

# Avaliação dos Programas de pós-graduação de Ciências Biológicas I e III e dos currículos dos professores / pesquisadores cadastrados

*Emanuel Victor A. Coelho 14/0137475, Jonas Prado 13/0117277, Maria Júlia G. R. Rodrigues 14/0153713 - Grupo 10*

*05/12/2018*

## Resumo

Este documento apresenta o relatório final da disciplina de tópicos Avançados em computadores - Turma D - 2018.2 do departamento de Ciência da Computação da universidade de Brasília, que trata da análise da produção científica e acadêmica na Universidade de Brasília, na área de Ciências Biológicas. Para isso, foi aplicado o modelo metodológico de mineração de dados CRISP-DM e através dele foram dispostas diversas informações sobre a evolução destes programas de pós-graduação entre os anos de 2010 e 2017 (Fernandes et. al., 2018).

## Introdução

Este trabalho visa, no contexto da pós-graduação do sentido amplo e restrito do Brasil, buscar relações e comparações entre os Programas de Pós-Graduação em Biologia Molecular, Biologia Animal, Biologia Microbiana e Patologia Molecular da Universidade de Brasília para sintetizar o panorama e o paradigma da pós-graduação nestes programas.

## CRISP-DM:

A metodologia para desenvolvimento do relatório é baseada no modelo de mineração de dados denominado CRISP-DM (Chapman et al., 2000, Mariscal et al., 2010). Este modelo é caracterizado por um projeto dividido em seis fases, as quais serão tratadas ao longo das seções seguintes.

- Fase 1) Entendimento do negócio: Busca definir um problema de mineração de dados com base nos objetivos e necessidades do projeto.
- Fase 2) Entendimento dos dados: Consiste na coleta e descrição dos dados, incluindo uma análise de qualidade e quantidade dos mesmos - necessários para avaliar a viabilidade do projeto.
- Fase 3) Preparação dos dados: Consiste na estruturação dos dados capturados, incluindo uma fase de limpeza de dados indesejados ou inconsistentes. Esta fase tem como objetivo produzir datasets prontos para análises estatísticas e produção de gráficos.
- Fase 4) Modelagem: Envolve a construção e revisão de modelos estatísticos de interesse para o projeto, além de testes para descarte dos modelos produzidos que não atingirem o grau desejado de confiabilidade.
- Fase 5) Avaliação: Consiste na avaliação dos resultados e do processo como um todo, revisando se todas as questões relevantes para o projeto foram abordadas adequadamente.
- Fase 6) Implementação: Consiste no planejamento e implementação dos entregáveis finais desenvolvidos pelo projeto, incluindo monitoramento e manutenção dos mesmos.

## **CRISP-DM Fase 1 - Entendimento do Negócio**

### **O que é o Sistema Nacional de Pós-Graduação?**

O Programa de Pós-Graduação é o conjugado formado pelos cursos de mestrado e/ou doutorado acadêmicos ou de mestrado profissional de uma Instituição de Ensino Superior - IES atuante numa mesma área do conhecimento (sua área básica) que partilha essencialmente o mesmo corpo docente e tem uma estrutura administrativa comum. Programas são divididos em acadêmicos ou profissionais, de acordo com os cursos (Leite, 2018).

O sistema nacional de pós-graduação engloba todo programa de pós graduação no Brasil e o CAPES e o responsável pelo processo de acompanhamento e avaliação da qualidade dos programas de pós-graduação. Cabe à CAPES constituir as normas específicas para a sistematização de avaliação, delineando as etapas, critérios e ferramentas utilizadas no processo (Silveira e Neto, 2016).

O sistema nacional de pós-graduação tem vários objetivos dentre eles: fortalecer as bases científica e tecnológicas de inovação; formação adequada de docentes para todos os níveis de ensino; formação de quadros para mercados não acadêmicos; diminuir as desigualdades acadêmicas entre as regiões do país; preservação do sistema nacional de avaliação de qualidade da pós-graduação brasileira, como um sistema de certificação e referência para a distribuição de bolsas e recursos para o fomento à pesquisa; avaliação da produção tecnológica e seu impacto e relevância para o setor econômico; incentivo à inovação; diversificar o sistema de avaliação de forma a permitir a análise de diferentes modelos de pós-graduação; introduzir procedimentos que contemplem aspectos qualitativos dos programas de pós-graduação; estimular à pós-graduação, apresentando constantes atualizações dos indicadores qualitativos e quantitativos, orientando a formação de recursos humanos e a pesquisa na direção das fronteiras do conhecimento e das precedências estratégicas do País (Nunes, 2009).

### **Os Colégios, Grandes áreas e áreas da Pós-Graduação Brasileira**

A partir de 2018, as diversas áreas da pós-graduação brasileira foram organizadas na forma de colégios, grandes áreas e áreas. Os colégios foram separados da seguinte forma:

- Colégio de Ciências da Vida - que contempla as grandes áreas de Ciências Agrárias, Ciências Biológicas e Ciências da Saúde.
- Colégio de Ciências Exatas, Tecnológicas e Multidisciplinar - que contempla as grandes áreas de Ciências Exatas e da Terra, Engenharias e Multidisciplinar.
- Colégio de Humanidades - que contempla as grandes áreas de Ciências Humanas, Ciências Sociais Aplicadas e Linguística, Letras e Artes.

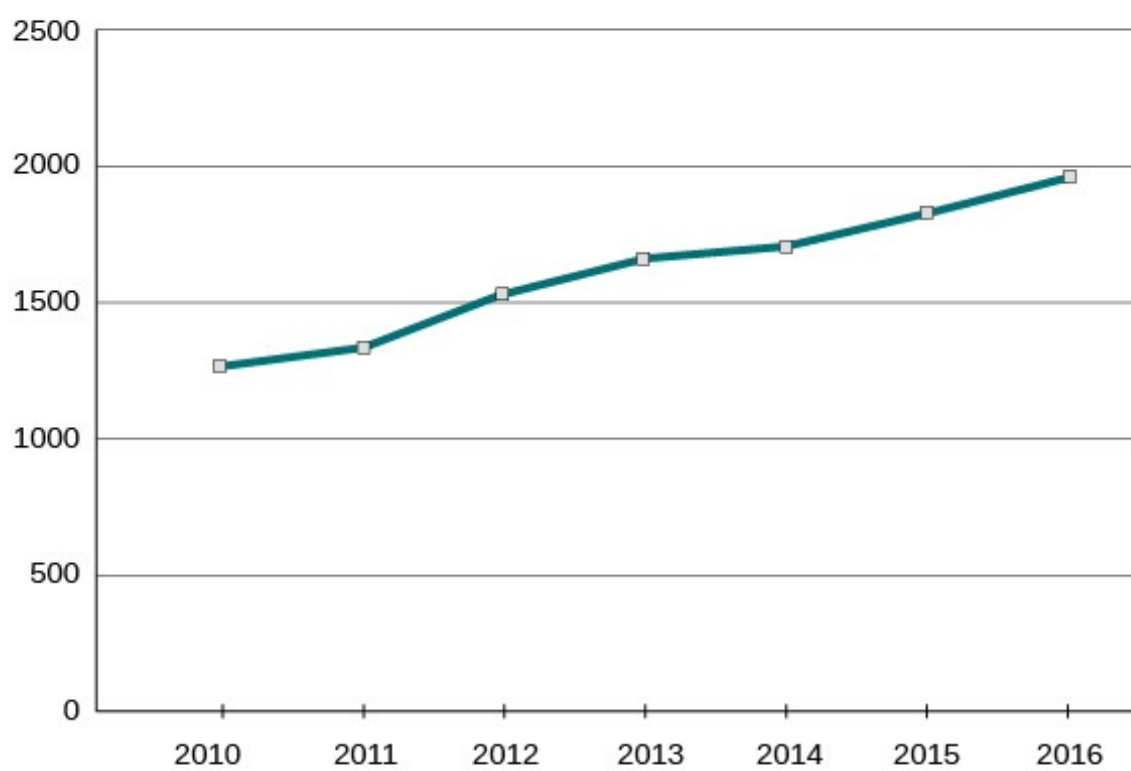
Cada um dos colégios, grandes áreas e áreas de conhecimento possui suas particularidades e não podem ser avaliadas de forma igual. Cada qual deve ser analisada e avaliada conforme o documento de área disponível em Capes (2018). (Fernandes et. al., 2018).

### **A UnB dentro do Sistema Nacional de Pós-Graduação**

A produção científica na UnB é umas das mais importantes do país. De acordo com o Plano de Internacionalização da Universidade de Brasília a produção aumentou em aspecto de qualidade e quantidade no período de 2011 a 2016. (SciVal, 2018).

Um impacto normalizado de 1,00 significa que a produção teve comportamento similar à média global. Um impacto superior a 1,00 indica maior citação que a média (por exemplo, um impacto de 1,50 indica 50% a mais de citação) enquanto um impacto inferior a 1,00 indica citação inferior à média (SeCom, 2018, Adaptado).

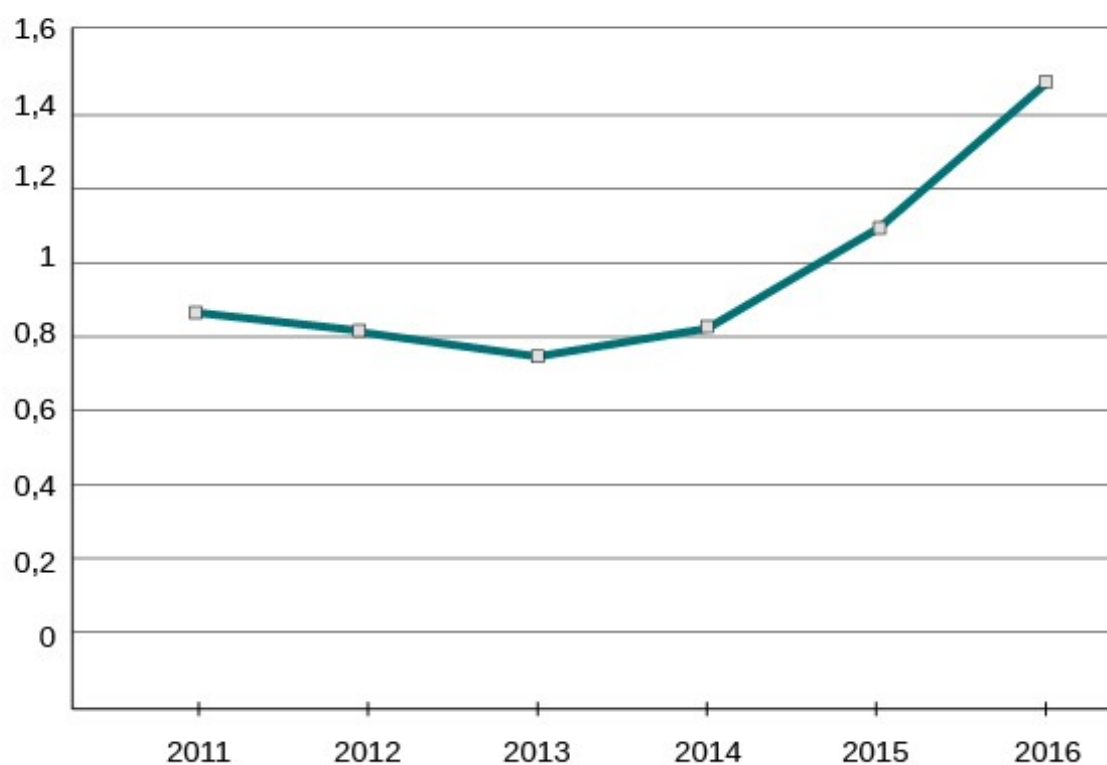
## Produção científica da UnB de 2010 a 2016



Fonte: SciVal - 2018

Figure 1: Produção científica.

## Impacto normalizado por citação da produção científica da UnB de 2010 a 2016<sup>[1]</sup>



Fonte: SciVal - 2018

<sup>[1]</sup> Um impacto normalizado de 1,00 significa que a produção teve comportamento similar à média global. Um impacto superior a 1,00 indica maior citação que a média (por exemplo, um impacto de 1,50 indica 50% a mais de citação) enquanto um impacto inferior a 1,00 indica citação inferior à média (um impacto de 0,91 indica um 9% a menos de citação que a média).

Figure 2: Impacto da produção.

Em 2011, o impacto normalizado por citação foi, de acordo com o gráfico, de aproximadamente 0.85, o que indica que os pesquisadores da UnB foram 15% menos citados do que a média global. Similarmente em 2013, houve uma redução branda no número de pesquisadores citados. Entretanto, de 2013 até 2016 houve um aumento considerável nesse impacto normalizado - os pesquisadores da UnB atingiram índice de impacto por citação de 1.45 em 2016, sendo 45% mais citados que a média global (SeCom, 2018).

Esse aumento na produção científica está diretamente associada ao aumento no número de docentes da Universidade. Entre os anos de 2009 e 2013, 577 professores ingressaram no quadro permanente da instituição, com um aumento de 32%. Entre 2013 e 2016, este aumento foi de 185 docentes (8%). (...) Com um quadro de professores maior espera-se naturalmente uma maior produção de artigos e outras matérias científicas pela UnB (SeCom, 2018).

## **Contextualização dos Programas**

### **Ciências Biológicas I**

A forma de avaliação da Capes para os programas de Pós-Graduação em Biologia Molecular e Biologia Animal da Universidade de Brasília é determinada pela área 06 que se trata sobre ciências biológicas com atuação nas áreas de Genética, Biologia Geral, Biologia Molecular, Biologia Celular, Biologia do Desenvolvimento, Bioinformática e Biologia de Sistemas.

Há pouco tempo à área de ciências biológicas I passou por mudanças no arranjo dos Programas de Pós-Graduação (PPG) componentes para atender padrões de qualificação e a especialização das Áreas. Isso demonstra algo muito bom, pois a CAPES tem formado novas Áreas que atendem melhor aos critérios característicos de avaliação, resultando em Áreas de Avaliação padronizadas e admitindo maior integração entre os PPG. (Capes, 2016).

A interdisciplinaridade é um fator importante e que caracteriza bem a área de ciências biológicas I. Os programas de pós-graduação que pertencem essa área são compostos por docentes com distintas formações profissionais principalmente em áreas com relação direta com a área de biológicas. O conjunto formado pela união desses profissionais reflete no progresso e desenvolvimento de extrema importância nessas áreas. Este lado multidisciplinar é importante e necessário ser mantido para se compreender e desenvolver melhor técnicas, pesquisas, metodologias, etc. A união destas áreas vem desenvolvendo os núcleos de pesquisa em pós-graduação de maneira muito positiva. (Capes, 2016).

Os referentes de maior peso da Ficha de Avaliação dos programas de pós-graduação na área de ciências biológicas I constituem-se dos tópicos: Corpo Docente, Corpo Discente, Teses, Dissertações e Produção Intelectual. De maneira que se busque o desenvolvimento acadêmico e tecnológico dos programas avaliados e da área de maneira geral. Todo o processo de avaliação dos programas de pós-graduação leva em consideração as atividades do conjugado de Docentes e Discentes. (Capes, 2016).

Abaixo apresenta-se a forma de avaliação e cada peso dos respectivos quesitos avaliados na área de Ciências Biológicas I.

As propostas do programa avaliado devem deixar claro e de forma objetiva os fins, as áreas de concentração e as linhas de pesquisa. As áreas de concentração devem refletir as áreas de especialização do seu corpo docente e refletir sobre as projeções futuras do programa. Os programas devem ter uma base concreta em seu núcleo de docentes permanentes para se garantir o total desenvolvimento das atividades de ensino, pesquisa e orientação do programa (Capes, 2016).

Os Programas devem debater e promover o oferecimento de disciplinas com conteúdo programático relacionados com conhecimento das linhas de pesquisa oferecidas pelo programa e de áreas relacionadas que podem promover um desenvolvimento maior. Linhas de pesquisa não atualizadas devem ser reavaliadas, como incentivo à formação interdisciplinar e multidisciplinar do discente e isto que torna a relação entre os programas possível (Capes, 2016).

<i>Quesitos</i>	<i>Pesos para Mestrado Acadêmico e Doutorado</i>	<i>Pesos para Mestrados Profissionais</i>
1.Proposta do Programa	-	-
2.Corpo Docente	20%	20%
3.Corpo Discente, Teses e Dissertações	35%	30%
4.Produção Intelectual	35%	30%
5.Inserção Social	10%	20%

Fonte: (Capes, 2016).]

Figure 3: Critérios de Avaliação em Biológicas I.

## **\*\* Biologia Animal**

Os programas de pós-graduação da CB I foram norteados, no Coleta 2016, a reavaliar e sintetizar as propostas do programa em pós-graduação em seus vários aspectos, descrevendo as alterações que aconteceram e os efeitos muito significativos que afetaram sua qualidade. Os programas de pós-graduação ainda descreveram mudanças no seu Corpo Docente, nas Disciplinas, além das diretrizes e discernimentos das deliberações adotadas. As propostas dos Programas demonstraram a assiduidade e também extensão da grande amplitude de Linhas de Pesquisa e Multidisciplinaridade particularidades da CB I, englobando as seguintes áreas: Genética (Humana, Animal, Vegetal e de Microrganismos); Biologia Geral (Comparada, Estrutural, Funcional, Toxicologia); Biologia Molecular; Biologia Celular; Biologia do Desenvolvimento; Bioinformática e Biologia de Sistemas. A área de Bioinformática mostrou ampliação na maioria dos programas de pós-graduação acontecendo um impacto destaque do edital de Biologia Computacional. Todas as propostas foram avaliadas adequadas. Contudo, em alguns programas de pós-graduação, houve precisão de readequações que foram assinaladas nos pareceres específicos (Capes, 2017).

As propostas do Programa são bem formuladas e descrevem de maneira satisfatória as suas particularidades mais importantes. Os objetivos expostos são claros e coerentes com o perfil acadêmico dos programas. A análise minudenciada da Proposta, em relação aos programas da área CBI, corroborou que o Programa avaliado apresentara desempenhos considerados bons. A observação delineada do perfil, adequação, distribuição de atividades e dedicação dos docentes no quadriênio, em relação aos programas da área CBI, demonstrou que o Programa analisados ofereceram desempenho positivo (Capes, 2017).

A importância da qualidade da formação acadêmica e da produção científica são objetivos prioritários das ações da CB I. Esse atributo pode ser aferido por meio de indicadores que vêm sendo aprimorados no conjunto internacional, e também no sistema CAPES de avaliação. O enfoque fundamental da produção científica na CB I é artigo em periódico. Essa produção cogita a qualidade da área que tem consideração internacional. (Capes, 2017).

“Os limites dos estratos do QUALIS têm sido crescentes na última década e neste quadriênio atingiram patamares elevados, comparáveis à produção qualificada de pesquisadores internacionais. Há ainda Programas que não atingiram os patamares de produção científica da área, além da distribuição heterogênea da produção entre os docentes do NP. Isso é reflexo natural do estágio de desenvolvimento e consolidação dos PPGs.” (Capes, 2017).

A análise particularizada do desenvolvimento científico e técnica do NP no quadriênio e sua repartição entre os docentes em relação aos programas da área CBI corroborou que os Programas analisados proporcionaram

desempenho avaliado positivamente (Capes, 2017).

Esse programa apresenta relevância regional, nacional e/ou internacional evidenciada por convênios, parcerias interinstitucionais. Nele notam-se formidáveis impactos tecnológicos, com o desenvolvimento de novas técnicas, produtos e metodologias. Os impactos em nível regional têm acontecido por meio das ações de extensão, com destaque para projetos na área da educação. O programa tem destacado esforços na exposição da ciência ao público leigo. O programa também buscado integrar discentes do ensino médio as pesquisas estimulando-os e fornecendo bolsas. Ele possui relevância regional e/ou nacional, adotando como base as pesquisas desenvolvidas pelos corpos docente e discente. A proposta destaca o papel do Programa na formação de recursos humanos especializados com potencial para desempenhar funções em órgãos públicos e privados (Capes, 2017).

## **\*\* Biologia Molecular**

No levantamento de 2017 o número de discentes para o Programa de Pós-Graduação em Biologia Molecular foi de 4.781 alunos sendo 2.863 de Mestrado e 1.918 de Doutorado; dos docentes, 27 são permanentes e 11 são colaboradores. As linhas de pesquisa seguem vertentes definidas: Genética (Humana, Animal, Vegetal e de Microrganismos); Biologia Geral (Comparada, Estrutural, Funcional, Toxicologia), que pleiteia áreas descritivas e experimentais de forma geral; Biologia Molecular, que estuda experimentalmente e teoricamente processos biológicos a nível molecular e atômico; Biologia Celular, que estuda o funcionamento da célula como um todo, ou de cada um de seus aparatos; Biologia do Desenvolvimento, que estuda a formação de organismos, dos mais simples aos complexos; Bioinformática, que delinea e gera modelos para padrões de expressão ou estuda dinâmicas moleculares; Biologia de Sistemas, que estuda a interação e a formação dos sistemas para organismos vivos. Os projetos de pesquisa do programa seguem diversas vertentes, desde a neuropsicológica até a física da interação de moléculas. Dos diversos projetos, há o enfoque em bioinformática que foca em simulações e modelagens e em biologia molecular, que estuda experimentalmente a interação entre biomoléculas.

A última avaliação quadrienal fomenta a progressão do PPG em Biologia Molecular, frisando as modificações que ocorreram no corpo docente, que alterou as tomadas de decisões e impactos na produção científica - positivamente - em relação a avaliação anterior. A avaliação mostra em seu resultado que os recursos obtidos pelos os docentes foram oriunda de procura a agencias de fomento à pesquisa. O conjunto de disciplinas apresentado, segundo a avaliação, é de acordo com a proposta, tangenciando o conhecimento necessário para a formação acadêmica do indivíduo de forma abrangente e consistente. A avaliação também frisa a presença de Double-Degrees com universidades estado unidenses e doutorados sanduíche com outras diversas instituições internacionais.

Quanto ao corpo docente em geral, o número destes se manteve significativamente constante em relação ao triênio anterior, tanto na quantidade de permanentes quanto a de associados. Não houve mudança significativa na composição destes, nem na abrangência de áreas - o que não desqualificou estes docentes quanto a sua adequação a proposta do programa. A avaliação do corpo docente fomentou que o número de bolsistas de produtividade diminuiu em relação a avaliação do triênio anterior. A proporção ode orientados em mestrado e doutorado é de aproximadamente um para um, onde o número mediano de orientados de mestrado e doutorado para cada orientador é o mesmo.

A avaliação dos discentes demonstrou que houve um aumento na matrícula e formação de doutores e a diminuição dos mesmos para mestrado em relação ao triênio anterior. A maioria dos professores orienta menos que doze alunos, sendo que um quinto dos docentes orienta somente um aluno. Alunos que publicam com em estratos de revistas de A1 a B1 tem em ao menos metade das vezes, a participação de um professor conjuntamente. Houve também discentes ganhadores de prêmios dentro desta PPG.

A avaliação da produção intelectual evidenciou claramente a melhoria da qualidade do conteúdo publicado pelos docentes e discentes do PPG como um todo. Onde cerca de três quartos de todos os artigos publicados por participantes desta PPG tiveram avaliação Qualis superior a B2. Toda esta produção de conhecimento resulta na valorização internacional desta área de conhecimento de de seus participantes.

O impacto social mais evidente avaliado foi a criação de novas ferramentas e métodos que auxiliam na mudança de paradigma do cientista da área, na qual, esta PPG vem contribuindo incessantemente - de acordo com a avaliação. É exaltada a participação de todos os docentes deste programa em intercâmbios para congressos, pesquisas e afins, que determina a importância regional, nacional e internacional que esta PPG tem sobre sua área de atuação, cooperando de forma eficaz com meios nacionais e internacionais a estar impelida na divulgação do conhecimento científico da área.

### Ciências Biológicas III

O programa de Pós-Graduação em Ciências Biológicas III compreende as seguintes áreas: Microbiologia, Imunologia e Parasitologia. Ela é considerada uma das mais tradicionais no Sistema Nacional de Pós-Graduação (SNPG), pois é desenvolvida a mais de 50 anos. Um fato importante sobre esta área é que ela aumenta de forma branda, pelo fato de apresentar uma dura avaliação das propostas dos cursos novos. Um desses quesitos é que haja uma semelhança entre as atividades propostas pelos cursos novos com os que já estão na área. Isso fez com que demorasse por volta de 15 anos para dobrar o número de programas, que passou de 18 para 36 programas em 2016 (Capes, 2016).

Essa área é muito voltada, desde o século XIX, para a proposta de soluções para epidemias e pandemias. Continuamente, no século passado o Brasil já apresentava uma qualidade e quantidade de produções científicas para solução de doenças tropicais principalmente nas áreas de Patologia e doenças Infecciosas e Parasitárias. Além disso, a área apresenta ótima visibilidade no cenário internacional, onde há projetos e publicações interdisciplinares que ajudam na expansão das pesquisas e tem um impacto muito bom na produção intelectual, visto que, essa interdisciplinaridade é muito importante para entender as metodologias para o progresso da área (Capes, 2016).

Os referentes de maior peso da Ficha de Avaliação dos programas de Pós-Graduação na área de Ciências Biológicas III constituem-se dos tópicos: Corpo Docente, Corpo Discente, Teses, Dissertações e Produção Intelectual. De maneira que se busque o desenvolvimento acadêmico e tecnológico dos programas avaliados e da área de maneira geral. Todo o processo de avaliação dos programas de pós-graduação leva em consideração as atividades do conjugado de Docentes e Discentes (Capes, 2016).

Abaixo apresenta-se a forma de avaliação e cada peso dos respectivos quesitos avaliados na área de Ciências Biológicas III para os programas de mestrado acadêmico, doutorado e mestrado profissional.

<i><b>Quesitos</b></i>	<i><b>Pesos para Mestrado Acadêmico e Doutorado</b></i>	<i><b>Pesos para Mestrados Profissionais</b></i>
<b>1.Proposta do Programa</b>	-	-
<b>2.Corpo Docente</b>	20%	20%
<b>3.Corpo Discente, Teses e Dissertações</b>	30%	30%
<b>4.Produção Intelectual</b>	40%	30%
<b>5.Inserção Social</b>	10%	20%

**Fonte: Capes, 2016.**

Figure 4: Critérios de Avaliação em Biológicas III.



## **\*\* Biologia Microbiana**

O programa de pós-graduação (mestrado e doutorado) em Biologia Microbiana pertence a área de Ciências Biológicas I. Este programa foi criado em 2012 na Universidade de Brasília. Suas áreas de pesquisa são – Biologia dos Patógenos, Microbiologia Ambiental e Microbiologia Aplicada – com um foco maior para a linha de Biologia dos Patógenos, apresentando 23 projetos dos 46 desta área. O corpo docente conta com 18 professores permanentes e 3 colaboradores, com diversos graus de experiência. Todos eles são doutores e possuem formações dentro e fora do Brasil. Suas formações são compatíveis com a área em que atuam e suas linhas de pesquisa. Entretanto, somente 11 docentes permanentes orientaram dissertações e teses nesse quadriênio, o que é ruim visto que a área possui 19 docentes permanentes. Há 11 docentes que são pesquisadores de produtividade do CNPq, sendo que 7 deles contem bolsa nível de 1 e 4 deles contem bolsa nível 2. As disciplinas ofertadas pelos discentes possuem um aspecto de diversidade e multidisciplinaridade que são características desta área de formação (Capes, 2017).

No que se refere a quantidade de formandos no quadriênio em questão, formou-se 20 alunos do mestrado e 4 do doutorado e foi produzido uma média de 20 artigos com discentes. Esses números são considerados baixos em relação aos outros programas, visto que, é um curso novo de pós-graduação na universidade. Além disso, há uma porcentagem de 42,5% de mestrandos formados em relação ao número de mestrandos matriculados, o que é considerado muito bom. A perspectiva de formação para esses docentes é muito boa, pois eles já são incorporados ao mercado de trabalho ou ingressam diretamente no doutorado (Capes, 2017). “Com a recente criação do curso de doutorado, este programa terá um impacto regional importante na formação de recursos humanos de alta qualidade que poderão atuar como docentes, pesquisadores ou profissionais da área de microbiologia.” (Capes, 2017).

A produção científica em periódicos ficou com uma quantidade (193) e qualidade muito boas, sendo a produção, entorno de, 39 extratos A1+A2 e 81 nos extratos A1+A2+B1. É importante frisar que a produção de artigos é produzida em sua maior parte pelos docentes permanentes, porem a produção feita pelos colaboradores não fica muito atrás e possui um peso considerável na produção do programa, apresentando 40 de um total de 233 artigos. A produção de capítulo de livros dos discentes e participação em patente é boa, apontando para 33 capítulos de livros e uma participação em patente feita (Capes, 2017).

## **\*\* Patologia Molecular**

O Programa de Pós-Graduação da Universidade de Brasília em Patologia Molecular, tem perspectivas de desenvolvimento incluindo a capacitação dos docentes, dos discentes e colaboradores internacionais. Esse programa tem uma linha multidisciplinar ligado à área da saúde. Eles vêm implementando medidas para aumentar as exigências na produção científica para contratação de orientadores e possuem uma boa infraestrutura para garantir o desenvolvimento do programa. O Programa contém 3 áreas de conhecimento que contou com cerca de 20 projetos no período. A maioria dos projetos foram financiados e possuem quase a totalidade de docentes permanentes em sua elaboração. Sendo 11 deles da área de Biologia Molecular e desses 11, 3 deles são pertencentes também ao Programa de Biotecnologia e Biodiversidade (Capes, 2017).

Em relação ao corpo docente o programa apresentou por volta de 20 docentes permanentes e 5 docentes colaboradores. Todos eles possuem doutorado e pertencem a diversas áreas do conhecimento, como: Morfologia, Bioquímica, Medicina, Genética, entre outros. E possuem formação em instituições nacionais e internacionais. Cerca de 76% deles são bolsistas produtividade CNPq, que são distribuídos nos níveis 1A até 1D, nível 2 e bolsa DTI2. Já o corpo discente contou com 37 mestre e 35 doutores formados, que foram orientados por 17 docentes permanentes da área. Somente 3 discentes colaborativos não tiveram orientados no período, mas apresentam formações em andamento. A produção em artigos com Qualis A1-B1 foi de 128, sendo 53% dos artigos em Qualis A1,A2 e B1 e 40% em artigos Qualis B2 a B5. O tempo de formação dos discentes também foi muito boa em relação ao período anterior, apresentando uma redução no tempo médio de titulação (Capes, 2017).

A produção intelectual dos docentes permanentes neste período foi de 266 artigos, sendo 66% deles em Qualis A1, A2 e B1 e 34% deles em Qualis B2 a B5. Em comparação com o quadriênio anterior houve uma melhora

considerável na produção do programa e a pontuação por docente ficou em torno de 697 pontos. Os capítulos de livros tiveram uma pontuação de 315, o que é considerado muito bom para o programa. Destaca-se que todos os docentes permanentes possuem publicações qualificadas no período em questão e que 8 patentes da área ganharam 510 pontos (Capes, 2017).

Já ao que se refere a inserção social o programa tem um bom número de doutores deste programa participando dos cursos de graduação e pós-graduação. Vários dos professores participam do programa Reuni e trabalham em órgãos públicos como: Secretaria de Educação do Distrito Federal, Agência do Meio Ambiente do Estado de Goiás, Ministérios, entre outros. Além do mais, é importante destacar que alguns destes docentes participam do programa de mestrado profissional em ensino de biologia em rede nacional (PROFBIO), que tem por objetivo a qualificação do professor de biologia do ensino básico. O programa ainda possui muitas colaborações nacionais e internacionais e alguns intercâmbios internacionais feitos por pós-doutorandos da área (Capes, 2017).

## Captura dos dados: Definindo bancos de dados para as PPGs a partir de arquivos JSON e CSV:

Aqui define-se todas as variáveis base que contém todos os dados utilizados para a realização deste trabalho. Este código estará explícito logo abaixo para demonstrar que utilizou-se dos perfis, dos dados de publicações e orientações de cada um dos docentes. Além disso utilizou-se dados de correlação para a geração de um grafo de correlação ao fim deste trabalho.

```
library(jsonlite) #Importado para carga dos arquivos JSON para o R
# Perfis
perfil.240 <- fromJSON("240BiologiaAnimal/240profile.json")
perfil.242 <- fromJSON("242BiologiaMicrobiana/242profile.json")
perfil.243 <- fromJSON("243BiologiaMolecular/243profile.json")
perfil.313 <- fromJSON("313PatologiaMolecular/313profile.json")
# Publicações
public.240 <- fromJSON("240BiologiaAnimal/240publication.json")
public.242 <- fromJSON("242BiologiaMicrobiana/242publication.json")
public.243 <- fromJSON("243BiologiaMolecular/243publication.json")
public.313 <- fromJSON("313PatologiaMolecular/313publication.json")
# Orientações
orient.240 <- fromJSON("240BiologiaAnimal/240advise.json")
orient.242 <- fromJSON("242BiologiaMicrobiana/242advise.json")
orient.243 <- fromJSON("243BiologiaMolecular/243advise.json")
orient.313 <- fromJSON("313PatologiaMolecular/313advise.json")
# Grafos
graphl.240 <- fromJSON("240BiologiaAnimal/240graph.json")
graphl.242 <- fromJSON("242BiologiaMicrobiana/242graph.json")
graphl.243 <- fromJSON("243BiologiaMolecular/243graph.json")
graphl.313 <- fromJSON("313PatologiaMolecular/313graph.json")
# Todas PPGs
df.prog <- read.table("UnBPosGeral/prof_prog.csv", sep = ",",
                     colClasses = "character", encoding = "UTF-8", header = TRUE)
```

## CRISP-DM Fase 2 - Entendimento dos Dados

A segunda parte do CRISP-DM consiste no entendimento dos dados. Para realizar análises significativas com os **datasets** disponíveis, é essencial ter um bom entendimento sobre a forma que estão organizados.

Os arquivos utilizados são provenientes da plataforma Elattes e compilam informações sobre a produção

científica dos professores do **Programa de Pós-graduação em Biologia Animal** no período de 2010 a 2017.

Os **datasets** que serão trabalhados consistem em: perfil profissional; orientações de mestrado e doutorado realizadas; produções bibliográficas e redes de colaboração entre os pesquisadores.

