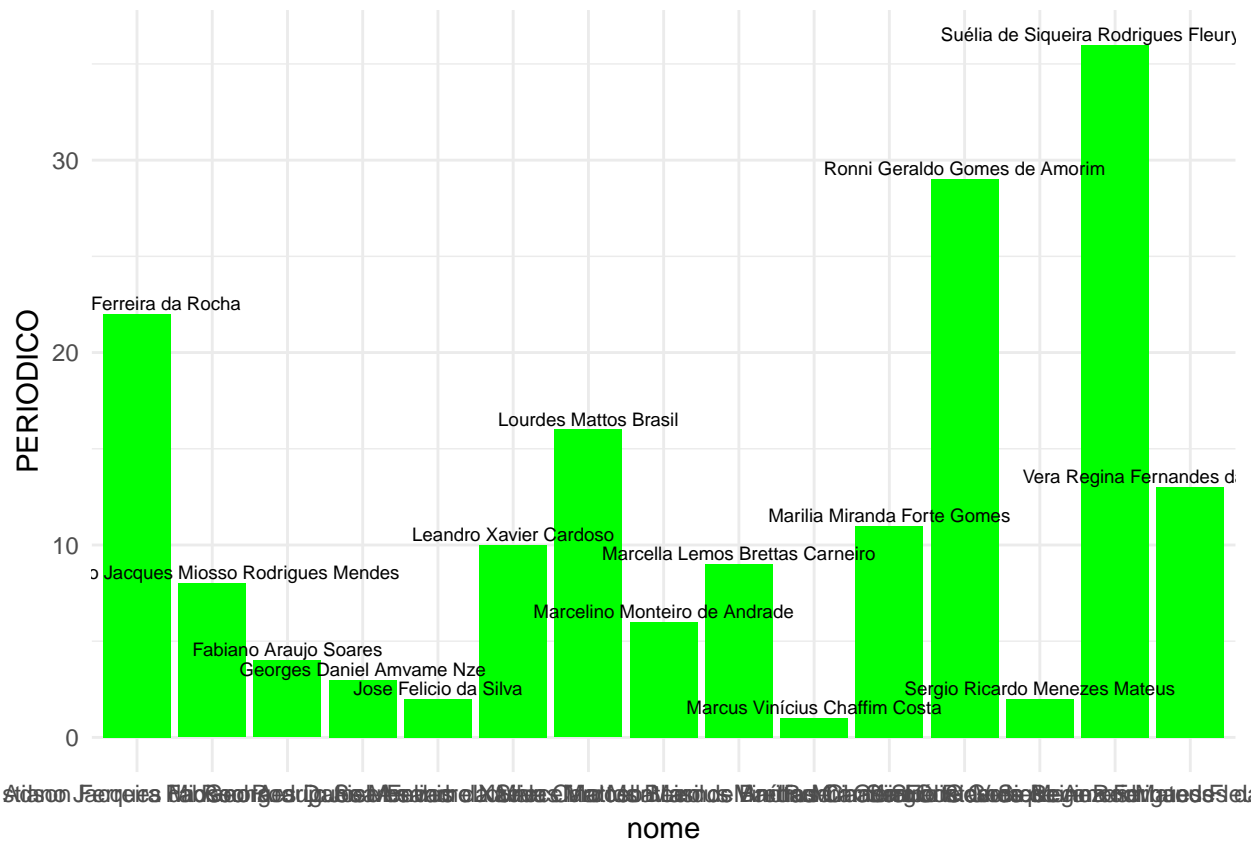
```
unb.prof.biomed.df %>%
  group_by(nome) %>%
  filter(!is.na(EVENTO))%>%
  ggplot(aes(x = nome, y = EVENTO)) +
  geom_bar(position = "stack", stat = "identity", fill = "red")+
  geom_text(aes(label=nome), vjust=-0.3, size=2.5)+
  theme_minimal()
```





