

Relatório 3 - Disciplina Ciência de Dados para Todos ou Ciência de Dados Aplicada

Mateus Hiro Nagata

31 de maio de 2018

Introdução e Contextualização

Introdução

Todo esse esforço para produção antropológica se baseia no fato de que a antropologia é uma área importante. A antropologia, de acordo com a universidade de Princeton é “Anthropology is the comprehensive study of human development, culture, and change throughout the world, past and present.” E ela se torna distinta na medida em que é extremamente intercultural e comparativa. Isso se deve ao fato de contrastar experiências familiares com outras de outros contextos sociais. Permitindo assim, eliminar teorias etnocêntricas ou viesadas apenas para o grupo cultural do sujeito que analisa. (PRINCETON, 2018)

Ainda, de acordo com a EASA(2015), a antropologia é a arte de “fazer familiar o que é exótico e fazer exótico o que é familiar”. Em termos técnicos, “Antropologia pode ser definida como o estudo comparativo dos humanos, de suas sociedades e de seus mundos culturais. Explorando simultaneamente o que eles têm em comum”

Contextualização

Qualidade

O presente trabalho tem como objetivo avaliar a situação docente dos professores e pesquisadores da UnB ativos em 2018 na área de antropologia.

O Departamento de Antropologia (DAN) na UnB abriga o Programa de Pós-Graduação em Antropologia Social. Curso que se destaca por receber a mais alta nota na avaliação CAPES de qualidade de curso universitário (DAN, 2018).

A expressividade na produção antropológica da UnB é tradição que vem da época em que Darcy Ribeiro, o primeiro reitor da universidade e doutor honoris causa dessa, fazia seus estudos aqui voltados ao entendimento da cultura indígena. A importância de Darcy se expressa na frase “A construção do campus brotou do cruzamento de mentes geniais. O inquieto antropólogo Darcy Ribeiro definiu as bases da instituição. O educador Anísio Teixeira planejou o modelo pedagógico. O arquiteto Oscar Niemeyer transformou as ideias em prédios.” (UnB, 2018)

A qualidade do ensino é justamente a importância dada a esta área do conhecimento. Como está escrito na sua descrição (site): “Praticamente a totalidade dos nossos estudantes de pós-graduação recebe financiamento para sua pesquisa sob a forma de bolsas de estudo e/ou verbas para realização de trabalho de campo e participação em eventos científicos.” Dessa forma, os recursos são o potencial dessa área de estudo na UnB. “O resultado dessa conjunção especial a unir qualidade de ensino e pesquisa, corpo docente altamente qualificado com elevada produção intelectual e infraestrutura física de boa qualidade só poderia ser um corpo discente bem formado, que tem recebido os principais prêmios científicos do país.”

Outro motivo de sua qualidade é pelo volume e atualização da sua biblioteca. Quantitativamente, " Além da aquisição regular de livros traduzidos e publicados originalmente no Brasil, entre 1976 e 1990 foram comprados um total de 1503 livros no exterior, sendo que no biênio 89/91 foram adquiridos 512 exemplares, portanto 33,2% do total" (DAN, 2018)

Um destaque dentre as produções do departamento é o “Anuário Antropológico”. “Seu corpo editorial e organizador é integrado por docentes do Departamento, que tem dele participado substantivamente, com artigos, ensaios e resenhas. O Anuário foi iniciado em 1976 tendo alcançado até o presente um total de 42 edições” (DAN, 2018)

Grupos de Pesquisa

Dentro da UnB existem diversos grupos de pesquisa atuando e produzindo conteúdo, que são (DAN, 2018): Antropologia da Ciência e da Técnica Estuda a ciência enquanto construção cosmológica, etnografando a prática científica. Pesquisa técnicas de grupos e atores sociais diversos, explorando a intersubjetividade das relações entre humanos e não humanos, como animais e artefatos.

