

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
CENTRO DE LETRAS E ARTES
ESCOLA DE BELAS ARTES
COMUNICAÇÃO VISUAL DESIGN

LUANA CAROLINA DA SILVA

DESIGN DE INCERTEZA EM VISUALIZAÇÃO DE DADOS
Uma avaliação sobre os recursos de design de dados incertos

Rio de Janeiro

2021

LUANA CAROLINA DA SILVA

DESIGN DE INCERTEZA EM VISUALIZAÇÃO DE DADOS

Uma avaliação sobre os recursos de design na representação de dados incertos

Trabalho de Conclusão de Curso apresentado à Escola de Belas Artes da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do grau de bacharel em Comunicação Visual Design.

Orientadora: Doris Kosminsky

Rio de Janeiro

2021

CIP - Catalogação na Publicação

d586d da Silva, Luana Carolina
Design de incerteza em visualização de dados /
Luana Carolina da Silva. -- Rio de Janeiro, 2021.
84 f.

Orientadora: Doris Kosminsky.
Trabalho de conclusão de curso (graduação) -
Universidade Federal do Rio de Janeiro, Escola de
Belas Artes, Bacharel em Comunicação Visual Design,
2021.

1. Design da informação. 2. Visualização de dados.
3. Incerteza. 4. Representação. I. Kosminsky, Doris,
orient. II. Título.

Agradecimentos

Dedico meus agradecimentos à minha família, especialmente a minha mãe que sempre segurou minhas mãos, mas me deu asas para voar. E voando, pude atravessar a capital de São Paulo até o estado do Rio de Janeiro para fazer a faculdade pela qual batalhei para estar. Obrigada dona Marly, você é meu guia, minha luz nos dias escuros junto à Deus. A meus irmãos, que nunca descreditaram de mim e sempre me perguntavam se eu precisava de alguma coisa. Cada qual de valores distintos, mas profusamente bons de coração, engoliram soluços e me trouxeram sorrisos: Paulinha, Neto, Marcinho e Fellipe.

Pelas cotas, ao sistema Id Jovem, ao governo de 2012 à 2015 que permitiram um sonho realizado, estar em uma Universidade pública de qualidade. À balbúrdia e aos prêmios ganhos com colegas que sempre levarei no coração, sendo estes; Marcela Lobo, Vitor Alves, Elson Teixeira, Barbara Frast, Cecília Ramos, Livia Guimarães e Leonardo Cesar, sem vocês eu não passaria sequer pelo segundo módulo de Comunicação Visual Design.

À minha esposa, Gabrielle Vilamar, pelo ombro para chorar, pelos momentos de alegria e os debates sobre minha monografia. Sua companhia foi imensurável durante esse período e espero que tenhamos feito um bom trabalho sobre o caminho que abraçamos sem medo das projeções futuras.

Aos membros do Laboratório da Visualidade e Visualização, LabVis, projeto que me tornei membro junto a Lucas Barcellos, Doris Kosminsky e Cláudio Esperança. A vida acadêmica não é fácil, mas dar retorno sobre nossos estudos para a sociedade é um prazer que sempre procurei ter. Satisfação essa que começou com o ArteCiência Brasil, laboratório interdisciplinar que se desfez em 2018, mas que sou imensamente grata aos antigos participantes: João Silveira, Denise Lannes, Hera Telles, Julia Jasmin e Maria G. Moura.

Aos meus professores, espelhos de quem sonho um dia me tornar e grandes resistentes às calamidades de um sistema público. Juntos somos resistência. Sem alunos e sem professores a UFRJ não seria uma parcela do que é hoje.

Resumo

Este projeto de conclusão de curso consiste na elaboração de uma análise – no âmbito do *design* da informação – de marcadores visuais capazes de representar diferentes tipos de dados incertos. Incertezas não são comumente representadas graficamente nas visualizações de dados, o que pode levar a uma leitura limitada delas. Seguindo boas práticas de *design* para melhor comunicar aspectos desconhecidos, imprecisos e inconsistentes nos dados, apresentaremos três visualizações de dados que exploram métodos visuais para representar a incerteza, materializando a teoria abordada ao longo dos próximos capítulos. Portanto, este projeto objetiva provocar *designers* da informação e entusiastas da área para que reflitam sobre a incerteza em dados e desenvolvam representações visuais que possam produzir uma melhor experiência comunicacional em seus projetos.

Palavras-chave: design da informação; visualização de dados; incerteza; representação.

Abstract

This project consists of the elaboration of an analysis – in the scope of information design – of visual markers capable of representing different types of uncertain data. Uncertainties are not commonly represented graphically in data visualizations, which can lead to a limited reading of them. Following good design practices to better communicate unknown, inaccurate and inconsistent aspects in the data, we will present three data visualizations that explore visual methods to represent uncertainty, materializing the theory covered over the next chapters. Therefore, this project aims to provoke information designers and enthusiasts in the area, so that they reflect on the uncertainty in data and develop visual representations that can produce a better communicational experience in their projects.

Keywords: information design; data visualization; uncertainty; representation.

Lista de figuras

Figura 1	O mapa da campanha russa de Napoleão	14
Figura 2	Fragmento do mapa da Cólera	15
Figura 3	Escala categórica, ordinal e numérica	16
Figura 4	Sistema de coordenadas	18
Figura 5	Escala linear, percentual e logarítmica	19
Figura 6	Escalas temporal contínua e discreta	19
Figura 7	Marcadores visuais	20
Figura 8	Cor matiz	21
Figura 9	Cor valor	22
Figura 10	Posição	22
Figura 11	Rotação	23
Figura 12	Tamanho (área ou volume)	24
Figura 13	Chicletes e margem de erros	27
Figura 14	Intervalo de confiança	31
Figura 15	Anotações gráficas de propriedades distributivas	32
Figura 16	Barras de erro	33
Figura 17	Gráfico de caixa	33
Figura 18	Gráfico de violino	34
Figura 19	Gráfico de gradiente	34
Figura 20	Gráfico de conjunto	35
Figura 21	<i>Hypothetical Outcome Plots</i>	36
Figura 22	Gráfico de quantis	37
Figura 23	Codificação visual da incerteza	38
Figura 24	<i>Project Ukko</i>	39

Figura 25	<i>2016 General Election: Trump vs. Sanders</i>	40
Figura 26	Barras de erro	41
Figura 27	Método híbrido	42
Figura 28	Gráfico de intervalo e densidade de probabilidade	43
Figura 29	Linha de tendência de novas mortes por covid-19 por país	45
Figura 30	Linha de tendência de novas mortes por covid-19 no Brasil	47
Figura 31	Desempenho dos candidatos a reeleição nas pesquisas	49
Figura 32	Intenção de voto para a prefeitura de João Pessoa	51
Figura 33	Furacão Irma	53
Figura 34	<i>Income Mobility chart</i>	56
Figura 35	<i>When will it be sweater weather in St. Louis?</i>	58
Figura 36	<i>Flamengo are the favorites to win the 2019 Brazilian</i>	60
Figura 37	Rascunhos - Eleição para prefeito de São Paulo	66
Figura 38	Rascunhos - Prisão do deputado Daniel Silveira	67
Figura 39	Rascunhos - O barco perdido	68
Figura 40	Painel semântico	70
Figura 41	Paleta de cores	71
Figura 42	Simulador de daltonismo	72
Figura 43	Estudo tipográfico	73
Figura 44	Tipografia	73
Figura 45	Eleições para prefeito de São Paulo	74
Figura 46	Prisão do deputado Daniel Silveira	75
Figura 47	O barco perdido	75

Lista de tabelas

Tabela 1 *Semiology of Graphics*

17

Lista de quadros

Quadro 1	Tipos de incerteza	29
Quadro 2	Quadro de análises	62

Lista de abreviações e siglas

LGPD	Lei Geral de Proteção de Dados Pessoais
SRAG	Síndrome Respiratória Aguda Grave
PSDB	Partido da Social Democracia Brasileira
IBOPE	Instituto Brasileiro de Opinião Pública e Estatística
<i>SWDChallenge</i>	<i>Storytelling With Data Challenge</i>
PSL	Partido Social Liberal

Sumário

1 INTRODUÇÃO	12
2 VISUALIZAÇÃO DE DADOS	13
2.1 Categorização de dados	15
2.1.1 <i>Sistema de coordenadas</i>	17
2.1.2 <i>Escalas</i>	18
2.1.3 <i>Marcadores visuais</i>	19
2.1.4 <i>Contexto</i>	20
2.2 Relação entre escalas e marcadores visuais	20
3 CATEGORIZAÇÃO DE INCERTEZA	24
3.1 Tipos de incertezas	27
3.2 Mapeamento e representação de dados incertos	29
3.2.1 <i>Anotações gráficas da incerteza</i>	30
3.2.2 <i>Codificação visual da incerteza</i>	37
3.2.3 <i>Método híbrido</i>	41
4 ANÁLISE DE VISUALIZAÇÕES COM DADOS INCERTOS	43
4.1 Saúde	44
4.2 Eleições	48
4.3 Meteorologia	52
4.4 Outros	55
4.5 Conclusões	61
5 DESENVOLVIMENTO	63
5.1 Visualizações incertas	64
5.1.1 Eleições para prefeitura de São Paulo	65
5.1.2 Prisão do deputado Daniel Silveira	66
5.1.3 O barco perdido	68
5.2 Identidade visual	69
5.3 Resultado final	73
6 CONSIDERAÇÕES FINAIS	76
7 REFERÊNCIAS	77
8 ANEXOS	81

1 INTRODUÇÃO

Este trabalho de conclusão de curso tem como principal objetivo analisar os diferentes métodos de representação da incerteza na visualização de dados. Com isso, pretendemos provocar *designers* da informação a pensarem com mais afinco nas escolhas visuais exclusivas para dados incertos, partindo do princípio que a maioria das visualizações de dados possuem pouca ou nenhuma representação da incerteza.

A visualização de dados, segundo Card, Mackinlay e Shneiderman (1999) consiste em representações digitais visuais de dados como forma de ampliar a cognição. Já a incerteza será definida neste projeto em sua forma quantificável, que pode ser representada visualmente.

Iremos abordar o mapeamento visual da incerteza de acordo com as representações tradicionais da estatística (composição de barras de erro, por exemplo), as alterações ou inserções de marcadores visuais (implementação das técnicas de gradiente, diminuição de opacidade, etc.) e a mesclagem de ambos os métodos anteriores para uma abordagem híbrida.

Para isso, exploraremos as distintas classes de incerteza (SPIEGELHALTER, 2017), os marcadores visuais propostos pela visualização de dados (YAU, 2013 apud MACKINLAY, 1986) e, por fim, os recursos visuais específicos para a incerteza (PADILLA, 2020). De tal forma, será possível parametrizar uma análise de referências com essas teorias e, assim, compor o produto final que terá a função de complementar o que será apresentado na base teórica.

O produto final deste projeto será a composição de três visualizações de dados individuais e baseadas em situações reais. As três visualizações produzidas terão como objetivo primário a materialização da monografia e irão se utilizar exclusivamente dos recursos de *design* estudados para a representação da incerteza. Em última instância, pretendemos detalhar o processo de desenvolvimento deste trabalho nos capítulos seguintes.

2 VISUALIZAÇÃO DE DADOS

O presente capítulo apresenta uma definição sobre categorização de dados e as relações entre escalas e marcadores visuais, baseado na leitura de Nathan Yau (2013). As

correlações feitas a seguir são importantes para este projeto por serem pilares impactados e modificados de acordo com o volume de incerteza presente nos dados.

A visualização de dados consiste no uso de representações digitais visuais de dados para ampliar a cognição (CARD, MACKINLAY E SHNEIDERMAN, 1999). Um dos principais objetivos da visualização é a obtenção de *insights* sobre os padrões exibidos, podendo apresentar conceitos complexos e/ou validar tendências. Ao produzirmos uma visualização de dados devemos ter em mente o objetivo desse projeto, já que as visualizações são uma abstração do mundo real que, assim como o que acontece com a escrita, contam histórias e direcionam caminhos.

Uma boa visualização, segundo Yau (2013), é uma representação de dados que ajuda o leitor a ver sob outra ótica detalhes que poderiam passar despercebidos caso estivessem em uma planilha. Interagir com uma ótima visualização pode entusiasmar o leitor a conseguir obter *insights* ou verificar conteúdos que não seriam identificáveis de outra forma.

O emprego de gráficos na representação e análise de dados não se trata de algo novo, como se pode observar a partir do diagrama de Charles Joseph Minard (Figura 1), engenheiro civil francês que construiu um mapa da campanha de Napoleão em 1812, demonstrando a perda dramática de seu exército no avanço e retirada em direção a Moscou. Minard exhibe os seguintes dados em muitas dimensões: o tamanho da tropa de Napoleão; a distância percorrida; temperatura; latitude e longitude; direção da viagem e localização relativa a datas. A largura da linha em bege ilustra o tamanho do exército de Napoleão durante seu avanço pelas fronteiras, e a linha em preto a sua retirada. Por meio da visualização e, especificamente, da variação da largura da linha, é possível sentir o horror da guerra e imaginar a angústia dos soldados.

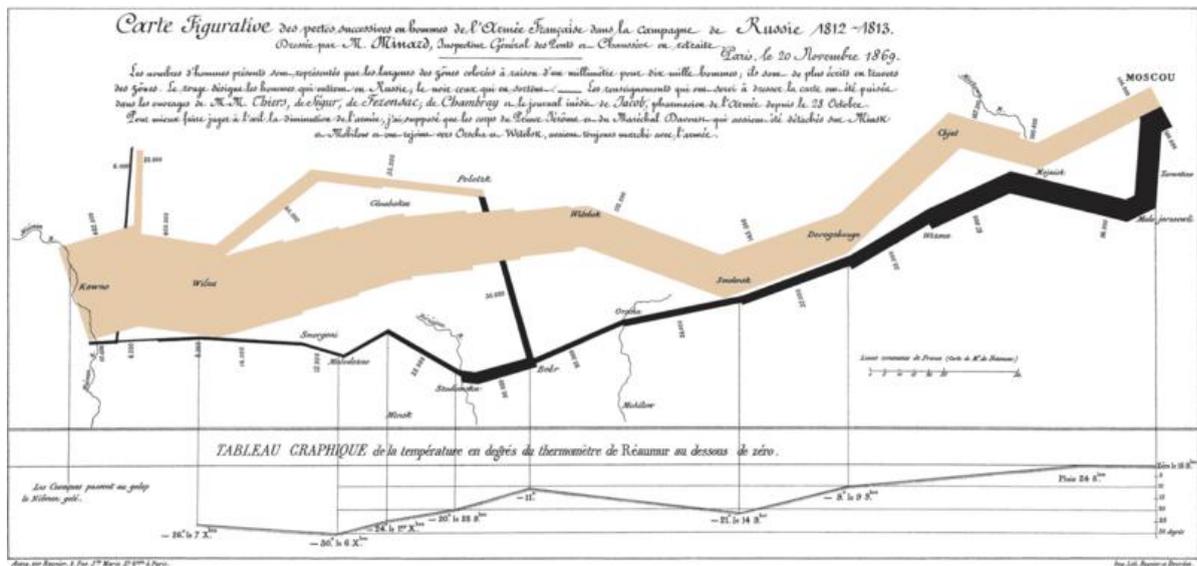


Figura 1: O mapa da campanha russa de Napoleão. Fonte: Charles Minard. Domínio público, 2020.

Outro exemplo clássico de representação visual de dados é encontrado no mapa da cólera (Figura 2), no qual o médico John Snow investigou a origem da doença na cidade de Londres, localizando o epicentro da epidemia em 1854. Na época, a ciência defendia a teoria miasmática, ou seja, o mau odor da cidade seria responsável pela cólera, causando vômitos e diarreia nos infectados. Snow procurou saber, além dos sintomas, onde e como as pessoas estavam se infectando e encontrou o motivo do problema na água potável das bombas públicas. No mapa cartográfico do bairro de Soho, o médico indicou com linhas pretas o número de pessoas contaminadas e com círculos, a localização das bombas de água. Snow tornou-se uma das maiores referências para a epidemiologia moderna ao provar, com o seu mapa, como ocorria a infecção pela cólera.



Figura 2: Fragmento do mapa da Cólera. Fonte: John Snow. Domínio público, 2020.

A partir do final do século XX e início do XXI, a visualização de dados avançou com o auxílio da informática, de forma a incluir o grande volume de dados produzido na sociedade contemporânea. Nos próximos subcapítulos, vamos abordar alguns dos fundamentos da representação de dados que servirão de base para a compreensão da incerteza no *design* de visualizações.

2.1 Categorização de dados

No *design* de uma visualização, os dados podem ser codificados de diversas maneiras, de acordo com o objetivo da visualização (ou sua função como, por exemplo, comparação, tendência ao longo do tempo, etc.), seu tipo (categórico, numérico ou ordinal), os elementos gráficos escolhidos (também conhecidos como marcas, tais como barras, linhas e pontos) e seus atributos visuais (posição, forma, tamanho e/ou cor). Apresentaremos, a seguir, dois autores que detalham o processo de representação de dados, de forma a refletirmos sobre as escolhas visuais no *design* da informação.

Os dados se categorizam por meio de três variáveis principais: dados categóricos (Figura 3), mais conhecido como variável qualitativa ou categórica, ou seja, não possuem

valores quantitativos. Ela pode indicar características dos dados, como por exemplo tipos de redes sociais ou diferentes aplicativos de videoconferência. Os dados ordinais constituem uma categoria na qual a ordem importa como, por exemplo, os estágios de uma doença (inicial, intermediária, terminal) ou de qualidade (bom, médio e ruim). E dados de categoria numérica, que é composta por dados que se referem a números, independente se positivos ou negativos (-2, -1, 0, 1, 2).

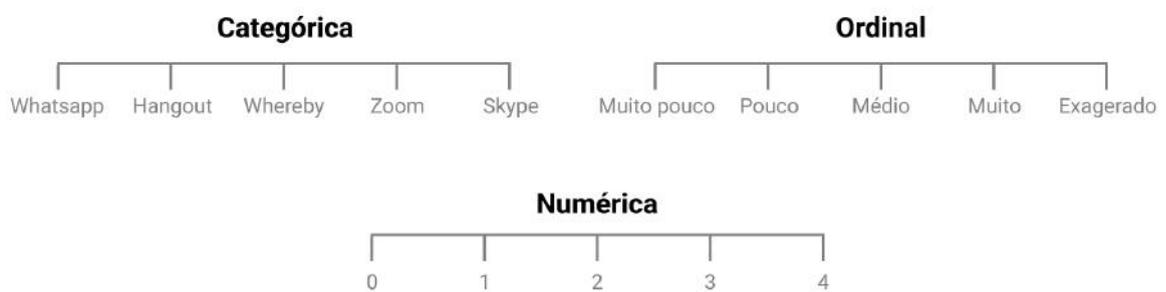


Figura 3: Escala categórica, ordinal e numérica. Fonte: O autor, 2020 apud Yau, 2013.

Quanto às escolhas visuais para a representação de dados, iniciaremos pela pesquisa do cartógrafo francês, Jacques Bertin (2010), que desenvolveu a semiologia gráfica considerando a maneira como podemos compor os dados. Bertin menciona três itens complementares: (a) a variável dos planos, (b) as variáveis visuais e (c) as modalidades de implementação dos planos. As variáveis do plano consideram a posição dos eixos x e do eixo y de um sistema cartesiano; as variáveis visuais são seis: orientação, tamanho (podendo ser área ou comprimento), forma, cor (matiz), granulação (textura) e valor (intensidade/luminância); e, por fim, as modalidades de implementação dos planos são três: ponto, linha e área. Todos esses elementos atuam para compor uma visualização. Na Tabela 1 vemos a relação dos modos de implementação e as diversas variáveis visuais.

		LEVEL OF THE RETINAL VARIABLES			
PLANAR DIMENSIONS		ASSOCIATION ≡ The marks can be perceived as SIMILAR	SELECTION ≠ The marks are perceived as DIFFERENT, forming families	ORDER ○ The marks are perceived as ORDERED	QUANTITY Q The marks are perceived as PROPORTIONAL to each other
SIZE					
VALUE					
TEXTURE					
COLOR					
ORIENTATION					
SHAPE					
					Conventions leading to the ELEMENTARY READING LEVEL

Tabela 1: *Semiology of Graphics*. Fonte: Jacques Bertin. *Semiology of Graphics: diagrams, networks, maps*, 2010.

Para Nathan Yau (2013) a visualização de dados se forma por quatro categorias complementares de componentes: (a) sistema de coordenadas, (b) escalas, (c) marcadores visuais (ou pistas visuais) e; (d) contexto. Qualquer visualização, independente de seu nível de abstração, pode ser construída baseada nesses quatro componentes. Eles são como ingredientes para uma receita e devem ser trabalhados de acordo com escolhas coerentes sobre a informação que se deseja representar visualmente. A seguir, iremos descrever a categorização desenvolvida por Yau (2013).