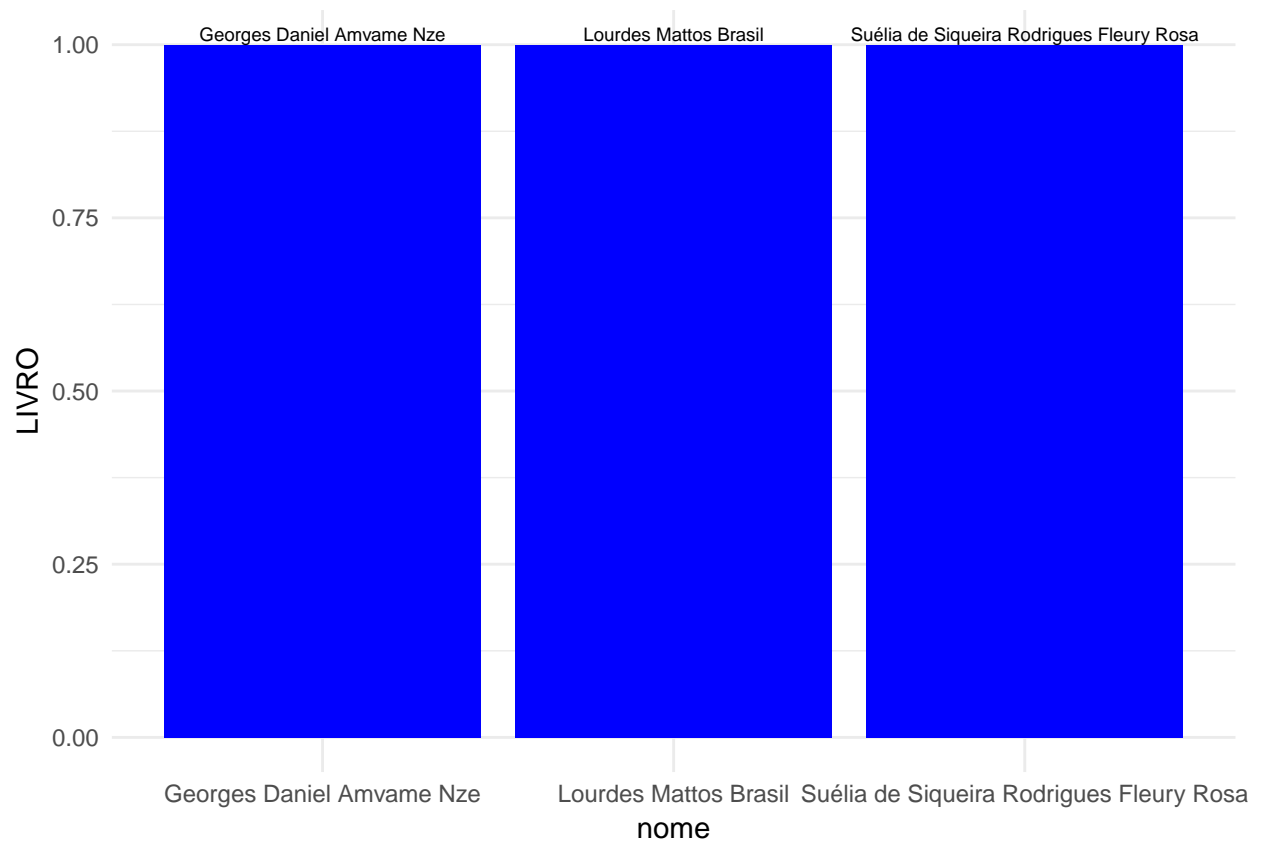
Analisando quantidade de periódicos por professor.

```
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()
```