Antropologia da Política, do Poder e do Direito Examina a política, o poder e a justiça de uma perspectiva etnográfica e comparada. Inclui estudos comparativos de processos de resolução de disputas das relações de poder e as dimensões sócio-culturais das estruturas hierárquicas.

Antropologia do Campesinato, dos Povos Tradicionais e do Meio Ambiente Conceitos e tradições teóricas no estudo de campesinato. Leitura de etnografias clássicas e recentes sobre o tema. As transformações contemporâneas do campesinato.

Antropologia do Corpo e da Saúde Investiga a construção social do corpo e da pessoa e dos processos de adoecer, tratar e curar e do manejo das condições de existência do humano. Privilegia as abordagens etnográficas e suas implicações teóricas para a compreensão desses fenômenos.

Antropologia Urbana, do Desenvolvimento e Globalização Estudos da cidade e na cidade; do campo do desenvolvimento, suas práticas e discursos; da circulação de pessoas, capitais, informações e coisas em diversas escalas; das relações locais/supra-locais e globais, do transnacionalismo e do cosmopolitismo.

Cultura Popular, Arte, Religião e Literatura Estudos comparativos dentro de uma perspectiva geral da antropologia simbólica de tradições populares, sistemas artísticos, tradições religiosas e textos literários e literatura oral.

Etnologia Indígena, Indigenismo e Relações Interétnicas Estudos etnográficos das sociedades indígenas do Brasil e da América do Sul. Sistemas regionais de interligações históricas, políticas e comerciais. Construção do indigenismo no Brasil e na América Latina.

Identities, Cidadania e Gênero Análise de construções das categorias de identidade, representações políticas, disputas por reconhecimento e direitos de cidadania, disposições e situações de classe, gênero, raça e etnia.

Integração e Conflito nas Sociedades Africanas e do Sudeste da Ásia Investiga processos de integração e conflito que atuam na configuração das sociedades nestas partes do mundo. A emergência de novas sociedades (crioulização), e os idiomas de expressão de conflitos são temas privilegiados.

Teoria e História da Antropologia Reflete sobre debates - rupturas e continuidades - na história do pensamento antropológico além de processos envolvendo controvérsias sobre práticas conceituais posteriormente estabelecidas e atualizadas na experiência etnográfica.

Grupos de Pesquisa

Existem também grupos de pesquisa especializados para diversos assuntos, nos quais os docentes da UnB exercem grande impacto na comunidade acadêmica, como (DAN, 2018):

Grupo de Pesquisa: Etnologia Indígena em Contextos Nacionais: Brasil-Austrália-Canadá Líder: Prof. Stephen Grant Baines Alunos da Graduação: Bernardo de Almeida Monteiro, Elisangela Moreira Menezes, Janaína Ferreira Fernandes, Leticia Maciel Bruno. Alunos da Pós-Graduação: Alessandro Roberto de Oliveira (DAN), Daniela Fernandes Alarcon (CEPPAC), Rodrigo Paranhos Faleiro (CEPPAC), Thiago Almeida Garcia (CEPPAC), Lucas Pereira das Neves Souza Lima (Ceppac), Raoni da Rosa (DAN) e Thaís Nogueira Brayner (DAN). Outros pesquisadores: Cristhian Teófilo da Silva, Ruben Caixeta de Queiroz, Gersem José dos Santos Luciano e Maxim Repetto.

Transformações no Mundo Contemporâneo: Territórios, Identidades, Ideologias e Coletividades Emergentes Líderes: Prof. Gustavo Lins Ribeiro e Kelly Cristiane da Silva Alunos da Graduação: Luciana de Almeida Pinto Coelho, Fernanda Silveira Anjos Alunos da Pós-Graduação: Adriana Coelho Saraiva, Elena Nava Morales, Rodrigo Augusto Lima de Medeiros, Yoko Nitahara Souza Outros pesquisadores: Flávia Lessa de Barros e Gonzalo Díaz Crovetto.