## Arquivos Analisados

Os arquivos com informações sobre os pesquisadores do Programa de **Biologia Animal**:

- **240BiologiaAnimal/240profile.json**: apresenta dados sobre o **perfil** de todos os pesquisadores.
- **240BiologiaAnimal/240publication.json**: apresenta dados sobre as **publicações** e **produções** bibliográficas geradas por todos os pesquisadores.
- **240BiologiaAnimal/240advise.json**: apresenta dados sobre **orientações de mestrado e doutorado** feitas por todos os pesquisadores.
- **240BiologiaAnimal/240graph.json**: apresenta dados sobre **produções bibliográficas colaborativas** feitas entre os pesquisadores.

## Análise estrutural dos dados

Para continuar com as análises, as seguintes bibliotecas são selecionadas:

```
#library(tidyverse) #Importado para manipulação de tibbles
library(listviewer) #Importado para análise dos arquivos JSON
library(igraph) #Importado para manipulação de grafo
library(dplyr) #Importado para uso do Operador Pipe
library(tidyr) #Importado para uso da função spread()
library(ggplot2) #Importado para visualizações com ggplot()
library(stringr) #Importado para manipulações de strings
setwd("~/Repository/DataScience/DS4A-BioAni-BioMic-BioMol-PatMol-master")
source("elattes.ls2df.R")
```

Após a importação e definição das bibliotecas utilizadas, podemos utilizar funções de tais pacotes para análise sistêmica dos dados, mas para isto, estes devem ser primeiramente descritos por meio de funções de descrição do pacote **dplyr** que facilita a visualização e manipulação dos dados:

*Número de docentes na base e a construção das listas dos mesmos:*

```
## [1] Biologia Animal: 19
## [1] Biologia Microbiana: 23
## [1] Biologia Molecular: 36
## [1] Patologia Molecular: 23
```

## Análise das listas:

*Número de áreas de atuação cumulativas:*

```
## [1] Biologia Animal: 78
## [1] Biologia Microbiana: 95
## [1] Biologia Molecular: 149
## [1] Patologia Molecular: 96
```

*Numero de areas de atuacao por pessoa*

```
## [1] Biologia Animal:
##
## 2 3 4 5 6
## 2 3 7 5 2

## [1] Biologia Microbiana:
##
## 2 3 4 5 6
## 3 6 4 5 5

## [1] Biologia Molecular:
##
## 1 2 3 4 5 6
## 2 3 6 9 9 7

## [1] Patologia Molecular:
##
## 1 2 3 4 5 6
## 1 2 3 6 8 3
```

*Numero de pessoas por grande area*

```
## [1] Biologia Animal:
##
##          CIENCIAS_AGRARIAS          CIENCIAS_BIOLOGICAS
##                  2                  68
## CIENCIAS_EXATAS_E_DA_TERRA          CIENCIAS_HUMANAS
##                  2                  5
##          ENGENHARIAS
##                  1

## [1] Biologia Microbiana:
##
##          CIENCIAS_AGRARIAS          CIENCIAS_BIOLOGICAS
##                  9                  81
##          CIENCIAS_DA_SAUDE CIENCIAS_EXATAS_E_DA_TERRA
##                  4                  1

## [1] Biologia Molecular:
##
##          CIENCIAS_AGRARIAS          CIENCIAS_BIOLOGICAS
##                  9                  129
##          CIENCIAS_DA_SAUDE CIENCIAS_EXATAS_E_DA_TERRA
##                  4                  3
##          CIENCIAS_HUMANAS          ENGENHARIAS
##                  3                  1

## [1] Patologia Molecular:
##
##          CIENCIAS_AGRARIAS          CIENCIAS_BIOLOGICAS
##                  2                  77
##          CIENCIAS_DA_SAUDE CIENCIAS_EXATAS_E_DA_TERRA
```

##	11	4
##	CIENCIAS_HUMANAS	ENGENHARIAS
##	1	1

*Numero de pessoas que produziram os especificos tipos de producao*

## [1] Biologia Animal:

##	
##	ARTIGO_ACEITO
##	3
##	CAPITULO_DE_LIVRO
##	14
##	DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA
##	1
##	EVENTO
##	15
##	LIVRO
##	4
##	PERIODICO
##	19
##	TEXTO_EM_JORNAIS
##	4

## [1] Biologia Microbiana:

##	
##	ARTIGO_ACEITO
##	2
##	CAPITULO_DE_LIVRO
##	21
##	DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA
##	4
##	EVENTO
##	22
##	LIVRO
##	7
##	PERIODICO
##	23
##	TEXTO_EM_JORNAIS
##	3

## [1] Biologia Molecular:

##	
##	ARTIGO_ACEITO
##	7
##	CAPITULO_DE_LIVRO
##	32
##	DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA
##	4
##	EVENTO
##	30
##	LIVRO
##	12
##	PERIODICO
##	36

```

##          TEXTO_EM_JORNAIS
##          7
## [1] Patologia Molecular:
##
##          ARTIGO_ACEITO
##          6
##          CAPITULO_DE_LIVRO
##          18
## DEMAIS_TIPOS_DE_PRODUCAO_BIBLIOGRAFICA
##          2
##          EVENTO
##          20
##          LIVRO
##          4
##          PERIODICO
##          23
##          TEXTO_EM_JORNAIS
##          5

```

*Numero de publicacoes por tipo*

```

## [1] ----- Biologia Animal:
## [1] Artigos Aceitos: 6
## [1] Capítulos de livros: 41
## [1] Livros: 5
## [1] Periódicos: 681
## [1] Textos em jornais: 4
## [1] ----- Biologia Microbiana:
## [1] Artigos Aceitos: 2
## [1] Capítulos de livros: 78
## [1] Livros: 7
## [1] Periódicos: 381
## [1] Textos em jornais: 10
## [1] ----- Biologia Molecular:
## [1] Artigos Aceitos: 8
## [1] Capítulos de livros: 114
## [1] Livros: 22
## [1] Periódicos: 1185
## [1] Textos em jornais: 10
## [1] ----- Patologia Molecular:
## [1] Artigos Aceitos: 10
## [1] Capítulos de livros: 57
## [1] Livros: 8

```

## [1] Periódicos: 891

## [1] Textos em jornais: 72

*Número de pessoas por quantitativo de produções por pessoa*

## [1] ----- Biologia Animal:

##		0		2		
##	Artigos Aceitos:	16		3		
##			0		1	
##	Capítulos de livros:		5		2	
##		2	3		4	
##		7	1		2	
##		5	9			
##		1	1			
##		0	1	2		
##	Livros:	15	3	1		
##			10	11	13	18
##	Periódicos:		1	1	1	3
##		21	23	26	27	33
##		1	1	2	1	1
##		44	57	68	103	104
##		1	1	1	1	1

## [1] Textos em jornais: 4

## [1] ----- Biologia Microbiana:

##		0		1		
##	Artigos Aceitos:	21		2		
##			0		1	
##	Capítulos de livros:		15		3	
##		2				
##		1				

## [1] Livros: 7

## [1] Periódicos: 381

##		0		1		
##	Textos em jornais:	15		4		

## [1] ----- Biologia Molecular:

##		0		1		2
##	Artigos Aceitos:	29		6		1
##			0		1	
##	Capítulos de livros:		4		10	
##		2	3		4	
##		4	6		5	
##		5	7		8	
##		2	1		1	
##		11				
##		3				
##		0	1	2	3	4
##	Livros:	24	7	1	3	1

##		9	14	15	16	17
## Periódicos:		1	1	1	1	1
##	19	20	21	22	23	25
##	2	1	2	1	3	1
##	26	27	29	31	32	34
##	1	2	2	1	1	1
##	36	38	40	44	45	47
##	1	3	1	2	1	1
##	49	63	77	104		
##	1	1	1	1		

##		0	1
## Textos em jornais:		29	5
##	2	3	
##	1	1	

## [1] ----- Patologia Molecular:

##		0	1	2
## Artigos Aceitos:		17	4	1
##	4			
##	1			

##		0	1
## Capítulos de livros:		5	7
##	2	3	4
##	2	5	2
##	5	18	
##	1	1	

##	0	1	2	4
## Livros:	19	2	1	1

##		9	14	16	18	19
## Periódicos:		1	1	2	1	1
##	20	21	22	23	25	26
##	1	1	1	1	1	1
##	29	31	34	44	45	49
##	3	1	1	1	1	1
##	51	88	233			
##	1	1	1			

##		0	1
## Textos em jornais:		18	3
##	2	67	
##	1	1	

*Número de produções por ano*

## [1] ----- Biologia Animal:

##		2016	2017
## Artigos Aceitos:		1	5

##		2010	2011
## Capítulos de livros:		7	9
##	2012	2013	2014
##	3	4	3
##	2015	2016	2017



```

##          4          5          6
##      2013  2014  2016
## Livros:      2      1      2

##      2010      2011      2012      2013      2014
## Periódicos:      81      81      81      90      90
##      2015      2016      2017
##      76      91      91

##      2012      2013
## Textos em jornais:      1      1
##      2014      2016
##      1      1

## [1] ----- Biologia Microbiana:

##      2013      2016
## Artigos Aceitos:      1      1

##      2010      2011
## Capítulos de livros:      2      5
##      2012      2013      2014
##      8      10      27
##      2015      2016      2017
##      10      11      5

##      2010  2012  2013  2015  2016
## Livros:      1      1      2      1      2

##      2010      2011      2012      2013      2014
## Periódicos:      37      41      41      45      63
##      2015      2016      2017
##      47      54      53

##      2010      2011
## Textos em jornais:      2      1
##      2012      2013      2015
##      3      2      2

## [1] ----- Biologia Molecular:

##      2013      2016      2017
## Artigos Aceitos:      1      2      5

##      2010      2011
## Capítulos de livros:      8      10
##      2012      2013      2014
##      11      10      28
##      2015      2016      2017
##      15      17      15

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

##      2010      2011      2012      2013      2014
## Periódicos:      140      165      126      160      139
##      2015      2016      2017
##      159      158      138

##      2010      2011

```

```

## Textos em jornais:          2          2
##           2012          2014          2015
##           2          1          2
##           2016
##           1

## [1] ----- Patologia Molecular:

##           2017
## Artigos Aceitos:          10

##           2010          2011
## Capítulos de livros:          3          8
##           2012          2013          2014
##           8          4          13
##           2015          2016          2017
##           5          10          6

##           2010    2011    2014    2015    2016
## Livros:          1      1      1      2      3

##           2010          2011          2012          2013          2014
## Periódicos:          66          95          97          127          121
##           2015          2016          2017
##           132          131          122

##           2010          2012
## Textos em jornais:          1          2
##           2014          2015          2016
##           7          27          20
##           2017
##           15

```

*Número de pessoas que realizaram diferentes tipos de orientações*

```
print(c('Biologia Animal:',length(unlist(sapply(perfil.240, function(x) names(x$orientacoes_academicas))
```

```
## [1] Biologia Animal: 102
```

```
print(c('Biologia Microbiana:',length(unlist(sapply(perfil.242, function(x) names(x$orientacoes_academi
```

```
## [1] Biologia Microbiana: 108
```

```
print(c('Biologia Molecular:',length(unlist(sapply(perfil.243, function(x) names(x$orientacoes_academic
```

```
## [1] Biologia Molecular: 204
```

```
print(c('Patologia Molecular:',length(unlist(sapply(perfil.313, function(x) names(x$orientacoes_academi
```

```
## [1] Patologia Molecular: 133
```

*Número de pessoas por tipo de orientação*

```
print(c('----- Biologia Animal:',table(unlist(sapply(perfil.240, function(x) names(x$orientacoes_acader
```

```

##
##           ----- Biologia Animal:
##           ORIENTACAO_CONCLUIDA_DOUTORADO
##           18
##           ORIENTACAO_CONCLUIDA_MESTRADO
##           19
##           ORIENTACAO_CONCLUIDA_POS_DOUTORADO

```

```

##                                7
##      ORIENTACAO_EM_ANDAMENTO_DOUTORADO
##                                16
## ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA
##                                10
##      ORIENTACAO_EM_ANDAMENTO_MESTRADO
##                                14
##      OUTRAS_ORIENTACOES_CONCLUIDAS
##                                18
print(c('----- Biologia Microbiana:',table(unlist(sapply(perfil.242, function(x) names(x$orientacoes_a

##
##      ----- Biologia Microbiana:
##      ORIENTACAO_CONCLUIDA_DOUTORADO
##                                13
##      ORIENTACAO_CONCLUIDA_MESTRADO
##                                21
##      ORIENTACAO_CONCLUIDA_POS_DOUTORADO
##                                6
##      ORIENTACAO_EM_ANDAMENTO_DOUTORADO
##                                18
##      ORIENTACAO_EM_ANDAMENTO_GRADUACAO
##                                1
## ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA
##                                15
##      ORIENTACAO_EM_ANDAMENTO_MESTRADO
##                                14
##      OUTRAS_ORIENTACOES_CONCLUIDAS
##                                20
print(c('----- Biologia Molecular:',table(unlist(sapply(perfil.243, function(x) names(x$orientacoes_a

##
##      ----- Biologia Molecular:
##      ORIENTACAO_CONCLUIDA_DOUTORADO
##                                33
##      ORIENTACAO_CONCLUIDA_MESTRADO
##                                36
##      ORIENTACAO_CONCLUIDA_POS_DOUTORADO
##                                24
##      ORIENTACAO_EM_ANDAMENTO_DOUTORADO
##                                35
## ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA
##                                18
##      ORIENTACAO_EM_ANDAMENTO_MESTRADO
##                                24
##      OUTRAS_ORIENTACOES_CONCLUIDAS
##                                34
print(c('----- Patologia Molecular:',table(unlist(sapply(perfil.313, function(x) names(x$orientacoes_a

##
##      ----- Patologia Molecular:
##      ORIENTACAO_CONCLUIDA_DOUTORADO
##                                22

```

```
##          ORIENTACAO_CONCLUIDA_MESTRADO
##                               23
##          ORIENTACAO_CONCLUIDA_POS_DOUTORADO
##                               15
##          ORIENTACAO_EM_ANDAMENTO_DOUTORADO
##                               22
## ORIENTACAO_EM_ANDAMENTO_INICIACAO_CIENTIFICA
##                               13
##          ORIENTACAO_EM_ANDAMENTO_MESTRADO
##                               15
##          OUTRAS_ORIENTACOES_CONCLUIDAS
##                               23
```

*Número de orientações concluídas*

```
print('----- Biologia Animal:', quote=FALSE)
```

```
## [1] ----- Biologia Animal:
```

```
  print(c('Orientações concluídas de mestrados:',sum(sapply(perfil.240, function(x) length(x$orientacoes))
```

```
## [1] Orientações concluídas de mestrados:
```

```
## [2] 136
```

```
  print(c('Orientações concluídas de doutorados:',sum(sapply(perfil.240, function(x) length(x$orientacoes))
```

```
## [1] Orientações concluídas de doutorados:
```

```
## [2] 92
```

```
  print(c('Orientações concluídas de pós-doutores:',sum(sapply(perfil.240, function(x) length(x$orientacoes))
```

```
## [1] Orientações concluídas de pós-doutores:
```

```
## [2] 39
```

```
print('----- Biologia Microbiana:', quote=FALSE)
```

```
## [1] ----- Biologia Microbiana:
```

```
  print(c('Orientações concluídas de mestrados:',sum(sapply(perfil.242, function(x) length(x$orientacoes))
```

```
## [1] Orientações concluídas de mestrados:
```

```
## [2] 130
```

```
  print(c('Orientações concluídas de doutorados:',sum(sapply(perfil.242, function(x) length(x$orientacoes))
```

```
## [1] Orientações concluídas de doutorados:
```

```
## [2] 45
```

```
  print(c('Orientações concluídas de pós-doutores:',sum(sapply(perfil.242, function(x) length(x$orientacoes))
```

```
## [1] Orientações concluídas de pós-doutores:
```

```
## [2] 27
```

```
print('----- Biologia Molecular:', quote=FALSE)
```

```
## [1] ----- Biologia Molecular:
```

```
  print(c('Orientações concluídas de mestrados:',sum(sapply(perfil.243, function(x) length(x$orientacoes))
```

```
## [1] Orientações concluídas de mestrados:
```

```
## [2] 275
```

```

print(c('Orientações concluídas de doutorados:',sum(sapply(perfil.243, function(x) length(x$orientacoes))

## [1] Orientações concluídas de doutorados:
## [2] 199

print(c('Orientações concluídas de pós-doutores:',sum(sapply(perfil.243, function(x) length(x$orientacoes))

## [1] Orientações concluídas de pós-doutores:
## [2] 107

print('----- Patologia Molecular:', quote=FALSE)

## [1] ----- Patologia Molecular:

print(c('Orientações concluídas de mestrados:',sum(sapply(perfil.313, function(x) length(x$orientacoes))

## [1] Orientações concluídas de mestrados:
## [2] 242

print(c('Orientações concluídas de doutorados:',sum(sapply(perfil.313, function(x) length(x$orientacoes))

## [1] Orientações concluídas de doutorados:
## [2] 152

print(c('Orientações concluídas de pós-doutores:',sum(sapply(perfil.313, function(x) length(x$orientacoes))

## [1] Orientações concluídas de pós-doutores:
## [2] 79

Número de pessoas por quantitativo de orientações por pessoa

## [1] ----- Biologia Animal:

##
## Orientações concluídas de mestrados:
## 2 5
## 2 3
## 6 7
## 2 2
## 8 9
## 3 1
## 10 11
## 1 2
## 12 13
## 1 1

##
## Orientações concluídas de doutorados:
## 0
## 1
## 1
## 4
## 2
## 2
## 3
## 2
## 5
## 1
## 6

```

```

##                2
##                7
##                2
##                8
##                3
##               10
##                1
##               13
##                1
##
## Orientações concluídas de pós-doutores:
##                0
##               12
##                2
##                2
##                3
##                1
##                5
##                1
##                6
##                1
##                7
##                1
##               14
##                1
##
## [1] ----- Biologia Microbiana:
##
## Orientações concluídas de mestrados:
##                1
##                1
##                3
##                3
##                1
##                5
##                2
##                7
##                2
##                9
##                1
##               11
##                1
##
## Orientações concluídas de doutorados:
##                0
##               10
##                1
##                4
##                2
##                3
##                3
##                1
##                5
##                1

```

```

##                                     6
##                                     2
##                                     7
##                                     1
##                                     8
##                                     1

##
## Orientações concluídas de pós-doutores:
##                                     0
##                                     17
##                                     1
##                                     2
##                                     2
##                                     2
##                                     4
##                                     1
##                                     17
##                                     1

## [1] ----- Biologia Molecular:

##                                     1
## Orientações concluídas de mestrados:
##                                     2
##                                     3
##                                     4
##                                     4
##                                     6
##                                     4
##                                     8
##                                     8
##                                     2
##                                     10
##                                     5
##                                     5
##                                     12
##                                     12
##                                     3
##                                     17
##                                     1

##
## Orientações concluídas de doutorados:
##                                     0
##                                     3
##                                     2
##                                     5
##                                     3
##                                     6
##                                     4
##                                     2
##                                     5
##                                     3
##                                     6
##                                     4
##                                     7
##                                     5
##                                     8
##                                     5
##                                     11

```

```

##                                     1
##                                     17
##                                     1
##                                     21
##                                     1
##
## Orientações concluídas de pós-doutores:
##                                     0
##                                     12
##                                     1
##                                     3
##                                     2
##                                     6
##                                     3
##                                     2
##                                     4
##                                     4
##                                     5
##                                     3
##                                     6
##                                     2
##                                     7
##                                     2
##                                     8
##                                     1
##                                     21
##                                     1
##
## [1] ----- Patologia Molecular:
##                                     4
## Orientações concluídas de mestrados:
##                                     3
##                                     5
##                                     6
##                                     3
##                                     8
##                                     1
##                                     10
##                                     3
##                                     12
##                                     1
##                                     27
##                                     1
##                                     41
##                                     1
##
## Orientações concluídas de doutorados:
##                                     0
##                                     1
##                                     2
##                                     3
##                                     3
##                                     4
##                                     4
##                                     2

```



```
##          5
##          3
##          6
##          2
##          7
##          2
##          8
##          2
##         11
##          1
##         14
##          1
##         22
##          2
```

```
##
## Orientações concluídas de pós-doutores:
##          0
##          8
##          1
##          1
##          2
##          3
##          3
##          1
##          4
##          4
##          5
##          2
##          7
##          1
##          8
##          1
##         14
##          2
```

*Número de orientações por ano*

```
print('----- Biologia Animal:', quote=FALSE)
```

```
## [1] ----- Biologia Animal:
```

```
print(c('Orientações concluídas de mestrados:',table(unlist(sapply(perfil.240, function(x) (x$orienta
```

```
##          2010
## Orientações concluídas de mestrados:          17
##          2011          2012
##          14          21
##          2013          2014
##          26          20
##          2015          2016
##          13          19
##          2017
##          6
```

```
print(c('Orientações concluídas de doutorados:',table(unlist(sapply(perfil.240, function(x) (x$orienta
```

```

##
## Orientações concluídas de doutorados:
##           2010
##           2
##           2011
##           15
##           2012
##           15
##           2013
##           4
##           2014
##           11
##           2015
##           15
##           2016
##           11
##           2017
##           19

print(c('Orientações concluídas de pós-doutores:',table(unlist(sapply(perfil.240, function(x) (x$orienta

##
## Orientações concluídas de pós-doutores:
##           2010
##           3
##           2011
##           4
##           2012
##           9
##           2013
##           5
##           2014
##           9
##           2015
##           3
##           2016
##           6

print('----- Biologia Microbiana:', quote=FALSE)

## [1] ----- Biologia Microbiana:

print(c('Orientações concluídas de mestrados:',table(unlist(sapply(perfil.242, function(x) (x$orienta

##           2010
## Orientações concluídas de mestrados:           9
##           2011           2012
##           16           17
##           2013           2014
##           17           31
##           2015           2016
##           18           9
##           2017
##           13

print(c('Orientações concluídas de doutorados:',table(unlist(sapply(perfil.242, function(x) (x$orienta

```

```
##
## Orientações concluídas de doutorados:
##           2010
##           4
##           2012
##           8
##           2013
##          13
##           2014
##           5
##           2015
##           2
##           2016
##           9
##           2017
##           4

print(c('Orientações concluídas de pós-doutores:',table(unlist(sapply(perfil.242, function(x) (x$orienta

##
## Orientações concluídas de pós-doutores:
##           2010
##           11
##           2011
##            1
##           2013
##            3
##           2014
##            6
##           2015
##            2
##           2016
##            4

print('----- Biologia Molecular:', quote=FALSE)

## [1] ----- Biologia Molecular:

print(c('Orientações concluídas de mestrados:',table(unlist(sapply(perfil.243, function(x) (x$orienta

##           2010
## Orientações concluídas de mestrados:           32
##           2011           2012
##           38           39
##           2013           2014
##           45           46
##           2015           2016
##           28           36
##           2017
##           11

print(c('Orientações concluídas de doutorados:',table(unlist(sapply(perfil.243, function(x) (x$orienta

##
## Orientações concluídas de doutorados:
##           2010
##           24
```

```
##          2011
##          17
##          2012
##          33
##          2013
##          27
##          2014
##          26
##          2015
##          22
##          2016
##          26
##          2017
##          24
```

```
print(c('Orientações concluídas de pós-doutores:',table(unlist(sapply(perfil.243, function(x) (x$orien
```

```
##
## Orientações concluídas de pós-doutores:
##          2010
##          11
##          2011
##           8
##          2012
##          13
##          2013
##          18
##          2014
##          18
##          2015
##          12
##          2016
##          17
##          2017
##          10
```

*Cria arquivo com dados quantitativos para análise*

```
## [1] "Biologia Animal"
```

```
##      idLattes      nome      resumo_cv
## Length:19      Length:19      Length:19
## Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character
##
##
##
##      senioridade      ORIENTACAO_CONCLUIDA_DOUTORADO
## Length:19      Min.   : 0.000
## Class :character 1st Qu.: 1.500
## Mode  :character Median : 5.000
##                  Mean   : 4.842
##                  3rd Qu.: 7.500
##                  Max.   :13.000
## ORIENTACAO_CONCLUIDA_MESTRADO ORIENTACAO_CONCLUIDA_POS_DOUTORADO
## Min.   : 1.000      Min.   : 0.000
```

```

## 1st Qu.: 5.000          1st Qu.: 0.000
## Median : 7.000          Median : 0.000
## Mean   : 7.158          Mean    : 2.053
## 3rd Qu.: 9.500          3rd Qu.: 2.500
## Max.    :13.000          Max.    :14.000
## OUTRAS_ORIENTACOES_CONCLUIDAS CAPITULO_DE_LIVRO      EVENTO
## Min.     : 0.0           Min.     :0.000      Min.     : 0.00
## 1st Qu.: 7.5             1st Qu.:0.500      1st Qu.: 3.50
## Median :17.0             Median :2.000      Median :11.00
## Mean    :16.0            Mean    :2.158      Mean    :17.42
## 3rd Qu.:21.5             3rd Qu.:2.500      3rd Qu.:20.00
## Max.     :36.0           Max.     :9.000      Max.     :60.00
##      LIVRO      OUTRAS_PRODUCOES      PERIODICO      TEXTO_EM_JORNAIS
## Min.     :0.0000   Min.     :0.00000   Min.     : 10.00   Min.     :0.0000
## 1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.: 19.00   1st Qu.:0.0000
## Median :0.0000   Median :0.00000   Median : 26.00   Median :0.0000
## Mean    :0.2632   Mean    :0.05263   Mean    : 35.84   Mean    :0.2105
## 3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.: 42.00   3rd Qu.:0.0000
## Max.     :2.0000   Max.     :1.00000   Max.     :104.00   Max.     :1.0000
## num_areas
## 1:5
## 2:5
## 3:6
## 4:2
## 6:1
##
## [1] "Biologia Microbiana"
##      idLattes      nome      resumo_cv
## Length:23      Length:23      Length:23
## Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character
##
##
##
## senioridade      ORIENTACAO_CONCLUIDA_DOUTORADO
## Length:23      Min.     :0.000
## Class :character 1st Qu.:0.000
## Mode  :character Median :1.000
##                  Mean    :1.957
##                  3rd Qu.:2.500
##                  Max.     :8.000
## ORIENTACAO_CONCLUIDA_MESTRADO ORIENTACAO_CONCLUIDA_POS_DOUTORADO
## Min.     : 0.000          Min.     : 0.000
## 1st Qu.: 2.500          1st Qu.: 0.000
## Median : 6.000          Median : 0.000
## Mean    : 5.652          Mean    : 1.174
## 3rd Qu.: 8.500          3rd Qu.: 0.500
## Max.     :11.000          Max.     :17.000
## OUTRAS_ORIENTACOES_CONCLUIDAS CAPITULO_DE_LIVRO      EVENTO
## Min.     : 0.00          Min.     : 0.000      Min.     : 0.00
## 1st Qu.: 6.50            1st Qu.: 1.000      1st Qu.: 8.50
## Median :10.00            Median : 3.000      Median :20.00
## Mean    :11.57           Mean    : 3.391      Mean    :25.13

```

```

## 3rd Qu.:17.50          3rd Qu.: 4.000      3rd Qu.:34.50
## Max.    :26.00          Max.    :13.000      Max.    :82.00
##      LIVRO      OUTRAS_PRODUCOES      PERIODICO      TEXTO_EM_JORNAIS
## Min.    :0.0000      Min.    :0.0000      Min.    : 3.00      Min.    :0.0000
## 1st Qu.:0.0000      1st Qu.:0.0000      1st Qu.: 8.50      1st Qu.:0.0000
## Median :0.0000      Median :0.0000      Median :11.00      Median :0.0000
## Mean    :0.3043      Mean    :0.4348      Mean    :16.57      Mean    :0.4348
## 3rd Qu.:1.0000      3rd Qu.:0.0000      3rd Qu.:17.50      3rd Qu.:0.0000
## Max.    :1.0000      Max.    :4.0000      Max.    :52.00      Max.    :6.0000
## num_areas
## 1:7
## 2:7
## 3:6
## 4:3
##
##
## [1] "Biologia Molecular"
##      idLattes      nome      resumo_cv
## Length:36      Length:36      Length:36
## Class :character      Class :character      Class :character
## Mode  :character      Mode  :character      Mode  :character
##
##
##
## senioridade      ORIENTACAO_CONCLUIDA_DOUTORADO
## Length:36      Min.    : 0.000
## Class :character      1st Qu.: 3.000
## Mode  :character      Median : 5.000
##      Mean    : 5.528
##      3rd Qu.: 7.000
##      Max.    :21.000
## ORIENTACAO_CONCLUIDA_MESTRADO ORIENTACAO_CONCLUIDA_POS_DOUTORADO
## Min.    : 1.000      Min.    : 0.000
## 1st Qu.: 5.000      1st Qu.: 0.000
## Median : 7.000      Median : 2.000
## Mean    : 7.639      Mean    : 2.972
## 3rd Qu.:10.000      3rd Qu.: 4.250
## Max.    :17.000      Max.    :21.000
## OUTRAS_ORIENTACOES_CONCLUIDAS CAPITULO_DE_LIVRO      EVENTO
## Min.    : 0.00      Min.    : 0.000      Min.    : 0.00
## 1st Qu.: 5.00      1st Qu.: 1.000      1st Qu.: 3.00
## Median : 9.00      Median : 2.500      Median :15.00
## Mean    :11.61      Mean    : 3.167      Mean    :20.83
## 3rd Qu.:16.50      3rd Qu.: 4.000      3rd Qu.:32.00
## Max.    :36.00      Max.    :11.000      Max.    :68.00
##      LIVRO      OUTRAS_PRODUCOES      PERIODICO      TEXTO_EM_JORNAIS
## Min.    :0.0000      Min.    :0.0000      Min.    : 9.00      Min.    :0.0000
## 1st Qu.:0.0000      1st Qu.:0.0000      1st Qu.:21.00      1st Qu.:0.0000
## Median :0.0000      Median :0.0000      Median :28.00      Median :0.0000
## Mean    :0.6111      Mean    :0.1667      Mean    :32.92      Mean    :0.2778
## 3rd Qu.:1.0000      3rd Qu.:0.0000      3rd Qu.:38.50      3rd Qu.:0.0000
## Max.    :4.0000      Max.    :2.0000      Max.    :104.00      Max.    :3.0000
## num_areas

```

```

## 1: 6
## 2:12
## 3:13
## 4: 5
##
##
## [1] "Patologia Molecular"

##      idLattes      nome      resumo_cv
## Length:23      Length:23      Length:23
## Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character
##
##
##
##      senioridade      ORIENTACAO_CONCLUIDA_DOUTORADO
## Length:23      Min.      : 0.000
## Class :character 1st Qu.: 3.000
## Mode  :character Median : 5.000
##                  Mean   : 6.609
##                  3rd Qu.: 7.500
##                  Max.   :22.000
## ORIENTACAO_CONCLUIDA_MESTRADO ORIENTACAO_CONCLUIDA_POS_DOUTORADO
## Min.      : 4.00      Min.      : 0.000
## 1st Qu.: 6.00      1st Qu.: 0.000
## Median : 9.00      Median : 2.000
## Mean   :10.52      Mean   : 3.435
## 3rd Qu.:11.00      3rd Qu.: 4.500
## Max.   :41.00      Max.   :14.000
## OUTRAS_ORIENTACOES_CONCLUIDAS CAPITULO_DE_LIVRO      EVENTO
## Min.      : 1.0      Min.      : 0.000      Min.      : 0.00
## 1st Qu.: 5.0      1st Qu.: 1.000      1st Qu.: 10.50
## Median : 9.0      Median : 1.000      Median : 20.00
## Mean   :12.7      Mean   : 2.478      Mean   : 30.52
## 3rd Qu.:17.0      3rd Qu.: 3.000      3rd Qu.: 34.00
## Max.   :42.0      Max.   :18.000      Max.   :213.00
##      LIVRO      OUTRAS_PRODUCOES      PERIODICO      TEXTO_EM_JORNAIS
## Min.      :0.0000      Min.      :0.0000      Min.      : 9.00      Min.      : 0.00
## 1st Qu.:0.0000      1st Qu.:0.0000      1st Qu.: 19.50      1st Qu.: 0.00
## Median :0.0000      Median :0.0000      Median : 26.00      Median : 0.00
## Mean   :0.3478      Mean   :0.1739      Mean   : 38.74      Mean   : 3.13
## 3rd Qu.:0.0000      3rd Qu.:0.0000      3rd Qu.: 39.00      3rd Qu.: 0.00
## Max.   :4.0000      Max.   :2.0000      Max.   :233.00      Max.   :67.00
## num_areas
## 1: 1
## 2: 9
## 3:12
## 4: 1
##
##

```

## CRISP-DM Fase 3 - Preparação dos Dados

A terceira fase do CRISP é conhecida por ser a parte de preparação dos dados. Tal fase possui como característica a execução de atividades para construir o conjunto final de dados a partir dos dados brutos iniciais. Pode-se separar essa etapa em cinco momentos que serão descritos nessa seção.

- Seleção dos dados
- Limpeza dos dados
- Construção dos dados
- Integração dos dados
- Formatação dos dados

Na etapa de seleção dos dados a entrada é o conjunto de dados bruto e nela ocorre a decisão dos dados a serem usados para análise. Os critérios incluem relevância para as metas de mineração de dados, qualidade e restrições técnicas, como limites no volume de dados ou tipos de dados. Então vem a fase da limpeza que recebe a seleção de dados úteis efetuada anteriormente e é efetuado um aumento na qualidade dos dados para o nível exigido pelas técnicas de análise selecionadas. Aqui pode haver o uso de técnicas mais elaboradas, como a estimativa de dados ausentes por modelagem e inserção de padrões adequados.

O terceiro passo é a construção dos dados. Essa tarefa inclui operações de preparação de dados construtivos, como a produção de atributos derivados, novos registros ou valores transformados para atributos existentes. A penúltima atividade é a integração dos dados. Este é o momento no qual as informações são combinadas de vários bancos de dados, tabelas ou registros para criar novos registros ou valores. Por fim, ocorre a tarefa de formatação dos dados, que é a realização de modificações na estrutura dos dados de forma que as operações planejadas possam ser efetuadas de forma conveniente.

Para tornar a análise mais fácil de ser feita e até mesmo para possibilitar a realização de comparações ao final, os mesmos procedimentos foram realizados para os três programas de pós-graduação. Além disso, é importante ressaltar que as variáveis e estruturas montadas foram nomeados de forma mnemônica permitindo a distinção de diferentes programas e aspectos, como orientações, publicações, entre outros.