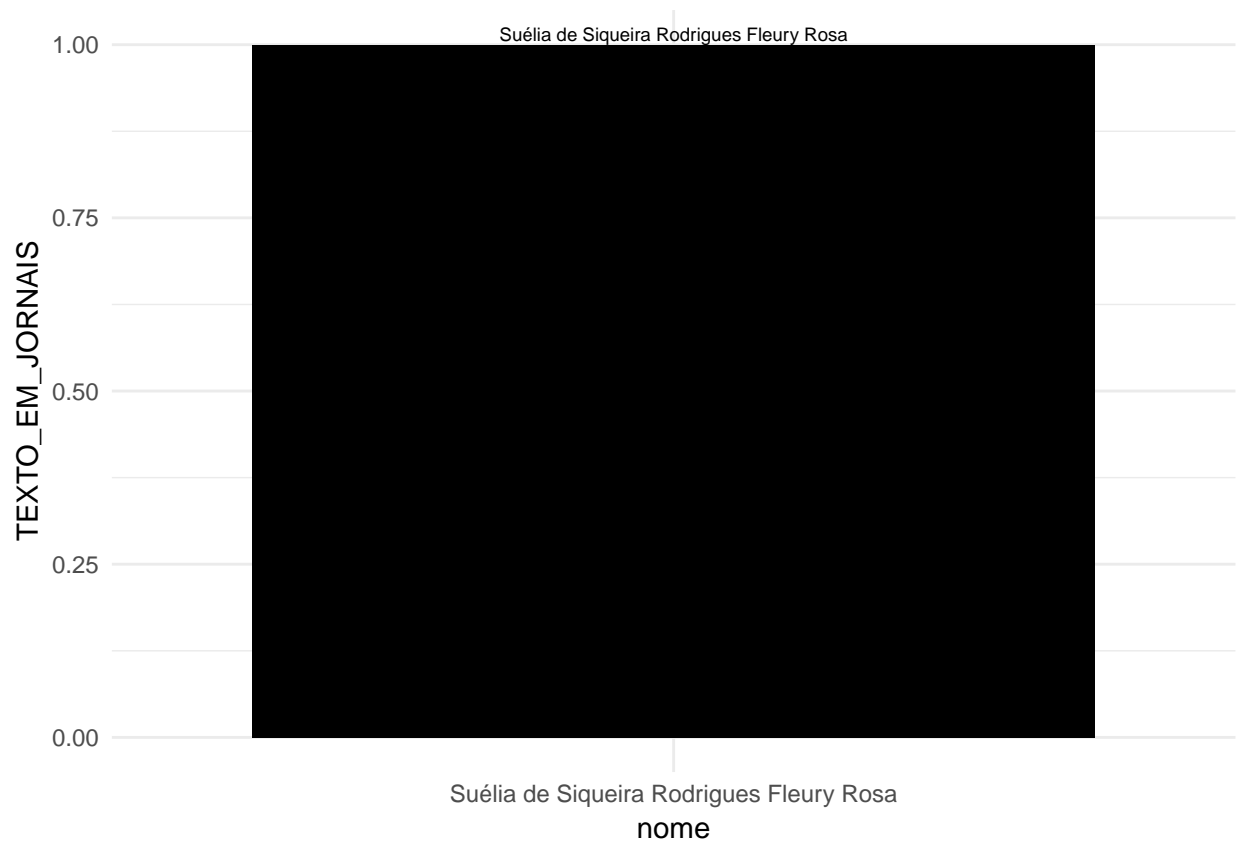
Analizando quantidade de livro por professor.

```
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()
```



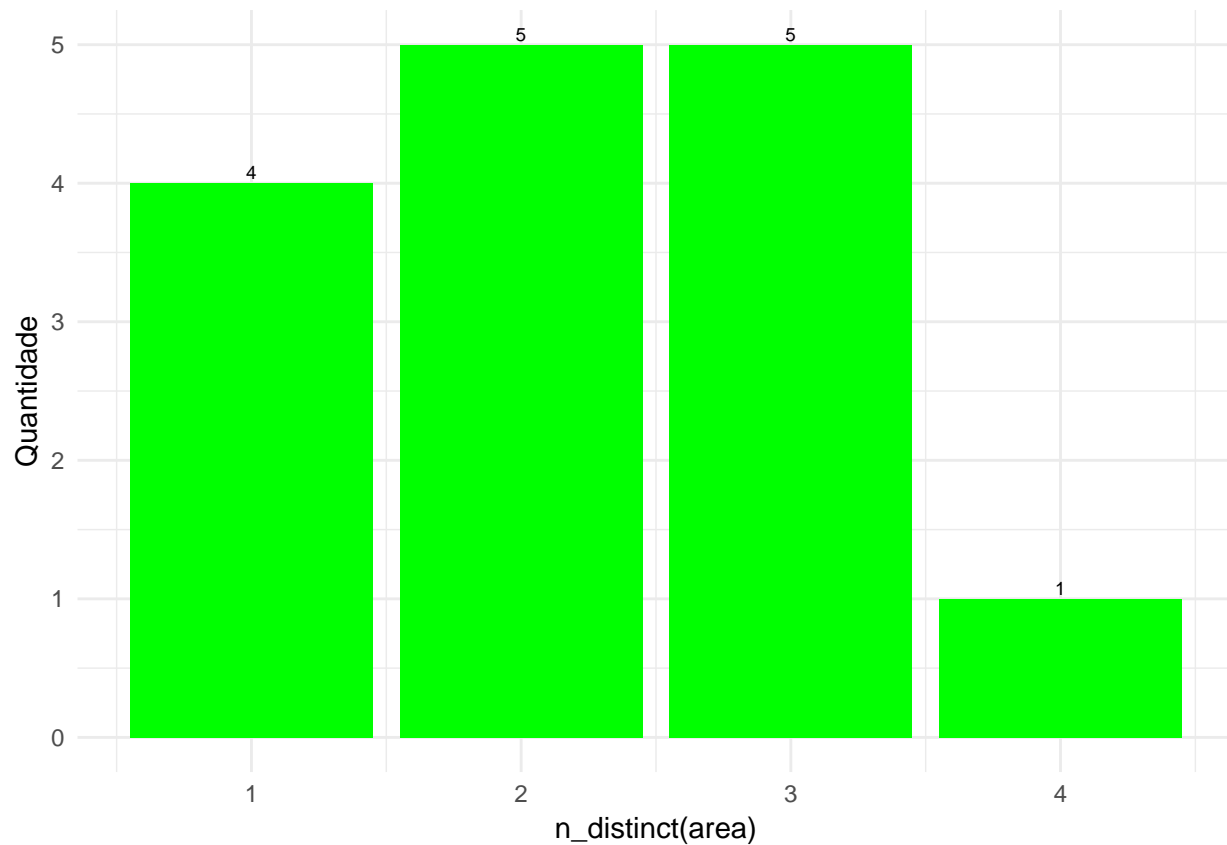
Analisando quantidade de texto em jornais por professor.

```
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()
```