Urbanidade e Estilos de Vida Líder: Cristina Patriota de Moura e Mariza Veloso Motta Santos Alunos da Graduação: Bruno Cesar Medeiros Cassemiro, Paloma Karuza Maroni da Silva, Ranna Iara de Pinho Chaves Almeida. Alunos da Pós-Graduação: Herika Christina Amador Chagas Outros pesquisadores: Joaze Bernardino Costa, Frederico Flósculo Pinheiro Barreto, Letícia Costa Rodrigues Vianna, Mariza Veloso Motta Santos, Jesse Samba Samuel Wheeler.

Antropologia Política da Saúde Líder: Carla Costa Teixeira Alunos da Graduação: Ana Paula de Araújo Barbolosa, Bruno Calisto de Carvalho, Gustavo Leonel Ferreira Angei, Luís Cláudio Rocha Henriques de Moura, Natalia Heringer Mendonça, Samira Correia Dias Alunos da Pós-Graduação: Diogo Neves Pereira, Rosana Maria Nascimento Castro Silva Outros Pesquisadores: Amanda Rodrigues Faria, Cristina Dias da Silva, Flavia do Bonsucesso Teixeira, João Marcos Alem, Mariana Magalhães Pinto Côrtes, Soraya Resende Fleischer.

Etnografia Contemporânea: Antropologia dos Rituais, do Direito e da Política Líder: Mariza Gomes e S. Peirano e Luís Roberto Cardoso de Oliveira Pesquisadores: Antonádia Monteiro Borges, Christine de Alencar Chaves, Daniel Schroeter Simião, Kelly Cristiane da Silva, Patrice Schuch

Saberes e Ideologias Tradicionais Líderes: Ellen F. Woortmann e Klaas Axel A.W. Woortmann Alunos da Graduação: Mariana Oliveira Ramos Alunos de Pós-Graduação: Roberto Alves de Almeida, Tainá Bacellar Zaneti Outros Pesquisadores: Carlos Alexandre Barboza Plínio dos Santos, Aderval Costa Filho, Sueli Pereira de Castro, Renata Menasche, Santra Regina Lestinge.

Laboratório de Antropologia da Ciência e da Técnica Líderes: Profs. Guilherme José da Silva e Sá e Carlos Emanuel Sautchuk Alunos da Graduação: Guilherme Moura Fagundes Alunos da Pós-Graduação: Fausto dos Anjos Alvim e Rafael Antunes Almeida Outros Pesquisadores: Eduardo Viana Vargas, Sayonará de Amorim Gonçalves Leal

Processos de Invenção, Transposição e Subversão da Modernidade Líder: Profa. Kelly Cristiane da Silva e Daniel Schroeter Simião Alunos da Graduação: Larissa Caetano Mizutani, Alexandre Jorge de Medeiros Fernandes Alunos da Pós-Graduação: Anderson Silva Vieira

Etnologia e Indigenismo Líderes: Profs. Alcida Rita Ramos e José Antonio Vieira Pimenta Alunos da Graduação: Guilherme Moura Fagundes, Julia de Alencar Arcanjo e Juliane Bazzo Alunos da Pós-Graduação: Flavia Oliveira Serpa Gonçalves, Julio Cesar Borges, Nádia Philippsen Fürbringer, Ney José de Brito Maciel, Patrik Thames Franco. Outros Pesquisadores: Carmen Lucia da Silva, Giovana Acácia Tempesta, Karenina Vieira Andrade, Luis Abraham Cayon Duran, Marcia Leila de Castro Pereira, Maria Inês Smiljanic Borges, Roque de Barros Laraia, Silvia Maria Ferreira Guimarães.

Internacionalização

Quanto à internacionalização, não há muitos indicadores, mas nos resultados serão apresentados o quanto a produção bibliográfica da UnB é publicada no exterior.

