



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
INSTITUTO METRÓPOLE DIGITAL
BACHARELADO EM TECNOLOGIA DA INFORMAÇÃO
IMD1153 - ANÁLISE VISUAL DE DADOS

Ian Mendes Fernandes Do Nascimento
Marcone Rafael Do Nascimento Júnior
Wellder Bernardo De Oliveira

**RELATÓRIO REFERENTE A VISUALIZAÇÃO E ANÁLISE DE DADOS DO PERFIL
DE VISITANTES DO MUSEU CÂMARA CASCUDO COLETADOS PELA
PLATAFORMA *PUBLICUS***

NATAL/RN
MAIO/2023

SUMÁRIO

LISTA DE FIGURAS.....	2
LISTA DE TABELAS.....	3
1. OBJETIVO.....	4
2. INTRODUÇÃO.....	4
3. METODOLOGIA.....	5
3.1. DESAFIOS INICIAIS.....	5
3.2. CRIAÇÃO ARTIFICIAL DE DADOS.....	6
3.3. PLANEJAMENTO.....	6
3.3.1. Ian:.....	6
3.3.2. Marcone:.....	9
3.3.3. Wellder:.....	11
3.4. ESCOLHAS E RESULTADOS.....	13
3.4.1. Ian.....	13
3.4.2. Marcone.....	15
3.4.2.1 Principais problemas:.....	16
3.4.2.2 Tentativa de solução:.....	17
3.4.2.3 Novos problemas.....	18
3.4.2.4 Novo gráfico.....	18
3.4.2.5 Acesso a geração de dados e ao gráfico interativo.....	19
3.4.3. Wellder.....	20
4. APÊNDICES.....	22
5. REFERÊNCIAS.....	30

LISTA DE FIGURAS

LISTA DE FIGURAS.....	2
LISTA DE TABELAS.....	3
Figura 1: Guia esquematizado para visualização de dados para análise.....	5
Figura 2: Dataframe criado utilizando Pandas.....	13
Figura 3 - Gráfico final (lan).....	14
Figura 4 - Gráfico 1 (Marccone).....	16
Figura 5 - Gráfico 2 (Marccone).....	17
Fonte: Autoria própria.....	18
Figura 6 - Novo gráficos solução, primeiro. (Marccone).....	18
Figura 7 - Detalhe da legenda.....	18
Figura 8 - Novo gráfico solução, segundo. (Marccone).....	19
Figura 9 - Detalhe da legenda.....	19
Figura 10: Esboço da representação geográfica.....	20
Figura 11: Esboço da representação 1.....	21
Figura 12: Gráfico da solução final.....	21

LISTA DE TABELAS

Tabela 1 - Perguntas geradoras de possíveis análises (lan).....	6
Tabela 2 - Tabela para representação gráfica (lan).....	7
Tabela 3 - Detalhamento, Marcone.....	9
Tabela 4 - Detalhamento, Wellder.....	12

1. OBJETIVO

Este documento tem como objetivo relatar o processo de visualização de dados orientada por objetivos de análise a partir do banco de dados providenciado pela plataforma *Publicus*, versão 1.0, referente ao perfil de visitação do Museu Câmara Cascudo (MCC), localizado na cidade de Natal, Rio Grande do Norte, nos anos de 2022 e 2023.

2. INTRODUÇÃO

Como consta em seu manual de uso, “*Publicus*”, em sua atual versão 1.0, é um “software de gestão de informações sobre públicos que visitam museus e instituições semelhantes.” (SILVA, 2023, p.5). É de propriedade da Universidade Federal do Rio Grande do Norte e, em sua arquitetura, conta com dois módulos separados: o módulo de contagem de público e o módulo de armazenamento de perfil de visitante. Enquanto o primeiro armazena variáveis quantitativas (com exceção da variável *área de visitação*) como data e horário de entrada, quantidade de pessoas visitando por vez e valor da entrada, o módulo perfil de visitante conta com quinze perguntas com quantidade de respostas variáveis e de diversos tipos.

Este relatório irá se aprofundar nas propostas de análises que foram selecionadas após o estudo dos dados coletados pelo *software*, referentes ao Museu Câmara Cascudo, localizado na cidade de Natal, Rio Grande do Norte. O estudo foi orientado pelo exame da qualidade dos dados fornecidos e usabilidade dos possíveis resultados obtidos para estratégias de gestão do Museu Câmara Cascudo. Limitações, como a organização dos dados em dois bancos de dados separados, foram importantes na tomada de decisões. Mais sobre o assunto será discorrido na subseção *Desafios*, em *Metodologia*.

3. METODOLOGIA

3.1. DESAFIOS INICIAIS

É importante salientar que a plataforma fornece recursos para a visualização automatizada de dados em gráficos e tabelas referente aos dois módulos para usuários autorizados. Esse fato propôs o primeiro desafio enfrentado, pois diminuiu bastante o escopo de análises originais a serem propostas. Por meio da leitura do manual é possível constatar que a maioria das análises pertinentes relacionadas ao módulo de contagem de público podem ser feitas rapidamente por um gestor, ademais, exemplos encontram-se no guia.

Figura 1: Guia esquematizado para visualização de dados para análise.



Fonte: Silva, 2023.

Outro desafio foi a impossibilidade do cruzamento entre dados dos dois módulos, devido à separação dos dados em bancos distintos. Cada integrante, portanto, teve que escolher uma única base com a qual trabalhar. Com isso em mente e após ponderação, o grupo decidiu limitar o estudo aos dados referentes ao módulo de perfil de grupo. Os fatores determinantes foram: a maior quantidade de comparações que poderiam ser feitas devido à diversidade de perguntas e respostas e a diversidade de tipos de variáveis que cada resposta oferecia.

Utilizando o *software* MySQL por meio da ferramenta phpMyAdmin para estudo da base de dados, constatou-se problema com a qualidade desses, como a ausência de respostas em algumas colunas ou a presença de respostas

incoerentes (ex.: no campo “gênero”, uma linha foi preenchida com “mamífero”). Essas intercorrências são antecipadas no manual de uso, ao enfatizar a importância do papel da recepção na coleta correta dos dados (ex.: cadastrar grupos corretamente) e ao alertar que “é importante lembrar que os visitantes podem pular uma pergunta do questionário de coleta de dados” (SILVA, 2023, p. 49). Isso permite que análises cheguem a conclusões equivocadas. A solução encontrada pelo grupo foi a criação de um banco de dados fictício.