### Análise dos dados no formato lista

#### *Número de Publicacoes em periódicos*

```
## [1] Biologia Animal: 529
## [1] Biologia Microbiana: 335
## [1] Biologia Molecular: 977
## [1] Patologia Molecular: 736
```

#### *Anos analisados*

```
## [1] "Biologia Animal:"
## [1] "2010" "2011" "2012" "2013" "2014" "2015" "2016" "2017"
## [1] "Biologia Microbiana:"
## [1] "2010" "2011" "2012" "2013" "2014" "2015" "2016" "2017"
## [1] "Biologia Molecular:"
## [1] "2010" "2011" "2012" "2013" "2014" "2015" "2016" "2017"
## [1] "Patologia Molecular:"
## [1] "2010" "2011" "2012" "2013" "2014" "2015" "2016" "2017"
```



20 revistas mais publicadas

```
## [1] "Biologia Animal:"
##
##                                     Plos One
##                                     18
##                                     Toxicon (Oxford)
##                                     14
##                                     Genetics and Molecular Biology (Impresso)
##                                     11
##                                     Journal of Biomedical Nanotechnology
##                                     8
##                                     Genetics and Molecular Research
##                                     7
##                                     International Journal of Nanomedicine (Online)
##                                     7
##                                     Protein and Peptide Letters
##                                     7
## Revista de la Universidad Industrial de Santander. Salud
##                                     7
##                                     Journal of Nanobiotechnology
##                                     6
##                                     Scientific Reports
##                                     6
##                                     Theriogenology
##                                     6
##                                     Behavioural Brain Research
##                                     5
##                                     Frontiers in Behavioral Neuroscience
##                                     5
##                                     Frontiers in Pharmacology
##                                     5
##                                     Genética na Escola
##                                     5
##                                     Journal of Nanomedicine & Nanotechnology
##                                     5
##                                     Nanomedicine
##                                     5
##                                     Small Ruminant Research
##                                     5
## Bulletin of Environmental Contamination and Toxicology
##                                     4
##                                     Journal of Nanoscience and Nanotechnology (Print)
##                                     4
## [1] "Biologia Microbiana:"
##
##                                     Archives of Virology
##                                     13
##                                     Plant Disease
##                                     9
##                                     Plos One
##                                     8
##                                     Virus Genes
```

```

##                                     8
##                               Microbial Ecology
##                                     6
##                               BMC Genomics
##                                     5
##                               Mycologia
##                                     5
##                               Mycological Progress
##                                     5
##                               Mycotaxon
##                                     5
##                               Biogeochemistry (Dordrecht)
##                                     4
##                               Frontiers in Microbiology (Online)
##                                     4
##                               Fungal Genetics and Biology (Print)
##                                     4
##                               Global Change Biology (Print)
##                                     4
##                               Microbiologia in Foco
##                                     4
##                               Applied Microbiology and Biotechnology
##                                     3
##                               BIOENERG RES
##                                     3
##                               Biotechnology for Biofuels
##                                     3
##                               Brazilian Journal of Microbiology (Impresso)
##                                     3
##                               Genetics and Molecular Research
##                                     3
##                               Journal of Industrial Microbiology & Biotechnology
##                                     3

## [1] "Biologia Molecular:"

##
##                               Plos One
##                               49
##                               BMC Genomics
##                               20
##                               BMC Proceedings
##                               20
##                               Archives of Virology
##                               19
##                               Toxicon (Oxford)
##                               12
##                               Frontiers in Microbiology (Online)
##                               11
##                               Genetics and Molecular Research
##                               10
##                               PLoS Neglected Tropical Diseases (Online)
##                               10
##                               Protein and Peptide Letters
##                               10

```

```

##          Scientific Reports
##          10
##          Virus Genes
##          10
##          Plant Disease
##          8
##          Tree Genetics & Genomes (Print)
##          8
##          BMC Plant Biology (Online)
##          7
##          Journal of Proteome Research (Print)
##          7
##          Proteomics (Weinheim. Print)
##          7
##          Acta Horticulturae
##          6
##          Applied Microbiology and Biotechnology
##          6
##          Journal of Biomedical Nanotechnology
##          6
##          Microbial Ecology
##          6
## [1] "Patologia Molecular:"
##
##          Plos One
##          44
##          Frontiers in Microbiology (Online)
##          22
##          Peptides (New York, N.Y. 1980)
##          16
##          PLoS Neglected Tropical Diseases (Online)
##          12
##          BMC Genomics
##          11
##          Scientific Reports
##          11
##          Protein and Peptide Letters
##          10
##          Biopolymers (New York. Print)
##          9
##          Journal of Proteome Research (Print)
##          9
##          Journal of Biomedical Nanotechnology
##          8
##          Antimicrobial Agents and Chemotherapy (Print)
##          7
##          Biochimica et Biophysica Acta. Biomembranes
##          6
##          BMC Microbiology (Online)
##          6
##          Current Protein and Peptide Science
##          6
##          Experimental Parasitology

```

```
## 6
## Fungal Genetics and Biology (Print)
## 6
## Journal of Proteomics
## 6
## Medical Mycology (Oxford. Print)
## 6
## International Journal of Nanomedicine (Online)
## 5
## Revista da Sociedade Brasileira de Medicina Tropical (Impresso)
## 5
```

*Análise dos dados no formato dataframe*

*Publicações por ano*

```
## [1] "Biologia Animal "
##
## 2010 2011 2012 2013 2014 2015 2016 2017
## 70 68 60 70 70 53 66 72
## [1] "Biologia Microbiana"
##
## 2010 2011 2012 2013 2014 2015 2016 2017
## 35 39 36 34 50 44 48 49
## [1] "Biologia Molecular"
##
## 2010 2011 2012 2013 2014 2015 2016 2017
## 123 136 101 126 116 127 130 118
## [1] "Patologia Molecular"
##
## 2010 2011 2012 2013 2014 2015 2016 2017
## 54 73 85 105 100 113 109 97
```

*20 revistas mais publicadas (análise com dataframe)*

```
## [1] "Biologia Animal "
##
## Plos One
## 18
## Toxicon (Oxford)
## 14
## Genetics and Molecular Biology (Impresso)
## 11
## Journal of Biomedical Nanotechnology
## 8
## Genetics and Molecular Research
## 7
## International Journal of Nanomedicine (Online)
## 7
## Protein and Peptide Letters
## 7
## Revista de la Universidad Industrial de Santander. Salud
```

```

##                                     7
##                               Journal of Nanobiotechnology
##                                     6
##                               Scientific Reports
##                                     6
##                               Theriogenology
##                                     6
##                               Behavioural Brain Research
##                                     5
##                               Frontiers in Behavioral Neuroscience
##                                     5
##                               Frontiers in Pharmacology
##                                     5
##                               Genética na Escola
##                                     5
##                               Journal of Nanomedicine & Nanotechnology
##                                     5
##                               Nanomedicine
##                                     5
##                               Small Ruminant Research
##                                     5
##                               Bulletin of Environmental Contamination and Toxicology
##                                     4
##                               Journal of Nanoscience and Nanotechnology (Print)
##                                     4
## [1] "Biologia Microbiana"
##
##                               Archives of Virology
##                                     13
##                               Plant Disease
##                                     9
##                               Plos One
##                                     8
##                               Virus Genes
##                                     8
##                               Microbial Ecology
##                                     6
##                               BMC Genomics
##                                     5
##                               Mycologia
##                                     5
##                               Mycological Progress
##                                     5
##                               Mycotaxon
##                                     5
##                               Biogeochemistry (Dordrecht)
##                                     4
##                               Frontiers in Microbiology (Online)
##                                     4
##                               Fungal Genetics and Biology (Print)
##                                     4
##                               Global Change Biology (Print)
##                                     4

```

```

##                               Microbiologia in Foco
##                               4
##       Applied Microbiology and Biotechnology
##                               3
##                               BIOENERG RES
##                               3
##       Biotechnology for Biofuels
##                               3
##       Brazilian Journal of Microbiology (Impresso)
##                               3
##       Genetics and Molecular Research
##                               3
## Journal of Industrial Microbiology & Biotechnology
##                               3

## [1] "Biologia Molecular"

##
##                               Plos One
##                               49
##                               BMC Genomics
##                               20
##                               BMC Proceedings
##                               20
##                               Archives of Virology
##                               19
##                               Toxicon (Oxford)
##                               12
##       Frontiers in Microbiology (Online)
##                               11
##       Genetics and Molecular Research
##                               10
## PLoS Neglected Tropical Diseases (Online)
##                               10
##       Protein and Peptide Letters
##                               10
##       Scientific Reports
##                               10
##       Virus Genes
##                               10
##       Plant Disease
##                               8
##       Tree Genetics & Genomes (Print)
##                               8
##       BMC Plant Biology (Online)
##                               7
##       Journal of Proteome Research (Print)
##                               7
##       Proteomics (Weinheim. Print)
##                               7
##       Acta Horticulturae
##                               6
##       Applied Microbiology and Biotechnology
##                               6
##       Journal of Biomedical Nanotechnology

```

##		6
##	Microbial Ecology	
##		6
##	[1] "Patologia Molecular"	
##		
##		Plos One
##		44
##	Frontiers in Microbiology (Online)	
##		22
##	Peptides (New York, N.Y. 1980)	
##		16
##	PLoS Neglected Tropical Diseases (Online)	
##		12
##		BMC Genomics
##		11
##		Scientific Reports
##		11
##	Protein and Peptide Letters	
##		10
##	Biopolymers (New York. Print)	
##		9
##	Journal of Proteome Research (Print)	
##		9
##	Journal of Biomedical Nanotechnology	
##		8
##	Antimicrobial Agents and Chemotherapy (Print)	
##		7
##	Biochimica et Biophysica Acta. Biomembranes	
##		6
##	BMC Microbiology (Online)	
##		6
##	Current Protein and Peptide Science	
##		6
##	Experimental Parasitology	
##		6
##	Fungal Genetics and Biology (Print)	
##		6
##	Journal of Proteomics	
##		6
##	Medical Mycology (Oxford. Print)	
##		6
##	International Journal of Nanomedicine (Online)	
##		5
##	Revista da Sociedade Brasileira de Medicina Tropical (Impresso)	
##		5

*Número de Orientações de Mestrado e Doutorado*

##	[1] Biologia Animal:	220
##	[1] Biologia Microbiana:	175
##	[1] Biologia Molecular:	460
##	[1] Patologia Molecular:	386

*Construção de dataframe de orientações*

*Construção de Grafo de Colaborações*

```
g.240 <- g.ls2ig(graphl.240)
df.240 <- as.data.frame(V(g.240)$name); colnames(df.240) <- "Idlattes"
df.240 <- left_join(df.240, df.prog, by = c("Idlattes" = "idLattes"))
g.242 <- g.ls2ig(graphl.242)
df.242 <- as.data.frame(V(g.242)$name); colnames(df.242) <- "Idlattes"
df.242 <- left_join(df.242, df.prog, by = c("Idlattes" = "idLattes"))
g.243 <- g.ls2ig(graphl.243)
df.243 <- as.data.frame(V(g.243)$name); colnames(df.243) <- "Idlattes"
df.243 <- left_join(df.243, df.prog, by = c("Idlattes" = "idLattes"))
g.313 <- g.ls2ig(graphl.313)
df.313 <- as.data.frame(V(g.313)$name); colnames(df.313) <- "Idlattes"
df.313 <- left_join(df.313, df.prog, by = c("Idlattes" = "idLattes"))
perfil.total.df = bind_rows(perfil.240.df, perfil.242.df, perfil.243.df, perfil.313.df)
```

Apenas para fins de análise inicial, foram retiradas do grafo as observacoes de pesquisadores duplicados quando inclusos em mais de um programa, vide código abaixo.

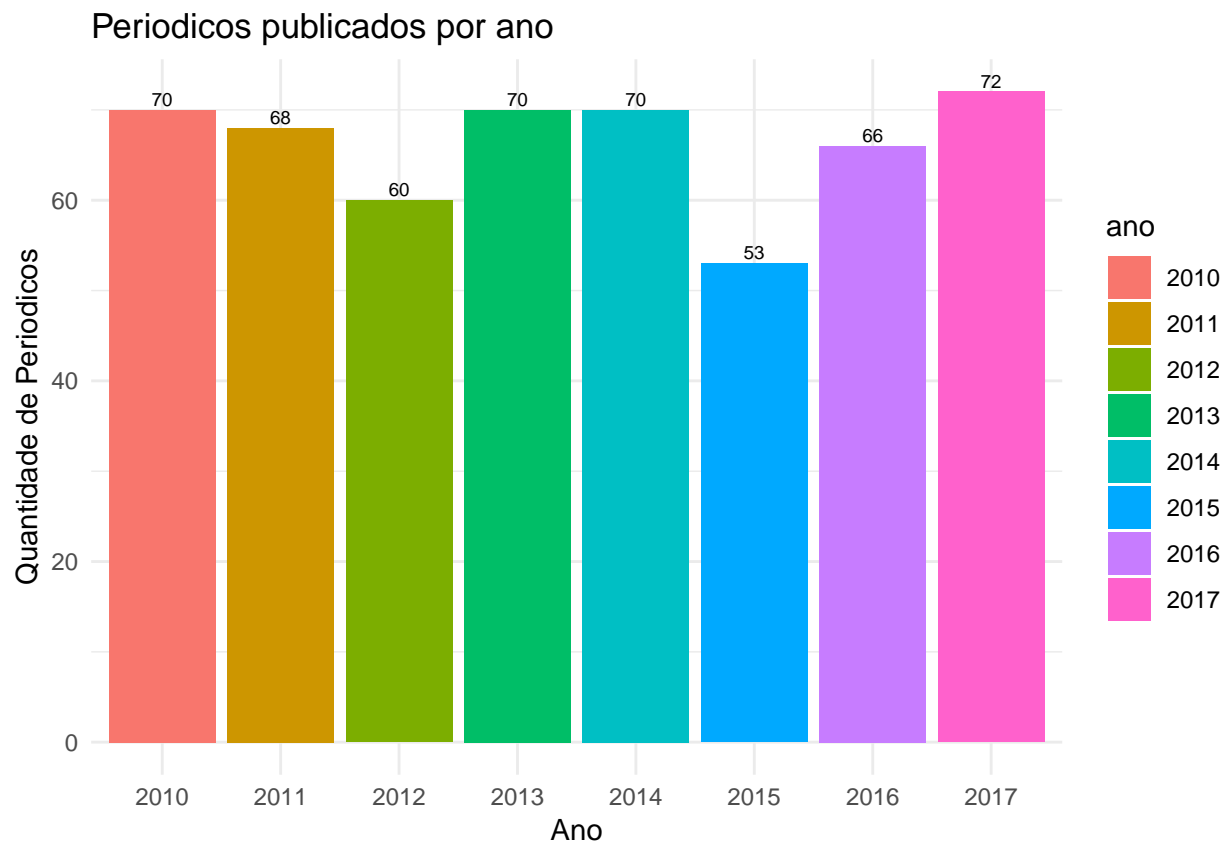
## CRISP-DM Fases 4 a 6 - Resultados e visualizações:

Foram escolhidos alguns resultados, em relação aos dados encontrados durante o processo, para serem plotados em gráficos.



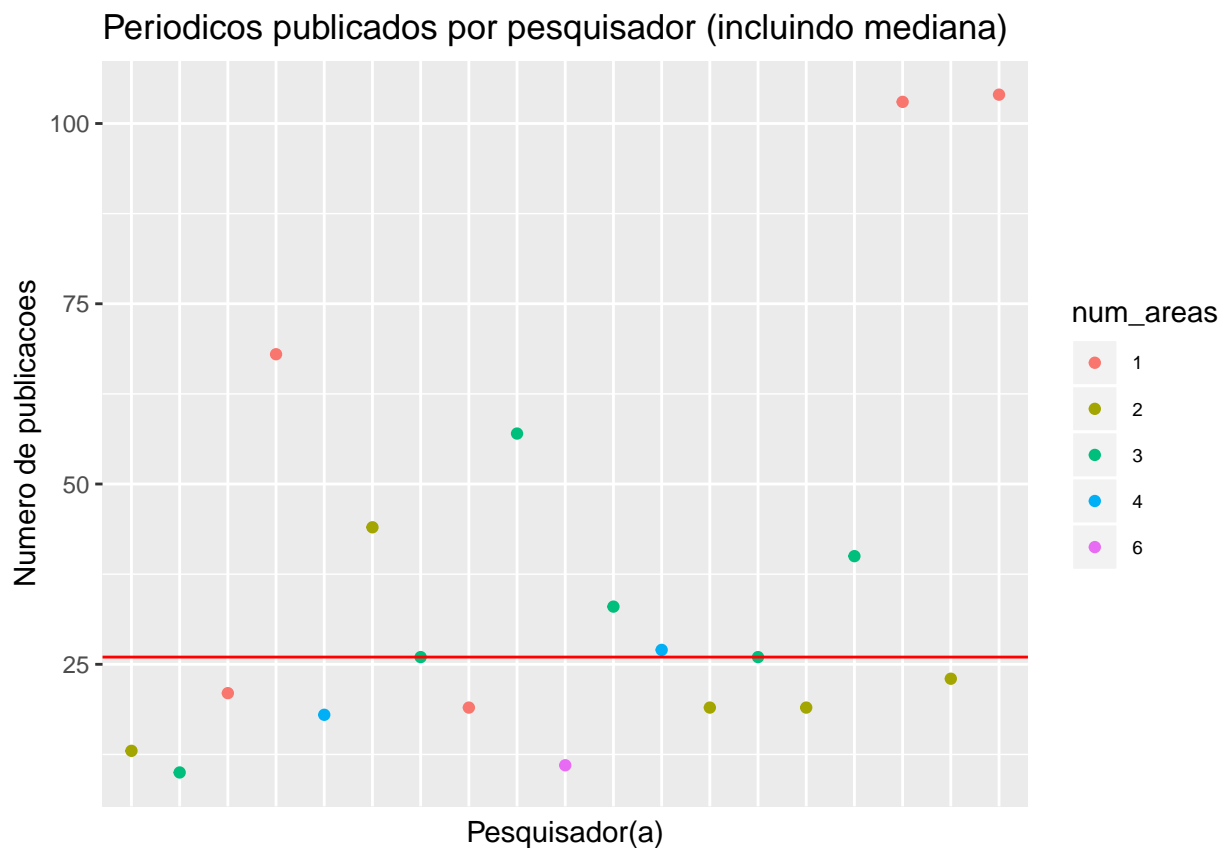
## Biologia Animal

Grafico de barras; periodicos por ano



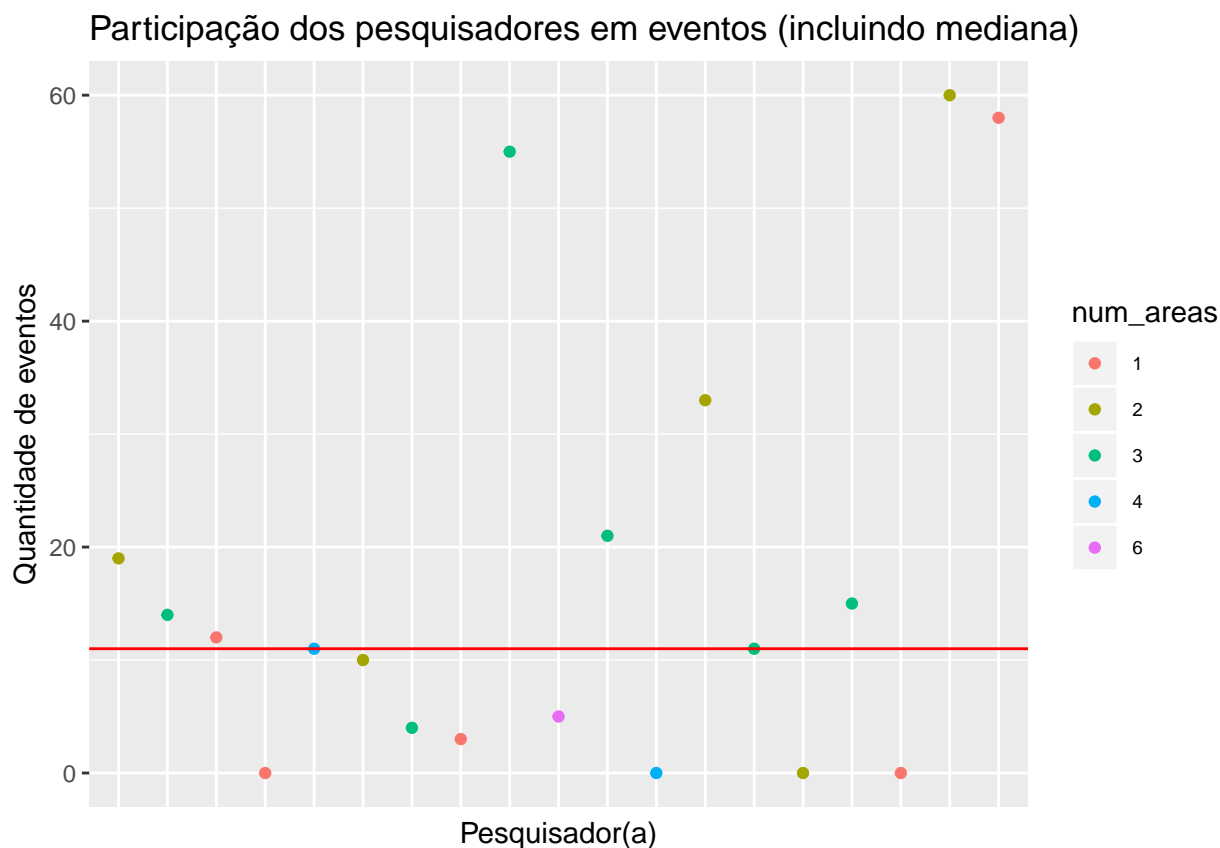
Sob uma perspectiva de publicações em periódicos, o Programa de Pós-Graduação de Biologia Animal apresentou pouca variação entre os anos de 2011 e 2017, registrando valores entre 60 e 72 publicações por ano. Um vale único foi encontrado fora deste intervalo, com número de publicações 18% menor que o ano anterior (2015). Ainda assim, este comportamento não foi propagado nos anos seguintes, caracterizando um ano atípico na amostra.

## Quantidade de periodicos publicados por professor(a) entre 2010 e 2017



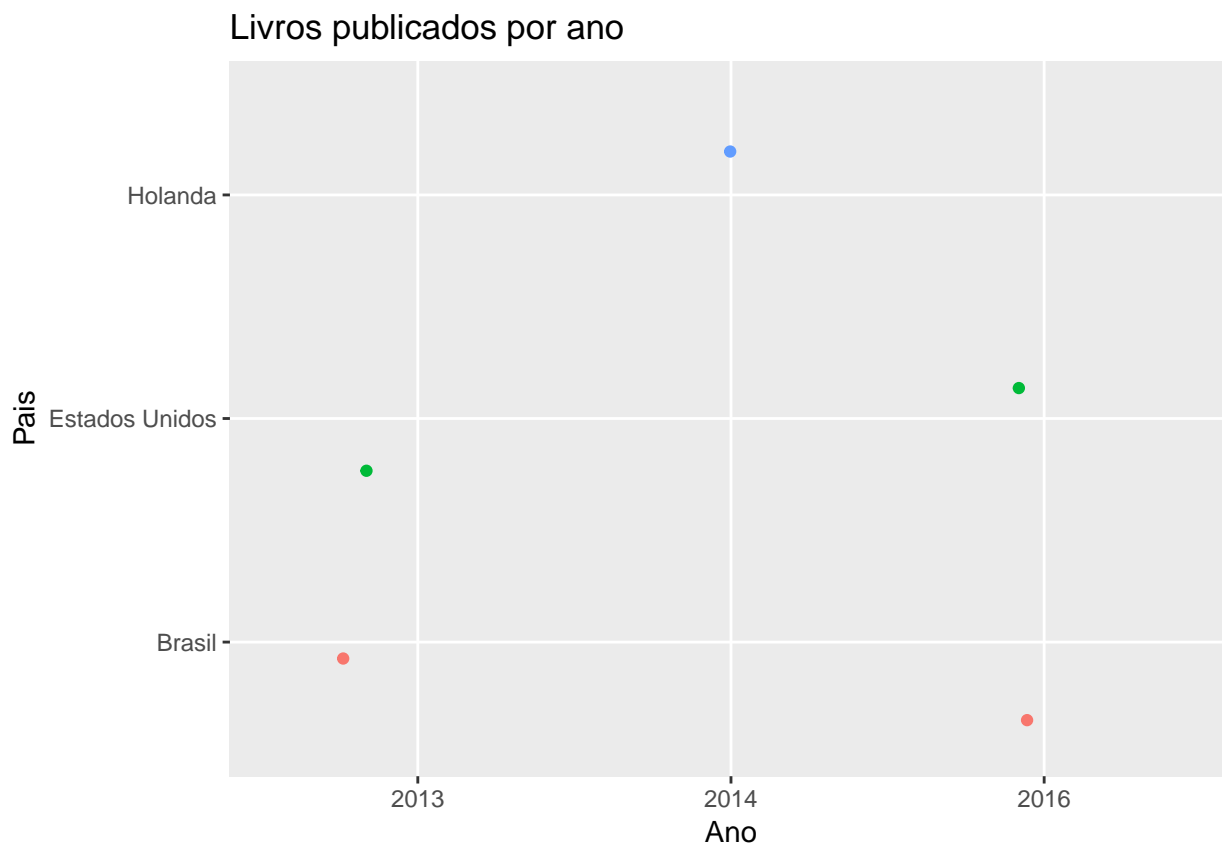
O gráfico acima demonstra o total de publicações registradas por cada pesquisador, demonstrando relativa homogeneidade na produção média por pesquisador, uma vez que grande parte das amostras se aproxima da mediana (26 publicações em periódicos). Nota-se, entretanto, um grande destaque para dois pesquisadores que apresentaram mais de 100 publicações no período (cerca de quatro vezes o valor da mediana encontrada). Outra consideração relevante é a constatação de que os 3 pesquisadores com maiores coeficientes no gráfico apresentam apenas uma única área de pesquisa cadastrada, enquanto os pesquisadores com mais áreas de pesquisa apresentaram um índice menor.

## Número de áreas de pesquisa e quantidade de eventos por pesquisador



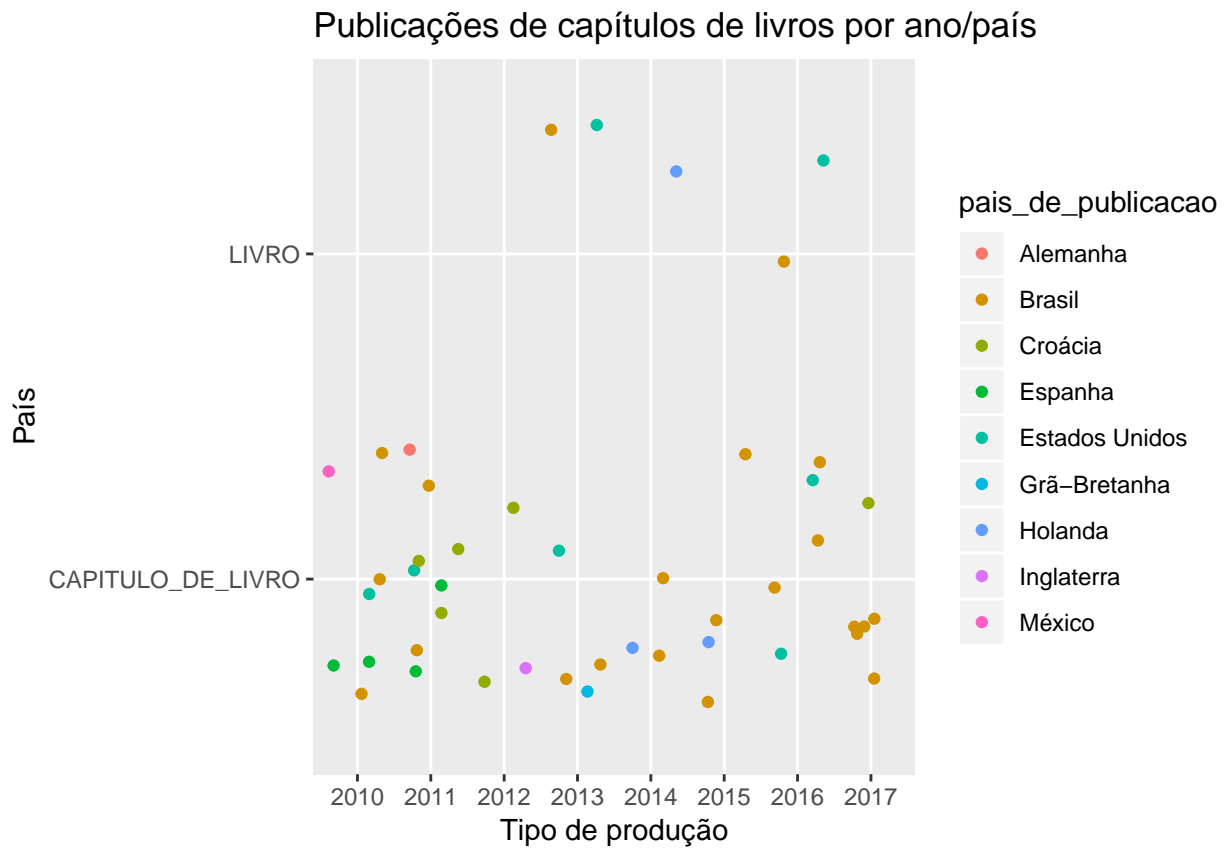
O gráfico acima demonstra o total de eventos registrados por cada pesquisador, demonstrando relativa homogeneidade na participação média de cada pesquisador em eventos, uma vez que grande parte das amostras se aproxima da mediana (11 participações). Ainda assim, alguns *outliers* apresentaram número expressivamente maiores, chegando a quintuplicar a mediana encontrada. É curioso notar que, assim como observado no gráfico de produção de periódicos, os pesquisadores que apresentaram um maior coeficiente de participação em eventos não foram aqueles que relataram um maior número de áreas de pesquisa, induzindo a noção de que pesquisadores especializados apresentam uma maior participação mais expressiva na comunidade acadêmica.

## Publicação de livros por país/ano



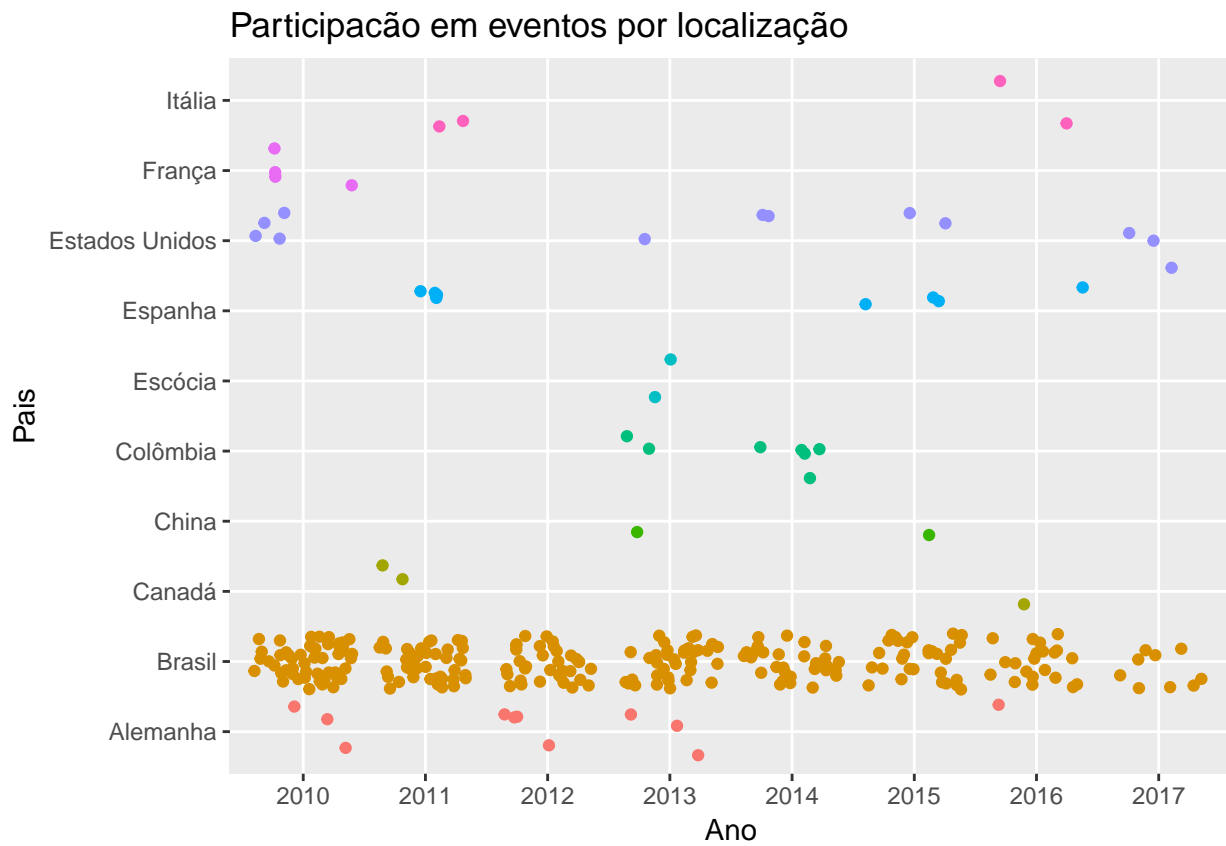
Denota-se homogeneidade na distribuição de livros publicados para cada país de publicação, onde os Estados Unidos se equiparam ao Brasil quanto a número de livros publicados para o PPG em Biologia Animal. Ainda assim, é curioso notar que menos da metade dos oito anos da base de dados registraram publicações de livros (2013, 2014, 2016).

