

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS, TECNOLOGIAS E SAÚDE DO CAMPUS ARARANGUÁ
CURSO DE ENGENHARIA DA COMPUTAÇÃO

KAUÊ BÜRCEL DA SILVA

INVESTIMENTOS EM TI: VISUALIZAÇÃO DOS DADOS DO CNPQ

Araranguá
2021

KAUÊ BÜRGE DA SILVA

INVESTIMENTOS EM TI: VISUALIZAÇÃO DOS DADOS DO CNPQ

Trabalho apresentado como requisito para obtenção do título de bacharel no Curso de Graduação em Engenharia de Computação do Centro de Ciências, Tecnologia e Saúde do Campus Araranguá da Universidade Federal de Santa Catarina.

Orientador(a): Dra. Luciana Bolan Frigo

Araranguá

2021

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

da Silva, Kauê Bürgel
Investimento em TI: Visualização dos dados do CNPq /
Kauê Bürgel da Silva ; orientador, Luciana Bolan Frigo,
2021.
64 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Campus Araranguá,
Graduação em Engenharia de Computação, Araranguá, 2021.

Inclui referências.

1. Engenharia de Computação. 2. CNPq. 3. Ciência de
Dados. 4. Mineração de Dados. 5. Investimentos. I. Frigo,
Luciana Bolan. II. Universidade Federal de Santa Catarina.
Graduação em Engenharia de Computação. III. Título.

KAUÊ BÜRCEL DA SILVA

INVESTIMENTOS EM TI: VISUALIZAÇÃO DOS DADOS DO CNPQ

Este trabalho de conclusão de curso foi julgado adequado para a obtenção de título de bacharel em Engenharia da Computação do Centro de Ciências, Tecnologia e Saúde do Campus Araranguá da Universidade Federal de Santa Catarina.

Orientadora: Profa. Dra. Luciana Bolan Frigo

Aprovado em 06 de abril de 2021.

BANCA EXAMINADORA:

Profa. Dra. Luciana Bolan Frigo
Orientadora

Profa. Dra. Andréa Sabedra Bordin
Avaliadora

Profa. Dra. Olga Yevseyeva
Avaliadora

Dedicado a Liane, meu porto seguro,
minha mãe.

AGRADECIMENTOS

Agradeço primeiramente a Deus, pois, sem ele nada iria existir. Sou grato a minha mãe Liane Terezinha Bürgel que fez do meu sonho possível, aguentou toda a barra comigo e jamais me deixou sozinho em nenhuma parte dessa jornada. A minha família que sempre esteve ao meu lado independente das circunstâncias e a distância. Agradeço a minha namorada por me aturar em todos os surtos e tensões que houve no decorrer do curso e ainda ficar ao meu lado. Agradeço aos professores por compartilhar o conhecimento e experiências de forma mais simples e clara e a minha orientadora que fez desse meu sonho possível.

RESUMO

Em torno de 90% das pesquisas científicas produzidas no Brasil decorrem de programas de pós-graduação de universidades públicas e o CNPq é uma agência de fomento à pesquisa em diferentes áreas com um papel essencial no financiamento das atividades desses pesquisadores. No setor de Tecnologias da Informação, os investimentos em pesquisa revertem em inúmeros benefícios, como o alcance de tecnologias para colocar o país em evidência, com maior competitividade produtiva, além de auxiliar as empresas a inserirem inovações em seu cotidiano sem que tenham que adquirir de fora do país os insumos para isso. Este estudo foi desenvolvido com o objetivo de permitir uma visualização gráfica dos dados referentes à oferta de bolsas e auxílios pagos pelo CNPq entre os anos 2002 a 2020. O trabalho apresenta uma pesquisa exploratória, na qual foram realizadas análises de dados históricos dos projetos e bolsas ofertados CNPq com um olhar mais criterioso para a área de Tecnologias da Informação. O estudo permitiu visualizar uma crescente até o ano de 2014, porém, uma crise econômica levou a cortes importantes nas mais diversas áreas de pesquisa, verificando-se uma queda em torno de 25% na quantidade de projetos atendidos e 50% em termos de valor investido. Observa-se também uma concentração de projetos nas capitais litorâneas. Houve um aporte grande de recursos durante o programa Ciências sem Fronteiras, que proporcionou muitos estudantes de graduação realizarem parte dos estudos em universidades no exterior, havendo uma concentração de bolsas em países europeus.

Palavras-chave: Tecnologia da informação. Ciência de dados. Dados abertos.

ABSTRACT

Around 90% of scientific research produced in Brazil comes from postgraduate programs at public universities. CNPq is an important funding agency that supports projects and researchers. In the Information Technology sector investments in research allow to develop industry and produce innovations. The goal of this study is to create a graphic visualization of the data about scholarships and grants paid by CNPQ between the years 2002 to 2020. It is an exploratory research, in which historical data is analysed and the Information Technologies area is explored. The study allowed to see an increase until 2014, however, with a drop of around 25% in the number of projects served and 50% in terms of value invested. There is also a concentration of projects in the Brazilian cost cities. There was a large contribution of resources during the Science without Borders program, which allowed many undergraduate students to study abroad, with a concentration of scholarships in European countries.

Keywords: Information technology. Data Science. Open data.

LISTA DE FIGURAS

Figura 1 – <i>Dataframe</i>	18
Figura 2 – Valores ausentes	19
Figura 3 – Função	21
Figura 4 – Geração de gêneros.....	21
Figura 5 – Latitude e longitude	22
Figura 6 – Geração de Mapas.....	23
Figura 7 – Linhas com valor de lista.....	23
Figura 8 – Investimento em Pesquisa Avançada por tipo de programa	27
Figura 9 – Recursos de 2011 a 2014 (em bilhões).....	30
Figura 10 – Investimentos entre 2000 e 2018 (em bilhões)	31
Figura 11 – Participação entre 2000 e 2018 em uma escala global (em %)	31
Figura 12 – Crescimento dos investimentos em P&D entre 2000 e 2018 (em %).....	32
Figura 13 – Total de projetos (todas as áreas do conhecimento).....	33
Figura 14 – Total de investimentos entre 2002 e 2020.....	34
Figura 15 – Total de projetos por gênero.....	35
Figura 16 – Total de projetos nas áreas de TI	36
Figura 17 – Valor investido nas áreas de TI	36
Figura 18 – Percentual de projetos por gênero	37
Figura 19 – Quantidade de projetos em TI por gênero – 2002 a 2010.....	39
Figura 20 – Quantidade de projetos em TI por gênero - 2011 a 2020.....	39
Figura 21 – Percentual da quantidade de projetos por área de TI	40
Figura 22 – Percentual do valor total investido por áreas de TI	40
Figura 23 –: Valor único por área da TI	41
Figura 24 – Quantidade de projetos por áreas da TI por gênero.....	42
Figura 25 – Quantidade de projetos por subáreas da TI – 2002 a 2006	43
Figura 26 – Quantidade de projetos por subáreas da TI – 2007 a 2011.....	43
Figura 27 – Quantidade de projetos por subáreas da TI – 2012 a 2016	44
Figura 28 – Quantidade de projetos por subáreas da TI – 2017 a 2020	44
Figura 29 – Quantidade de projetos por subáreas da TI por gênero.....	45
Figura 30 – Quantidade de projetos por subáreas da TI – 2002 a 2020	45
Figura 31 – Percentual de valor investido por subáreas da TI – 2002 a 2020.....	46

Figura 32 – Valor único por subáreas da TI – 2002 a 2020.....	46
Figura 33 – Quantidade de projetos por subáreas da TI no exterior – 2002 a 2020 .	47
Figura 34 – Percentual do valor investido por subáreas da TI no exterior– 2002 a 2020	48
Figura 35 – Quantidade de projetos no exterior por áreas da TI - 2012 a 2017	49
Figura 36 – Percentual de projetos no exterior no programa Ciências sem Fronteiras por subáreas da TI – 2012 a 2020	50
Figura 37 – Quantidade de projetos no exterior no programa Ciências sem Fronteiras das grandes áreas da TI – 2012 a 2020.....	51
Figura 38 – Quantidade de projetos no exterior no programa Ciências sem Fronteiras top 10 áreas da TI – 2012 a 2017	51
Figura 39 – Quantidade total de projetos no exterior no programa Ciências sem Fronteiras top 10 áreas da TI– 2012 a 2017	52
Figura 40 – Quantidade total de projetos no exterior no programa Ciências sem Fronteiras todas as áreas de acordo com a modalidade – 2012 a 2017.....	53
Figura 41 – Quantidade total de projetos no exterior no programa Ciências sem Fronteira áreas de TI de acordo com a modalidade – 2012 a 2017	54
Figura 42 – Mapa de calor por países.....	55
Figura 43 – Mapa de calor por estado.....	56
Figura 44 – Mapa de calor da região Sul por cidades	57

SUMÁRIO

1 INTRODUÇÃO	11
1.1 CONTEXTUALIZAÇÃO	11
1.2 JUSTIFICATIVA	13
1.3 OBJETIVOS	14
1.3.1 Objetivo geral	14
1.3.2 Objetivos Específicos	14
1.4 METODOLOGIA.....	15
1.4.1 Bibliotecas	16
1.4.2 Importação dos dados	17
1.4.3 Detalhamento dos dados.....	17
1.4.4 Limpeza dos dados	18
1.4.5 Geração estimativa dos gêneros.....	21
1.4.6 Geração dos mapas de calor	22
1.4.7 Preparação dos <i>dataframes</i>	23
1.5 ORGANIZAÇÃO DO ESTUDO.....	24
2 TRABALHOS RELACIONADOS	25
3 ANÁLISE DOS DADOS	33
3.1 ANÁLISE DE TODAS AS ÁREAS	33
3.2 ANÁLISE DAS ÁREAS DE TI.....	35
3.2.1 Geral.....	35
3.2.2 Áreas	38
3.2.3 Subáreas	42
3.3 ANÁLISE DAS ÁREAS DE TI NO EXTERIOR	47
3.3.1 Ciências sem fronteiras	48
3.4 MAPAS DE CALOR	54
3.4.1 Áreas de TI no Exterior – por países	54
3.4.2 Áreas de TI no Brasil – por estados	55
3.4.3 Áreas de TI na Região Sul (RS SC PR) – por cidades.....	56
4 CONSIDERAÇÕES FINAIS	58
REFERÊNCIAS	61

1 INTRODUÇÃO

Este capítulo aborda a contextualização do tema, a justificativa do estudo, seus objetivos, a metodologia utilizada e a organização deste trabalho.

1.1 CONTEXTUALIZAÇÃO

As universidades públicas, em seus programas de pós-graduação, respondem por 90% das pesquisas realizadas no país. Pesquisas fomentam o processo de desenvolvimento da ciência, ao mesmo tempo em que permitem aos esforços capitalistas incorporar seus resultados tecnológicos para o alcance de seus objetivos. Atualmente, o CNPq é uma instituição de fomento à pesquisa científica nas mais diversas áreas, com foco na expansão de conhecimentos direcionados para a sociedade, investindo em áreas diversas, dentre elas a área de Tecnologia da Informação – TI (RIBEIRO *et al.*, 2020).

A existência de um portal de transparência com dados a respeito dos investimentos em pesquisa é de grande importância, já que a ciência somente será valorizada e terá seu papel reconhecido quando for confiável em todas as suas atividades. Assim, transparência na área de pesquisas é um meio essencial para que esses investimentos sejam acompanhados e possam se tornar cada vez maiores (MARTINS, 2020).

Dados apontam que os investimentos em TI são importantes em função dessa área permitir a tomada de decisões em diferentes setores do cotidiano de forma organizada e precisa, com base em dados agrupados e claros que oferecem uma visão clara sobre oportunidades, desafios, etc. Se trata de uma área que jamais deixa de crescer, porém, os investimentos nela ainda não acompanham seu potencial de desenvolvimento (MARINHO; VALLANDRO; HOPPEN, 2015).

De longa data existe o conhecimento de que os investimentos em TI são essenciais para o desenvolvimento de empresas, governos, instituições educacionais, entre outras. Em um cenário globalizado, no qual as tecnologias da informação permitem acesso a informações, fornecedores e outras facilidades em tempo real e em qualquer lugar do mundo, quanto maiores os investimentos nessa área, maiores as chances de acompanhamento das inovações que ocorrem todos os dias (MALAQUIAS; ALBERTIN, 2011).

Os investimentos em TI trazem evolução, renovação e inovação para as mais diversas áreas, com isso elevam-se os lucros, mas também a possibilidade de concorrer com outras empresas, se modernizar, assumir uma nova imagem diante dos clientes e do mercado, dar acesso a um público cada vez maior, enfim, os resultados positivos são inúmeros (FARIAS *et al.*, 2011).

Apesar dessa realidade, porém, o Brasil ainda conta com investimentos limitados em pesquisas na área, eles ocorrem de forma mais pontual, por algumas empresas que desejam melhorar seus produtos ou serviços, mas não são tão amplas quanto seria necessário e benéfico.

Alguns conjuntos de dados são maiores do que as ferramentas de software comuns são capazes de gerir, chamados de big data, ainda que não haja uma definição específica de qual a proporção de dados que se enquadra como tal. Nesse sentido, big data não se trata de um conceito fechado quanto a um número específico, mas da ideia de que há um banco de dados de amplas proporções e que carrega em si informações essenciais e que não podem ser perdidas por serem amplas demais. O conjunto de todas essas informações, a forma como são armazenadas e utilizadas, seu desenvolvimento e as melhorias direcionadas às demandas do usuário configura-se como data Science, ou seja, a ciência dos dados desenvolvida e aplicada dentro de um cenário específico (SCHLAUCHER, 2019).

Entender os padrões de investimento em pesquisas é necessário, considerando-se que a pesquisa é atividade de grande valor para o desenvolvimento científico de uma nação. Quanto maiores os investimentos na pesquisa, em todas as áreas, melhores serão seus resultados em conhecimento e qualificação dos profissionais atuantes, além de gerar resultados mais amplos que extrapolam os limites das universidades e alcançam o cotidiano de todos os cidadãos (MASSI; QUEIROZ, 2015).

Paula, Jorge e Moraes (2019, p. 2) afirmam que a pesquisa, em geral, envolve a sistematização de diversos processos através dos quais se formula um problema a ser respondido, com base nas implicações de cada pesquisador, bem como nas necessidades científicas e sociais que se deseja atender. É essencial definir com clareza qual será a metodologia aplicada para seu desenvolvimento, que estratégias serão aplicadas para que se propaguem seus resultados e quais as recomendações a serem definidas quanto ao intuito de transformar as melhores práticas em políticas.

Nesse sentido, as pesquisas devem ser compreendidas como parte importante do processo de desenvolvimento em áreas variadas, levando à definição de políticas públicas mais efetivas para educação, saúde, etc.

Alves e Oliveira (2014) afirmam que os investimentos em pesquisa crescem em países que percebem que com o aumento da competitividade, o investimento em desenvolvimento científico e tecnológico é necessário para manter a nação entre as melhores do mundo. Quanto mais evoluída a pesquisa, com projetos bem formulados e resultados sólidos, melhores os resultados para o país no qual ela é realizada.

1.2 JUSTIFICATIVA

Este estudo justifica-se pela necessidade de compreender de que forma os dados de pesquisa podem ser coletados, agrupados, interpretados e avaliados para que seus resultados possam, de fato, ser compreendidos e aplicados positivamente para maior efetividade da pesquisa em todas as áreas do conhecimento e das áreas de TI, fomentando o envolvimento e o direcionamento de recursos para as áreas que estejam recebendo recursos mais limitados.

Dados abertos são informações que visam à transparência e esclarecimento em muitas atividades e, assim, permitem o acesso público para consulta, avaliação e verificação de pontos duvidosos, caso existam. As instituições que atuam com recursos municipais, estaduais ou Federais devem manter páginas de acesso a dados abertos, considerando-se que estes refletem como foram gastos os recursos que os contribuintes recolheram em determinado período (CREATIVE COMMONS, 2011).

Estão disponíveis diferentes áreas de atuação dentro das quais são importantes possuir dados abertos, acessíveis e claros. Diversos grupos de indivíduos e organizações podem ser amplamente beneficiados da disponibilidade de informações precisas a partir de práticas de transparência, iniciando-se pelo próprio governo. Além disso, não se pode saber com precisão como e onde os dados abertos serão mais importantes, já que ao permitir a inovação em diversas áreas, novas formas de uso acabarão por surgir de locais e situações não muito previsíveis (CREATIVE COMMONS, 2011, p. 7-8).

Compreende-se, assim, que os dados abertos são fontes de informação e

conhecimentos que podem ser aproveitados para finalidades diversas, mas seu intuito é o mesmo, transparência e acessibilidade, permitindo aos cidadãos terem uma percepção específica de como, quando e onde os recursos foram alocados.

No Brasil, o que é comum é que haja divulgação de alguns dados selecionados, porém, não se pode dizer que são realmente abertos. No máximo, alguns podem ser visualizados e copiados, mas as barreiras técnicas para um acesso mais amplo são consideráveis. O fato é que dado aberto é toda aquela informação que pode ser livremente utilizada, reutilizada e redistribuída por qualquer um, o que demonstra que a questão de dados abertos no Brasil ainda demanda de melhorias (CREATIVE COMMONS, 2011).

Tamanha a importância dos dados abertos para a transparência e confiabilidade que existe uma percepção de que a democracia participativa apenas poderá se configurar de forma integral quando os governos compreenderem o que são os dados abertos e atuarem para que sejam disponibilizados para fácil acesso de todos os seus cidadãos, o que assegura clareza e possibilidade de fiscalização (BRASIL, 2011).

1.3 OBJETIVOS

1.3.1 Objetivo geral

Este estudo foi desenvolvido com o objetivo de permitir uma visualização dos dados referentes à oferta de bolsas e auxílios pagos pelo CNPQ de 2002 até 2020, que são os dados disponibilizados no portal.

1.3.2 Objetivos Específicos

- Gerar a visualização gráfica dos dados;
- Classificar as áreas de TI;
- Analisar os dados de investimentos em pesquisa ao longo do tempo;
- Identificar as áreas com maior direcionamento de recursos;
- Buscar relações com o cenário político-econômico que justifiquem os investimentos.

1.4 METODOLOGIA

Este estudo aborda conjuntos de dados extremamente amplos, que encampam aproximadamente 20 anos de levantamento de dados a respeito da oferta de bolsas de pesquisa e, assim, utiliza-se de metodologias necessárias para a identificação, coleta, organização e agrupamento desses dados, para que possam ser avaliados.

O trabalho realizado caracteriza-se como uma pesquisa exploratória, em que foram realizadas análises de dados históricos dos projetos e bolsas ofertados pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) de todas as áreas de forma geral e de forma mais detalhada envolvendo as subáreas de TI, buscando informações relevantes sobre as quantidades e valores investidos nos projetos e áreas e localizações mais recebidas no Brasil e no exterior. Como os dados estão organizados por grande área, área e subáreas, foram selecionadas algumas subáreas mais próximas as Tecnologias da Informação e Comunicação, de forma a fornecer ao trabalho de pesquisa um foco em TIC.:

- Análise de Algoritmos e Complexidade de Computação
- Análise de Dados
- Banco de Dados
- Ciência da Computação
- Ciência da Informação
- Engenharia de Software
- Estatística
- Matemática da Computação
- Matemática Aplicada
- Matemática Discreta e Combinatória
- Metodologia e Técnicas da Computação
- Probabilidade e Estatística Aplicadas
- Sistemas de Computação
- Técnicas de Recuperação de Informação

- Tecnologia da Informação e Comunicação
- Tecnologias Digitais
- Teleinformática
- Teoria da Computação
- Teoria da Comunicação
- Teoria da Informação
- Teoria da Informação
- Teoria dos Grafos

1.4.1 Escolha das Bibliotecas

Com os dados definidos, a etapa seguinte consiste em utilizar a plataforma Google Colab, uma ferramenta gratuita oferecida pelo Google para realizar a programação na linguagem Python de forma online sem a necessidade de utilizar a memória e processamento do seu computador, sendo possível processar grandes volumes de dados de forma rápida. Essa linguagem permite trabalhar com Ciência de Dados e para isso existem bibliotecas que facilitam a programação. Neste trabalho as bibliotecas utilizadas foram:

- **Pandas:** a principal e mais completa biblioteca para análise de dados. Fornece ferramentas e estruturas de dados de alta performance, foi escolhida para trabalhar com *dataframes* importados dos arquivos CSV.
- **Seaborn:** é uma biblioteca de visualização de dados Python baseada em matplotlib. Ele fornece uma interface de alto nível para desenhar gráficos estatísticos atraentes e informativos. Foi escolhida por trazer os gráficos com mais robustez.
- **Plotly:** fornece gráficos, análises e ferramentas estatísticas para indivíduos e empresas, bem como bibliotecas de gráficos científicos para Python, R, MATLAB, Perl, Julia, Arduino e REST. Foi escolhida por trazer os gráficos interativos e mais elegantes.
- **Matplotlib:** Matplotlib é uma biblioteca abrangente para a criação de visualizações estáticas, animadas e interativas em Python. Utilizada para setar algumas configurações de layout dos gráficos.