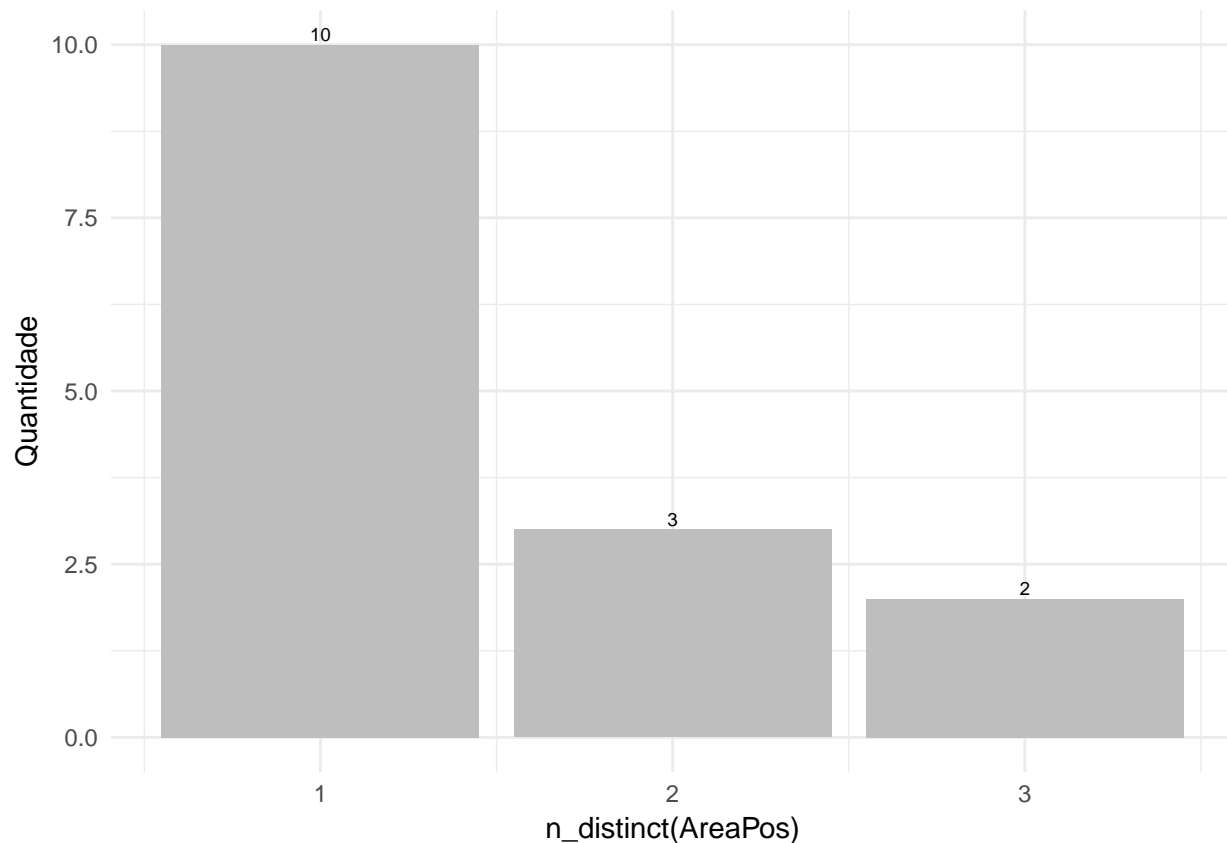
Analisando quantidade de professor por maior área e posteriormente o área pós-graduação.

```
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"
```