2.1.1 Sistema de coordenadas

O **sistema de coordenadas** (Figura 4) trata-se de um espaço estruturado no qual os elementos visuais são distribuídos de acordo com certas regras. Existem diversos sistemas, dos quais destacamos o cartesiano, o polar e o geográfico. O sistema cartesiano, coordenado pelo eixo x-y, é conhecido pela ampla utilização de gráficos de linha e de barra; o sistema polar é utilizado majoritariamente em gráficos de pizza e *donut* nos quais as coordenadas são medidas pelo raio r e ângulo θ ; e o sistema geográfico é aquele cujas localidades do mundo físico são identificadas de acordo com a latitude e a longitude. Como o planeta Terra é redondo e a representação do seu espaço é realizada sobre uma superfície bidimensional, existem diversas possibilidades de projeção. A mais utilizada é a de Mercator e foi criada no século XVI.

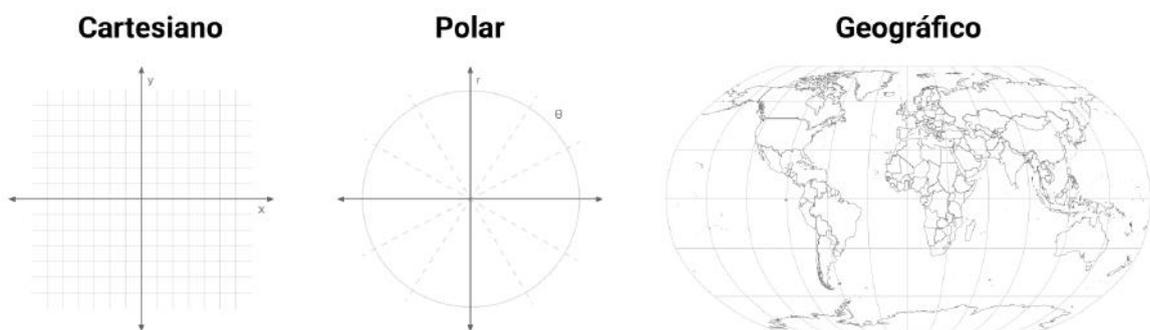


Figura 4: Sistemas de coordenadas. Fonte: O autor, 2020 apud Yau, 2013.

2.1.2 Escalas

Enquanto o sistema de coordenadas indica as dimensões e espaço disponível para a visualização, a escala indica a colocação dos dados. As escalas podem ser agrupadas em duas categorias: numérica e temporal.

A escala numérica (Figura 5) pode ser linear, onde os dados são mapeados uniformemente; percentual, que representa partes de um todo; e a escala logarítmica, geralmente usada quando há uma grande exponencialidade nos valores, alterando a representação para que ela seja mais facilmente visualizada.

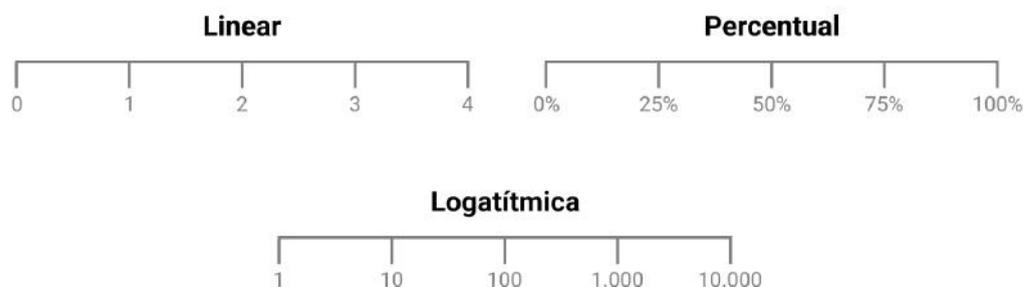


Figura 5: Escala linear, percentual e logarítmica. Fonte: O autor, 2020 apud Yau, 2013.

A escala temporal (Figura 6), como seu nome diz, representa a passagem de tempo e pode ser: contínua ou discreta. Na temporal contínua, a unidade de medida pode ser infinitamente dividida em unidades menores, como hora, minuto, segundo, etc. A escala temporal discreta atua de modo que os dados são escalonáveis por categorias, como dias de uma semana.



Figura 6: Escalas temporal contínua e discreta. Fonte: O autor, 2020 apud Yau, 2013.

2.1.3 Marcadores visuais

Os **marcadores visuais** (YAU, 2013 apud MACKINLAY, 1986), também chamados de pistas visuais, são, em resumo, a transformação dos dados crus em uma visualização, podendo ser alternados de acordo com a finalidade de seu projeto. Os marcadores (Figura 7) são: posição (ditando onde a marca estará inserida em um espaço); comprimento (o quão longa será esta forma); ângulo (onde as formas podem ter uma rotação entre vetores); direção (para onde essas formas irão, caso vão a algum lugar como uma linha ou uma seta); área (o quanto de um espaço 2-D a forma ocupa); volume (o quanto de um espaço 3-D a forma ocupa); valor (saturação de determinada matiz); e cor (qual matiz representará cada conjunto de dados).



Figura 7: Marcadores visuais. Fonte: Nathan Yau, 2013 apud Mackinlay, 1986)

2.1.4 Contexto

Por último, mas não menos importante, temos como categoria o **contexto**, que são as informações que nos auxiliam a compreender melhor os nossos dados. Devemos deixar que os leitores olhem para os dados e respondam *quem, o quê, quando, onde e por que* os dados estão da forma que estão. O contexto pode tornar os dados mais claros ao leitor, fazendo-o lembrar sobre o que se trata a visualização alguns dias depois de tê-la visto. A maneira mais simples de se aplicar um contexto é por meio de títulos descritivos, legendas e *tooltips*. Ou seja, recursos que tragam mais informações sobre os dados, dando a possibilidade do leitor poder explorar entre os números e as palavras, além das formas.

2.2 Relação entre escalas e marcadores visuais

Ao visualizar os dados, é necessário representá-los com uma combinação de pistas visuais. Ou seja, pensar sobre posicionamento, cores e escalas que dependem dos valores representados. As formas em diferentes posicionamentos podem representar valores maiores ou menores, assim como formas mais claras que outras ou mesmo maiores em tamanho. Visto que é simples pensar em geometria e cores, pois nossos cérebros estão programados para encontrar padrões e desvios, variando entre números e pistas visuais, conseguimos obter um bom entendimento de um gráfico. Mas qual relação podemos obter entre marcadores visuais e escalas para se representar um conjunto de dados? A seguir, apresentaremos a relação entre marcadores visuais e escalas (ou categorias de dados visuais), baseado na leitura do autor Nathan Yau (2013 apud MACKINLAY, 1986);

Iniciaremos a relação pelo marcador visual **cor** (Figura 8), utilizado principalmente em marcas como codificação para categorias de dados. É necessário que tenhamos cuidado ao utilizar o matiz como única forma de codificação, pois ele pode ser ineficiente para quem possui problemas de visão como daltonismo. Uma solução prática seria inseri-la em conjunto com uma escala categórica, como letras e/ou palavras.

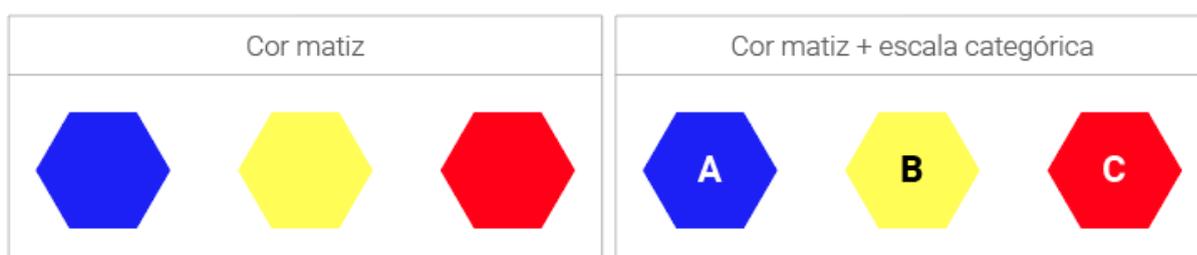


Figura 8: Cor matiz. Fonte: O autor, 2020.

O marcador **valor**, ou a intensidade de uma cor, é utilizado como codificação para dados que mudam de quantidade. A saturação pode indicar também uma escala categórica ordinal, como de bom, razoável e ruim (Figura 9). Esse marcador representa a variação de um único matiz, mas é necessário se ter cuidado porque, caso utilizado com muita nuance, pode passar despercebida aos olhos do leitor. Por exemplo, se tivéssemos vinte linhas de um

conjunto de dados que quiséssemos representar apenas pelo valor de um matiz, seria quase impossível observar com precisão a diferença entre a cor da marca 1 com a cor da marca 2 caso seus valores fossem pouco alterados.

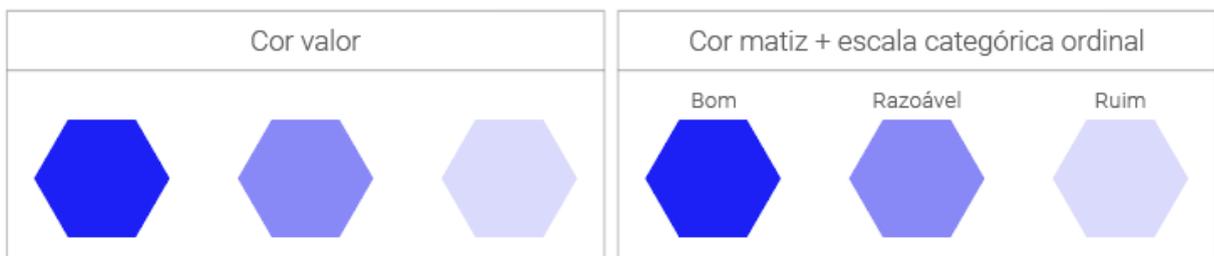


Figura 9: Cor valor. Fonte: O autor, 2020.

A **posição** (Figura 10) de um dado pode ser determinante para indicar seu valor, mas também pode indicar localidade caso estejamos em um sistema cartográfico. Em suma, a posição se alinha com diversos sistemas de coordenadas e é um dos marcadores visuais com maior pregnância, sendo o mais utilizado em visualizações de dados. Comumente podemos vê-lo em conjunto com as escalas numérica, categórica e temporal.

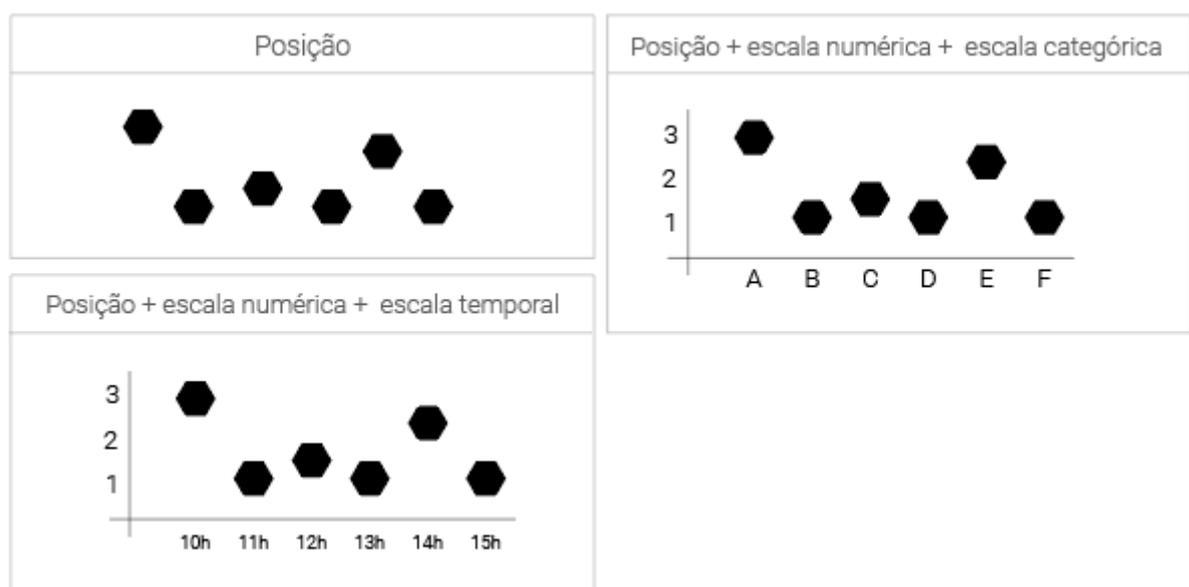


Figura 10: Posição. Fonte: O autor, 2020.

Com **ângulo** podemos falar também de **rotação**. Sendo o primeiro, uma codificação comum para gráficos de pizza, e o segundo um indicativo de escala categórica ou temporal. Dependendo da visualização, o ângulo pode contar como uma boa solução visual. Neste caso, a forma pode ser rotacionada em torno de si mesma, ou abrir vetores no raio de um círculo como apresentamos na figura 11.



Figura 11: Rotação. Fonte: O autor, 2020.

Com **comprimento, área e volume** (Figura 12) podemos estabelecer uma relação pelo formato do marcador visual se alternar de acordo com o conjunto de dados apresentado, diminuindo ou aumentando nos três casos. Iremos chamar este marcador de **tamanho** para melhor associar a escalas, visto que essa pista visual pode ser percebida com mais facilidade em gráficos do nosso cotidiano. O tamanho de uma forma pode causar impacto ao leitor, assim como seu comprimento e espessura, dependendo se estamos projetando um gráfico de barras ou um fluxo de linhas. Normalmente o tamanho vem acompanhado de uma legenda, ou seja, uma escala numérica, temporal e/ou categórica para que possamos identificar a área ou volume com mais facilidade. Essa é uma boa prática pois não podemos notar uma grande diferença entre áreas ou volumes, devido ao cálculo envolvido. Em uma escala percentual, representando partes de um todo, ficaria mais simples de fazermos comparações, já que temos noção do todo.

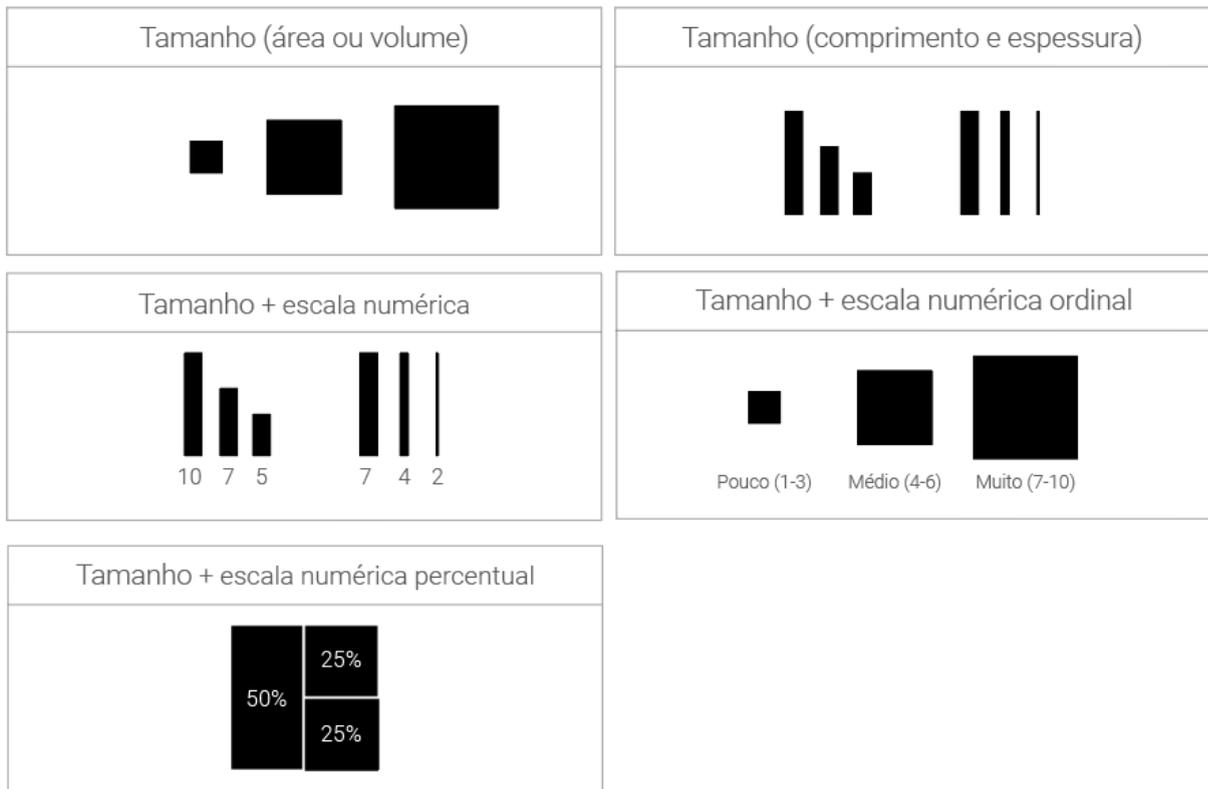


Figura 12: Tamanho (área ou volume). Fonte: O autor, 2020.

Contudo, as relações são infinitas; podemos utilizar mais de um marcador e mais de uma escala. Os exemplos apresentados acima foram algumas das relações que poderíamos atribuir aos dados, mas contaremos também com o contexto e o sistema de coordenadas para que a visualização esteja coerente e com o devido valor empregado tanto aos números, quanto ao seu aspecto gráfico.

3 CATEGORIZAÇÃO DE INCERTEZA

Neste capítulo, abordaremos a categorização de incertezas segundo acadêmicos da área que procuraram entender a origem de um dado incerto. Também iremos detalhar a representação de incertezas na visualização de dados, com o objetivo de tornar clara a implementação visual de dados incertos com recursos de *design*.

O termo incerteza é empregado em diversas áreas do conhecimento, tais como estatística, psicologia e ciências sociais. No presente projeto, o nosso foco será no campo do *design*, especificamente na relação entre incerteza e as diferentes técnicas para visualizar dados. Empregaremos o termo incerteza para fazer referência à incerteza quantificada, que pode ser visualizada mais comumente em uma distribuição de probabilidade. A representação visual da incerteza é comum em certos contextos, como por exemplo em gráficos de pesquisa eleitoral. No entanto, em outros contextos, predominantemente em visualizações de acompanhamento diário de eventos, como o número de novos casos de Covid-19, a incerteza tende a ser pouco ou mesmo não representada.

A maioria das visualizações foi projetada na suposição de que os dados representados visualmente são livres de incerteza. No entanto, raramente é o caso. Visualizar a incerteza é essencial se quisermos melhorar a forma como as pessoas entendem as estatísticas.

(YAU, 2018)

Diante da busca pela transparência relacionada às novas políticas públicas de dados (LGPD, 2020), a incerteza já deveria ser considerada como parte inerente aos dados. Nesse contexto, cabe ao *designer* desenvolver novas técnicas, marcadores e recursos para evidenciar a parte dos dados onde há incerteza, de forma a garantir a compreensão do que é apresentado. Para isso é necessário que se faça uma análise de correlação entre os tipos de incerteza e as técnicas visuais de representá-las, promovendo assim, mais assertividade e conforto ao invés de confusão.

A incerteza, assim como a probabilidade, tratam de conceitos inerentes ao ser humano (DE LAPLACE, 1812). Essa ideia é explicada com o exemplo da espera diária por um ônibus em seu ponto. Sabemos mais ou menos o horário em que o ônibus pode passar, mas não temos tanta certeza, porque podem ocorrer atrasos no tráfego, problemas no motor, ou qualquer dificuldade no trajeto anterior (HULLMAN, 2016). Entendemos que o ônibus pode possuir uma variedade de horários, mas temos como uma espécie de instinto a certeza que o ônibus em algum momento virá, mas sem uma medida temporal exata.

O cartógrafo Mark Monmonier (1995) lista em seu livro, *Mapping it Out*, as habilidades que todos deveríamos desenvolver para facilitar a transmissão de mensagens,

tornando mais fluida a comunicação. Esta lista contém os seguintes itens: (a) Alfabetização - Habilidade em ler e escrever; (b) Articulação - Capacidade de se articular utilizando palavras faladas; (c) Raciocínio numérico - Capacidade de lidar não necessariamente com estatísticas, mas com raciocínio quantitativo: pensar com números; (d) Graficidade - Ser capaz de interpretar e criar gráficos. As quatro habilidades citadas são importantes para qualquer projeto de visualização de dados, mas para a compreensão da representação da incerteza, destacamos o raciocínio numérico, que possibilita compreender a lógica dos dados e suas oscilações. A incerteza está relacionada, na maioria das vezes, a um dado quantitativo que pode variar ao depender de uma série de fatores que mencionaremos na próxima seção. Portanto, a habilidade de conseguir lidar e pensar com os números é essencial, mas não somente ela. É importante termos cuidado para que a mensagem não seja confusa e/ou mal interpretada, o que pode estar relacionado às suas características visuais.

Exemplos de conjuntos de dados que lidam com a incerteza são previsões sobre o futuro (ex.: gráficos eleitorais), fenômenos naturais (ex.: gráficos de furacões, meteorologia), mercado financeiro (ex.: bolsa de valores), dentre outros. Podemos observar na figura 13, logo abaixo, a composição de uma amostra cujos dados são representados por chicletes coloridos. A jarra ilustra uma população e, em seguida, vemos que podemos adivinhar o que está na jarra baseado no tamanho de uma amostra. Quanto mais chicletes retirarmos da jarra para a amostra, menor sua margem de erro. Já com menos chicletes, maior nossa margem de erro, o que pode ser aplicado para outros estudos e projetos. Portanto, quanto maior a variedade dessa população presente na amostra (cores dos chicletes), menor será a incerteza.



