

Relatório final

Luiza, Oswaldo e Pedro

07/07/2018

Introdução e contextualização

Ciências Médicas

História

O curso de Medicina da UnB foi criado em 1966, então parte da Faculdade de Ciências Médicas. Com a criação de outros cursos na área da saúde, criou-se a Faculdade de Ciências da Saúde, englobando todos os cursos de saúde, e a Faculdade de Ciências Médicas deixou de existir. Na década de 90, porém, por um desejo de maior autonomia dos professores, servidores e alunos, passou-se a se discutir a criação da Faculdade de Medicina (FM), que acabou sendo criada em 1999.

Já existiam cursos de pós-graduação ligados à Medicina sob a Faculdade de Ciências da Saúde, os de Medicina Tropical, Patologia Molecular e Clínica Médica, ainda assim, faltava um curso, de mestrado e doutorado, que pudesse atender à todos docentes e alunos da FM. Assim, a Faculdade de Medicina começou a propor a criação de um novo curso.

Esse curso deveria ser abrangente, com linhas de pesquisa que acolhessem os professores de todas as áreas da FM e que pudesse ser atendido tanto por alunos da FM quanto por alunos de outras áreas, valorizando o trabalho multidisciplinar. Além disso, o curso deveria atender ao requisitos mínimos para a criação de um curso de mestrado e doutorado pela CAPES.

Assim, em 2002, foi criado o Curso de Pós-graduação em Ciências Médicas, com níveis de Mestrado e Doutorado, de acordo com as propostas apresentadas. Após um primeiro semestre no qual foram só feitos ajustes operacionais e um processo seletivo, o curso passou a receber alunos no segundo semestre de 2002. Desde então o curso tem crescido significativamente, se tornando um polo de formação de mestres e doutores no Distrito Federal e na região Centro-Oeste.

Informações

O Programa de Pós-Graduação em Ciências Médicas da UnB é parte da Faculdade de Medicina da UnB e oferece cursos de mestrado e doutorado. O Curso de Mestrado possui duração mínima de dois e duração máxima de quatro semestres letivos, enquanto o Curso de Doutorado possui duração mínima de quatro e duração máxima de oito semestres letivos.

O curso possui dois objetivos principais produzir conhecimento novo mediante pesquisa científica de qualidade e socialmente relevante e capacitar recursos humanos competentes para a prática docente. Além disso, o curso é dividido em duas áreas de concentração: Medicina e Ciências Aplicadas a Saúde. Assim o programa atende tanto profissionais médicos quanto não-médicos de áreas com as quais é possível estabelecer uma relação com a atividade médica.

O programa segue sete linhas de pesquisas, sendo elas:

- Aspectos Clínicos, Epidemiológicos, Experimentais, Microbiológicos, Patológicos, Terapêuticos e Profiláticos das Doenças Crônico-Degenerativas
- Aspectos Clínicos, Epidemiológicos, Experimentais, Microbiológicos, Patológicos, Terapêuticos e Profiláticos das Doenças Infecciosas

- Distúrbios do Metabolismo
- Estudo de plantas, fitoativos e fármacos
- Instrumentação médica e processamentos de sinais biomédicos
- Métodos Epidemiológicos nas Ciências Médicas
- Oncologia e Biologia Molecular Aplicadas à Medicina

Ainda, o programa conta com laboratórios próprios, além de laboratórios mantidos pelos docentes e pesquisadores na Faculdade de Medicina, todos muito bem equipados.

Entre os laboratórios próprios temos laboratórios de Cirurgia Experimental, de Pesquisa em Doença de Chagas, de Genética Médica, de Imunologia Celular, de Pesquisas em Pediatria, de Genética Animal, de Biologia Molecular, de Patologia Clínica, de Anatomia Patológica, de Endocrinologia e Farmacologia Molecular, de Bioquímica e Química de Proteínas, de Virologia Molecular, Medicina Tropical, Dermatômica, Parasitologia e, finalmente, de Fisiologia Pulmonar.

Além destes, temos mantidos pelos docentes e pesquisadores os laboratórios de Nefrologia, Fisiologia Cardiovascular, de Doenças Auto-imunes e Infeciosas, de Gastroenterologia, de Endocrinologia, de Ginecologia, de Psico-farmacologia, de Morfologia e Histologia, de Patologia Geral e de Doenças Auto-imunes.

Finalmente, o Programa de Pós-Graduação em Ciências Médicas tem em seu quadro professores doutores, possui trinta e um docentes permanentes e sete docentes colaboradores, com nenhum docente visitante.

Saúde Coletiva

Informações

A pós graduação em saúde coletiva da UnB começou em 2010 e continua em vigência até hoje. Inicialmente era um programa unicamente de mestrado acadêmico e profissionalizante, hoje o programa conta também com doutorado e pós-doutorado.

Atualmente, o programa possui quatro linhas de pesquisa, essas são:

- Política, Planejamento, Gestão e Atenção à saúde - Visa integrar a análise de formulação, as implementações das políticas de saúde e da organização dos sistemas e serviços de saúde para tentar encontrar mudanças no modelo de atenção a saúde ao mesmo tempo procurando contribuir para o desenvolvimento de estudos das práticas de gestão envolvendo a população.
- Saúde, Cultura e Cidadania - Busca procurar os motivos socioeconômicos, culturais e políticos que causam influência no processo de saúde-doença, analisando casos históricos, sociológicos, políticos e antropológicos para assim promover de uma forma geral a saúde e a cidadania.
- Epidemiologia, Ambiente e Trabalho - Desenvolve estudos voltados a frequência e a distribuição de fenômenos da saúde-doença para assim auxiliar na tomada de decisão na saúde pública
- Pesquisa translacional em Saúde Coletiva - pesquisas interdisciplinares para acelerar trocas entre as pesquisas biomédicas, tecnológicas, clínicas, epidemiológicas e dos programas e políticas de saúde para facilitar sua implantação. É voltado para as inovações em produtos e processos.