- **Folium**: se baseia nos pontos fortes do ecossistema Python e nas forças de mapeamento da biblioteca Leaflet.js. Isso significa que podemos manipular os dados usando linguagem de programação e visualizar o resultado imediatamente em um mapa interativo. Escolhida para realizar os mapas de calor.
- **Name2GenderBR**: possibilita a classificação de gênero com base nos primeiros nomes, com base na frequência de nomes encontrados no Censo 2010, escolhida para gerar os gêneros dos beneficiários (BRASIL, 2021).

1.4.2 Importação dos dados

Os arquivos foram encontrados no site do Portal Brasileiro de Dados Abertos, uma ferramenta disponibilizada pelo governo para que todos possam encontrar e utilizar os dados e as informações públicas. Os dados do CNPq estão no link http://dadosabertos.cnpq.br/pt_BR/dataset e são separados por ano de investimento, cada arquivo contém o nome investimento_cnpq_20XX, onde, XX é o ano referente.

Essa base de dados contém os anos de 2002 até 2020, ou seja, 19 anos de investimentos separados em 19 arquivos no formato CSV.

Com o pandas, é utilizado o `read_csv` para ler cada arquivo e armazenar em uma tabela (*dataframe*) e após isso se utiliza o comando `pd.concat` para concatenar todas as tabelas em uma única. Após uma visualização dos dados gerais foram filtrados os dados referentes às subáreas de TI e outra para projetos no exterior.

1.4.3 Detalhamento dos dados

Analisando o *dataframe* gerado, utilizou-se o comando `dataframe.shape` para extrair a quantidade de linhas e colunas, em todas as áreas foram encontrados 2,37 milhões de linhas e nas áreas de TI 121 mil, ambos separados em 21 colunas contendo dados referente ao CNPq de 2002 a 2020. Abaixo segue as principais colunas:

Figura 1 – Dataframe

Ano Referência	Processo	Beneficiário	Linha de Fomento	Modalidade	Programa CNPq	Grande Área	Área	Subárea	Instituição Destino	Sigla Instituição Destino	Cidade Destino	Sigla UF Destino	País Destino	Valor Pago	
0	2002	130942/2000-6	Abad Oliveira Souza Júnior	Bolsas de Mestrado	GM - Mestrado	Programa Básico de Geociências: Geofísica e Ge...	Ciências Exatas e da Terra	Geociências	Geofísica Aplicada	Universidade de São Paulo	USP	São Paulo	SP	BRA - Brasil	1449.04
1	2002	131008/2001-3	Abel Batista de Oliveira	Bolsas de Mestrado	GM - Mestrado	PROGRAMA BASICO DE AGRONOMIA	Ciências Agrárias	Agronomia	Ciência do Solo	Universidade Federal Rural de Pernambuco	UFRPE	Recife	PE	BRA - Brasil	7245.20
2	2002	131008/2001-3	Abel Batista de Oliveira	Bolsas de Mestrado	GM - Mestrado	PROGRAMA BASICO DE AGRONOMIA	Ciências Agrárias	Agronomia	Ciência do Solo	Universidade Federal Rural de Pernambuco	UFRPE	Recife	PE	BRA - Brasil	1449.04
3	2002	132920/2002-6	Abel Bemvenuti	Bolsas de Mestrado	GM - Mestrado	PROGRAMA BASICO DE BIOFISICA	Ciências Biológicas	Biofísica	Biofísica	Universidade Federal do Rio de Janeiro	UFRJ	Rio de Janeiro	RJ	BRA - Brasil	3622.60
4	2002	141578/2002-5	Abel Guilhermino da Silva Filho	Bolsas de Doutorado	GD - Doutorado	PROGRAMA BASICO DE CIENCIA DA COMPUTACAO	Ciências Exatas e da Terra	Ciência da Computação	Sistemas de Computação	Universidade Federal de Pernambuco	UFPE	Recife	PE	BRA - Brasil	5364.45

Fonte: Dados do estudo (2021).

Para saber os valores únicos de cada coluna, foi utilizado o `dataframe.value_counts()`, sendo as principais colunas:

Todas as Áreas:

- 11 Grandes Áreas
- 135 Áreas
- 1.389 Subáreas
- 23 Linhas de Fomentos
- 76 Modalidades
- 79 Países
- 388 Programas CNPq

Áreas de TI:

- 4 Grandes Áreas
- 8 Áreas
- 21 Subáreas
- 22 Linhas de Fomentos
- 57 Modalidades
- 41 Países
- 140 Programas CNPq

1.4.4 Limpeza dos dados

Com os dados já coletados, a próxima etapa é realizar a análise dos dados faltantes e preenchê-los de acordo com cada caso. O comando `dataframe.isnull()` retorna a quantidade de valores nulos em cada coluna, pegando esse valor e dividindo pela quantidade de dados `dataframe.shape[0]` e multiplicando por 100, obtemos o percentual de valores faltantes em cada coluna:

Figura 2 – Valores ausentes

```
# Percentual dos Valores Ausentes
valor = (dataframe.isnull().sum() / dataframe.shape[0] * 100)
valor.sort_values(ascending=False).round(2).astype(str) + '%'
```

Categoria/Nível	74.48%
Sigla UF Origem	10.73%
Instituição Origem	9.49%
Sigla UF Destino	4.21%
Nome Chamada	2.41%
Cidade Destino	0.64%
Sigla Instituição Destino	0.53%
Sigla Instituição Macro	0.52%
Instituição Destino	0.52%
Área	0.38%
Valor Pago	0.04%
Beneficiário	0.0%
Processo	0.0%
Linha de Fomento	0.0%
Modalidade	0.0%
Subárea	0.0%
Programa CNPq	0.0%
Grande Área	0.0%
País Destino	0.0%
País Origem	0.0%
Ano Referência	0.0%

Fonte: Dados do estudo (2021).

É possível observar que a Categoria/Nível possui mais de 70% de seus dados faltantes, seguido pela Sigla UF Origem com quase 11%, porém, não iremos utilizar todas essas colunas, então, na primeira etapa foram removidas as colunas que não serão utilizadas: Processo, Categoria/Nível, Instituição Origem, Nome Chamada, Instituição Destino, Sigla Instituição Destino, Sigla Instituição Macro, Sigla UF Origem, País Origem, Linha de Fomento. Para removê-las, utiliza-se o comando `dataframe.drop(columns=['coluna'])`.

Para o preenchimento de dados irregulares ou faltantes é possível utilizar a moda, ou seja, o valor que mais aparece na coluna, já para dados numéricos é utilizado a mediana, valor central do conjunto numérico.

- **Mediana:** `dataframe['coluna'].median()`
- **Moda:** `dataframe['coluna'].mode()`

Com os valores respectivos de cada campo, utiliza-se a função `dataframe['coluna'].fillna('valor', inplace=True)` que irá preencher os campos vazios da coluna com o valor escolhido. Abaixo, segue as colunas e os respectivos tratamentos dos dados para as todas as áreas, sendo que as subáreas de ti não apresentam dados faltantes ou irregulares:

- **Cidade Destino**
 - Dados faltantes: 15.211
 - Moda: Rio de Janeiro
- **Sigla UF Destino**
 - Dados Faltantes: 100.015
 - Moda: SP
- **Área**
 - Dados faltantes: 8.975
 - Moda: Matemática
- **Valor Pago**
 - Dados faltantes: 923
 - Mediana: 2.800,00

Após preencher todos os dados faltantes, é hora de analisar os dados preenchidos, pois, é possível que contenha erros de digitação dos dados.

Uma forma de verificar é utilizando a função `dataframe['coluna'].value_counts()`, que retorna o valor único de cada coluna e com isso é possível ver se todos os valores únicos estão de acordo com o esperado. Já para verificar os valores numéricos o `dataframe['coluna'].describe()` nos fornece dados como o valor mínimo, valor máximo, média, entre outros, sendo assim possível verificar se há valores negativos.

Na coluna Ano Referência foi encontrado o ano 206 e substituído por 2006 e na coluna Valor Pago foi encontrado um valor negativo e substituído pela mediana. Nesses casos para substituir os valores é necessário utilizar uma função que localiza um registro no `dataframe` e substitua pelo valor novo, segue a função:

Figura 3 – Função

```
dataframe.loc[dataframe['coluna'] == valor, 'coluna'] = novo valor
```

Fonte: Dados do estudo (2021).

Com o conjunto de dados preparados, tornou-se necessário separar o *dataframe* de todos os dados nos seguintes *dataframes*: todas as áreas no brasil e no exterior, áreas de ti no brasil, no exterior e na região sul do brasil (RS,SC,PR) e também criar um *dataframe* que contenha somente dados do Programa CNPq – Ciências sem Fronteiras de todas as áreas e das áreas de ti.

1.4.5 Geração da estimativa dos gêneros

O conjunto de dados não contém a coluna Gêneros, e para se obter uma estimativa desses valores, foi utilizado a biblioteca *Name2GenderBR* para obter o gênero de acordo com o nome do beneficiário do programa, esse valor encontra-se na coluna Beneficiário. Segue o código utilizado:

Figura 4 – Geração de gêneros

```
# Instala o Pacote Name2GenderBR
!pip install Name2GenderBR

from Name2GenderBR import GenderClassifier

# Criando nova coluna Generos
dataframe = dataframe.assign(Gênero = "")

# Preenche cada linha com o Gênero
for indice,linha in dataframe.iterrows():
    dataframe.loc[indice,'Gênero'] = GenderClassifier().get_gender(linha['Beneficiário'])
```

Fonte: Dados do estudo (2021).

O código acima funciona da seguinte forma, é criado no *dataframe* uma coluna nova Gêneros e preenchida com valores vazios, após isso, é utilizado uma estrutura de repetição *for* para percorrer cada linha do *dataframe* e preenche-las com o retorno da função *GenderClassifier().get_gender('nome')* que retornará os valores “M” para sexo masculino e “F” para sexo feminino.

1.4.6 Geração dos mapas de calor

Para realizar o mapa de calor, foi feita uma busca das coordenadas das cidades, estados e país, onde contenham a latitude e longitude. Os dados foram exportados para um *dataframe* auxiliar utilizando a função *dataframe.read_csv*. Os dados foram encontrados no GitHub, em formato CSV, facilitando a importação. Segue abaixo os links:

- **Países:**
<https://raw.githubusercontent.com/google/dspl/master/samples/google/canonical/countries.csv>
- **Estados:** <https://raw.githubusercontent.com/kelvins/Municipios-Brasileiros/main/csv/estados.csv>
- **Cidades:** <https://raw.githubusercontent.com/alanwillms/geoinfo/master/latitude-longitude-cidades.csv>

Os mapas de calor foram divididos em mapa de calor no exterior, que é separado de acordo com o país, mapa de calor do Brasil, que é separado por estado e por último mapa de calor da região sul (RS|SC|PR), que é separado de acordo com a cidade. Segue o código que preenche as latitudes e longitudes no *dataframe*:

Figura 5 – Latitude e longitude

```
# Criando novas colunas
dataframe = dataframe.assign(Latitude = "", Longitude = "")

# Preenche as Longitudes e Latitudes
for indexSub, rowSub in dataframe.iterrows():
    for indexEst, rowEst in dataframeAux.iterrows():
        if rowSub['Valor'] == rowEst['Valor']:
            dataframe.loc[indexSub, 'Latitude'] = dataframeAux.loc[indexEst, 'latitude']
            dataframe.loc[indexSub, 'Longitude'] = dataframeAux.loc[indexEst, 'longitude']
```

Fonte: Dados do estudo (2021).

Na primeira linha do código são geradas duas colunas novas vazias no *dataframe*, com o nome de Latitude e Longitude, após isso são utilizadas duas estruturas de repetição *for*, que para cada linha do *dataframe* ele percorra todo *dataframeAux* e através do *if* para verificar se a linha é igual em ambos *dataframes* de acordo com a coluna escolhida, então ele preenche os valores da Latitude e

Longitude. Esse código está de forma genérica, no campo coluna, é utilizada a coluna de acordo com o código e no *dataframeAux* é utilizado o *dataframe* de acordo com a coluna escolhida, ou seja, foi feito um código para cada mapa de calor.

Para geração do mapa é utilizado a biblioteca *folium*, utilizando o *dataframe* e acordo com mapas específicos, sendo eles dos Países, Estados e Cidades. Segue o código:

Figura 6 – Geração de Mapas

```
from folium import plugins
sys.setrecursionlimit(10**6)

# Percorre o dataframe e pega as latitude e longitude
coordenadas=[]
for lat,lng in zip(dataframe['Latitude'].values,dataframe['Longitude'].values):
    coordenadas.append([lat,lng])

# Gera o Mapa de Calor
mapa = folium.Map(location=[0,0],zoom_start=1, min_zoom=1, tiles='CartoDB Positron')
mapa.add_child(plugins.HeatMap(coordenadas))

# Imprime o Mapa
mapa
```

Fonte: Dados do estudo (2021).

1.4.7 Preparação dos *dataframes*

A geração dos *dataframes* específicos é feita utilizando uma lista com os valores que se deseja filtrar e através da função *dataframe.isin* são filtradas as linhas que possuem o valor da lista de acordo com a coluna específica:

Figura 7 – Linhas com valor de lista

```
# Código para Gerar Dataframe Especifico
lista = ["Valores..."]
dataframenovo = dataframe[dataframe['coluna'].isin(lista)]
```

Fonte: Dados do estudo (2021).

Com o código acima foram criados os *dataframes* das áreas de TI escolhidas e dos programas do CNPq referentes ao programa Ciências sem Fronteiras. Dessa forma, todos os dados estão 100% preenchidos e tratados validando assim nossos *dataframes* para serem analisados.

1.5 ORGANIZAÇÃO DO ESTUDO

Para a melhor evolução do estudo, este foi organizado em forma de capítulos da seguinte forma: no capítulo 2 são apresentados os trabalhos relacionados, no capítulo 3 tem-se a análise dos dados e por fim, no capítulo 4 as considerações finais.

2 TRABALHOS RELACIONADOS

O CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico atua, desde a década de 1970, distribuindo bolsas de produtividade em pesquisa, disponibilizadas em edital, para pesquisadores brasileiros que tenham produtividade científica considerável. Tais bolsas, além de serem de grande auxílio aos pesquisadores, para que possam dedicar-se à pesquisa, melhorando seus conhecimentos e alcançando níveis ainda maiores de formação, estimulam o desenvolvimento científico no país (SPILKI, 2013).

No mundo, os programas de oferta de bolsas e estímulo à pesquisa são muitos, em sua maioria com uma destinação de recursos muito maior do que ocorre no Brasil, por haver no cenário internacional uma tradição muito mais antiga de investimento na pesquisa e no desenvolvimento dos profissionais, especialmente em níveis superiores, como as áreas de especialização. O Brasil ainda está construindo seus caminhos nesse esforço e o fato é que conhecer os indicadores de investimentos em pesquisa é importante para demonstrar qual foi o crescimento e o que pode ser feito para que se torne ainda maior no futuro (CASTIONE; MELO, 2020).

Em todas as áreas nas quais as pesquisas são conduzidas ocorre uma expansão da base de conhecimentos e a possibilidade de que esses novos saberes sejam compartilhados com outros acadêmicos, professores ou profissionais de diversas outras áreas. No Brasil, o financiamento da pesquisa com recursos públicos não é uma prática tão ampla quanto seria ideal e, assim, o CNPq representa um esforço para o avanço das ciências em suas mais diversas áreas e com benefícios mais amplos do que se imagina. Em 1975 o Brasil contava com 370 programas de mestrado, que passaram para 784 em 1985, 1.629 em 2014, 1.995 em 2016 e 2.077 em 2017. Em 1975 os professores envolvidos em programas de pós-graduação eram 7.500, em 1985 eram 20.900, 40.725 em 2004, 85.418 em 2014, 90.130 em 2015, 95.182 em 2016 e 100.287 em 2017 (FERREIRA; ARAÚJO, 2019).

Após a promulgação da Constituição Federal de 1988, que estabeleceu que a educação é um direito de todos e dever do Estado de ofertar meios de acesso a todos, o que se verificou foi uma mudança no cenário de oferta de bolsas e incentivo à especialização em diversas em diversas áreas, sendo que ao longo dos últimos 30 anos, depois da promulgação da Constituição Federal de 1988, as mudanças foram

intensas no cenário da Educação Superior no país, não apenas na graduação como também na pós graduação (CASTIONI; MELO, 2020, p. 2).

O número de alunos aumentou em mais de 6 vezes, o sistema de ensino superior brasileiro passou a ser o quinto maior do mundo, com mais de 8 milhões de alunos (atrás da China, Índia, Estados Unidos e África do Sul). Assim durante esse período, o país passou de um sistema de elite para um sistema de acesso de massa (CASTIONI; MELO, 2020, p. 2).

É preciso esclarecer que muitas dessas participações em programas de pós-graduação (mestrado, doutorado, etc.) só foram possíveis em função do financiamento através de bolsas de estudo, o que demonstra que essa prática gera ao país a disponibilidade de profissionais mais bem preparados para atuar em suas áreas do conhecimento. As pesquisas financiadas pelo CNPq visam incentivar esforços e promover o desenvolvimento científico e tecnológico, permitindo que sejam desenvolvidas pesquisas necessárias para que se alcance progresso social, econômico e cultural em todo o país (FERREIRA; ARAÚJO, 2019).

Cátion e Melo (2020) afirmam que o ensino superior no país, com foco nos programas de qualificação e especialização, apresentou uma importante elevação de qualidade, em função de políticas públicas efetivas e bem desenvolvidas, mas também em face de um sistema de inovações que envolve o desenvolvimento por meio da pesquisa, diferente do que ocorre nas etapas da educação básica, por exemplo.

A pesquisa é essencial para o desenvolvimento científico das mais diversas áreas do conhecimento, como a educação. Sem que haja possibilidade de desenvolver projetos de pesquisa aprofundados e claro, muitos vieses e tendências podem deixar de ser percebidos e, assim, a evolução dos saberes na área tornam-se mais lentos ou desviam-se dos padrões identificados em um cenário real (FERREIRA; ARAÚJO, 2019).

Moura e Camargo Júnior (2017) ressaltam que entre 2007 e 2014 os financiamentos de pesquisa foram mais acentuados, pela disponibilidade maior de recursos para essa área, enquanto em 2015 uma crise fez com que os recursos destinados à pesquisa fossem reduzidos na maioria dos estados brasileiros. Com isso, não apenas a pesquisa reduziu sua produtividade, com as áreas em que esses pesquisadores atuam passaram por uma queda importante na qualificação de profissionais e sua oferta ao mercado, com benefícios para o desenvolvimento do

país de forma geral.