Figura 13: Chicletes e margem de erros. Tradução do autor. Fonte: Nathan Yau. Data Points, 2013.

Uma pesquisa de opinião realizada pela Gallup em 2012 nos Estados Unidos ilustra esta correlação. Essa pesquisa mostrou que 48% da população estadunidense desaprovava o trabalho de Barack Obama. No entanto, a margem de erro era de 3%. Com essa margem aplicada, podemos ver mais da metade da população desaprovando o trabalho de Obama (51%) e/ou menos da metade (45%). Essa estimativa também costuma acontecer com pesquisas realizadas durante o período eleitoral, quando se apresentam dados numéricos com uma margem de erro que pode alterar o resultado final da eleição. Logo, este breve conhecimento estatístico nos dá uma base para compreender o conceito de uma amostra e a importância de conhecer a origem dos dados, considerando a incerteza e seu contexto.

3.1 Tipos de incertezas

Assim como existem diversas possibilidades de representação visual de dados, também existem diferentes tipos de incerteza que demandam técnicas variadas para representá-las.

Alguns acadêmicos da estatística podem explicar a origem de uma incerteza, e é neste subcapítulo que abordaremos três pesquisadores que consideramos mais relevantes a este projeto. Segundo o geógrafo Alan MacEachren (2005), a divisão dos tipos de incerteza se dá de três formas: (a) de acordo com a origem/proveniência dos dados, em relação aos erros de

amostragem que podem afetar a qualidade do que está sendo apresentado; (b) a origem dos dados, que pode indicar incerteza dependendo da informação disponível sobre o processo de coleta dos dados, e por fim, (c) os erros de registro no banco de dados ou a falta de acesso direto às suas fontes podem sinalizar pouca qualidade.

Como segunda referência, temos o estatístico Simon French (2015), que mapeou uma lista com os tipos de incerteza observados na probabilidade estatística quando se lida com uma quantidade considerável de dados: (a) Incertezas estocásticas (aleatoriedade física), (b) incertezas epistemológicas (falta de conhecimento científico), (c) incertezas de desfecho (quando o desfecho é mal definido), (d) incertezas de julgamento (por exemplo, a configuração de valores de parâmetro em códigos), (e) incertezas computacionais (cálculos imprecisos) e (f) erros de modelagem (por melhor que seja o modelo, ele não se ajustará perfeitamente ao mundo real ou, se parecer que sim, provavelmente terá pouco poder preditivo). Existem outras incertezas relacionadas a ambiguidades, sendo elas: (g) julgamentos de valor parcialmente formados; e (h) incertezas sociais e éticas (por exemplo, o modo como as recomendações de especialistas são formuladas e implementadas na sociedade, o valor ético de uma decisão e suas consequências).

Por fim, o estatístico David Spiegelhalter (2017) utiliza três dos conceitos anteriores para compor suas categorias, sendo elas: (a) aleatória (aleatoriedade irreduzível inerente a um processo), (b) epistêmica (incerteza por falta de conhecimento que teoricamente poderia ser reduzida com mais informações) e (c) incerteza ontológica (incerteza sobre a precisão com que a modelagem descreve a realidade, que só pode ser descrita subjetivamente).

Ao mapear os tipos de incerteza por autor em um quadro (Quadro 1), na qual cada cor indica semelhança por aproximação de conceito, observamos que muitos dos conceitos são correlatos ou se repetem. Por se tratar do estudo mais recente e por melhor resumir os tipos de dados incertos dentro de três abrangentes tópicos, utilizaremos Spiegelhalter (2017) como referência. Desse modo, a maior parte das incertezas que aparecem neste trabalho são classificadas como ontológica, epistêmica ou aleatória.

Tipos de incerteza		
Simon French (2015)	Alan MacEachren (2005)	David Spiegelhalter (2017)
Estocástica	Qualidade dos dados	Aleatória
Epistêmica	Credibilidade / Proveniência	Epistêmica
Erros de modelagem	Erro de amostragem	Ontológica
De julgamento		
Computacional		
De desfecho		
Julgamento de valor		
Sociais e éticas		

Quadro 1: Tipos de incerteza. Fonte: O autor, 2020.

Quando tratamos de incertezas sociais e éticas, como cita French (2015) em seus tipos de incerteza, estamos nos referindo à manipulação dos dados por uma pessoa (ou pessoas), fazendo com que os gráficos distorçam a informação por um conflito de interesses (CAIRO, 2019). No entanto, por mais que esse tema se relacione à incerteza, consideramos que esse assunto extrapola a proposta deste trabalho, além de não ser possível de ser solucionado pela escolha de recursos visuais. Considerando que os dados são manipulados conscientemente em sua origem ou em sua “tradução visual”, a proposta deste projeto é de categorizar dados incertos ficcionais que representam situações reais, desenvolvendo boas práticas para representá-los, sem considerar possíveis manipulações e/ou má fé.

3.2 Mapeamento e representação de dados incertos

Neste item apresentaremos algumas representações visuais que têm sido utilizadas para codificar a incerteza no âmbito do *design*. Utilizaremos como referência de categorização de marcadores visuais o formato que Padilla (2020) propôs de duas amplas vertentes exclusivas para a incerteza: as anotações gráficas de propriedades distributivas e a codificação

visual da incerteza. Dada essa separação explicaremos brevemente sobre como cada recurso é percebido para que possamos ter em mente a abrangência de um dado incerto sendo visualizado e o impacto que este pode ter. O objetivo deste subcapítulo é fornecer a infraestrutura teórica necessária para que se possa avaliar criticamente técnicas de visualização, tanto no contexto de seu público quanto às restrições de *design* (PADILLA, 2020).

3.2.1 Anotações gráficas da incerteza

As anotações gráficas podem ser utilizadas para representar as propriedades distributivas de dados incertos nas visualizações de dados, como a inserção de barras de erro e intervalos de confiança (PADILLA, 2020). Ou seja, em sua maioria, trata-se da inserção de métodos tradicionais de se representar a incerteza em um gráfico. Chamaremos as anotações gráficas de propriedades distributivas como **anotações gráficas da incerteza**, ou somente anotações gráficas, com o objetivo de tornar mais claro esse termo (propriedade distributiva) da estatística e trazê-lo para o campo do *design* da informação.

A propriedade distributiva é a característica gráfica que mostra o nível de incerteza dos dados na visualização. Para a estatística, esse termo faz parte do processo de se montar uma amostra ou intervalo de confiança, já que os dados incertos surgem como o valor x de uma variável aleatória, que foge do padrão. Em resumo, esse valor x será inserido em uma equação para obtermos uma propriedade distributiva, e assim formaremos um intervalo de números confiáveis diante da incerteza. Portanto, as **anotações gráficas da incerteza** são métodos visuais de inserção de um intervalo de números confiáveis (surgidos em uma amostra, por exemplo) na visualização.

Algumas definições consideradas importantes:

Amostra: Uma amostra é um conjunto de dados coletados de determinada população por um procedimento definido. Tipicamente, a população é muito grande, portanto fazer um censo ou uma enumeração completa de todos os valores na população é pouco prático ou impossível. A amostra geralmente representa um subconjunto de tamanho manejável. Amostras são coletadas e estatísticas são calculadas a partir das amostras, de modo que se possam fazer inferências ou extrapolações da amostra à população;

Média amostral: A média populacional representa a média de todos os indivíduos ou objetos que estão sendo estudados. Mas geralmente, nem todos os indivíduos podem ser medidos. Em geral, somente uma amostra de todos os indivíduos está disponível para nós e a média baseada nesta amostra, é usada para estimar a média populacional (PORTAL ACTION, 2021);

Intervalo de confiança: É uma estimativa de um intervalo utilizado na estatística, que contém um parâmetro populacional. Esse parâmetro desconhecido é encontrado através de um modelo de amostra calculado a partir dos dados recolhidos.

Exemplo: a média de uma amostra recolhida “a” pode ou não coincidir com a verdadeira média populacional “b”. Para isso, é possível considerar um intervalo de médias amostrais onde essa média populacional possa estar contida. Quanto maior esse intervalo, maior a probabilidade de isso ocorrer.

O intervalo de confiança é expresso em porcentagem (Figura 14), denominadas por nível de confiança, sendo 90%, 95% e 99% as mais indicadas. Na imagem abaixo, por exemplo, temos um intervalo de confiança de 90% entre seus limites superior e inferior (a e -a) (SIGNIFICADOS, 2021).

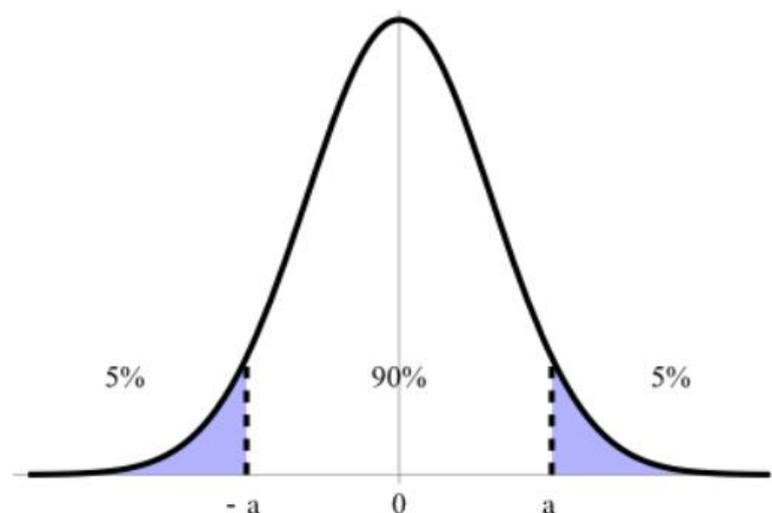


Figura 14: Intervalo de confiança. Fonte: Significados, acesso em janeiro de 2021.

Diversas visualizações utilizam as codificações apresentadas na figura abaixo para representar a incerteza. Algumas destas utilizam elementos provenientes da estatística, como é o caso da inserção de barras de erros ou gráficos de caixa, que acrescentam elementos gráficos à visualização para representar a incerteza. Já nas demais visualizações, os dados incertos são codificados nos próprios marcadores visuais; citamos como exemplo o gráfico de gradiente, onde as cores menos opacas indicam maior grau de incerteza.

A seguir, iremos listar uma sequência de anotações gráficas que representam a incerteza na visualização de dados.

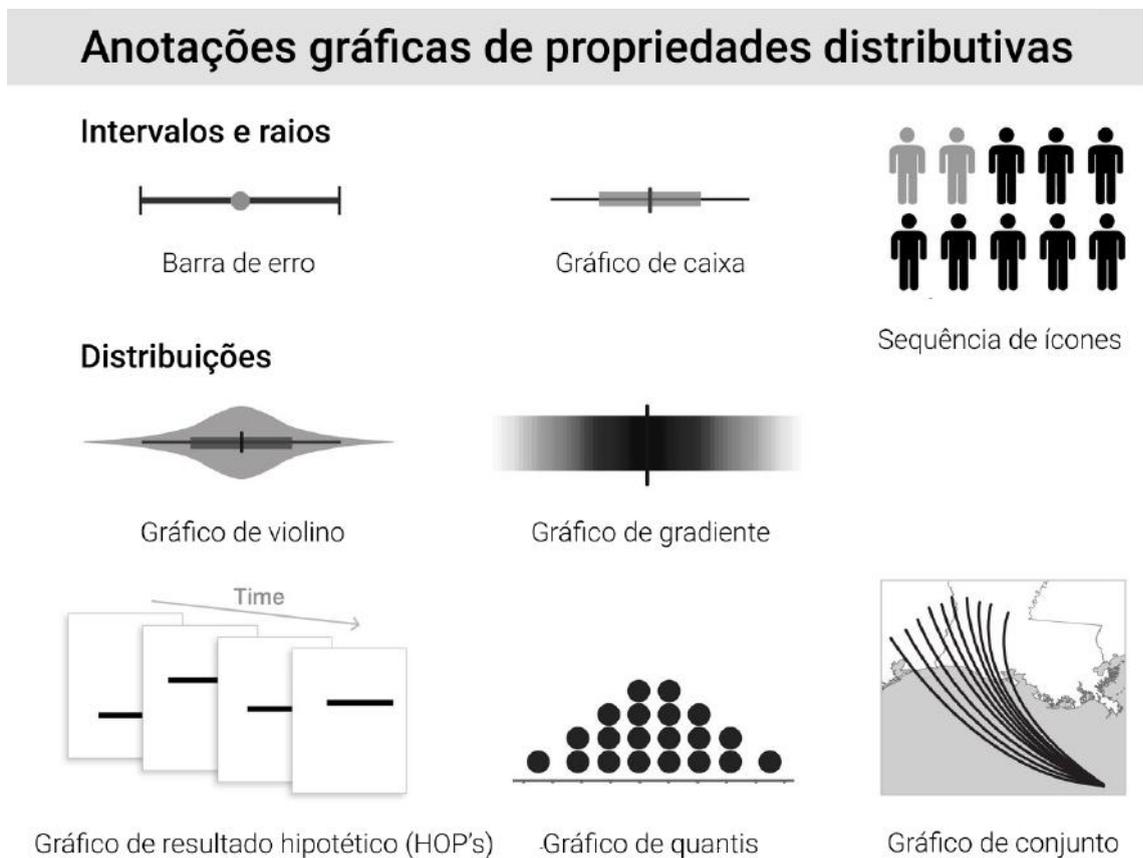


Figura 15: Anotações gráficas de propriedades distributivas. Tradução do autor. Fonte: Lace Padilla, Matthew Kay e Jessica Hullman. *Uncertainty Visualization*, 2020.

As barras de erro (Figura 16) e gráficos de caixa (Figura 17) possuem certa semelhança por utilizarem marcadores para representar diretamente a incerteza. A diferença que pode ser vista está presente no segundo item, que nos traz os quartis (que representa qualquer um dos valores que divide o conjunto ordenado de dados em quatro partes iguais, e assim cada parte representa $\frac{1}{4}$ da amostra realizada) representados nas bases da caixa, dentro da barra de distribuição. Os quartis podem trazer mais clareza de onde a incerteza está mais “concentrada”, mas ainda não trazem maior detalhamentos sobre os dados representados.

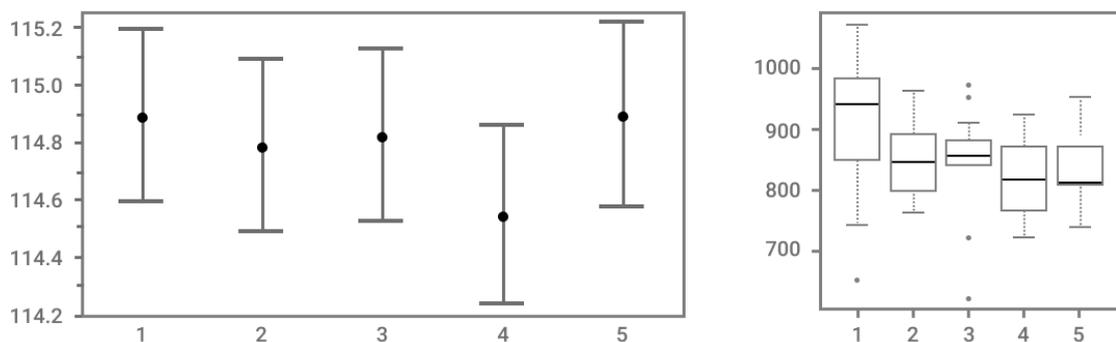


Figura 16 e 17: Barras de erro e gráfico de caixa. Fonte: O autor, 2020.

Os gráficos de violino (Figura 18) operam de maneira semelhante ao intervalo de confiança, onde a maior probabilidade se encontra na parte mais larga do violino. Já os gráficos de gradiente (Figura 19) funcionam de maneira que, onde se está mais opaco, maior a quantidade de certeza. A transparência é uma técnica bastante utilizada para comunicar incerteza, já que tendemos a acreditar que o que está mais transparente, pode estar mais distante ou incerto demais para estar em sua forma "completa". A linha que acompanha a barra gradiente demarca o meio desta e pode estar em diferentes formatos acompanhando um eixo x e y . Sendo tanto o gradiente, quanto o gráfico de violino, distribuições inerentes a visualização pela modificação dos marcadores visuais.

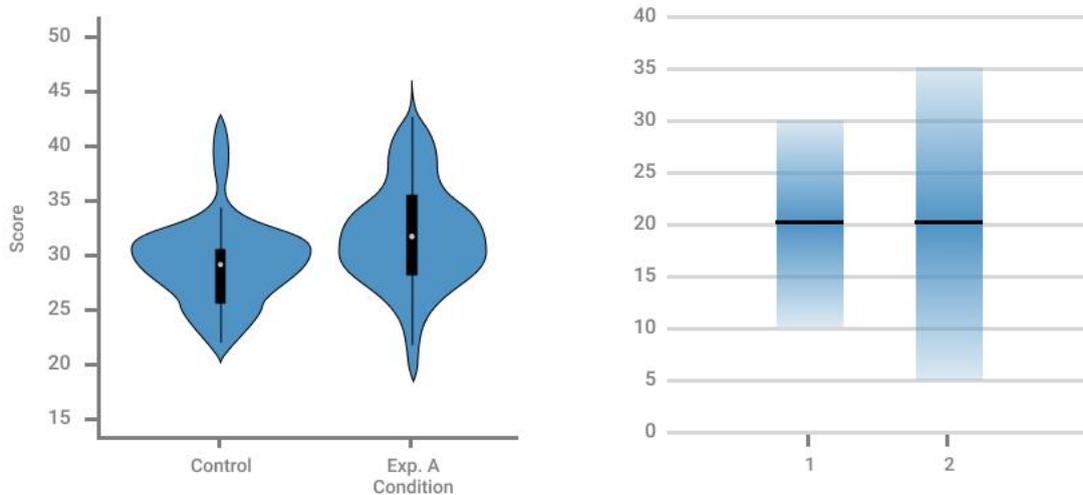


Figura 18 e 19: Gráfico de violino e gráfico de gradiente. Fonte: O autor, 2020.

Essas quatro codificações são as mais tradicionais, advindas da estatística, e também as principais referências para se visualizar a incerteza. Apesar de serem difíceis de compreender num contexto cotidiano, ainda são as bases para o estudo sobre dados incertos. Mais ainda, segundo Yau (2013), os métodos tradicionais de se visualizar a incerteza são limitados. De um lado, porque as formas tradicionais geralmente não consideram a cognição e a percepção visual; de outro, os pesquisadores científicos propuseram as tradicionais metodologias para lidar com este assunto, mas sem as conectarem aos problemas com os quais o público em geral se preocupa. Essas metodologias comumente dependem que os leitores compreendam previamente a base estatística para então absorver a visualização.

Gráficos de conjunto - *ensemble plots* - (Figura 20) são, em sua maioria, utilizados em previsões meteorológicas. Em resumo, um conjunto (linhas) são vários modelos executados ao mesmo tempo individualmente, possuindo condições iniciais e de contorno diferentes. Cada uma dessas execuções é uma linha e quando as linhas estão mais próximas, a previsão é razoavelmente mais confiável do que quando estão separadas.

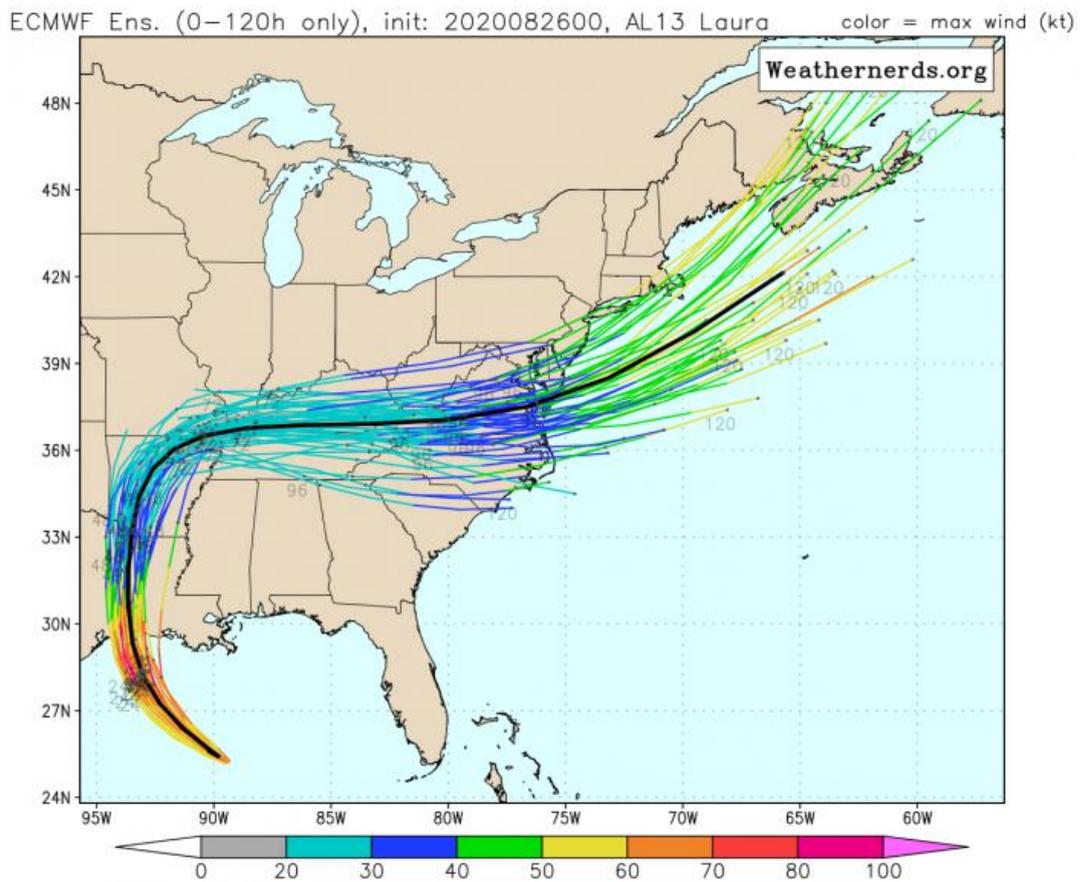


Figura 20: Gráfico de conjunto. Fonte: Wheternerds, acesso em fevereiro de 2021.