A falta de conhecimento prévio de ferramentas para a programação de visualização de dados também foi um fator a ser superado. Foram realizados estudos de bibliotecas e tutoriais para a realização deste trabalho, com enfoque nas bibliotecas *Pandas*, *NumPy* e a ferramenta *Plotly*.

3.2. CRIAÇÃO ARTIFICIAL DE DADOS

Devido à escolha da criação artificial de dados, os gráficos gerados não traduzem a realidade do perfil de visitação do MCC. No entanto, a proposta de estruturas poderá inspirar a criação de análises baseadas na realidade.

Os dados fictícios foram criados usando as plataformas *Jupyter* e *Google Colab*. Para a criação dos *dataframes* foi utilizada a linguagem *Python*, juntamente com as bibliotecas *Pandas*, *NumPy* e *Random*. Para a visualização, foram usadas as ferramentas *Plotly*, e a biblioteca *Seaborn*.

3.3. PLANEJAMENTO

3.3.1. Ian:

Três questões foram formuladas inicialmente, as quais encontram-se na tabela abaixo bem como no relatório referente à primeira unidade deste curso. Dentre elas, foi escolhida a seguinte questão: como o nível de escolaridade, relacionado à faixa etária, motivo de visitação e fonte de informação (como soube do MCC) se expressa relacionados aos meses de 2022, tendo como base os dados coletados de perfil de público?

Tabela 1 - Perguntas geradoras de possíveis análises (Ian)

Perguntas

I.	Qual estado e região brasileira representa a maior quantidade de visitantes do MCC levando em consideração os meses e tendo como base os visitantes que responderam ao questionário de visita ao MCC entre 2022 e o primeiro trimestre de 2023?
II.	Qual a variação da quantidade total de visitantes por semana nos anos e meses, tendo como base os dados de fluxo de visitação desde 2017?
III.	Como o nível de escolaridade, relacionado à faixa etária, motivo de visitação e fonte de informação (como soube do MCC) se expressa relacionados aos meses de 2022, tendo como base os dados coletados de perfil de público?

Fonte: Autoria própria

A pergunta de número II foi descartada devido à presença de análises muito similares no manual de uso do *Publicus*. A escolha da questão norteadora se deu devido à potencialidade de impacto da análise nas escolhas de gestão: a pergunta III oferece um panorama da efetividade de estratégias de divulgação do museu em cada faixa etária e nível de escolaridade, além de relacionar às motivações de cada grupo. Torna-se visível como comparar essas variáveis pode ser benéfico para as resoluções do MCC.

Após a escolha da pergunta norteadora, para o planejamento da visualização, foi preenchida uma tabela de representações. Segue abaixo:

Tabela 2 - Tabela para representação gráfica (lan)

Variável	Tipo de dado	Possíveis valores	qtde. de valores	Representação 1		Representação 2	
				Símbolo 1	Características visuais 1	Símbolo 2	Características visuais 2
Nível de Escolaridade	Ordem em faixas	<ul style="list-style-type: none"> • Ensino Fundamental Incompleto • Ensino Fundamental Completo • Ensino Médio Incompleto • Ensino Médio Completo • Ensino 	7	Barras	Comprimento (posição horizontal) e cor	Área	Círculo

		<ul style="list-style-type: none"> Superior Incompleto • Ensino Superior Completo • Pós-graduação 					
Faixa Etária	Ordem	<ul style="list-style-type: none"> • 10 a 14 anos • 15 a 19 anos • 20 a 29 anos • 30 a 39 anos • 40 a 49 anos • 50 a 59 anos • 60 a 64 anos • Mais de 65 anos. 	8	Barras	Posição em linha horizontal	Área	Quadrado
Motivo	Categórico	<ul style="list-style-type: none"> • Para nos conhecer • Por lazer em geral • Acompanhando alguém • Conhecer nossas novidades • Visitar alguma exposição específica • Participar de algum evento • Realizar estudo/pesquisa • Outros motivos. 	8	Área	Diferentes cores	linha	posição horizontal
Fonte	Categórico	<ul style="list-style-type: none"> • Redes sociais, • TV / rádio / jornal • notícias na internet • site do museu • panfletos • amigos / familiares e outros. 	7	Área	Diferentes símbolos	linha	posição vertical
Meses Ano de 2022	Ordem	Janeiro, Fevereiro, Março, Abril, Maio, Junho, Julho, Agosto, Setembro, Outubro, Novembro, Dezembro	12	Pontos	Posição horizontal em uma linha	Pontos	Posição vertical

Quantidade de respostas total	Quantidade	Indefinido	Indefinido	Área	Figura geométrica	Pontos	Posição vertical
-------------------------------	------------	------------	------------	------	-------------------	--------	------------------

Fonte: Autoria própria

Vale salientar que a anotação de possibilidades em tabela serviu apenas como norte para possibilidades de representação. Não necessariamente as opções foram acatadas para o resultado final.

Alguns rascunhos foram feitos como possibilidades e encontram-se no anexo B, levando em consideração quais aspectos deveriam ser priorizados e quais variáveis iriam ser dispostas nos eixos. Para esse planejamento, o tipo de dado e o conhecimento de tipos de gráficos foram consideradas características de interesse, pois, por exemplo, dados quantitativos contínuos são mais facilmente dispostos em eixos caso o interesse seja criar um gráfico de linhas.

Nessa etapa, foram experimentadas disposições diferentes de variáveis nos eixos. Até o momento, a inclinação era programar um gráfico de dispersão (ou de bolhas ou de pontos).

3.3.2. **Marcone:**

Buscando resolver o seguinte questionamento “ Qual é a área de formação mais comum entre os visitantes do museu relacionada com a faixa média de visitas a qualquer museu e ao museu câmara cascudo em 2022?”, foi identificada a necessidade das seguintes variáveis: Quantidade de visitantes, Área de formação, faixa de visitas, Museu Câmara Cascudo, faixa de visitas a outros museus e mês.

Para o planejamento do gráfico foi feito a seguinte tabela com as possíveis representações de cada variável envolvida:

Tabela 3 - Detalhamento, Marcone.

