



Universidade de Brasília
Instituto de Ciências Exatas - IE
Departamento de Estatística - EST

Visualização de deslocamentos espaciais usando mapas e vídeos

Aïtchéou Gauthier Zountchégnon
Thomaz Santana

Projeto apresentado para obtenção do título
de Bacharel em Estatística

Aïtchéou Gauthier Zountchégnon e Thomaz Santana

Visualização de deslocamentos espaciais usando mapas e vídeos

Orientador:
Prof. Guilherme Souza Rodrigues

Projeto apresentado para obtenção do título
de Bacharel em Estatística

Brasília
2020

Agradecimentos

- **Gauthier**

À minha família, que é a minha fonte de motivação, que mesmo longe se esforça para me ajudar de todas as formas possíveis.

Ao meu orientador, Prof. Guilherme, que me ensinou bastante coisas durante esse TCC, e teve a paciência e a sabedoria necessária para orientar esse trabalho.

A todos os meus Professores do departamento, pelos ensinamentos, pelas ajudas nas matérias e pelas motivações.

Aos meus superiores de estágio no TSE, Felipe, Josian e Thaís, que me ajudaram a aprofundar os meus conhecimentos em programação, e pelos conselhos no dia a dia de trabalho.

Aos meus colegas, Gabriel, Letícia, Rodrigo, pelas ajudas e bons momentos durante a graduação. Graças a eles, eu não tive uma vida de solidão na UnB.

À Luana Carvalho, que me ajudou na realização desse TCC.

Ao meu colega Thomaz pela parceria na realização desse TCC.

A todas as pessoas que me ajudaram em algum momento na minha graduação.

- **Thomaz**

Diante de uma longa jornada, gostaria de agradecer às pessoas que foram fundamentais na realização desse projeto e que estiveram ao meu lado nos momentos difíceis e nos momentos alegres.

À minha família, que acreditaram e me apoiaram no decorrer dessa caminhada, principalmente aos meus pais, que me deram todas as condições para realização deste trabalho.

À Gabi, minha namorada, que me ajudou e incentivou em todos os momentos, me mantendo sempre focado em buscar o melhor.

Ao Gauthier, com quem tive o prazer de compartilhar essa experiência.

Ao Departamento de Estatística, professores e funcionários, que sempre estiveram disponíveis para os alunos.

Um agradecimento especial ao meu orientador Prof. Guilherme, com quem aprendi bastante nesse tempo de convivência. Sempre muito dedicado e solícito. Sem dúvidas, um dos melhores profissionais com quem trabalhei.

O meu muito obrigado à todos vocês.

Resumo

Visualização de dados é uma área em crescimento da Ciência de Dados que desenvolve técnicas de representar dados brutos de forma visual, de maneira que as relações entre os mesmos sejam melhor entendidas ou que novas informações possam ser percebidas. Este trabalho apresenta, em primeiro lugar, alguns conceitos importantes sobre a visualização de dados. Em seguida, descrevemos o processo de construção de mapas de conexão, usando o *software* R, por meio de duas aplicações: uma referente ao fluxo de bicicletas compartilhadas do programa *+BIKE* e outra sobre a chegada de estrangeiros à UnB. Os resultados foram apresentados em mapas estáticos, dinâmicos (vídeos) e em *dashboards* interativos.

Palavras-Chave: Visualização de dados. Ciência de Dados. Representações visuais. Dados brutos. Mapa de conexão. Software R. Mapas estáticos. Vídeos. Dashboard interativo.

Abstract

Data visualization is a fast-developing area of Data Science that develops techniques to visually represent raw data in a way that the relationships between them are better understood or that new information can be perceived. This work presents, in the first place, some essential concepts of data visualization. Later, we describe the process of creating connection maps using the software R through two applications: the flow of shared bicycles of the program *+BIKE* and the arrival of foreigners at Unb. The results are presented in static and dynamic (videos) maps and an interactive dashboard.

Key Words: Data visualization. Data Science. Connection map. Software R. Static maps. Videos. Interactive dashboard.

Conteúdo

1	Revisão Literária	7
1.1	Introdução	7
1.1.1	Visualização de dados	8
1.1.2	Conceito de tomada de decisão	9
1.2	Histórico	10
1.2.1	História	10
1.2.2	Visualização nos dias atuais	11
1.3	Aspectos estéticos e cognitivos	11
1.3.1	Passos para fazer uma visualização	13
1.3.2	Cérebro x Visualização	14
1.3.3	Aspectos que ajudam uma boa visualização	14
1.4	Visualização de dados espaciais	17
1.4.1	O que são dados espaciais	17
1.4.2	Tipos de dados	17
1.5	Imagens e vídeos	18
1.6	Frameworks de visualização de dados no R	21
1.6.1	Software R	21
1.6.2	Pacotes	21
1.7	Outras ferramentas de visualização de dados	21
2	Metodologia	24
2.1	Tráfego de bicicletas em Brasília	24
2.1.1	Dados do +Bikes	24
2.1.2	Dados das Rotas	25
2.1.3	Desenvolvimento	26
2.2	Chegada de Estrangeiros à UnB	28
2.2.1	Dados do sistema da UnB	28
2.2.2	Desenvolvimento	30
3	Resultados	34
3.1	Tráfego de bicicletas em Brasília	34
3.2	Chegada de Estrangeiros à UnB	38
4	Conclusão	41

Lista de Figuras

1	Campanha Russa de Napoleão 1812.	10
2	Processo de visualização de dados	13
3	Visualização da Comunidade LGBT	15
4	Exemplo de pontos	17
5	Exemplo de linhas	18
6	Exemplo de curvas	18
7	Imagens e vídeos 1	19
8	Imagens e vídeos 2	19
9	Imagens e vídeos 3	20
10	(a) Python (b) Power BI (c) Tableau (d) SAS	23
11	Imagem da NASA	31
12	Imagem Vetorizada	31
13	Mapa dos Estrangeiros	32
14	Mapa_Melhorado dos Estrangeiros	32
15	Mapa das Bicicletas	34
16	Mapa_Final das Bicicletas	35
17	Sequências do Video	36
18	Dashboard das Bicicletas	37
19	Mapa_Final dos estrangeiros	38
20	Visualização por Continente	39
21	Sequências do Gif	40

Lista de Tabelas

1	Dados das Bicycletas	24
2	Distribuição por Sexo	24
3	Dados dos Estrangeiros	28
4	Distribuição por Ano	29
5	Pontos a serem Representados	30

1 Revisão Literária

1.1 Introdução

Dados estão em todos os lugares. Há milênios as pessoas os juntam, processam e tentam entendê-los. Hoje em dia, grandes volumes de dados são gerados todos os segundos e quantidades enormes deles são coletados, armazenados e acessados pelas empresas de forma ágil para fins de visualização e análise com o intuito de obter de informações valiosas, visto que os dados só tem valor quando ajudam a resolver problemas ou tomar decisões importantes.