Já os *Hypothetical Outcome Plots*, *HOPs* são gráficos onde a distribuição é representada em uma sequência de quadros de animação (Figura 21). Os gráficos de resultados hipotéticos fizeram parte de um estudo realizado pela pesquisadora Jessica Hullman (2015), que utilizou de tecnologias conhecidas para animar codificações tradicionais que, de forma estática, não levavam em consideração a percepção visual do leitor ou sua cognição. Então, ao invés da barra central estar estática em um gráfico com uma barra de erros por exemplo, esta barra se move ao longo do intervalo de confiança, demonstrando que o resultado pode variar. Partindo do pressuposto que os leitores adquirem uma impressão mais intuitiva acerca da incerteza enquanto assistem à sequência de quadros.

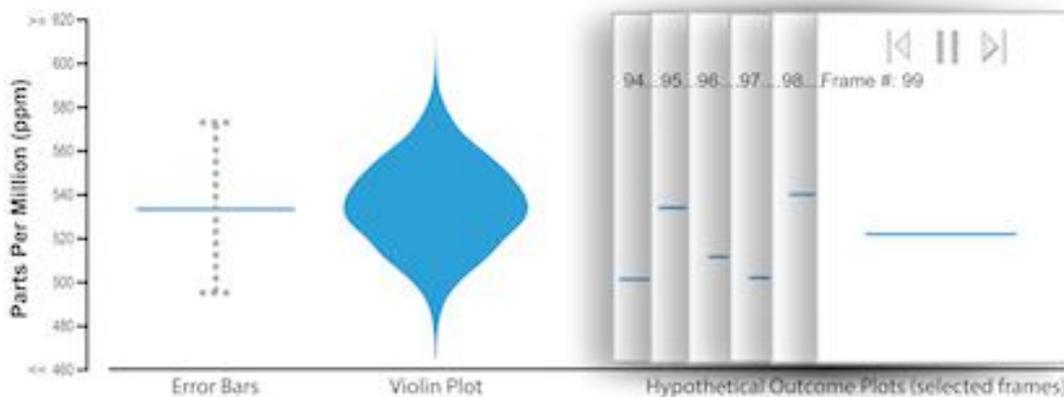


Figura 21: *Hypothetical Outcome Plots*. Fonte: Lace Padilla, Matthew Kay e Jessica Hullman. *Uncertainty Visualization*, 2020.

Por fim, o gráfico de quantis (Figura 22) é uma maneira de representar a distribuição contando com o segmento de uma curva de probabilidade e circunferências empilhadas em seu eixo x. Essa técnica se dá pela transformação de um resultado de dados discretos (antes existentes em uma curva de intervalo de confiança, que mostraremos a seguir), em dados contínuos, onde subdividimos a área total sob a curva em unidades de tamanhos iguais e desenhamos cada unidade como um círculo (WILKE, 2020). Desta forma, podemos quantificar pelo números de bolinhas presentes, e não pela especulação da altura ou média de uma curva. Essa indicação de quanto mais dados, maior quantidade de coisas representadas visualmente é proposta pela metodologia vianense chamada *ISOTYPE* (NEURATH, 1980), criada por Otto Neurath, onde usamos uma sequência de ícones para representar dados quantitativos. Esta técnica pode ser mais eficiente para quem não possui raciocínio numérico (MONMONIER, 1995) desenvolvido.

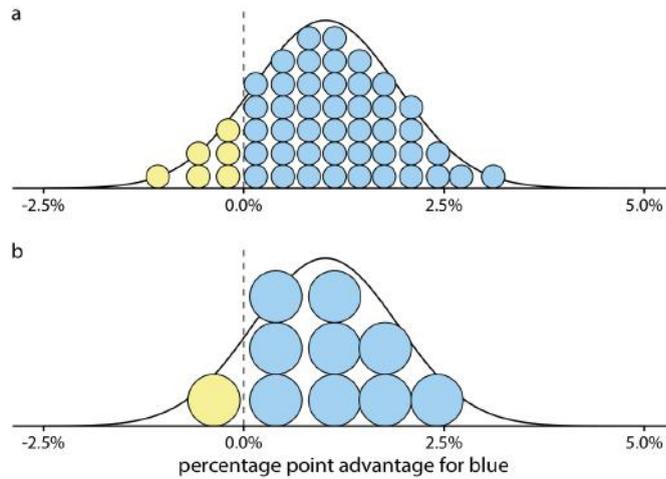


Figura 22: Gráfico de quantis. Fonte: Claus O. Wilke. Fundamentals of data visualization, 2021.

3.2.2 Codificação visual da incerteza

A segunda grande área de estudo da representação visual da incerteza proposta por Padilla (2020), é composta pela codificação visual da incerteza (Figura 23), que consiste em uma técnica de mapeamento onde se define a aparência dos marcadores, também descritos como marcas (barras, pontos, linhas, entre outros) ou pistas visuais, usando controles como cor, posição e transparência. A vantagem desta técnica é a edição imediata de um componente que já está em uso, por exemplo, aumentando a sua transparência para indicar uma maior incerteza. Portanto, neste subcapítulo mencionaremos apenas algumas codificações que tradicionalmente são vistas no âmbito da visualização de dados.

Codificações visuais da incerteza

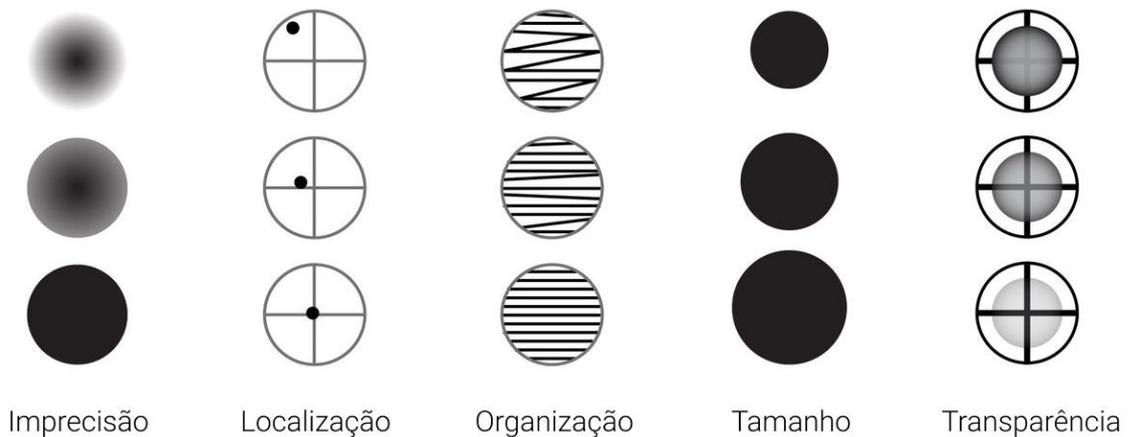


Figura 23: Codificação visual da incerteza. Tradução do autor. Fonte: Lacey Padilla, Matthew Kay e Jessica Hullman. *Uncertainty Visualization*, 2020.

A primeira codificação visual que citaremos é a imprecisão de um marcador visual, que se trata de uma maneira de mostrar aos leitores que os dados não são exatamente precisos. Quanto mais incerta for a estimativa, mais difícil será vê-la em comparação a dados mais concretos que podem compor a visualização. Com a imprecisão, mencionamos aqui falta de opacidade, imagens borradas, cores não saturadas, entre outras formas de variar um componente a ponto de deixá-lo menos evidente.

Na figura 24 podemos ver que as linhas do dashboard representam a direção do vento, mas a opacidade reflete a precisão deste dado. Por conta disso, a codificação de opacidade foi utilizada para obter o efeito de imprecisão sobre a incerteza. O que está mais claro e vívido comporta mais presença que o que não ocupa tanto espaço na nossa visão. Poucas vezes o aspecto de imprecisão é utilizado de uma maneira que podemos facilmente identificar (YAU, 2018) e pode ser acompanhado de uma ou mais codificações que reforcem seu aspecto.

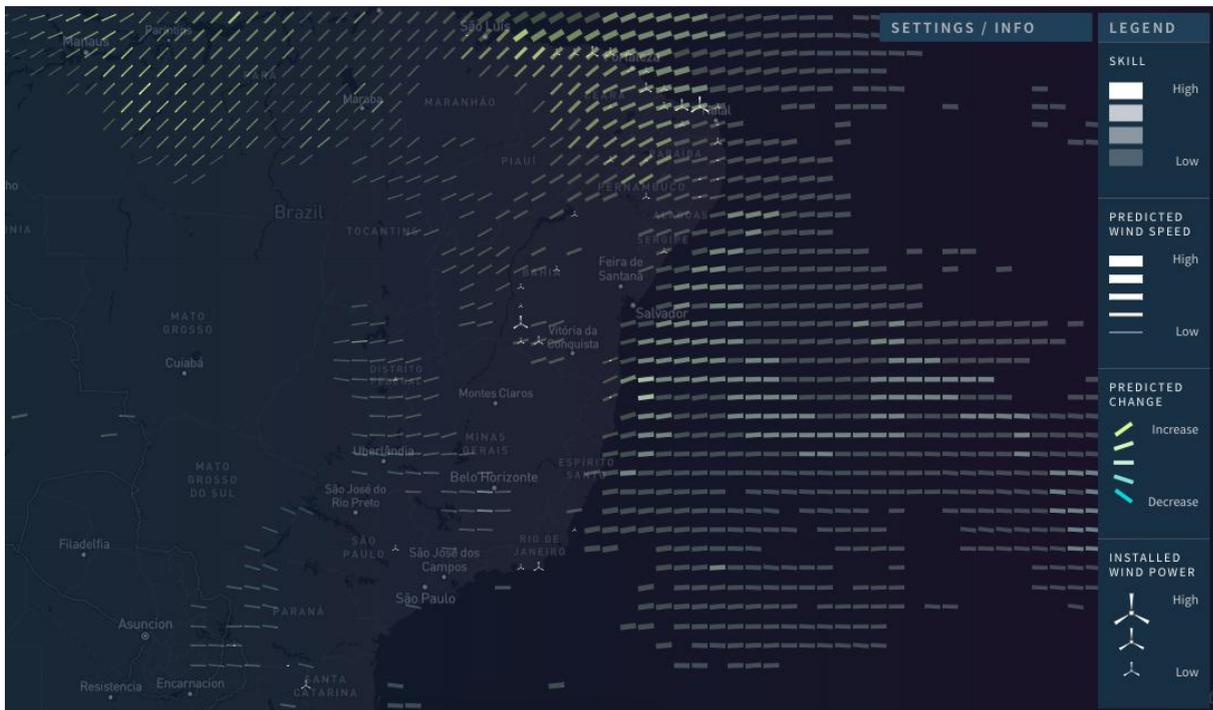


Figura 24: *Project Ukko*. Fonte: Moritz Stefaner. BSC - EUPORIAS, 2019.

A localização de um dado pode influenciar seu grau de precisão também, como podemos ver na figura 25, onde os pontos que se encontram a certa distância das linhas, representam a variação de votos. As áreas que envolvem as linhas de tendência representam as margens de erro de cada candidato que, quando sobrepostas, levam a um possível empate técnico. Foram utilizadas duas codificações para representar a incerteza nesse caso: a primeira sendo os pontos fora da tendência (linhas e margens de erro) e a própria margem de erro que pode levar a uma grande variação na média das pesquisas, dependendo do nível de confiança que poderíamos depositar na pesquisa.

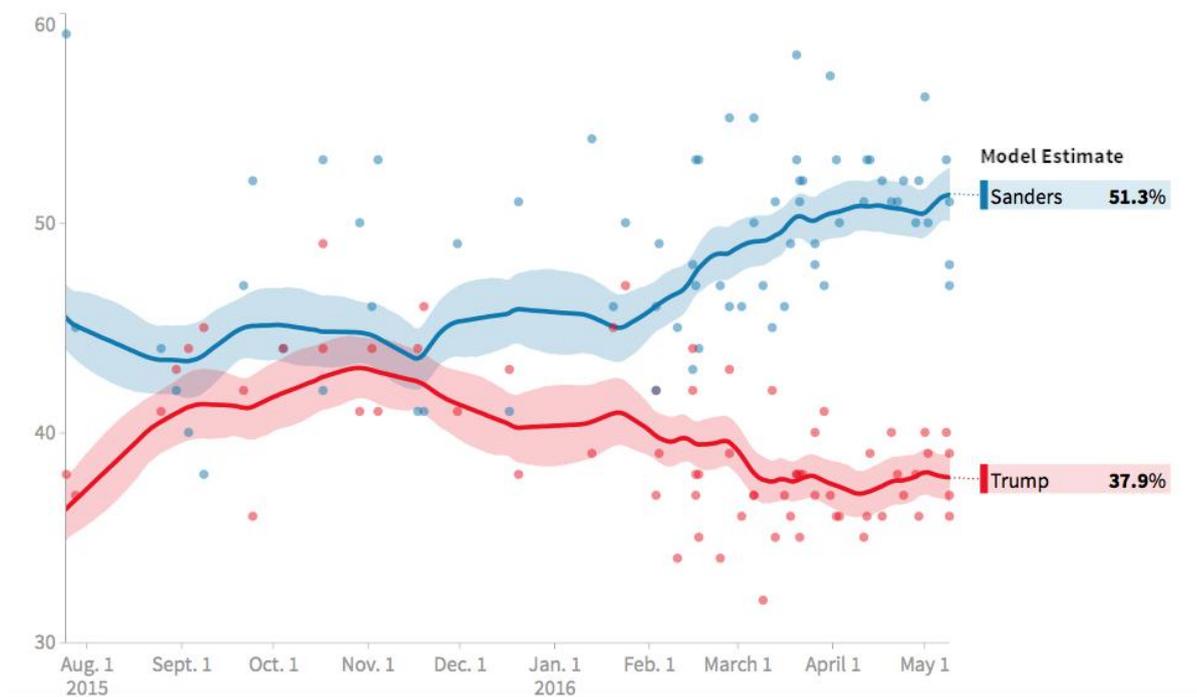


Figura 25: 2016 *General Election: Trump vs. Sanders*. Fonte: HuffPost Pollster, 2020.

Por fim, o tamanho também é um marcador visual utilizado para representar a incerteza, mas pode ser confusa sua comparação entre elementos, uma vez que não percebemos algumas medidas de imediato. O tamanho de uma barra de erro (Figura 26) determina a proporção de incerteza de uma amostra, mas, como visto anteriormente, esse modo de representação tradicional pode não ser facilmente compreendido por qualquer público. Além do que, o comprimento de um marcador usualmente é uma medida que representa dados contínuos, mas não seria correto interpretarmos que a barra de erro possui crescente (ou igual) grau de certeza de sua base até seu topo. Ela apenas representa qual a média de certeza do conjunto de dados, que pode variar de seu meio em igual proporção para cima ou para baixo.

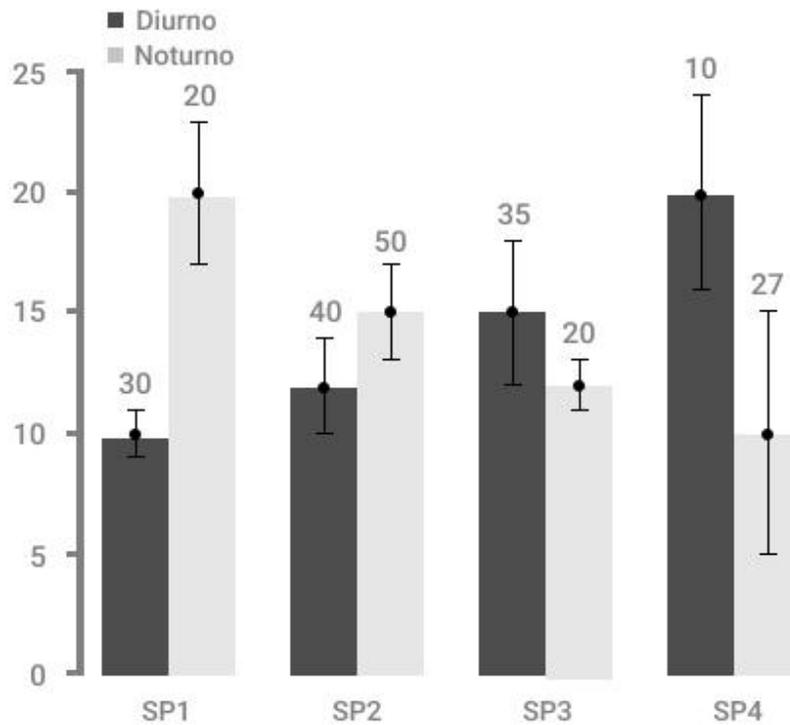


Figura 26: Barras de erro. Fonte: O autor, 2020.

3.2.3 Método híbrido

Anotações gráficas distributivas e codificações visuais podem ser combinadas para criar abordagens híbridas como os gráficos de caixa de contorno (Figura 27). Deste modo, as anotações gráficas (barras de erro, gráfico de contorno) e a personalização de marcadores (transparência, localização) podem se complementar para que os leitores tenham uma noção mais detalhada da visualização. A seguir, apresentamos duas técnicas bastante comuns que representam a abordagem deste subcapítulo.

Métodos híbridos

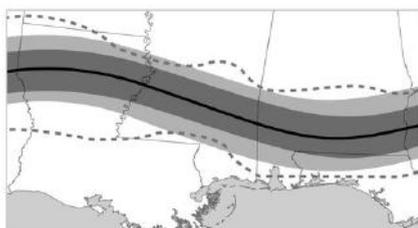


Gráfico de caixa contornada

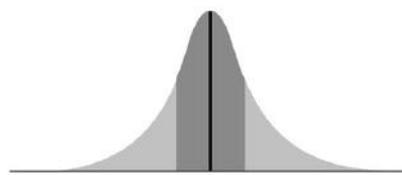


Gráfico de intervalo e densidade de probabilidade

Figura 27: Método híbrido. Tradução do autor. Fonte: Lace Padilla, Matthew Kay e Jessica Hullman. *Uncertainty Visualization*, 2020.

O gráfico de caixa contornada, dentre as várias técnicas de visualização de incerteza, é de grande interesse neste método, pois muitas vezes a forma com que os dados incertos são visualizados não captura exatamente o comportamento do fenômeno em estudo (como a passagem de um furacão sobre determinado território). O contorno das linhas representam médias e números em conjunto que se aproximam da certeza dentro de três parâmetros: uma área mais escura, para o que está mais concreto, uma mais clara, e um pontilhado para que seja possível visualizar uma margem de erro.

O gráfico de intervalo e densidade de probabilidade atua de maneira semelhante a um intervalo de confiança, onde as cores mais centralizadas representam maior grau de certeza sobre aquela amostra, dividindo-a na variação de um mesmo matiz, no caso da figura 28, dividindo em cores azuis para recortar a curva do intervalo e tornar sua média mais visível ao leitor. O grande desafio é visualizar informações estatísticas robustas e a variabilidade de um conjunto (WHITAKER, 2014).

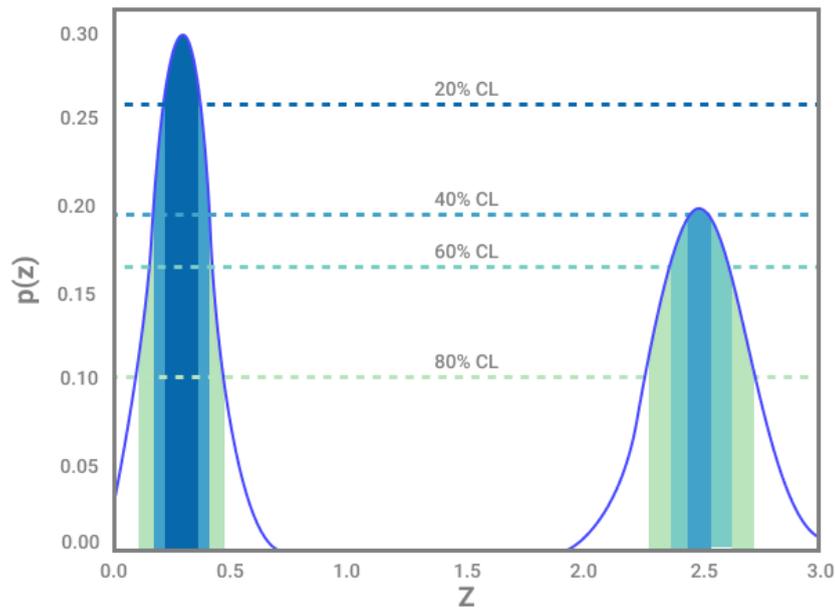


Figura 28: Gráfico de intervalo e densidade de probabilidade. Fonte: O autor, 2020.