Variável	Tipo de dado	Possíveis valores	Representação 1		Representação 2	
			Símbolo 1	Características visuais 1	Símbolo 2	Características visuais 2

Quantidade de visitantes	Quantidade	De 0 a 40	Linha	posição diagonal	Área	faixa de cores
Área de formação	Catégorico	<ul style="list-style-type: none"> • Ciências agrárias - (x %) • Ciências biológicas • Ciências exatas e da terra • Ciências humanas, letras e artes • Ciências da saúde • Ciências sociais aplicadas • Engenharias • Outra área 	ponto	rótulo	Área	forma
Faixa de visitas ao MCC	Quantidade	<ul style="list-style-type: none"> • Esta é a primeira vez - (x %) • Entre 2 e 5 vezes • Entre 6 e 10 vezes • Mais de 10 vezes 	Número	Posição vertical	Linha	Posição horizontal

Faixa de visitas a outros museus	Catégorico	<ul style="list-style-type: none"> • Esta é a primeira vez - (x %) • Entre 2 e 5 vezes • Entre 6 e 10 vezes • Mais de 10 vezes 	Ponto	Posição horizontal	Ponto	Posição vertical
Mês	Ordem	<ul style="list-style-type: none"> • Janeiro • Fevereiro • Março • Abril • Maio • Junho • Julho • Agosto • Setembro • Outubro • Novembro • Dezembro 	ponto	rótulo	Área	Escala de cor

Fonte: Autoria própria.

3.3.3. Wellder:

Proveniente às análises desenvolvidas na unidade passada em relação aos dados do MCC, a escolhida para a progressão da atividade do desenvolvimento do gráfico foi a de número 2:

- Dentre os dados coletados do ano de 2022 e analisando seus meses. Dos visitantes que responderam o questionário, qual é o nível de retorno ao museu comparado a pessoas que visitam pela primeira, relacionada à potiguares ?

Entretanto passou por uma adequação para melhorar o nível de comparação. De início, a ideia seria comparar a de visitação de cada cidade do Rio Grande do Norte, porém, não foi possível se adequar a esse grau de especificação, e também surgiu a necessidade de comparar com as cidades do Brasil. Para solucionar esse imbróglio foi necessário essas cidades em regiões, no caso do RN, uma divisão em quatro Mesorregiões, Central Potiguar, Leste Potiguar, Oeste Potiguar e Agreste Potiguar, e no caso do Brasil, uma divisão em suas regiões, Norte, Nordeste, Centro-Oeste, Sudeste e Sul.

A seguir, para o planejamento do gráfico temos as possíveis representações visuais pensadas, entretanto não necessariamente foram as soluções finais do problema.

Tabela 4 - Detalhamento, Wellder.

Variável	Tipo de dado	Possíveis valores	Representação 1		Representação 2	
			<i>Símbolo 1</i>	<i>Características visuais 1</i>	<i>Símbolo 2</i>	<i>Características visuais 2</i>
Frequência de visitação	Categórico	<ul style="list-style-type: none"> Esta é a primeira vez Entre 2 e 5 vezes Entre 6 e 10 vezes Mais de 10 vezes 	linha	cor característica	linha	
Proporção	Quantitativo	0% - 100%	Área	Barras	área	cores
Mês/ano	Ordem	<ul style="list-style-type: none"> Janeiro 2022 Fevereiro 2022 Março 2022 Abril 2022 Maio 2022 Junho 2022 Julho 2022 Agosto 2022 Setembro 2022 Outubro 2022 Novembro 2022 Dezembro 2022 	Ponto	rótulo na barra horizontal	linha	posição
Estados do Brasil e	Categórico	<ul style="list-style-type: none"> 5 regiões 4 Mesorregiões 	Área	Cores	área	símbolos

Mesorregiões do Rn						
--------------------	--	--	--	--	--	--

Fonte: Autoria própria.

3.4. ESCOLHAS E RESULTADOS

3.4.1. Ian

Após a geração do *dataframe* com dados artificiais, era a hora de experimentação e realização de escolhas.

Figura 2: Dataframe criado utilizando Pandas.

	mes	esc	fonte	motivo	quantidade
0	Jan	Ens Fun I	Redes Sociais	Para nos conhecer	15.0
1	Jan	Ens Fun C	Redes Sociais	Para nos conhecer	28.0
2	Jan	Ens Med I	Redes Sociais	Para nos conhecer	37.0
3	Jan	Ens Med C	Redes Sociais	Para nos conhecer	24.0
4	Jan	Ens Sup I	Redes Sociais	Para nos conhecer	45.0
...
4699	Dez	Ens Med I	Outros	Outros	3.0
4700	Dez	Ens Med C	Outros	Outros	26.0
4701	Dez	Ens Sup I	Outros	Outros	12.0
4702	Dez	Ens Sup C	Outros	Outros	31.0
4703	Dez	Pos	Outros	Outros	26.0

4704 rows x 5 columns

Fonte: Autoria própria.

O grande desafio nessa etapa mostra ser a quantidade expressiva de variáveis categóricas e ordinais qualitativas. A categoria “fonte de informação” possui sete possibilidades, enquanto “motivo” possui oito. Nível de escolaridade, faixa etária e meses são qualitativas ordinais, com sete, oito e doze possíveis valores respectivamente. O excesso de possíveis valores impacta a densidade visual dos gráficos e ameaça tornar a representação visual confusa.

Segue-se que, com o intuito de diminuir a quantidade de informação representada, ou seja, “enxugar” o gráfico, algumas variáveis deveriam ser priorizadas em detrimento de outras. Para tal, foi trazido de volta e como foco o objetivo de análise inicial. Como a intenção por trás deste trabalho era encontrar uma forma de analisar a eficácia das diferentes fontes de informação priorizando

como aspecto a escolaridade na dimensão do tempo, as variáveis “motivo de visitação” e “faixa etária” foram desconsideradas.

Ainda assim, o fato da maioria das variáveis serem qualitativas ordinais ou nominais impede que as informações se distribuam pelo gráfico, ficando a representação “presa” a interseções entre valores, gerando sobreposição de pontos, por exemplo. Na primeira tentativa, ao programar um gráfico de pontos, percebe-se que o gráfico ainda sofria de um excesso visual (apêndice C). O primeiro gráfico também relacionava apenas três variáveis, portanto deveria ser descartado.