Em 2014, Alves e Oliveira (2014) avaliaram investimento em pesquisa na área de Ciências e Tecnologia (C&T) entre 2002 e 2010. As pesquisas na área no período avaliado receberam investimentos crescentes, porém, deve-se ressaltar que financiamentos privados aproximaram-se em percentual dos financiamentos públicos, o que indica que as pesquisas na área ainda são financiadas, em boa proporção, por iniciativas privadas que utilizarão os resultados para seus interesses de competitividade e desenvolvimento financeiro e tecnológico.

Os autores verificaram, ainda, que em um ranking mundial de 20 países, em 2014, o Brasil ocupava o 13º lugar em produção científica na área de C&T, com EUA, China e Reino Unido ocupando as 3 primeiras posições, a Polônia ocupava o 20º lugar. Em 2010 o Brasil produziu mais de 54% dos artigos da América Latina publicados na base Scopus, uma base reconhecida mundialmente, em 2000 essa produção era de 43,3% (ALVES; OLIVEIRA, 2014).

Os achados dos autores quanto aos valores investidos no período constam da Figura 8.

Figura 8 – Investimento em Pesquisa Avançada por tipo de programa

Período	Formação e Capacitação de Recursos Humanos para CT&I	Promoção da Pesquisa e do Desenvolvimento Científico e Tecnológico	CT&I para Política Industrial, Tecnol. e de Comércio Exterior - PITCE	Gestão da Política de Ciência, Tecnologia e Inovação	Ciência, Tecnologia & Inovação Aplicada aos Recursos Naturais	Outras Ações	Total Geral
2002	432.937	0	0	1.584	0	269.247	703.768
2003	461.338	0	0	4.616	0	410.693	876.647
2004	545.151	124.366	0	4.473	0	272.622	946.612
2005	598.591	105.456	0	4.746	0	316.981	1.025.774
2006	626.994	128.599	154.933	5.632	35.991	67.732	1.019.881
2007	657.062	147.075	177.368	5.958	41.877	173.950	1.203.290
2008	588.700	144.642	419.042	4.990	32.891	233.358	1.423.623
2009	748.121	162.406	405.673	4.388	35.644	310.163	1.666.395
2010	864.722	162.252	630.164	2.192	27.193	260.966	1.947.489

Fonte: Alves e Oliveira (2014, p. 167).

Cândido, Santos e Rocha (2016) coletaram dados de investimento em pesquisas na área de química, avaliando a concessão de bolsas do CNPq em 2013, avaliando um período de 10 anos (2003 a 2013, data de seleção) com o intuito de verificar o perfil dos pesquisadores que obtêm incentivos à pesquisa. De acordo com os pesquisadores, os dados levantados a partir de pesquisas no site do CNPq, bem como dados recebidos da instituição demonstram uma predominância de pesquisadores do sexo masculino com projetos aprovados para financiamento em

sua área de especialização. O número de mulheres com financiamento na área vem aumentando, porém, ainda é bastante inferior ao número de homens. A produção de artigos foi considerável no período, com algumas áreas sendo mais investigadas do que outras, possivelmente em função da especialização dos bolsistas e sua área de atuação (CÂNDIDO; SANTOS; ROCHA, 2016).

Cruz, Silva e Fraga (2016) avaliaram os investimentos do CNPq em bolsas de mestrado e doutorado na área de educação, por meio de mapeamento de dados no site da instituição. Os dados foram coletados para demonstrar as bolsas de mestrado e doutorado na grande área do conhecimento de Ciências Humanas, sendo que a área de Educação foi o foco, com a demonstração dos resultados quantitativos por regiões brasileiras e seus estados. Os dados referem-se a março de 2016 quando estavam ativas 9.189 bolsas de mestrado financiadas pelo CNPq, dessas 1.793 eram para a área de engenharias, 1.419 para ciências agrárias e 1.355 para ciências humanas (cerca de 14,75% do total). Linguística, letras e artes é a área com menor número de bolsas de mestrado, somente 4,45% do total. Dentro da área de ciências humanas (1.355 bolsas), 285 eram de mestrado em educação, 275 em psicologia e 229 em história.

Cátion e Melo (2020) procederam de um levantamento a respeito da oferta de bolsas pelo CNPq, porém, somente na área de educação, visando compreender o quanto vem sendo investido no desenvolvimento de professores brasileiros. Os procedimentos envolveram, primeiramente, a coleta dos dados: os dados foram coletados a partir do site Investimentos do CNPq em C&T3 – no qual foram identificados os dados do pesquisador, título do seu projeto, a universidade de vínculo e o período da bolsa – e da Plataforma Lattes, visando a complementar as informações sendo extraídos os dados geográficos do pesquisador/instituição, número de publicações e orientações. Depois ocorreu a categorização dos projetos: etapa responsável por classificar o projeto, por meio da análise do título, dentro de cinco categorias definidas (quadro 1): Desenvolvimento Conceitual, Educação Básica, Docente (formação, práticas pedagógicas), Educação Superior, Política Pública e Metodologia de Ensino. Por fim, os autores procederam da análise dos resultados: responsável por identificar correlações dos dados analisados conforme o objetivo do trabalho (CASTIONI; MELO, 2020).

Os dados cobriram o período de 2018. Os autores verificaram que a região sudeste recebeu 56% das bolsas, a região sul 30%, e somente 14% do total foram

distribuídas entre centro-oeste (4%), nordeste (8%) e norte (2%). Desse total de bolsas, a produtividade avaliada através das publicações dos participantes de artigos em periódicos, livros e capítulos e orientações em andamento, esclarecendo que as produções se concentram, em sua maioria, nas regiões dos financiamentos, o que demonstra que as demais áreas poderiam produzir mais, porém, não recebem incentivos para isso.

Das diversas categorias nas quais as pesquisas podem enquadrar-se, 30% delas tiveram foco no desenvolvimento conceitual, 25% no ensino superior, 10% na educação básica, 18% na formação docente e 10% em políticas públicas e 7% na metodologia de ensino. Apesar de haver um investimento importante em desenvolvimento de estudos na área de educação básica, o país ainda apresenta indicadores de qualidade relativamente baixos nessa área, mostrando o subaproveitamento desses conhecimentos para a evolução na área de educação (CASTIONI; MELO, 2020).

Como limitações do estudo destaca-se o fato de que os dados levantados concentram-se apenas em uma área, ainda que fosse o foco do estudo verificar os dados de investimentos em pesquisas na área de educação, cruzar dados com outras áreas do conhecimento poderia indicar se, de fato, há um déficit de investimentos em uma área tão importante para o desenvolvimento social ou se há similaridade com outras áreas do conhecimento.

O direcionamento dos recursos segue parâmetros de atendimento aos critérios dos editais e, assim, podem ser considerados efetivos e dentro dos parâmetros de legalidade de acordo com as leis vigentes no país (RIBEIRO et al., 2020).

Cada país conta com suas próprias políticas de investimento em pesquisas, além de haver, em muitos locais, incentivo para que empresas privadas financiem parte do desenvolvimento tecnológico que possa ser aplicado em suas próprias atividades. Não se pode ignorar, porém, os impactos econômicos e tecnológicos da pesquisa para nações com maiores índices de investimento, considerando-se que os conhecimentos resultantes desses programas trazem novas visões, percepções e possíveis soluções a problemas diversos, que além de beneficiarem as pessoas podem se tornar atrativos ao mercado consumidor (BLOCH; SORENSEN, 2015).

Investimentos em pesquisa e desenvolvimento (P&D) desempenham um papel central nas economias avançadas, em áreas como crescimento econômico e

criação de empregos, competitividade industrial, segurança nacional, energia, agricultura, transporte, saúde pública e bem-estar, proteção ambiental e expansão das fronteiras de compreensão do conhecimento humano. Assim, empresas, governos, universidades, organizações sem fins lucrativos e outros em todo o mundo fazem investimentos substanciais em P&D. Desde 2000, os gastos globais totais com P&D mais do que triplicaram em dólares, de US \$676 bilhões para US \$2,0 trilhões em 2018 (CONGRESSIONAL RESEARCH SERVICE, 2020).

Os Estados Unidos emergiram como um líder global em ciência e tecnologia na segunda metade do século XX período no qual os investimentos públicos e privados dos EUA em P&D cresceram rapidamente e ajudaram a impulsionar o país para uma posição de liderança econômica global. Em 1960, os Estados Unidos respondiam por aproximadamente 69% do financiamento mundial de P&D. Em 2018, no entanto, a participação dos EUA nos gastos globais em P&D havia caído para em torno de 28%. O declínio dos EUA na participação em P&D global não é o resultado de uma redução nos investimentos em P&D dos EUA - na verdade, a P&D pública e privada cresceu fortemente durante esse período, é o resultado de aumentos ainda maiores nos investimentos dos governos e indústrias de outros países, que reconheceram a importância da P&D para sua inovação e competitividade industrial (CONGRESSIONAL RESEARCH SERVICE, 2020).

Entre os 20 países no mundo com maiores investimentos em P&D, o Brasil não é elencado em nenhuma colocação, demonstrando que há necessidade urgente de maiores investimentos na área, conforme se percebe na Figura 92.

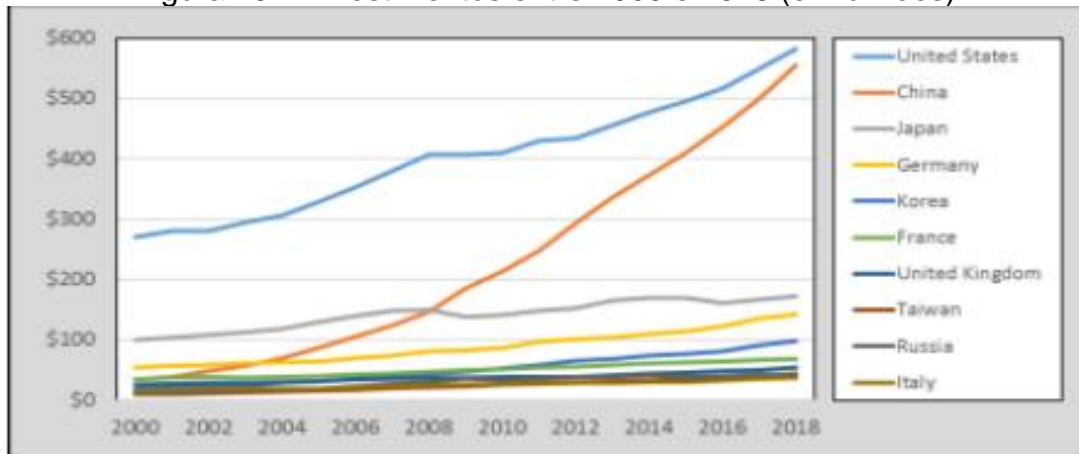
Figura 9 – Recursos de 2011 a 2014 (em bilhões)

Rank	Country	Amount	Rank	Country	Amount
1	United States	\$581.6	11	Canada	29.0
2	China	554.3	12	Spain	23.6
3	Japan	171.3	13	Australia	22.6
4	Germany	141.4	14	Turkey	21.7
5	South Korea	98.5	15	Netherlands	21.5
6	France	\$68.4	16	Switzerland	19.1
7	United Kingdom	53.1	17	Sweden	18.1
8	Taiwan	43.3	18	Israel	17.7
9	Russia	41.5	19	Belgium	16.5
10	Italy	36.0	20	Austria	16.0

Fonte: Congressional Research Service (2020, p. 2).

Os 10 primeiros países em investimentos na área de pesquisa e desenvolvimento no mundo investem de acordo com a Figura 10.

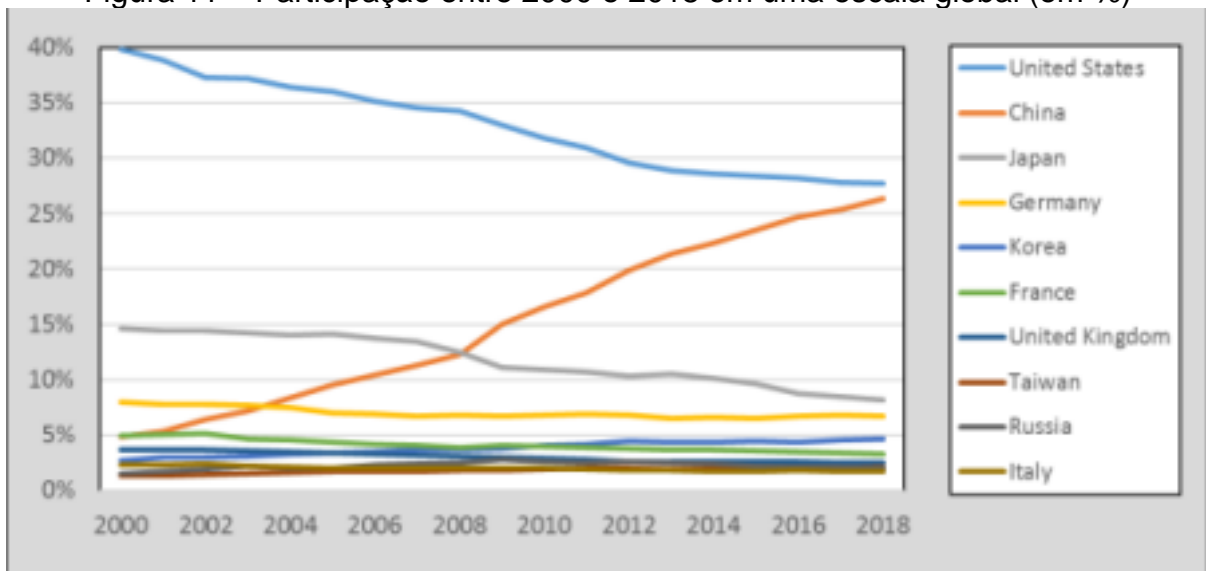
Figura 10 – Investimentos entre 2000 e 2018 (em bilhões)



Fonte: Congressional Ressarce Service (2020, p. 2).

Para compreender a participação desses países no total de investimento mundial em P&D no período de 2000 a 2018, a Figura 11 traz importantes esclarecimentos.

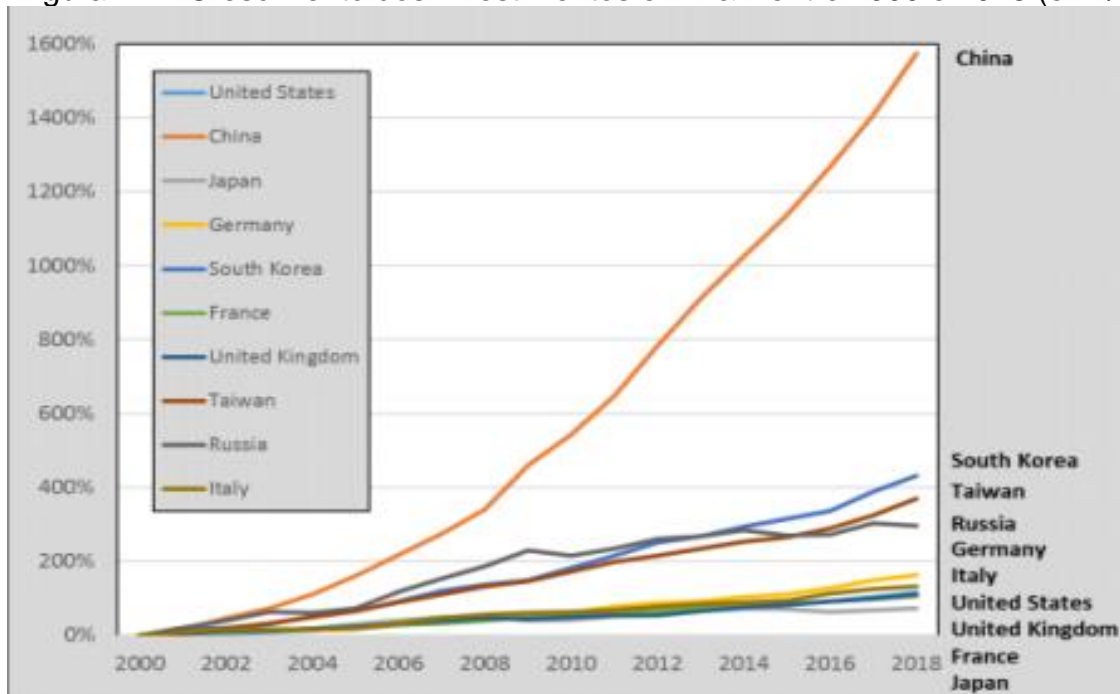
Figura 11 – Participação entre 2000 e 2018 em uma escala global (em %)



Fonte: Congressional Ressarce Service (2020, p. 2).

Por fim, para entender como ocorreu o crescimento dos investimentos em P&D nesses países entre 2000 em 2018, apresenta-se a Figura 12.

Figura 12 – Crescimento dos investimentos em P&D entre 2000 e 2018 (em %)



Fonte: Congressional Researce Service (2020, p. 4).

Verifica-se que os países mais fortes e representativos na economia mundial são aqueles com amplos investimentos em P&D nas mais diversas áreas, já que os resultados das pesquisas tendem a ser melhorias capazes de tornar mais efetivos os produtos, serviços e as políticas adotadas nesses países.

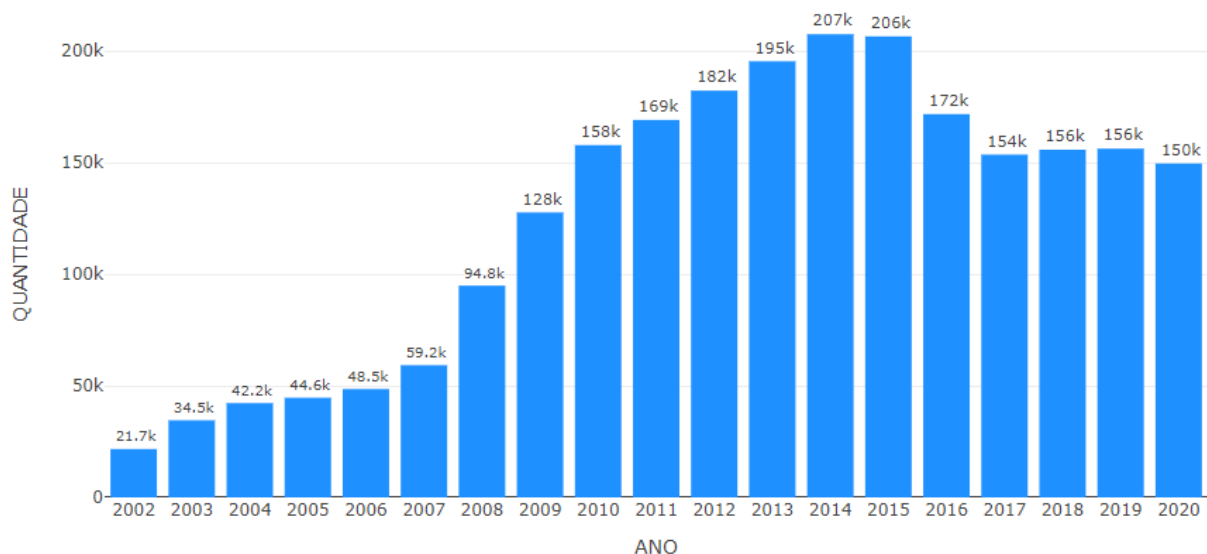
3 ANÁLISE DOS DADOS

3.1 ANÁLISE DE TODAS AS ÁREAS

A análise foi dividida em duas partes, na primeira etapa os dados foram analisados de forma geral, abrangendo todas as áreas existentes no programa CNPq, na segunda etapa foram analisadas somente as áreas de TI.

No primeiro gráfico (Fig.13) dos programas é possível observar que houve um aumento no número de projetos ao decorrer dos anos, chegando a seu pico máximo em 2014, logo após, diminuindo e estabilizando no valor em torno de 150.000 projetos.