Referencial

Tendo em vista que o presente artigo trata-se de uma pesquisa sobre como está sendo a pesquisa na instituição escolhida (UnB), ele pode ser classificado como um trabalho sobre logologia, ou seja, “a ciência das ciências” (WIKIPEDIA, 2018).

Recentemente, o advento de uma estrutura tecnológica de comunicação instantânea acarretou num grande influxo de informações que, caso não sejam devidamente organizados e analisados, tornam-se ruídos. A

condição tecnológica atual permite a geração de dados instantaneamente e num volume nunca antes produzido antes. Dessa forma, a aptidão de manusear esses dados de forma apropriada é condição fundamental para não se tornar obsoleto no meio de um ambiente interconectado e competitivo, onde informações são produzidas e disponibilizadas instantaneamente. O relatório da IBM (MILLER, 2017) traz previsões de grande crescimento nessa área de ciência de dados até 2020, como um aumento de demanda de emprego de 28% nessa área, comparado à 2015.

Essa tendência de aumento exponencial de informação indica a importância fundamental da manipulação dos dados. Como diria o New York Times: “dados são meramente a matéria-prima do conhecimento” (WHEELAN, 2016, p.11). Assim, com mais dados, é possível um maior progresso científico e técnico em todas as áreas do conhecimento.

A qualidade da informação, assim como a completeza é muito importante para chegar a conclusões científicas de como as coisas funcionam. “A falta de informações frequentemente leva à concorrência entre diferentes interpretações” (MLODINOW, 2009, p.8). Tendo os dados, é possível criar e provar teorias. Sem ela, por outro lado, não há como averiguar a veracidade de uma teoria. Nesse tocante, tudo indica que a disponibilidade de dados proporcionada pelo desenvolvimento tecnológico vai produzir mais avanços técnicos.

Tendo entendido a importância da ciência, é necessário ter em mente que este é um sistema que precisa de eventuais correções e monitoramentos. Isto é, o estudo da ciência da ciência é um tipo de trabalho que dá pouca retribuição direta, mas que quer tentar corrigir a cultura científica de seus erros (HU, 2016).

Metodologia

Para o presente trabalho foi utilizado a metodologia “CRISP-DM”, que é um acrônimo para “CRoss-Industry Standard Process for Data Mining”, uma metodologia de mineração de dados. Essa metodologia é de uso recorrente no mundo de mineração de dados, e muito de seu sucesso, de acordo com seus autores, é “CRISP-DM succeeds because it is soundly based on the practical, real-world experience of how people conduct data mining projects”. (NCR, 1999). Antes da descrição do método, é necessário uma compreensão do que é mineração de dados. De acordo com a empresa de business analytics SAS (2018): “Mineração de dados (em inglês, data mining) é o processo de encontrar anomalias, padrões e correlações em grandes conjuntos de dados para prever resultados”.

De forma geral, essa metodologia é dividida nos seguintes 6 passos em cadeia(NCR, 1999):

- 1- Business understanding. Essa etapa envolve determinar os objetivos da mineração de dados, averiguar a situação dos dados, determinar os objetivos da mineração e o plano de projeto de produção. Nesse ponto inicial, a formulação da pergunta é necessária para explorar com objetividade o trabalho que se segue.
- 2- Data understanding. Essa etapa consiste em uma coleta inicial dos dados, em seguida descrevê-lo, explorá-lo e verificar a qualidade dos dados.
- 3- Data preparation. Selecionar os dados a serem utilizados, limpá-los, construir dados, integrá-los e formatá-los em formato adequado. Feito isso, descreve-se o dataset como um todo
- 4- Modeling. Essa etapa consiste em selecionar as técnicas de modelagem adequadas, gerar um teste de design, construir um modelo e testá-lo.
- 5- Evaluation. Essa etapa consiste em avaliar os resultados, fazer uma revisão e determinar os próximos passos. Essa é uma etapa importante para decidir se todo o processo anterior foi adequado, comparando com os objetivos iniciais propostos no business understanding.
- 6- Deployment. Essa etapa consiste em checar os desdobramentos dos resultados obtidos. Faz-se isso com um plano de desdobramentos, em seguida um monitoramento e manutenção desse plano, depois um relatório final e por fim, uma revisão do projeto inteiro.