Mudando a abordagem, uma alternativa que evidencia a quantidade de fontes de informação são gráficos de coluna com as diferentes fontes agrupadas, distribuídas por quantidade em y e por meses em x. Isso significou um passo à frente na realização da análise correta. Restava apenas descobrir como evidenciar a escolaridade. Esse paradigma foi resolvido a partir da *subplotagem* do gráfico baseado em escolaridade. O resultado final foi um gráfico com sete *subplots* e barras agrupadas, diferenciadas por cor.

Figura 3 - Gráfico final (lan).



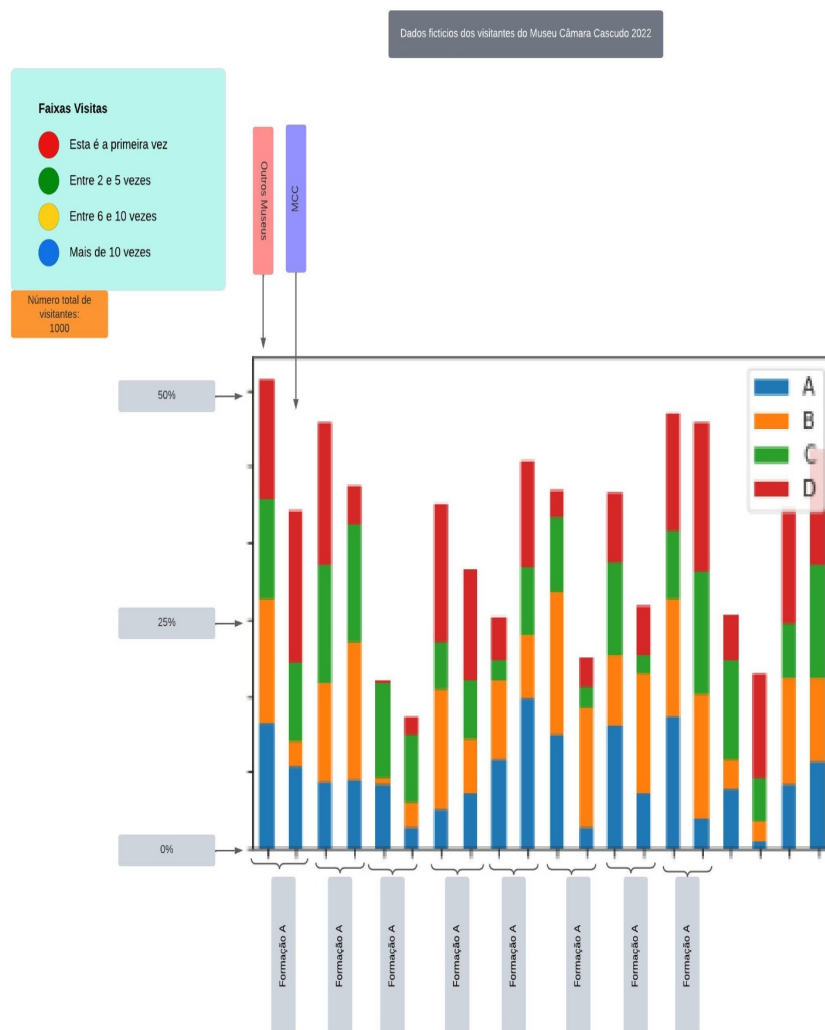
Fonte: Autoria Própria

Uma imagem em melhor qualidade do gráfico acima encontra-se em apêndice D. Mais pode ser visto no collab, presente em: [Arquivos de lan](#).

3.4.2. Marcone

Para a construção do gráfico, usamos os mesmos artifícios tecnológicos usados na criação dos dados junto com um aplicativo de diagramação online chamado lucidchart, e foram esboçados os modelos candidatos para a análise. A seguir veremos cada um deles, seus benefícios e suas dificuldades.

Figura 4 - Gráfico 1 (Marcone)



Fonte: Autoria própria.

No gráfico acima, temos, com empilhamento de colunas destacando a variável formação definida no planejamento o eixo x representado pelas 8 possíveis formações, o eixo y composto pela porcentagem de visitação por faixa entre museus, as barras são compostas de 2 em duas para representar essa porcentagem para os museu em geral e também entre o museu Câmara Cascudo, há também uma legenda explicando o significado de cada faixa de cor representada na coluna.

3.4.2.1 Principais problemas:

O gráfico não apresenta nenhuma escala temporal, o que impossibilita uma análise justa sabendo que as visitas e a formação das pessoas que a fazem variam entre os meses.