Alguns aspectos dos dados podem ser percebidos somente por meio de uma representação gráfica. Segundo Kirk (2012) e Silva (2019), a representação e a apresentação de dados exploram nossa capacidade de discernimento visual com a finalidade de aumentar nosso entendimento. Assim, o poder da visualização de dados está em sua capacidade de transformar dados brutos e abstratos em representações visuais de fácil compreensão. Uma visualização bem feita consegue transformar números ou palavras em histórias. Uma visualização feita a partir da análise tem a capacidade não só de exaltar vários pontos valiosos a serem observados, como também tem a possibilidade de estragar um trabalho que pode ter custado muitos esforços. “Neste contexto a visualização de dados representa o imaginário, reproduz a realidade e suas simulações, e trazem apoio para informações textuais. As capacidades comunicacionais e técnicas se juntam e se reforçam mutuamente, tornando possível a tríade Informação/Comunicação/Técnica” (Silva, 2019). O objetivo é auxiliar no entendimento de uma determinada situação que, sem uma visualização, ficaria difícil de entender (Oliveira et al., 2018). Em algumas situações, a visualização ajuda a extrair novas informações “escondidas” nos dados brutos.

Ao longo dos anos, com o desenvolvimento das ferramentas computacionais, as técnicas foram se aperfeiçoando e a visualização de dados se tornou mais fácil e mais utilizada. Ela vem tomando um lugar de destaque entre as principais formas de comunicação, pois ter uma representação de qualidade pode ser decisivo na tomada de decisão e isso a torna uma ferramenta fundamental no cenário mundial de grandes bancos de dados e “informações confusas”.

As visualizações de dados, que antigamente eram feitas à mão passaram a ser produzidas majoritariamente pela ferramenta *Excel* por conta de sua acessibilidade e da sua facilidade de uso. Considerando a quantidade de dados produzida pela atividade da Internet e o número crescente de sensores no ambiente como satélites e câmeras de tráfego, ferramentas mais poderosas foram desenvolvidas e lançadas todos os dias para atender a área de visualização que está em constante evolução. Atualmente, existem várias ferramentas disponíveis, de forma gratuita ou paga, que permitem a visualização estática, interativa, ou dinâmica dos dados. Com a grande quantidade de ferramentas disponíveis, é difícil reconhecer quais são as mais adequadas e eficazes para usos e necessidades específicas. Todos esses recursos de visualização de dados conferem mais capacidade de interação e de compreensão dos dados, o que torna a visualização de dados uma das áreas mais importantes da Ciência de Dados. Alguns *softwares* mais usados hoje em dia são: *R*, *Python*, *PowerBI*, *Tableau* e *Excel*.

No percurso deste trabalho será usado o *software* *R*, que é um ambiente de programação livre, para realizar gráficos e cálculos estatísticos. Serão investigadas e exploradas as mais modernas técnicas de visualização de dados no *Software R* e estas serão aplicadas em dois bancos de dados para que possam ser testadas, reproduzidas e documentadas.

Este trabalho foi iniciado criando visualizações sofisticadas e poderosas a partir de um banco de dados referente aos aluguéis de bicicletas na cidade de Brasília. A intenção foi tornar mais prática e eficiente a disposição das estações distribuídas na cidade, de acordo com as possíveis rotas mais comuns entre as estações e o horário de maior fluxo, buscando facilitar todo o processo de aluguel, devolução, etc.

Em seguida, procurou-se formas de apresentar o fluxo de alunos e docentes estrangeiros para a Universidade de Brasília. Quanto maiores e melhores as informações (bancos de dados), maior qualidade as representações gráficas e os gráficos interativos podem ter. Sendo assim, maiores informações deverão ser extraídas desses bancos de dados e representadas de forma sucinta sem perda de qualidade.

1.1.1 Visualização de dados

Novas tecnologias como *Machine Learning* são usadas pelas empresas para analisar grandes volumes de dados. Os dados podem ser complexos e extensos, necessitando de tratamentos e análises. Se essas tarefas forem realizadas de forma veloz e eficiente, tornam-se muito mais lucrativas. Porém, isso necessita do entendimento e clareza das informações de maneira que faça sentido para o negócio. É nesse momento que os dados tornam-se valiosos. De acordo com Assis (2019), ao invés de ler valores de forma individual, como em tabelas ou texto, por meio de uma representação gráfica é possível identificar e entender vários aspectos de uma só vez. Visualização de dados são técnicas de representações gráficas que se tornaram bem usuais, pois os gráficos e *dashboards* simplificam grandes bancos de dados e, conseqüentemente, tornam a informação mais acessível. Segundo Freitas et al. (2001 apud Oliveira et al., 2018), a visualização de dados permite que as informações sejam apresentadas em formas gráficas de modo que o usuário consiga utilizar sua percepção visual para entender e analisar as informações, trata dados abstratos como relacionamentos ou informações inferidas com base nos dados mensurados. A visualização deixa mais fácil a análise dos resultados, ajuda no processo de detecção de padrões e de tomada de decisão.

Segundo Lazaro e Alvim (2014), da definição de visualização de dados podemos tirar dois grandes objetivos: compreensão e comunicação. A visualização de informação pode ser igualmente chamada de visualização exploratória de dados. É importante levar em consideração esse fato no processo de realização de uma visualização de sucesso. Tendo por objetivo a exploração e a compreensão dos dados, a visualização deve ser de natureza exploratória. A representação gráfica é geralmente eficiente porque o ser humano é bastante adaptado para a análise visual. Escolher a visualização certa pode influenciar a interpretação, porque nossos olhos são atraídos por cores e padrões. São vários os recursos que podem ser utilizados como gráficos, tabelas, mapas, infográficos, painéis e assim por diante.

Segue abaixo alguns benefícios da visualização de dados para as empresas segundo o site Neomove (2019).