Outro aspecto importante é sua divisão hierárquica. Assim, cada atividade é pertencente a um nível de abstração que são (da mais geral para a mais específica): 1- Phases. Essas são as atividades genéricas de segundo nível; 2- Generic task. 3- Specialized task level. 4- Process instance.

No caso desse documento, nosso objetivo é tentar entender as características dos perfis dos docentes e do contexto em que está a antropologia na UnB.

O software estatístico R, em sua plataforma RStudio, foi utilizado para a manipulação dos dados disponíveis e, a partir de pacotes como dplyr, ggplot2, jsonlite, entre outros, foi possível obter os resultados descritos abaixo.

Resultados

Internacionalização

Primeiramente exploramos aqui a quantidade de publicações de capítulos de livros pelo departamento e também o país que este foi publicado. Essa é uma boa forma de averiguar o quanto a universidade tem internacionalizado e contribuído com a produção científica além do país. Como esperado, o Brasil é o país que se destina a maior parte das publicações, e pode se ver que há uma diversidade de países que receberam publicações da UnB, esses são pontuais e não se vê uma continuidade.

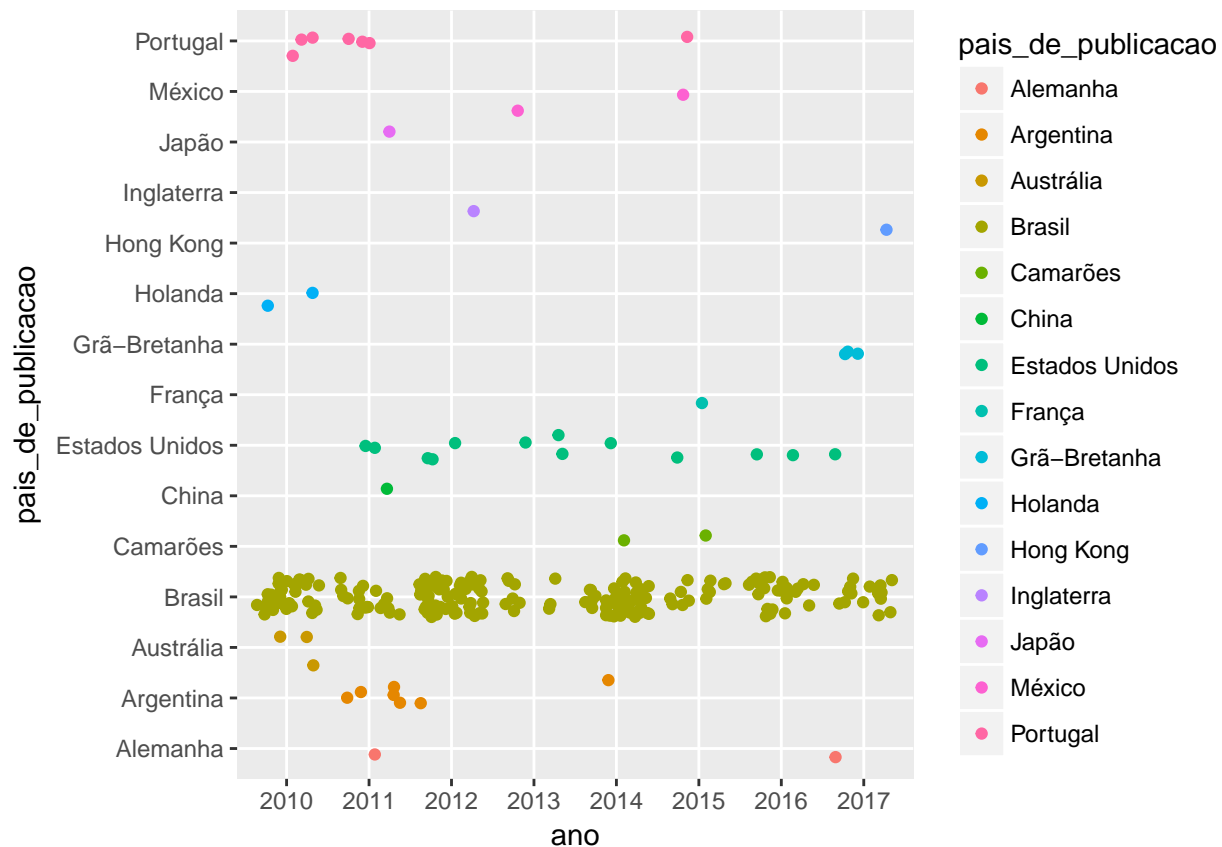
```
library(ggplot2)
library(jsonlite)
library(tidyverse)
library(readxl)
library(dplyr)
library(tidyr)

setwd("~/R/Data Science 4 All/Relatório 3")

antro <- fromJSON("Antropologia.profile.json")

PROD <- lapply(antro, function(x) x$producao_bibliografica)
PAIS <- lapply(PROD, function(x) x$CAPITULO_DE_LIVRO)
PAIS_bind <- bind_rows(PAIS[[1]], PAIS[[2]], PAIS [[3]], PAIS[[4]], PAIS [[5]], PAIS [[6]], PAIS[[7]], PAIS[[8]], PAIS[[9]], PAIS[[10]])

ggplot(PAIS_bind, aes(x = ano, y = pais_de_publicacao, col = pais_de_publicacao)) + geom_jitter()
```