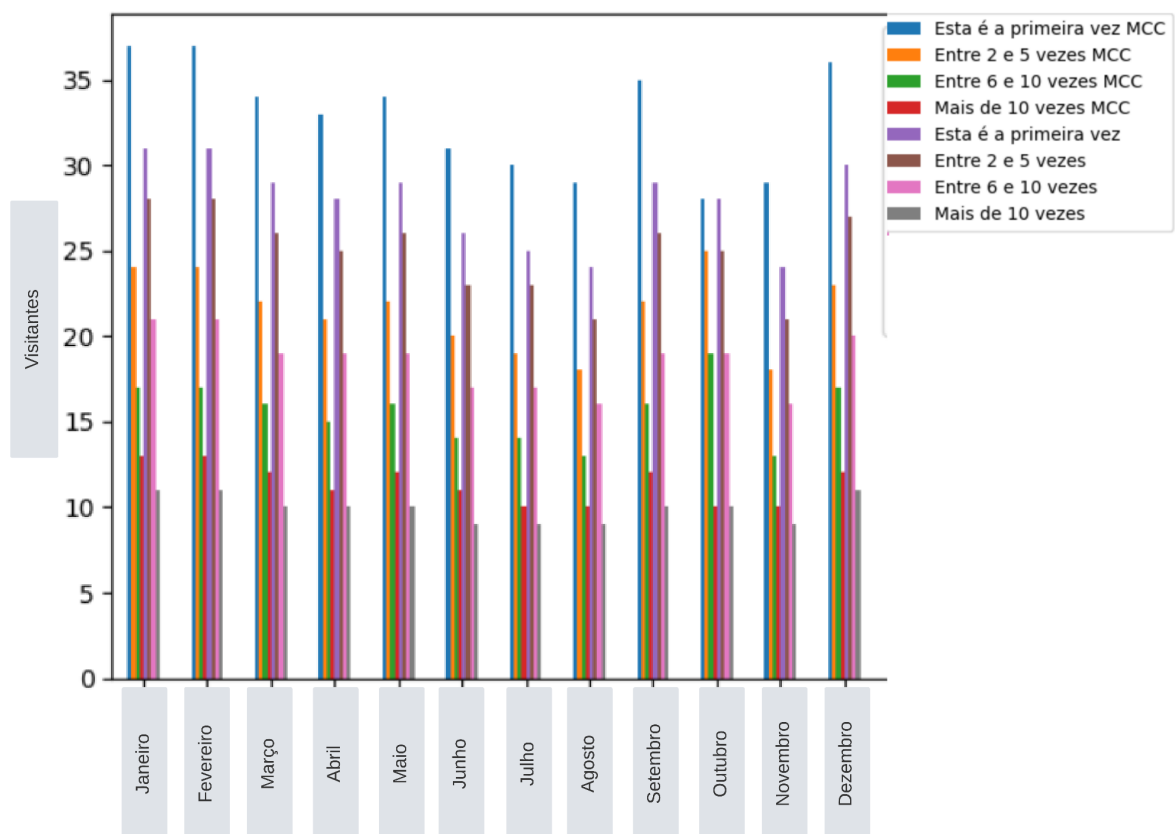
As faixas de cores entre a representação nas colunas dos museus em geral e do museu Câmara Cascudo estão iguais, o que pode dificultar ou até mesmo impossibilitar a distinção deles e a compreensão do gráfico.

Se os valores do eixo forem iguais para mais de uma barra que representa a faixa de visitação, as barras podem ficar sobrepostas e impossibilitar a visualização e consequentemente o entendimento.

3.4.2.2 Tentativa de solução:

Na tentativa de solucionar os problemas anteriormente mencionados removi a variável formações, adicionei a variável meses e reconstruí o gráfico obtendo o seguinte resultado:

Figura 5 - Gráfico 2 (Marcane)



Fonte: Autoria própria.

3.4.2.3 Novos problemas

Com o novo gráfico eu resolvi os principais problemas do gráfico anterior, porém surgiram novos problemas: As cores são parecidas e as barras muito finas dificultando a visualização e a escala numérica está errada.

3.4.2.4 Novo gráfico

Após o problema anterior eu repensei meu gráfico e em vez de remover uma variável, decidi fazer 2 gráficos diferentes usando uma nova estrutura.

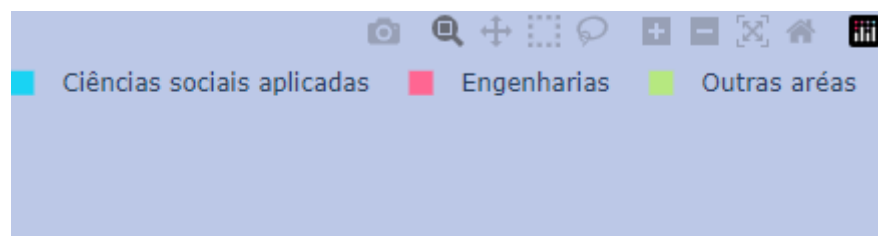
O primeiro analisa a quantidade de visitantes de cada formação em um mês no *museu Câmara Cascudo* por faixa de visitação:

Figura 6 - Novo gráficos solução, primeiro. (Marcone)



Fonte: Autoria própria.

Figura 7 - Detalhe da legenda



Fonte: Autoria própria.

O segundo analisa a quantidade de visitantes de cada formação em um mês para *qualquer museu* por faixa de visitação:

Figura 8 - Novo gráfico solução, segundo. (Marcone)



Fonte: Autoria Própria

Figura 9 - Detalhe da legenda



Fonte: Autoria própria

3.4.2.5 Acesso a geração de dados e ao gráfico interativo

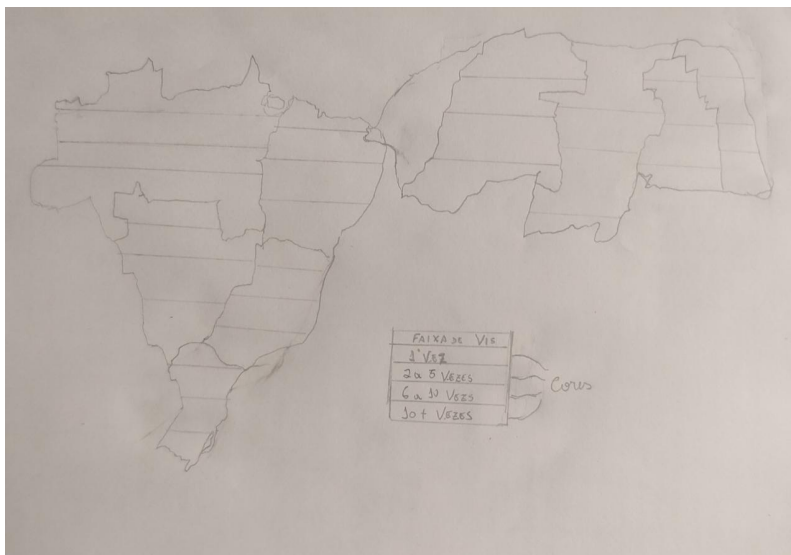
As imagens foram recortadas para facilitar a visualização mas na plataforma [Google colab](https://colab.research.google.com/) onde esses gráficos se encontram eles são fáceis de ver além de serem interativos. Onde também se encontram os códigos usados na geração dos meus dados fictícios.

3.4.3. Wellder

Levando em consideração a utilização de dados fictícios, a tabela de dados tem se sua geração a partir de um do dataframe gerado por códigos com um nível de aleatoriedade para a implementação do gráfico e resolução da problemática.