Figura 13 – Total de projetos (todas as áreas do conhecimento)



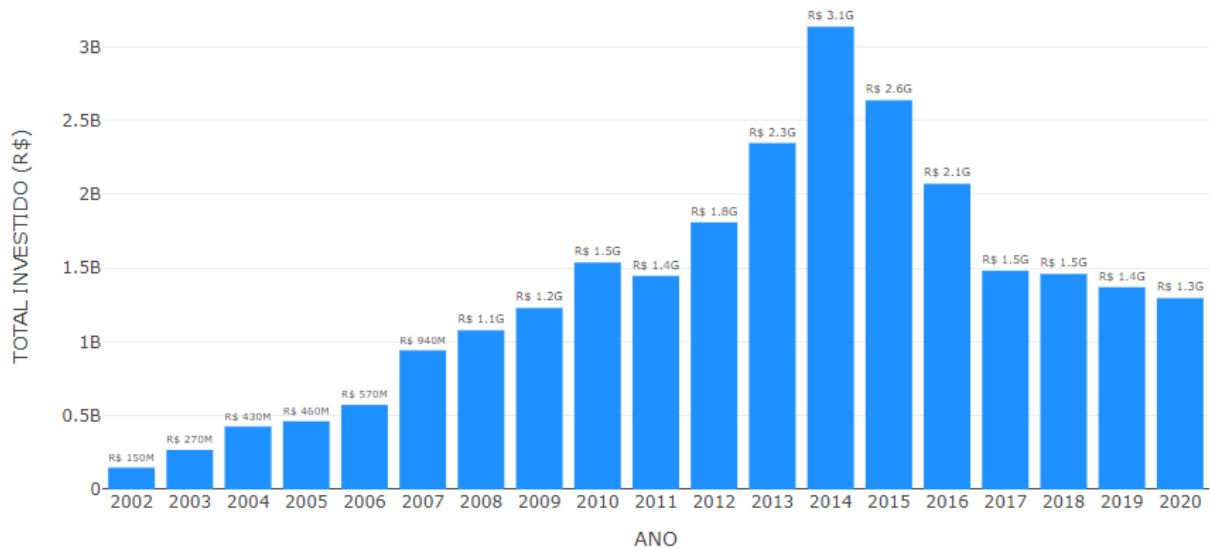
Fonte: Dados do estudo (2021).

Coletando o valor total investido (Fig. 13), coincide com o pico máximo em 2014 (3 Bilhões), porém, vale ressaltar que no valor investido a queda foi de 50% entre os anos de 2017 a 2020. O gráfico acima aponta, porém, que a quantidade diminuiu apenas 25%, o que permite concluir que o valor médio por projeto diminuiu com o passar dos anos.

Tais dados corroboram com o estudo de Moura e Camargo Júnior (2017) que afirmam que o período de 2007 e 2014 foi marcado por elevada disponibilidade de recursos nos diversos programas de financiamento de pesquisas no Brasil, ciclo interrompido abruptamente pela crise econômica de 2015.

Os valores investidos na área no período de 2002 a 2020 são apresentados na Figura 14, a seguir.

Figura 14 – Total de investimentos entre 2002 e 2020



Fonte: Dados do estudo (2021).

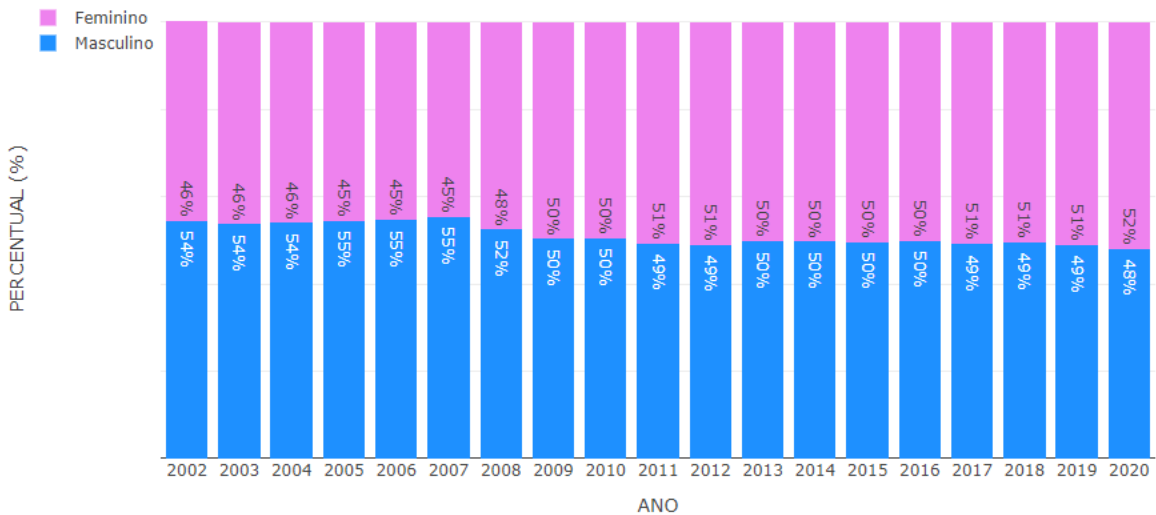
Marques (2017) ressalta que o ano de 2014 apresenta o maior volume de dados a respeito dos investimentos em pesquisa, provavelmente por se tratar do período em que esses investimentos foram os mais elevados. Os investimentos em P&D em 2014 corresponderam a 1,27 do PIB, totalizando R\$ 73,6 bilhões. Entre 2000 e 2013 o investimento em P&D no Brasil, tanto do setor público quanto privado, em milhões de dólares, pode ser descrito da seguinte maneira: o setor público passou de US\$ 8,5 bilhões em 2000 para US\$ 22,9 bilhões em 2013, com uma taxa de crescimento acentuada a partir de 2005, já o setor privado aumentou os investimentos de US\$ 8 bilhões para US\$ 16,8 bilhões.

Tenório, Mello e Viana (2017) afirmam que no Brasil a parcela mais significativa de investimento em P&D decorre da administração pública, enquanto em países em desenvolvimento o financiamento das atividades de pesquisa na área é amplamente realizado pelo setor privado.

Na análise referente ao sexo (Fig.15), vale ressaltar que os dados são apenas uma estimativa baseada no código *name2Gender*, que define o sexo pelo primeiro nome de acordo com o censo de 2010, ou seja, não são dados originários da nossa base de dados, eles foram gerados e não se tem a informação referente ao erro. Nesses dados, conforme Figura 13, mostra que até 2010 os homens lideravam os

projetos, após isso, as mulheres tomaram a frente e permanecem até hoje. Isso retrata nossa realidade, realmente as mulheres vem ganhando espaço no mercado em todas as áreas e em todos os níveis hierárquicos, assumindo até a presidência da república que foi exatamente no ano de 2010, quando a Dilma Rousseff assumiu a presidência da república (PRESSE, 2010).

Figura 15 – Total de projetos por gênero



Fonte: Dados do estudo (2021).

Olinto (2011) afirma que apesar da importância das mulheres para a construção socioeconômica de todas as sociedades no mundo, os investimentos em pesquisa ainda têm uma tendência de beneficiar os homens, de modo que muitas mulheres desistem de buscar a construção de uma carreira nas áreas de ciência e tecnologia, marcadamente no Brasil, pela percepção de que dificuldades consideráveis poderão ocorrer quando da busca por financiamentos para suas atividades e estudos.

Na sequência são apresentados os dados das análises com foco nas áreas de TI.

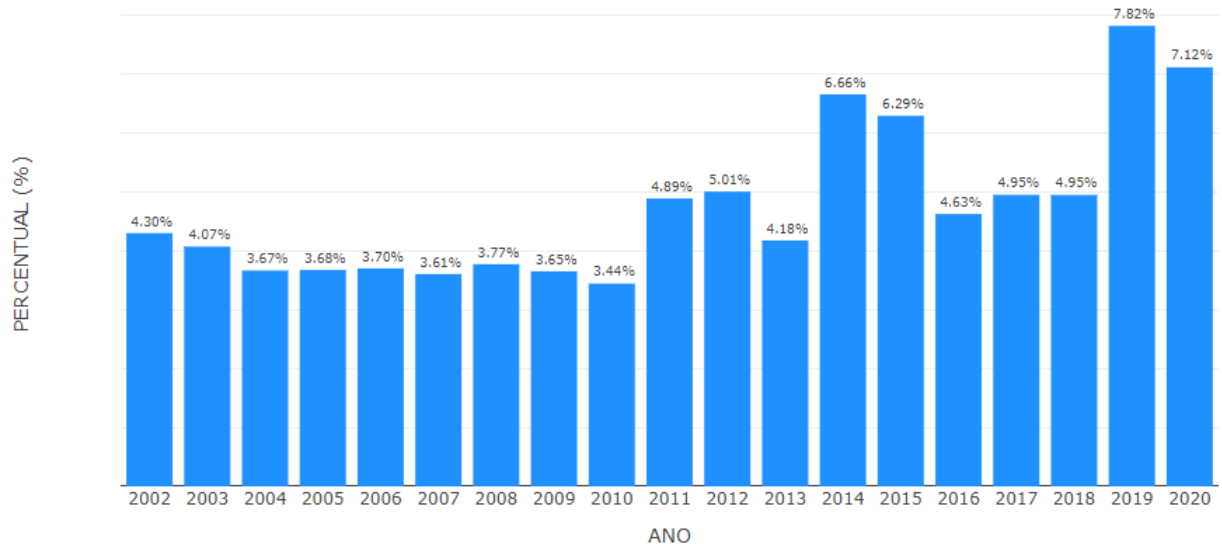
3.2 ANÁLISE DAS ÁREAS DE TI

3.2.1 Geral

São apresentados nesta seção, os gráficos gerados especificamente das áreas de TI selecionadas.

De forma a ser mais fácil a comparação com todas as áreas, foram criados dois gráficos com base nos percentuais da quantidade de projetos (Fig. 16) e do valor investido nas áreas de TI de acordo com todas as áreas (Fig. 16), ou seja, descobrir a fatia do bolo que a TI obteve de todos os benefícios.

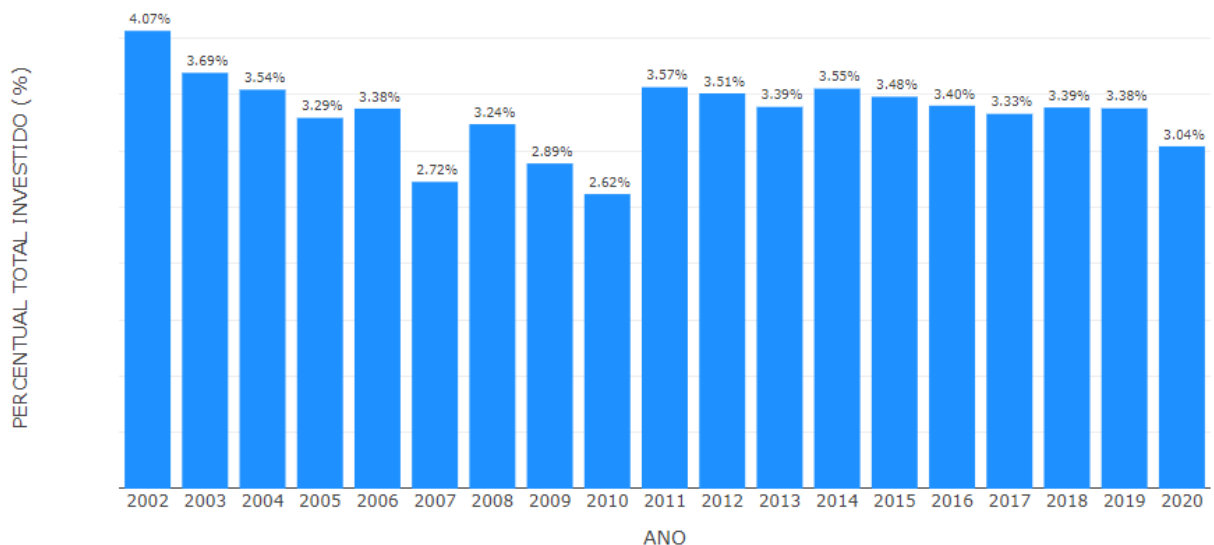
Figura 16 – Total de projetos nas áreas de TI



Fonte: Dados do estudo (2021).

A figura 16 evidência que as áreas de TI são pouco beneficiadas, entre 3% a 8%, sendo que nos últimos 2 anos se obteve as máximas históricas de investimento na áreas de TI, mas o valor investido em comparação às outras áreas, a área de TI se mantém entre 3% a 4%.

Figura 17 – Valor investido nas áreas de TI



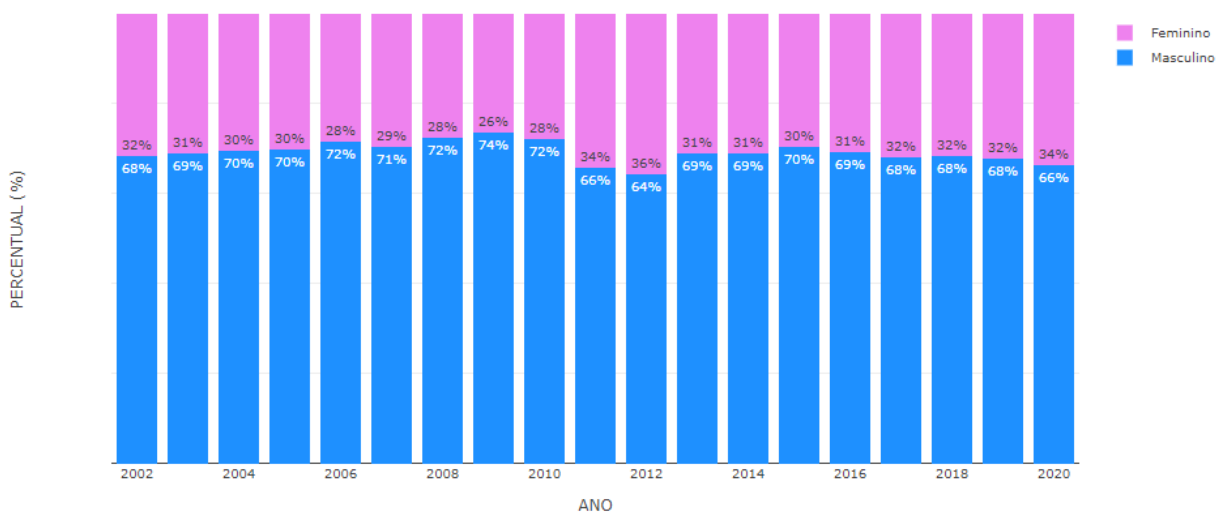
Fonte: Dados do estudo (2021).

Neste ponto é preciso destacar que dos R\$ 73,6 bilhões investidos em pesquisas em 2014, somente R\$ 5,6 bilhões (7,6% do total) foram decorrentes do MCTIC – Ministério de Ciência, Tecnologia, Inovação e Comunicação (MARQUES, 2017).

Para Arbix (2017), o Brasil ainda precisa alterar grandemente seus padrões de investimentos na área de pesquisas tecnológicas, pois somente assim será capaz de competir com países desenvolvidos, nos quais os investimentos nessa área são expressivamente maiores. A pesquisa envolve custos, porém, esses custos são revertidos em retornos financeiros decorrente do desenvolvimento tecnológico. Quando não há investimento em P&D na área das TICs, não existem custos relacionados, porém, o país depende de adquirir toda tecnologia de que necessita, a altos custos, dos países que investem continuamente nos esforços nessa importante área.

Com o código *name2GenderBR* foram extraídos o sexo de acordo com o nome nas áreas de TI (Fig. 18), obtendo uma estimativa de que no ramo da educação nas áreas de TI o sexo masculino ainda assume cerca de 70% dos benefícios no CNPq.

Figura 18 – Percentual de projetos por gênero áreas de TI



Fonte: Dados do estudo (2021).

Rosa e Quirino (2016) afirmam que a questão de haver uma participação maior de homens do que mulheres nas áreas de ciência e tecnologia é cultural e se desenvolveu no perpassar da história. Homens foram e são mais valorizados do que as mulheres em diversas áreas, conseguem ascender a cargos importantes com

mais facilidade, enquanto as mulheres são vistas como mais adequadas para determinadas carreiras nas quais os homens buscam menor participação.

De modo semelhante, Olinto (2011) já havia ressaltado que a escolha da carreira profissional de muitas mulheres baseia-se na percepção de que algumas áreas apresentam mais obstáculos, como ocorre na área de pesquisa científica e, assim, acabam por tomar decisões baseadas no receio das dificuldades e na percepção de que serão mais aceitas, respeitadas e valorizadas em outros setores sociais.

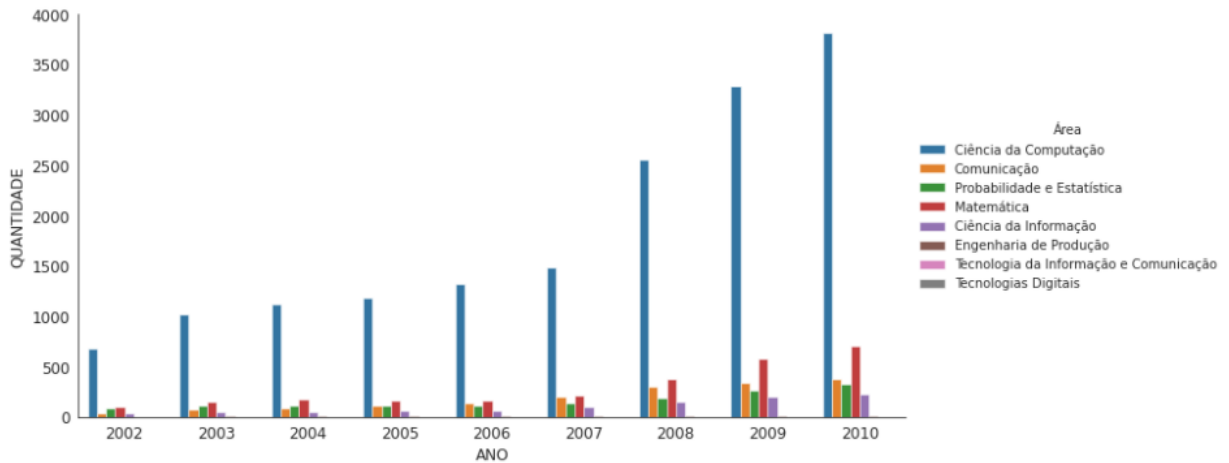
3.2.2 Áreas

A análise de forma mais detalhada será iniciada pela análise das áreas de acordo com as subáreas de TI selecionadas anteriormente, sendo as áreas:

- Ciência da Computação
- Matemática
- Comunicação
- Probabilidade e Estatística
- Ciência da Informação
- Tecnologia da Informação e Comunicação
- Engenharia de Produção
- Tecnologias Digitais

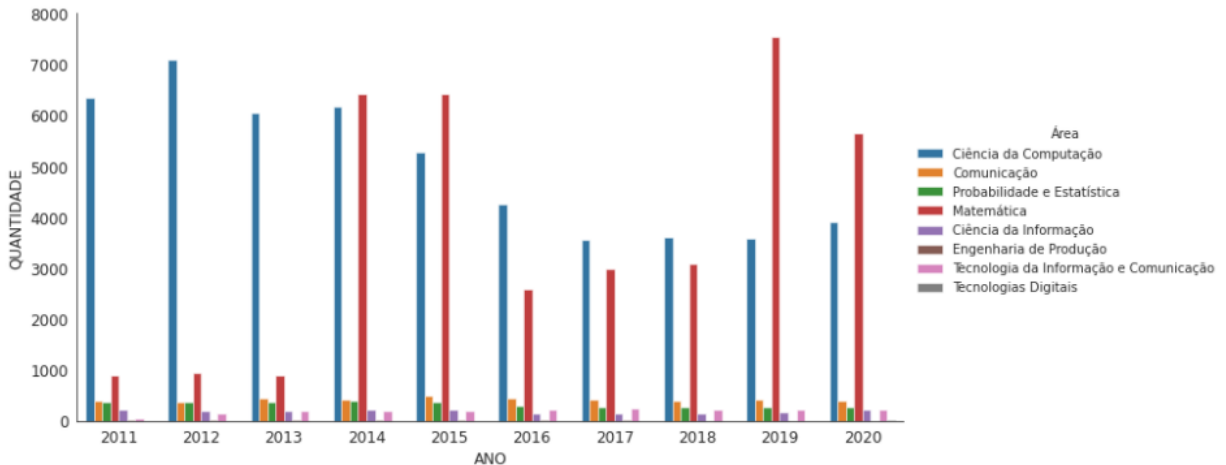
Analisando a evolução no decorrer dos anos, divididos em dois gráficos para melhor visualização dos dados (Fig. 19 e Fig. 20), podemos analisar que a Ciência da Computação sempre foi a área macro com mais quantidade de projetos até o ano de 2014, onde a área de Matemática teve um salto enorme ultrapassando a Ciências da Computação sendo assim as duas áreas com mais projetos de 2014 em diante.