## Publicações de capítulos de livros por ano/país



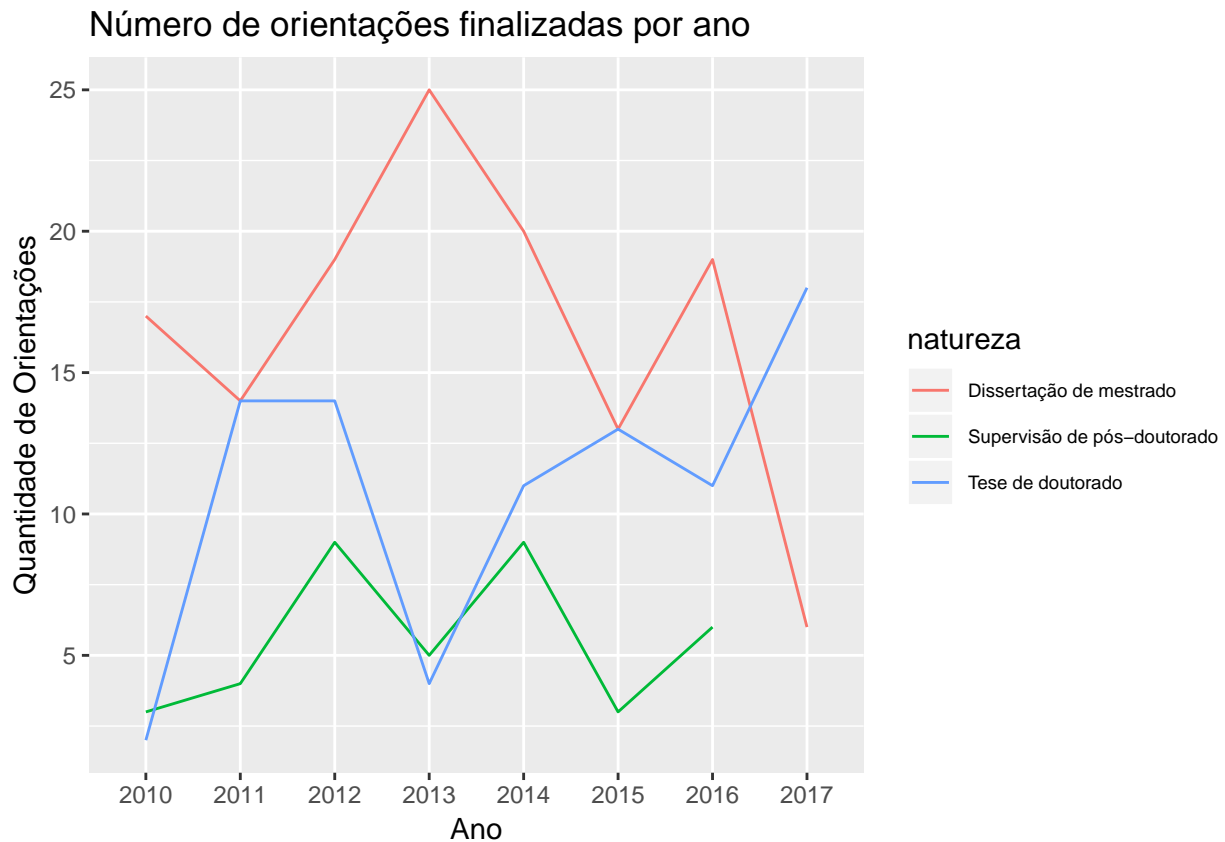
Trazendo a análise de publicações de capítulos de livros, há um aumento expressivo no número de amostras. Todos os anos analisados (2010-2017) registraram mais de um capítulo de livro publicado pelo PPG em Biologia Animal. Entretanto, a publicação de capítulos não apresenta o mesmo nível de internacionalização que o gráfico anterior. Uma observação curiosa é a grande participação de países estrangeiros no início da amostra, enquanto os anos 2016 e 2017 tem publicações majoritariamente nacionais.

## Eventos nacionais e internacionais



Considerando eventos, o Programa de Biologia Animal teve comparecimento maior em eventos no Brasil que em países estrangeiros. Ainda assim, o gráfico mostra que o programa esteve presente em eventos de sede internacional em todos os anos registrados na base de dados, principalmente nos Estados Unidos.

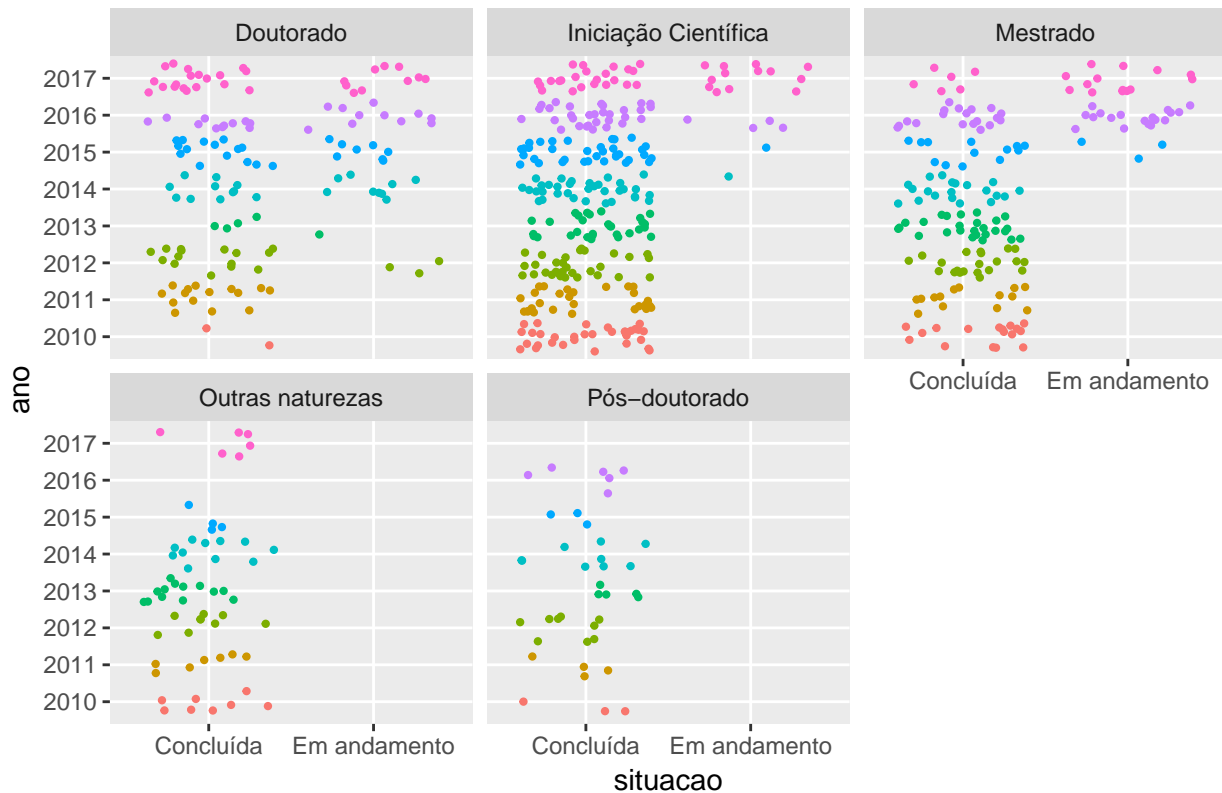
## Orientacoes completas por ano e natureza



Observando a evolução do número de orientações completas ao longo dos anos, percebe-se que o Programa de Pós-Graduação em Biologia Animal cresceu na natureza de mestrado após 2010, mas teve regressões em 2015 e 2017. Entretanto, não há um comportamento linear na evolução do número de orientações. Além disso, 2013 despontou no número de dissertações de mestrado, mas teve queda vertiginosa nas demais naturezas. Enquanto as orientações de mestrado do Programa parecem bem estabelecidas, as supervisões de pós-doutorado ainda parecem incipientes, tendo em vista que não despontaram em nenhum dos anos pesquisados e chegaram a zero registros em 2017 - enquanto o doutorado atingiu valor máximo neste mesmo ano.

## Evolução temporal de orientações em andamento e concluídas

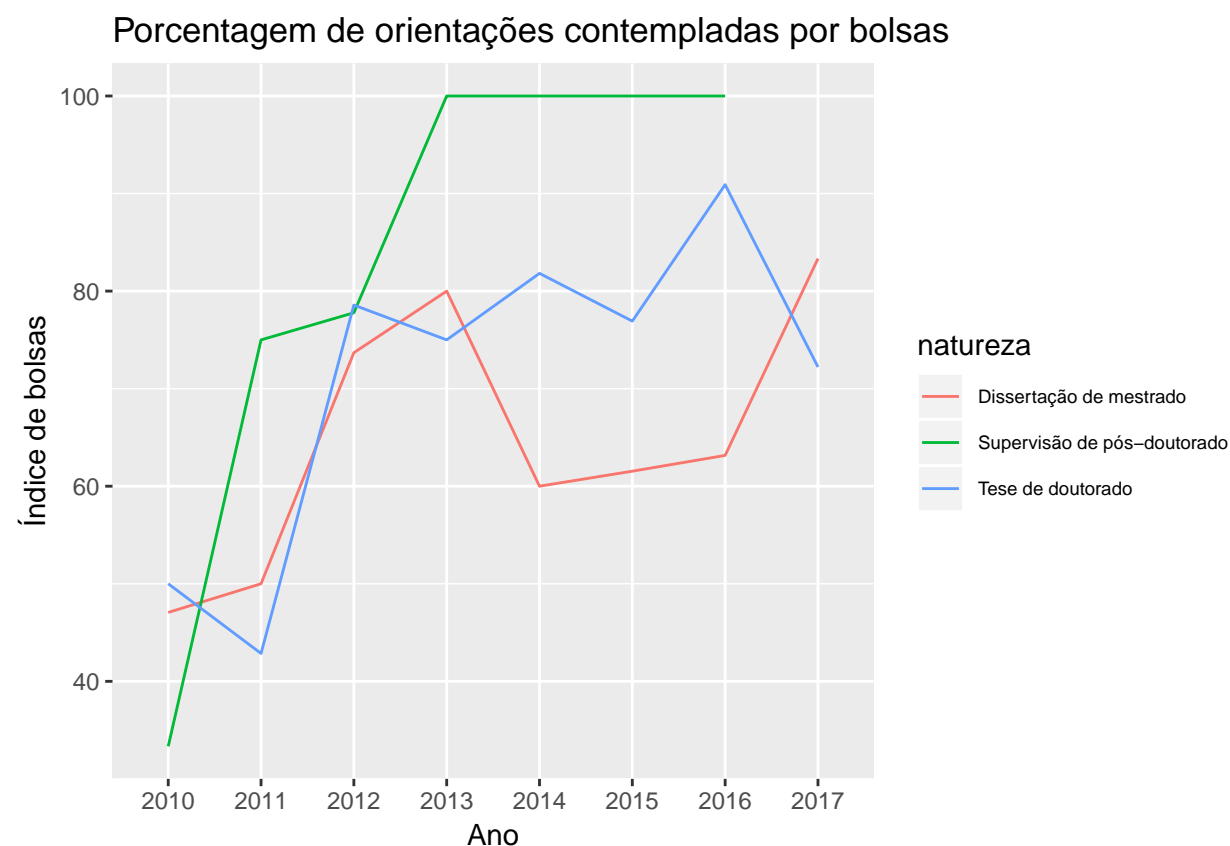
### Disposição de orientações por ano



Observou-se uma quantidade praticamente linear no número de iniciações científicas, mestrados e doutorados concluídos no período 2010 – 2017. Conjuntamente a isso, verifica-se um aumento significativo no número de pós-graduações e iniciações científicas em andamento a partir de 2012. Em contrapartida, os pós-doutorados e outros tipos de especialidades tiveram queda no período em questão. Este comportamento, com diversas orientações finalizadas e poucas em andamento, sugere que o PPG em Biologia Animal contemplava poucos cadastros de orientações em andamento e está intensificando esta prática nos anos mais recentes - o que tornará análises futuras ainda mais enriquecedoras.

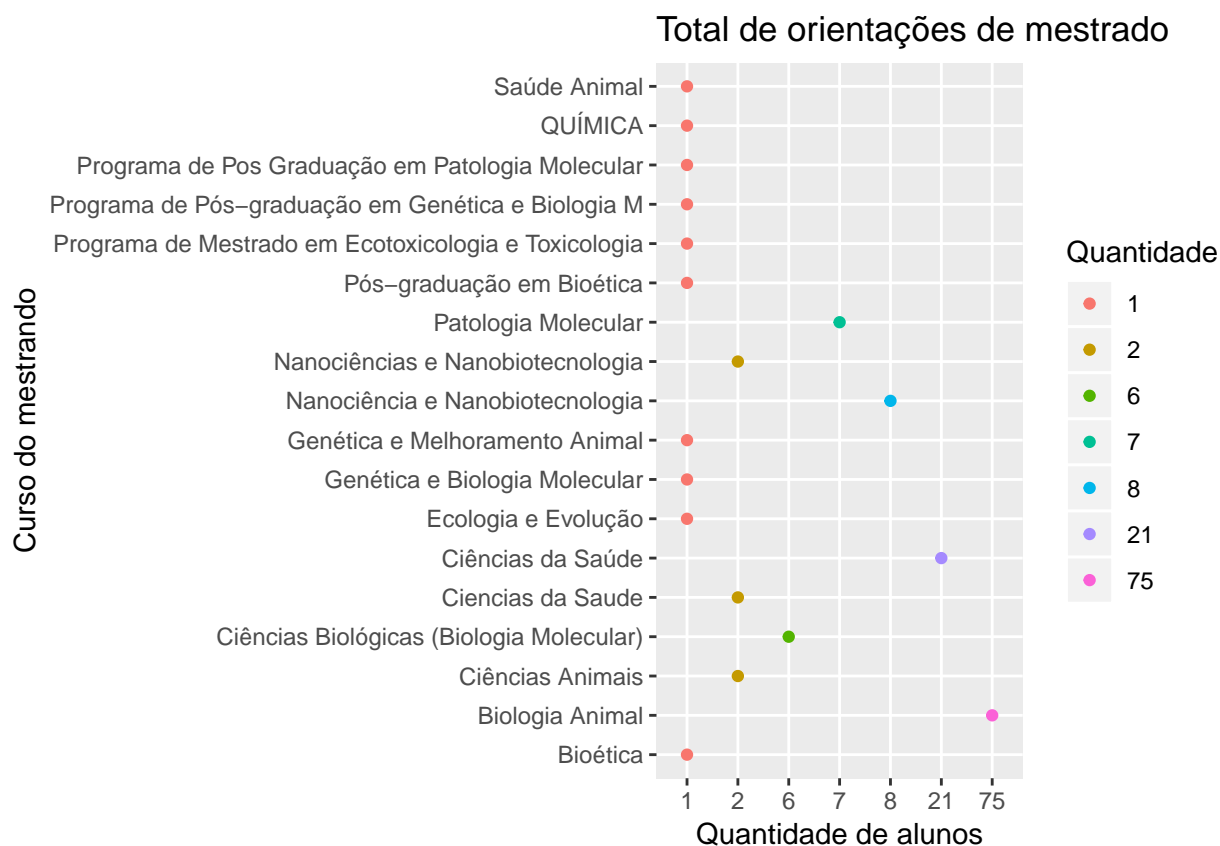


## Índice de bolsas entre naturezas e evolução temporal



Comparando os gráficos de orientações completas e de bolsas, é possível perceber que o número de bolsas oferecidas para o Programa acompanhou o total de orientações de maneira satisfatória ao longo dos anos. Durante todo o período, todas as naturezas apresentaram um índice de pelo menos 50% de bolsas, chegando a 100% em alguns casos. Além disso, as teses de pós-doutorado se mostram a natureza de pesquisa melhor contemplada pelas agências financiadoras, visto que apenas uma das observações não recebeu bolsa. Por fim, é possível observar que o pico no número de orientações de mestrado registrado em 2014 não foi tão expressivo em número de bolsas, caracterizando um ano com elevado número de alunos não bolsistas.

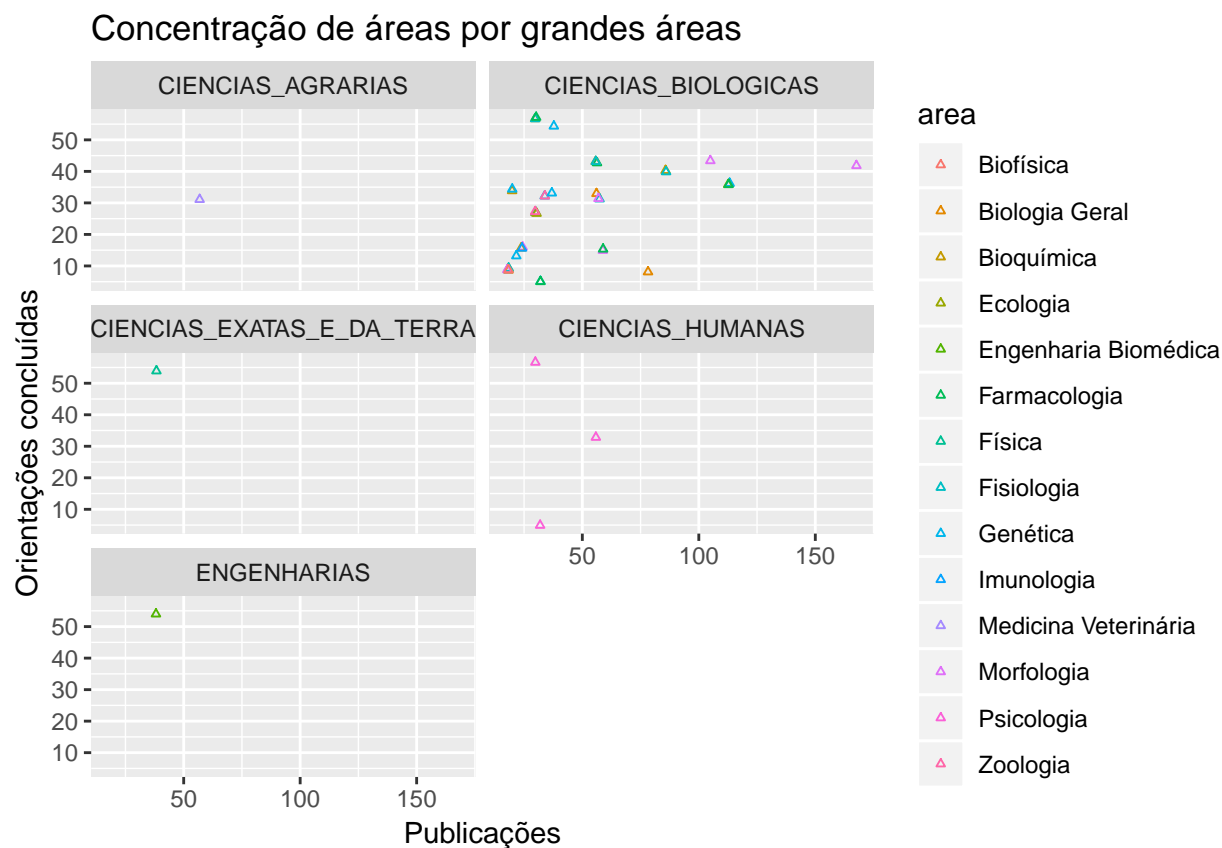
## Cursos de Mestrado e total de alunos no período (2011-2017)



A análise de cursos dos alunos de mestrado orientados pelos professores do PGG de Biologia Animal apresenta um pequeno índice de correlação com outras áreas. No gráfico são listados diversos programas de pós-graduação com orientandos cadastrados - não se limitando aos que foram analisados neste trabalho.

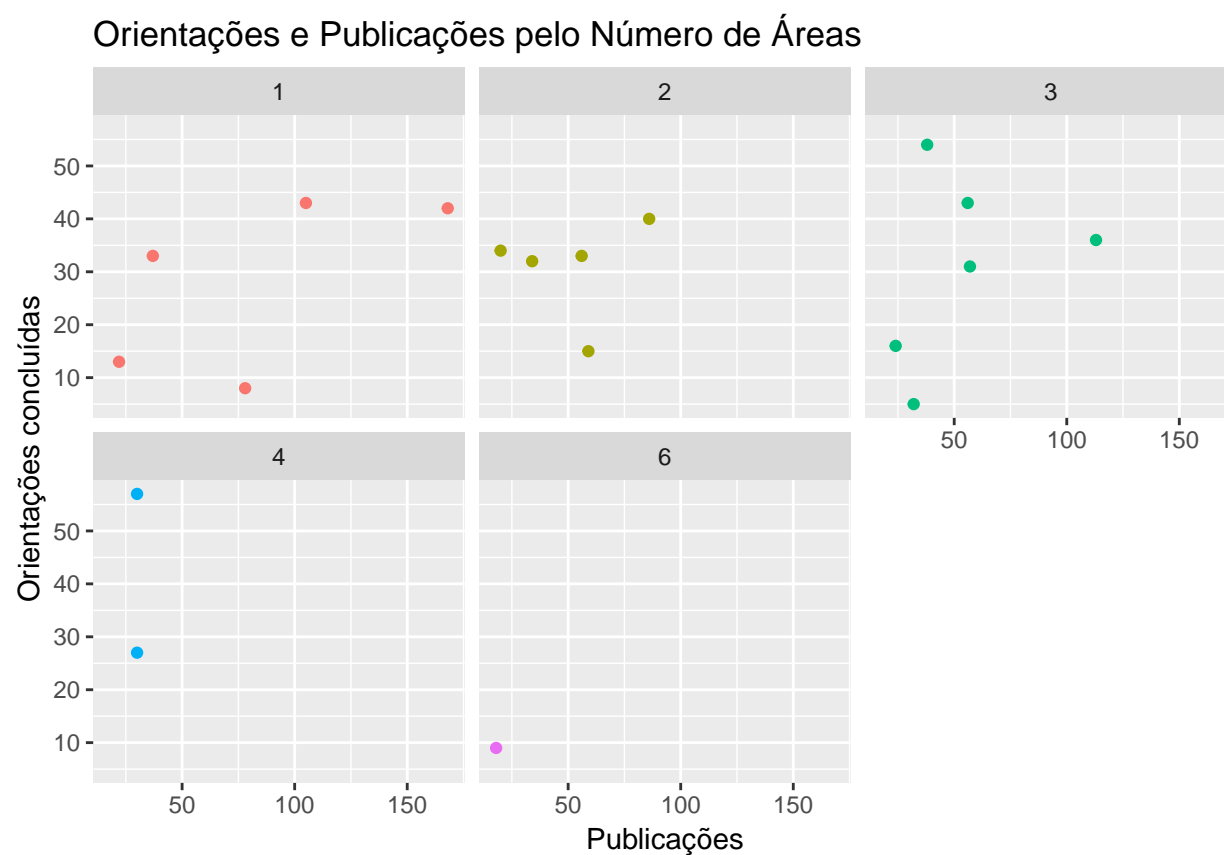
O curso com maior número de orientações do programa é ele mesmo, como esperado (75 orientações). Em segundo lugar está o curso de Ciências da Saúde, com 21 orientações. Dos Programas de Pós-Graduação analisados neste trabalho, o que apresentou maior número de orientações foi o de Patologia Molecular, com apenas 7 registros. Os gráficos produzidos sobre os demais PPGs contemplados por este trabalho apresentaram índices de correlação muito maiores entre si, indicando um certo distanciamento da Biologia Animal dos demais programas.

Concentração de áreas por grandes áreas:



Uma análise superficial do gráfico apresentado já evidencia uma grande concentração de pesquisadores do PPG de Biologia Animal na grande área de Ciências Biológicas. Em comparação com os demais PPGs analisados neste trabalho, o programa de Biologia Animal se destaca pela presença de poucos docentes com pesquisas em outras grandes áreas. Estes poucos docentes não tiveram grande expressão em número de publicações, mas se destacaram no número de orientações - os docentes que apresentaram maior número de orientações na amostra trabalhada estão justamente entre aqueles que se diversificaram quanto à grande área.

## Quantidade de Orientações e Publicações pelo Número de Áreas



Trazendo uma perspectiva sobre as áreas de pesquisa, há uma diversidade um pouco maior entre os docentes do PPG de Biologia Animal, mas poucos pesquisadores contemplam mais de 3 áreas. Além disso, nenhum dos pesquisadores que ultrapassaram esta marca teve destaque em número de publicações. Os destaques em número de publicações se concentram principalmente em pesquisadores com poucas áreas de pesquisa, comportamento este que se repetiu na análise de outros PPGs.

### Orientações concluídas x Participação em congressos internacionais

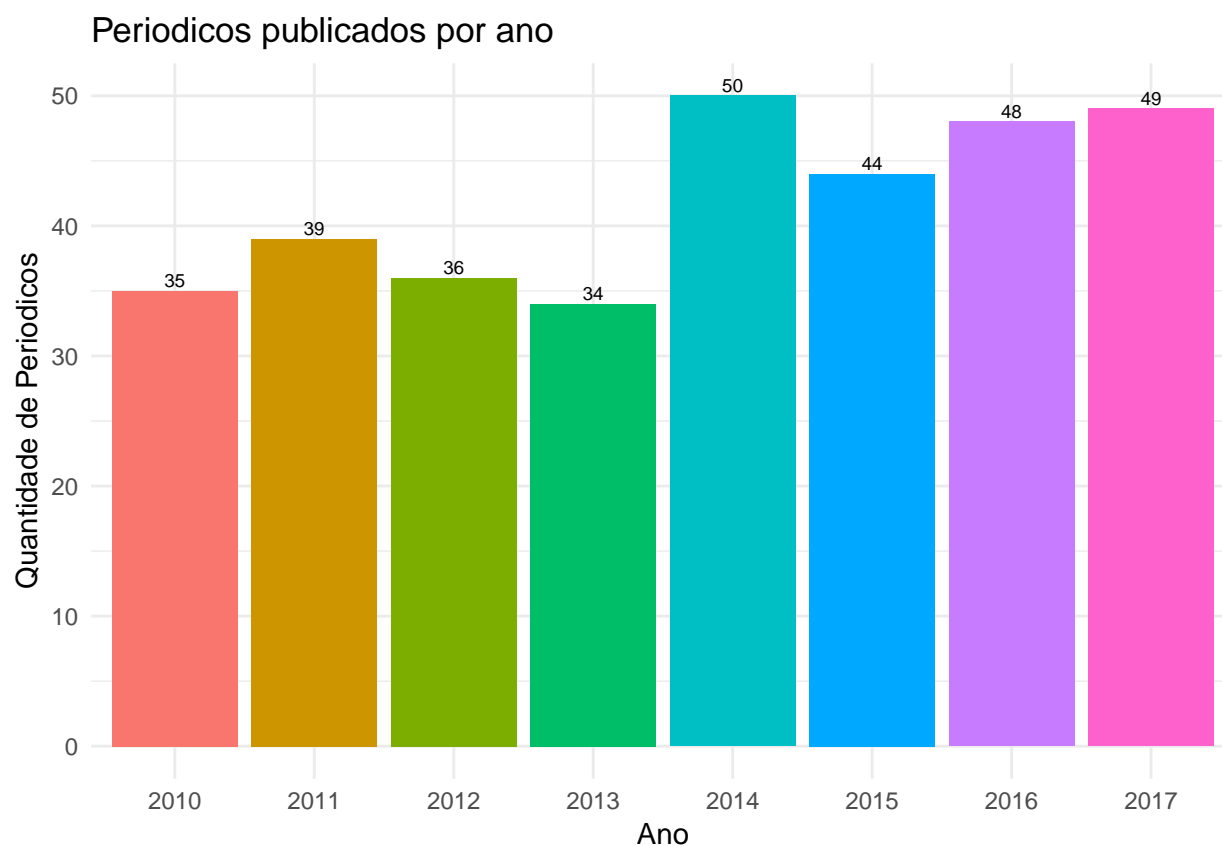
The scatter plot displays the relationship between international congress participation and the number of completed orientations. The x-axis, labeled 'Participacao internacional', has two categories: 'Não' (No) and 'Sim' (Yes). The y-axis, labeled 'Orientações concluídas', ranges from 0 to 60. Data points are represented by colored dots (blue, orange, green, pink, and purple). The plot shows a general upward trend, with more completed orientations observed for those who participated in international congresses, although there is significant individual variation.

Participacao internacional	Orientações concluídas
Não	13
Não	34
Não	58
Sim	8
Sim	16
Sim	36
Sim	54
Sim	55
Sim	56
Sim	57
Sim	58
Sim	42
Sim	41
Sim	43
Sim	40
Sim	32
Sim	31
Sim	33
Sim	32
Sim	33
Sim	34
Sim	31
Sim	32
Sim	33
Sim	34
Sim	35
Sim	36
Sim	37
Sim	38
Sim	39
Sim	40
Sim	41
Sim	42
Sim	43
Sim	44
Sim	45
Sim	46
Sim	47
Sim	48
Sim	49
Sim	50
Sim	51
Sim	52
Sim	53
Sim	54
Sim	55
Sim	56
Sim	57
Sim	58
Sim	59
Sim	60

53

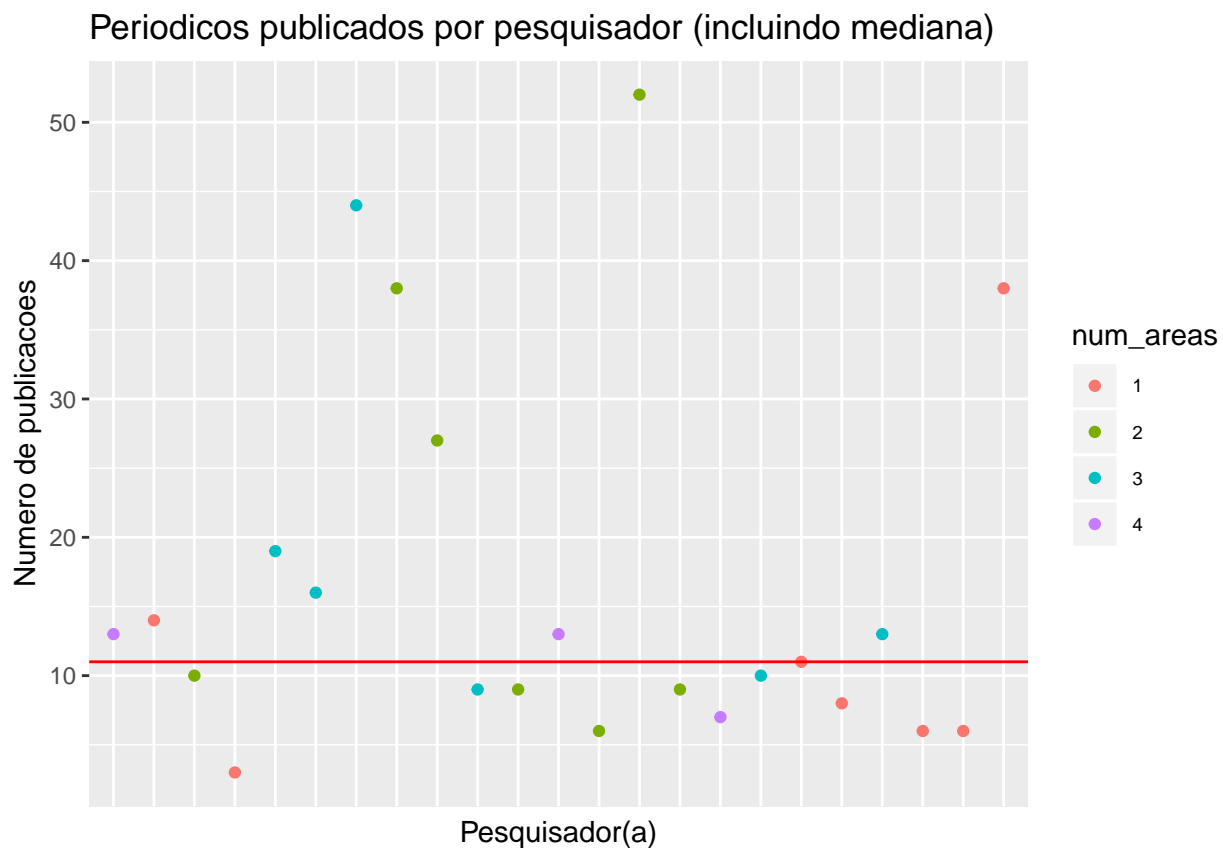
## Biologia Microbiana

Grafico de barras; periodicos por ano



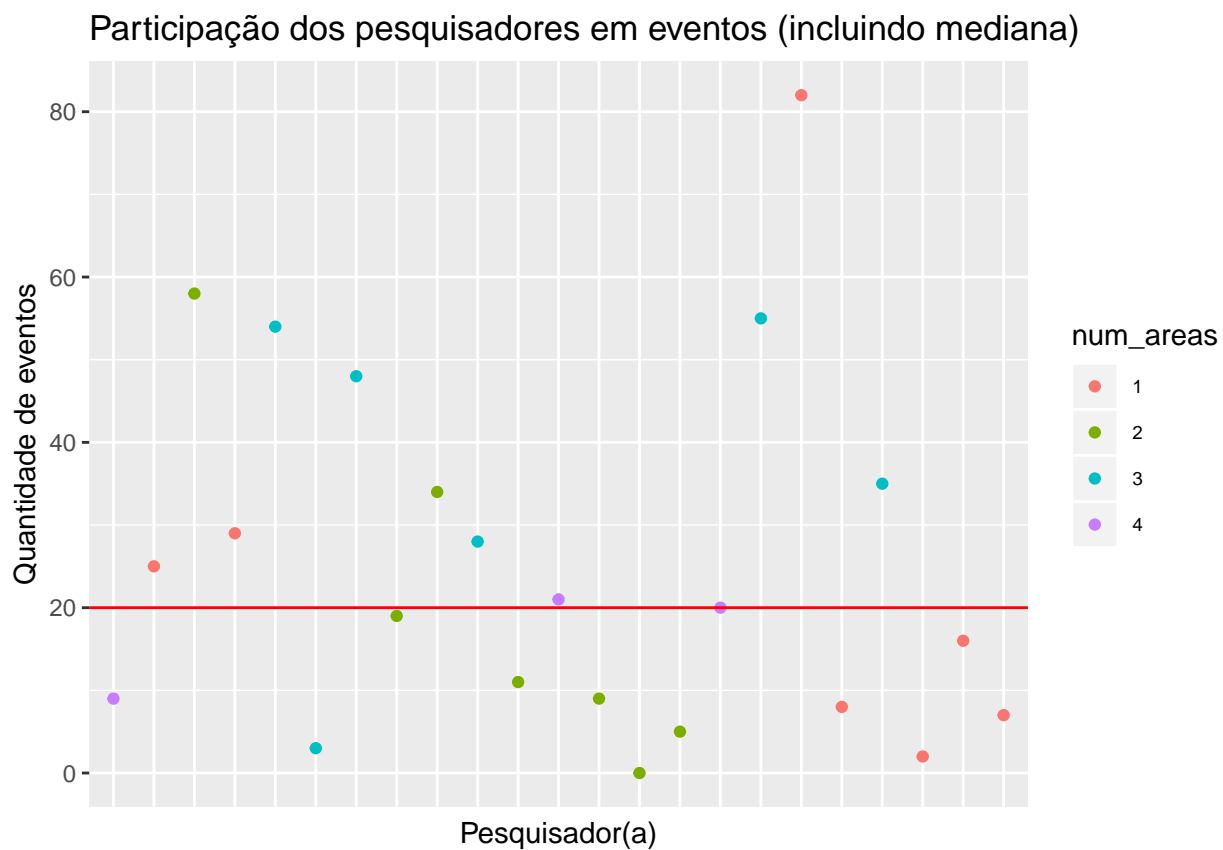
No programa de Biologia Microbiana verifica-se que há um aumento na quantidade de publicação de periódicos por ano durante o período apresentado. Houve um considerável aumento no número de periódicos entre 2013 e 2014, crescendo quase duas vezes em sua quantidade.