Partindo disso paralelo as características visuais geradas, foi sugerido a utilização de um gráfico geográfico pensando em tentar suprir a variável de regiões explicada no tópico anterior. Dada a ideia, os esboços foram criados tentando adequar as variáveis propostas, o mapa representa a variável de regiões, as divisões representa a variável de faixa de visitas e os tamanhos seriam as proporções:

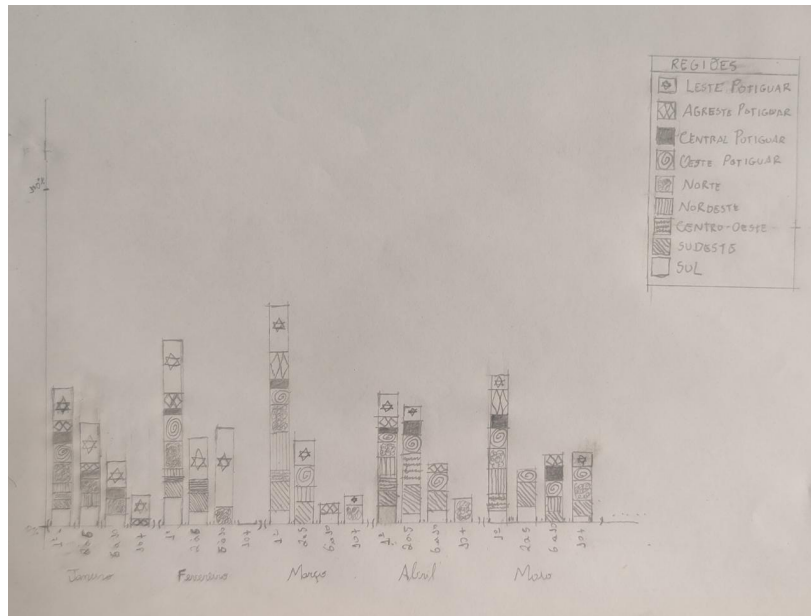
Figura 10: Esboço da representação geográfica



Como mostrado, surgiu a dificuldade em encaixar a variável de tempo, ou seja, os meses do ano, e também a representação de áreas com incidências de zero, sendo assim, a ideia foi descartada.

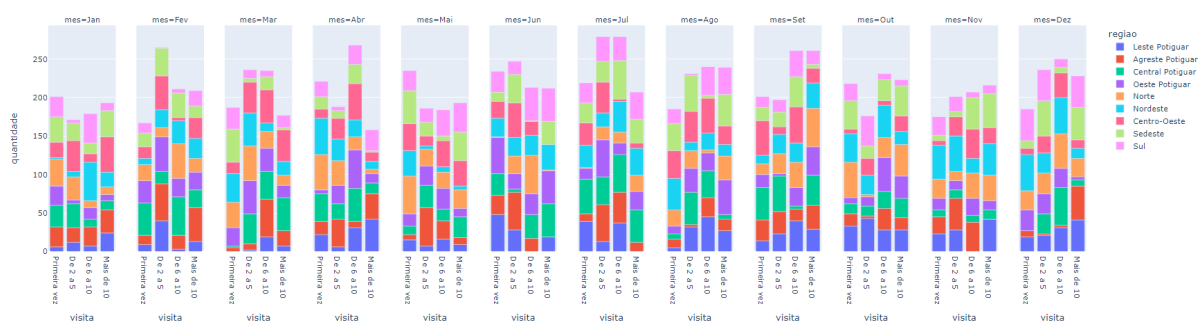
Seguindo a primeira representação gráfica da tabela, porém com algumas alterações, foi possível esboçar uma ideia viável de resolução, de modo que, a variável de proporção foi representada no eixo y, faixa de visita representadas por barras, meses representadas seguindo o eixo x e a variável das regiões representadas por divisões nas barras para representar a proporção de visita de acordo com sua região.

Figura 11: Esboço da representação 1



Para simular cores, símbolos foram colocados no lugar mas não fazem parte da ideia inicial. Partindo para a programação, foi proveniente que a variável proporcional fosse substituída por uma representação relacionada a valores absolutos, sendo assim foram montadas de acordo com o número de quantidades puros. Sendo assim, gerando o gráfico final que representa a resolução da pergunta proposta e suas análises paralelas que podem ser observadas a partir dele.

Figura 12: Gráfico da solução final



Para facilitar a visualização é possível visualizar melhor na plataforma que foram gerados os dados fictícios, [colab](#). Nele é possível interagir com o gráfico e observar os valores de cada uma das variáveis.

4. APÊNDICES

APÊNDICE A - CÓDIGO EM PYTHON PARA GERAÇÃO DE DATAFRAME (IAN)

```
#importação de bibliotecas
%matplotlib inline
import seaborn as sns
import pandas as pd
import numpy as np
import random as rd
import plotly.express as px
import plotly
import plotly.graph_objects as go

#criação de vetores
dmes = np.array([])
dfonte = np.array([])
dmotivo = np.array([])
desc = np.array([])
dqt = np.array([])

#criação de listas com variáveis (meses x escolaridade x qtde. total x
fonte de informação x motivo)
for mes in range (1,13):
    for fonte in range(1,8):
        for motivo in range (1,9):
            for esc in range (1,8):
                #meses de 2022
                match mes:
                    case 1:
                        dmes = np.append(dmes, 'Jan')
                    case 2:
                        dmes = np.append(dmes, 'Fev')
                    case 3:
                        dmes = np.append(dmes, 'Mar')
                    case 4:
                        dmes = np.append(dmes, 'Abr')
                    case 5:
                        dmes = np.append(dmes, 'Mai')
                    case 6:
                        dmes = np.append(dmes, 'Jun')
```

```

        dmes = np.append(dmes, 'Jul')
    case 8:
        dmes = np.append(dmes, 'Ago')
    case 9:
        dmes = np.append(dmes, 'Set')
    case 10:
        dmes = np.append(dmes, 'Out')
    case 11:
        dmes = np.append(dmes, 'Nov')
    case 12:
        dmes = np.append(dmes, 'Dez')

match fonte:
    case 1:
        dfonte = np.append(dfonte, 'Redes Sociais')
    case 2:
        dfonte = np.append(dfonte, 'TV / rádio / jornal')
    case 3:
        dfonte = np.append(dfonte, 'Amigos/Familiares')
    case 4:
        dfonte = np.append(dfonte, 'Site do Museu')
    case 5:
        dfonte = np.append(dfonte, 'Notícias da Internet')
    case 6:
        dfonte = np.append(dfonte, 'Panfleto')
    case 7:
        dfonte = np.append(dfonte, 'Outros')

match motivo:
    case 1:
        dmotivo = np.append(dmotivo, 'Para nos conhecer')
    case 2:
        dmotivo = np.append(dmotivo, 'Por lazer em geral')
    case 3:
        dmotivo = np.append(dmotivo, 'Acompanhando alguém')
    case 4:
        dmotivo = np.append(dmotivo, 'Conhecer nossas novidades')
    case 5:
        dmotivo = np.append(dmotivo, 'Visitar alguma exposição
específica')
    case 6:
        dmotivo = np.append(dmotivo, 'Participar de algum evento')