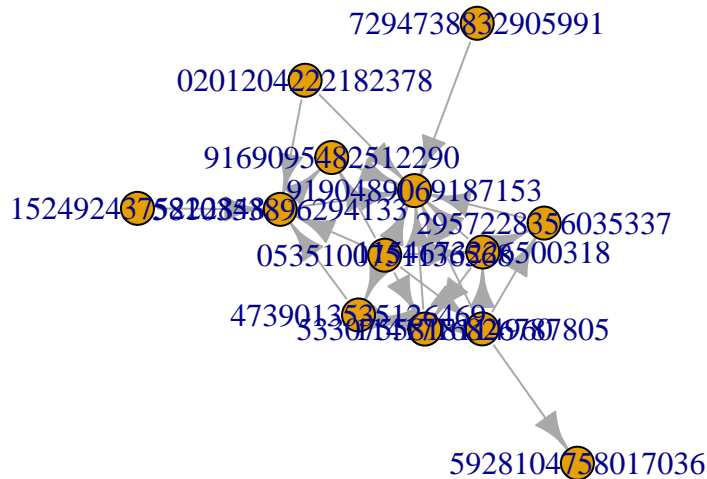
```
g_biomed<-graph(list_biomed, directed=TRUE)
g_biomed[ 1:length(g_biomed) ]
```

```
## 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"

### nuvem de palavras do curriculum
#docs <- Corpus(VectorSource(unb.prof.biomed.df$resumo_cv))
unb.prof.biomed.df$resumo_cv <- iconv(unb.prof.biomed.df$resumo_cv,from="UTF-8", to = "ASCII//TRANSLIT")
df_wc <- data.frame(doc_id = unb.prof.biomed.df$idLattes, text = unb.prof.biomed.df$resumo_cv, stringsAsFactors = FALSE)
docs <- Corpus(DataframeSource(df_wc))
docs <- tm_map(docs, content_transformer(tolower))
docs <- tm_map(docs, removeNumbers)
docs <- tm_map(docs, removeWords, stopwords("pt"))
docs <- tm_map(docs, removeWords, c("universidade", "federal","eletrica","posgraduacao","engenheiro","e"))
docs <- tm_map(docs, removePunctuation)
docs <- tm_map(docs, stripWhitespace)

dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
set.seed(1234)

wordcloud(words = (d$word), freq = d$freq, min.freq = 3,
           max.words=30,scale = c(4, 0.2), random.order=FALSE, rot.per=0.35,
           colors=brewer.pal(8, "Dark2"))

```



Vê-se que as palavras mais usadas são relacionadas a saúde, como “saúde”, “nanotecnologia” e “biomédica”.

```
### nuvem de palavras do título
unb.adv.biomed.df$titulo <- iconv(unb.adv.biomed.df$titulo,from="UTF-8", to = "ASCII//TRANSLIT")
df_wc <- data.frame(doc_id = unb.adv.biomed.df$idLattes1, text = unb.adv.biomed.df$titulo, stringsAsFactors=FALSE)
docs <- Corpus(DataframeSource(df_wc))
docs <- tm_map(docs, content_transformer(tolower))
docs <- tm_map(docs, removeNumbers)
docs <- tm_map(docs, removeWords, stopwords("pt"))
docs <- tm_map(docs, removeWords, stopwords("english"))
docs <- tm_map(docs, removeWords, c("universidade", "federal", "eletrica", "posgraduacao", "engenheiro", "e"))
docs <- tm_map(docs, removePunctuation)
docs <- tm_map(docs, stripWhitespace)

dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
set.seed(1234)

wordcloud(words = (d$word), freq = d$freq, min.freq = 3,
           max.words=30,scale = c(4, 0.2), random.order=FALSE, rot.per=0.35,
           colors=brewer.pal(8, "Dark2"))
```



As palavras mais usadas nos títulos são relacionadas a sinais “processamento, sinais, tratamento” e saúde.