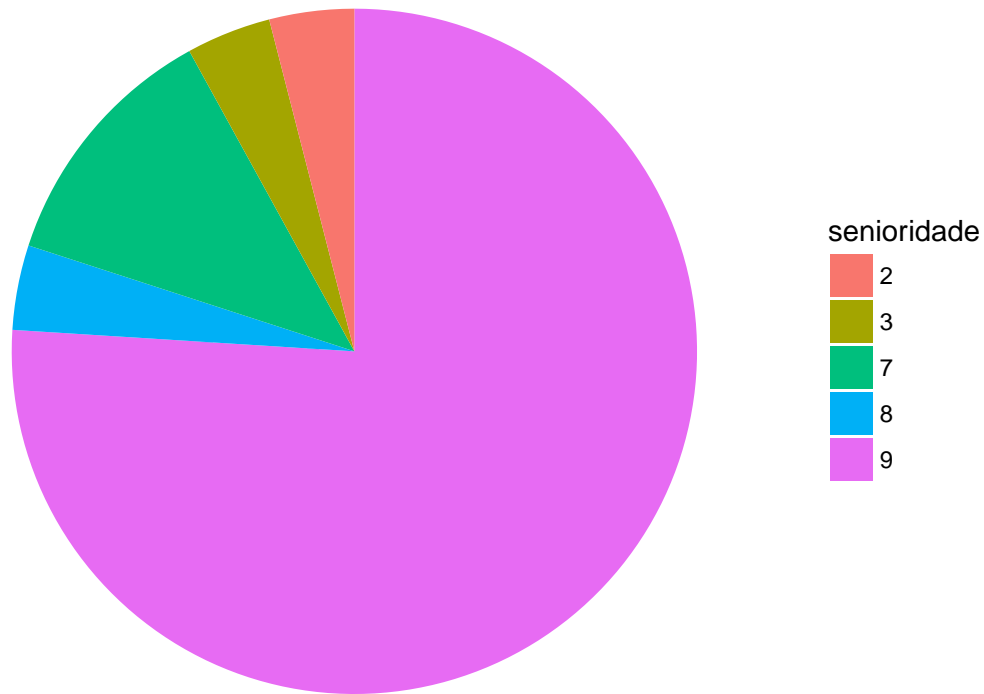


Senioridade

Nesse segundo momento, analisa-se a senioridade dos docentes desse departamento. O gráfico em forma de “pizza” mostra a distribuição de senioridade em relação ao total. Senioridade é uma variável que mede o tempo de contribuição do docente à instituição, então quanto maior, mais experiência e empenho institucional. Nos resultados percebe-se que muitos tem uma alta senioridade, indicando uma boa fidelidade e um corpo de docentes muito experiente.

```