```



```

        case 7:
            dmotivo = np.append(dmotivo, 'Realizar estudo/pesquisa')
        case 8:
            dmotivo = np.append(dmotivo, 'Outros')

    match esc:
        case 1:
            desc = np.append(desc, 'Ens Fun I')
        case 2:
            desc = np.append(desc, 'Ens Fun C')
        case 3:
            desc = np.append(desc, 'Ens Med I')
        case 4:
            desc = np.append(desc, 'Ens Med C')
        case 5:
            desc = np.append(desc, 'Ens Sup I')
        case 6:
            desc = np.append(desc, 'Ens Sup C')
        case 7:
            desc = np.append(desc, 'Pos')

    valor = rd.randint(0,50)
    dqt = np.append(dqt, valor)

defs = pd.DataFrame({'mes': dmes,
                     'esc': desc,
                     'fonte': dfonte,
                     'motivo': dmotivo,
                     'quantidade': dqt
                     })

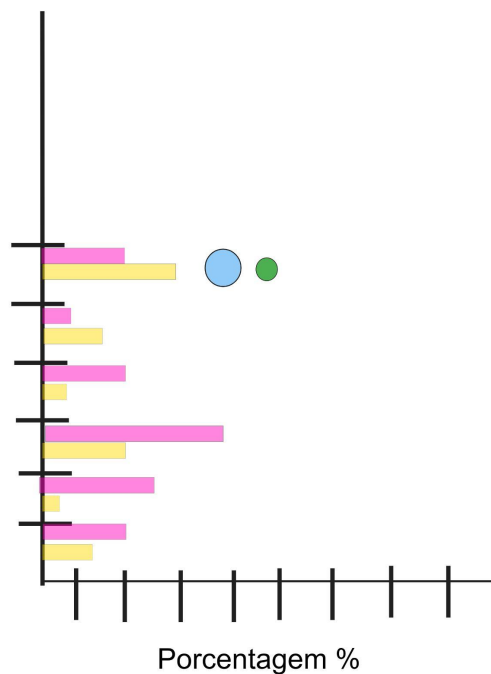
#retirar todos os valores zeros
defs.replace(0, np.nan, inplace=True)
defs

```

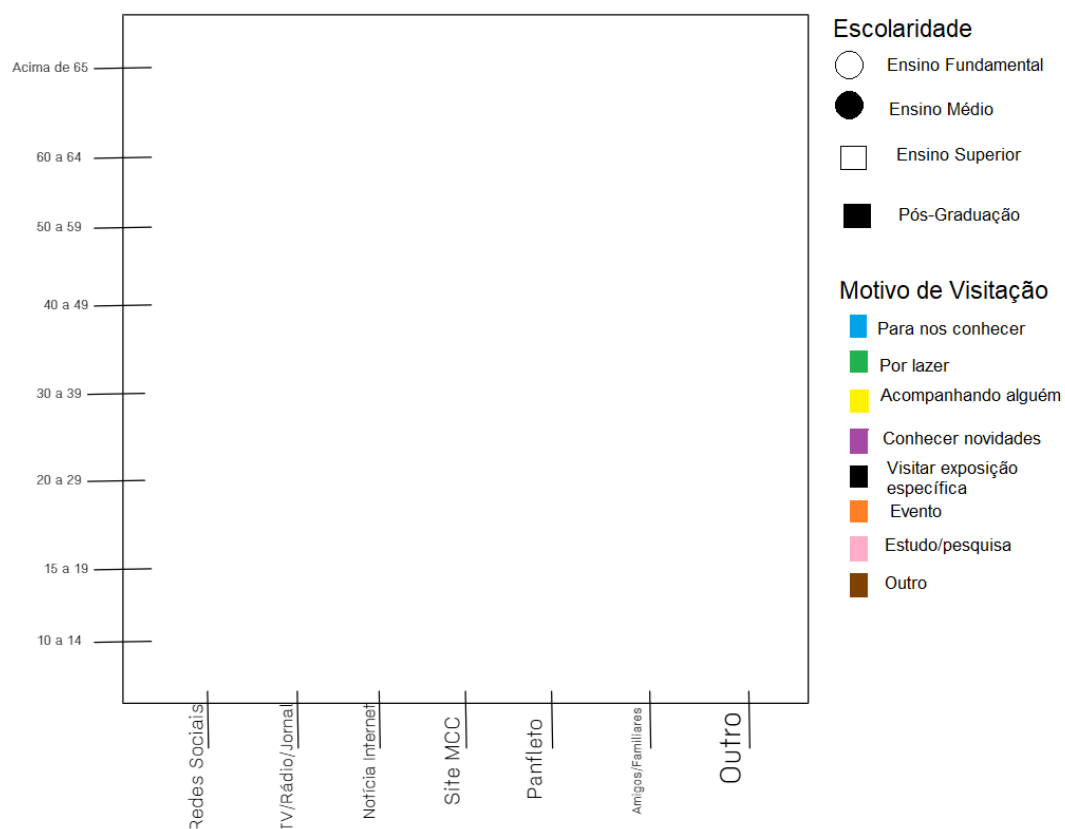
APÊNDICE B - RASCUNHOS DE PLANEJAMENTO (IAN)

Primeiro rascunho: Quantidade de respostas representada em porcentagem no eixo x. Escolaridade e faixa etária diferenciadas por cor, utilização de gráfico de barras agrupadas. Motivo de visita o representado por  rea em c rculo, diferenciando os tr s principais motivos por tamanho e cor. Fonte de informa o representada por  rea em estrela, diferenciado tamb m por cor.