Atualmente o programa conta com o total de 23 docentes, abrangendo tanto doutorado quanto mestrado.

Referencial

Ciencia da Ciência

O significado de Ciência ainda causa divergências, pode significar: conhecimento verdadeiro por oposição ao conhecimento errado ou duvidoso; o resultado de experiências em contraste com o que sabemos pelo senso comum; conhecimento medido, quantificado, e não aquele que adquirimos intuitivamente; a Verdade, com V maiúsculo em contraste com as verdades menores; um privilégio dos sábios e iniciados, nunca acessível às massas; um fator da produção, como o capital, o trabalho e a tecnologia; aquilo que fazem os cientistas. Dentre tantas definições é possível perceber que ciência é um tema bem complexo cuja definição varia de acordo com o tempo.

Com a difusão dos meios computacionais na civilização moderna, o acesso e a troca de informações se tornaram extremamente simples e com isso passamos a lidar com uma quantidade de dados inimaginável até algum tempo atrás. Com isso, vivemos em um mundo no qual a informação é cada vez mais ubíqua, o que nos possibilita pensar em novas formas de se entender o mundo.

Nesse contexto, com o crescente aumento da disponibilidade de dados sobre a produção científica, ciência da ciência surgiu do esforço para compreender o significado de ciência e poder agir sobre ele. É conhecido como sociologia da ciência, onde trata de examinar o fenômeno científico como um fator social, sendo a filosofia a mais antiga das ciências da ciência. Esta surge no mundo atual com um enfoque multidisciplinar cujo o objetivo é obter um melhor entendimento de como se dá a produção científica em suas mais diversas formas, tornando possível avaliar de uma forma mais direta o avanço da produção Científica e de grupos de pesquisas.

Lattes

O Lattes é uma plataforma do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) criada como uma forma de integrar e uniformizar os dados dos pesquisadores brasileiros. O uso da plataforma é bem amplo e, uma vez que a visualização é pública, o Lattes é muito utilizado como currículo profissional online. Esse currículo surgiu em 1999, seguindo tentativas anteriores do CNPq de se padronizar o registro dos currículos, e cresceu rapidamente, sendo hoje o principal instrumento de avaliação de pesquisadores, professores e alunos e utilizado pelas principais universidades, institutos e centros de pesquisas.

Por ser um registro padronizado e uma base de dados bem ampla e completa, além de ter as informações disponíveis atualizadas continuamente, o Lattes é muito utilizado para pesquisa e estudos sobre dados acadêmicos pela comunidade científica e tecnológica, sendo fácil avaliar pesquisadores, selecionar consultores e especialistas e gerar estatísticas sobre a produção científica no país. Gerando assim resultados interessantes na esfera social.

Assim, com o crescimento da Ciência da Ciência, é possível prever que a Plataforma Lattes, por oferecer um sistema de informações de grande confiabilidade e abrangência, será de extrema importância para o desenvolvimento desse campo no país.

Estudos de Ciência e Tecnologia

Os estudos sociais da ciência e da tecnologia, também denominados estudos sobre ciência, tecnologia e sociedade (CTS) se dedicam a entender a relação entre essas 3 áreas. Não se trata de uma disciplina específica, mas de uma perspectiva ou enfoque metodológico aplicado ao processo de ensino-aprendizagem na área de ciência e tecnologia em geral, aplicado no processo de conhecimento técnico e científico.

Esta área emergiu por volta do século XIX, influenciado pelo problema dos elementos geradores e decisivos na construção, sustentação e transformação do conhecimento, onde foram primeiramente analisados pelo ponto de vista filosófico e sociológico. Atualmente há novas pesquisas que se diferem pelas investigações voltadas às influências do contexto social na determinação do conteúdo do conhecimento científico. O amadurecimento

dos estudos sobre ciência e tecnologia é destacado pela institucionalização de grupos de pesquisa, linhas temáticas em programas de pós-graduação, periódicos e congressos especializados e por uma multiplicidade de abordagens.

Metodologia

A metodologia Cross Industry Standard Process for Data Mining (CRISP-DM) é uma metodologia especificamente desenhada para processos de mineração de dados que descreve abordagens comumente usadas por especialistas para atacar problemas utilizando um processo que fosse confiável e replicável e foi utilizada neste trabalho. As etapas estão demonstradas no Figura 1

No CRISP-DM o trabalho se dá em um ciclo de seis fases, como se observa na figura acima, sendo elas:

1. Entendimento do Negócio.

Procura-se entender melhor o que precisamos analisar, buscando detalhes de como as coisas funcionam e como diversos elementos da pesquisa se relacionam. É importante, também, nesta etapa, entender bem os objetivos e as expectativas quanto ao trabalho. No caso deste projeto busca-se entender a necessidade do Sistema Nacional de Pós-Graduação do Brasil de produzir análises de alta qualidade de suas pós-graduações, com baixo custo.

Além disso, é nesta fase que são identificados os recursos e dificuldades que podem influenciar no desenvolvimento do projeto. Tendo sido identificado o tempo especificado para produzir resultados tendo outros trabalhos e outras disciplinas a serem cursadas ao mesmo tempo. Além disso tem o risco de abandono dos integrantes do grupo ou até mesmo a falta de dedicação necessária.

2. Coompreensão dos Dados.

Aqui procuramos entender e descrever todos os dados disponíveis. Assim, é fundamental que se tenha um bom entendimento de quais dados serão relevantes e quais não serão.