Figura 19 – Quantidade de projetos em TI por área – 2002 a 2010



Fonte: Dados do estudo (2021).

Figura 20 – Quantidade de projetos em TI por área - 2011 a 2020

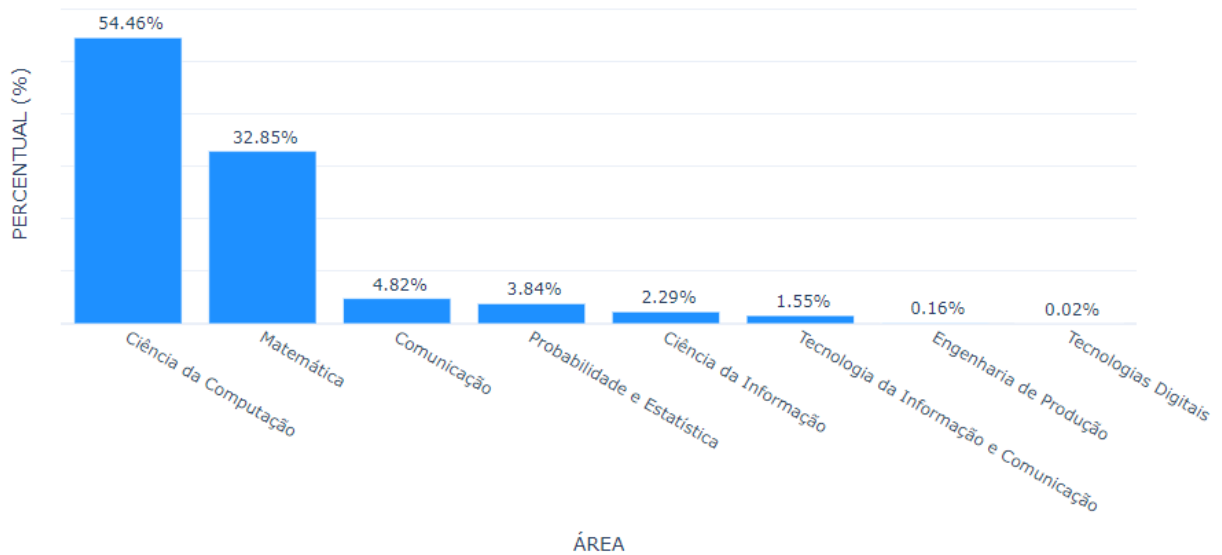


Fonte: Dados do estudo (2021).

Todas as áreas de TI são importantes para o desenvolvimento de diferentes indicadores e setores socioeconômicos. As pesquisas na área deveriam, de fato, manter-se em elevados números em função das contribuições que os projetos na área podem trazer para toda a sociedade (ALMEIDA; RODAS; MARQUES, 2020).

Com a soma do total de projetos na área de TI divididos pela quantidade total de projetos em cada área (Fig. 21), é possível confirmar que a área de Ciência da Computação tem mais da metade dos projetos recebidos, seguida por matemática.

Figura 21 – Percentual da quantidade de projetos por área de TI

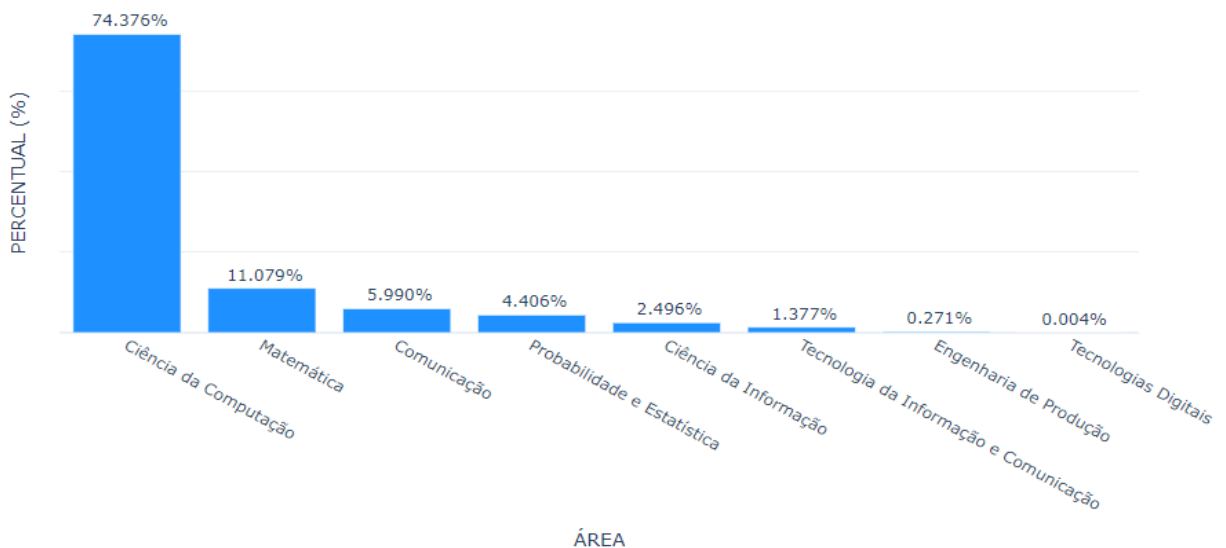


Fonte: Dados do estudo (2021).

Qualquer país que almeje ser competitivo e destacar-se no mercado, seja nacional ou internacional, deve ter acesso aos avanços tecnológicos e ser capaz de produzir as tecnologias que permeiam o mercado e que conquistam a preferência dos consumidores. Baixos investimentos nas pesquisas na área conduzem à dependência pelos resultados de outras nações (ALMEIDA; RODAS; MARQUES, 2020).

O percentual do valor investido em Ciências da Computação totaliza 74% do total investido em todas as áreas de TI, mostrando que é a área predominante (Fig. 22).

Figura 22 – Percentual do valor total investido por áreas de TI

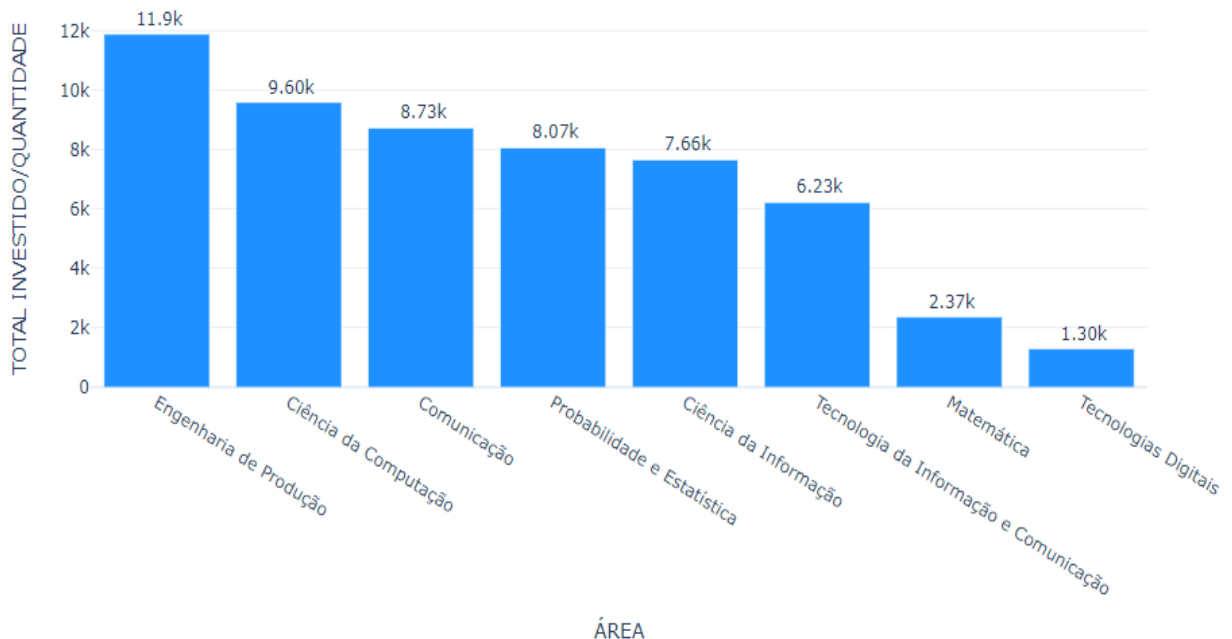


Fonte: Dados do estudo (2021).

Levando-se em consideração o valor total dividido pela quantidade é possível obter o valor único, ou seja, um valor por projeto. Com esses dados alcança-se uma informação importante, mostra que na Engenharia de produção foi a área com o valor de projeto mais caro (Fig. 23).

As políticas de fomento aos investimentos na pesquisa dentro das áreas e subáreas da TI no país ainda são insuficientes, apesar da percepção de sua importância. É preciso encontrar formas de elevar os investimentos públicos e privados em pesquisas que podem auxiliar na melhoria de todas as áreas (ALMEIDA; RODAS; MARQUES, 2020).

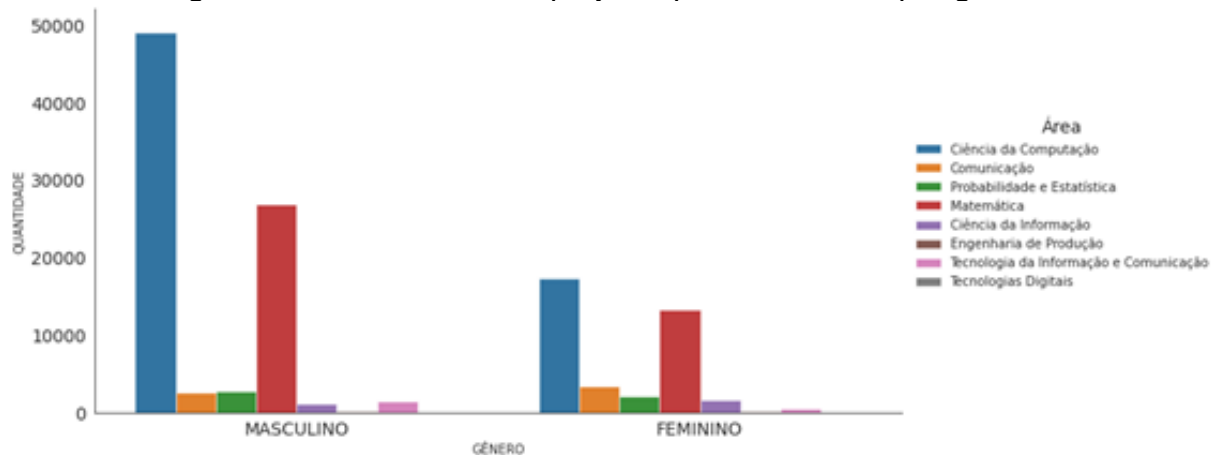
Figura 23 – Valor único por área da TI



Fonte: Dados do estudo (2021).

Separando por gênero de uma forma geral (Fig. 24), somente na área de Comunicação e Ciência da Informação as mulheres possuem mais projetos que os homens.

Figura 24 – Quantidade de projetos por áreas da TI por gênero



Fonte: Dados do estudo (2021).

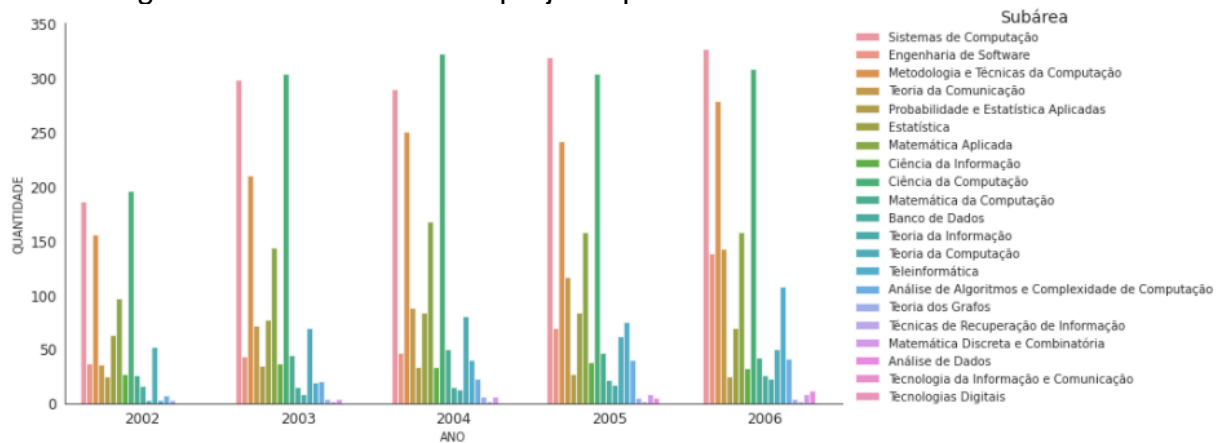
De acordo com Oliveira, Melo e Rigolin (2020), teoricamente a mulher já tem a garantia de igualdade de direitos, porém, na prática ainda se percebe que a segregação é acentuada. Na área de tecnologia, ainda persiste uma visão histórica de que as máquinas e projetos em áreas tecnológicas são melhores conduzidos por homens, enquanto a visibilidade das mulheres nesse setor ainda é bastante limitada. As tecnologias contribuíram e seguem dando importante suporte para a causa da igualdade entre gêneros, no entanto, a participação da mulher na construção dessas tecnologias ainda é limitada.

Na sequência são destacados os dados por subáreas selecionadas para este estudo.

3.2.3 Subáreas

Nas subáreas de TI que foram selecionadas, no decorrer dos anos foram separados em quatro gráficos para facilitar a visualização. No primeiro gráfico que vai de 2002 a 2006 (Fig. 25), é possível observar que Sistemas de Computação, Metodologias e Técnicas da Computação e Ciências da Computação são as subáreas com mais projetos.

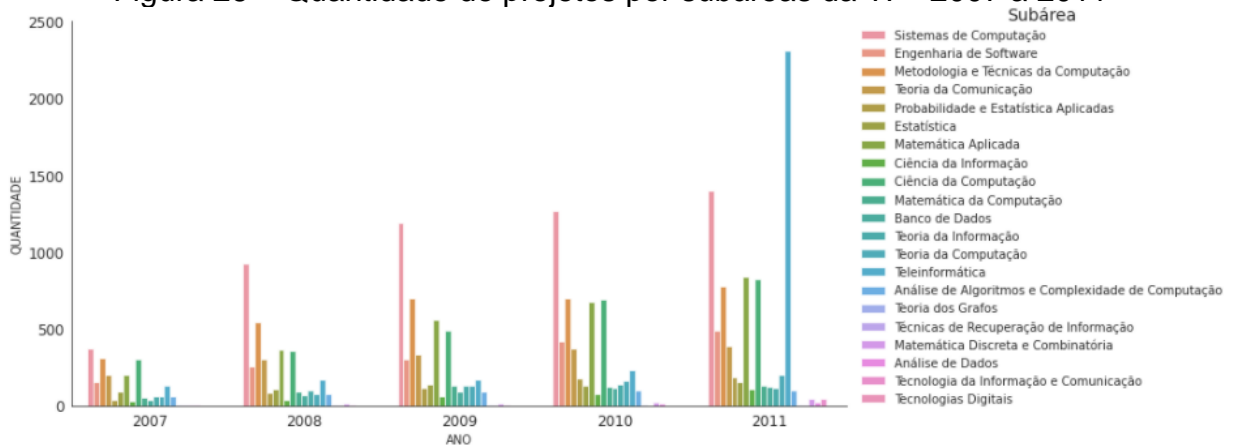
Figura 25 – Quantidade de projetos por subáreas da TI – 2002 a 2006



Fonte: Dados do estudo (2021).

Após 2007 (Fig. 26), apenas a subárea Sistema da Computação permaneceu com as quantidades altas de projetos e em 2011 e 2012 a subárea de Teleinformática deu um salto considerável na quantidade de projetos.

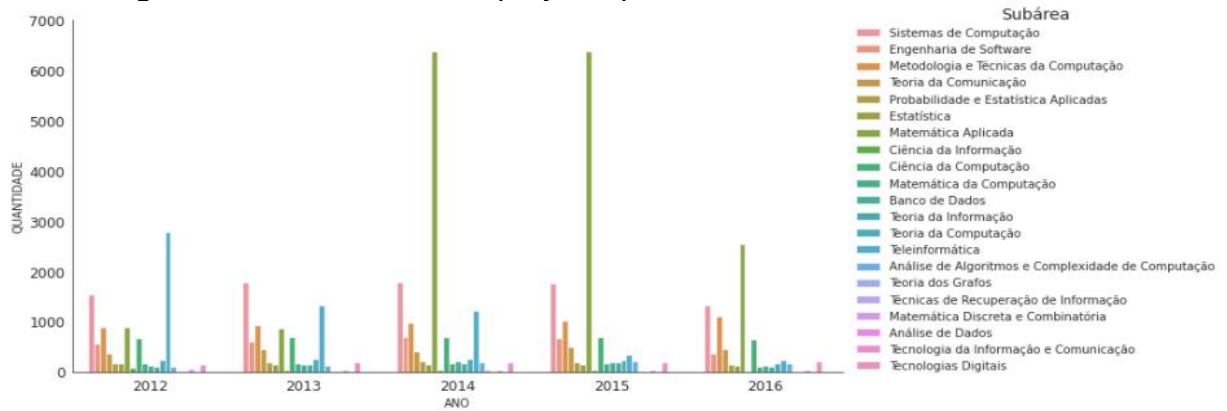
Figura 26 – Quantidade de projetos por subáreas da TI – 2007 a 2011



Fonte: Dados do estudo (2021).

Em 2014 (Fig. 27), a Matemática Aplicada ganhou força e obteve um salto para 2500 projetos.

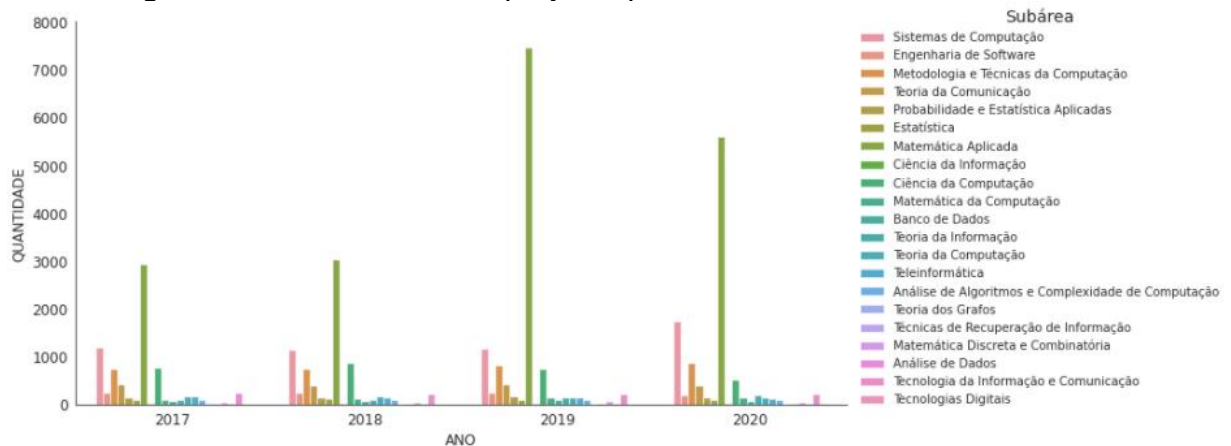
Figura 27 – Quantidade de projetos por subáreas da TI – 2012 a 2016



Fonte: Dados do estudo (2021).