# Nome de cada professor
nome <- sapply(antro, function(x) x$nome)
# Senioridade de cada professor
senioridade <- sapply(antro, function(x) x$senioridade)
# Interpreta senioridade como um fator
senioridade <- as.factor(senioridade)
# Cria um dataframe com nome e senioridade
df <- data.frame(nome, senioridade)
# Plota um histograma da senioridade

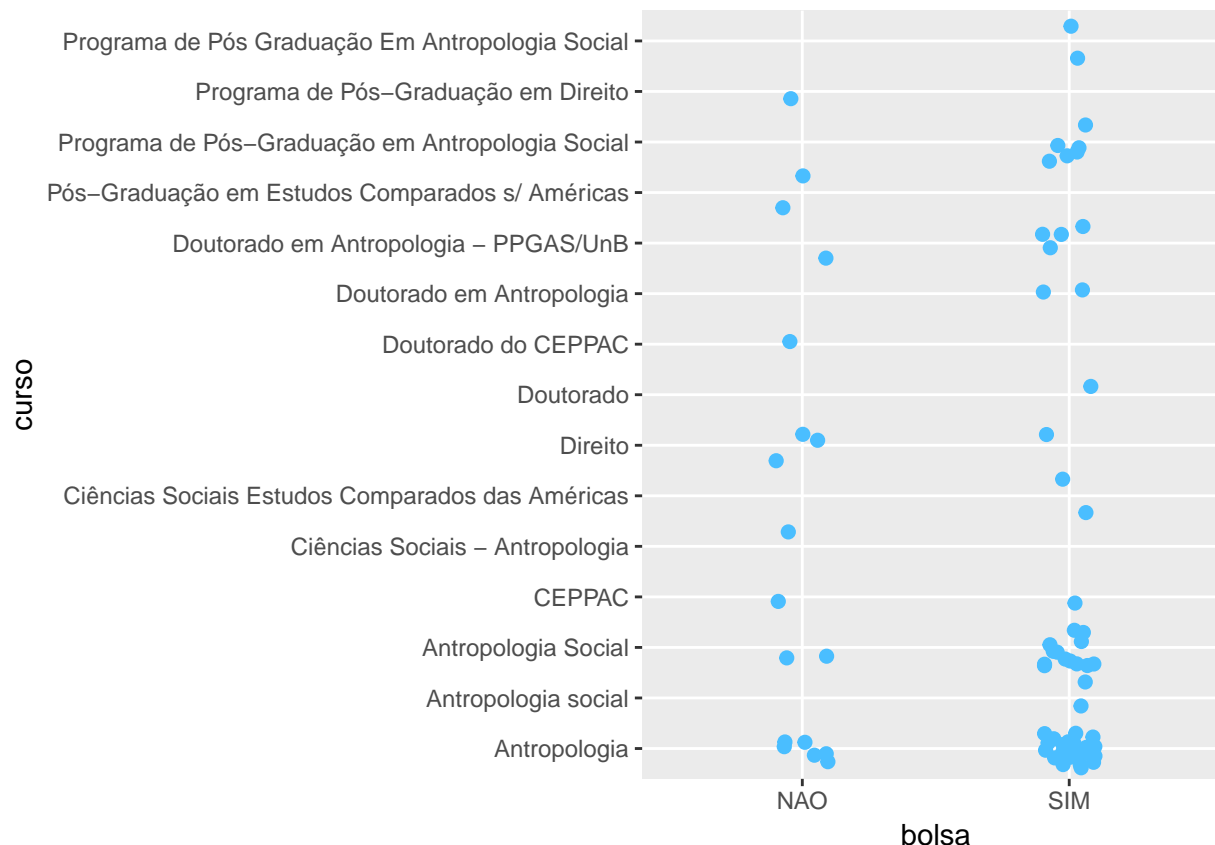
ggplot(df, aes(x = 1, fill = senioridade)) +
  geom_bar(position = "fill", width = 2) +
  coord_polar(theta = "y") + # Coordinates
  theme_void()
```