- **Absorção rápida das informações:** as informações apresentadas usando elementos visuais são muito mais claras e mais organizadas, facilitando assim a interpretação das informações e o processo de encontrar padrões e tendências nos dados.
- **Facilita a tomada de decisão:** depois de identificar padrões por meio da visualização é mais fácil planejar os próximos passos da empresa, reduzir tempo, erros e esforços para a tomada de decisão. Esse processo ajuda a empresa a tomar decisões mais assertivas.
- **Conexões:** a visualização não é útil somente para tomar decisão, ela ajuda igualmente a encontrar ligações entre as informações.
- **Apresentação mais envolvente:** uma apresentação é envolvente e atraente quando ela é feita de elementos visuais, desse jeito as pessoas ficam mais interessadas quando você estiver apresentando resultados.
- **Maior compartilhamento:** com o uso das novas tecnologias, é muito mais prático compartilhar resultados de visualização de dados.

Visualização de dados é importante para o negócio, mas não é em todas as situações que ela é útil. Segundo Zeviani (2018), a visualização deve ser feita nos casos seguintes:

- absorver informação de forma rápida e criativa;
- compreender relações, padrões e tendências;
- identificar e agir sobre emergentes tendências rapidamente;
- manipular e interagir diretamente com os dados;
- melhor compreensão dos fenômenos;
- geração e aperfeiçoamento de hipóteses e modelos conceituais;
- maior impacto na comunicação dos resultados;
- cativação dos revisores e da audiência.

1.1.2 Conceito de tomada de decisão

De maneira habitual, diversas decisões são tomadas por grupos sociais formados por pessoas com uma série de tarefas administrativas, que se integram a partir de uma estrutura organizada com a meta de exercer certos objetivos daquilo que se propõe.

Decisão é uma determinação ou resolução, ou seja, uma atitude que está relacionada com a ação e o efeito de resolver determinado acontecimento. Pode-se começar ou encerrar uma situação, bem como, estabelecer uma transformação de estado. Já, numa tomada de decisão, “é o processo de escolha da melhor alternativa” (Certo, 2005, p.123) ou a que mais beneficiará a organização. Uma vez que, para o autor, decisão é a escolha feita entre duas ou mais alternativas disponíveis.

Alguns elementos podem exercer influência no resultado de uma decisão, tais como, percepção, intuição, e racionalidade. A intuição, por exemplo, “nasce da experiência e de sentimentos a respeito de estímulos” (Maximiano, 2009), a conduta totalmente racional é utópica, algumas dificuldades, unicamente não serão resolvidas por meio de normas, nem sempre é possível ter acesso a todos os dados, desta forma:

[...] A diferença entre racionalidade e intuição está na proporção de informação, de um lado, e opinião e sentimentos, de outro. Quanto maior a base de informação, mais racional é o processo. Quanto maior a proporção de opiniões e sentimentos, mais intuitivo se torna. A racionalidade e a intuição são atributos humanos complementares e não concorrentes. (Idem)

Na verdade, as grandes decisões não são baseadas no acaso. As ações de um administrador envolvem uma gama de decisões, ou seja, a quem será destinado essa produção (nicho de mercado), o que se vai produzir e de que forma será feito. Portanto, para se tomar uma decisão, faz-se necessário distinguir, recolher e analisar dados para avaliar problemas de diversas naturezas. Na medida do possível deve-se gerar e avaliar melhores alternativas, avaliando sempre os resultados.

Levando-se em conta que o agente tomador de decisão é o indivíduo que faz essas escolhas, dentre as quais as já citadas acima, e, por estas estarem relacionadas a uma série de fatos, como evidenciais e subjetivas. Diga-se subjetivas, pelo fato de envolver características pessoais, arraigadas por percepções, não se pode descartar a possibilidade dessa atividade ser suscetível a erros.

Por esses e outros motivos, bem como os critérios de subjetividade, além da análise e visualização de dados, os autores Certo (2005), Chiavenato (2010), Maximiano (2009) e Robbins (2010), sugerem um processo organizado e metódico na tentativa de reduzir a proporções mínimas dessas possíveis falhas e chegar a um resultado satisfatório. Assim recomendam alguns passos a serem adotados:

- 1) descobrir um problema que já existe;
- 2) listar soluções para esse problema;
- 3) selecionar a solução mais eficiente;
- 4) colocar em ação a solução;
- 5) verificar se a escolha alcançou os resultados esperados.

1.2 Histórico

1.2.1 História

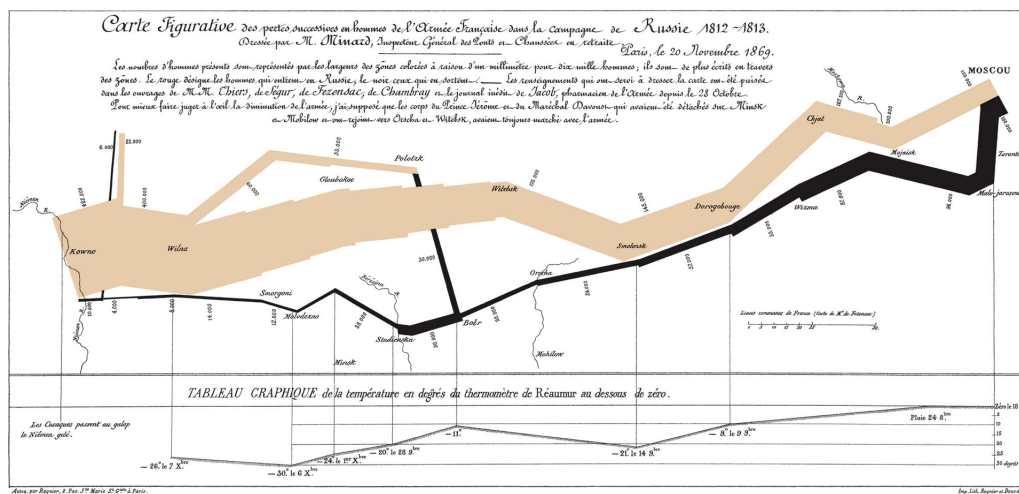
Acredita-se que as primeiras formas de visualizar dados foram feitas por William Playfair em 1786. Ainda com uma quantidade reduzida de dados, visualizações como linhas, barra e gráficos eram criadas por Playfair. Porém, muitos recursos como esses ainda são importantes e bastante utilizados nas análises de dados atuais.

Em 1854, o médico britânico John Snow descobriu a fonte de transmissão de cólera representando as ocorrências dos óbitos em uma mapa.

Uns anos depois, Florence Nightingale, enfermeira britânica durante a guerra da Crimeia (1853 a 1856), realizou o “coxcomb diagrams” que exaltou o efeito de doenças sobre o exército britânico.