No gráfico referente ao ano de 2017 a 2020 (Fig. 28), mostra que a Matemática Aplicada permaneceu com a quantidade de projetos elevada e se manteve a área de TI com mais projetos.

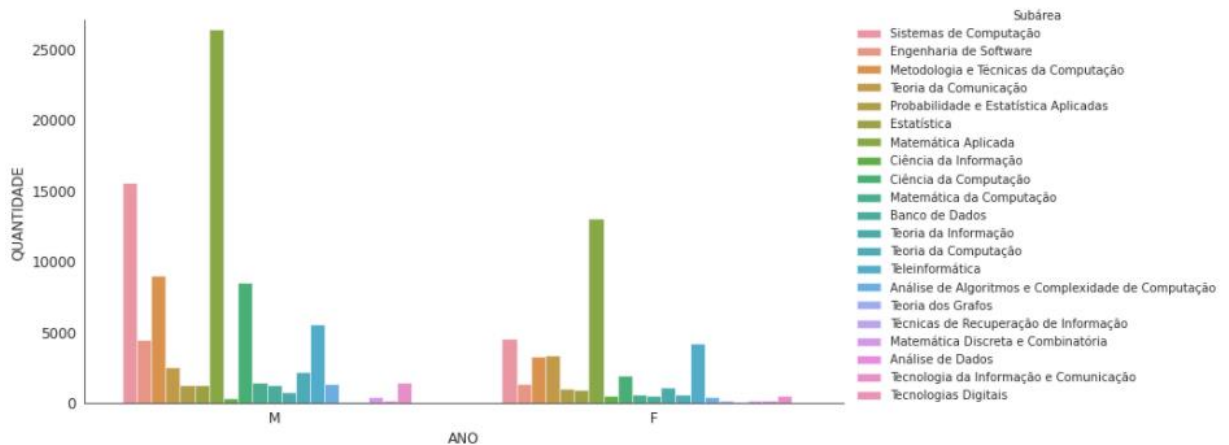
Figura 28 – Quantidade de projetos por subáreas da TI – 2017 a 2020



Fonte: Dados do estudo (2021).

No gênero dessas áreas de TI (Fig. 29), mostram que a Teoria da Comunicação e Probabilidade e Estatística Aplicada as mulheres conseguiram superar a quantidade de homens, mas no restante os homens seguem liderando.

Figura 29 – Quantidade de projetos por subáreas da TI por gênero

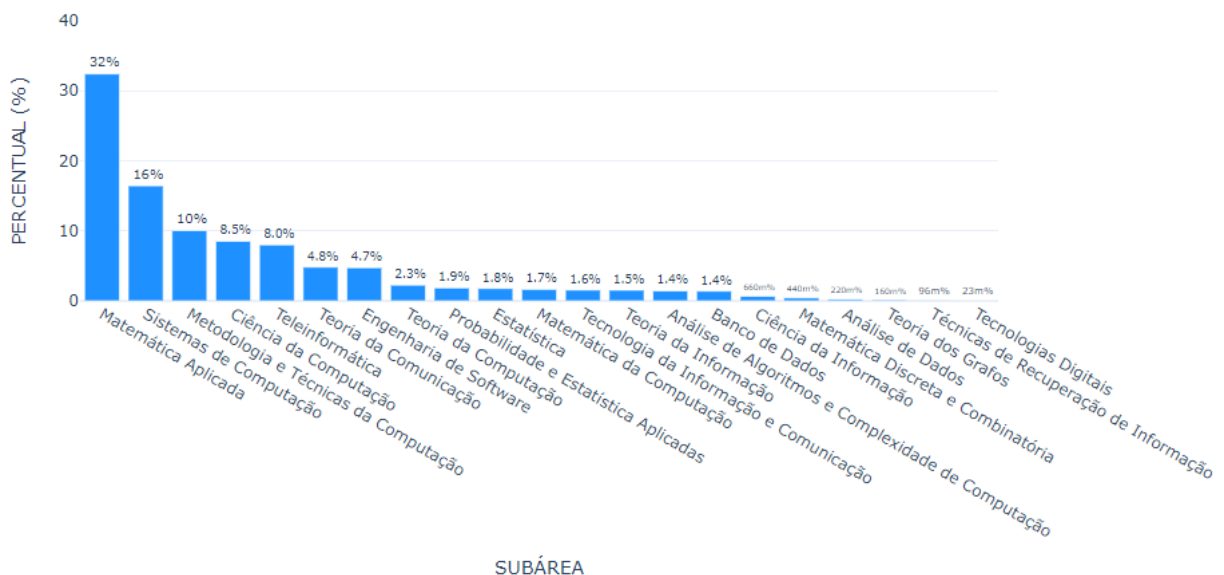


Fonte: Dados do estudo (2021).

Neste ponto é importante destacar que assim como ocorre em inúmeras outras áreas dentro da pesquisa científica, os projetos entre o gênero feminino foram menores, fato que tem relação com o receio das mulheres em investirem seu tempo em desenvolvimento de projetos na área e acabarem por não serem aceitos em face de seu gênero (OLINTO, 2011; OLIVEIRA; MELLO; RIGOLIN, 2020).

No gráfico do percentual da quantidade de cada subárea (Fig. 30), podemos perceber que a Matemática Aplicada tem 32% de projetos, sendo seguido pelo Sistema da Informação.

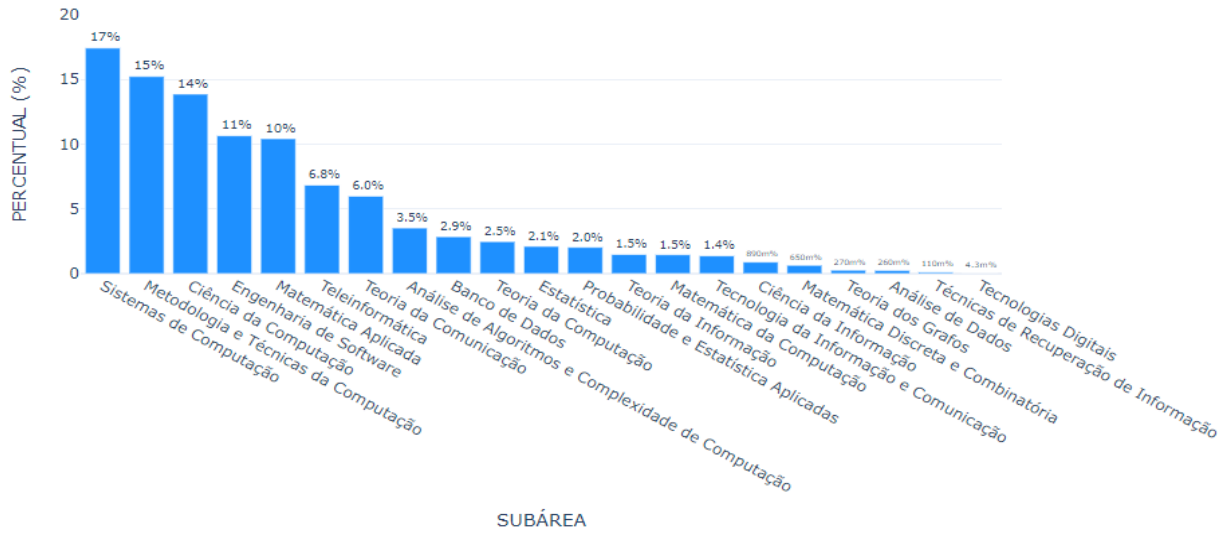
Figura 30 – Quantidade de projetos por subáreas da TI – 2002 a 2020



Fonte: Dados do estudo (2021).

Mas em questão de valor (Fig. 31), Sistemas da Computação e Metodologia e Técnicas da Computação, tem a maior porcentagem de valor investido.

Figura 31 – Percentual de valor investido por subáreas da TI – 2002 a 2020



Fonte: Dados do estudo (2021).

Avaliando o valor total investido na subárea dividido pela quantidade de projetos (Fig. 32) alcança-se o valor médio pago por projeto. Análise de Algoritmos e Complexidade de Programação é a subárea com o valor mais alto por projeto, seguido por Engenharia de Software e Banco de Dados.

Figura 32 – Valor único por subáreas da TI – 2002 a 2020



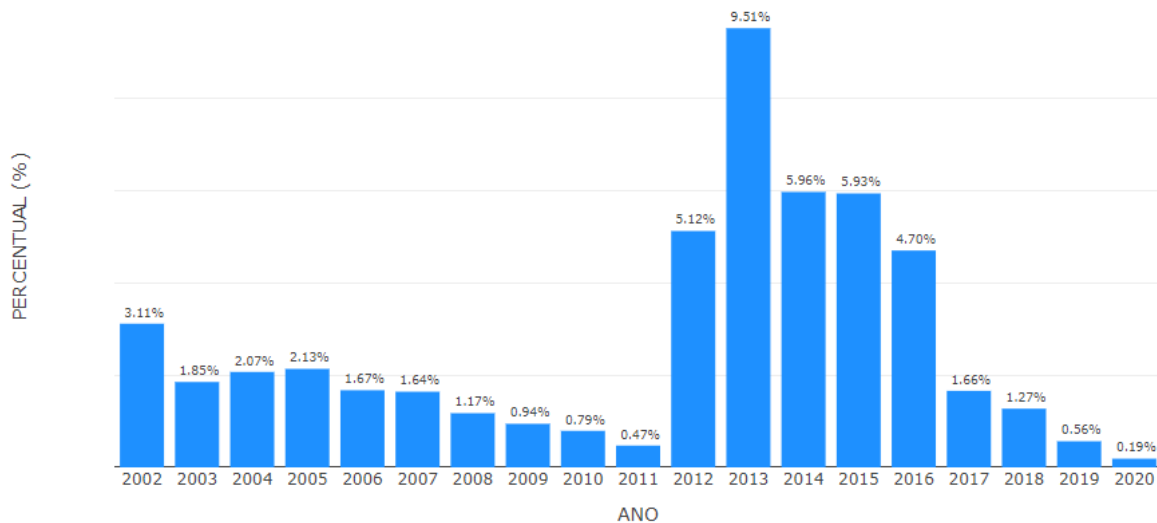
Fonte: Dados do estudo (2021).

3.3 ANÁLISE DAS ÁREAS DE TI NO EXTERIOR

Na análise exploratória dos dados do CNPq, observou-se que existem bolsas que não são do Brasil, então, foi gerado um *dataframe* com os dados de bolsas foram para o exterior.

Nesse gráfico (Fig. 33) mostra que na média geral a quantidade gira em torno de 0,2% a 3%, exceto, pelo grande estouro entre os anos de 2012 a 2017. Analisando somente esses anos foi descoberto que isso se dá ao projeto do governo Ciências sem Fronteiras, que será discutido mais detalhado em outra seção.

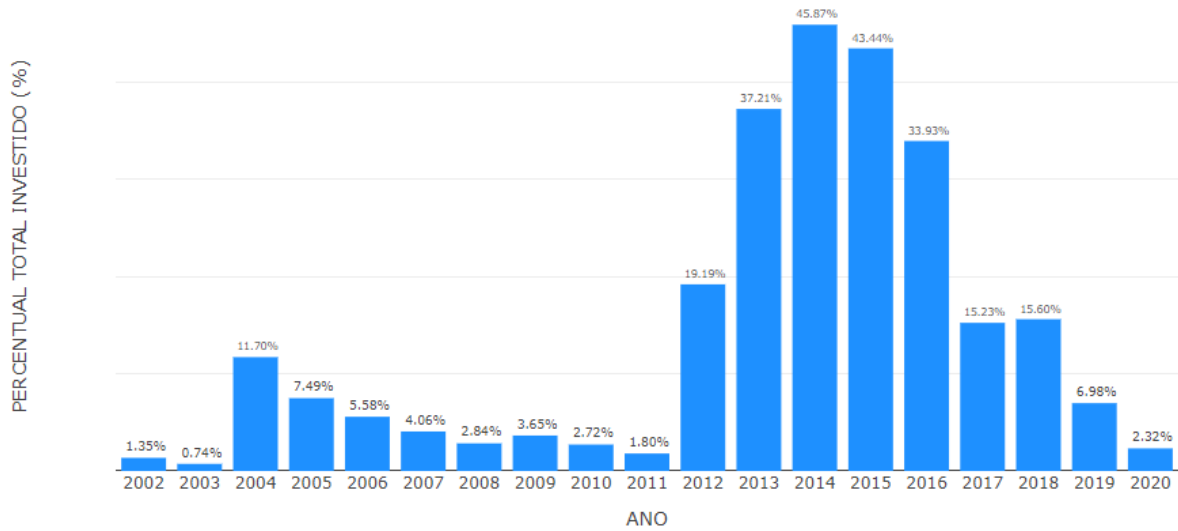
Figura 33 – Quantidade de projetos por subáreas da TI no exterior – 2002 a 2020



Fonte: Dados do estudo (2021).

Já sobre o valor investido (Fig. 34), o aumento foi absurdamente alto, chegando quase a 50%, enquanto em quantidade de projetos subiu apenas 10%, mostra que os valores foram altíssimos para que o programa fosse executado, com razão, já que as moedas Dólar e Euro tem poder maior sobre o Real.

Figura 34 – Percentual do valor investido por subáreas da TI no exterior– 2002 a 2020



Fonte: Dados do estudo (2021).

Manços e Coelho (2017) ressaltam que é importante que o país desenvolva políticas na área de Ciência, Tecnologia e Inovação (CT&I), além de fomentar a internacionalização, fatores de grande importância para um país emergente como no caso do Brasil e no desenvolvimento de inovações tecnológicas que possam tornar suas atividades mais competitivas.

3.3.1 Ciências sem fronteiras

Ciência sem Fronteiras é um programa que busca promover a consolidação, expansão e internacionalização da ciência e tecnologia, da inovação e da competitividade brasileira por meio do intercâmbio e da mobilidade internacional. A iniciativa é fruto de esforço conjunto dos Ministérios da Ciência, Tecnologia e Inovação (MCTI) e do Ministério da Educação (MEC), por meio de suas respectivas instituições de fomento – CNPq e Capes –, e Secretarias de Ensino Superior e de Ensino Tecnológico do MEC. O programa iniciou-se em 2011, tendo como uma de suas metas o envio de 75.000 bolsistas brasileiros até o final do ano de 2014 (CAPES; CNPQ, 2011, p. 1; MANÇOS; COELHO, 2017).

Para facilitar a visualização dos dados foram agrupados os programas CNPq que pertenciam ao Ciências sem Fronteiras em um único programa nomeado PROGRAMA CIÊNCIAS SEM FRONTEIRAS. Segue os programas que foram agrupados:

- Programa Ciência sem Fronteiras
- Programa CSF - Ciências sem Fronteiras (PADCT)
- Programa Fluxo Contínuo de Bolsas - CsF - (COICF)
- Programa CSF - Programa Ciência sem Fronteiras (CONCF)
- Programa CSF - Programa Ciência sem Fronteiras (COICF)
- Programa CSF - Ciências sem Fronteiras (COENE)
- Programa CSF - Biologia e Ciências Biomédicas
- Programa CSF - Tecnologia Aeroespacial)

O aumento das bolsas no exterior foi devido ao programa Ciências sem Fronteiras, conforme mostra o gráfico abaixo (Fig. 35).

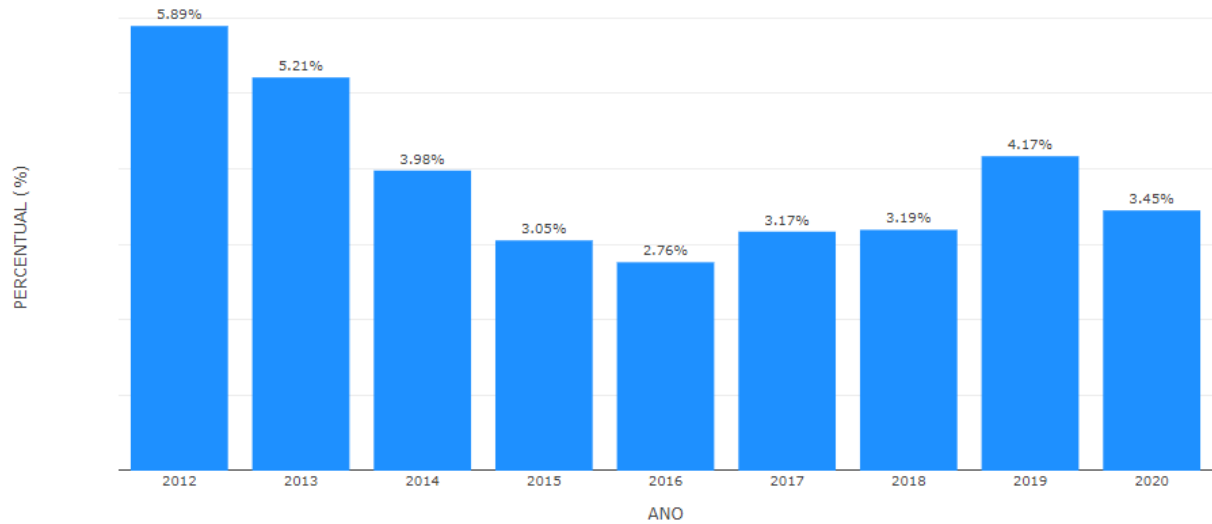
Figura 35 – Quantidade de projetos no exterior por áreas da TI - 2012 a 2017



Fonte: Dados do estudo (2021).

Mas as áreas de TI no programa representaram uma pequena fatia dos projetos e bolsas, algo em torno entre 2% e 6% (Fig. 36).

Figura 36 – Percentual de projetos no exterior no programa Ciências sem Fronteiras por subáreas da TI– 2012 a 2020



Fonte: Dados do estudo (2021).

Dados gerais do Programa ciências sem Fronteiras extraídos do conjunto de dados:

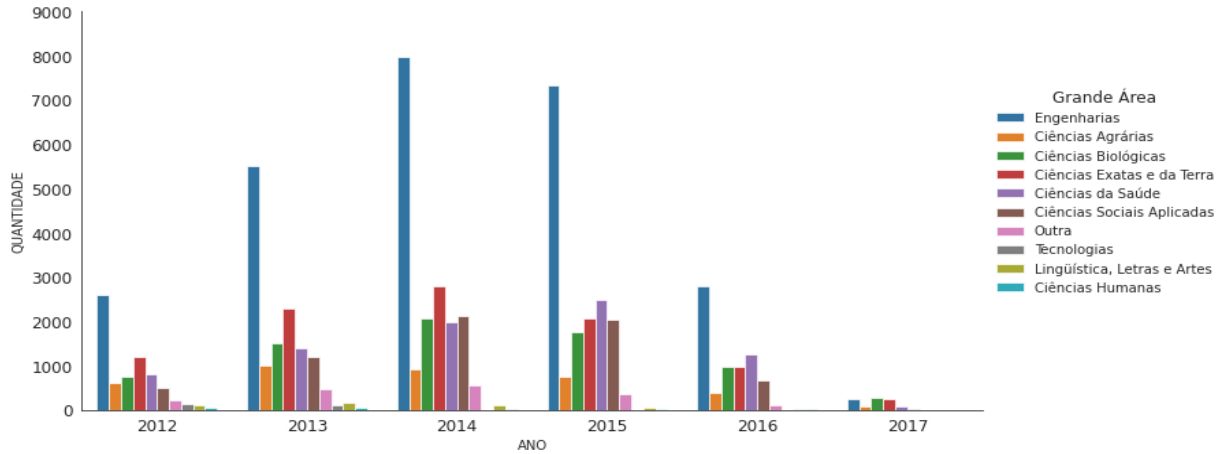
- **Quantidade de projetos**
 - Todas as áreas: 64.799
 - Áreas de TI: 2.625
- **Valor Pago Total**
 - Todas as áreas: R\$ 3.091.114.586
 - Áreas de TI: R\$ 122.810.171

De acordo com esse Mazzeto (2017), “das 92.880 bolsas concedidas entre 2011 e 2016, 79% (73.353) foram destinadas aos alunos de graduação. Ao todo, os gastos com o programa foram entre R\$ 12 a 15 bilhões nesse período.”