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

```
perfisCM <- fromJSON("PPG Ciências Medicas.profile.json")
perfisPSC <- fromJSON("PPG Profissional Saúde Coletiva.profile.json")
perfisSC <- fromJSON("PPG Saúde Coletiva.profile.json")

adviseCM <- fromJSON("PPG Ciências Medicas.advise.json")
advisePSC <- fromJSON("PPG Profissional Saúde Coletiva.advise.json")
adviseSC <- fromJSON("PPG Saúde Coletiva.advise.json")

publicationCM <- fromJSON("PPG Ciências Medicas.publication.json")
publicationPSC <- fromJSON("PPG Profissional Saúde Coletiva.publication.json")
publicationSC <- fromJSON("PPG Saúde Coletiva.publication.json")

#unb.BDT <- fromJSON("unb.BDTD.json")
#unb.Oasis <- fromJSON("oasisbr_unb_all.json")

DGP_Grupos <- read_excel("unb.DGP.xls", sheet = 1, skip = 2)
DGP_Pesquisa <- read_excel("unb.DGP.xls", sheet = 2)
DGP_Parcerias <- read_excel("unb.DGP.xls", sheet = 3)
```

```
DGP_Participantes <-read_excel("unb.DGP.xls", sheet = 4)
DGP_Producao <- read_excel("unb.DGP.xls", sheet = 5)
```

Tanto o arquivo “PPG Ciências Medicas.profile.json”, “PPG Profissional Saúde Coletiva.profile.json” e “PPG Saúde Coletiva.profile.json” apresentam o perfil dos docentes vinculados a programas das áreas de Ciências Médicas, Saúde Coletiva Profissional e Saúde Coletiva, respectivamente.

Já os arquivos “PPG Ciências Medicas.advise.json”, “PPG Profissional Saúde Coletiva.advise.json” e “PPG Saúde Coletiva.advise.json” apresentam dados sobre as orientações de mestrado e doutorado feitas por todos os docentes vinculados a programas de pós-graduação das áreas de Ciências Médicas, Saúde Coletiva Profissional e Saúde Coletiva, respectivamente.

Finalmente, os arquivos “PPG Ciências Medicas.publication.json”, “PPG Profissional Saúde Coletiva.publication.json” e “PPG Saúde Coletiva.publication.json” apresentam dados sobre a produção bibliográfica gerada por todos os docentes vinculados a programas das áreas de Ciências Médicas, Saúde Coletiva Profissional e Saúde Coletiva, respectivamente.

Descrição e análise exploratória dos dados

A análise foi realizada no RStudio utilizando a linguagem, para leitura e interpretação dos dados serão usados as seguintes bibliotecas abaixo.

Perfis

Os arquivos “PPG Ciências Medicas.profile.json”, “PPG Profissional Saúde Coletiva.profile.json” e “PPG Saúde Coletiva.profile.json” contêm dados que caracterizam o perfil profissional de todos os docentes do grupo sob análise. Com isso temos o nome, as áreas de atuação, endereço profissional, as produções bibliográficas, as orientações acadêmicas e a senioridade de cada professor. Todos os três arquivos tem a mesma estrutura, diferenciando-se apenas nos docentes presentes.

```
summary(perfisCM$`0056945699042193`)
```

##	Length	Class	Mode
## nome	1	-none-	character
## resumo_cv	1	-none-	character
## areas_de_atuacao	4	data.frame	list
## endereco_profissional	8	-none-	list
## producao_bibliografica	2	-none-	list
## orientacoes_academicas	5	-none-	list
## senioridade	1	-none-	character

```
summary(perfisPSC$`0153088935731412`)
```

##	Length	Class	Mode
## nome	1	-none-	character
## resumo_cv	1	-none-	character
## areas_de_atuacao	4	data.frame	list
## endereco_profissional	8	-none-	list
## producao_bibliografica	5	-none-	list
## orientacoes_academicas	2	-none-	list
## senioridade	1	-none-	character

```
summary(perfisSC$`0153088935731412`)
```

##	Length	Class	Mode
## nome	1	-none-	character

```
## resumo_cv          1      -none-    character
## areas_de_atuacao   4      data.frame list
## endereco_profissional 8      -none-    list
## producao_bibliografica 5      -none-    list
## orientacoes_academicas 2      -none-    list
## senioridade        1      -none-    character
```

Nos arquivos, encontramos um total de 38 docentes da Ciências Médicas, 27 docentes do Profissional de Saúde Coletiva e 24 docentes de Saúde Coletiva.

```
# Numero de docentes
length(perfisCM)
```

```
## [1] 38
```

```
length(perfisPSC)
```

```
## [1] 27
```

```
length(perfisSC)
```

```
## [1] 24
```

Podemos notar também que temos 5 grandes áreas no programa de Ciências Médicas, 8 no de Saúde Coletiva Profissional e 9 no de Saúde Coletiva Profissional. Para áreas, temos 22, 24 e 24, respectivamente. Já para sub-áreas, temos 75, 64 e 49, respectivamente. Finalmente, para especialidades, temos 40, 21 e 21, respectivamente.

```
# Quantidade de areas, sub-areas e especialidades
data.frame(Reduce(rbind, perfisCM)) %>%
  unnest(areas_de_atuacao) %>%
  summarise(qts_grande_areas = n_distinct(grande_area),
            qtd_areas = n_distinct(area),
            qtd_sub_areas = n_distinct(sub_area),
            qtd_especialidades = n_distinct(especialidade))
```

```
##   qts_grande_areas qtd_areas qtd_sub_areas qtd_especialidades
## 1                5        22           75                40
```

```
data.frame(Reduce(rbind, perfisPSC)) %>%
  unnest(areas_de_atuacao) %>%
  summarise(qts_grande_areas = n_distinct(grande_area),
            qtd_areas = n_distinct(area),
            qtd_sub_areas = n_distinct(sub_area),
            qtd_especialidades = n_distinct(especialidade))
```

```
##   qts_grande_areas qtd_areas qtd_sub_areas qtd_especialidades
## 1                8        24           64                21
```

```
data.frame(Reduce(rbind, perfisSC)) %>%
  unnest(areas_de_atuacao) %>%
  summarise(qts_grande_areas = n_distinct(grande_area),
            qtd_areas = n_distinct(area),
            qtd_sub_areas = n_distinct(sub_area),
            qtd_especialidades = n_distinct(especialidade))
```

```
##   qts_grande_areas qtd_areas qtd_sub_areas qtd_especialidades
## 1                9        49           49                21
```