Outro marco histórico na área das visualizações foi protagonizado pelo engenheiro civil francês Charles Joseph Minard, que divulgou um diagrama retratando as baixas no exército de Napoleão na marcha para Rússia em 1861.

Figura 1: Campanha Russa de Napoleão 1812.



Fonte: <https://sciencenorway.no/blog-blog-from-numbers-to-graphics-statistics/charles-joseph-minards-map-of-napoleons-flawed-russian-campaign-an-ever-current-classic/1618695>

O mapa acima retrata todo o cenário vivido na época, o tamanho do exército e suas baixas ao longo do percurso. A retirada do exército de Napoleão de Moscou quando se deparou com as cidades vazias e sem suprimento. O mapa ainda retrata as temperaturas vivenciadas e a cronologia da guerra. Um mapa rico em detalhes e capaz de elucidar todo um conto histórico.

Depois de uns anos, exatamente em 1914, o engenheiro americano Willard Brinton, publicou o primeiro livro de visualização para negócios chamado “Graphic Methods for presenting Facts”. E seguindo a mesma linha, Mary Eleanor Spear se baseando nas suas experiências de serviço no governo americano publicou em 1952 o “Practical Charting Statistics”. Um livro relatando as boas práticas de visualização de dados.

Outro livro sobre a visualização de dados foi publicado em 1967 pelo cartógrafo francês Jacques Bertin. O livro chamado “Sémiologie Graphique” foi o primeiro livro sobre teoria da visualização. Nesse livro, o autor descreveu 7 variáveis visuais (posição, tamanho, forma, cor, saturação, orientação e textura) e definiu 2 princípios: o da expressividade e da efetividade.

Nos anos 1970, o matemático americano John Tukey, foi o protagonista na utilização de computadores para a realização de visualização de dados e popularizou ideia de visualização exploratória e confirmatória. Ele foi o criador do gráfico *boxplot*.

Outra obra importante na história da visualização foi o livro “The visual display of quantitative information” de Edward Tufte publicado em 1983. Nesse livro, o autor combina rigor estatísticos com clareza e princípios de *design*. Um ano depois, os primeiros artigos sobre percepção gráfica foram publicados por William Cleveland e Robert McGill, eles procuraram entender como fazer visualizações efetivas.

O trabalho do francês Bertin foi para era digital quando Jock Mackinlay divulgou a sua tese “Automating the design of graphical presentations of relational information” de PhD em 1986.

Em 1999, o livro “The Grammar of Graphics” foi lançado por Leland Wilkinson. Nesse livro, o autor definiu uma gramática concisa para descrever os elementos de um gráfico.

De 1990 até 2010, houve uma evolução nos diferentes tipos de abordagens para visualização de dados: abordagens computacionais, abordagens orientadas ao *design* e à comunidade científica. O progresso da tecnologia gerou aumento na produção de dados e facilitou a realização de representações gráficas.

1.2.2 Visualização nos dias atuais

Com os avanços tecnológicos, as visualizações de dados ganharam grande destaque. Atualmente, a análise e o armazenamento de grandes massas de dados se tornou comum, um fenômeno conhecido como “big data”. Dessa forma, tornar a informação algo acessível e prático não é das tarefas mais fáceis. Segundo Cairo (2019), a popularidade da visualização de dados, a representação de informações por gráficos, tabelas e mapas, cresceu bastante nos últimos anos graças ao aumento da disponibilidade dos dados governamentais em formato aberto, não-governamental ou em plataformas corporativas, além de ferramentas de software que, em muitos casos, são de livre acesso ou em código aberto.

De acordo com Lazaro e Alvim (2014) o *Excel* era a ferramenta preferida das empresas e ainda é muito utilizada em alguns lugares, mas quando entramos na era do “big data” as empresas começaram a investir pesado em inovações computacionais pois os métodos “clássicos” já não traziam as informações tão precisas e na velocidade necessária.

Assim, as plataformas tradicionais de *Business Intelligence* (BI) começaram a ser usadas. Essas plataformas possuem recursos de apresentação de dados e tem a capacidade de realizar relatórios e *dashboards* cada vez mais interativos. Mas uns anos depois, as plataformas tradicionais de BI não atendiam mais as demandas da área de visualização, que está sempre se reinventando. Por isso, ferramentas simples para representação gráfica são áreas de BI com o desenvolvimento mais acelerado e as ferramentas mais tradicionais estão sempre inovando para tentar acompanhar. Além disso, por meio de computadores, tablets e celulares, as pessoas podem interagir com os seus dados em qualquer momento e em qualquer lugar, é por isso que Silva (2019) afirma: “internet e demais mídias digitais permitiram que informações e ideias fluíssem livremente, facilitando o aprendizado de outras pessoas”.

Segundo Zeviani (2019) é possível resumir a situação atual da visualização de dados em 5 pontos.

- Recursos/ferramentas em inovação e crescimentos.
- Representações gráficas mais apropriadas são desenvolvidas.
- Novas dimensões para as representações gráficas: animação, reatividade e interatividade.
- Monitoramento com *dashboards*.
- Velocidade e acesso:
 - Atualização em tempo real.
 - Relatórios web dinâmicos.
 - *Smart things*: *smartwatch*, aplicativos esportivos

Hoje em dia, a visualização de dados é importante na área da ciência, mas segundo Silva (2019), ela não resolve todos os problemas relacionados a dados e possui algumas limitações:

- ferramentas de visualização de dados mostram, mas não explicam;
- a simplificação excessiva dos dados;
- as limitações humanas na hora de realizar os algoritmos;
- dificuldade de escolher a ferramenta adequada;

1.3 Aspectos estéticos e cognitivos

Os aspectos cognitivos estão ligados ao poder das pessoas de processarem uma informação, isso é, a capacidade de pensar e compreender do cérebro. Já os aspectos estéticos estão ligados ao poder das pessoas de perceber e notar uma informação, sem contar que, no aspecto estético, busca-se a beleza na apresentação, ou seja, tornar atraente ao interessado na visualização. Haja vista que a “estética é uma ciência que remete à beleza e também aborda o sentimento que a coisa bela desperta dentro de cada indivíduo”. Portanto, ela está profundamente vinculada ao conceito de beleza. Extraído do Site Significados (<https://www.significados.com.br/estetica/>)

Os dois aspectos são fundamentais no processo de compreensão e trabalham de forma simultânea, um sendo responsável por atrair o interesse e o outro responsável por desenredar a informação. O sucesso de uma visualização está ligado ao comprometimento do projeto com esses dois aspectos.