## Expondo engenharia automação

### Expondo engenharia elétrica p8

```
### nuvem de palavras do curriculum
unb.prof.ele_p8.df$resumo_cv <- iconv(unb.prof.ele_p8.df$resumo_cv, from="UTF-8", to = "ASCII//TRANSLIT")
df_wc <- data.frame(doc_id = unb.prof.ele_p8.df$idLattes, text = unb.prof.ele_p8.df$resumo_cv, stringsAsFactors = FALSE)
docs <- Corpus(VectorSource(unb.prof.ele_p8.df$resumo_cv))

docs <- tm_map(docs, content_transformer(tolower))

## Warning in tm_map.SimpleCorpus(docs, content_transformer(tolower)):
## transformation drops documents

docs <- tm_map(docs, removeNumbers)

## Warning in tm_map.SimpleCorpus(docs, removeNumbers): transformation drops
## documents

docs <- tm_map(docs, removeWords, stopwords("pt"))

## Warning in tm_map.SimpleCorpus(docs, removeWords, stopwords("pt")):
## transformation drops documents

docs <- tm_map(docs, removeWords, c("universidade", "federal", "eletrica", "posgraduacao", "engenheiro", "e"))
```





```

docs <- tm_map(docs, content_transformer(tolower))

## Warning in tm_map.SimpleCorpus(docs, content_transformer(tolower)):
## transformation drops documents

docs <- tm_map(docs, removeNumbers)

## Warning in tm_map.SimpleCorpus(docs, removeNumbers): transformation drops
## documents

docs <- tm_map(docs, removeWords, stopwords("pt"))

## Warning in tm_map.SimpleCorpus(docs, removeWords, stopwords("pt")):
## transformation drops documents

docs <- tm_map(docs, removeWords, c("universidade", "federal", "eletrica", "posgraduacao", "engenheiro", "e"))

## Warning in tm_map.SimpleCorpus(docs, removeWords, c("universidade",
## "federal", : transformation drops documents

docs <- tm_map(docs, removePunctuation)

## Warning in tm_map.SimpleCorpus(docs, removePunctuation): transformation
## drops documents

docs <- tm_map(docs, stripWhitespace)

## Warning in tm_map.SimpleCorpus(docs, stripWhitespace): transformation drops
## documents

dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m), decreasing=TRUE)
d <- data.frame(word = names(v), freq=v)
set.seed(1234)

wordcloud(words = (d$word), freq = d$freq, min.freq = 3,
          max.words=30, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))

```



```
### nuvem de palavras do título
unb.adv.ele_p7.df$titulo <- iconv(unb.adv.ele_p7.df$titulo,from="UTF-8", to = "ASCII//TRANSLIT")
df_wc <- data.frame(doc_id = unb.adv.ele_p7.df$idLattes1, text = unb.adv.ele_p7.df$titulo, stringsAsFactors=FALSE)
docs <- Corpus(DataframeSource(df_wc))
docs <- tm_map(docs, content_transformer(tolower))
docs <- tm_map(docs, removeNumbers)
docs <- tm_map(docs, removeWords, stopwords("pt"))
docs <- tm_map(docs, removeWords, stopwords("english"))
docs <- tm_map(docs, removeWords, c("universidade", "federal", "eletrica", "posgraduacao", "engenheiro", "engenharia"))
docs <- tm_map(docs, removePunctuation)
docs <- tm_map(docs, stripWhitespace)

dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
set.seed(1234)

wordcloud(words = (d$word), freq = d$freq, min.freq = 3,
           max.words=30,scale = c(4, 0.2), random.order=FALSE, rot.per=0.35,
           colors=brewer.pal(8, "Dark2"))
```



Muitas palavras ligada a análise, sistema e também há a palavra “business intelligence”.

## 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 decorrendo 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