Em todos os três programas temos que ciências da saúde é a grande área de maior prevalência. Em ciências médicas, podemos ver ainda uma grande concentração na grande área de ciências biológicas.

```
# Quantidade de pessoas por grande area
data.frame(Reduce(rbind, perfisCM)) %>%
  unnest(areas_de_atuacao) %>%
  group_by(grande_area) %>%
  tally() %>% arrange(desc(n)) %>% na_if("")
```

```
## # A tibble: 5 x 2
##   grande_area      n
##   <chr>          <int>
## 1 CIENCIAS_DA_SAUDE      86
## 2 CIENCIAS_BIOLOGICAS    42
## 3 CIENCIAS_EXATAS_E_DA_TERRA    6
## 4 ENGENHARIAS            3
## 5 CIENCIAS_HUMANAS        2
```

```
data.frame(Reduce(rbind, perfisPSC)) %>%
  unnest(areas_de_atuacao) %>%
  group_by(grande_area) %>%
  tally() %>% arrange(desc(n)) %>% na_if("")
```

```
## # A tibble: 8 x 2
##   grande_area      n
##   <chr>          <int>
## 1 CIENCIAS_DA_SAUDE      81
## 2 CIENCIAS_HUMANAS      11
## 3 CIENCIAS_SOCIAIS_APLICADAS    9
## 4 OUTROS                  2
## 5 <NA>                    1
## 6 CIENCIAS_AGRARIAS         1
## 7 CIENCIAS_EXATAS_E_DA_TERRA    1
## 8 LINGUISTICA_LETRAS_E_ARTES    1
```

```
data.frame(Reduce(rbind, perfisSC)) %>%
  unnest(areas_de_atuacao) %>%
  group_by(grande_area) %>%
  tally() %>% arrange(desc(n)) %>% na_if("")
```

```
## # A tibble: 9 x 2
##   grande_area      n
##   <chr>          <int>
## 1 CIENCIAS_DA_SAUDE      71
## 2 CIENCIAS_SOCIAIS_APLICADAS    10
## 3 CIENCIAS_HUMANAS          7
## 4 CIENCIAS_BIOLOGICAS        6
## 5 OUTROS                      2
## 6 <NA>                        1
## 7 CIENCIAS_AGRARIAS          1
## 8 CIENCIAS_EXATAS_E_DA_TERRA    1
## 9 LINGUISTICA_LETRAS_E_ARTES    1
```

Em todos os arquivos, temos que a área de maior prevalência é a do curso principal do programa, Medicina no caso de Ciências Médicas e Saúde Coletiva nos outros dois.

```
# Quantidade de pessoas por area
data.frame(Reduce(rbind, perfisCM)) %>%
  unnest(areas_de_atuacao) %>%
  group_by(area) %>%
  tally() %>% arrange(desc(n)) %>% na_if("")
```

```
## # A tibble: 22 x 2
##   area      n
##   <chr>    <int>
## 1 Medicina    66
## 2 Farmácia    10
## 3 Parasitologia    8
## 4 Genética      6
## 5 Morfologia      6
## 6 Fisiologia      5
## 7 Imunologia      5
## 8 Química        5
## 9 Bioquímica      4
## 10 Farmacologia    4
## # ... with 12 more rows
```

```
data.frame(Reduce(rbind, perfisPSC)) %>%
  unnest(areas_de_atuacao) %>%
  group_by(area) %>%
  tally() %>% arrange(desc(n)) %>% na_if("")
```

```
## # A tibble: 24 x 2
##   area      n
##   <chr>    <int>
## 1 Saúde Coletiva    54
## 2 Enfermagem        8
## 3 Farmácia          7
## 4 Nutrição          5
## 5 Medicina          4
## 6 Economia          3
## 7 Psicologia        3
## 8 Antropologia      2
## 9 Ciência Política  2
## 10 Comunicação      2
## # ... with 14 more rows
```

```
data.frame(Reduce(rbind, perfisSC)) %>%
  unnest(areas_de_atuacao) %>%
  group_by(area) %>%
  tally() %>% arrange(desc(n)) %>% na_if("")
```

```
## # A tibble: 24 x 2
##   area      n
##   <chr>    <int>
## 1 Saúde Coletiva    50
## 2 Nutrição          11
## 3 Farmácia          4
## 4 Medicina          4
## 5 Bioquímica        3
## 6 Ciência Política  3
```



```
## 7 Economia 3
## 8 Antropologia 2
## 9 Comunicação 2
## 10 Direito 2
## # ... with 14 more rows
```

Nas sub-áreas, temos que nos dois programas da Saúde Coletiva a opção mais comum é não ter sub-área, enquanto que essa opção é a segunda em Ciências Médicas, perdendo apenas para Clínica Médica. Em todos os casos, as três primeiras opções são bem mais populares do que as demais.

```
# Quantidade de pessoas por sub-área
data.frame(Reduce(rbind, perfisCM)) %>%
  unnest(areas_de_atuacao) %>%
  group_by(sub_area) %>%
  tally() %>% arrange(desc(n)) %>% na_if("")
```

```
## # A tibble: 75 x 2
##   sub_area n
##   <chr>   <int>
## 1 Clínica Médica 23
## 2 <NA> 19
## 3 Cirurgia 11
## 4 Anatomia Patológica e Patologia Clínica 4
## 5 Fisiologia de Órgãos e Sistemas 3
## 6 Imunologia Aplicada 3
## 7 Farmacologia Bioquímica e Molecular 2
## 8 Genética Humana e Médica 2
## 9 Pediatria 2
## 10 Pneumologia 2
## # ... with 65 more rows
```

```
data.frame(Reduce(rbind, perfisPSC)) %>%
  unnest(areas_de_atuacao) %>%
  group_by(sub_area) %>%
  tally() %>% arrange(desc(n)) %>% na_if("")
```

```
## # A tibble: 64 x 2
##   sub_area n
##   <chr>   <int>
## 1 <NA> 23
## 2 Epidemiologia 9
## 3 Saúde Pública 9
## 4 Ensino na Saúde 2
## 5 Medicina Preventiva 2
## 6 Políticas Públicas 2
## 7 PROMOÇÃO E EQUIDADE EM SAÚDE 2
## 8 Tratamento e Prevenção Psicológica 2
## 9 Análise da Gestão das Políticas e dos Serviços de Saúde 1
## 10 Análise Nutricional de População 1
## # ... with 54 more rows
```

```
data.frame(Reduce(rbind, perfisSC)) %>%
  unnest(areas_de_atuacao) %>%
  group_by(sub_area) %>%
  tally() %>% arrange(desc(n)) %>% na_if("")
```

```
## # A tibble: 49 x 2
```

```
##   sub_area                                n
##   <chr>                                <int>
## 1 <NA>                                23
## 2 Saúde Pública                        13
## 3 Epidemiologia                       12
## 4 Análise Nutricional de População     3
## 5 Políticas Públicas                   3
## 6 Bioética                             2
## 7 Medicina Preventiva                  2
## 8 Administração Pública                1
## 9 Aleitamento Materno                  1
## 10 Análise da Gestão das Políticas e dos Serviços de Saúde 1
## # ... with 39 more rows
```