Então, os processos cognitivos são um conjunto das funções mentais que nos deixam receber, registrar e conferir todas as informações que chega do nosso ambiente. Processo essencial e de fundamental importância

para os seres humanos, pois, sem ele não seria possível observar, compreender e interagir com o mundo externo.

Para a psicóloga Marissa Glover, do site Psicologia Online, os processos cognitivos dividem-se entre os básicos e superiores. “Os processos cognitivos básicos são muito importantes, já que são a base que nos permite receber a informação, armazená-la e usá-la para depois poder executar os processos cognitivos superiores”. E, neste caso foi dividido da seguinte forma:

“*Sensação e percepção*: Se trata da forma de sentir e perceber o mundo através dos nossos sentidos sem esforço aparente. Podemos receber informação do que nos rodeia e também do nosso próprio corpo, o qual nos permite interpretar tudo o que percebemos e compreender de forma coerente e organizada o ambiente e a nossa própria realidade. *Memória*: Como todos sabemos, a memória permite encontrar respostas a certas questões, como, por exemplo, saber a nossa própria data de nascimento, recordar eventos que aconteceram no passado, saber o que acabaram de nos dizer há um momento, entre muitas outras coisas. *Atenção*: Este processo cognitivo básico faz referência à nossa capacidade de concentração em um grande número de estímulos, assim como nos que são particularmente do nosso interesse”.

Já os processos cognitivos superiores, comumente são os processos conscientes e demandam um empenho intelectual maior para serem realizados. Assim discorre Glover:

“*Pensamento*: É um processo mental que se encarrega de raciocinar, tomar decisões, resolver problemas da vida cotidiana, entre muitas mais funções, devido à sua complexidade e heterogeneidade. *Linguagem*: A linguagem é a capacidade com que contamos para emitir e compreender distintos tipos de sons e palavras que vêm do exterior, em combinação com um grande número de frases e letras, com a finalidade de entender e comunicar com as outras pessoas”.

(<https://br.psicologia-online.com/processos-cognitivos-basicos-e-superiores-exemplos-e-tipos-212.html>)

1.3.1 Passos para fazer uma visualização

Depois de entender a importância de uma boa visualização é preciso entender os passos para fazer uma visualização útil. O primeiro passo é o motivo pelo qual se deseja fazer uma visualização. Entender a motivação é fundamental para seu sucesso. Nessa etapa é importante definir o objetivo final.

Após definir o motivo, o próximo passo é adquirir os dados. Os dados são a alma da visualização, afinal sem eles não seria possível o projeto. É importante que nessa etapa seja verificado se esses dados estão completos e se são confiáveis.

A próxima etapa é filtrar os dados. Nessa fase, todos os dados são convertidos em um formato que seja possível a utilização. Nessa parte também que é feita a filtragem, ou seja, descarta-se todo dado que não seja relevante para a visualização.

Só então, os dados estão prontos para a representação gráfica. Inicialmente, é importante que seja feito alguns gráficos intermediários para reavaliar o objeto principal. Após essa etapa, caso não haja outros desafios, deve-se partir para a representação mais refinada. Nessa etapa de refinamento, são feitos os toques finais da visualização, como por exemplo vídeos, gifs, dentre outros.

Por último, deve-se reavaliar o projeto final e comparar com o objetivo principal. Se atender a todas as expectativas o produto pode ser divulgado ou publicado. Caso contrário, deverá retornar à etapa em que houve a falha e revisá-la.

A Figura 2 resume bem o processo de criação de um projeto de visualização de dados :

Figura 2: Processo de visualização de dados

Processo	Habilidade	Resumo	
Definir objetivo		Entender a motivação e definir um objetivo final apropriado. Esta etapa é fundamental para o êxito da visualização. Depois, com base no status atual, determinar a etapa do processo a partir da qual é apropriado iniciar.	
Adquirir	Ciência da computação	Adquirir os dados relevantes e garantir que sejam tão completos e precisos quanto possível. Se relevante, complementar com dados públicos.	
Formatar	Ciência da computação	Analisar e formatar os dados adquiridos em uma forma passível de uso. Se existem diversos conjuntos de dados, garantir que estejam integrados.	Exploratória
Filtrar	Statistics and data mining	Filtrar os dados, de modo que o conjunto de dados inclua apenas os dados necessários.	
Analisar	Estatística e data mining	Selecionar a(s) ferramenta(s) apropriada(s) para a análise. Depois, realizar data mining, modelar e analisar os dados. Utilizar visualização exploratória de dados, se apropriado. Se outros interagirem com os dados, continuar no processo. De outro modo, esta é a última etapa; reiterar as etapas acima, se aplicável.	
Representar	Designer gráfico	Selecionar a(s) ferramenta(s) apropriada(s) para representar a visualização de dados. Depois, projetar e criar a visualização de dados ou infográfico.	
Refinar	Designer gráfico	Refinar a visualização de dados ou infográfico, de modo a se adequarem ao público-alvo.	Exploratória ou explanatória
Interagir	Interação homem-máquina	Publicar, implantar e interagir com a visualização de dados. Se o produto final for insatisfatório, determinar a etapa do processo à qual retornar e então repetir as etapas seguir. De outro modo, se o objetivo foi alcançado, o processo terminou.	

Fonte: Lazaro e Alvim (2014), disponível em:

<https://www.accenture.com/acnmedia/PDF-45/Accenture-Entendido-De-Dados.pdf>

Segundo Lazaro e Alvim (2014), cada etapa do processo deve ser finalizada antes da próxima, não porque o processo é linear mas porque uma etapa pode parecer concluída, mas futuramente pode aparecer a necessidade de revê-la, algumas etapas podem exigir muitas iterações. Quando não sabe do processo ou pula alguma etapa por achar demorada ou inútil, o resultado pode ser um produto de baixa qualidade ou ineficiente. Por isso é importante que, ao final de todo o processo, a visualização esteja a mais completa possível tendo somente as mais relevantes informações.