4 ANÁLISE DE VISUALIZAÇÕES COM DADOS INCERTOS

Este capítulo visa complementar o que foi apresentado até agora, observando algumas representações visuais de dados incertos. Os exemplos a seguir se subdividem nas seguintes temáticas: saúde, eleições, meteorologia e outros. Dessa forma, podemos destacar os gráficos que se integram ao nosso cotidiano por serem amplamente divulgados por veículos jornalísticos. Propomos analisar se os recursos visuais abordados neste projeto estão presentes nas soluções escolhidas pelos autores dessas visualizações e, por fim, se atendem ao contexto de seu público.

O objetivo principal é oferecer uma base teórica, sugerindo um modelo de análise construído em formato de quadro, como um método a ser aplicado em visualizações de dados incertos. Pois pensando nas restrições de *design* em seus projetos, *designers* podem desenvolver aplicações exclusivas para a incerteza.

É importante ressaltar que não existe uma única abordagem de visualização de incerteza que garanta a correta representação gráfica destes dados, nem que ao usar estes recursos estaremos melhorando esta representação e/ou aumentando a confiança e o julgamento por parte dos leitores. Desse modo, é importante que os *designers* da informação atentem para suas escolhas visuais, do contrário, corre-se o risco de adicionar mais confusão a um processo já complexo (PADILLA, 2020).

4.1 Saúde

Devido a pandemia causada pela Covid-19, ocorreu um aumento das visualizações de dados na área da saúde. Em decorrência desse aumento e da constante atualização dos dados, algumas vezes nos deparamos com bancos de dados defasados. Um exemplo são os números de pacientes diagnosticados com Síndrome Respiratória Aguda Grave (SRAG), que podem não ter sido testados para o novo coronavírus.

Podemos assumir que boa parte das visualizações de dados da Covid-19 é repleta de uma **incerteza epistêmica** (trata de uma incerteza causada por falta de informação), que, por exemplo, poderia ser solucionada, neste caso, com um maior número de testes.

No Brasil, os veículos de mídia fizeram um consórcio para a divulgação da pesquisa de novas mortes pela doença e o aumento de casos por região demográfica, movimento que tem auxiliado a população a ter mais clareza sobre os impactos da pandemia. Temos como exemplo o Jornal Nexo, demonstrando **a linha de tendência de novas mortes por Covid-19 por país**, presente na figura 29.

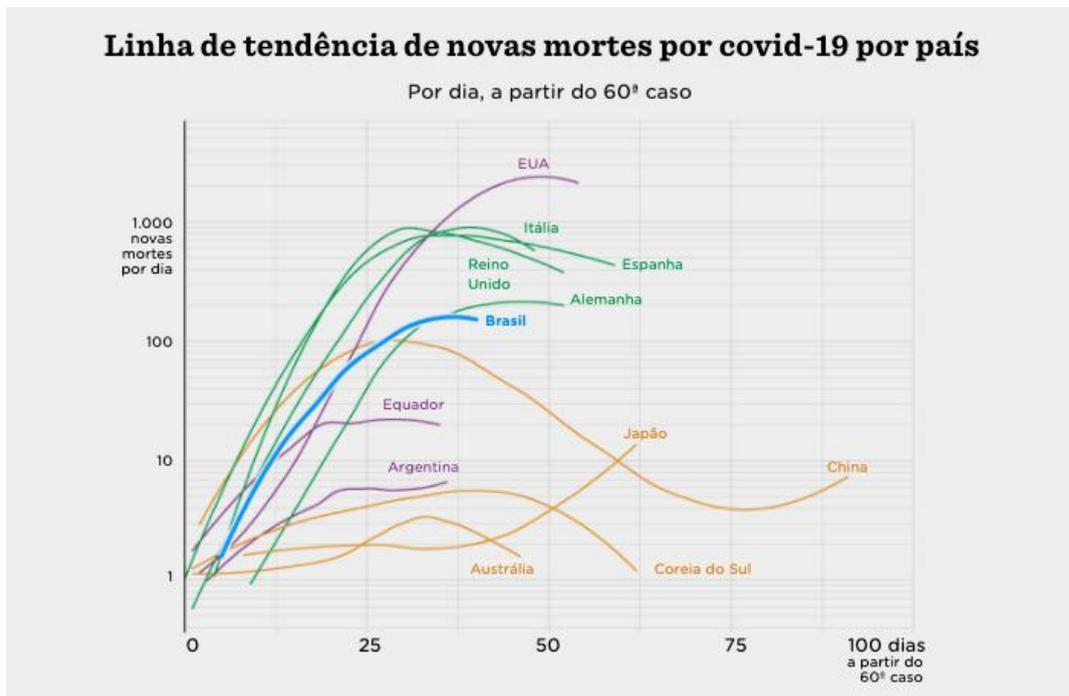


Figura 29: Linha de tendência de novas mortes por covid-19 por país¹. Fonte: Jornal Nexo, 2020.

Os **marcadores visuais** utilizados foram a posição, a linha e a cor matiz de cada país, em um gráfico de eixo x-y em **escala** logarítmica. Podemos notar que as cores demarcam países localizados em regiões próximas e que a linha azul de maior espessura retrata o Brasil. O aglomerado de linhas horizontais demonstra a frequência da escala logarítmica, que, quanto mais próximas as linhas estão, mais aglomerados os números de novas mortes por dia. Por exemplo, entre o número 1 e o número 10 no eixo y, vemos a frequência maior das linhas horizontais por espaço. Duas vezes o gráfico reafirma que o número de dias foi contado a partir do 60º caso de Covid-19: uma vez abaixo de seu título, e outro momento em seu eixo x, sendo sua **contextualização** para a disposição dos dados.

Podemos notar, com essa análise, as seguintes características da visualização:

Tipo de incerteza: Epistêmica;

Sistema de coordenadas: Plano cartesiano;

¹ Disponível em:

<https://www.nexojornal.com.br/grafico/2020/04/22/Covid-19-a-evolu%C3%A7%C3%A3o-dos-casos-e-das-mortes-em-diferentes-pa%C3%ADses>. Acesso em: 21 de nov. 2020.

Escala: Logarítmica;

Contexto: Legendas;

Marcadores visuais: Posição, linhas e cores matizes.

Apesar do jornal ter o cuidado de demarcar o 60º caso de morte por dia, ainda ocorrem ruídos nesta **incerteza epistêmica**, por faltar a confirmação de que as mortes foram causadas pela doença. Ou seja, o leitor não sabe o quanto desses números são óbitos confirmados para Covid-19 e o quanto ainda são suspeitos pois os testes ainda não foram concluídos. O gráfico faz comparação entre países durante o período de 100 dias, mas poderiam ser acrescentadas informações por meio de um **botão de informação** ou um *mouseover* (ação de passar o *mouse* por cima de um componente e obter informações acerca do mesmo). A inserção de componentes interativos poderia adicionar mais clareza sobre quantos óbitos foram registrados por síndrome respiratória aguda grave e quantos óbitos foram causados pela covid-19.

Podemos, então, acrescentar como uma **anotação gráfica de incerteza sugerida** um botão de informação ou *mouseover* para evidenciar os dados incertos.

Na mesma matéria, vemos o recorte para o Brasil (Figura 30) que exhibe uma linha de incerteza em cinza, onde o número absoluto faz um zigzague por detrás da linha azul e leva o leitor a uma compreensão sobre o tracejar da linha de tendência. A escala logarítmica mudou para um número de 50 dias antes do 60º caso e foi evidenciado dois marcos importantes – tanto para o número de casos, quanto para as mortes – sendo um a partir do momento de controle de viagens internacionais e outro para o fechamento de escolas no estado de São Paulo.

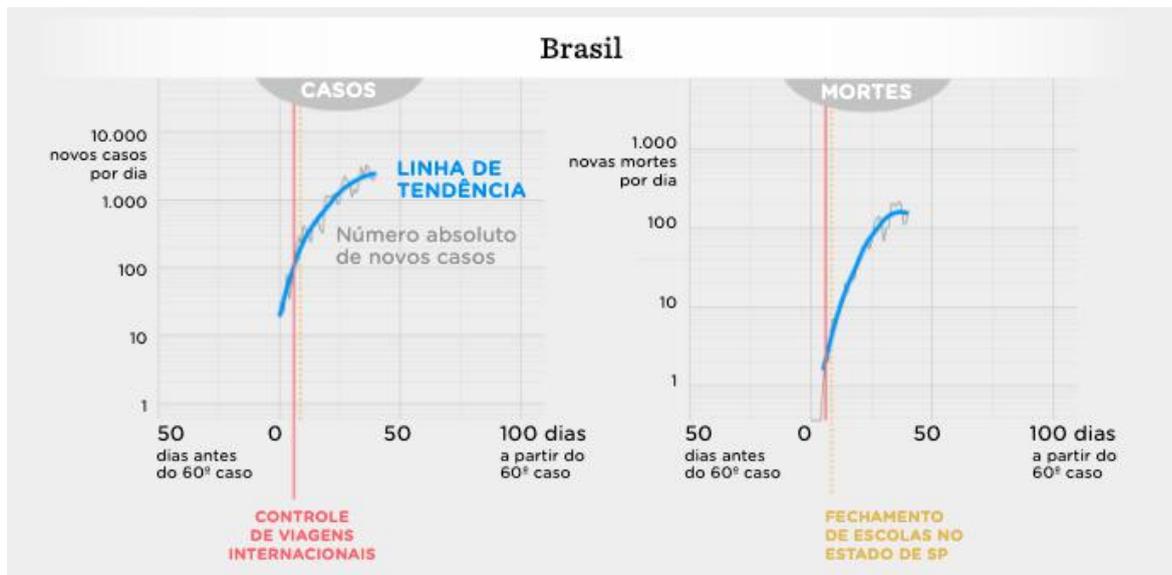


Figura 30: Linha de tendência de novas mortes por Covid-19 no Brasil². Fonte: Jornal Nexo, 2020.

Podemos notar as seguintes características da visualização:

Tipo de incerteza: Epistêmica;

Sistema de coordenadas: Plano cartesiano;

Escala: Logarítmica;

Contexto: Legendas;

Marcadores visuais: Posição, linhas e cores matizes;

Anotação gráfica da incerteza: Posição, cor e linha.

O jornal Nexo evidenciou pontos importantes em seus gráficos sobre Covid-19, sendo eles: as ações governamentais que tiveram impacto sobre o curso da pandemia, os números absolutos por detrás de seus números contínuos e permitiu uma comparação entre duas ações governamentais.

² Disponível em:

<https://www.nexojornal.com.br/grafico/2020/04/22/Covid-19-a-evolu%C3%A7%C3%A3o-dos-casos-e-das-mortes-em-diferentes-pa%C3%ADses>. Acesso em: 21 de nov. 2020.

4.2 Eleições

As eleições para prefeitos e vereadores no Brasil em 2020 foram amplamente debatidas e estudadas em sua véspera, anterior ao dia 15 de novembro. Diversas pesquisas de opinião pública surgem nesta época, assim como a análise de candidatos e a avaliação de suas plataformas. Podemos notar uma crescente de palavras como pontos percentuais, amostra de pesquisa, margem de erro e índice de confiança, surgidas em matérias de jornais. Estes conjuntos de dados, vindos de opinião pública, são compostos por uma **incerteza aleatória**, inerente ao processo de recolhimento de amostra porque conta com opiniões de um número limitado de pessoas para a previsão de um futuro próximo.

Um exemplo, foi a equipe do Jornal Nexô que divulgou inúmeras pesquisas sobre o desempenho dos candidatos e análises dos mesmos, por região demográfica. Na figura 31 podemos ver a pesquisa de reeleição de candidatos por capital e a porcentagem das intenções de voto de acordo com uma amostra de eleitores que foi recolhida até o dia onze de novembro de 2020.



Figura 31: Desempenho dos candidatos à reeleição nas pesquisas³. Fonte: Jornal Nexo, 2020.

O principal **marcador visual** do gráfico é o círculo em uma barra branca dividida em uma **escala percentual** de 0%, 25%, 50% e 75%, sendo evidenciado o seu centro (50%). A linha tracejada representa 50% dos eleitores em uma escala percentual de votos em cada capital, distribuída em um **plano cartesiano** de eixo x-y. O azul mais vívido sendo a média das pesquisas, e o azul mais transparente, as pesquisas. Nota-se que a pesquisa sofreu uma grande variação em sua média, como por exemplo no Tocantins, onde a candidata Cinthia do PSDB possuía uma pesquisa mais distante de seu ponto azul.

Podemos notar as seguintes características da visualização:

³ Disponível em:

<https://www.nexojornal.com.br/grafico/2020/11/12/O-desempenho-dos-prefeitos-que-buscam-reeleicao-nas-capitais>. Acesso em: 21 de nov. 2020.

Tipo de incerteza: Aleatória;

Sistema de coordenadas: Plano cartesiano;

Escala: Percentual;

Contexto: Legendas;

Marcadores visuais: Posição, círculos e cores matizes;

Anotação gráfica da incerteza: Posição e imprecisão (cor menos opaca).

O Jornal Nexo foi prático ao demonstrar os dados incertos sobre o desempenho dos candidatos à reeleição nas pesquisas, inserindo a própria pesquisa como marcador visual de um azul menos saturado. Sendo comum que gráficos eleitorais tomem tal posição, porque estão sugerindo um futuro próximo, não declarando um acontecimento.

Em uma pesquisa sobre intenção de voto para a prefeitura de João Pessoa (Figura 32), o G1 divulgou dados com uma margem de 4 pontos percentuais para mais ou para menos, sendo de 95% as chances do resultado desta amostra representar a realidade eleitoral para esta cidade. São exibidos cinco candidatos e a fonte dos dados foi retirada do Instituto Brasileiro de Opinião Pública e Estatística (IBOPE), órgão responsável pela maioria das amostras populacionais retiradas para pesquisas de opinião pública.



Figura 32: Intenção de voto para a prefeitura de João Pessoa. Fonte: Jornalismo G1⁴, 2020.

Neste gráfico, podemos ver que a margem de erro é representada por um **marcador visual** cinza que abrange uma parte da barra azul e um pouco para cima desta, este recurso foi utilizado para os 4 pontos percentuais da disputa entre os candidatos. É comum essas pesquisas aparecerem como um gráfico de barras com uma **escala percentual** em um **plano cartesiano**. A diferença nesta visualização pode ser vista pelos números ao lado da caixa cinzenta, demarcando o ponto de seu índice mais baixo, e o mais alto, permitindo que o leitor interprete que há uma possibilidade da maioria dos candidatos estarem empatados de acordo com a média da amostra recolhida pelo IBOPE.

O G1 utiliza da matéria como suporte ao gráfico (sendo seu **contexto**) divulgando a quantidade de pessoas que foram ouvidas, neste caso sendo 602, a data da pesquisa, entre os dias 5 e 6 de outubro, a margem de erro, 4 pontos, e o nível de confiança, de 95%, que podemos depositar neste resultado.

⁴ Disponível em:

<https://g1.globo.com/pb/paraiba/eleicoes/2020/noticia/2020/10/05/pesquisa-ibope-em-joao-pessoa-cicero-18percent-nilvan-15percent-ricardo-12percent-wallber-10percent.ghtml>. Acesso em: 21 de nov. 2020.

Podemos notar as seguintes características da visualização:

Tipo de incerteza: Aleatória;

Sistema de coordenadas: Plano cartesiano;

Escala: Percentual;

Contexto: Matéria e legendas;

Marcadores visuais: Posição, barras e cor matiz;

Anotação gráfica da incerteza: Margem de erro, imprecisão (cor menos opaca).

4.3 Meteorologia

A meteorologia é uma ciência que estuda a atmosfera terrestre, com foco de estudo em processos atmosféricos, como a previsão de formação de furacões, e a previsão do tempo. Essa ciência possui uma **incerteza ontológica** envolvida em seu processo pela falta de precisão de modelagem gráfica sobre as ações da natureza.

O jornal Newsweek publicou uma matéria em 2017 sobre a passagem do furacão Irma na Baía de Guantánamo, campo de detenção dos Estados Unidos ao sul de Cuba. Um gráfico (Figura 33) foi feito para dar suporte a matéria, colocando em destaque a categoria da tempestade e sua passagem por áreas densamente populosas. Até então, o furacão tinha feito dez vítimas em ilhas vizinhas e estava ameaçando passar por Cuba de maneira mais devastadora. Desta forma, o jornal relatou especificamente o ocorrido onde os Estados Unidos não tinha a intenção de evacuar os presos em segurança máxima nos próximos dias, e talvez estes sofressem pela tempestade que passaria ao norte do local.

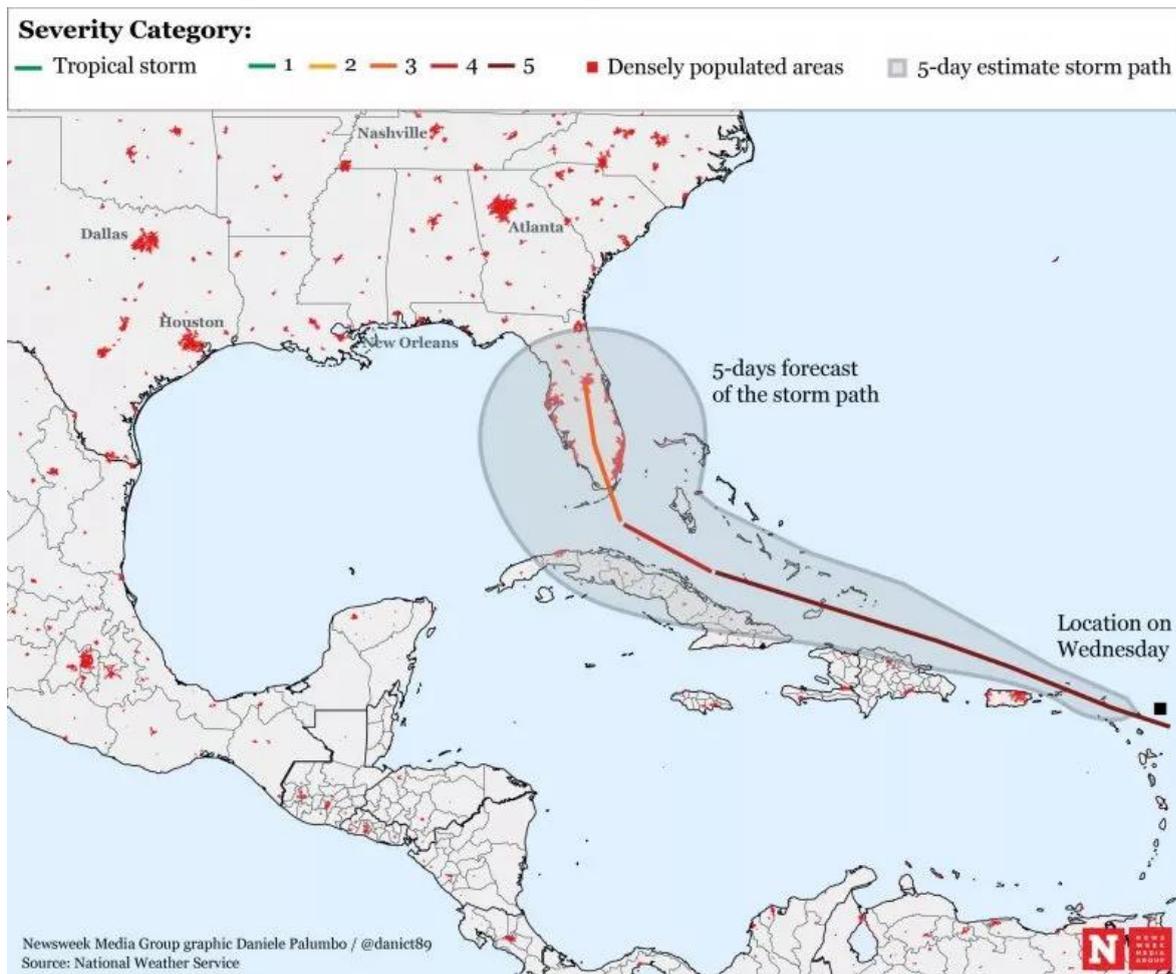


Figura 33: Furacão Irma⁵. Fonte: Newsweek, 2017.

Foi mapeado na visualização um ponto, ao início da linha - da direita para a esquerda, o local em que a tempestade estava em um dia e outro ponto, ao final da mesma linha, onde ela estaria dali 5 dias. A área que cerca a linha, onde imaginamos ser a passagem mais previsível da tempestade, gera certa confusão, pois os leitores podem especular que o furacão aumente de tamanho e se forme conforme a demarcação da área cinza no mapa cartográfico.

Os **marcadores visuais** utilizados para essa visualização foram: a direção e posição da linha que traça o caminho do furacão, sua cor que varia em gradiente, do verde para o vermelho, os pontos vermelhos que demarcam regiões mais populosas e a área cinza. Este

⁵ Disponível em: <https://www.newsweek.com/hurricane-irma-guantanamo-bay-evacuations-terrorists-660963>. Acesso em: 10 de fev. 2021.