Para as especialidades, temos que a esmagadora maioria dos casos não possuem especialidade e que as especialidades que existem aparecem poucas vezes.

```
# Quantidade de pessoas por especialidades
data.frame(Reduce(rbind, perfisCM)) %>%
  unnest(areas_de_atuacao) %>%
  group_by(especialidade) %>%
  tally() %>% arrange(desc(n)) %>% na_if("")
```

```
## # A tibble: 40 x 2
##   especialidade                                n
##   <chr>                                <int>
## 1 <NA>                                89
## 2 Doenças Infecciosas e Parasitárias      3
## 3 Cardiologia                             2
## 4 Cirurgia Experimental                    2
## 5 Cirurgia Otorrinolaringológica           2
## 6 Cirurgia Proctológica                    2
## 7 Dermatologia                           2
## 8 Fisiologia Cardiovascular                2
## 9 Pneumologia                             2
## 10 Química dos Produtos Naturais           2
## # ... with 30 more rows
```

```
data.frame(Reduce(rbind, perfisPSC)) %>%
  unnest(areas_de_atuacao) %>%
  group_by(especialidade) %>%
  tally() %>% arrange(desc(n)) %>% na_if("")
```

```
## # A tibble: 21 x 2
##   especialidade                                n
##   <chr>                                <int>
## 1 <NA>                                87
## 2 Antropologia da Saúde                    1
## 3 Avaliação Epidemiológica de Intervenções 1
## 4 Bioestatística                           1
## 5 Bioética                                 1
## 6 Desigualdades Em Saúde                    1
## 7 Determinantes sociais de saúde            1
## 8 Direito Sanitário                         1
## 9 Distúrbios Psicossomáticos                1
## 10 Educação Em Saúde                        1
```

```
## # ... with 11 more rows
```

```
data.frame(Reduce(rbind, perfisSC)) %>%  
  unnest(areas_de_atuacao) %>%  
  group_by(especialidade) %>%  
  tally() %>% arrange(desc(n)) %>% na_if("")
```

```
## # A tibble: 21 x 2  
##   especialidade      n  
##   <chr>          <int>  
## 1 <NA>          80  
## 2 Antropologia da Saúde      1  
## 3 Avaliação Epidemiológica de Intervenções      1  
## 4 Bioestatística            1  
## 5 Bioética e Enfermagem      1  
## 6 Bioética; Bioética e Gênero; Direitos Humanos      1  
## 7 Desigualdades Em Saúde      1  
## 8 Direito Sanitário          1  
## 9 Educação Em Saúde          1  
## 10 Epidemiologia            1  
## # ... with 11 more rows
```