1.3.2 Cérebro x Visualização

Composto por mais de um bilhão de neurônios, o cérebro é um órgão controlado pelo Sistema Nervoso Central (SNC), cuja complexidade se assemelha a uma máquina, a um processador de computador. Por meio do córtex cerebral se faz o controle do [...] “pensamento, a aprendizagem, a solução de problemas, as emoções, a memória, a fala, a leitura, a escrita e os movimentos voluntários” (Sanchis, 2020). Sendo que:

“A unidade funcional do sistema nervoso é o neurônio. Sua função consiste em receber mensagens de outras células, decidir se envia ou não a informação e fazer isso de maneira eficiente. Existem neurônios sensoriais, que recebem informação dos sistemas sensoriais e as levam para o Sistema Nervoso Central (SNC); neurônios motores, que transmitem informação de ativação motora, do cérebro aos músculos e os chamados interneurônios, que transmitem informação a nível local, dentro do SNC”.

(https://br.psicologia-online.com/como-funciona-o-cerebro-humano-430.html#anchor_3)

Relacionar uma visualização de dados com a capacidade de absorção da informação pelas pessoas que a acessam, deve ser um dos pilares no processo de construção dessa visualização. Segundo o site Datapine, “90% da informação transmitida ao cérebro é visual”. O site ainda destaca que:

“se uma alegação científica for apresentada apenas em palavras ou números, 68% das pessoas considerarão as informações precisas e verdadeiras. Mas se as mesmas informações ou métricas forem apresentadas em um gráfico simples, o número aumentará para 97%.”

O site continua dizendo que “As pessoas têm 30 vezes mais probabilidade de absorver infográficos de alta qualidade do que texto simples.”

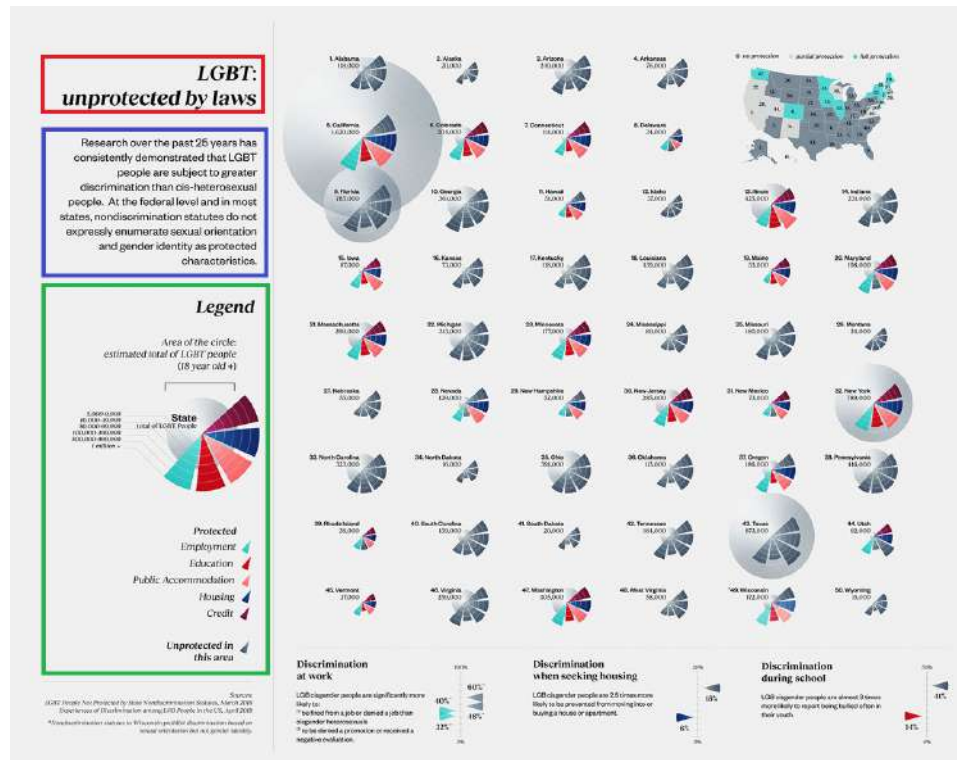
Fica evidente a importância da visualização de dados. Não só por questão estética ou uma questão de simplificar os dados, pois ter a informação não é tudo. É preciso ter a preocupação de como a informação chegará ao público alvo. Não tem como negar que somente com uma boa visualização é possível alcançar tais feitos. Fonte: <https://www.datapine.com/blog/best-data-visualization-books/>

De acordo com Brum (1998), 75% do conhecimento que adquirimos vem através da visão, da forma como visualizamos e interpretamos as coisas.

1.3.3 Aspectos que ajudam uma boa visualização

Algumas ferramentas ajudam a boa compreensão das visualizações. É importante que a visualização de dados seja autoexplicativa, ou seja, as imagens ou vídeos não precisam de outros fatores para passar a mensagem, além dos que já estão na visualização. Aspectos esses, como título, subtítulo, legenda e cores, são apresentados no exemplo a seguir:

Figura 3: Visualização da Comunidade LGBT



Fonte: <https://visme.co/blog/best-data-visualizations/>

Como exemplo de uma boa visualização, a designer e artista plástica Gabrielle Merite criou uma visualização sobre como a comunidade LGBT está desprovida de leis que os amparem nos Estados Unidos da América. Essa visualização foi destaque no ano de 2019.

O **título**, que está destacado pelo retângulo vermelho na imagem, é o cartão de visita. Nele contém a informação mais direta do que vem a ser a visualização. Nesse caso, o título deve ser atraente a ponto de causar interesse de quem está olhando a visualização. Por esse motivo, o título deve ser simples e objetivo.

O título tem a necessidade de informar. Tomando-se o cuidado minucioso, caso haja necessidade, com as pontuações, tais como exclamação e interrogação. Primeiramente recomenda-se criar o texto e/ou conteúdos (imagens) antes e, a posteriori, o título. Desta forma, ele discorrerá precisamente sobre o que foi publicado. Portanto, a atenção dos visualizadores está relacionada a um título que seja preciso, característico, breve e importante.

O **subtítulo**, que está sendo destacado pelo retângulo azul, serve para especificar a forma como foi feita a visualização ou simplesmente para dar uma informação adicional ao leitor que fará a interpretação da representação gráfica.

De acordo com esta pesquisa da Nielsen Norman Group, empresa norte-americana de pesquisa, treinamento e consultoria, 79% dos usuários da internet usam a escaneabilidade (a famosa “leitura dinâmica”) para avaliar uma página em vez de ler o seu conteúdo do início ao fim. Só então eles decidem se vão prosseguir ou fechar o navegador. Extraído do site <https://comunidade.rockcontent.com/subtitulos/>.