último marcador pode também ser chamado de “cone de incerteza”, tratando-se de uma anotação gráfica da incerteza, por seu aspecto de abordar todos os caminhos pelos quais o furacão poderia passar que não fossem a própria linha. A legenda traz o **contexto** para as cores, assim como o texto sobre o próprio gráfico. A **escala** apresentada é temporal contínua, por contar com um intervalo de tempo e o **sistema de coordenadas** é geográfico, por visualizarmos o mapa de Cuba.

O jornal Newsweek demonstrou todos os pontos importantes para representar a previsão meteorológica que estava deixando Cuba em alerta. O autor Alberto Cairo (2017) apresentou um estudo sobre cones de incerteza (área cinza da figura 33) e enfatizou sua ineficiência de representação visual, sendo de 66% o nível de confiança neste tipo de marcador, segundo seu estudo.

Podemos notar as seguintes características da visualização:

Tipo de incerteza: Ontológica;

Sistema de coordenadas: Mapa cartográfico;

Escala: Temporal contínua;

Contexto: Legendas;

Marcadores visuais: Linha, direção, posição, pontos e cores.

Anotação gráfica da incerteza: Área demarcada em cinza (cone de incerteza).

Mas ainda assim, reiteramos a importância do nível de confiança, 66%, depositado neste último tópico (CAIRO, 2017). Para serem melhor compreendidos, o jornal Newsweek poderia ter seguido outras alternativas e por isso, adicionamos abaixo outro item para esta análise.

Sugestão de representação visual da incerteza: Substituição do cone de incerteza por outras marcações visuais como outras linhas, animação sobre a rota prevista do furacão ou reiteração textual (em visualizações estáticas) sobre a imprevisibilidade do modelo fornecido, onde o curso da tempestade pode ser inexato.

4.4 Outros

Agora que mencionamos os tipos de incerteza em cada exemplo citado nos subcapítulos anteriores, temos a proposta de uma avaliação que não é filtrada por categorias (saúde, eleições e meteorologia), para podermos incluir outros campos que trabalham dados incertos. Então, o atual subcapítulo analisa outras visualizações que não se restringem a assuntos específicos. Tornando-se uma avaliação focada na representação visual e na diversidade de características das visualizações a seguir.

O jornal New York Times realizou em 2018 um estudo sobre o alcance punitivo do racismo nos Estados Unidos. Sendo avaliado, assim, a mobilidade econômica de pessoas de diferentes etnias, da fase infantil à fase adulta. Foi constatado, em primeiro momento, que crianças negras (do sexo masculino) nascidas em famílias ricas possuem mais chances de mudarem sua situação econômica em sua fase adulta que crianças brancas. Essa incerteza é composta por uma média de dados que pode variar bastante, como, por exemplo, de acordo com a idade ou período de tempo em que essas pessoas viveram. Então, podemos avaliar que estes são dados que se formam pela **incerteza aleatória**, onde temos uma média que nos leva a *insights* sobre o que acontece na realidade.

A figura 34 apresenta um *frame* de uma animação, onde os quadrados se movem da fase infantil até a fase adulta, se divergindo de acordo com seu grupo econômico. Trouxemos apenas um dos resultados do estudo, eventualmente o jornal elaborou mais três matérias sobre o assunto, permitindo inclusive que o leitor cruze os dados que prefere para ver as animações que se formam de acordo com suas escolhas.

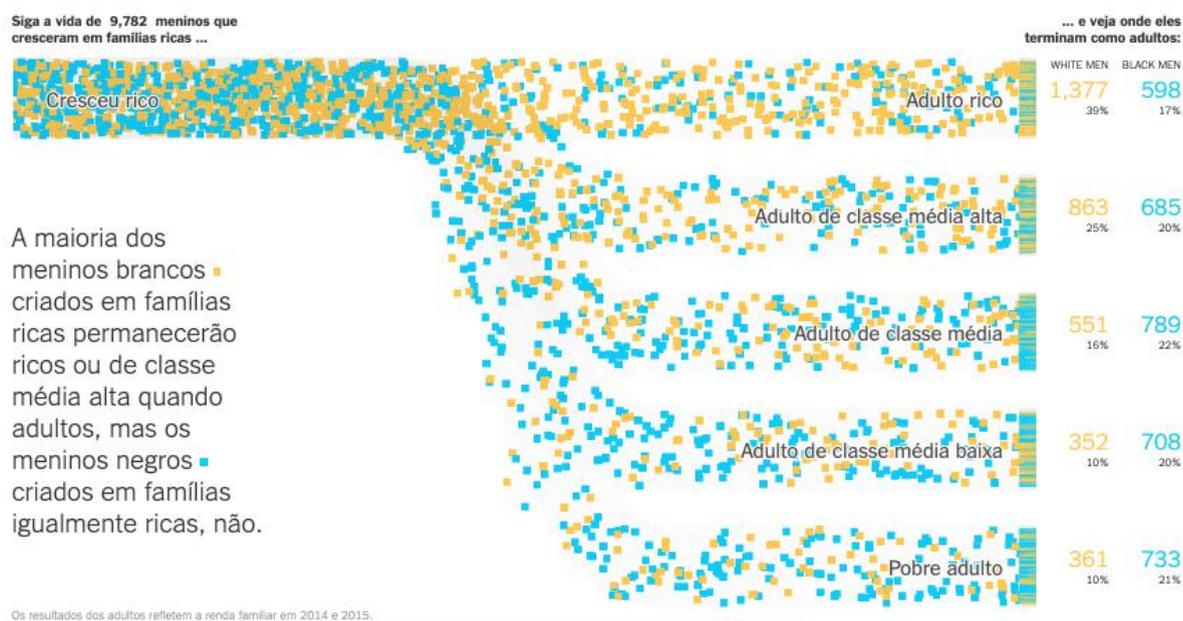


Figura 34: *Income Mobility chart*.⁶ Fonte: New York Times, 2018.

Os **marcadores visuais** utilizados para essa visualização são as cores e a posição dos quadrados que se movem de um grupo a outro, sendo o **sistema de coordenadas** um diagrama de fluxo que apresenta o dado “se movendo” de um ponto “a” a um ponto “b”, não se aplicando aos sistemas propostos por Yau (2013).

O **contexto** é inserido através da legenda e a **escala** apresentada é tanto percentual, quanto linear, porque temos os números totais, e coloridos, acima da porcentagem. A incerteza é inserida na própria animação, onde o resultado do estudo não é definitivo e completo quando representado. Ou seja, para que o leitor veja o resultado completo do estudo, ele precisa ficar um período de tempo em frente a animação para que, individualmente, cada quadrado saia de seu ponto de início e pare em um dos grupos propostos no lado direito do diagrama. Cada dado é individualmente apresentado e está margeado de muitos outros, a imprecisão durante todo o movimento, até o seu final, pode ser considerado uma forma de representar a incerteza pela codificação visual da análise.

⁶ Disponível em: <https://www.nytimes.com/interactive/2018/03/19/upshot/race-class-white-and-black-men.html>. Acesso em: 22 de fev. 2021.

Podemos notar as seguintes características da visualização:

Tipo de incerteza: Aleatória;

Sistema de coordenadas: Não se aplica por ser um diagrama de fluxo;

Escala: Linear e percentual;

Contexto: Legendas e título;

Marcadores visuais: Posição e cores;

Anotação gráfica da incerteza: Movimento dos dados durante a animação.

Podemos constatar que o objetivo da matéria foi demonstrar uma tendência e comprovar a teoria sobre o alcance punitivo do racismo, já que mesmo crianças com *backgrounds* iguais tiveram tanta divergência em sua fase adulta. O jornal explicou o assunto com bastante clareza, e a proposta de fazer uma animação para a visualização de um extenso banco de dados os auxiliou com a incerteza.

Em uma proposta distinta, a *designer* Claire Santoro participou do *SWDChallenge* (*Storytelling With Data Challenge*, desafio de visualização de dados com narrativa que aconteceu em setembro de 2019) com seu gráfico (Figura 35) sobre o clima “*Sweater Weather*” em St. Louis nos Estados Unidos. Este tipo de clima, de acordo com a autora é “Quando você pode usar um suéter confortavelmente o dia todo sem escondê-lo sob uma parca. Para mim, isso significa altas temperaturas diárias em algum lugar entre 4° a 16° graus Celsius” (SANTORO, 2019).

The other day, in scorching heat, a friend who just moved to town asked me:

"When will it be sweater weather in St. Louis?"

* When you can comfortably wear a sweater all day long without hiding it under a parka.
For me, this means daily high temperatures somewhere between 40° and 60°F.

My answer: Maybe in **October**. Almost certainly in **November**. But you never know... every day of the year has been bad sweater weather at some point in the last 10 years.

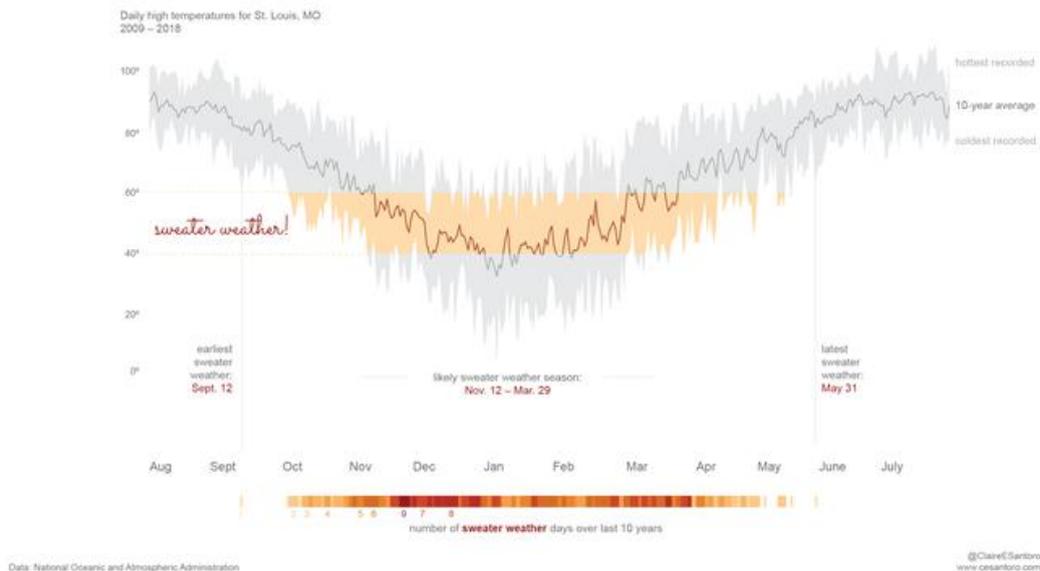


Figura 35: *When will it be sweater weather in St. Louis?*⁷. Fonte: Claire Santoro, 2019.

Na figura 35 podemos ver a média de temperatura de agosto a junho de 2008 a 2018, representada pela linha, envolta por uma faixa cinzenta que demonstra a variação de temperatura de acordo com a passagem de anos. Sendo assim, a linha, a posição, a cor alaranjada e vermelha, a posição dos elementos, e a faixa cinza, os **marcadores visuais** do gráfico, alinhados pelo **sistema de coordenadas** cartesiano em **escala** temporal contínua. A intenção de Santoro foi observar em qual dos períodos apresentados estaria este clima perfeito para *"Sweater Weather"*, sendo de 4° e 16° graus Celsius. Desta forma, foi inserido um marcador laranja entre os meses de outubro e novembro, onde a média costumava variar para sua temperatura ideal. A incerteza presente nestes dados é **aleatória**, pois consta em um futuro próximo, mas que se baseia na média de dados anteriores para compor essa suposição.

⁷ Disponível em:

<http://www.storytellingwithdata.com/blog/2019/9/24/swdchallenge-an-expression-of-ignorance>. Acesso em 22 de fev. 2021.

Além disso, diversos itens do gráfico nos leva a uma percepção de menor formalidade, sendo estes: a tipografia curvilínea utilizada, as cores que compõem o gráfico e as observações da própria autora.

Podemos notar as seguintes características da visualização:

Tipo de incerteza: Aleatória;

Sistema de coordenadas: Plano cartesiano;

Escala: Temporal contínua;

Contexto: Legendas;

Marcadores visuais: Linha, posição e cores;

Anotação gráfica da incerteza: Intervalo distributivo (faixa cinza) e marcador laranja/vermelho sobre o período de agosto a novembro do gráfico.

Ambas as visualizações anteriores possuem caráter singular e maneiras diferentes para apresentar a incerteza aleatória. Mas podemos também observar que foram codificações visuais específicas para cada caso que não interferem diretamente em sua percepção. O Jornal New York Times contou com recursos da animação para contar uma história sobre mobilidade econômica e racismo. Já Claire Santoro, compartilhou um ponto de vista individual sobre em qual época do ano estaria mais confortável com o clima de St. Louis, Missouri.

Outro participante do *SWDChallenge* foi o engenheiro de *software* Augusto Oazi, que propôs uma previsão sobre as posições de cada time de futebol no Campeonato Brasileiro de 2019 (Figura 36). Os dados também são compostos pela **incerteza aleatória**, contando com os resultados anteriores da performance de cada time. Mas como cada time pode mudar de um mês para o outro, tanto de jogadores, quanto técnicos, ainda seria difícil constatar, com total certeza, quem iria ganhar aquele ano.

Flamengo are the favorites to win the 2019 Brazilian League.

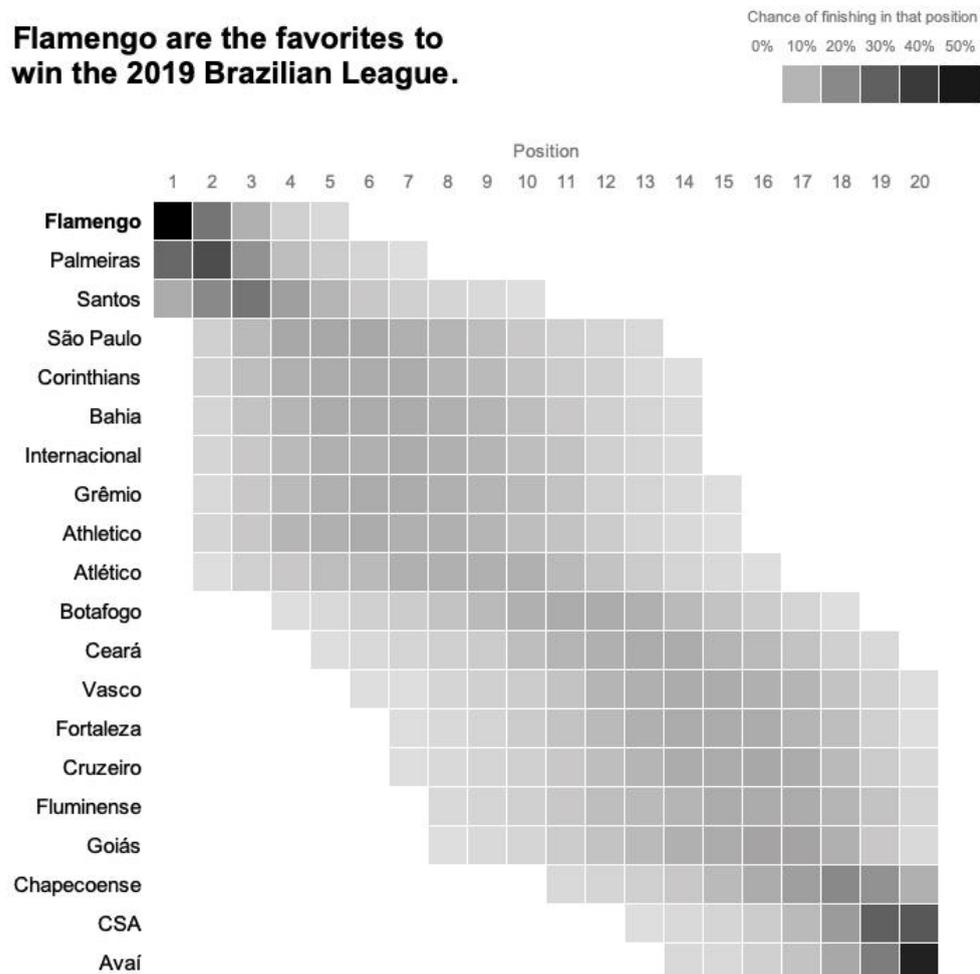


Figura 36: *Flamengo are the favorites to win the 2019 Brazilian*⁸. Fonte: Augusto Oazi, 2019.

A visualização acima foi composta por um **sistema de coordenadas** cartesiano, com seu eixo x sendo as posições de classificação dos times, e o y os times que concorrem ao campeonato do ano. Os quadrados, a posição destes, e a mudança no valor da cor cinza, são os **marcadores visuais** e a incerteza é composta pela variação da escala tonal. A legenda nos dá o **contexto** das chances (em porcentagem, sendo sua **escala**) de cada time se classificar, sendo de 0% a 50%. Ao todo, a visualização forma uma faixa composta de forma modular para cada resultado individual dos times e pode variar bastante quanto maior o número de quadrados;

⁸ Disponível em:

<http://www.storytellingwithdata.com/blog/2019/9/24/swdchallenge-an-expression-of-ignorance>. Acesso em 22 de fev. 2021.

Como observamos no time Atlético que pode estar tanto em primeiro quanto em quinto lugar no Campeonato Brasileiro de 2019, mas com mais chances entre a sétima e décima posição.

A granularidade do gradiente leva ao maior índice de incerteza sobre algum resultado específico, mas também nos dá um panorama geral sobre a especulação da posição de cada time de futebol, que pode fugir completamente deste panorama em particular.

Podemos notar as seguintes características da visualização:

Tipo de incerteza: Aleatória;

Sistema de coordenadas: Plano cartesiano;

Escala: Percentual;

Contexto: Legendas;

Marcadores visuais: Quadrados, posição e valor;

Anotação gráfica da incerteza: Codificação visual composta pelo valor tonal de cinza e quantidade de quadrados para cada time de futebol representado.

Com a apresentação destes exemplos e compostas as análises para cada um, podemos obter um panorama de métodos de representação da incerteza, dentro da visualização de dados, importantes para a continuidade do projeto em voga. A intenção, como mencionada anteriormente, é de compor uma base de visualizações que podem ser lidas conforme a teoria apresentada. Assim, podemos validar a importância de cada item específico para questionar e abranger os dados incertos como parte importante das visualizações.

4.5 Conclusões

Com base no estudo de análises dos subcapítulos anteriores, elaboramos o quadro 2 com a finalidade de parametrizar as visualizações de acordo com os conceitos apresentados até aqui. O quadro é dividido em categorização de incertezas (SPIEGELHALTER, 2017), categorização de dados (YAU, 2013) e representação visual da incerteza (PADILLA, 2020). Sua parte esquerda contém as características estudadas, enquanto a direita conta com a lista de exemplos, de acordo com a ordem que estes foram apresentados. Os exemplos que utilizam-se dos conceitos descritos no quadro são marcados com um círculo preto, indicando quais características estão presentes naquela visualização e N/A caso não se apliquem.

			Exemplo 1	Exemplo 2	Exemplo 3	Exemplo 4	Exemplo 5	Exemplo 6	Exemplo 7	Exemplo 8
Categorização de incertezas <small>(SPIEGELHALTER, 2017)</small>	Ontológica						●			
	Epistêmica		●	●						
	Aleatória				●	●		●	●	●
Categorização de dados <small>(YAU, 2013)</small>	Variável	Catégorica					●			
		Ordinal					●			
		Numérica	●	●	●	●			●	●
	Sistema de coordenadas	Cartesiano	●	●	●	●		N/A	●	●
		Polar						N/A		
		Geográfico					●	N/A		
	Escala	Linear						●		
		Percentual			●	●		●		●
		Logarítmica	●	●						
		Temporal contínua						●		●
		Temporal discreta								
	Marcadores visuais <small>(YAU, 2013 apud MACKINLAY, 1986)</small>	Posição	●	●	●		●	●	●	●
		Comprimento				●				
		Ângulo								
		Direção					●			
		Área								
		Volume								
Valor									●	
Cor		●	●	●	●	●	●	●	●	
Contexto	●	●	●	●	●	●	●	●		
Representação visual da incerteza <small>(PADILLA, 2020)</small>	Anotações gráficas		●	●	●	●				
	Codificação visual						●			
	Método híbrido							●	●	
	Não possui	●								

Quadro 2: Quadro de análises. Fonte: O autor, 2020.

O quadro acima foi criado em sistema binário, cruzando a base teórica apresentada com os exemplos do capítulo 4. Logo, tivemos de adaptar as análises anteriores, a primeira adaptação foi a inserção do contexto da visualização como “existente” ou “não existente”, já que não podemos colocar de maneira descritiva qual contexto, ou qual formato de contexto, foi aplicado a cada exemplo. Nas análises anteriores, incluímos elementos gráficos (círculos,

linhas, quadrados, etc) como marcadores visuais. Porém, no quadro, contemplamos apenas o estudo realizado por Yau (2013 apud MACKINLAY, 1986), onde os marcadores visuais são as características que modificam essas formas, como o valor, a posição e a direção desta.

Foi necessário que repensássemos a sessão de representação visual da incerteza, que havíamos abordado nas análises anteriores pelo formato (marcador ou anotação gráfica) que a incerteza estava sendo visualizada. O resultado no quadro foi uma configuração onde os exemplos se enquadram, ou não, em um dos grupos de representação visual: anotações gráficas, codificação visual e método híbrido. Como existem três grupos, e um dos exemplos não possui a representação visual da incerteza, propusemos o grupo “Não possui”, com o objetivo de tornar clara a falta desta representação.