Orientações

```
names(adviseCM)
```

```
## [1] "ORIENTACAO_EM_ANDAMENTO_DE_POS_DOUTORADO"  
## [2] "ORIENTACAO_EM_ANDAMENTO_DOUTORADO"  
## [3] "ORIENTACAO_EM_ANDAMENTO_MESTRADO"  
## [4] "ORIENTACAO_EM_ANDAMENTO_GRADUACAO"  
## [5] "ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA"  
## [6] "ORIENTACAO_CONCLUIDA_POS_DOUTORADO"  
## [7] "ORIENTACAO_CONCLUIDA_DOUTORADO"  
## [8] "ORIENTACAO_CONCLUIDA_MESTRADO"  
## [9] "OUTRAS_ORIENTACOES_CONCLUIDAS"
```

```
names(advisePSC)
```

```
## [1] "ORIENTACAO_EM_ANDAMENTO_DE_POS_DOUTORADO"  
## [2] "ORIENTACAO_EM_ANDAMENTO_DOUTORADO"  
## [3] "ORIENTACAO_EM_ANDAMENTO_MESTRADO"  
## [4] "ORIENTACAO_EM_ANDAMENTO_GRADUACAO"  
## [5] "ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA"  
## [6] "ORIENTACAO_CONCLUIDA_POS_DOUTORADO"  
## [7] "ORIENTACAO_CONCLUIDA_DOUTORADO"  
## [8] "ORIENTACAO_CONCLUIDA_MESTRADO"  
## [9] "OUTRAS_ORIENTACOES_CONCLUIDAS"
```

```
names(adviseSC)
```

```
## [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"

## adivise da base de dados de Ciências Médicas
advise <- adviseCM
andamento <- data.frame()
for( i in 1:length(advise$ORIENTACAO_EM_ANDAMENTO_MESTRADO))
andamento <- rbind(andamento, advise$ORIENTACAO_EM_ANDAMENTO_MESTRADO[[i]])
for( i in 1:length(advise$ORIENTACAO_EM_ANDAMENTO_DE_POS_DOUTORADO))
andamento <- rbind(andamento, advise$ORIENTACAO_EM_ANDAMENTO_DE_POS_DOUTORADO[[i]])
for( i in 1:length(advise$ORIENTACAO_EM_ANDAMENTO_DOUTORADO))
andamento <- rbind(andamento, advise$ORIENTACAO_EM_ANDAMENTO_DOUTORADO[[i]])
for( i in 1:length(advise$ORIENTACAO_EM_ANDAMENTO_GRADUACAO))
andamento <- rbind(andamento, advise$ORIENTACAO_EM_ANDAMENTO_GRADUACAO[[i]])
for( i in 1:length(advise$ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA))
andamento <- rbind(andamento, advise$ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA[[i]])

concluido <- data.frame()
for( i in 1:length(advise$ORIENTACAO_CONCLUIDA_MESTRADO))
concluido <- rbind(concluido, advise$ORIENTACAO_CONCLUIDA_MESTRADO[[i]])
for( i in 1:length(advise$ORIENTACAO_CONCLUIDA_POS_DOUTORADO))
concluido <- rbind(concluido, advise$ORIENTACAO_CONCLUIDA_POS_DOUTORADO[[i]])
for( i in 1:length(advise$ORIENTACAO_CONCLUIDA_DOUTORADO))
concluido <- rbind(concluido, advise$ORIENTACAO_CONCLUIDA_DOUTORADO[[i]])
for( i in 1:length(advise$OUTRAS_ORIENTACOES_CONCLUIDAS))
concluido <- rbind(concluido, advise$OUTRAS_ORIENTACOES_CONCLUIDAS[[i]])

concluidoCM <- concluido
andamentoCM <- andamento

## adivise da base de dados de Saude Coletiva
advise <- adviseSC
andamento <- data.frame()
for( i in 1:length(advise$ORIENTACAO_EM_ANDAMENTO_MESTRADO))
andamento <- rbind(andamento, advise$ORIENTACAO_EM_ANDAMENTO_MESTRADO[[i]])
for( i in 1:length(advise$ORIENTACAO_EM_ANDAMENTO_DE_POS_DOUTORADO))
andamento <- rbind(andamento, advise$ORIENTACAO_EM_ANDAMENTO_DE_POS_DOUTORADO[[i]])
for( i in 1:length(advise$ORIENTACAO_EM_ANDAMENTO_DOUTORADO))
andamento <- rbind(andamento, advise$ORIENTACAO_EM_ANDAMENTO_DOUTORADO[[i]])
for( i in 1:length(advise$ORIENTACAO_EM_ANDAMENTO_GRADUACAO))
andamento <- rbind(andamento, advise$ORIENTACAO_EM_ANDAMENTO_GRADUACAO[[i]])
for( i in 1:length(advise$ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA))
andamento <- rbind(andamento, advise$ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA[[i]])

concluido <- data.frame()
for( i in 1:length(advise$ORIENTACAO_CONCLUIDA_MESTRADO))
concluido <- rbind(concluido, advise$ORIENTACAO_CONCLUIDA_MESTRADO[[i]])
for( i in 1:length(advise$ORIENTACAO_CONCLUIDA_POS_DOUTORADO))
concluido <- rbind(concluido, advise$ORIENTACAO_CONCLUIDA_POS_DOUTORADO[[i]])
for( i in 1:length(advise$ORIENTACAO_CONCLUIDA_DOUTORADO))
concluido <- rbind(concluido, advise$ORIENTACAO_CONCLUIDA_DOUTORADO[[i]])
for( i in 1:length(advise$OUTRAS_ORIENTACOES_CONCLUIDAS))
concluido <- rbind(concluido, advise$OUTRAS_ORIENTACOES_CONCLUIDAS[[i]])

```

```

andamentoSC <- andamento
concluidoSC <- concluido
## adivise da base de dados de Profissional Saude Coletiva
advise <- advisePSC
andamento <- data.frame()
for( i in 1:length(advise$ORIENTACAO_EM_ANDAMENTO_MESTRADO))
andamento <- rbind(andamento, advise$ORIENTACAO_EM_ANDAMENTO_MESTRADO[[i]])
for( i in 1:length(advise$ORIENTACAO_EM_ANDAMENTO_DE_POS_DOUTORADO))
andamento <- rbind(andamento, advise$ORIENTACAO_EM_ANDAMENTO_DE_POS_DOUTORADO[[i]])
for( i in 1:length(advise$ORIENTACAO_EM_ANDAMENTO_DOUTORADO))
andamento <- rbind(andamento, advise$ORIENTACAO_EM_ANDAMENTO_DOUTORADO[[i]])
for( i in 1:length(advise$ORIENTACAO_EM_ANDAMENTO_GRADUACAO))
andamento <- rbind(andamento, advise$ORIENTACAO_EM_ANDAMENTO_GRADUACAO[[i]])
for( i in 1:length(advise$ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA))
andamento <- rbind(andamento, advise$ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA[[i]])

concluido <- data.frame()
for( i in 1:length(advise$ORIENTACAO_CONCLUIDA_MESTRADO))
concluido <- rbind(concluido, advise$ORIENTACAO_CONCLUIDA_MESTRADO[[i]])
for( i in 1:length(advise$ORIENTACAO_CONCLUIDA_POS_DOUTORADO))
concluido <- rbind(concluido, advise$ORIENTACAO_CONCLUIDA_POS_DOUTORADO[[i]])
for( i in 1:length(advise$ORIENTACAO_CONCLUIDA_DOUTORADO))
concluido <- rbind(concluido, advise$ORIENTACAO_CONCLUIDA_DOUTORADO[[i]])
for( i in 1:length(advise$OUTRAS_ORIENTACOES_CONCLUIDAS))
concluido <- rbind(concluido, advise$OUTRAS_ORIENTACOES_CONCLUIDAS[[i]])

andamentoPSC <- andamento
concluidoPSC <- concluido

```

Da Ciências Médicas foram encontrados 214 orientações em andamento e 873 orientações concluídas

```
dim(andamentoCM)
```

```
## [1] 214 13
```

```
dim(concluidoCM)
```

```
## [1] 873 13
```