Comparando os dados observamos que em Total de Projetos há uma diferença de quase 9 mil projetos e em valor pago há uma diferença entre R\$ 9 e 12 bilhões.

Explorando mais detalhadamente o programa (Fig. 37), mostrou que as Grandes Áreas que mais receberam o auxílio foram as Engenharias, seguido por Ciências Exatas e da Terra.

Figura 37 – Quantidade de projetos no exterior no programa Ciências sem Fronteiras das grandes áreas da TI – 2012 a 2020

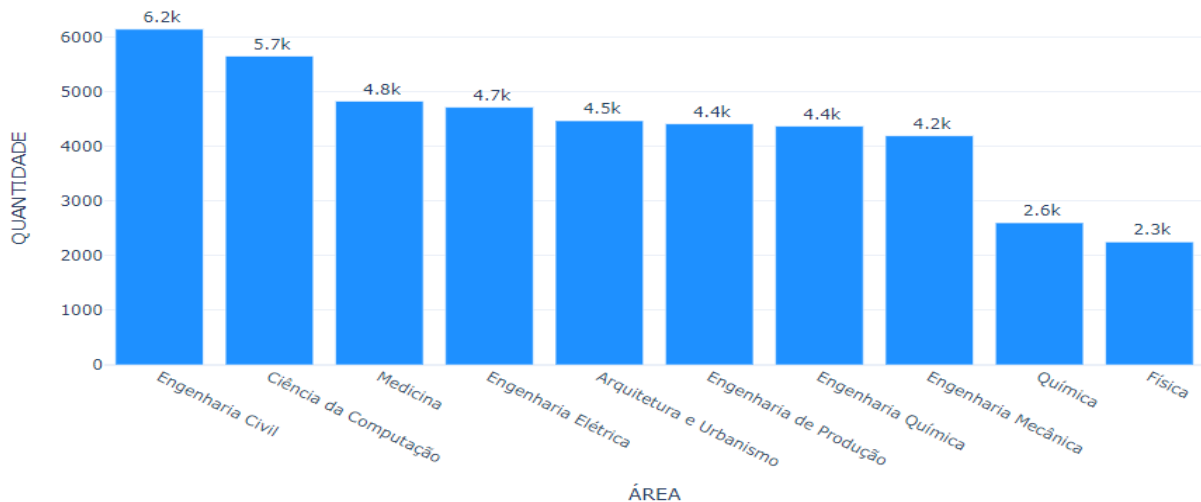


Fonte: Dados do estudo (2021).

As subáreas de TI se encontram em quatro dessas Grandes Áreas, segue em porcentagem a quantidade de projetos de acordo com cada uma das Grandes Áreas: Ciências exatas e da Terra (96,99%), Tecnologias (1,71%), Ciências sociais aplicadas (0,99%) e Engenharias (0,30%).

Nas áreas (Fig. 38) é possível verificar que Engenharia Civil é a que possui maior quantidade de projetos, seguido por Ciências da Computação e Medicina.

Figura 38 – Quantidade de projetos no exterior no programa Ciências sem Fronteiras top 10 áreas da TI – 2012 a 2017

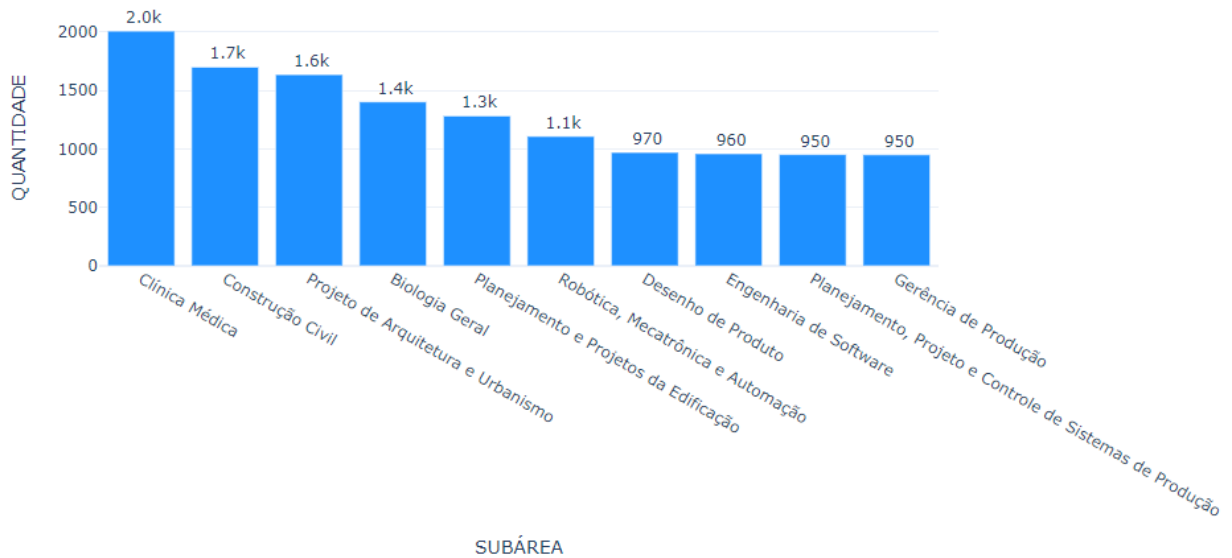


Fonte: Dados do estudo (2021).

Das áreas acima, nossas subáreas de TI estão presentes em seis, segue abaixo o percentual da quantidade de projetos por Área: Ciência da computação (88,11%), Matemática (4,57%), Probabilidade e estatística (4,30%), TIC (1,71%), Comunicação (0,99%) e Engenharia de produção (0,30%).

Já nas subáreas (Fig. 39) a Clínica Médica lidera, seguido pela Construção Civil e Projeto de Arquitetura e Urbanismo. Nesse TOP 10, a única posição das subáreas de TI está na Engenharia de Software.

Figura 39 – Quantidade total de projetos no exterior no programa Ciências sem Fronteiras top 10 áreas da TI– 2012 a 2017



Fonte: Dados do estudo (2021).

Para entender nossas subáreas de TI em percentual, segue a relação delas em percentual: Engenharia de software (28,57%), Metodologia e técnicas da computação (21,52%), Sistemas de computação (20,42%), Banco de dados (5,14%), Matemática aplicada (4,19%), Análise de algoritmos e complexidade de computação (3,81%), Teleinformática (2,70%), Ciência da computação (2,59%), Teoria da computação (2,51%), Estatística (2,32%), Probabilidade e estatística aplicadas (1,90%), Tecnologia da informação e comunicação (1,71%), Teoria da comunicação (0,99%), Matemática da computação (0,84%), Matemática discreta e combinatória (0,38%), Teoria dos grafos (0,30%), Análise de dados (0,08%).

No conjunto de dados é possível analisar as Modalidades, para facilitar a visualização foram agrupadas conforme:

PÓS GRADUAÇÃO: GDE - Doutorado no Exterior, SWE - Doutorado Sanduíche no Exterior, PDE - Pós-doutorado no Exterior, AVG - Participação em Eventos Científicos, ESN - Estágio Sênior no Exterior, PQ - Produtividade em Pesquisa, DEJ - Desenvolvimento Tecnológico e Inovação Ext Junior, SPE -

Estagio/Especializacao no Exterior, DES - Desenvolvimento Tecnológico e Inovação Ext Sênior.

GRADUAÇÃO: SWG - Graduação Sanduíche no Exterior, IC - Iniciação Científica.

TREINAMENTO NO EXTERIOR: BSP - Bolsa Treinamento no Exterior, APE - Aperfeiçoamento no Exterior, DTI - Desenvolvimento Tecnológico Industrial.

AUXILIO PESQUISA: APQ - Auxílio a Pesquisa, PV - Pesquisador Visitante, BEV - Bolsa Especialista Visitante, AED - Auxílio a Editoração e Publicação, ARC - Promoção de Eventos Científicos.

MESTRADO: MPE - Mestrado Profissional no Exterior, GM – Mestrado.

DOCTORADO: GD – Doutorado, SWP - Doutorado-Sanduiche no País.

PÓS DOCTORADO: PD - Pós-doutorado, PDJ - Pós-doutorado Júnior.

ESTÁGIO: EJR - Estágio Júnior no Exterior

Analisando as modalidades de todas as áreas (Fig. 40), encontramos que a graduação segue como a mais requisitada.

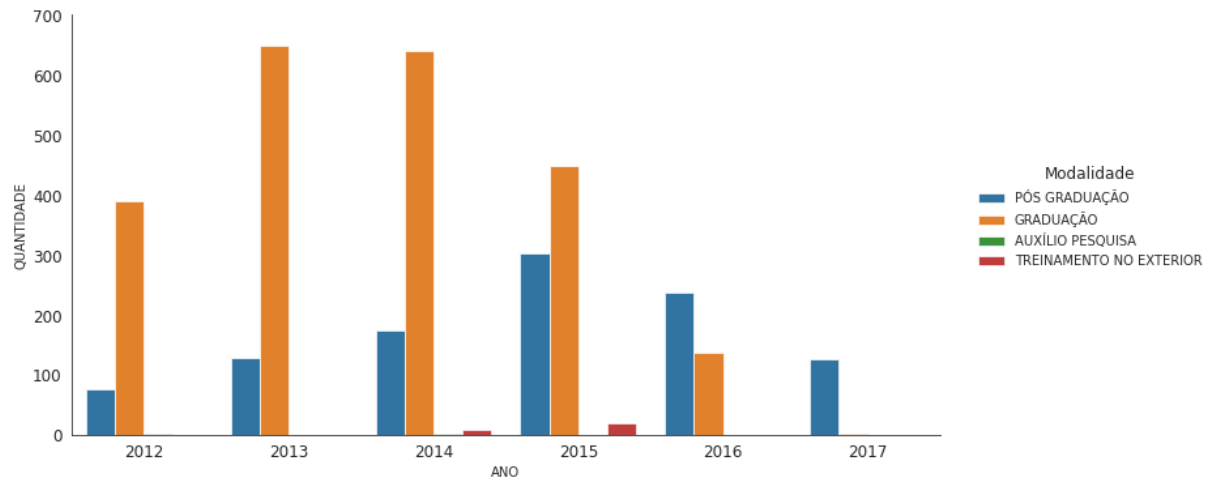
Figura 40 – Quantidade total de projetos no exterior no programa Ciências sem Fronteiras todas as áreas de acordo com a modalidade – 2012 a 2017



Fonte: Dados do estudo (2021).

E nas nossas subáreas de TI (Fig. 41), se mantém o mesmo padrão, sendo Graduação e pós-graduação.

Figura 41 – Quantidade total de projetos no exterior no programa Ciências sem Fronteira áreas de TI de acordo com a modalidade – 2012 a 2017



Fonte: Dados do estudo (2021).

Manços e Coelho (2017) ressaltam que apesar da importância do programa Ciências sem Fronteiras, sua produção de conhecimentos e colaboração científica internacional ainda não foram exploradas de forma detalhada e, assim, há um conhecimento limitado que pode gerar uma desvalorização do programa, não por seus gastos ou limitação de contribuição, mas pela falta de dados detalhados, claros e confiáveis sobre seu papel no cenário nacional e internacional da pesquisa científica.

3.4 MAPAS DE CALOR

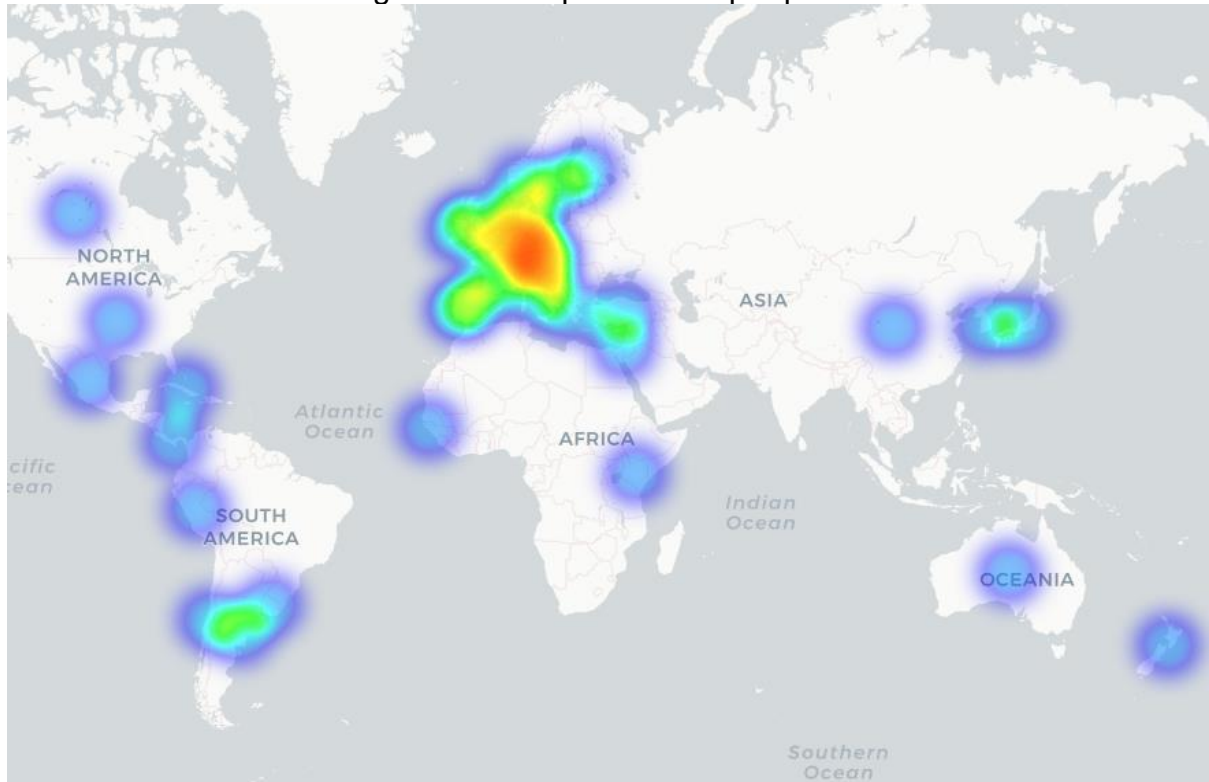
Os *heatmaps* ou, em português, mapas de calor são representações gráficas que indicam por meio de uma escala de cores que, geralmente, vai do vermelho (mais quente) ao azul (mais frio), onde há mais concentração ou não de dados. Nesse trabalho os dados são as quantidades de projetos, maior quantidade representa cores mais quentes e menor quantidade cores mais frias, os mapas são baseados nas áreas de TI selecionadas.

3.4.1 Áreas de TI no Exterior – por países

Esse é um mapa de calor da quantidade de projetos das áreas de TI de acordo com os países no exterior, ele tem as latitudes e longitudes dos Países e é

definido de acordo com a quantidade de projetos e bolsas em cada País. Esse mapa de calor mostra que a maior concentração está na Europa (Fig. 42).

Figura 42 – Mapa de calor por países

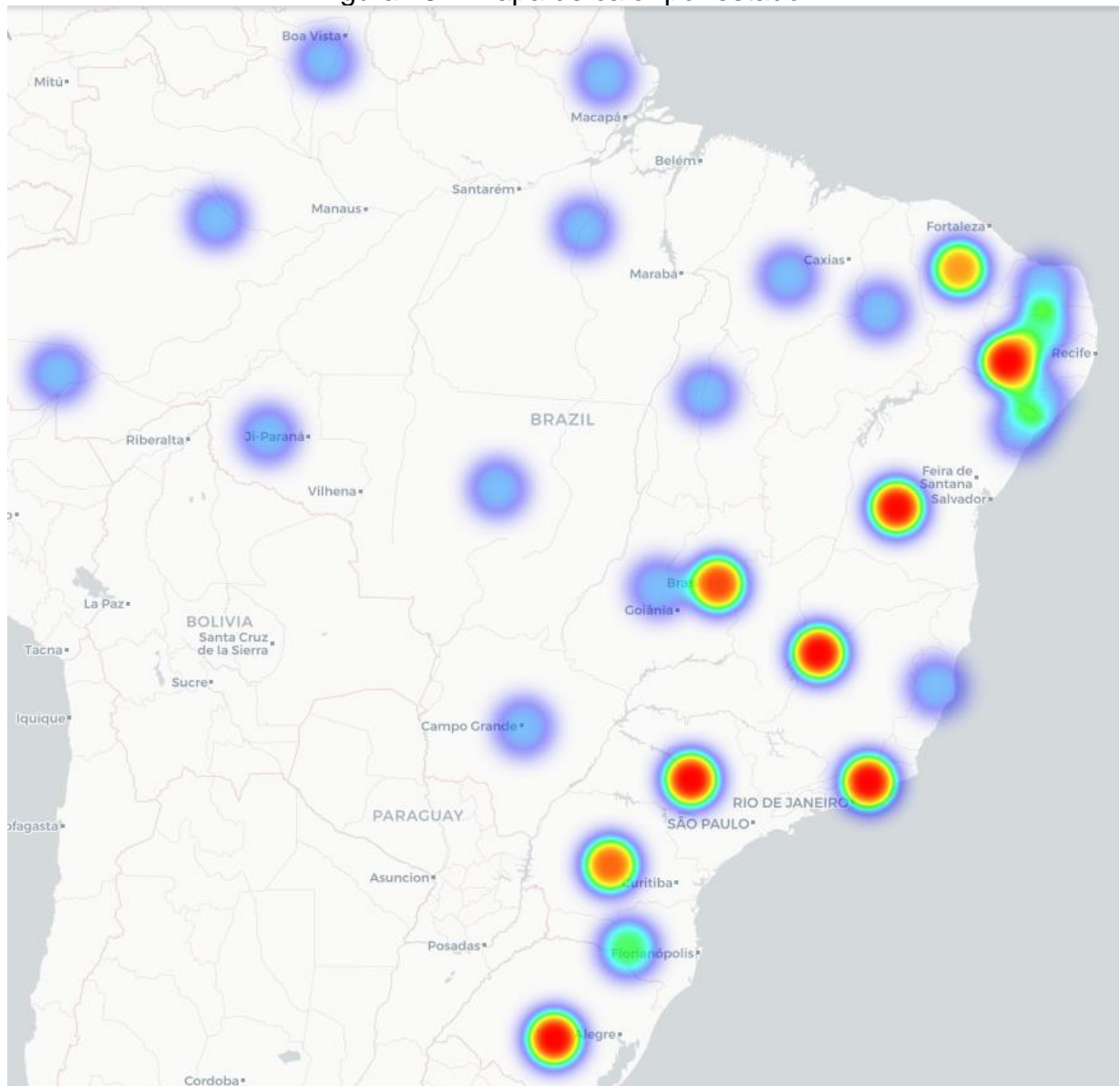


Fonte: Dados do estudo (2021).