Contudo, podemos observar que as visualizações que possuem qualquer grau de incerteza em seus dados, geralmente são compostas por características diferentes, mas com tendência a utilizar os mesmos métodos tradicionais de representação da incerteza (barras de erro, intervalo de confiança, etc). Já os exemplos que foram citados no subcapítulo 4.4 se diferem da maioria por possuírem uma característica singular de não representarem eventos comuns do cotidiano, como acontece em visualizações de dados sobre saúde, eleições e meteorologia.

Certamente, oito exemplos não abrangem todos os espectros de dados incertos sendo exibidos visualmente, mas nos trazem alguma base comparativa dentro de seus contextos. De tal forma, o projeto que será abordado no próximo capítulo conta com o desenvolvimento de três visualizações fictícias que utilizam-se de soluções de *design* para abordar tanto os tipos de incerteza presente nos dados, quanto o métodos visuais de representá-las.

5 DESENVOLVIMENTO

Após o fim da análise em que abordamos as categorizações de incertezas, características dos dados e os métodos de representar esses dois pilares visualmente, descreveremos as etapas do projeto de *design* desenvolvido. Para isso, elaboramos três visualizações de dados de contextos distintos e fictícios, tendo como público-alvo os

profissionais da área de *design* da informação que buscam desenvolver visualizações de dados mais assertivas e fidedignas à incerteza presente na realidade.

As três visualizações propostas partem de dados ficcionais sobre situações reais para cumprir a proposta de dar destaque a incerteza. Sendo o objetivo principal do produto dar suporte a pesquisa teórica realizada para esta monografia.

Cada visualização atua individualmente, sendo a primeira uma corrida eleitoral para a prefeitura da cidade de São Paulo, onde uma amostra da população conta como seriam os resultados caso votassem três meses antes da data de votação para primeiro turno. A segunda aborda o caso do deputado do PSL-RJ, Daniel Silveira, onde sua sentença de prisão foi votada pela câmara dos deputados. E, por último, temos uma visualização sobre o desaparecimento de dois marinheiros, baseada na história do pescador José Salvador Alvarenga, que ficou à deriva e foi encontrado quatorze meses depois de seu desaparecimento nas ilhas Marshall.

Nos próximos subcapítulos, iremos destrinchar o processo de produção das visualizações, descrevendo o escopo dos desenhos realizados, a criação da identidade visual e a implementação dos métodos de visualizar a incerteza em cada caso.

5.1 Visualizações incertas

Neste subcapítulo iremos abordar a última etapa antes da construção do produto final. Apresentaremos o processo de pesquisa, os tipos de incerteza que compuseram o produto e os recursos visuais que utilizamos para representá-lo.

É importante frisar, antes de darmos continuidade, que a incerteza epistêmica não foi representada. Esse tipo de incerteza depende da obtenção de mais informações sobre os dados que já estão inseridos na visualização para ser resolvida, o que pode dificultar o processo de representação por meio de marcadores visuais.

A incerteza epistêmica é melhor resolvida na visualização por meio da inserção de um contexto, ou seja, a adição de mais explicações sobre os dados faltantes, como por botões de informação ou *tooltips*. Essa é a melhor maneira de se resolver essa incerteza por conta da falta de dados que não poderiam ser representados visualmente, pois não foram quantificados no processo de pesquisa. Por exemplo, podemos ter uma estimativa de quantas pessoas foram diagnosticadas com SRAG em um gráfico que apresenta os óbitos por Covid-19, mas não

poderíamos quantificar quantos casos não foram contados por não terem sido testados antes do indivíduo ir à óbito.

Então, iremos apresentar três visualizações que apresentam dois tipos de incerteza: a incerteza aleatória e a ontológica. Nesses casos os dados estão quantificados e podem ser visualizados, utilizando como referência o mapeamento da incerteza que abordamos nos capítulos anteriores.

5.1.1 Eleições para prefeitura de São Paulo

A primeira visualização demonstra o evento das eleições para a prefeitura de São Paulo. As eleições ocorreram em dezembro de 2020 no Brasil, e os dados utilizados para este projeto foram de uma pesquisa realizada pelo IBOPE e apresentada, primeiramente, pelo G1 (G1, 2020).

Para os rascunhos procuramos uma solução mais comum ao cotidiano, como por exemplo um gráfico de barras que demonstra a tendência de votação para cada candidato. A intenção de trazer um gráfico de “uso comum” foi trazer a referência da realidade para um produto que fala sobre incerteza em visualizações de dados. Este tipo de gráfico (barras ou linha com margens de erro) é muito utilizado, e exemplificado por jornais locais nos meses que antecedem as eleições. Não faria sentido o desconectar de um contexto já consolidado porque, fazendo isso, podemos levar confusão aos leitores ou atrasar sua interpretação.

Esses dados contêm um exemplo de incerteza aleatória, por se tratar de um resultado parcial. Depois de alguns testes feitos no papel, chegamos ao protótipo de um gráfico de linhas com uma margem de confiança, mas com algumas alterações em vista de deixar evidente que se trata de uma pesquisa, e não a definição de um futuro.

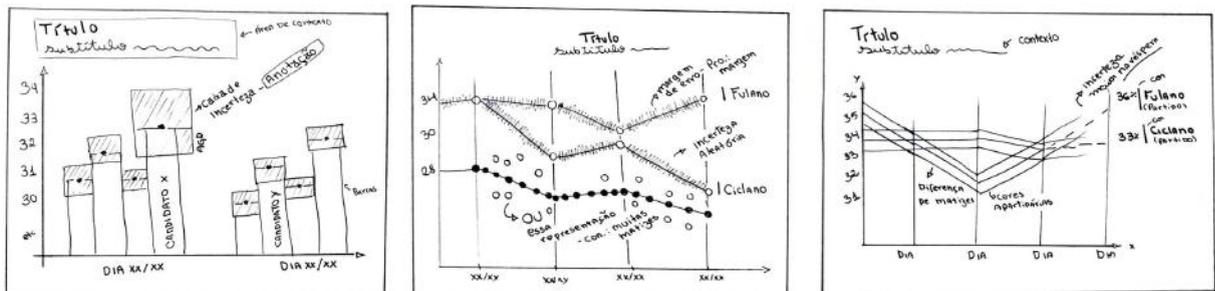


Figura 37: Rascunhos - Eleições para prefeito de São Paulo . Fonte: O autor, 2021.

Na pesquisa realizada pelo IBOPE, utilizada como referência para essa visualização, havia uma grande probabilidade de empate técnico entre três dos candidatos, o que poderia alterar os resultados do segundo turno. Segundo o próprio IBOPE, é considerado empate técnico quando a diferença entre os candidatos se encontra dentro das margens de erro das pesquisas, ou seja, quando há superposição dos respectivos intervalos de confiança dos candidatos (POVO, 2020).

As estimativas feitas na Figura 37 contam apenas com o primeiro turno das eleições e foram desenvolvidas para representar o primeiro método de representação visual da incerteza proposto por Padilla (2020): a inserção da incerteza por meio de uma anotação gráfica. Procuramos, então, aprimorar a inserção dessa anotação gráfica, a diferenciando, visualmente, dos métodos tradicionais propostos pela estatística de se confeccionar uma distribuição.

Seguem alguns detalhes técnicos sobre os dados: A margem de erro é de 3 pontos percentuais para cima ou para baixo, ou seja, se um candidato estiver marcado com 13% em determinado dia da pesquisa, esse resultado pode não ser apenas 13%, mas variar entre 16% e 10% das intenções de voto; 1.204 eleitores foram ouvidos; podendo ser depositada uma confiança de 95% nos resultados parciais para essa eleição.

5.1.2 Prisão do deputado Daniel Silveira

A segunda visualização é contextualizada em 2021, quando o deputado federal Daniel Silveira foi acusado de fake news, defesa do AI-5 e críticas aos ministros do Supremo Tribunal Federal. Sua prisão foi colocada em votação na Câmara dos Deputados por divergências na lei que poderia considerar o caso como ataque à liberdade de expressão (CAMARA, 2021). A votação ocorreu anteriormente ao término do presente projeto, portanto os votos já estavam definidos e os dados já estavam disponíveis. No entanto, ocorreram alguns casos de traição partidária que são considerados uma incerteza aleatória, pois é inerente ao processo de uma escolha individual, que poderia se alterar de acordo com afinidades ou forças externas.

É de bom senso que cada partido aconselhe seus deputados nas votações para irem contra ou a favor do caso, dependendo do objetivo partidário. Porém, como dito anteriormente, as escolhas individuais poderiam mudar o curso da câmara (CONGRESSO EM FOCO, 2021). Com isso, a visualização de dados foi pensada para que se apresentasse uma tendência sobre o rumo de Daniel Silveira, anterior à votação que de fato ocorreu no Rio de Janeiro.

Após alguns rascunhos, apresentados pela figura 38, foi definida a utilização de uma sequência de ícones. Assim, foi possível demonstrar cada voto individualmente e alterar marcadores específicos para a incerteza. Apresentando, dessa forma, o segundo método de representação visual proposto por Padilla (2020): a codificação visual da incerteza.

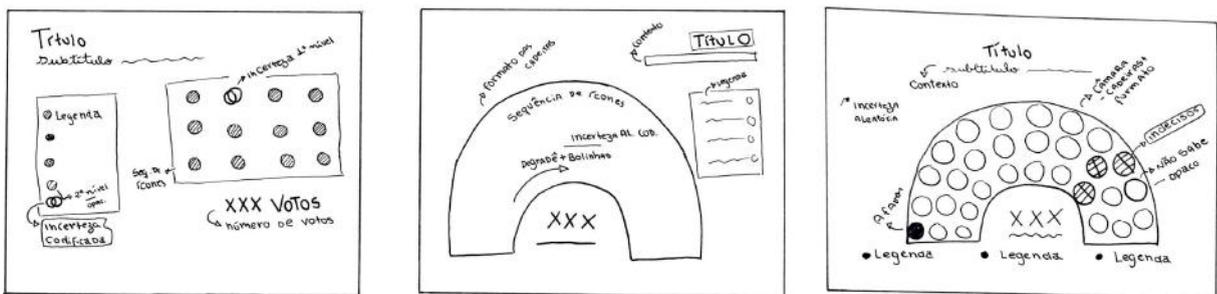


Figura 38: Rascunhos - Prisão do deputado Daniel Silveira. Fonte: O autor, 2021.

A princípio, tivemos o desafio de desenhar todos os votos. Foram mais de 400 deputados e a solução encontrada foi de aproximar a sequência de ícones da realidade: a contornando no formato da Câmara dos Deputados, ao invés de representá-la em um molde retangular. Deste modo, aproximamos novamente a visualização da realidade a colocando em um formato que se associa ao que está sendo representado, levando contexto ao leitor.

Por fim, cada ícone pode representar um voto individual de cada deputado, sendo agrupados nas categorias: “contra”, “a favor”, “indecisos”, “possível traição partidária a favor” e “possível traição partidária contra”.

5.1.3 O barco perdido

A terceira visualização conta a história de um barco que se perdeu no caminho de São Paulo a Alagoas. Dois homens da marinha teriam saído do Iate Clube de Itanhaém e pedido socorro próximo ao litoral do Rio de Janeiro. Nessa história, uma tempestade teria os atingido, fazendo que perdessem qualquer equipamento de geolocalização para se comunicarem com terra firme.

O contexto desta visualização foi baseada no caso do pescador José Alvarenga, que teria passado quatorze meses à deriva por quebra de seus equipamentos. A incerteza que está intrínseca à visualização é a incerteza ontológica, pela falta de precisão que podemos ter sobre as ações da natureza.

Foram feitos alguns rascunhos, presentes na figura 39, para contextualizar os dados que seriam demonstrados. A princípio, tínhamos a ideia de mapear a possibilidade de locais pelos quais o barco poderia ter se dirigido, de acordo com dados como do vento e da superfície marítima. Mas, como a incerteza está justamente na falta desta precisão, descartamos o caso para não compor dados inexistentes e desvirtuar a história. Foram feitos testes com interatividade para compor três telas distintas de representação, mas a ideia também foi descartada por acreditarmos que o produto deveria ter apenas uma interface por visualização.

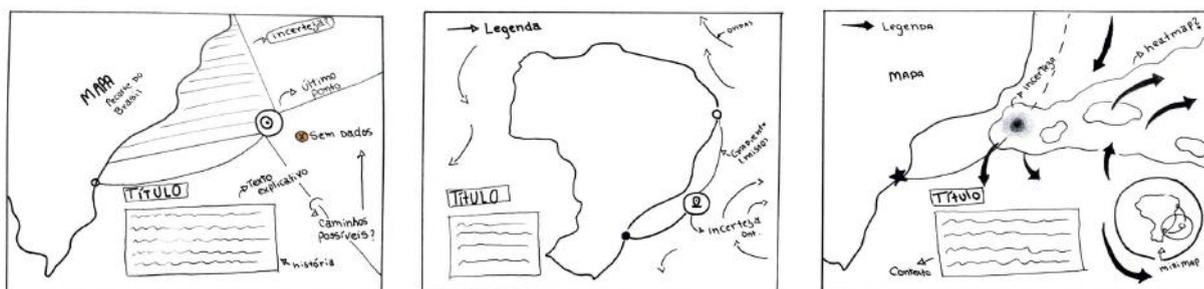


Figura 39: Rascunhos - O barco perdido. Fonte: O autor, 2021.

Foram definidos que os dados representados seriam: a última localização da dupla, obtida pelo rastreamento da ligação telefônica, mas que possui uma margem de erro de 10km

de diâmetro próximo ao litoral do Rio de Janeiro; e também os dados sobre as correntes marítimas, para deixar a interpretação à escolha do leitor sobre o desaparecimento do barco. Por fim, a visualização utilizou um sistema cartográfico de coordenadas para localizar geograficamente a história sobre como o mar poderia ter levado os marinheiros à deriva.

5.2 Identidade visual

O presente subcapítulo aborda a identidade visual das visualizações de dados, como a implementação de suas cores, o pensamento por trás das tipografias e parte do estudo semântico produzido para colocar em destaque alguns padrões e variáveis consolidadas na visualização da incerteza. O objetivo da elaboração de uma identidade visual é promover consistência ao projeto, fazendo com que cada parte individual componha uma família facilmente reconhecível por suas características.

Para iniciar o mapeamento da identidade visual, foi composto um painel semântico (Figura 40) com diferentes modos de representar a incerteza, assim como imagens de parte de estudos realizados por outros autores. O objetivo do painel foi identificar pontos comuns entre os objetivos deste projeto e os objetivos de projetos semelhantes, traçando relações de concordância e absorvendo características visuais principais como a utilização de cores, formas e tipografias.

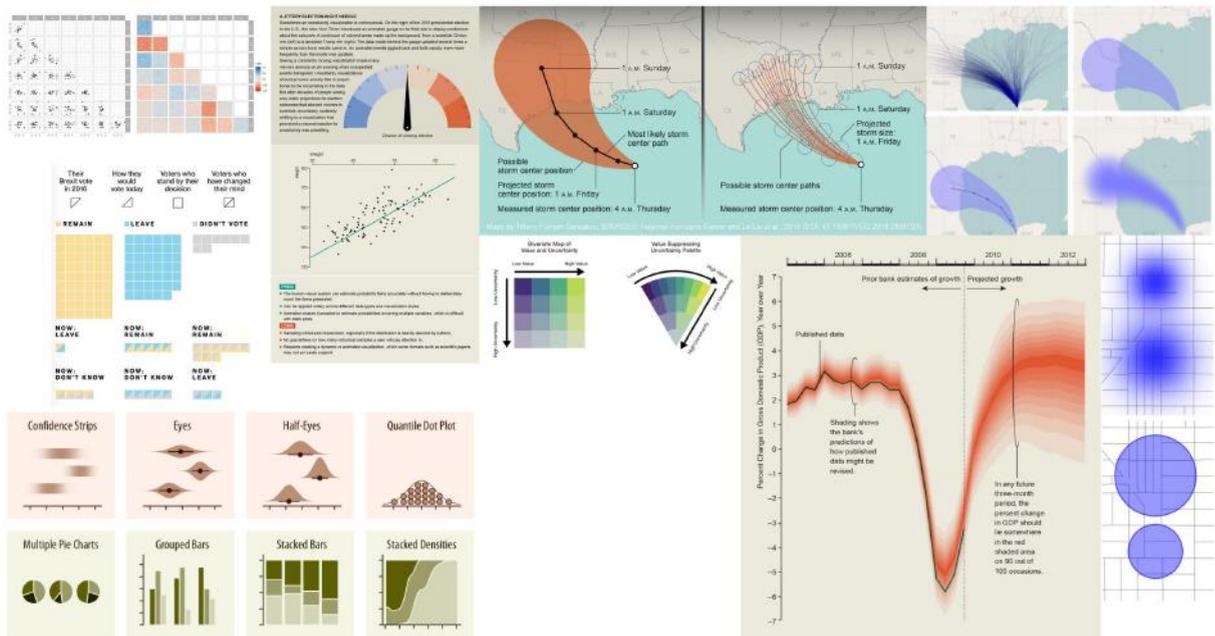


Figura 40: Painel semântico. Fonte: O autor, 2020.

Sendo assim, pudemos notar de acordo com as análises do capítulo 4 e a produção do painel semântico da figura 40, as seguintes características de *design*:

- Exploração da ferramenta gradiente;
- Diferenças na opacidade dos marcadores visuais;
- Uso de cores neutras;
- Contraste entre cores análogas, mesmo que neutras;
- Variação de uma ou duas famílias tipográficas sem serifa;
- Informações adicionais sobrepostas à visualização como anotações ou contextualizadas em forma de legenda.

Sendo assim, algumas dessas características foram aproveitadas para a produção das visualizações e outras foram descartadas, dado que não promoveriam consistência ou identificação.

A partir do desenvolvimento do painel e a captura das principais correlações entre as imagens, foi elaborada uma série de paletas de cores que tinham em comum a utilização de meio-tons e cores neutras. As matizes foram retiradas de algumas partes do painel, e foram

pouco modificadas posteriormente conforme a elaboração das visualizações. As seleções de cores deveriam seguir os seguintes princípios: (a) se comportarem como cores neutras, para não se tornarem partidárias, e (b) possuírem ao menos um par de cores complementares para gerar contraste à visualização e promover destaque. Sendo assim, foi composta a paleta presente na figura 41, reunindo os princípios apresentados anteriormente e sendo parte estrutural das visualizações de dados.



Figura 41: Paleta de cores. Fonte: O autor, 2020.

A paleta de cores selecionada foi validada em um simulador digital de daltonismo (NICHOLS, 2021), para assim observarmos questões como contraste entre figura e fundo, destaques para os dados incertos e concordância entre as escalas tonais. Observamos na figura 42 algumas cores que podem não ser tão contrastantes, portanto evitamos codificações de dados divergentes para não causar confusão. Cada visualização conta com, no máximo, quatro cores, assim podemos explorar a paleta e produzir diferentes combinações desta.

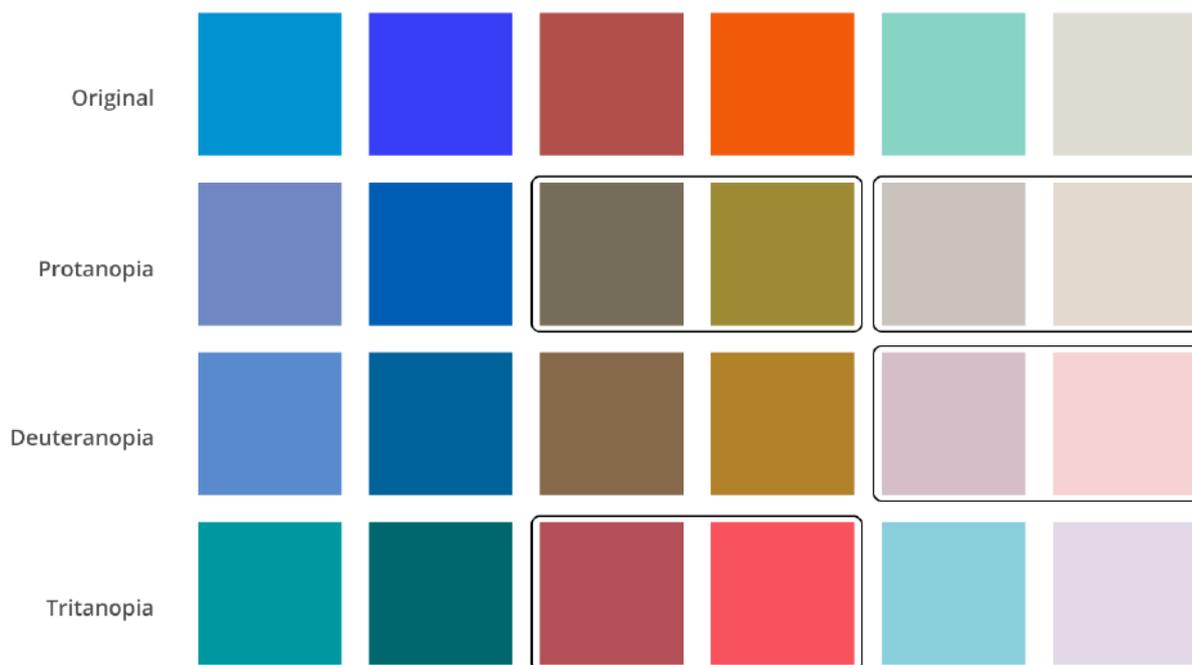


Figura 42: Simulador de daltonismo. Fonte: O autor, 2020.