Quantidade de periodicos publicados por professor(a) entre 2010 e 2017



Observa-se no gráfico que não existe nenhum padrão linear entre Número de Publicações e Numero de Áreas do pesquisador. Verifica-se apenas que quando o pesquisador está apenas em uma área de pesquisa tende a estar perto da mediana de publicações e quanto mais alto o número de áreas do pesquisador mais se observa uma aleatoriedade no número de publicações. Além do mais, há uma diferença do pesquisador que mais publicou em relação a mediana de cinco vezes mais publicações.

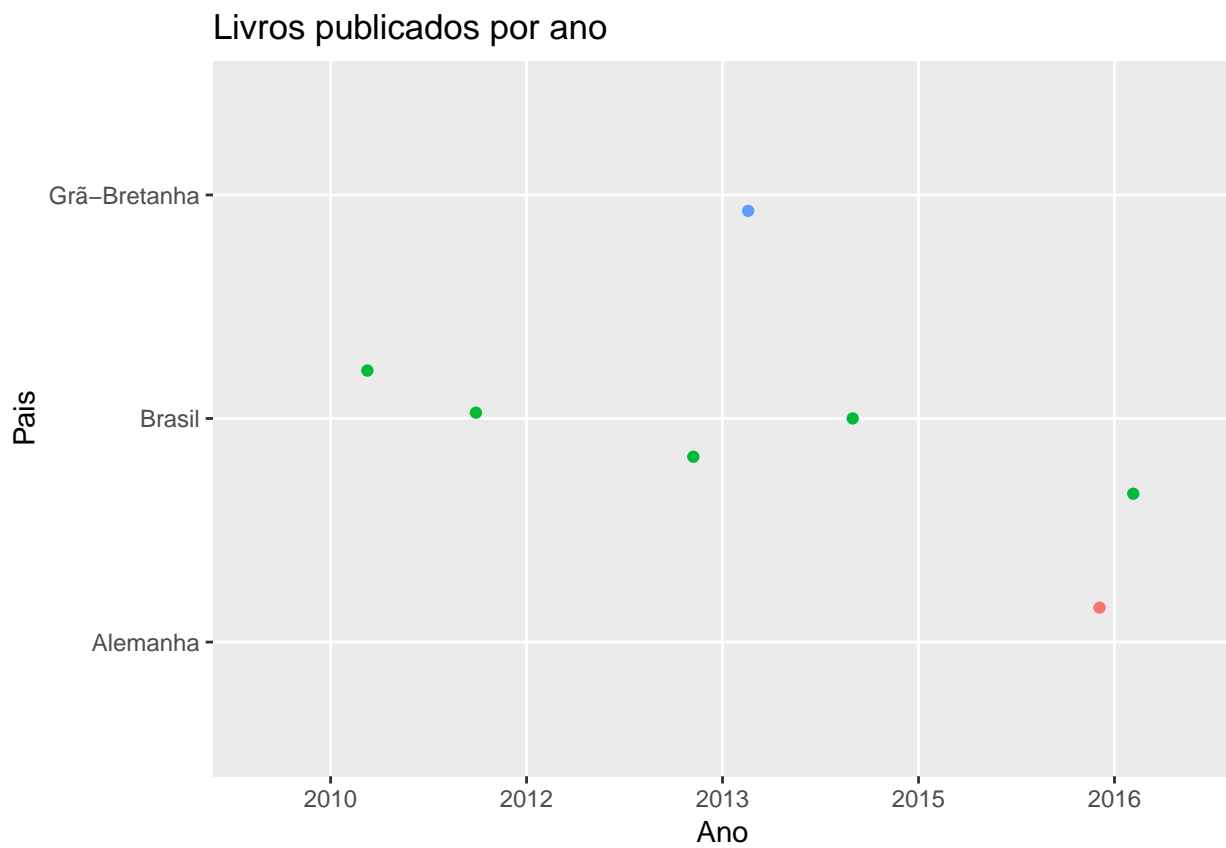
## Número de áreas de pesquisa e quantidade de eventos por pesquisador



Como constatado também no gráfico anterior, não há um padrão definido entre a Quantidade de Participação em Eventos e o Número de Áreas em que cada pesquisador está inserido. A quantidade fica em torno de 20 eventos e há um pesquisador um pouco discrepante aos demais que foi em mais de 80 eventos durante o período.

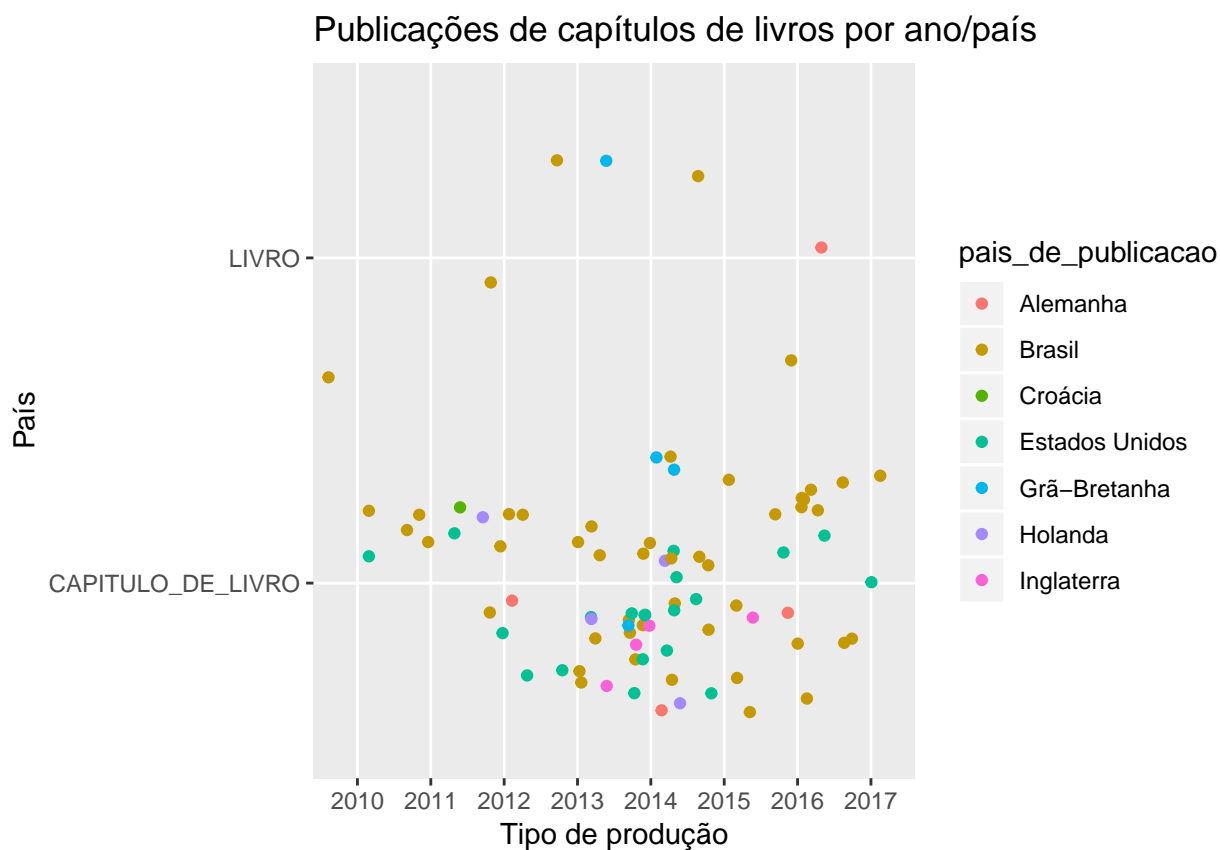


## Publicação de livros por país/ano



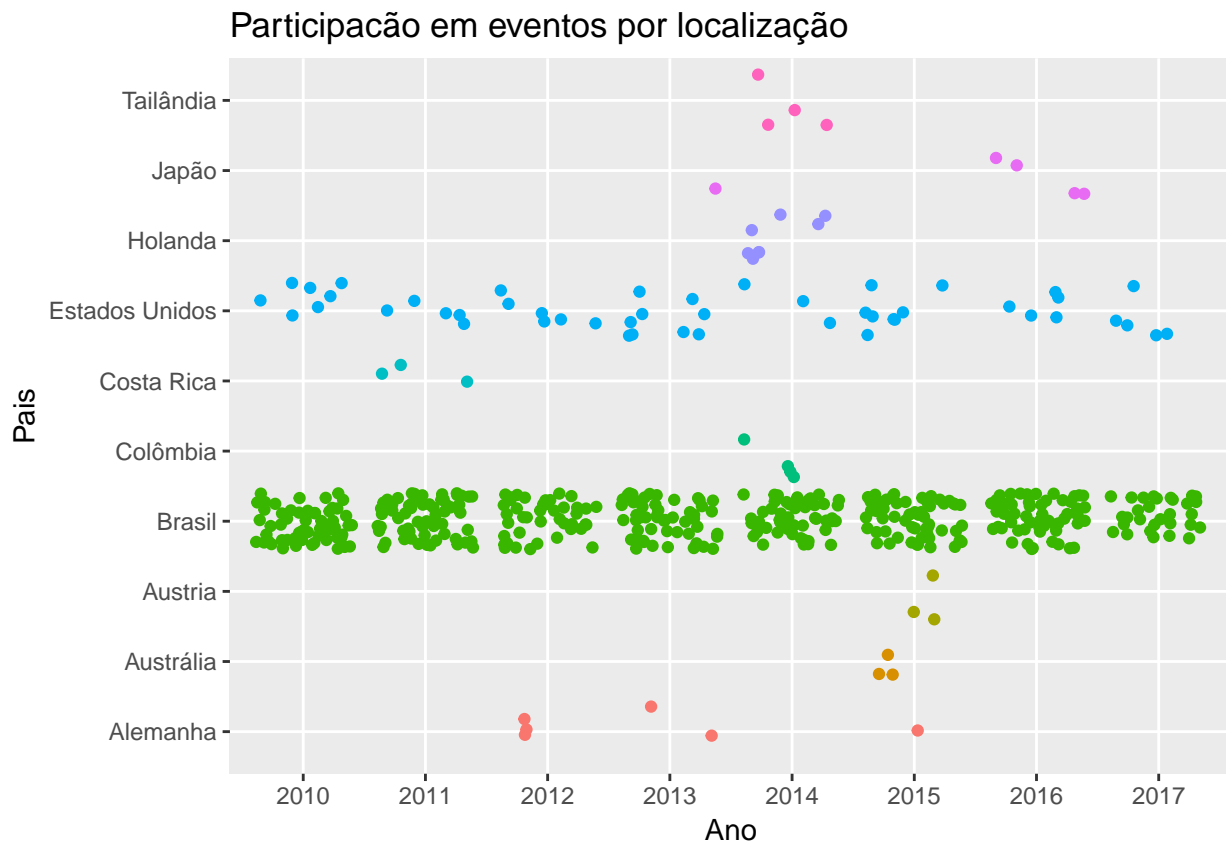
Em relação a publicação de livro, houve uma maior quantidade de publicações feitas no Brasil. Sendo divulgadas apenas em 2013 e 2015 na Grã-Bretanha e na Alemanha, respectivamente.

## Publicações de capítulos de livros por ano/país



Os docentes do programa de Biologia Microbiana apresentaram uma quantidade muito boa de publicações de capítulo de livros durante estes 7 anos. Houve uma maior concentração de publicações em 2014, com destaque para Brasil, mas não se deixou de publicar nos outros países também como: Alemanha, Croácia, Estados Unidos, entre outros. Já no que se refere a publicação de livros, essa se deu de forma branda e com concentração maior no Brasil.

## Eventos nacionais e internacionais



O Brasil teve em disparado um maior número de participação em eventos feito pelos docentes da área de Biologia Microbiana da Universidade de Brasília. Em segundo lugar ficou a participação em eventos nos Estados Unidos, seguido da Alemanha. Nos outros países não tiveram um volume considerável de participações. Além disso, a participações em eventos ficaram lineares durante os anos nos países, somente havendo um numero um pouco maior de ida a esses eventos em 2014 e 2015.

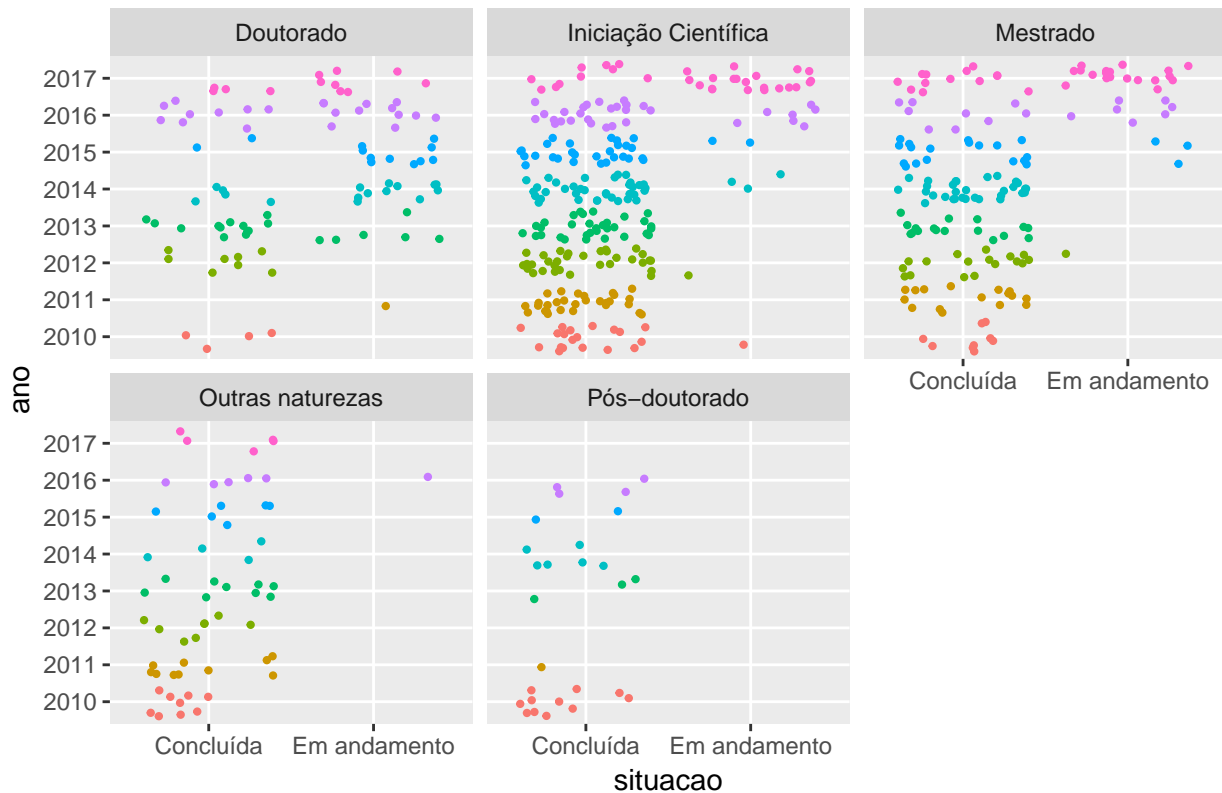
## Orientações completas por ano e natureza



Nota-se um aumento expressivo no número de orientações de mestrado finalizadas de 2010 até 2014, depois disso, houve um declínio nesse número e uma pequena recuperação na quantidade de orientações finalizadas no ano de 2016. Conjuntamente a isto, o número de orientações finalizadas em teses de doutorado em supervisões de pós-graduação teve oscilações durante todo período, apresentando queda de 2016 para 2017 e aumento de 2015 para 2016, em teses de doutorado e supervisão de pós-doutorado, respectivamente.

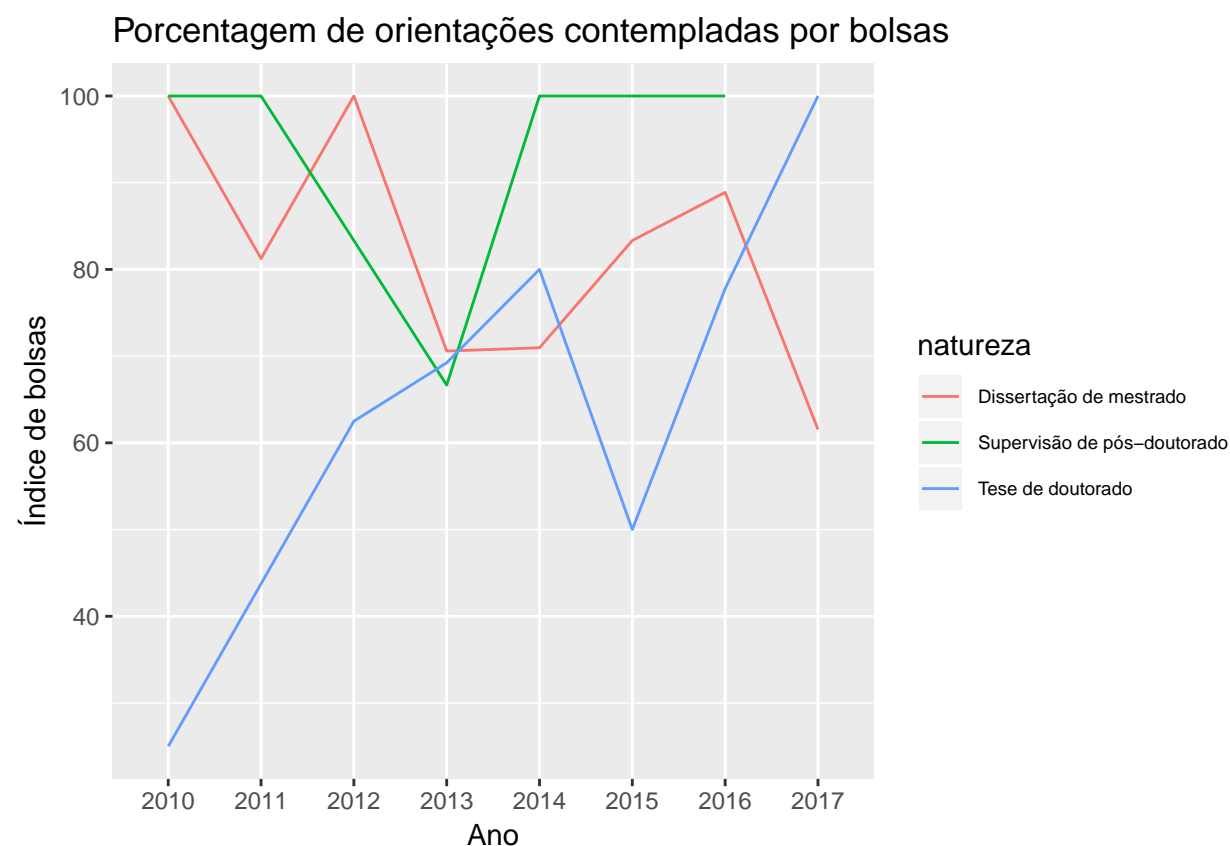
## Evolução temporal de orientações em andamento e concluídas

### Disposição de orientações por ano



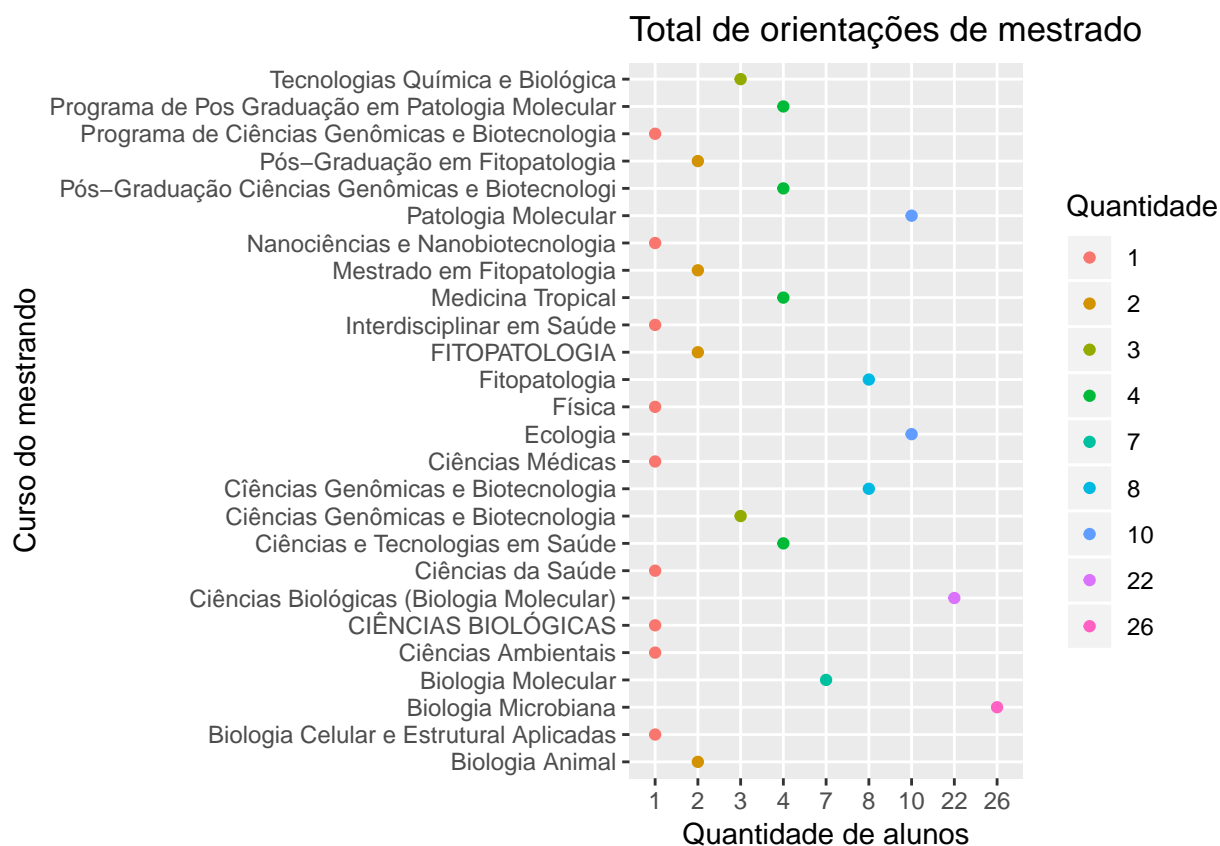
Verifica-se uma maior quantidade de orientações concluídas na Iniciação Científica e no Mestrado do programa de Biologia Microbiana, durante todo o período apresentado. É importante destacar que o Doutorado, a Iniciação Científica e o Mestrado vêm mostrando aumento no número de orientações em andamento. Já os outros campos de pós-graduação e outras naturezas tiveram uma diminuição branca em suas conclusões e quase não apresentam orientações em andamento.

## Índice de bolsas entre naturezas e evolução temporal



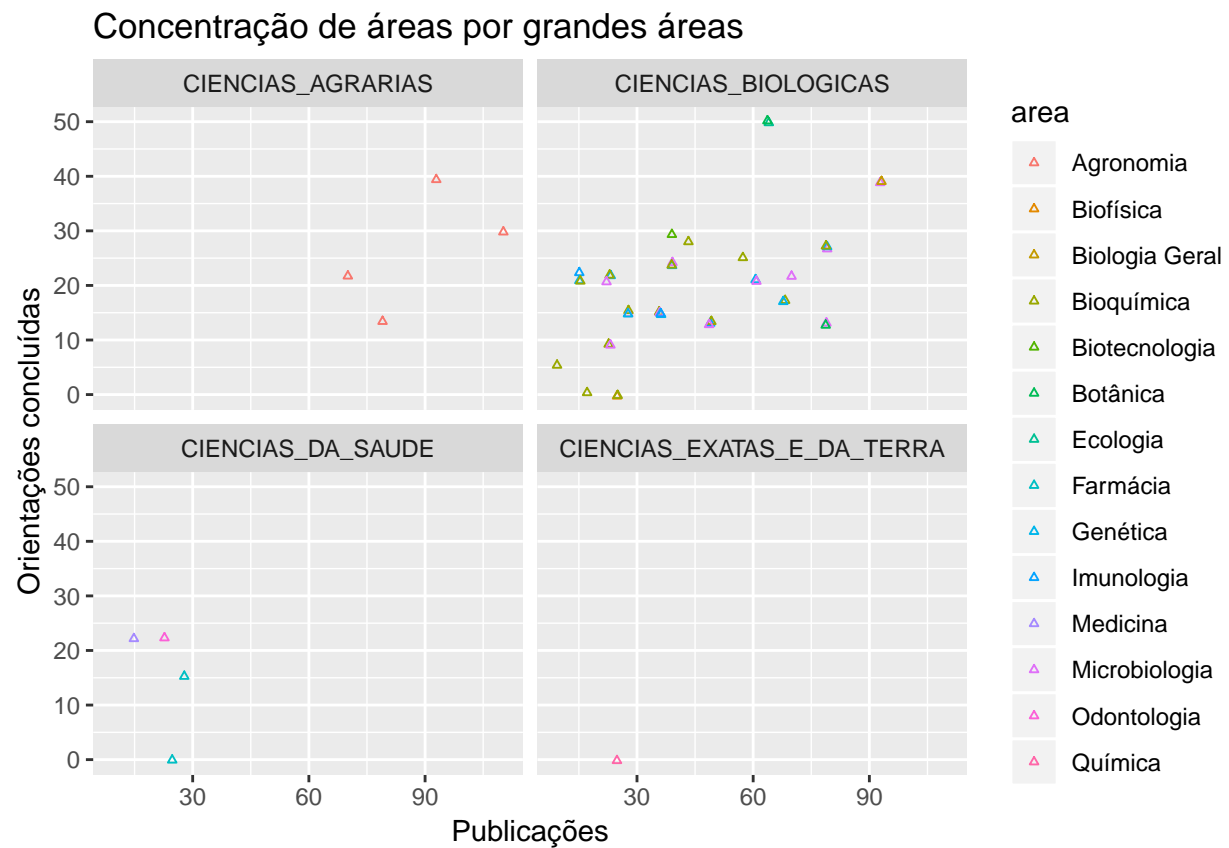
É considerável notar o aumento expressivo no número de bolsas de teses de doutorado, que em 2017 já apresentava um índice de 100%. Ademais, esse índice de bolsas para os discentes de pós-graduação apresentou queda em 2011 mas em 2014 já retomaram o índice de 100%. Entretanto, as bolsas em dissertações de mestrado vêm diminuindo, com oscilações, como mostra o gráfico.

## Cursos de Mestrado e total de alunos no período (2011-2017)



Este gráfico demonstra que para os docentes dentro da PPG de Biologia Microbiana, há alunos de diversas outras PPGs que compõem sua contagem de orientação. Naturalmente a quantidade de alunos para a própria PPG é maior que para outras. Porém é notável que áreas conexas, tal como a Biologia Molecular e a Patologia Molecular, tem muitos alunos sendo orientados por alunos da PPG de Biologia Microbiana.

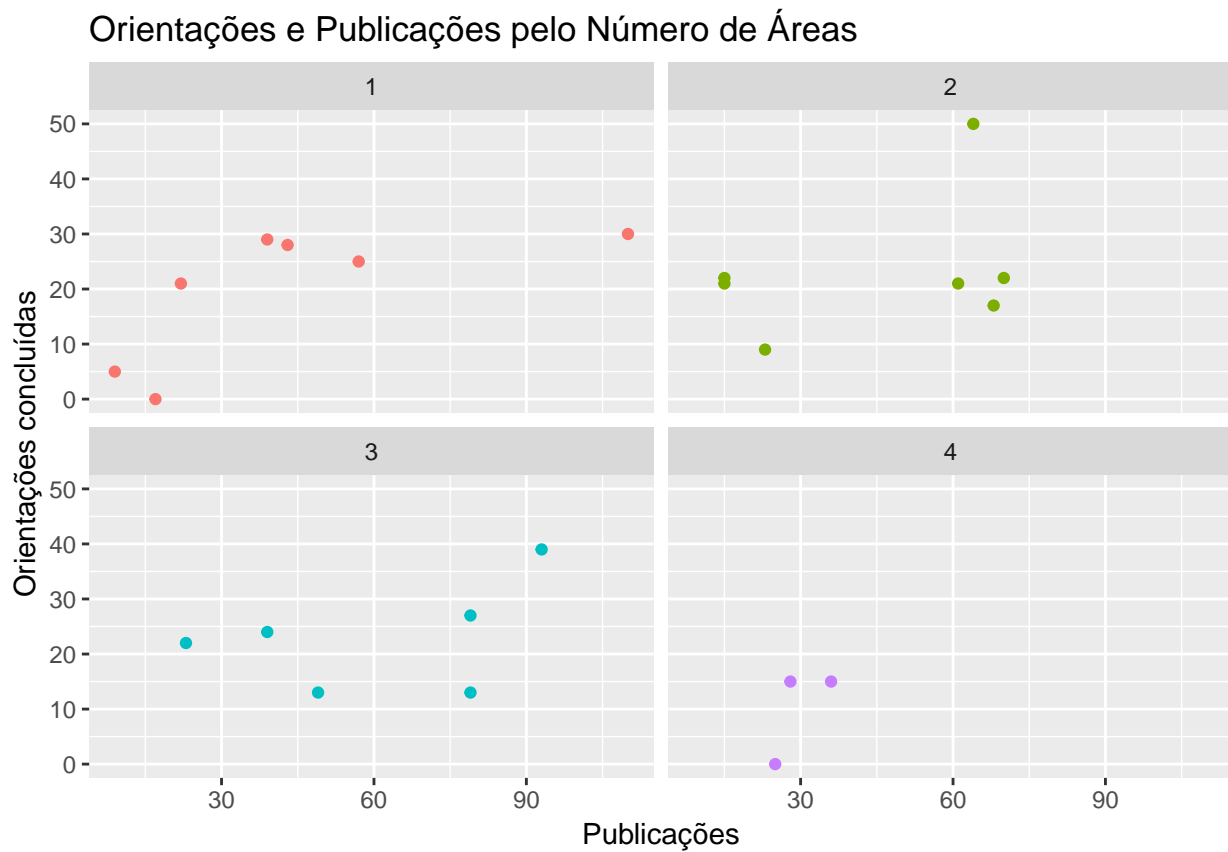
Concentração de áreas por grandes áreas:



É notável que a concentração de áreas esteja mais aglomerada dentro da grande área de Ciências Biológicas, uma vez que esta grande área engloba todas as PPGs aqui estudadas. Por mais que a predominância desta grande área seja opressora sobre as outras, ainda há a presença de outras grandes áreas com áreas específicas nas publicações.

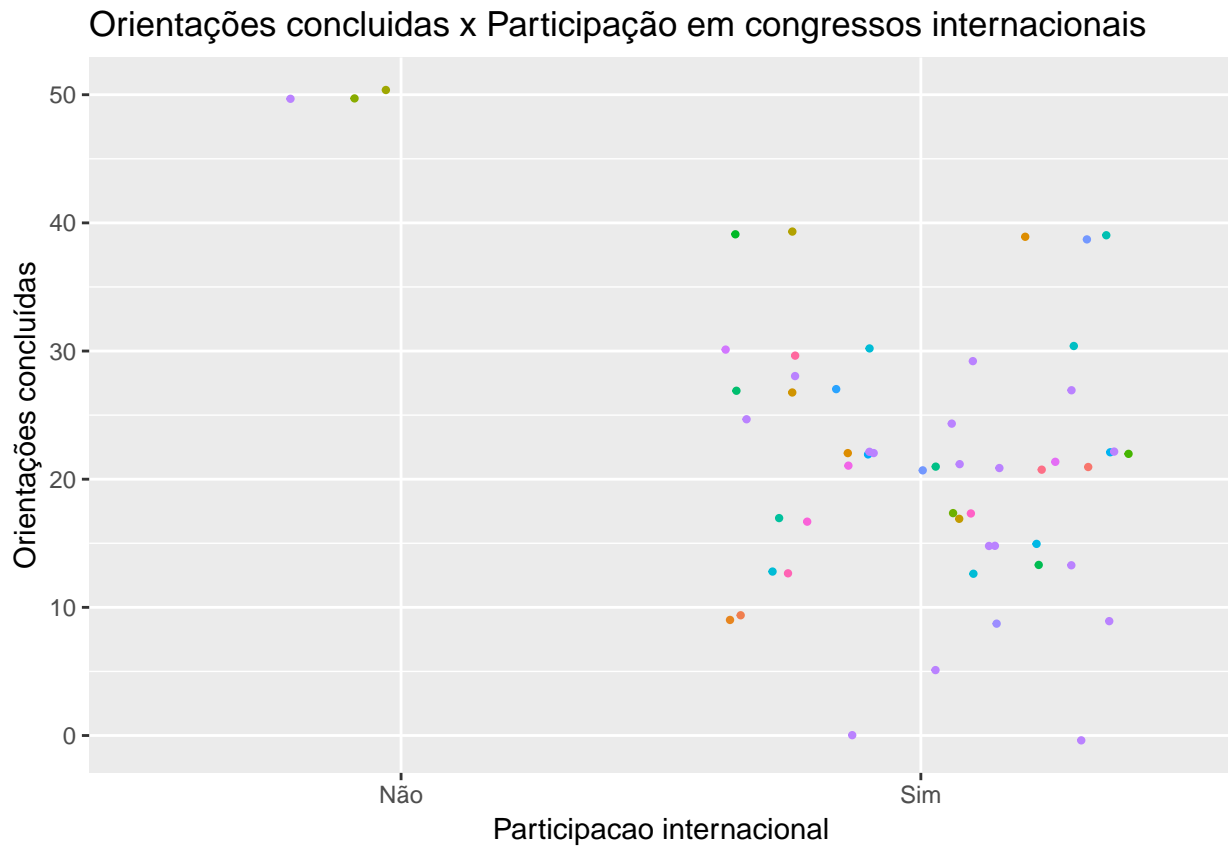


## Quantidade de Orientações e Publicações pelo Número de Áreas



De acordo com a figura anterior, nesta PPG, quem trabalha em mais áreas orienta menos pessoas e publica menos. Há uma notável tendência de diminuição do número de orientações e do número de publicações em pesquisadores que trabalham em cada vez mais áreas.