Depois do título e subtítulo, a **legenda**, que está sendo destacada pelo retângulo verde, auxiliam o leitor a diferenciar os parâmetros analisados nas visualizações. Geralmente as legendas apresentam os dados a serem comparados, como na visualização da Gabrielle Merite.

Para o geógrafo Ronaldo Decicino, na página “Pesquisa Escolar Geografia”, site Uol Educação, dar-se o nome de “legenda” [...] “às simbologias utilizadas para representar um fenômeno qualquer no mapa. A simbologia existente nas legendas – em um mapa de orientação básica – permite que qualquer pessoa, independente do seu país de origem, possa entender a representação gráfica de uma determinada região”. Fonte: <https://educacao.uol.com.br/disciplinas/geografia/legendas-convencoes-graficas-ajudam-a-compreender-os-mapas.htm>

Em se tratando de fotografias, entre os fotógrafos ouviu-se muito a expressão de que “a fotografia tem que se auto-explicar”. Pois, uma fotografia sem legenda poderá dar margem para várias interpretações dentro de um subjetivismo; ou, de repente, deixar de ser um registro para se tornar uma imagem perdida no tempo e no espaço. Uma foto sem história é uma imagem sem profundidade, sem substância, que nada contém.

As **cores**, em uma visualização, assumem, muitas vezes, o papel de diferenciar o que está sendo analisado. Outra função das cores é chamar atenção para as visualizações, em muitos casos, são os toques das cores empregadas na visualização que fazem a função de atrair o leitor.

São inúmeros os aspectos que auxiliam numa boa visualização. Por isso, a importância de saber para quem e qual o objetivo deseja-se alcançar com a visualização. Não há uma receita para uma boa visualização e o grande desafio é justamente saber qual ferramenta utilizar para que seus objetivos sejam alcançados.

1.4 Visualização de dados espaciais

1.4.1 O que são dados espaciais

Dados espaciais ou dados geográficos são dados localizados no espaço geográfico, são dados que possuem posição. Em outras palavras possuem coordenadas na superfície terrestre. Essas coordenadas são chamadas de latitudes e longitudes. Segundo João Henrique Quoos.

“A latitude é a distância ao Equador medida ao longo do meridiano de Greenwich. Esta distância mede-se em graus, podendo variar entre 0° e 90° para Norte (N) ou para Sul (S). E a longitude é a distância ao meridiano de Greenwich medida ao longo do Equador. Esta distância mede-se em graus, podendo variar entre 0° e 180° para Leste (E) ou para Oeste (W).”

Fonte: (http://coral.ufsm.br/cartografia/index.php?option=com_contentview=articleid=43Itemid=39)

Dessa forma, análises poderão ser feitas a partir desses dados espaciais. Análises essas que poderão investigar fenômenos e buscar padrões ou tendências nos dados.

1.4.2 Tipos de dados

Os dados apresentam-se de diversas maneiras, podendo ser eles pontos, curvas, superfície ou área.

- **Pontos**

Pontos são utilizados para representar uma ocorrência isolada, nesse caso um ponto apresenta uma latitude e longitude. Um exemplo seria plotar uma determinada capital. Se forem plotadas várias capitais o mapa apresentará um ponto para cada capital. Como no exemplo a seguir:

Figura 4: Exemplo de pontos

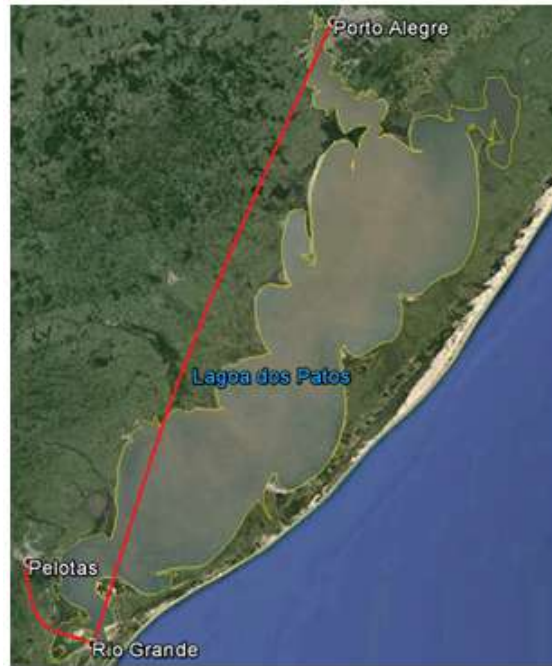


Fonte: https://rstudio-pubs-static.s3.amazonaws.com/176768_ec7fb4801e3a4772886d61e65885fbdd.html

- **Linhas**

Linhas representam a ligação de vários pontos, podendo representar por exemplo a rota de um avião, rios, lagoas, distância entre dois pontos no mapa, entre outros. Como no exemplo a seguir, que representa a rota de um avião.

Figura 5: Exemplo de linhas



Fonte: <https://www.varig-airlines.com/pt/rotas.htm>

- **Curvas**

Curvas são arredondamentos nas linhas, como por exemplo, a curva de uma rua ou até mesmo a parábola de um avião ao decolar e pousar na cidade desejada.

Figura 6: Exemplo de curvas



Fonte: Autores.

1.5 Imagens e vídeos

Vídeos curtos e gifs ganharam notável destaque nos tempos modernos por conta do grande protagonismo das mídias sociais. Existem várias maneiras de ser fazer um vídeo. Além da gravação de câmera, é possível criar um vídeo a partir de imagens fixas. Quando se trata de uma animação, que busca retratar um acontecimento ao longo do tempo, torna ainda mais desafiadora sua criação. Uma forma de se retratar o acontecimento ao longo do tempo é tratando cada instante como um quadro, tendo assim, cada instante em uma imagem. Desta forma, a criação do vídeo é uma questão de juntar essas imagens de forma ordenada ao longo do período.

Figura 7: Imagens e vídeos 1



Fonte: Autores.

Essa é uma imagem das várias utilizadas para criação do vídeo. Nessa imagem é possível identificar o dia e o horário que ela representa do banco de dados. Dessa forma, quando em conjunto com as demais, formam o vídeo com os dias e horários em sequência. Abaixo, outros exemplos desse mesmo vídeo.

Figura 8: Imagens e vídeos 2



Fonte: Autores.

Figura 9: Imagens e vídeos 3



Fonte: Autores.

A relação tempo versus quantidade de imagens do vídeo é dada pelos *frames*, podendo, então, o criador do vídeo determinar o tempo do vídeo a partir da quantidade de *frames* por segundo. Então, esse controle se dá primeiro pela quantidade de tempo que se espera ter no vídeo e segundo pela quantidade de imagens que se tem no banco de dados.