Definida a paleta de cores do projeto, estudamos qual família tipográfica seria flexível o suficiente para se encaixar em todas as partes. Os critérios de escolha foram baseados em: (a) tipografias que não possuem serifas, para que dialoguem melhor com o ambiente da *internet*; (b) famílias que possuíam grande variação de tipos para que pudessem compor títulos, subtítulos, legendas e outros.

Portanto, como podemos observar na figura 43, foram selecionadas as tipografias Lato, Roboto, Open Sans, Helvetica Neue, Din Alternate e Geneva. Tanto a tipografia Din Alternate quanto a Geneva não obedeciam o segundo critério de possuírem uma família tipográfica extensa, por isso foram descartadas. Já a Lato possui problema em seu *Kerning* da letra T, que se "junta" a qualquer vogal quando regular ou *light*.

Lato

abcdefghijklmnopqrstuvwxy
z
0123456789
.,;?!@#%&:"

Roboto

abcdefghijklmnopqrstuvwxy
z
0123456789
.,;?!@#%&:"

Open Sans

abcdefghijklmnopqrstuvwxy
z
0123456789
.,;?!@#%&:"

Helvetica Neue

abcdefghijklmnopqrstuvwxy
z
0123456789
.,;?!@#%&:"

Din Alternate

abcdefghijklmnopqrstuvwxy
z
0123456789
.,;?!@#%&:"

Geneva

abcdefghijklmnopqrstuvwxy
z
0123456789
.,;?!@#%&:"

Figura 43: Estudo tipográfico. Fonte: O autor, 2020.

Com essa análise de critérios, definimos a utilização das famílias tipográficas Open Sans e Helvetica Neue (Figura 44), que podem ser utilizadas para fins acadêmicos. Caso o projeto saia do escopo acadêmico, a Helvetica Neue será substituída ou comprada para uso comercial. Essas fontes são geométricas e sem ornamentos, para assim promover legibilidade e não tirem a atenção dos marcadores visuais.

Títulos e legendas

Open Sans

abcdefghijklmnopqrstuvwxy
z
0123456789
.,;?!@#%&:"

Subtítulo e corpo de texto

Helvetica Neue

abcdefghijklmnopqrstuvwxy
z
0123456789
.,;?!@#%&:"

Figura 44: Tipografia. Fonte: O autor, 2020.

Traçadas as características visuais principais do projeto, este tomou forma e ganhou uma identidade necessária para que ocorra congruência entre as visualizações.

5.3 Resultado final

O resultado do projeto foi fiel à pesquisa apresentada, trazendo referência dos estudos abordados nos capítulos anteriores. As visualizações podem ser utilizadas tanto individualmente, quanto em conjunto, servindo como suporte à compreensão do que são

dados incertos, em plataformas diferentes de publicação (*webview*, impressão, dispositivos móveis, entre outros).

A falta de interatividade ou animações visa demonstrar que os dados incertos representados podem ser fechados em imagens estáticas, como boa parte dos exemplos trazidos aqui. Assim como a representação visual desses dados não necessita, obrigatoriamente, do acréscimo de métodos tradicionais da estatística, como inserção de barras de erros ou intervalos de confiança. As visualizações exploram as diversas formas de promover a compreensão da incerteza e correspondem às expectativas.

Seguem abaixo as imagens das visualizações de dados finais do projeto:

Eleições para prefeito de São Paulo

Veja o resultado parcial do primeiro turno dessa corrida eleitoral.

Fonte: Instituto Brasileiro de Opinião Pública e Estatística (IBOPE)

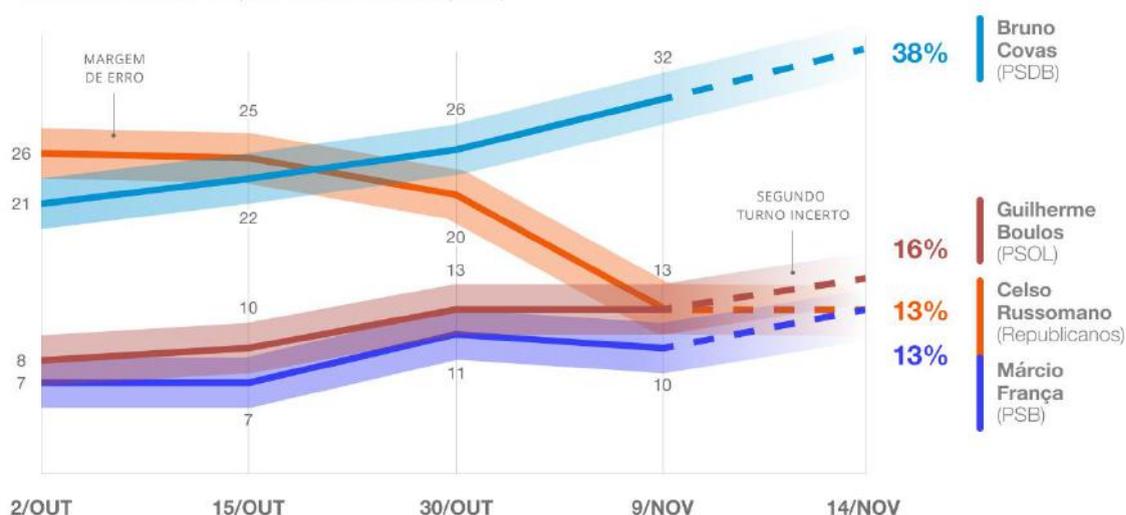


Figura 45: Eleições para prefeito de São Paulo . Fonte: O autor, 2021.

Prisão do deputado Daniel Silveira

Entenda a expectativa de votos pela câmara dos deputados sobre a acusação de fake news, defesa do AI-5 e críticas aos ministros do supremo, do caso do deputado do Partido Social Liberal do Rio de Janeiro.

Fonte: Congresso em Foco

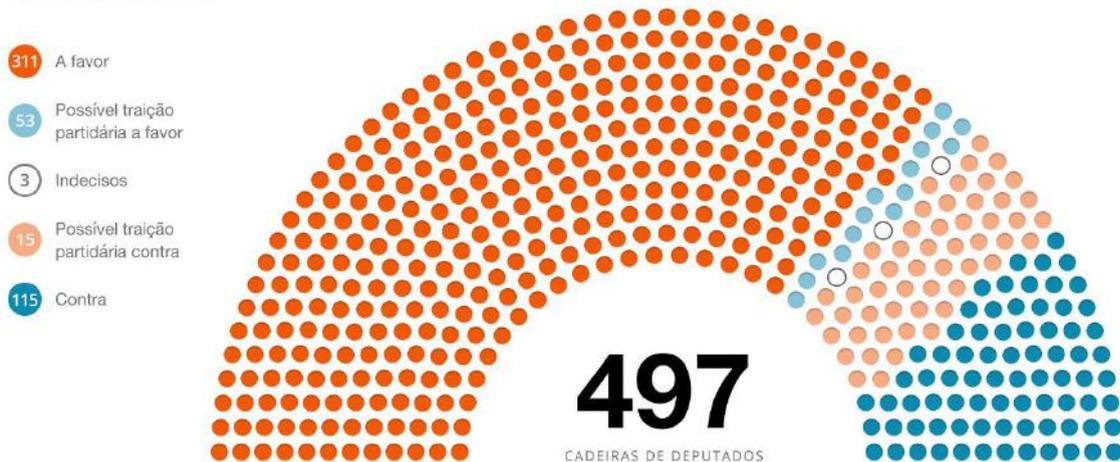


Figura 46: Prisão do deputado Daniel Silveira. Fonte: O autor, 2021.



Figura 47: O barco perdido. Fonte: O autor, 2021.

6 CONSIDERAÇÕES FINAIS

O tema incerteza surgiu da reflexão sobre os gráficos que circulavam na internet durante a pandemia. Traçar um paralelo entre a incerteza e a visualização de dados foi um desafio. Para isso, foi necessário promover uma pesquisa teórica sobre uma área pouco abordada pelo campo do design, já que se encontra relacionada majoritariamente à estatística.

Além disso, as pesquisas que estudam o uso de métodos visuais para representar a incerteza são recentes e feitas por pesquisadores de áreas correlatas à estatística. Até então, não existe um método implementado que tenha sido validado para testar a eficácia dos marcadores visuais que representam dados incertos. Portanto, acredito que este projeto futuramente possa ser testado com usuários e estudado por diferentes pesquisadores, para, assim, promover mais assertividade sobre o que poderá ser implementado na visualização de dados, partindo do conhecimento sobre os tipos de incerteza.

Foi complicado, como investigadora, fechar o escopo. Inicialmente, as ideias pareciam abstratas e abertas. Surgiam perguntas sobre como esse projeto poderia contribuir para o *design* da informação. Para concluir meus objetivos, passei por um intenso processo de pesquisa sobre a incerteza, selecionando recortes que interessassem ao *design* e, por fim, esquematizando completamente a parte teórica (parte dela no Miro, no papel e no Figma).

O desenvolvimento desse projeto requisitou esforço, tempo e paciência. Foram várias noites em claro e mudanças de ideia; a combinação de tudo o que eu tinha estudado até então e a busca pelo que eu ainda não entendia muito bem. Tive que ter foco, organização e ainda passar por todas as fases de mudança na minha vida que ocorreram no processo.

Em conclusão, espero ainda poder trabalhar mais com o que comecei a criar, e também auxiliar *designers* a pensarem e planejarem seus projetos contando com a incerteza. Dessa forma, espero que a incerteza passe a ser considerada nos métodos de produção de visualização de dados, assim como pelos leitores desses trabalhos.

7 REFERÊNCIAS

BERTIN, J. **Semiology of Graphics: diagrams, networks, maps**. California: Esri Press, 2010.

CAIRO, A. **Uncertainty and graphicacy How should statisticians, journalists, and designers reveal uncertainty in graphics for public consumption?**. p. 12, [s.d.] Disponível em:

<<https://ec.europa.eu/eurostat/cros/powerfromstatistics/OR/PfS-OutlookReport-Cairo.pdf>>.

Acesso em: 6 de mar. 2021.

CAIRO, A. **‘Make it visual and people will understand’**. Disponível em:

<<https://www.maartenlambrechts.com/2015/05/27/alberto-cairo-make-it-visual-and-people-will-understand.html>>. Acesso em: 15 out. 2020.

CAMARA. **Câmara inicia sessão para decidir sobre prisão do deputado Daniel Silveira - Notícias..** Disponível em:

<<https://www.camara.leg.br/noticias/729146-camara-inicia-sessao-para-decidir-sobre-prisao-do-deputado-daniel-silveira/>>. Acesso em: 28 mar. 2021.

CARD, S.K.; MACKINLAY, J.D.; SHNEIDERMAN, **Readings in information visualization: using vision to think**. San Francisco, Calif: Morgan Kaufmann Publishers, 1999. (The Morgan Kaufmann series in interactive technologies).

CHALABI, M. **Três modos de identificar uma estatística ruim, [s.d.]**. Disponível em:

<https://www.ted.com/talks/mona_chalabi_3_ways_to_spot_a_bad_statistic?language=pt-br>.

Acesso em: 6 mar. 2021

CONGRESSO EM FOCO. **68 deputados traíram partidos em votação de Daniel Silveira**. Disponível em:

<<https://congressoemfoco.uol.com.br/legislativo/deputados-contra-partidos-daniel-silveira/>>. Acesso em: 28 mar. 2021.

DE LAPLACE, P.S. **Théorie analytique des probabilités**. [s.l.]: Imprimerie royale, 1812. (Oeuvres de Laplace). Disponível em:

<<https://books.google.com.br/books?id=yewSAQAAMAAJ>>.

FRENCH, S. Cynefin: Uncertainty, small worlds and scenarios. **Journal of the Operational Research Society**, v. 66, 2015.

G1. **Pesquisa Ibope em São Paulo, votos válidos: Covas, 38%; Boulos, 16%; Russomanno, 13%; França, 13%. Eleições 2020 em São Paulo.** 2020 Disponível em: <<https://g1.globo.com/sp/sao-paulo/eleicoes/2020/noticia/2020/11/14/pesquisa-ibope-em-sao-paulo-votos-validos-covas-38percent-boulos-16percent-russomanno-13percent-franca-13percent.ghml>>. Acesso em: 28 mar. 2021.

HULLMAN, J. **The Visual Uncertainty Experience - Jessica Hullman.** [s.l.: s.n., s.d.]. Disponível em: <<https://www.youtube.com/watch?v=pTVAn4oLvbc&t=145s>>. Acesso em: 1 nov. 2020.

HULLMAN, J.; RESNICK, P.; ADAR, E. **Hypothetical Outcome Plots Outperform Error Bars and Violin Plots for Inferences about Reliability of Variable Ordering.** PLOS ONE, v. 10, n. 11, p. e0142444, 16 nov. 2015.

LGPD. **Lei Geral de Proteção de Dados**, 3 set. 2020. Disponível em: <<https://www.gov.br/defesa/pt-br/aceso-a-informacao/lei-geral-de-protecao-de-dados-pessoais-lgpd/lei-geral-de-protecao-de-dados-lgpd-1>>. Acesso em: 21 mar. 2021.

MACKINLAY, Jock. Automating the design of graphical presentations of relational information. **ACM Transactions on Graphics**, v. 5, n. 2, p. 110–141, 1986. Disponível em: <<https://dl.acm.org/doi/10.1145/22949.22950>>. Acesso em: 28 mar. 2021.

MONMONIER, M. **Mapping It Out.** The University Of Chicago Press, 1995.

NEURATH, O. **International Picture Language.** Reading: The University of Reading Press, 1980.

NICHOLS, D. **Coloring for Colorblindness.** Disponível em: <<http://www.davidmathlogic.com/colorblind/>>. Acesso em: 21 mar. 2021.

PADILLA, L.; KAY, M.; HULLMAN, J. **Uncertainty Visualization.** [s.l.]: PsyArXiv, 2020. Disponível em: <<https://osf.io/ebd6r>>. Acesso em: 3 set. 2020.

PORTAL ACTION. **2 - Distribuições Amostrais - Inferência.** Portal Action. Disponível em: <<http://www.portalaction.com.br/inferencia/distribuicoes-amostrais>>. Acesso em: 17 jan. 2021.

POVO, O. **Eleições 2020: entenda o que é um empate técnico nas pesquisas.** Política, 28 out. 2020. Disponível em: <<https://www.opovo.com.br/noticias/politica/2020/10/28/eleicoes-2020--entenda-o-que-e-um-empate-tecnico-nas-pesquisas.html>>. Acesso em: 2 abr. 2021.

SANTORO, C. **According to my calculations, yesterday could have been sweater weather in St. Louis. But given that it was 93 and sunny, I guess there's a bit of #uncertainty to account for. #SWDchallenge #DataViz #SweaterWeather** <https://t.co/GppNTUqHlo@ClaireESantoro>, 2 jan. 2019. Disponível em: <<https://twitter.com/ClaireESantoro/status/1172677318852169728/photo/1>>. Acesso em: 6 mar. 2021

SIGNIFICADOS. **Significado de Intervalo de Confiança (O que é, Conceito e Definição).** Significados. Disponível em: <<https://www.significados.com.br/intervalo-de-confianca/>>. Acesso em: 17 jan. 2021.

SPIEGELHALTER, D. Risk and uncertainty communication, **Annual Review of Statistics and Its Application**, vol. 4, pp. 31–60, 2017.

THOMSON, J.; HETZLER, E.; MACEACHREN, A.; *et al.* A typology for visualizing uncertainty. *In*: ERBACHER, Robert F.; ROBERTS, Jonathan C.; GROHN, Matti T.; *et al* (Orgs.). San Jose, CA: [s.n.], 2005, p. 146. Disponível em: <<http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.587254>>. Acesso em: 1 nov. 2020.

TUFTE, E.R., 1942-. **The Visual Display of Quantitative Information.** - 1942. Cheshire, Conn. :Graphics Press, 2001.

WHITAKER, R., MIRZARGAR, M., and kirby, R., “**Contour Boxplots: A Method for Characterizing Uncertainty in Feature Sets from Simulation Ensembles**”, 2014.

WILKE, C.O. **Fundamentals of Data Visualization.** [s.l.: s.n., s.d.]. Disponível em: <<https://clauswilke.com/dataviz/visualizing-uncertainty.html>>. Acesso em: 7 nov. 2020.

YAU, N. **Visualizing the Uncertainty in Data FlowingData**, 8 jan. 2018. Disponível em: <<https://flowingdata.com/2018/01/08/visualizing-the-uncertainty-in-data/>>. Acesso em: 10 ago. 2020.

YAU, N. **Data Points: Visualization That Means Something**. - 2013. John Wiley & Sons, Inc., Indianapolis, Indiana

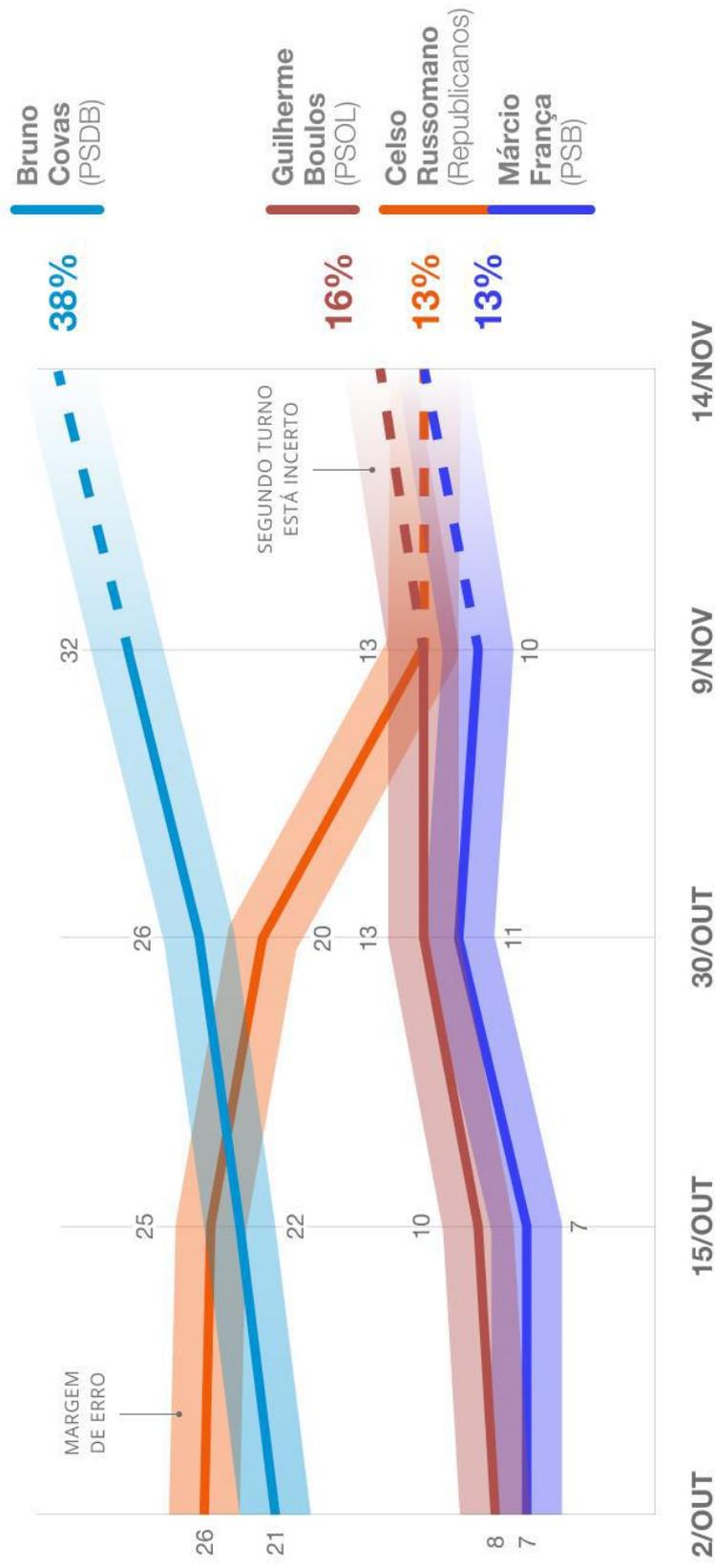
YU, Y. **Visualizing uncertainty : what we talk about when we talk about uncertainty**. Northeastern University, 2018. Disponível em: <<http://hdl.handle.net/2047/D20290236>>. Acesso em: 5 nov. 2020.

8 ANEXOS

Eleições para prefeito de São Paulo

Veja o resultado parcial do primeiro turno dessa corrida eleitoral.

Fonte: Instituto Brasileiro de Opinião Pública e Estatística (IBOPE)



Prisão do deputado Daniel Silveira

Entenda a expectativa de votos pela câmara dos deputados sobre a acusação de fake news, defesa do AI-5 e críticas aos ministros do supremo, do caso do deputado do Partido Social Liberal do Rio de Janeiro.

Fonte: Congresso em Foco