Bolsas

Por fim, explora-se a proporção de bolsas por orientação concluída de doutorado em relação ao curso feito. A ideia é tentar estimar se existe algum tipo de viés para a concessão de bolsas discriminado por curso. Comparando, por exemplo, “Programa de Pós Graduação em Antropologia Social”, que todos tiveram bolsa, e “Pós-Graduação em Estudos Comparados s/ Américas”, que nenhum teve bolsa, pode indicar que talvez exista esse efeito, mas é necessário uma pesquisa mais aprofundada com um espaço amostral maior para tirar qualquer conclusão.

```
my_color <- "#4ABEFF"
ORI <- lapply(antro, function (x) return (x$"orientacoes_academicas"))
ORI_CONCL <- lapply(ORI, function (x) return (x$"ORIENTACAO_CONCLUIDA_DOUTORADO"))
Lolo <- bind_rows(ORI_CONCL[[2]], ORI_CONCL [[3]], ORI_CONCL[[4]], ORI_CONCL [[5]], ORI_CONCL [[6]], ORI_CONCL [[7]])
ggplot(Lolo, aes(x = bolsa, y = curso)) + geom_jitter(width = 0.1, size = 2, shape = 21, color = my_color)
```



Referência Bibliográfica

PRINCETON. Why Study Anthropology?. **Princeton University**, 2018. Disponível em: <https://anthropology.princeton.edu/undergraduate/why-study-anthropology>. Acesso em: 31 mai. 2018

EASA. Why anthropology matters. **European Association of Social Anthropologists**, 2018. Disponível em: https://www.easaonline.org/publications/policy/why_en.shtml. Acesso em: 31 mai. 2018

DAN. Corpo Docente. **Departamento de Antropologia da UnB**. Disponível em: <http://www.dan.unb.br/dan-corpodocente>. Acesso em: 31 mai. 2018

DAN. História do DAN/UnB. **Departamento de Antropologia da UnB**. Disponível em: <http://www.dan.unb.br/dan-historia>. Acesso em: 31 mai. 2018 UnB.

UNB. História. **UnB**. Disponível em: <https://www.unb.br/a-unb/historia>. Acesso em: 31 mai. 2018

WIKIPEDIA. Logology. **Wikipedia**, 2018. Disponível em: [https://en.wikipedia.org/wiki/Logology_\(science_of_science\)](https://en.wikipedia.org/wiki/Logology_(science_of_science))

MILLER, S., & Hughes, D. (2017). The Quant Crunch: How the demand for data science skills is disrupting the job market.

WHEELAN, C. (2016). Estatística: o que é, para que serve, como funciona. Zahar.

MLODINOW, L. (2009). O andar do bêbado. Rio de Janeiro: Jorge Zahar.

HU, J. Studying the science of science. **Science**, 2016. Disponível em: <http://www.sciencemag.org/careers/2016/03/studying-science-science>

NCR, P. C., Clinton, J., Ncr, R. K., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (1999). CRISP-DM 1.0.

SAS. Mineração de Dados. **SAS**, 2018. Disponível em: https://www.sas.com/pt_br/insights/analytics/mineracao-de-dados.html