3.4.2 Áreas de TI no Brasil – por estados

Esse é um mapa de calor da quantidade de projetos das áreas de TI no Brasil de acordo com os estados, ele tem as latitudes e longitudes dos estados do Brasil e é definido de acordo com a quantidade de projetos e bolsas em cada estado. Através desse mapa é possível analisar que a região do litoral possui a maior concentração de bolsas (Figura 43).

Figura 43 – Mapa de calor por estado

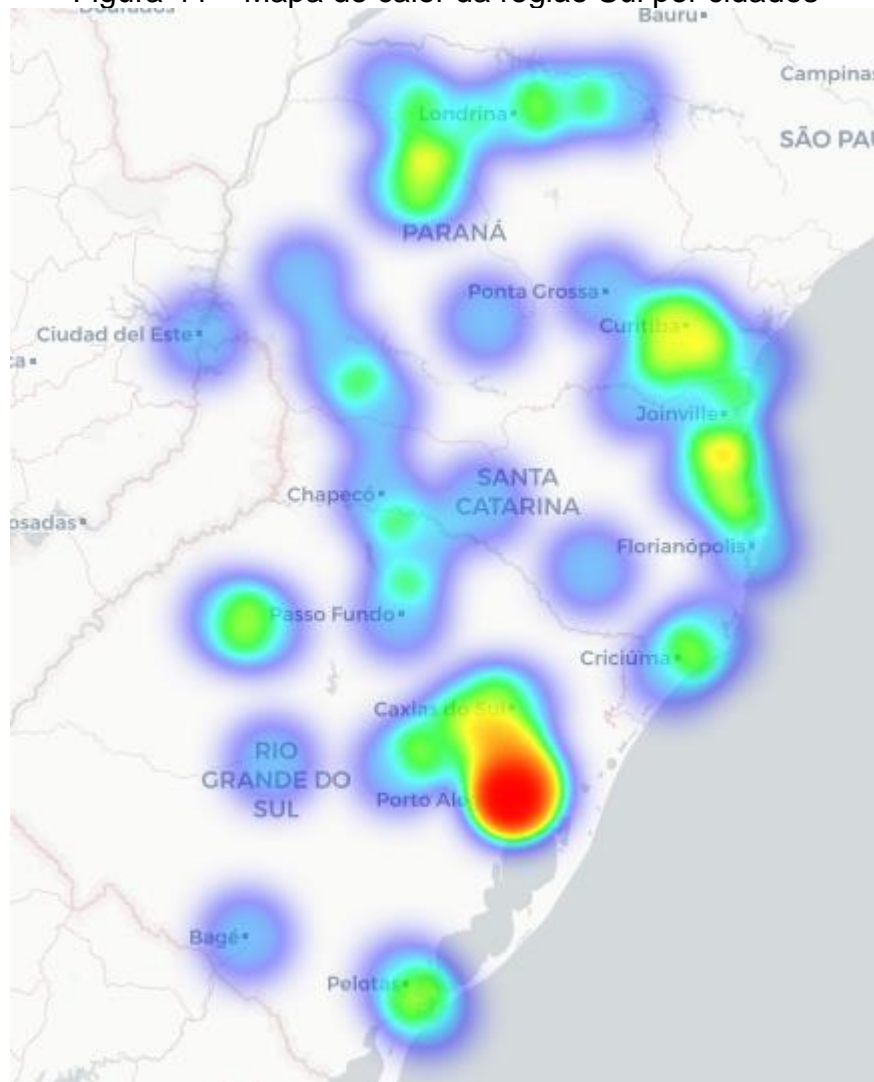


Fonte: Dados do estudo (2021).

3.4.3 Áreas de TI na Região Sul (RS|SC|PR) – por cidades

Esse é um mapa de calor da quantidade de projetos das áreas de TI de acordo com a região sul do Brasil, nos estados do Rio Grande do Sul, Santa Catarina e Paraná. É possível verificar que no Paraná está concentrado mais ao Oeste enquanto em Santa Catarina e Rio Grande do Sul está na região litorânea (Fig. 44).

Figura 44 – Mapa de calor da região Sul por cidades



Fonte: Dados do estudo (2021).

Os dados deixam evidente, assim, que existe uma diferença importante entre locais nos quais há maior oferta de bolsas para as diferentes áreas do conhecimento avaliadas.

4 CONSIDERAÇÕES FINAIS

A pesquisa é uma parte essencial da construção do conhecimento e deve ser fomentada em todas as áreas do conhecimento, por contribuir para o desenvolvimento do pesquisador, da ciência e dos leitores desses estudos. Atualmente, há ampla disponibilidade das pesquisas e seus resultados por meio das ferramentas tecnológicas de informação e comunicação existentes, eliminando barreiras no acesso aos dados e na disseminação de novas conclusões.

As limitações do trabalho envolveram os gêneros que foram criados a partir de um código que é passível de erro, por isso, enfoca sempre que é apenas uma estimativa e não dados concretos e também a parte do preenchimento dos dados utilizando a mediana e a moda tendo em vista que essa pode não ser a melhor prática para preenchimento de dados faltantes ou irregulares sendo passível de erros.

Diante disso, este estudo teve o objetivo central de obter uma visualização dos dados referentes à oferta de bolsas e auxílios pagos pelo CNPQ de 2002 até 2020 com enfoque nas áreas de TI, que são os dados disponibilizados no portal. Essa disponibilização tem relação direta com a transparência; um dos princípios da administração pública, com o intuito de permitir que a população tenha acesso aos dados que demonstram como são investidos recursos dos cofres públicos.

O estudo demonstrou que o crescimento das bolsas foi até o ano de 2014, onde foi seu pico máximo, e após diminuíram em torno de 25% em questão de quantidade e 50% em questão de valor investido. Desses valores o percentual nas áreas de TI está em torno de 2% a 8%, sendo os últimos dois anos 2019-2020, a maior quantidade de projetos nas áreas de TI (7,5%). O mapa de calor no Brasil mostrou que a concentração está na região litorânea, nas capitais, e na região sul de acordo com as cidades, destaca-se as capitais: Porto Alegre, Florianópolis e Curitiba.

Os gêneros de acordo com a estimativa, de forma geral, mostraram que as mulheres ultrapassaram os homens após 2010 em torno de 1% a 2%, mas nas áreas de TI ainda existem muitas discrepâncias dos dados, onde mostra que as mulheres estão presentes em torno de 30% dos projetos. Isso não significa que a instituição tenha políticas segregacionistas, mas que muitas mulheres desistem de

desenvolver projetos em áreas que ainda são vistas como predominantemente masculinas, justamente pelo receio de não serem aceitas, e optam por áreas culturalmente mais escolhidas entre as mulheres.

Explorando as áreas de TI, foi possível observar que as maiores áreas em quantidade de projetos foram: ciências da computação (54,4%) e matemática (32,8%), porém em valor investido ciências da computação obteve 74,3% enquanto matemática 11,8%. Um Fato relevante considerando o valor único por projetos, entra a área de engenharia de produção onde se obteve o maior valor por projeto. As áreas de acordo com o gênero, somente em comunicação e TIC as mulheres ficaram à frente dos homens.

Já nas subáreas de TI, as subáreas com maior quantidade de projetos foram: matemática aplicada (32%), sistemas de computação (16%), metodologia e técnicas da computação (10%), ciência da computação (8,5%), teleinformática (8%). Porém, o valor investido não acompanha a quantidade: sistemas da computação (17%), metodologia e técnicas da computação (15%), ciência da computação (14%). Um ponto interessante é que a subárea de TI com maior valor por projeto é análise de algoritmos e complexidade de computação e engenharia de software.

Entrando na análise no exterior, o programa ciências sem fronteiras proporcionou um aumento de investimentos em seus anos vigentes (2012-2017). Com base em todos os projetos do programa, os projetos voltados para a área de TI representaram entre 1,6% e 9,5%, porém, em valor investido de 15% a 45,6%. As áreas que mais obtiveram as bolsas foram engenharia civil, ciências da computação e medicina nas modalidades de graduação seguido por pós-graduação. Através dos mapas de calores foi possível observar que as bolsas no exterior foram bem distribuídas em todos os países.

Os dados demonstram que é possível agrupar informações a desenvolver dados estatísticas claros, precisos e confiáveis com base em diferentes temas, desde que sejam utilizadas as ferramentas adequadas para os objetivos que se busca alcançar. Essas ferramentas permitem compreender qual a real situação de um tema sobre o qual se busca mais conhecimentos e esclarecimentos.

Trabalhos futuros, considero utilizar bibliotecas do python com inteligência artificial ou machine learning com intuito de encontrar padrões ou prever dados futuros. Considero também explorar mais os dados, colunas como universidade e estados não foram utilizadas, e por último realizar uma pesquisa mais aprofundada

sobre cada comportamento dos dados com referências mais concretas para agregar mais informações e gráficos correlacionados.

REFERÊNCIAS

- ALMEIDA, Neuler André Soares de; RODAS, Saulo Erick Rocha; MARQUES, Wiston Muniz Ramos. Investimento em pesquisa e inovação tecnológica: um estudo de caso para o Brasil. **Estudo & Debate**, Lajeado, v. 27, n. 1, p. 7-28, 2020.
- ALVES, Camila Guimarães Monteiro de Freitas; OLIVEIRA, Murilo Alvarenga. Análise de investimento e produção em C&T no Brasil entre 2002 e 2010. **FACECLA**, Campo Largo – PR, Brasil. *RECADM*, v. 13, n. 2, p. 156-171, maio Ago./2014.
- ARBIX, Glauco. Dilemas da inovação no Brasil. *In*: TURCHI, Lenita Maria; MORAIS, José Mauro de (orgs). **Políticas de apoio à inovação tecnológica no Brasil: avanços recentes, limitações e propostas de ações**. Brasília: Ipea, 2017.
- BLOCH, Carter; SORENSEN, Mas P. The size of research funding: Trends and implications. **Science and Public Policy**, 2015; 42 (1), 30-43. Disponível em: https://www.researchgate.net/publication/261704307_The_size_of_research_funding_Police/link/5817557208aeffbed6c319ac/download. Acesso em: 24 mar. 2021.
- BRASIL. Ministério das Relações Exteriores. **Dados abertos para a Democracia na Era Digital**. Brasília: Fundação Alexandre de Gusmão, 2011. 84 p. Disponível em: http://funag.gov.br/biblioteca/download/817-Dados_Abertos_para_a_Democracia_na_Era_Digital.pdf. Acesso em: 27 mar. 2021.
- BRASIL. **Gênero dos nomes**. Disponível em: <https://brasil.io/dataset/genero-nomes/grupos/>. Acesso em: 10 abr. 2021.
- CANDIDO, Lucilene F. O.; SANTOS, Natacha C. F.; ROCHA, João B. T da. Perfil dos bolsistas de produtividade em pesquisa nas subáreas da química do CNPq. **Quím. Nova**, São Paulo, v. 39, n. 3, p. 393-405, abr. 2016. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-40422016000300393&lng=en&nem=isso. Acesso em: 24 mar. 2021.
- CAPES; CNPQ. **Ciência Sem Fronteiras: um programa especial de mobilidade internacional em Ciência, Tecnologia e Inovação**. Brasília: CAPES/CNPQ, jul. 2011. Disponível em: . Acesso em: 12 abr. 2021.
- CASTIONI, Remi; MELO, Adriana Almeida Sales de; AFONSO, Maria da Conceição Lima. Bolsa produtividade do CNPq na área de Educação: uma análise com foco na Educação Básica. **Educação e Pesquisa**, São Paulo, v. 46, e221210, 2020. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1517-97022020000100566&lng=en&nem=isso. Acesso em: 19 mar. 2021.
- CONGRESSIONAL RESEARCH SERVICE. **Global Research and Development: Fast Set Update** abril 29, 2020. Disponível em: <https://fas.org/sgp/crs/misc/R44283.pdf>. Acesso em: 25 mar. 2021.
- CREATIVE COMMONS. **Manual dos dados abertos**: governo. 2011. Disponível em:

https://www.w3c.br/pub/Materiais/PublicacoesW3C/Manual_Dados_Abertos_WEB.pdf. Acesso em: 27 mar. 2021.

CRUZ, Cleide Ane Barbosa da; SILVA, Simone Maria Rodrigues da; FRAGA, Érica Emília Almeida. **Educação**: mapeamento de bolsas mestrado e doutorado em educação do CNPq. 8º Encontro Internacional de Formação de Professores. 9º Fórum Permanente de Inovação Educacional. 2016. Disponível em: <https://eventos.set.edu.br/enfope/article/view/1867/546>. Acesso em: 25 mar. 2021.

FARIAS, Everton da Silveira et al. **Investimentos em Ti, Além do Retorno Financeiro**: um Estudo de Caso sobre os Benefícios Intangíveis, Compatibilidade e Adoção da Ti no Processo de Trabalho de uma Empresa. VIII Simpósio de Excelência em Gestão e Tecnologia – SEGET. 2011. Disponível em: Acesso em: 28 mar. 2011.

FAZZIO, Adalberto. Uma breve análise do financiamento da pesquisa no Brasil. **Pesquisa ABC**. Abr. 2017, ed. 19. Disponível em: <https://www.ufabc.edu.br/artigos/uma-breve-analise-do-financiamento-da-pesquisa-no-brasil>. Acesso em: 21 abr. 2021.

FERREIRA, Luciana Rodrigues; ARAÚJO, Jose Nilson Guilherme de. Papel do CNPq no fomento à pesquisa em educação: análise sobre o perfil do bolsista produtividade em pesquisa. **Revista Eletrônica de Educação**, v.13, n.3, p. 1013-1031, set. /dez. 2019. Disponível em: <http://dx.doi.org/10.14244/198271993553%7CRevista>. Acesso em: 24 mar. 2021.

MALAQUIAS, Rodrigo Fernandes; ALBERTIN, Alberto Luiz. Por que os Gestores Postergam Investimentos em Tecnologia da Informação? Um Estudo de Caso. **RAC**, Curitiba, v. 15, n. 6, art. 8, pp. 1120-1136, nov./dez. 2011. Disponível em: <https://www.scielo.br/pdf/rac/v15n6/09.pdf>. Acesso em: 28 mar. 2021.

MANÇOS, Guilherme de Rosso; COELHO, Fernando de Souza. Internacionalização da Ciência Brasileira: subsídios para avaliação do programa Ciência sem Fronteiras. **RBP – Revista Brasileira de Políticas Públicas e Internacionais**, v. 2, n. 2, p. 52-82, dez. 2017.

MARINHO, Antônio Ricardo Monteiro; VALLANDRO, Luiz Felipe Jostmeier; HOPPEN, Norberto. Investimentos em tecnologia da informação: m estudo bibliométrico em eventos e periódicos brasileiros. **Revista Eletrônica de Sistemas de Informação**, [S.l.], v. 14, n. 3, dez. 2015. ISSN 1677-3071. Disponível em: <http://www.periodicosibepes.org.br/index.php/reinfo/article/view/2145>. Acesso em: 16 abr. 2021.

MARQUES, Fabrício. Financiamento em crise. **Pesquisa FAPESP**. Jun. 2017, ed. 256. Disponível em: <https://revistapesquisa.fapesp.br/financiamento-em-crise/>. Acesso em: 21 abr. 2021.

MARTINS, Henrique Castro. A importância da Ciência Aberta (Open Science) na pesquisa em Administração. **Rev. adm. contemp.**, Curitiba, v. 24, n. 1, p. 1-2, Jan. 2020. Disponível em:

http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-6552020000100001&lng=en&nrm=iso. Acesso em: 19 abr. 2021.

MASSI, Luciana; QUEIROZ, Salete Linhares (org.). **Iniciação científica: aspectos históricos, organizacionais e formativos da atividade no ensino superior brasileiro** [online]. São Paulo: Editora UNESP, 2015, 160 p. Disponível em: <https://static.scielo.org/scielobooks/s3ny4/pdf/massi-9788568334577.pdf>. Acesso em: 23 mar. 2020.

MAZZETO, André. **O fim do Ciência sem Fronteiras: Uma análise dos dados**. 2017. Disponível em: <https://medium.com/revista-subjetiva/o-fim-do-ci%C3%A4ncia-sem-fronteiras-uma-an%C3%A1lise-dos-dados-d65f6f88cada#:~:text=Vamos%20ent%C3%A3o%20aos%20dados%3A,a%2015%20bilh%C3%B5es%20nesse%20per%C3%ADodo..> Acesso em: 23 abr. 2021.

MOURA, Egberto Gaspar de; CAMARGO JÚNIOR, Kenneth Rochel de. A crise no financiamento da pesquisa e pós-graduação no Brasil. **Cad. Saúde Pública**. 2017; v. 33, n. 4: e00052917. Disponível em: http://cadernos.ensp.fiocruz.br/static/arquivo/csp_0529_17_Crise-pt.pdf. Acesso em: 20 abr. 2021.

OLINTO, Gilda. A inclusão das mulheres nas carreiras de ciência e tecnologia no Brasil. **Inc. Soc.**, Brasília, DF, v. 5 n. 1, p.68-77, jul./dez. 2011.

OLIVEIRA, Jussara Ribeiro de; MELLO, Lívia Coelho; RIGOLIN, Camila Carneiro Dias. Participação feminina na pesquisa sobre tecnologia da informação no Brasil: grupos de pesquisa e produção científica de teses e dissertações. **Cad. Pagu**, Campinas, n. 58, e205804, 2020. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-83332020000100503&lng=en&nrm=iso. Acesso em: 22 abr. 2021.

PAULA, Milena Lima de; JORGE, Maria Salete Bessa; MORAIS, Jamie Borges de. O processo de produção científica e as dificuldades para utilização de resultados de pesquisas pelos profissionais de saúde. **Interface - Comunicação, Saúde, Educação**, 2019; v. 23, e190083. Disponível em: <https://doi.org/10.1590/Interface.190083>. Acesso em: 25 mar. 2021.

PRESSE, France. **Mulheres no poder no mundo**. Out. 2010. Disponível em: <http://g1.globo.com/mundo/noticia/2010/10/mulheres-no-poder-no-mundo.html>. Acesso em: 5 abr. 2021.

RIBEIRO, Daniella Borges et al. Financiamento à ciência no Brasil: distribuição entre as grandes áreas do conhecimento. **Rev. entre as**, Florianópolis, v. 23, n. 3, p. 548- 561, dez. 2020. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1414-49802020000300548&lng=na Educação=isso. Acesso em: 23 mar. 2021.

RIBEIRO, Daniella Borges et al. Financiamento à ciência no Brasil: distribuição entre as grandes áreas do conhecimento. **Rev. katálysis**, Florianópolis, v. 23, n. 3, p. 548-561, Dez. 2020. Disponível em:

http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1414-49802020000300548&lng=en&nrm=iso. Acesso em: 19 abr. 2021.

ROSA, Mislene Aparecida Gonçalves; QUIRINO, Raquel Gonçalves. Relações de Gênero na Ciência e Tecnologia (C&T): estudo de caso de um Centro Federal de Educação Tecnológica. **Diversidade E Educação**, 2017; v. 4, n. 8, p. 42–55. Disponível em: <https://periodicos.furg.br/divedu/article/view/6739>. Acesso em: 22 abr. 2021.

SPIPKI, Fernando Rosado. Perfil dos bolsistas de produtividade do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) na área de Medicina Veterinária. **Pesq. Vet. Bras.**, Rio de Janeiro, v. 33, n. 2, p. 205-213, fev. 2013. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-736X2013000200012&lng=neném=isso. Acesso em: 22 mar. 2021.

TENÓRIO, Marge; MELLO, Guilherme Arantes; VIANA, Ana Luiza D'Ávila Viana. Políticas de fomento à ciência, tecnologia e inovação em saúde no Brasil e o lugar da pesquisa clínica. **Ciências e Saúde Coletiva**. 2017, v. 22, n. 5, p. 1441-1454. Disponível em: <https://www.scielo.br/pdf/csc/v22n5/1413-8123-csc-22-05-1441.pdf>. Acesso em: 22 abr. 2021.