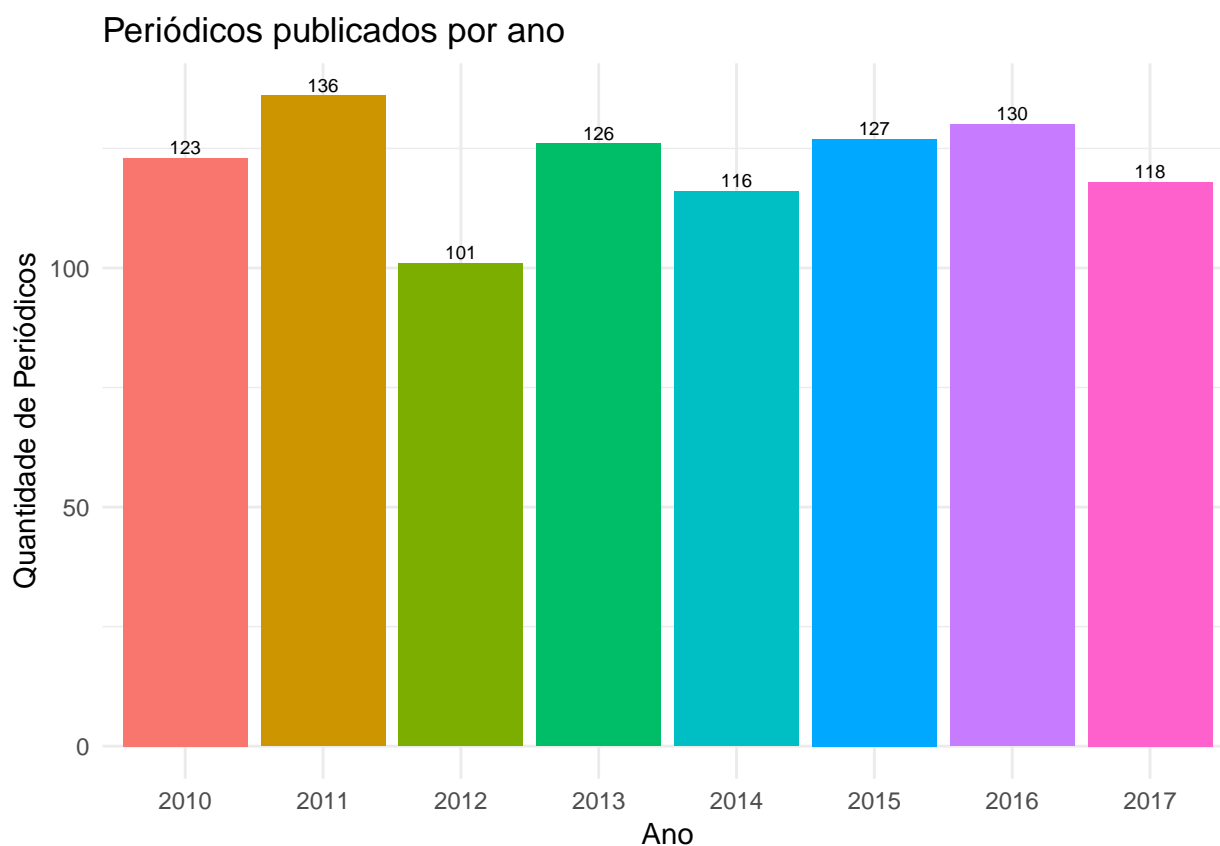
### Presença em congressos internacionais por número de orientações



Por mais que o número de pesquisadores que não participam de congressos internacionais ser muito menor, é possível visualizar que estes conseguem concluir mais orientações que os participantes destes congressos. Isso deve dar-se pela a maior presença deles com seus orientandos, portanto melhor acompanhamento.

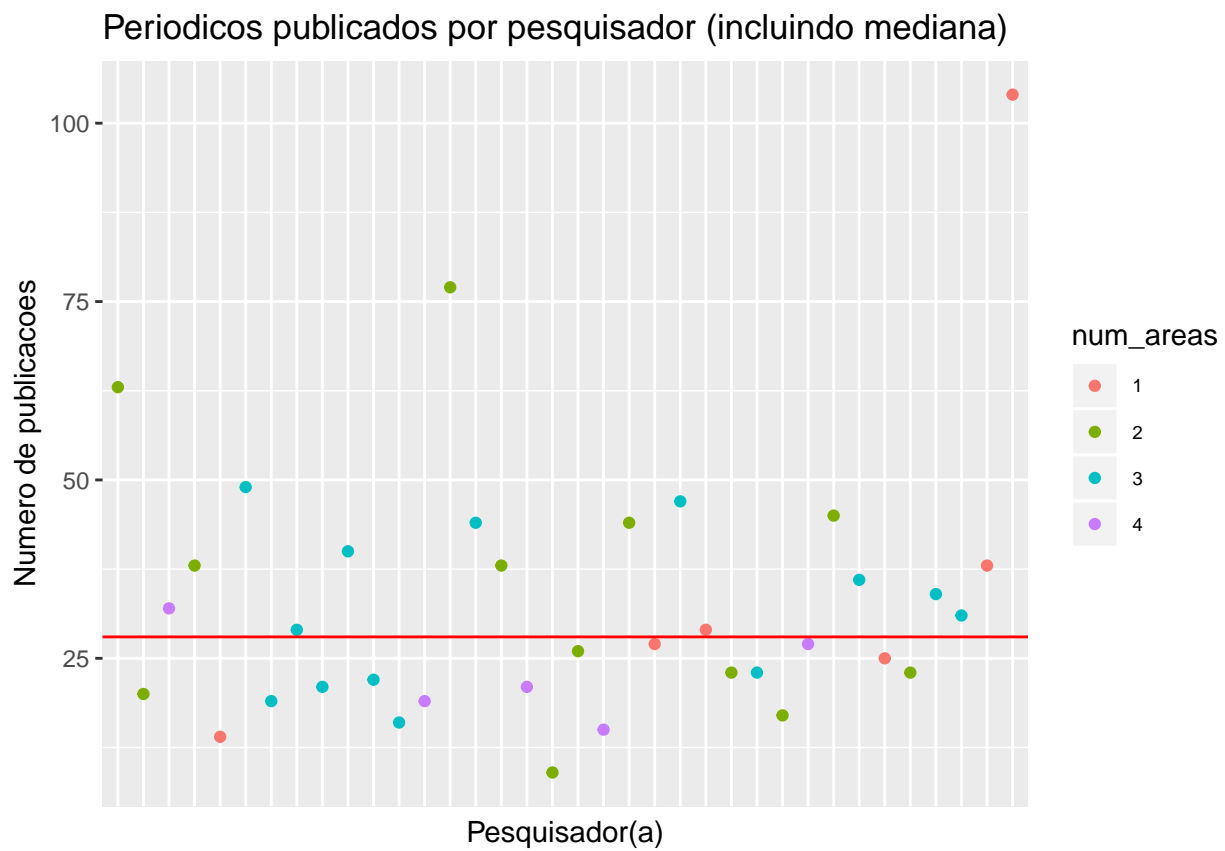
## Biologia Molecular

Gráfico de barras; periodicos por ano



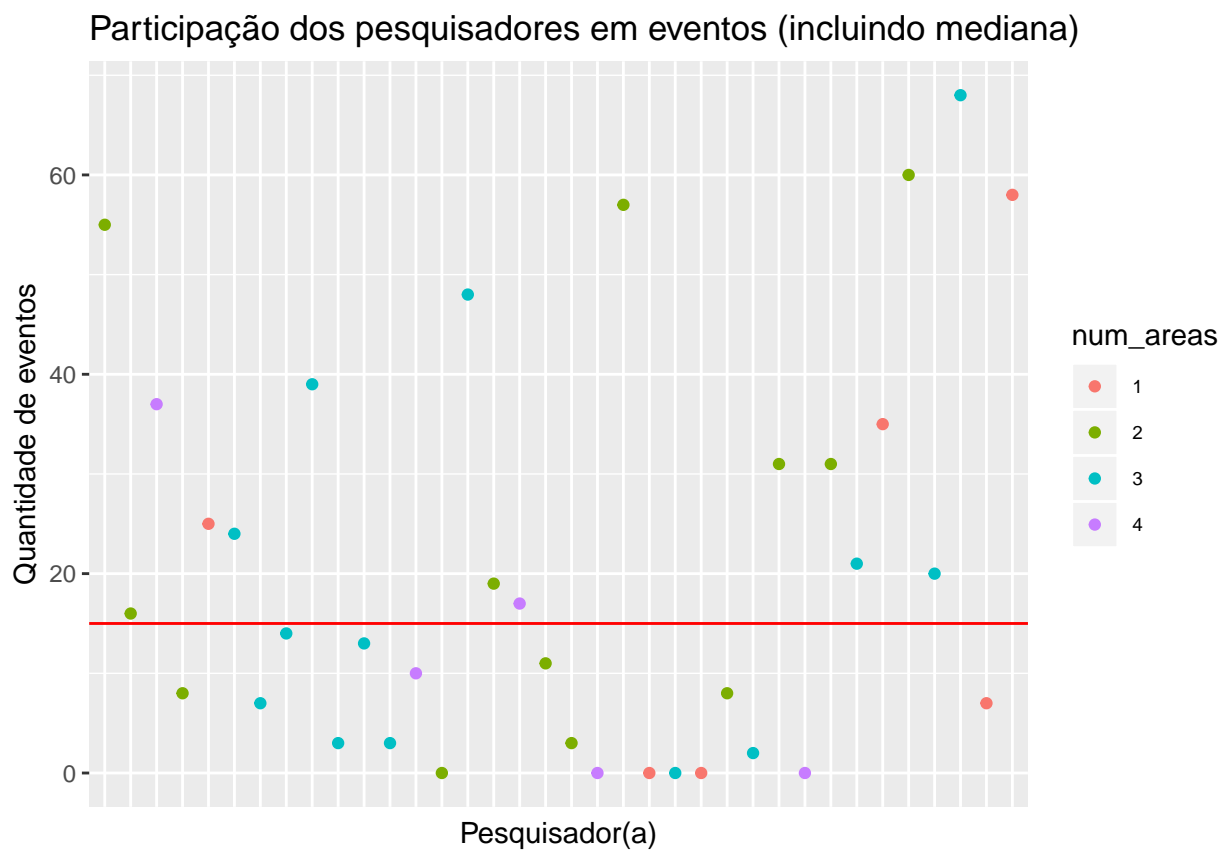
Este gráfico demonstra a quantidade de periódicos por ano para a PPG de Biologia molecular. É possível notar que por uma margem, as publicações tem uma quantidade constante, a despeito de 2012 que foi significativamente menor que os outros anos.

Quantidade de periodicos publicados por professor(a) entre 2010 e 2017



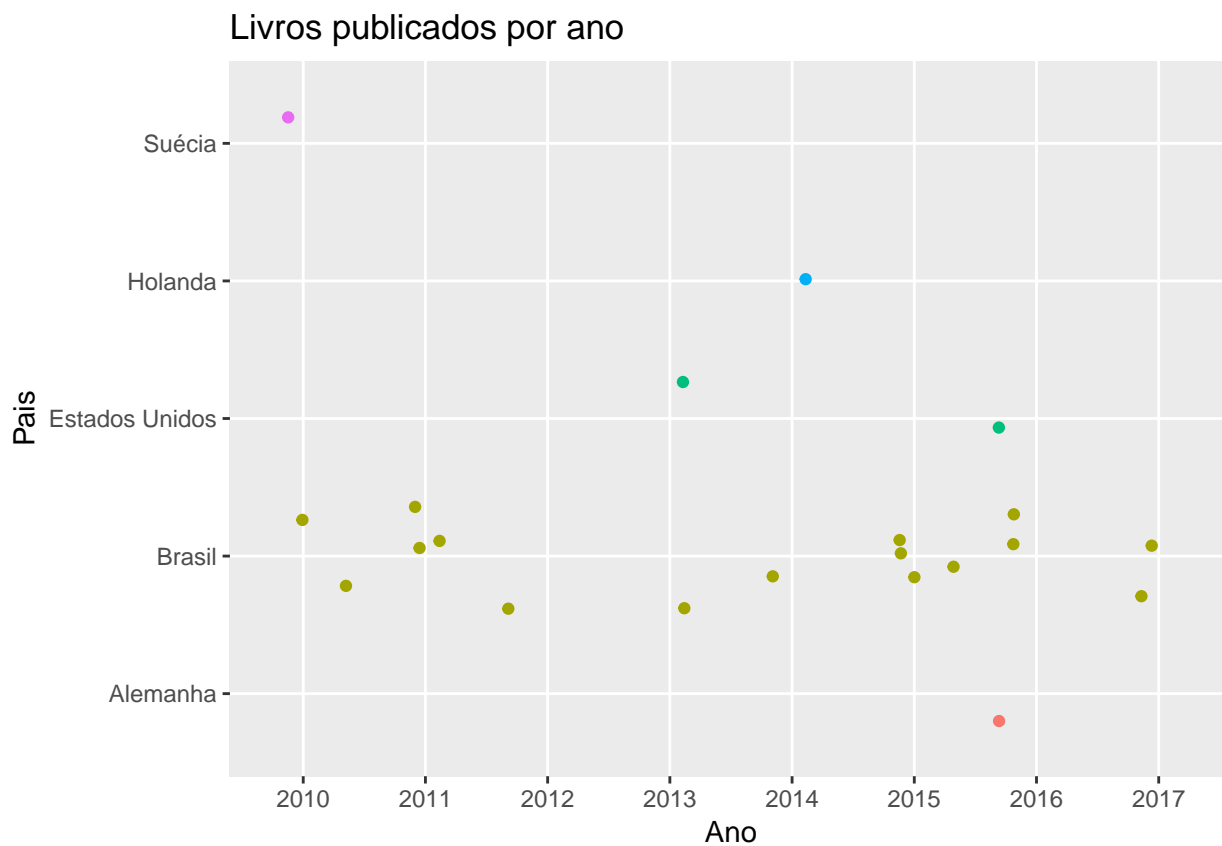
Este gráfico de dispersão demonstra para quantas área pertence (indicada pelas cores dos pontos), há uma homogeneidade na quantidade de publicações, exceto para alguns poucos pontos que deslocam a média para cima. A barra vermelha demonstra a mediana da quantidade de publicações.

## Número de áreas de pesquisa e quantidade de eventos por pesquisador



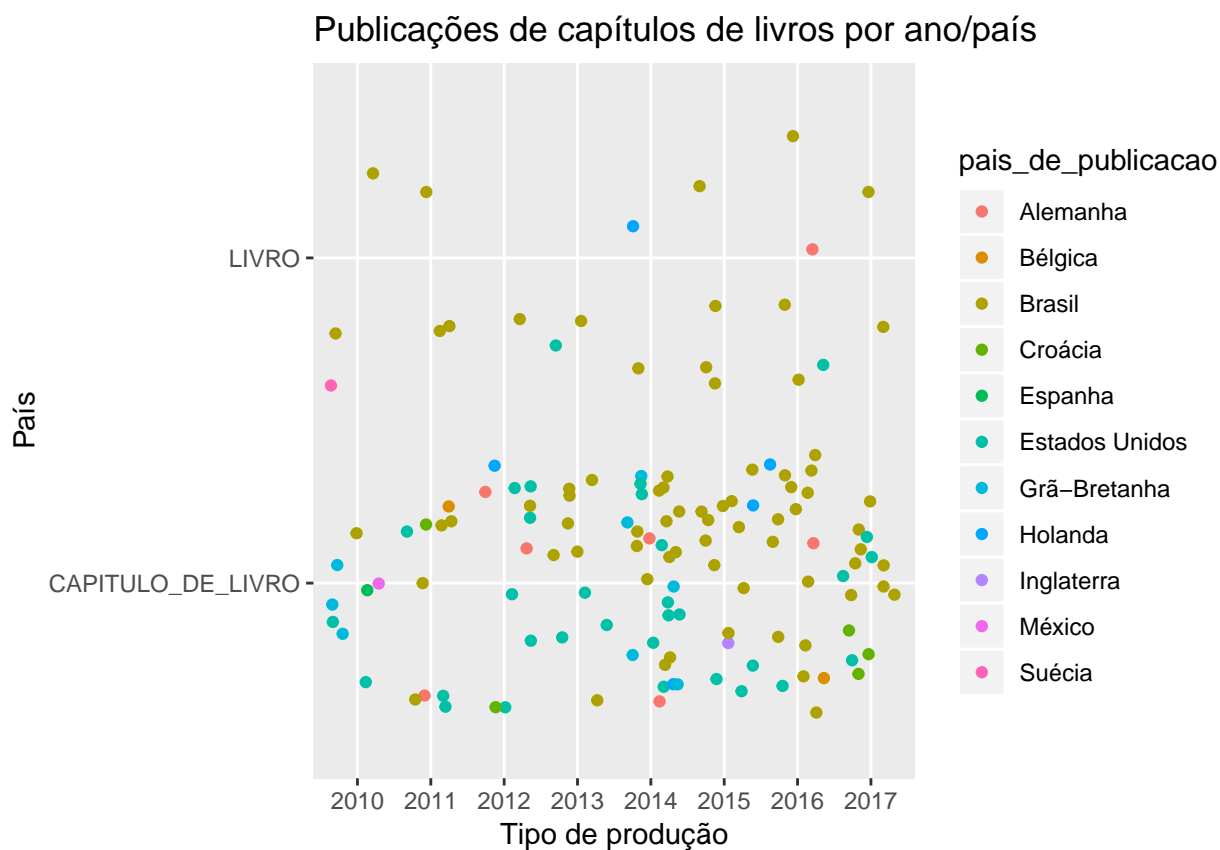
Como o gráfico anterior, para cada pesquisador e pela quantidade de áreas, mostra-se em quantos eventos foram presentes. Nesta figura, denota-se a heterogeneidade, onde há uma grande quantidade de pesquisadores que participam de poucos eventos, bem como pesquisadores que participam de muitos, porém com uma dispersão muito maior e menor agrupamento em torno da mediana.

## Publicação de livros por país/ano



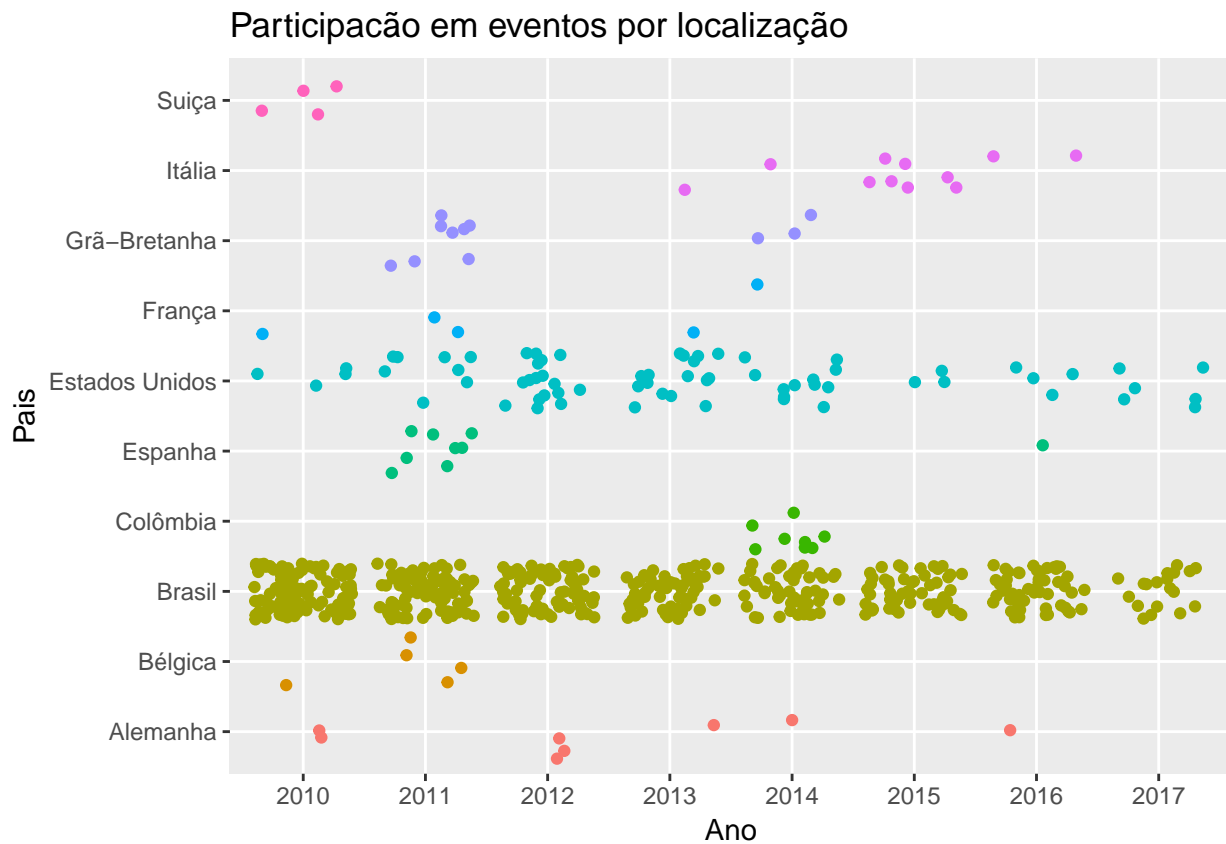
Percebe-se por esta figura, que para a PPG de Biologia Molecular, há uma predominância de livros publicados no Brasil e uma carência de difusão de informação para outros países, uma vez que poucos livros foram publicados no exterior.

## Publicações de capítulos de livros por ano/país



Esta figura como a anterior, mostra que a maioria de capítulos de livros e a maioria de livros inteiros é publicada no Brasil. Porém para a quantidade de capítulos, há uma maior homogeneidade de publicações no exterior, uma vez que há uma boa quantidade de publicações na Holanda e nos Estados Unidos.

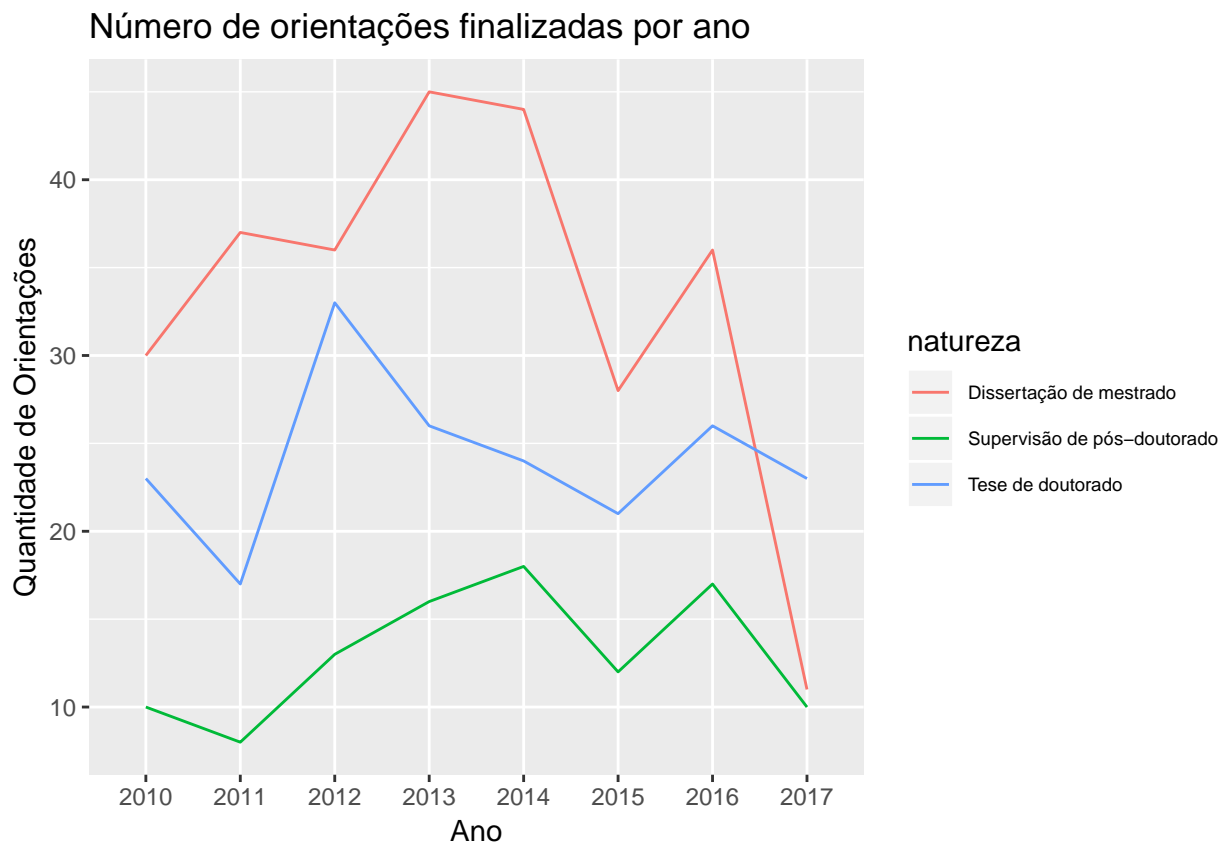
## Eventos nacionais e internacionais



Dos eventos participados, a maioria exorbitante foi no Brasil com uma expressividade no exterior quase constante para os Estados Unidos, e temporária para outros países, como a Bélgica e a Espanha (que tiveram eventos somente no ano de 2011), Grã-Bretanha (que tiveram eventos nos anos de 2011 e 2014 somente), a Suíça (que teve eventos somente em 2010) e a Colômbia (que teve eventos somente em 2014).



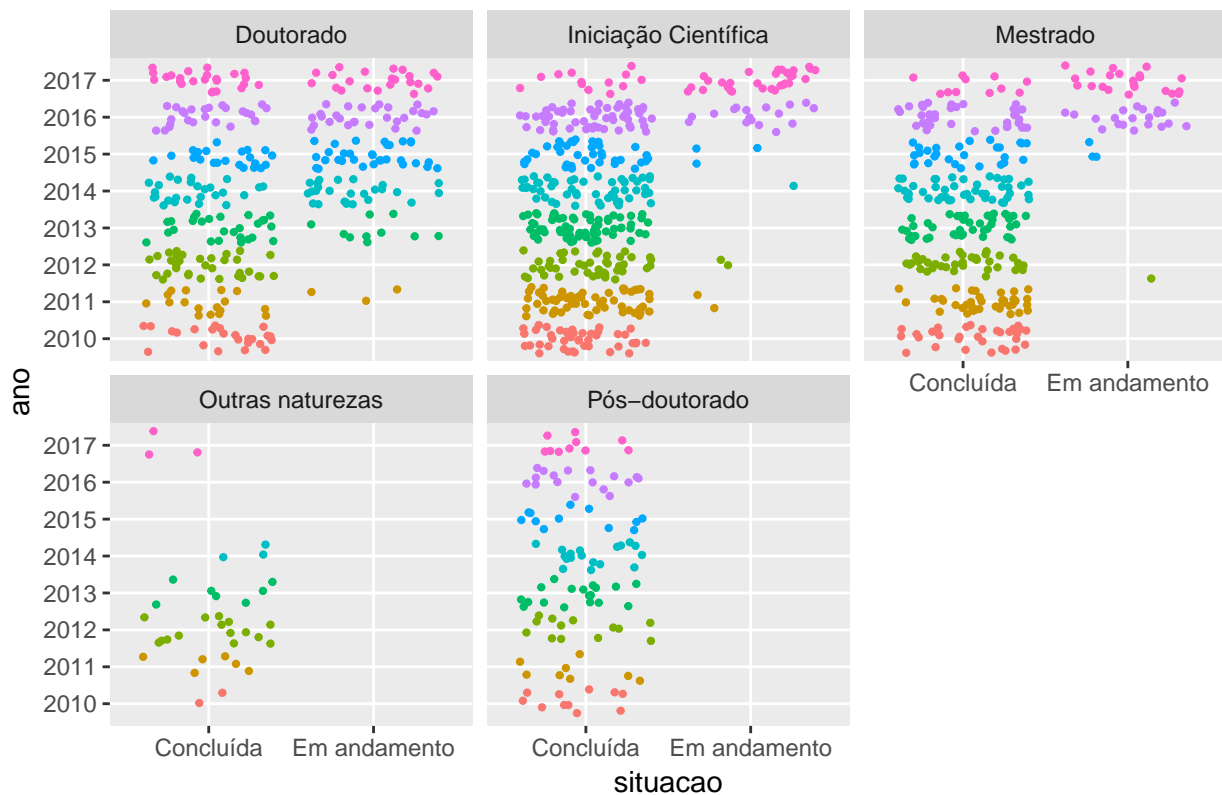
## Orientações completas por ano e natureza



Este gráfico de tendência compara a quantidade de orientações e a natureza destes por ano para a PPG de Biologia Molecular. É notável que o aumento até o ano de 2014 era expressivo para as dissertações de mestrado e supervisões de doutorado, tendência que se alterou para a queda a partir de 2014, onde em 2014 houveram mais que 40 dissertações de mestrado, enquanto em 2017 caiu para 11. A quantidade de teses de doutorado caiu desde 2012 e manteve-se constante em uma margem até 2017.

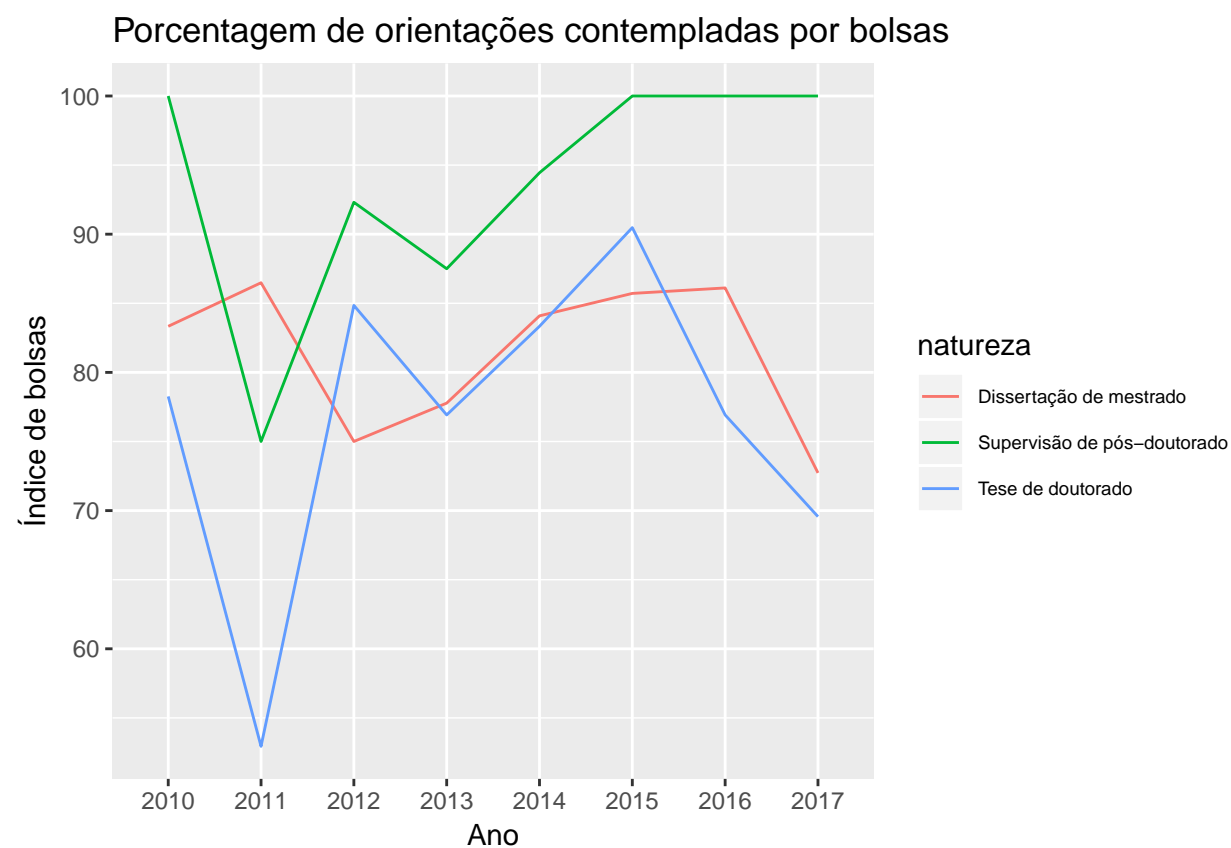
## Evolução temporal de orientações em andamento e concluídas

### Disposição de orientações por ano



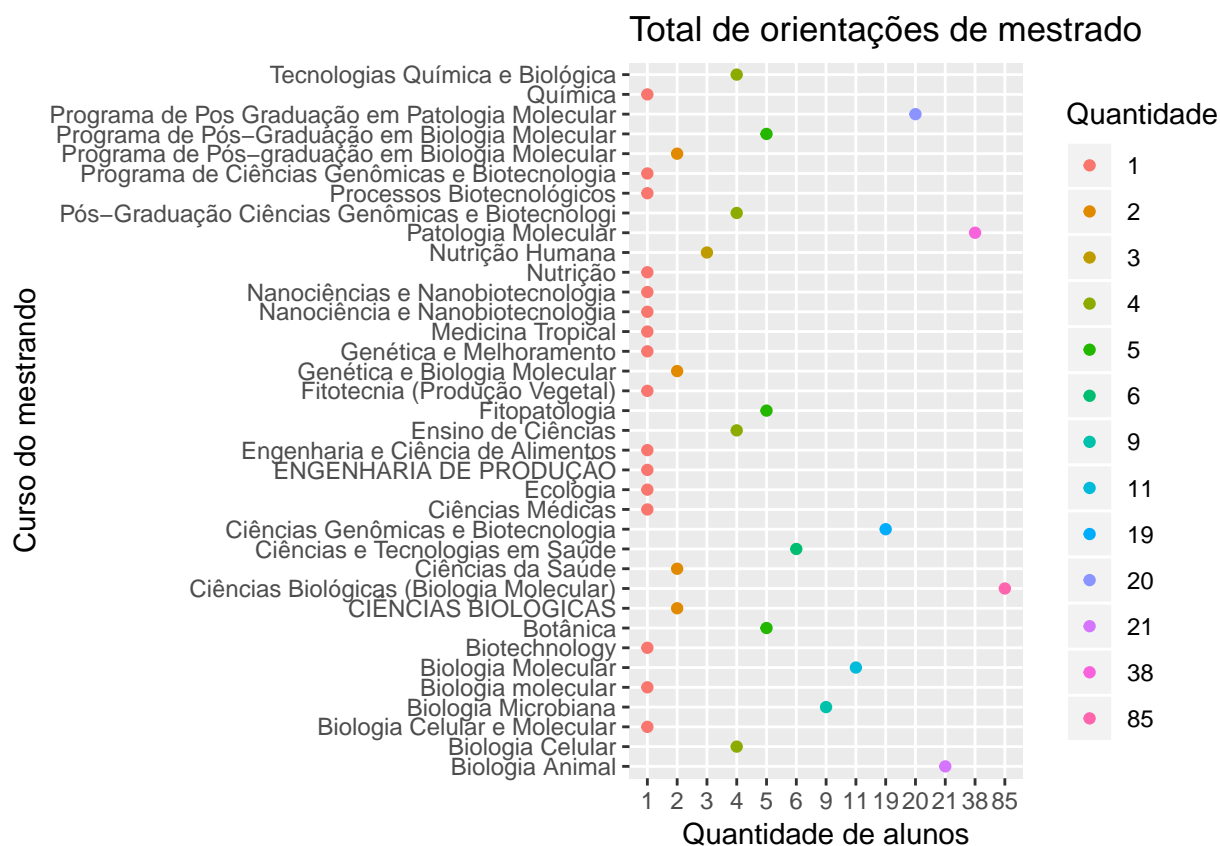
A disposição das orientações e das naturezas de orientações é descrita na figura acima. Nesta figura, percebe-se a tendência de aumento de orientações em andamento em todas as naturezas de pós-graduações. É notável também que mesmo não havendo orientações em andamento para alguns anos, ainda haviam dissertações concluídas. Este fato se deve pela a recepção dos docentes para alunos de outras instituições, nas quais estes participavam das bancas e das orientações destes alunos.