Da Saúde Coletiva foram encontrados 111 orientações em andamento e 737 orientações concluídas

```
dim(andamentoSC)
```

```
## [1] 111 13
```

```
dim(concluidoSC)
```

```
## [1] 737 13
```

Da Saúde Coletiva foram encontrados 112 orientações em andamento e 682 orientações concluídas

```
dim(andamentoPSC)
```

```
## [1] 112 13
```

```
dim(concluidoPSC)
```

```
## [1] 682 13
```

Publicações

As publicações são divididas em 7 tipos (periódicos, livros, capítulos de livros, jornais, eventos, artigos, demais), todos esses tipos estavam agrupados por ano.

```
names(publicationCM)
```

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

```
names(publicationPSC)
```

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

```
names(publicationSC)
```

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

```
publication_periodicoCM <- publicationCM$PERIODICO %>% ldply(data.frame)  
publication_livroCM <- publicationCM$LIVRO %>% ldply(data.frame)  
publication_cap_livroCM <- publicationCM$CAPITULO_DE_LIVRO %>% ldply(data.frame)  
publication_jornaisCM <- publicationCM$TEXTO_EM_JORNAIS %>% ldply(data.frame)  
publication_eventoCM <- publicationCM$EVENTO %>% ldply(data.frame)  
publication_artigosCM <- publicationCM$ARTIGO_ACEITO %>% ldply(data.frame)  
publication_demaismCM <- publicationCM$DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA %>% ldply(data.frame)
```

```
publication_periodicoSC <- publicationSC$PERIODICO %>% ldply(data.frame)  
publication_livroSC <- publicationSC$LIVRO %>% ldply(data.frame)  
publication_cap_livroSC <- publicationSC$CAPITULO_DE_LIVRO %>% ldply(data.frame)  
publication_jornaisSC <- publicationSC$TEXTO_EM_JORNAIS %>% ldply(data.frame)  
publication_eventoSC <- publicationSC$EVENTO %>% ldply(data.frame)  
publication_artigosSC <- publicationSC$ARTIGO_ACEITO %>% ldply(data.frame)  
publication_demaismSC <- publicationSC$DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA %>% ldply(data.frame)
```

```
publication_periodicoPSC <- publicationPSC$PERIODICO %>% ldply(data.frame)  
publication_livroPSC <- publicationPSC$LIVRO %>% ldply(data.frame)  
publication_cap_livroPSC <- publicationPSC$CAPITULO_DE_LIVRO %>% ldply(data.frame)  
publication_jornaisPSC <- publicationPSC$TEXTO_EM_JORNAIS %>% ldply(data.frame)  
publication_eventoPSC <- publicationPSC$EVENTO %>% ldply(data.frame)  
publication_artigosPSC <- publicationPSC$ARTIGO_ACEITO %>% ldply(data.frame)
```

```
publication_demaisPSC <- publicationPSC$DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA %>% ldply(data.frame)
```

Entretanto, cada tipo de publicação possui dados diferentes, como é o caso dos periódicos e dos eventos, por exemplo. Os dados dos periódicos são os seguintes:

```
names(publication_periodicoCM)
```

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

```
names(publication_periodicoPSC)
```

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

```
names(publication_periodicoSC)
```

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

Já os dados dos eventos são os seguintes:

```
names(publication_eventoCM)
```

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

```
names(publication_eventoPSC)
```

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

```
names(publication_eventoSC)
```

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

Também é interessante reparar na quantidade de publicações cada programa possui.

Para o programa de Ciencias Medicas temos:

```
nrow(publication_periodicoCM)
```

```
## [1] 1173
```

```
nrow(publication_livroCM)
```

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

```
nrow(publication_cap_livroCM)
```

```
## [1] 87
```

```
nrow(publication_jornaisCM)
```

```
## [1] 61
```

```
nrow(publication_eventoCM)
```

```
## [1] 756
```

```
nrow(publication_artigosCM)
```

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

```
nrow(publication_demaisCM)
```

```
## [1] 8
```

Para o programa de Saude coletiva temos:

```
nrow(publication_periodicoSC)
```

```
## [1] 772
```

```
nrow(publication_livroSC)
```

```
## [1] 48
```

```
nrow(publication_cap_livroSC)
```

```
## [1] 133
```

```
nrow(publication_jornaisSC)
```

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

```
nrow(publication_eventoSC)
```

```
## [1] 319
```

```
nrow(publication_artigosSC)
```

```
## [1] 27
```

```
nrow(publication_demaisSC)
```

```
## [1] 36
```

Para o programa de Saude coletiva Profissional temos:

```
nrow(publication_periodicoPSC)
```

```
## [1] 578
```

```
nrow(publication_livroPSC)
```

```
## [1] 43
```

```
nrow(publication_cap_livroPSC)
```

```
## [1] 130
```



```
nrow(publication_jornaisPSC)
```

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

```
nrow(publication_eventoPSC)
```

```
## [1] 277
```

```
nrow(publication_artigosPSC)
```

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

```
nrow(publication_demaisPSC)
```

```
## [1] 30
```

3. Preparação dos Dados.

Uma vez que exista um bom entendimento dos dados, precisamos preparar esses dados para a análise, definindo em qual formato esses dados serão mais úteis, bem como escolhendo quais dados são relevantes para a análise. Nesta etapa é preciso selecionar e escolher os atributos dos dados que serão trabalhados. Tendo sido selecionado as base de dados BDTD, DGP e Oasis para complementar os dados obtidos do Lattes sobre Saude Coletiva. Sendo realizado a integração dessas bases, fazendo merges e aggregations com as funções `inner_join()`, `rbind()`.