Escolaridade █ Motivo de visitaç o (3 mais escolhidos em cada faixa de escolaridade) ●
 Faixa Et ria █
 Percentagem a partir da quantidade total de respostas
 Fonte de informa  o ★

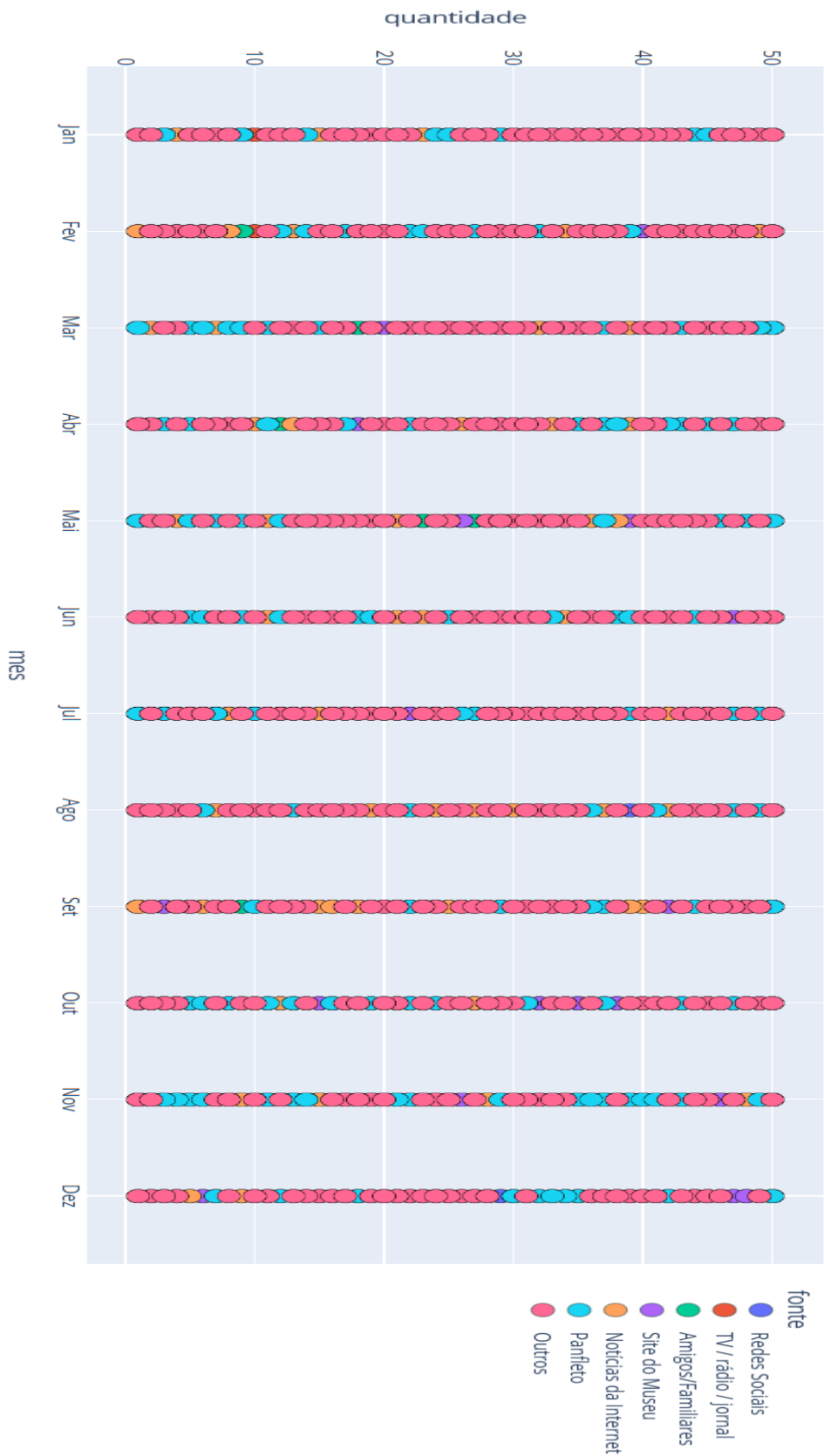


Segundo rascunho: Gr fico de dispers  o. Experimenta   o com eixos: vari vel categ rica no eixo x (fonte de informa   o), tipo qualitativo ordinal no eixo y (faixa et ria). Motivo de visita   o indicado por cor e escolaridade - no rascunho, transformada em faixas - por  rea e formato. Inten   o de fazer um gr fico ou de dispers   o ou de pontos.



spiral

APÊNDICE C - GRÁFICO TESTE (IAN)



APÊNDICE D - RESULTADO (IAN)



5. REFERÊNCIAS

#01 - Introdução ao Plotly | Plotly com Python. Disponível em:

<<https://www.youtube.com/pyplot>>.

ANACONDA. Anaconda. Disponível em: <<https://www.anaconda.com/>>.

C, P. A clean style for Plotly charts. Disponível em:

<<https://towardsdatascience.com/>>.

Crie seu Mapa Mental Online | Lucidchart. Disponível em:

<<https://www.lucidchart.com/pages/pt/exemplos/mapa-mental-online>>.

ECYY. Data Visualisation — How to Plot a Scatter Bubble Chart by Plotly. Disponível em: <<https://ecyy.medium>>.

ELIAS, D. Gerador de Referências ABNT [Atualizado para 2021]. Disponível em:

<<https://www.mybib.com/pt/ferramentas/gerador-referencias-abnt>>.

JUPYTER. Project Jupyter. Disponível em: <<https://jupyter.org/>>.

MATPLOTLIB. Matplotlib: Python plotting — Matplotlib 3.1.1 documentation.

Disponível em: <<https://matplotlib.org/>>.

NUMPY. Overview — NumPy v1.19 Manual. Disponível em:

<<https://numpy.org/doc/stable/>>.

Pandas documentation — pandas 1.0.1 documentation. Disponível em:

<<https://pandas.pydata.org/docs/>>.

PLOTLY. Plotly Python Graphing Library. Disponível em: <<https://plotly.com/python/>>.

PYTHON SOFTWARE FOUNDATION. Welcome to Python.org. Disponível em:

<<https://www.python.org/doc/>>.

SEABORN. seaborn: statistical data visualization — seaborn 0.9.0 documentation.

Disponível em: <<https://seaborn.pydata.org/>>.