## Índice de bolsas entre naturezas e evolução temporal



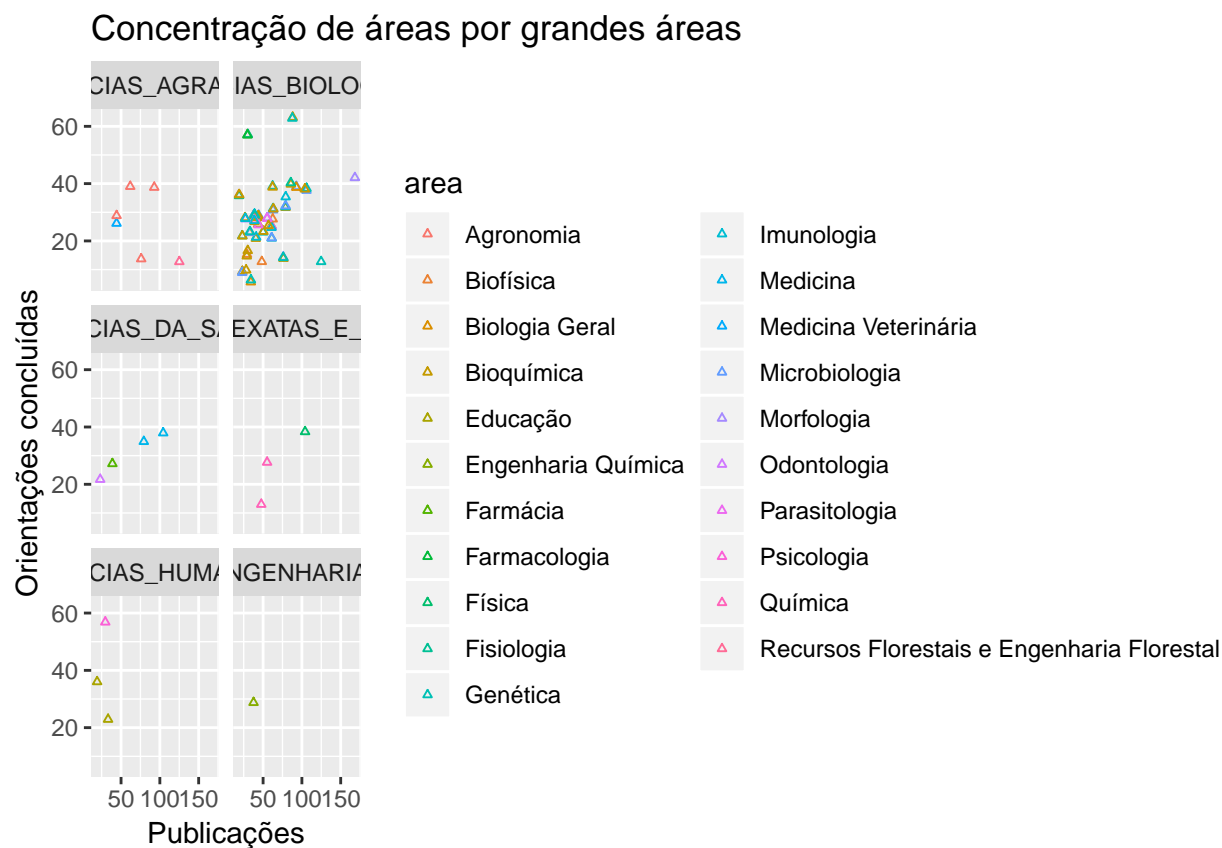
Na figura acima, demonstra-se a quantidade de bolsas por ano para cada natureza de orientação. É notável a queda para 2011 nas teses de doutorado e supervisões de pós doutorado, algo que não se repetiu em nenhum dos outros anos, que continuaram em sua margem de variação normal.

## Cursos de Mestrado e total de alunos no período (2011-2017)



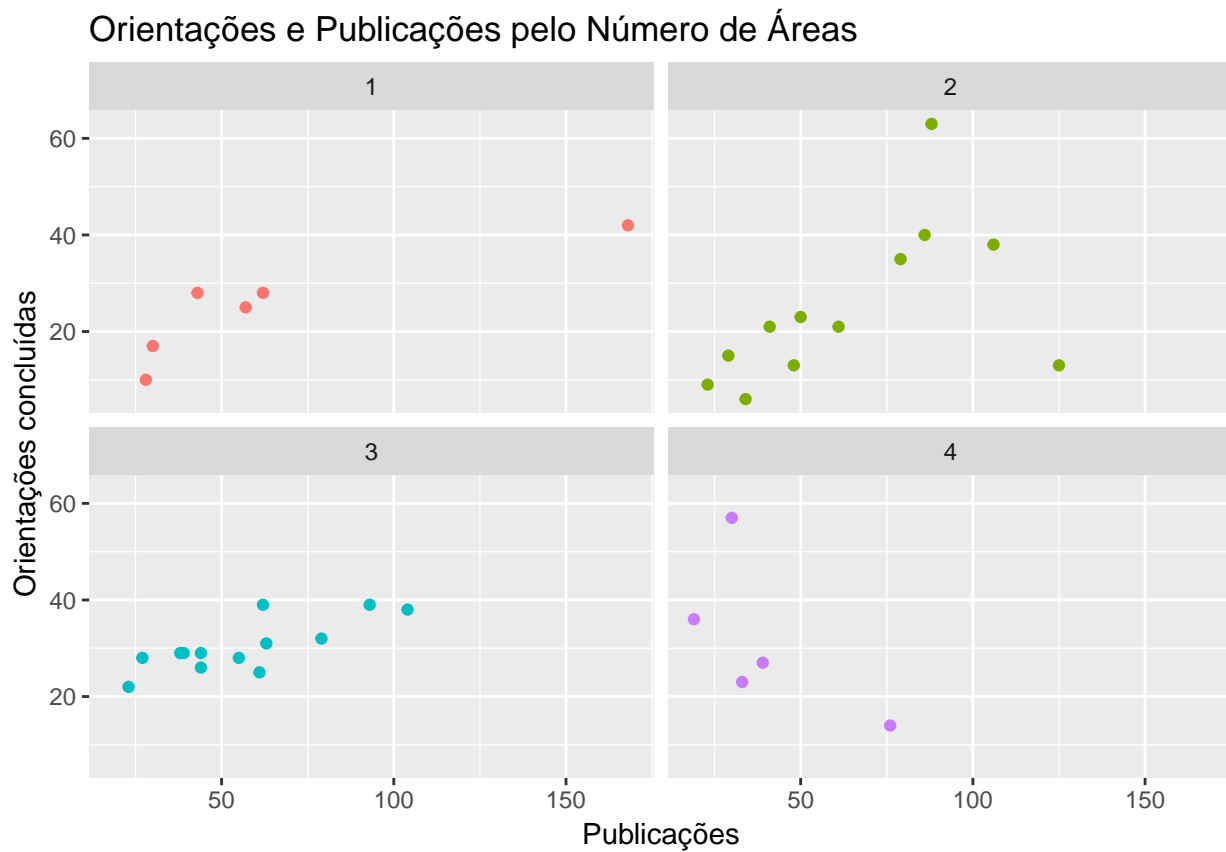
A figura acima demonstra no eixo ordenado que um mesmo orientador pode ter alunos em diversos PPGs e compara a quantidade de alunos para cada PPG. Há uma concentração para uma grande quantidade de alunos no programa de Biologia Molecular - o que era esperado e de áreas correlacionadas, que por coincidência são justamente os programas de Biologia Animal e Patologia Molecular. Observe que no eixo abcisso, os números são tomados como fatores para uma quebra de escala, que não seria possível em valores numéricos.

Concentração de áreas por grandes áreas:



A figura de gráficos acima explicita que a maior concentração de publicações é dentro da grande área de Ciências Biológicas, como o esperado, dado que ainda ocorrem orientações em outras grandes áreas. É visto que os números das áreas que se destacam em cada grande área condiz com sua natureza, uma vez que psicologia é de fato mais presente em Ciências Humanas, como exemplo. Lembrando que um pesquisador possui mais de uma grande área, o pesquisador se repete entre diferentes gráficos.

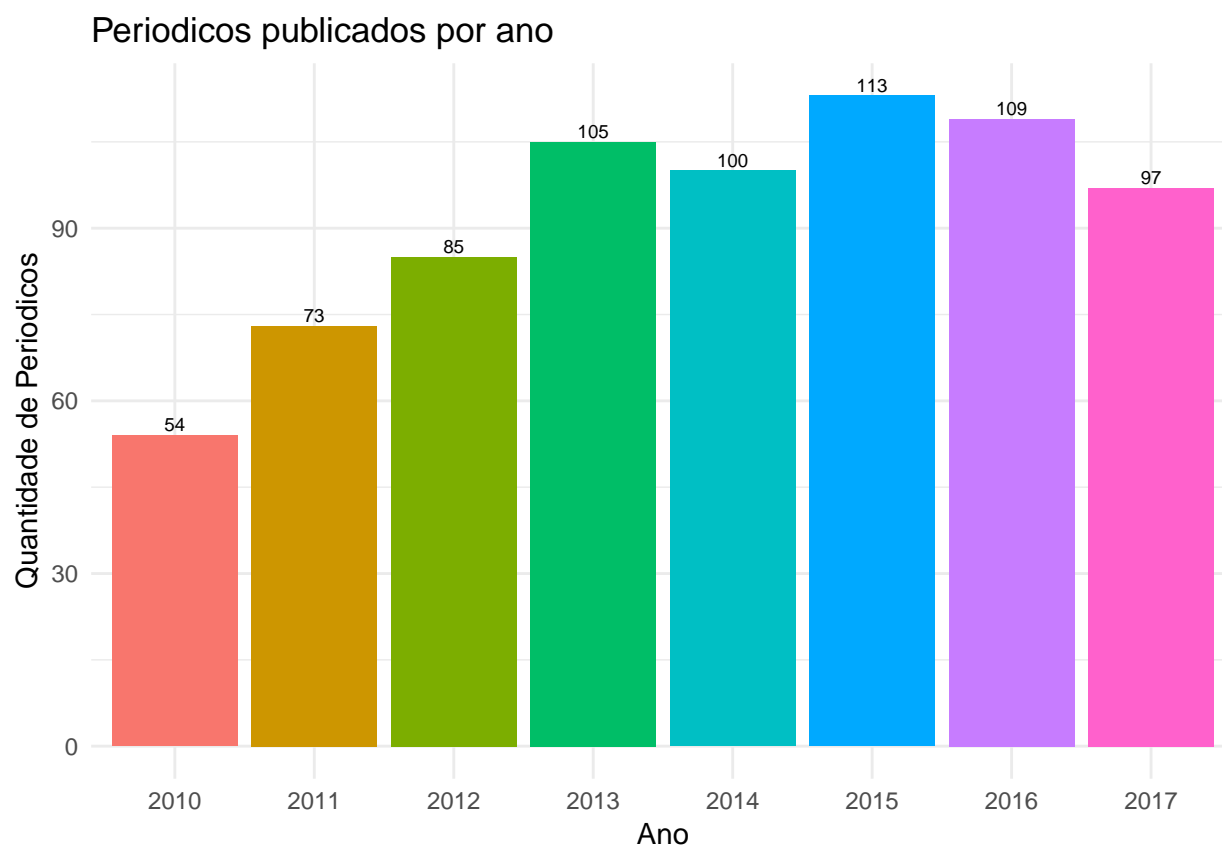
## Quantidade de Orientações e Publicações pelo Número de Áreas



Quem trabalha em mais áreas diferentes publica/orienta mais? A figura acima mostra que não, uma vez que quem trabalha em mais áreas orienta mais pessoas, porém publica menos em relação a quem trabalha em menos áreas.

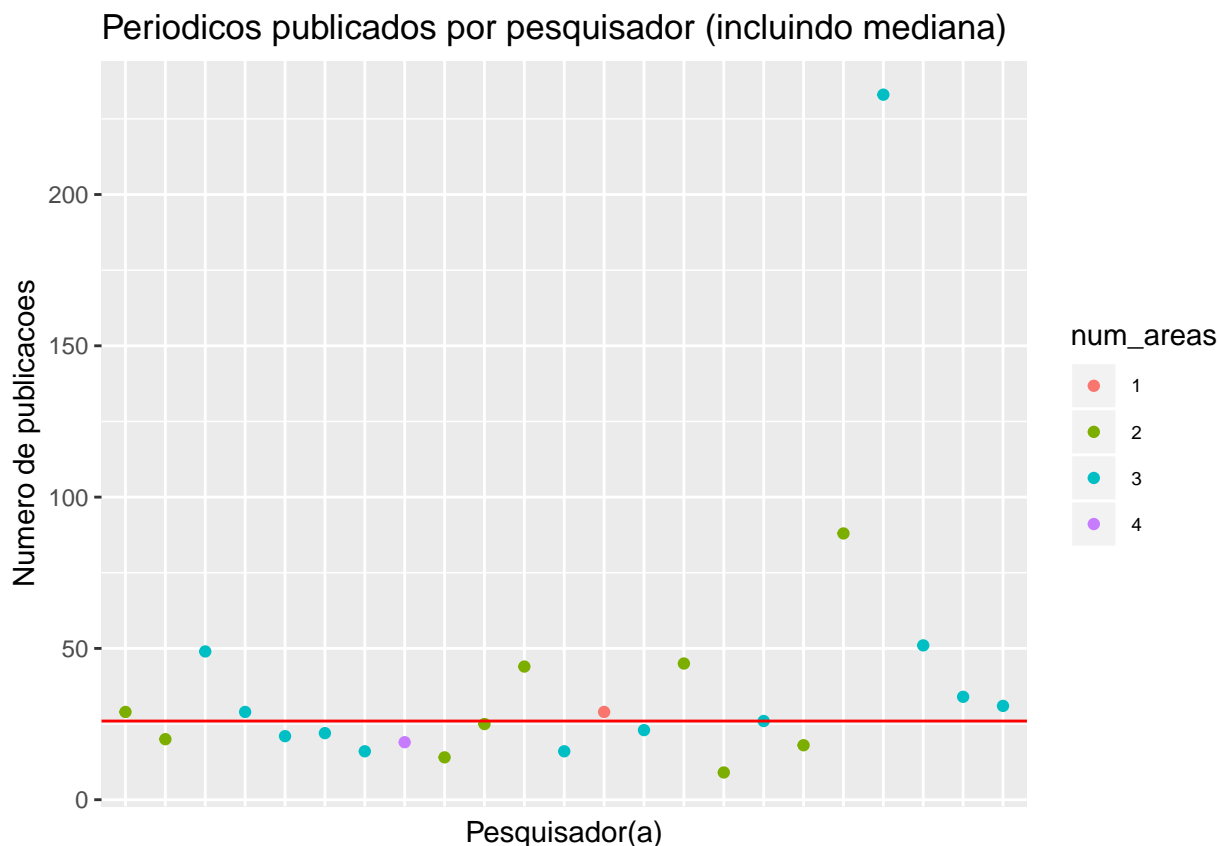
## Patologia Molecular

Grafico de barras; periodicos por ano



Sob uma perspectiva de publicações em periódicos, o Programa de Pós-Graduação de Patologia Molecular apresentou um grande crescimento entre os anos de 2011 e 2015, evoluindo gradualmente a ponto totalizar um aumento de 109% na produção anual neste intervalo (de 54 para 113). Após atingir este pico em 2015, o PPG teve redução no número de publicações nos anos seguintes. Considerando a participação de 23 docentes no programa em 2017, foi atingida uma média de 4,22 publicações em periódicos por professor no ano em questão.

## Quantidade de periodicos publicados por professor(a) entre 2010 e 2017

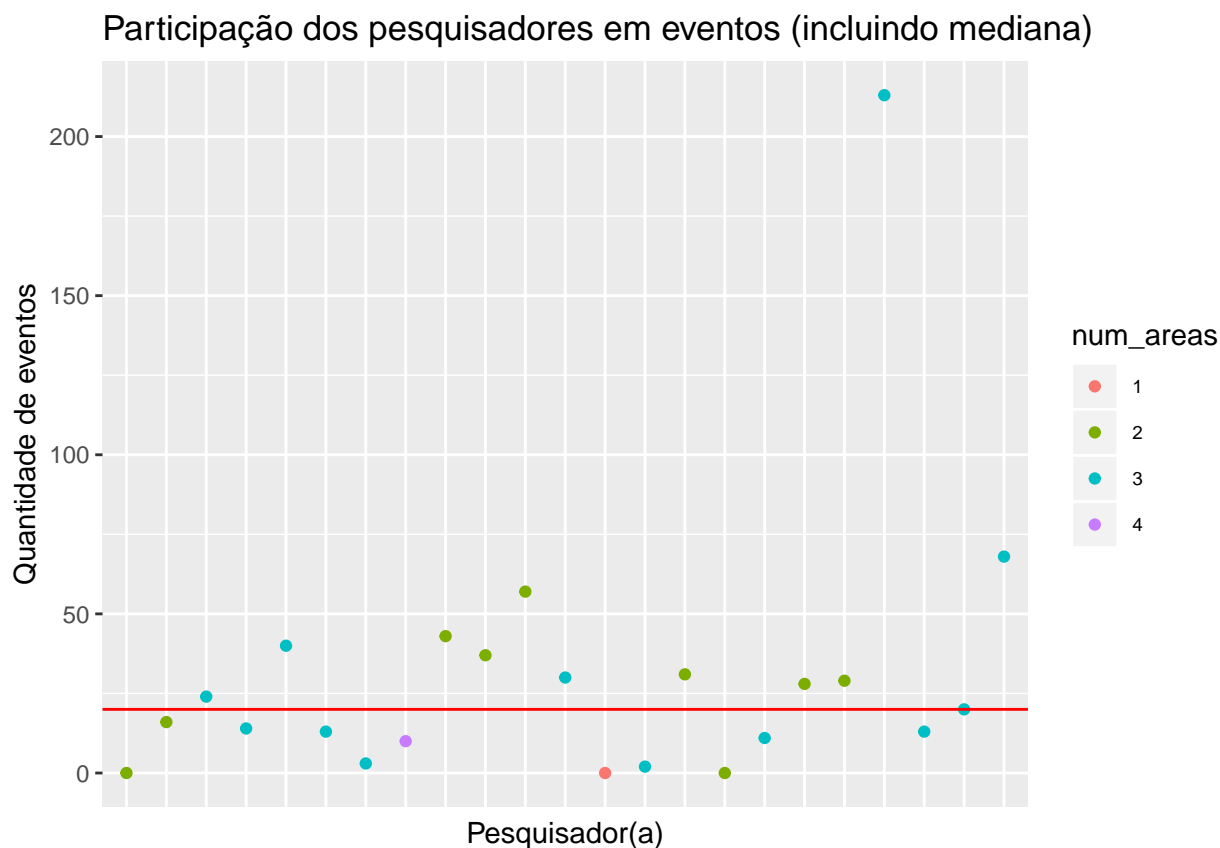


O gráfico acima demonstra o total de publicações registradas por cada pesquisador do PPG de Patologia Molecular, demonstrando relativa homogeneidade na produção média por pesquisador, uma vez que grande parte das amostras se aproxima da mediana (26 publicações em periódicos). É curioso observar que, apesar de os gráficos se comportarem de maneira diferente, a mediana foi a mesma obtida pelo Programa de Biologia Animal, apresentado anteriormente.

Nota-se aqui um grande destaque para um pesquisador que apresentou 233 publicações, quase 9 vezes o valor obtido pela mediana. Além disso, 22 dos 23 pesquisadores do PPG de Patologia estão relacionados a mais de uma área de pesquisa - revelando um foco menor do programa em especialização e maior em multidisciplinaridade.

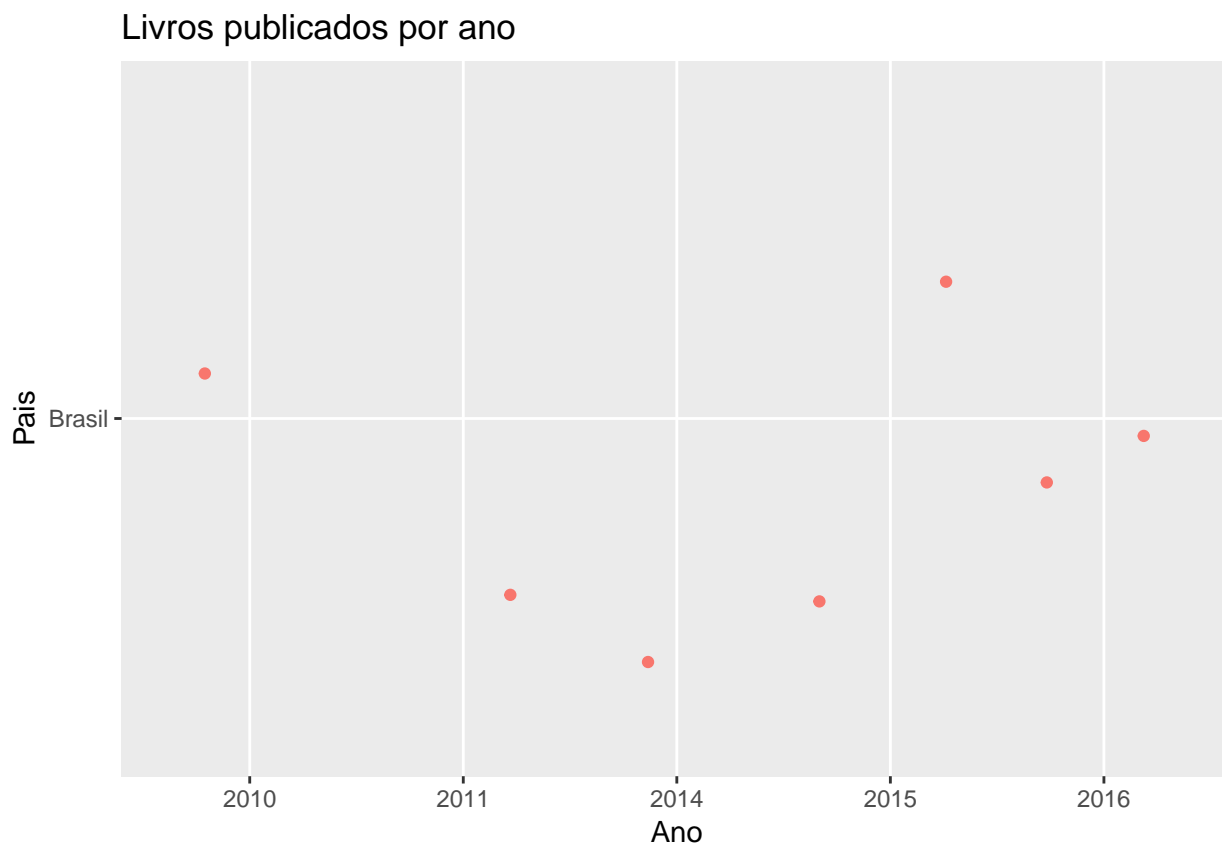


## Número de áreas de pesquisa e quantidade de eventos por pesquisador



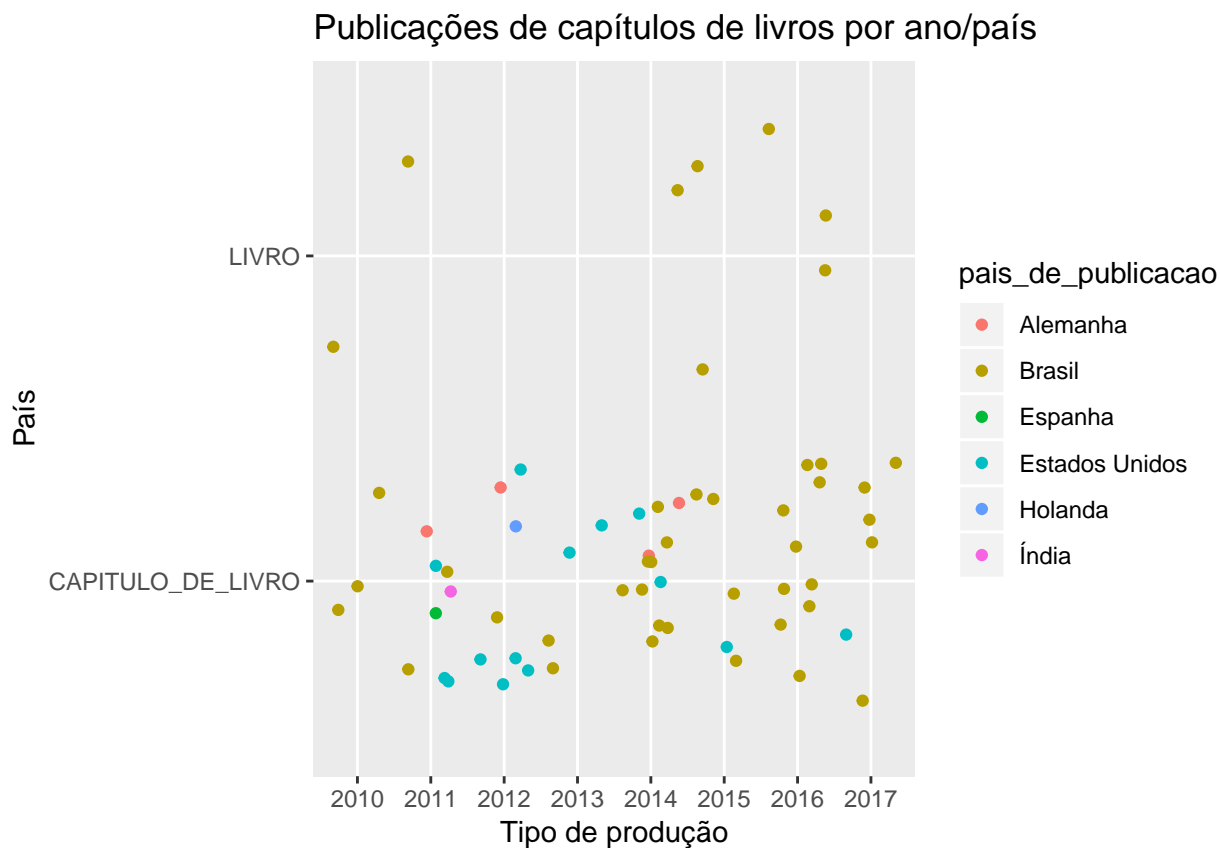
O gráfico acima demonstra o total de eventos registrados por cada pesquisador, demonstrando relativa homogeneidade na participação média de cada pesquisador em eventos, uma vez que grande parte das amostras se aproxima da mediana (20 participações). Assim como na relação de periódicos, há um pesquisador com muito mais participações registradas que seus colegas de programa (10 vezes o valor da mediana), sendo este o mesmo docente que atingiu o pico no número de periódicos. É curioso notar que o único pesquisador cadastrado com apenas uma área de pesquisa não apresentou nenhuma participação em eventos, reiterando uma proposta de multidisciplinaridade do PPG de Patologia Molecular.

## Publicação de livros por país/ano



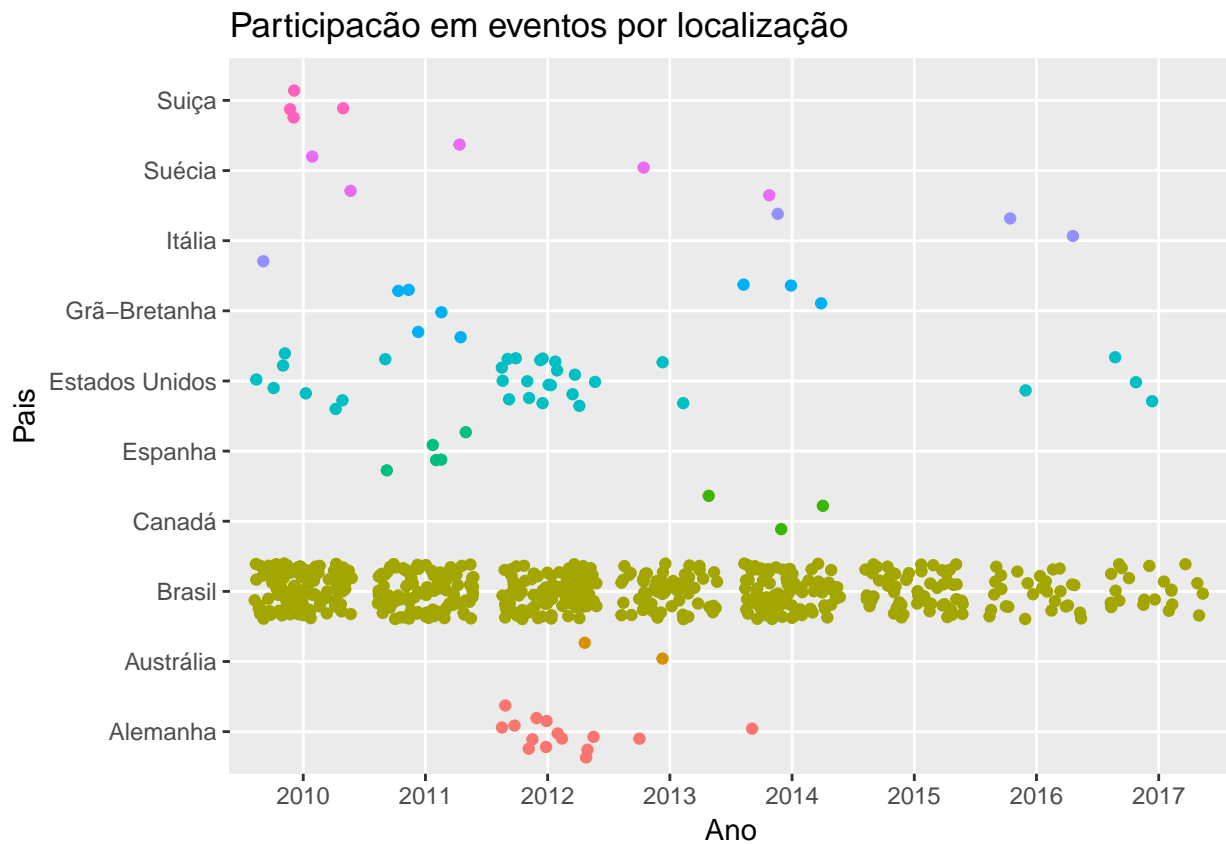
Denota-se no PPG de Patologia Molecular uma certa homogeneidade temporal na distribuição de livros publicados para cada país de publicação, com sete publicações em cinco anos diferentes. Curiosamente, nota-se que este é o único programa que não apresentou nenhum livro publicado fora do Brasil no período analisado. Uma das possíveis justificativas para este baixo índice comparativo é o fato de ser o programa de criação mais recente entre os programas aqui analisados, revelando neste um processo de maturação em andamento no aspecto de internacionalização.

## Publicações de capítulos de livros por ano/país



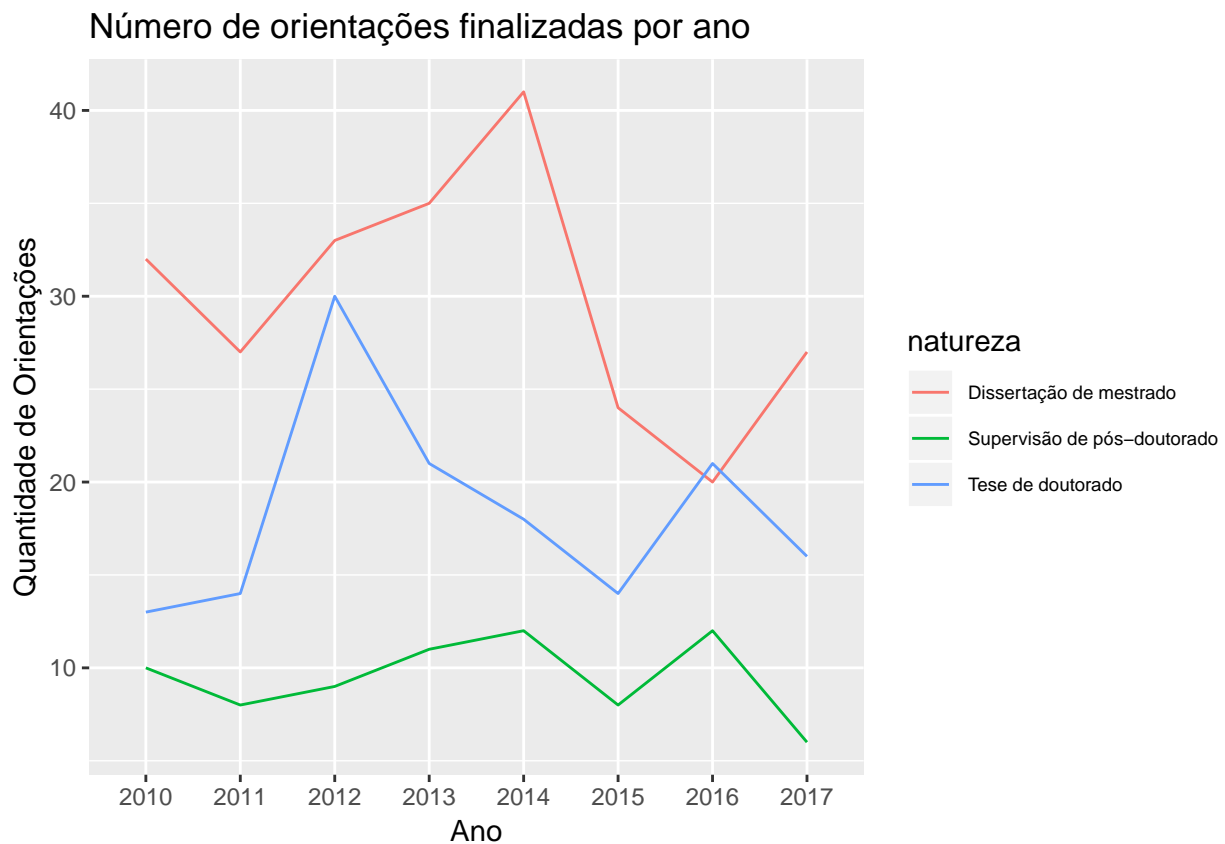
Trazendo a análise de publicações de capítulos de livros, há um aumento expressivo no número de amostras, em relação a considerar apenas livros. Todos os anos analisados (2010-2017) registraram mais de um capítulo de livro publicado pelo PPG em Patologia Molecular. Neste eixo, o nível de internacionalização foi consideravelmente maior que o anterior, com a menção a diversos países da América e Europa. Entretanto, a redução da participação internacional do programa em livros é evidente nos anos 2016 e 2017, com publicações nacionais, principalmente - assim como nos demais programas analisados.