Na base de dados do DPG existem informações sobre as linhas de pesquisa, as parcerias, os participantes, os numeros de participação e grupos de pesquisa.

```
excel_sheets("unb.DGP.xls")
```

```
## [1] "Grupos" "Linhas de Pesquisa" "Parcerias"
```

```
## [4] "Participantes" "Produção"
```

Para uma futura junção entre as bases do Lattes com as outras foi criada uma tabela de cada uma das bases de dados contendo o nome e o id lattes dos professores

```
perfis_tblCM <- data.frame(Reduce(rbind, perfisCM)) #carrega lista de perfis
idsCM <- names(perfisCM) # carrega os ids do lattes de cada professor
nomes_id_tblCM <- perfis_tblCM %>% mutate(ids = idsCM) %>%
select(name = nome, id_lates = ids) %>% as_tibble() # junta o nome e id
```

```
perfis_tblSC <- data.frame(Reduce(rbind, perfisSC)) #carrega lista de perfis
idsSC <- names(perfisSC) # carrega os ids do lattes de cada professor
nomes_id_tblSC <- perfis_tblSC %>% mutate(ids = idsSC) %>%
select(name = nome, id_lates = ids) %>% as_tibble() # junta o nome e id
```

```
perfis_tblPSC <- data.frame(Reduce(rbind, perfisPSC)) #carrega lista de perfis
idsPSC <- names(perfisPSC) # carrega os ids do lattes de cada professor
nomes_id_tblPSC <- perfis_tblPSC %>% mutate(ids = idsPSC) %>%
select(name = nome, id_lates = ids) %>% as_tibble() # junta o nome e id
```

Em seguida, nos dados do DGP, foi aplicado um filtro para trabalharmos somente com os valores mais recentes, isso é, o do ano de 2016

```
DGP_Producao <- DGP_Producao %>%
filter(`Ano Censo` == 2016) %>%
select(-`Ano Censo`)
```

```

DGP_Grupos <- DGP_Grupos %>%
filter(`Ano Censo` == 2016) %>%
select(-`Ano Censo`)

DGP_Participantes <- DGP_Participantes %>%
filter(`Ano Censo` == 2016) %>%
select(-`Ano Censo`)

DGP_Parcerias <- DGP_Parcerias %>%
filter(`Ano Censo` == 2016) %>%
select(-`Ano Censo`)

DGP_Pesquisa <- DGP_Pesquisa %>%
filter(`Ano Censo` == 2016) %>%
select(-`Ano Censo`)

```

Com o filtro de ano aplicado podemos agora juntar os dados dos grupos de pesquisas, que estão armazenados em DGP_Grupos, com os outros, assim, aplicamos um left join com as produções, parcerias e com os participantes.

```

DGP_Producao_Grupos <- left_join( DGP_Producao, DGP_Grupos,
by = "Token Grupo Pesquisa")

DGP_Parcerias_Grupos <- left_join( DGP_Parcerias, DGP_Grupos,
by = "Token Grupo Pesquisa")

DGP_Participantes_Grupos <- left_join( DGP_Participantes, DGP_Grupos,
by = "Token Grupo Pesquisa")

```

4. Modelagem.

Nesta fase são selecionadas e aplicadas as técnicas de mineração de dados mais apropriadas, dependendo dos objetivos identificados na primeira fase. Ou seja, foi realizada a escolha, construção e execução do modelo de modelagem utilizado.

Assim, foi aplicado um filtro dos para trabalharmos somente com os grupos de pesquisa que possuem professores dos programas de pós graduação de Saúde coletiva e Ciências Médicas.

Para isso foi armazenado o token de pesquisa de cada professor

```

token_profCM <-DGP_Participantes %>%
filter(is.element(`Nome Participante`, nomes_id_tblCM$name)) %>%
select(`Token Grupo Pesquisa`)

token_profSC <-DGP_Participantes %>%
filter(is.element(`Nome Participante`, nomes_id_tblSC$name)) %>%
select(`Token Grupo Pesquisa`)

token_profPSC <-DGP_Participantes %>%
filter(is.element(`Nome Participante`, nomes_id_tblPSC$name)) %>%
select(`Token Grupo Pesquisa`)

```

Em seguida, foram selecionado os grupos em que os professores trabalharam

```

prof_producaoCM <- DGP_Producao_Grupos %>%
filter(is.element(`Token Grupo Pesquisa`,token_profCM$`Token Grupo Pesquisa`))

prof_producaoSC <- DGP_Producao_Grupos %>%
filter(is.element(`Token Grupo Pesquisa`,token_profSC$`Token Grupo Pesquisa`))

prof_producaoPSC <- DGP_Producao_Grupos %>%
filter(is.element(`Token Grupo Pesquisa`,token_profPSC$`Token Grupo Pesquisa`))

prof_parceriasCM <- DGP_Parcerias_Grupos %>%
filter(is.element(`Token Grupo Pesquisa`,token_profCM$`Token Grupo Pesquisa`))

prof_parceriasCM <- DGP_Parcerias_Grupos %>%
filter(is.element(`Token Grupo Pesquisa`,token_profCM$`Token Grupo Pesquisa`))

prof_parceriasCM <- DGP_Parcerias_Grupos %>%
filter(is.element(`Token Grupo Pesquisa`,token_profCM$`Token Grupo Pesquisa`))

```

5. Avaliação.

A quinta fase pede o acompanhamento dos resultados objetivos e a avaliação da aplicabilidade confiável dos insights e conhecimentos obtidos. Ocorrendo um processo de revisão e definição para os próximos passos, verificando se existem questões relacionadas à organização que não foram suficientemente abordadas. Deve-se refletir se o uso repetido do modelo criado pode trazer algum “efeito colateral” para a organização.

6. Implantação.

Finalmente, aplicamos o conhecimento obtido dos resultados do processo nas tomadas de decisão. Ou seja, se realiza o planejamento de implantação dos produtos desenvolvidos (scripts, no caso do executado nesta disciplina) para o ambiente operacional, para seu uso repetitivo, envolvendo atividades de monitoramento e manutenção do sistema (script) desenvolvido. A fase de implantação se dá como concluída com a produção e apresentação do relatório final com os resultados do projeto.