## Eventos nacionais e internacionais



Considerando eventos, o Programa de Patologia Molecular teve comparecimento muito maior em eventos no Brasil que em países estrangeiros, como esperado. Ainda assim, o gráfico mostra que o programa teve boa presença em eventos de sede internacional nos anos iniciais da base de dados, com destaque para 2012. Além disso, é notável a queda no número de participações a partir do ano de 2015, tanto nacionais quanto internacionais - revelando pouca correlação entre a participação de eventos e publicação de periódicos (mostrada anteriormente).

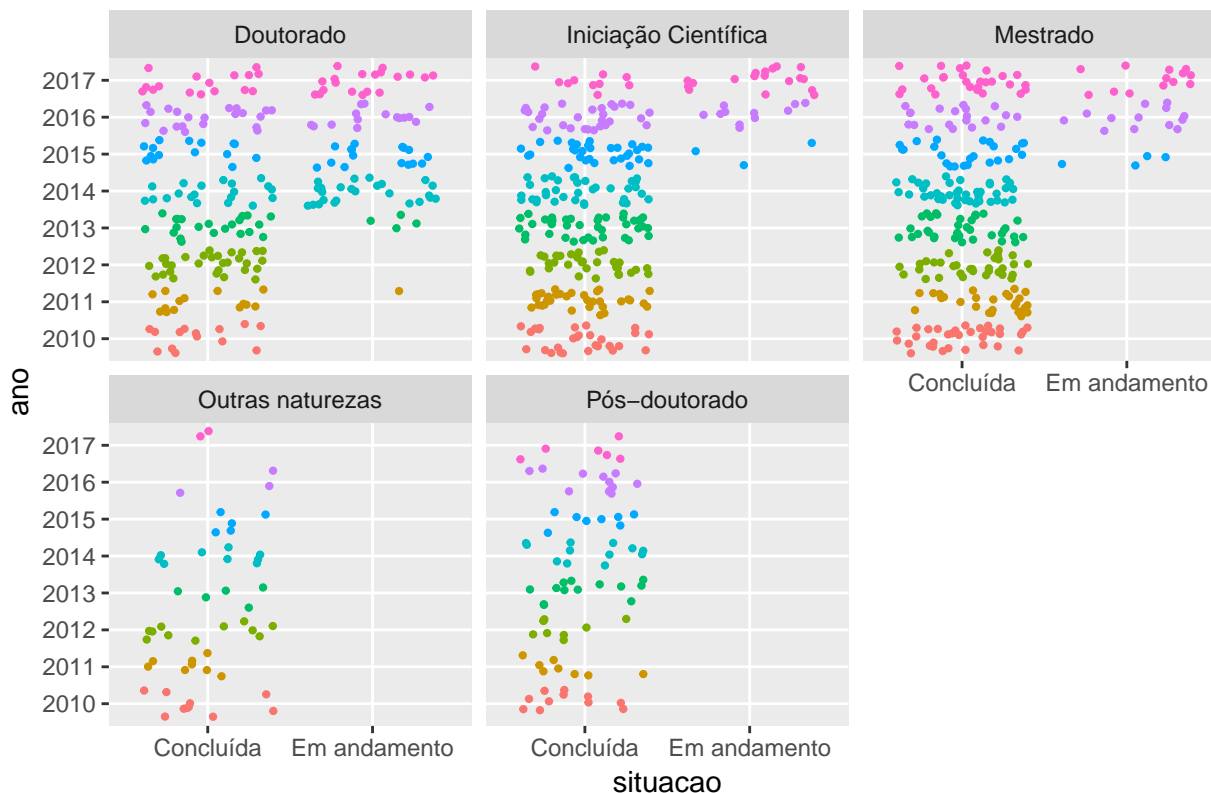
## Orientacoes completas por ano e natureza



Observando a evolução do número de orientações completas ao longo dos anos, percebe-se que o Programa de mestrado cresceu consideravelmente entre 2010 e 2014, mas apresentou regressão nos períodos posteriores, chegando ao seu menor índice em 2016. Em contrapartida, o número de orientações de pós-doutorado apresentou maior estabilidade ao longo da amostra, mesmo que com valores reduzidos. Todas as naturezas de orientação do Programa parecem bem estabelecidas, visto que mesmo os índices de doutorado apresentaram ao menos 15 observações/ano. Ainda assim, é notável a queda no número de orientações de 2014 para 2015 - o que corrobora com as análises anteriores de queda de participação em eventos no ano em questão.

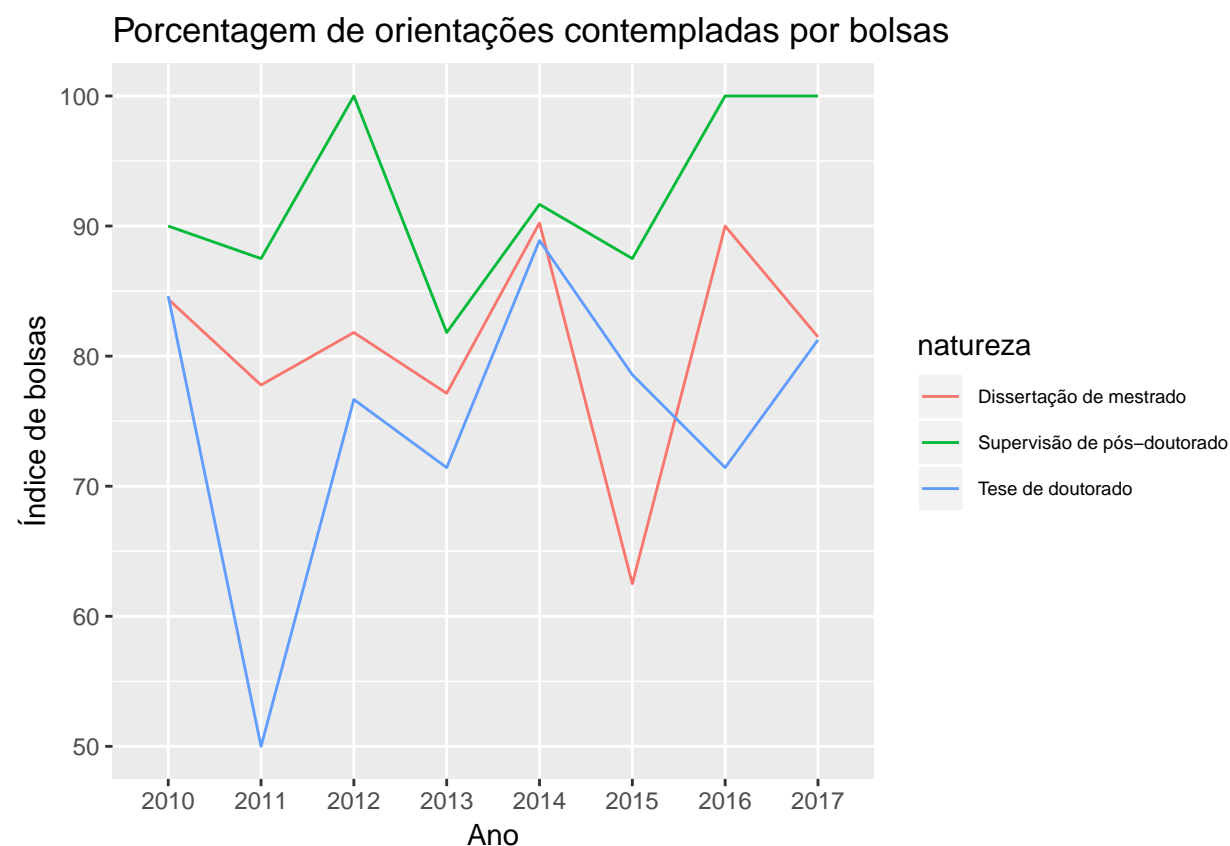
## Evolução temporal de orientações em andamento e concluídas

### Disposição de orientações por ano



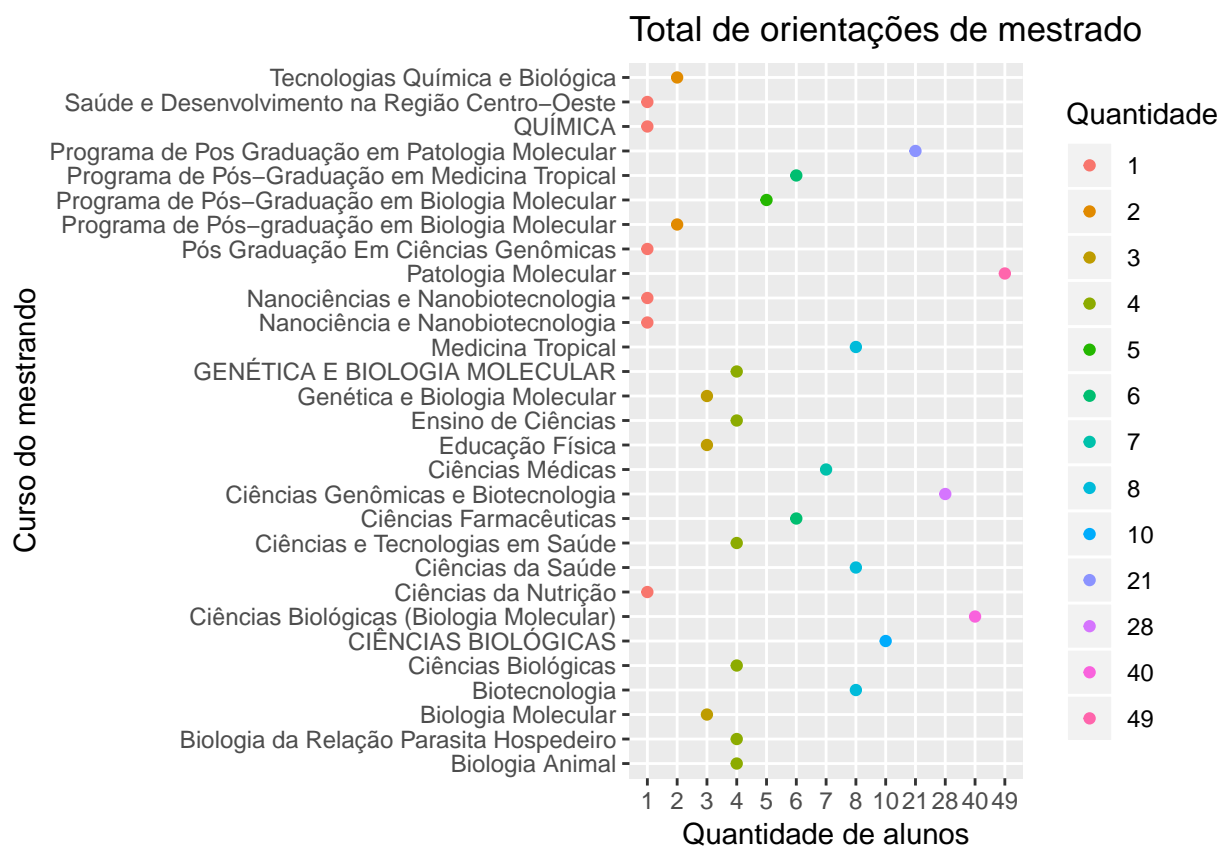
Observou-se uma quantidade praticamente linear no número de iniciações científicas, mestrados e doutorados concluídos no período 2010–2014, com uma queda relevante após 2015. As iniciações científicas tiveram queda vertiginosa em 2017, o que sinaliza uma queda no número de pessoas envolvidas na pesquisa também na graduação. Conjuntamente a isso, verifica-se um aumento significativo no número de orientações em andamento a partir de 2014. O comportamento prévio a este ano, com diversas orientações finalizadas e poucas em andamento, sugere que o PPG em Patologia Molecular contemplava poucos cadastros de orientações em andamento e está intensificando esta prática nos anos mais recentes - o que tornará análises futuras ainda mais enriquecedoras.

## Índice de bolsas entre naturezas e evolução temporal



Observando o gráfico de porcentagem de bolsas do Programa de Patologia Molecular, é possível perceber que o número de bolsas oferecidas se comportou de maneira não linear ao longo dos anos. Durante todo o período, todas as naturezas apresentaram um índice de pelo menos 50% de bolsas. Além disso, as teses de pós-doutorado se mostram a natureza de pesquisa melhor contemplada pelas agências financiadoras, visto que foram a única natureza de pós-doutorado que atingiu 90% em algum momento da amostra. Por fim, é possível observar que o pico no número de orientações de mestrado registrado em 2014 foi acompanhado por um aumento de 15% no índice de alunos contemplados com bolsa, revelando um ano de grande crescimento na área de mestrado do PPG de Patologia Molecular.

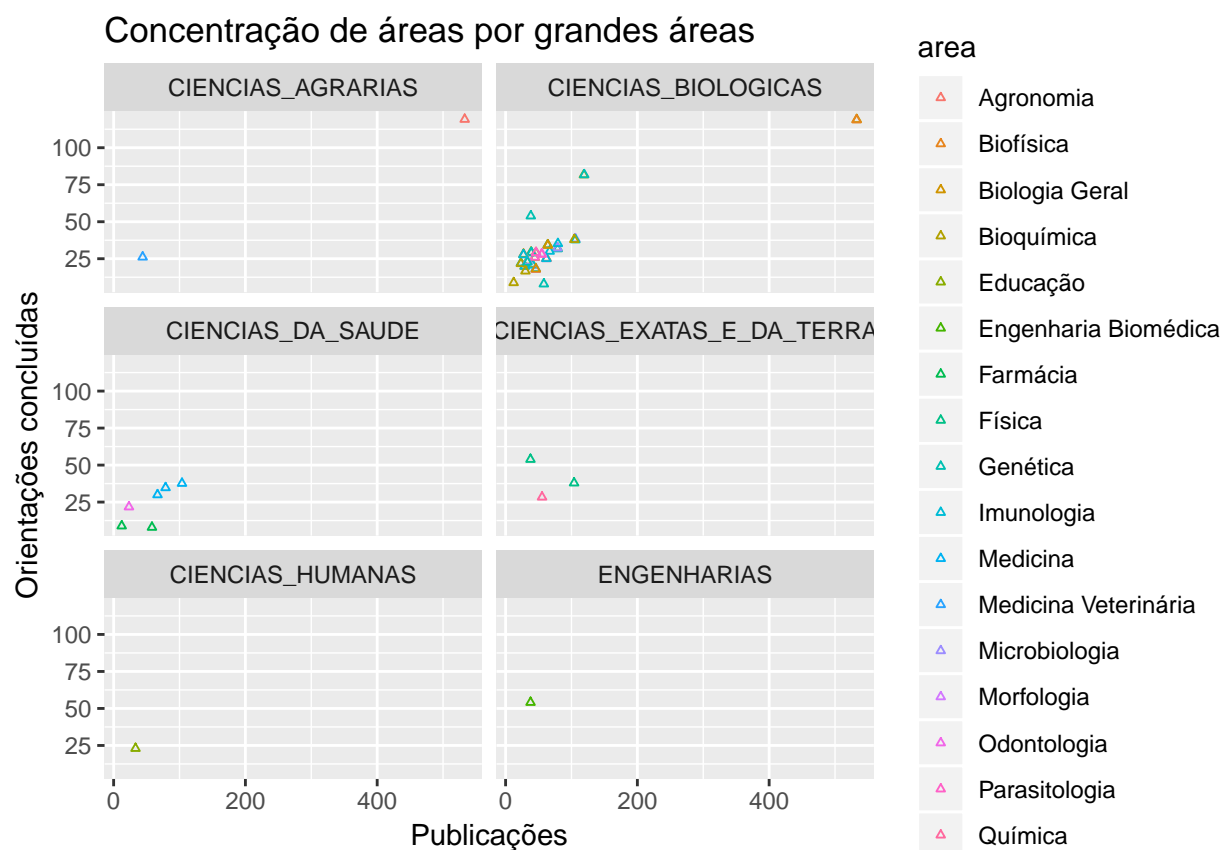
## Cursos de Mestrado e total de alunos no período (2011-2017)



A análise de cursos dos alunos de mestrado orientados pelos professores do PGG de Patologia Molecular apresenta um grande índice de correlação com outras áreas. O curso com maior número de orientações do programa é ele mesmo, como esperado (49 orientações). Entretanto, é visível a relação íntima dos docentes com o Programa de Biologia Molecular, com 40 orientações. Em seguida, são listados diversos outros programas de pós-graduação com quantidade significativa de orientados cadastrados - não se limitando aos que foram analisados neste trabalho.

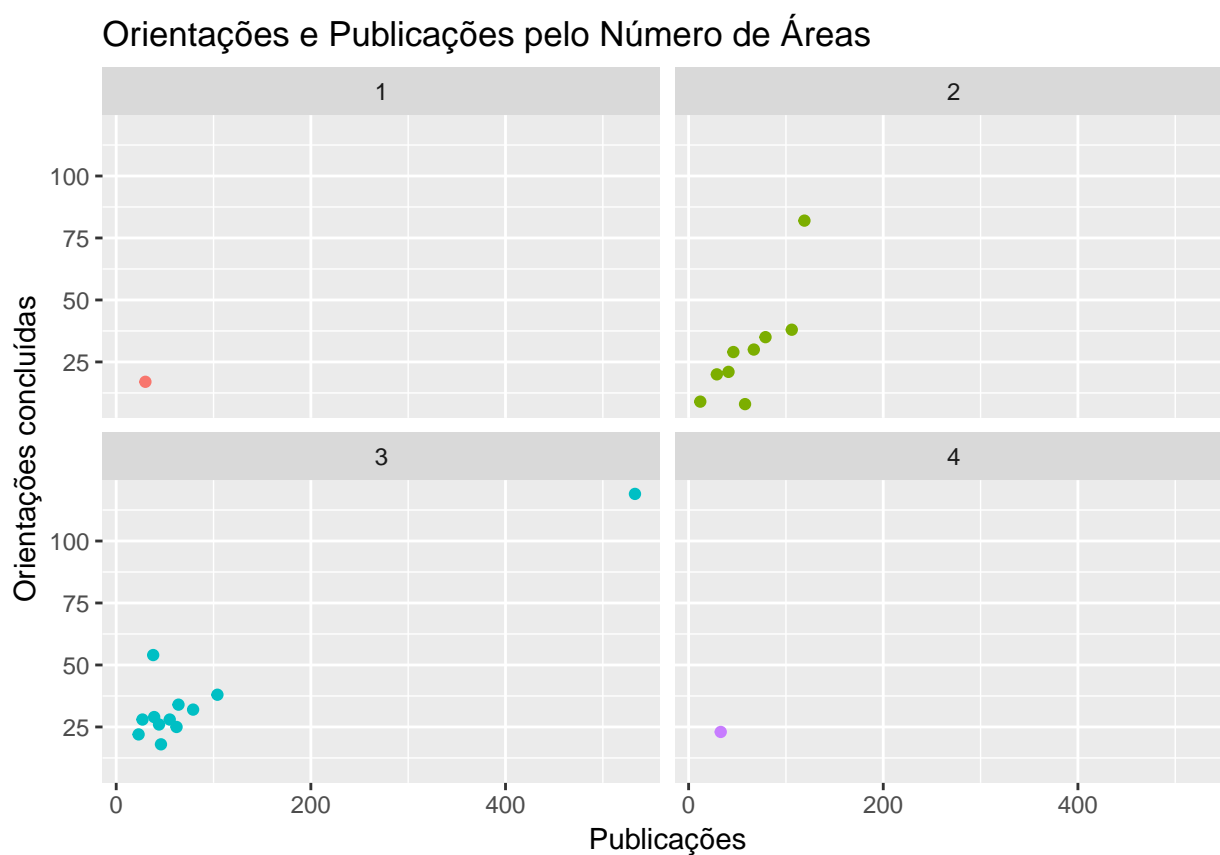


Concentração de áreas por grandes áreas:



Uma análise superficial do gráfico apresentado já evidencia uma grande concentração de pesquisadores do PPG de Patologia Molecular na grande área de Ciências Biológicas. Entretanto, é curioso observar que a maioria dos pesquisadores de maior produtividade (em orientações e publicações) também estão integrados a outras grandes áreas. O principal expoente do programa nestes aspectos está associado à grande área de Ciências Agrárias, por exemplo.

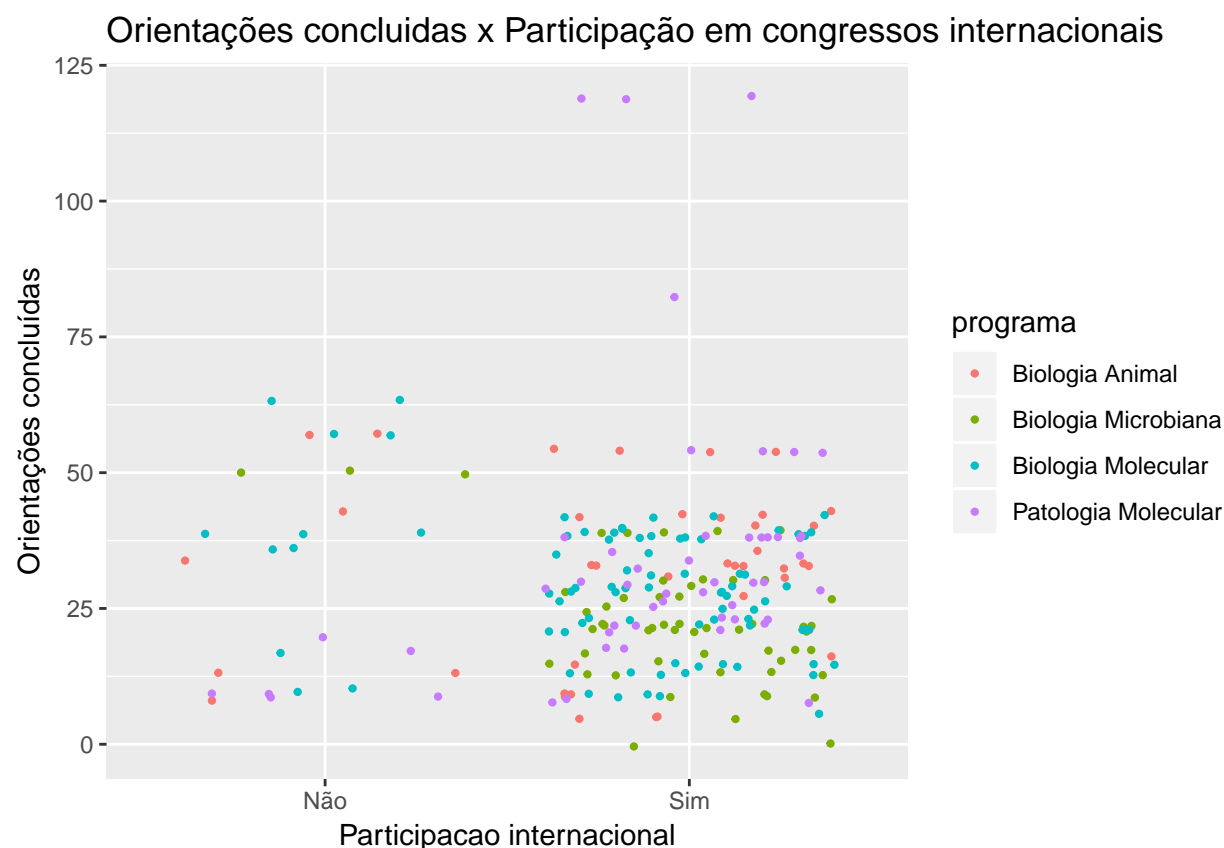
## Quantidade de Orientações e Publicações pelo Número de Áreas



Em um primeiro momento, observa-se que o pesquisador *outlier* comentado anteriormente apresenta valores discrepantes tanto para número de publicações quanto para orientações concluídas. A ideia discutida até aqui sobre a valorização de pluralidade de disciplinas, entretanto, não implica necessariamente em maior produtividade no PPG - o único pesquisador a trabalhar em 4 grandes áreas não apresentou número de orientações e/ou publicações destoante dos demais docentes.

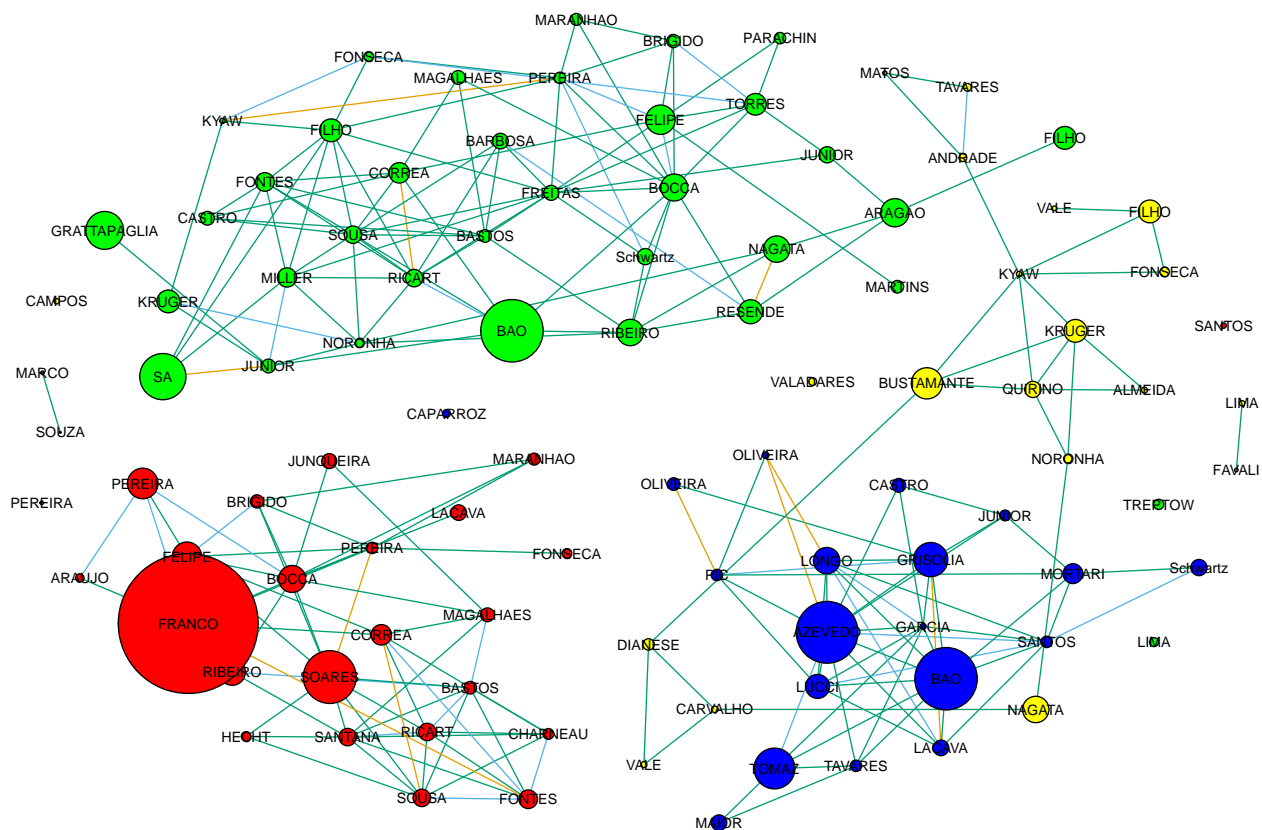
## Análises globais

### Quantidade de Orientações e Participação internacional por PPG



O gráfico acima traz uma comparação entre os pesquisadores dos diversos PPGs analisados neste trabalho e uma relação entre a participação internacional dos docentes e orientações concluídas dos mesmos. É curioso observar que, de todos os programas, o PPG de Patologia Molecular foi o que apresentou mais pontos fora do padrão, com docentes que ultrapassaram 100 orientações no período analisado. Além disso, é possível observar que o PPG de Biologia Microbiana foi o que apresentou maior índice de participação internacional entre seus membros, com grande concentração de suas amostras no lado direito do gráfico. Por fim, é possível perceber que a grande massa pesquisadores ativos internacionalmente se encontra na parte de baixo do gráfico - sinalizando que experiências internacionais não implicam em maior número de orientações.

## Grafo de proximidade entre pesquisadores do Programa de Pós-Graduação



Este grafo demonstra a rede de cooperação, peso, quantidades de publicações e senioridade dos indivíduos e da cooperação entre eles. Cada círculo representa um docente dentro de um PPG indicado pela cor, onde: azul é a PPG de Biologia Animal; amarelo é PPG de Biologia Microbiana; verde corresponde a PPG de Biologia Molecular; e vermelho a PPG de Patologia Molecular. Já para cada aresta, as cores determinam a senioridade entre os dois pontos, onde verde é 9, azul é 8 e laranja é 7; quanto ao diâmetro é dado pela a quantidade de áreas um vértice compartilha com o outro; já para a ligação da aresta, esta é dada pelos os periódicos em comum para cada vértice, indicando cooperação. A partir deste grafo é possível entender que cada programa de pós graduação se mantém praticamente isolado, exceto pelos os docentes que se repetem para cada PPG. É notável que a maioria destas PPGs tem muita intracomunicação, deixando poucos docentes isolados exceto pela a PPG de Biologia Microbiana que tiveram seis docentes completamente isolados de qualquer outra PPG e dela mesma, tendo o maior número de isolamentos. É possível perceber também que a docente de sobrenome Cynthia Maria Kyaw é uma comunicadora entre PPGs, pois intercomunica as PPGs de Biologia Molecular e Biologia Microbiana e além disso previne o isolamento entre sua própria PPG formando uma ponte entre um e outro grupo de docentes que estariam isolados se não colaborassem com esta. Em cada uma das PPGs há docentes em destaque pela a quantidade de publicações (indicada pelo o diâmetro do vértice), mas é possível perceber que o PPG de Patologia Molecular tem uma dominância de um docente que publica aproximadamente o triplo que o anterior em publicações, o que resulta em um grande vértice central. Este dominante também está com a maioria das interligações de seu PPG, sendo portanto um canal de informação dentro de seu grupo.

## Bibliografia

“SciVal Metrics Guidebook.” ELSEVIER, 2014.

Chapman, Pete, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, e Rüdi-ger Wirth. “CRISP-DM 1.0: Step-by-Step Data Mining Guide”. USA: CRISP-DM Consortium, 2000. Disponível em: <https://www.the-modeling-agency.com/>, visitado em Novembro, 2019.

Fernandes, Jorge H C, Ricardo Barros Sampaio, e João Ribas de Moura. “Ciência de Dados para Todos (Data Science For All) - 2018.1 - Análise da Produção Científica e Acadêmica da Universidade de Brasília - Modelo de Relatório Final da Disciplina - Departamento de Ciência da Computação da UnB”. Disciplina 116297 - Tópicos Avançados em Computadores, turma D, do semestre 2018.1, do Departamento de Ciência da Computação do Instituto de Ciências Exatas da Universidade de Brasília, 13 de junho de 2018

Secretaria de Comunicação da UnB (SeCom). 2018. “Impacto da produção científica da UnB aumenta mais de 100% em seis anos.” Universidade de Brasília, Junho de 2018, Disponível em <https://www.noticias.unb.br/>, visitado em Novembro de 2018.

Azevedo, Mário Luiz Neves de, João Ferreira de Oliveira, e Afrânio Mendes Catani. “O Sistema Nacional de Pós-Graduação (SNPG) e o Plano Nacional de Educação (PNE 2014-2024): regulação, avaliação e financiamento”. Revista Brasileira de Política e Administração da Educação 32, nº 3 (2016). <http://dx.doi.org/10.21573/vol32n32016.68576>.

Leite, Fernando César Lima. “Considerações básicas sobre a Avaliação do Sistema Nacional de PósGraduação”. Comunicação Pessoal (slides). Universidade de Brasília, abril de 2018.

CAPES. “Documentos de Área”. CAPES.gov.br. Acessado 12 de junho de 2018. Disponível em: <http://avaliacaoquadrienal.capes.gov.br/>, visitado em Dezembro de 2018.

PORTAL DE PERIODICOS UNISUL, O Processo de Avaliação do Sistema Nacional de Pós-Graduação Identifica a Capes e a Diferencia de outras Agencias. Disponível em: <http://www.portaldeperiodicos.unisul.br/index.php/Poiesis/article/view/1961/2708>, visitado em: 01/12/2018.

Pátaro, Cristina Saitê de Oliveira, e Frank Antonio Mezzomo. “Sistema Nacional de Pós-Graduação no Brasil: estrutura, resultados e desafios para política de Estado - Lívio Amaral”. Revista Educação e Linguagens 2, nº 3 (julho de 2013): 11–17.

Fernandes, Jorge H C, e Ricardo Barros Sampaio. “DataScienceForAll”. Zotero, 13 de junho de 2018