Usando essa forma, qualquer alteração no vídeo, seja ela estética ou não, precisa ser feita nas imagens primeiro. Somente após as alterações será possível a junção das imagens.

1.6 Frameworks de visualização de dados no R

1.6.1 Software R

Segundo o site `r-project`:

“O R é um projeto GNU semelhante à linguagem e ao ambiente S que foi desenvolvido na Bell Laboratories (anteriormente ATT, agora Lucent Technologies) por John Chambers e colegas.”

O R é um software computacional estatístico e gráfico. Nele é possível ter acesso a uma grande variedade de técnicas estatísticas como modelagem, testes estatísticos, série temporal e técnicas gráficas como mapas, tabelas e infográficos.

Uma das grandes vantagens de se usar o R é o fato dele ser um software de código aberto, livre e grátis. Em outras palavras, os usuários podem programar de acordo com a demanda, isso é possível a partir de criação de pacotes e funções próprias. Tornando o R cada mais completo.

“Muitos usuários pensam em R como um sistema de estatísticas. Preferimos pensar nisso como um ambiente no qual as técnicas estatísticas são implementadas. R pode ser estendido (facilmente) por meio de pacotes. Existem cerca de oito pacotes fornecidos com a distribuição R e muitos mais estão disponíveis por meio da família *CRAN* de sites da Internet, cobrindo uma ampla gama de estatísticas modernas.”

Fonte: <https://www.r-project.org/about.html>

Nesse sentido os pacotes são muito importantes na extensão das capacidades do R, fazendo com que certas implementações ocorram. Para a criação de mapas foi utilizado pacotes prontos e que são bem conhecidos na criação de mapas.

Sobre as desvantagens do software R, por ser de código aberto e ser fomentado por seus usuários não possui qualquer suporte técnico quanto aos problemas de código.

1.6.2 Pacotes

Os pacotes são um conjunto de comandos que possuem funções específicas. Na criação dos mapas alguns pacotes são importantíssimos como *ggplot2*, *gganimate*, *leaflet*, *magic*, *shiny*. Eles trazem a possibilidade de gerar imagens de alta qualidade com poucas linhas de código. Esse assunto será abordado brevemente a seguir.

- **Ggplot2** Criado por Hadley Wickham em 2005, esse pacote revolucionou a forma de se apresentar a visualização de dados. Ele cria gráficos mais bonitos e complexos de forma fácil e intuitiva. Além de permitir a criação de inúmeras possibilidades de gráficos com poucas linhas de código.
- **Gganimate** É uma extensão do pacote *ggplot2*, sendo responsável pela animação. Mostra os objetos plotados variando de acordo com o tempo.
- **Leaflet** É um vasto pacote para mapas interativos. Com ele é possível mexer em praticamente todos os parâmetros do mapa, como por exemplo escolher o *template* do mapa, colocar título, dar zoom, acrescentar marcadores, colocar linhas, etc.
- **Magic** Pacote responsável pelas transições das imagens que geram o vídeo, de forma a suavizar as passagens entre elas. Nesse pacote é possível pegar as características de cada foto e mesclá-las, formando uma única imagem.
- **Shiny** Pacote que permite a construção de aplicativos na web que tem interação com R. Nele é possível que o usuário crie de forma fácil parâmetros para a construção do mapa de forma rápida e interativa.

1.7 Outras ferramentas de visualização de dados

Além do R, explorado para a realização dos mapas deste trabalho, existem outras ferramentas consolidadas no mercado que fazem função semelhante. A escolha da ferramenta é muito particular, tendo em cada uma delas vantagens e desvantagem. O produto final, quando bem elaborado, traz informações muito similares independente de qual ferramenta escolhida.

Será abordado brevemente as linguagens *Python*, *Power BI*, *Tableau* e *SAS*.

- **Python**

O Python foi criado no final dos anos 1980 por Guido van Rossum. Uma de suas vantagens é que são códigos fáceis de implementar e sua sintaxe simples de aprender. Com ele é possível produzir muito em pouco tempo.

Segundo o site Computer Word:

“Além de uma poderosa linguagem de programação, capaz de fazer uma variedade de programas complexos, *Python* também pode ser usado como uma linguagem de *scripts*. Um *script* feito em *Python* pode automatizar diferentes tarefas que reduzem o tempo de processos, economizando energia e em alguns casos até dinheiro.”

(<https://computerworld.com.br/carreira/python-10-motivos-para-aprender-a-linguagem-em-2019/>: :text=Python%20%

O Python é uma linguagem livre e está presente em várias outras plataforma, ou seja, é possível rodar o mesmo código em outras plataformas sem fazer qualquer tipo de alteração.

Empresas conhecidas mundialmente já aderiram a linguagem de programação *Python*, tais como: Instagram, Google, Netflix, Mozilla, Uber, Dropbox, entre outras.

- **Power BI**

Power BI é uma ferramenta da Microsoft relativamente nova, tendo sua liberação sido realizada em 2015.

Segundo o site da Microsof:

”O *Power BI* é uma coleção de serviços de software, aplicativos e conectores que trabalham juntos para transformar suas fontes de dados não relacionadas em informações coerentes, visualmente envolventes e interativas. Os dados podem estar em uma planilha do *Excel* ou em uma coleção de *data warehouses* híbridos locais ou baseados na nuvem. Com o *Power BI*, você pode se conectar facilmente a fontes de dados, visualizar e descobrir conteúdo importante e compartilhá-lo com todas as pessoas que quiser.”

(<https://docs.microsoft.com/pt-br/power-bi/fundamentals/power-bi-overview>)

- **Tableau**

Criada em 2003 por Christian Chabot, Pat Hanrahan e Chris Stolte. Tableau é um desenvolvedor de softwares para visualização de dados. O padrão fácil das implementações e o controle preciso da geração dos gráficos a tornam uma das principais ferramentas de visualizações de dados.

Segundo o site Five Acts:

“A visualização dos dados na ferramenta é direcionada à compreensão ágil e clara dos usuários e à aplicação das informações na estratégia empresarial.”

(<https://www.fiveacts.com.br/tableau-software-como-funciona-e-quais-sao-os-beneficios/>)

- **SAS**

No mercado de 1976, o *SAS* é um dos software mais usado e seguros no mercado. Muitas empresas mundialmente conhecidas usam esse software.

Segundo o site do Departamento de Estatística da Universidade Federal de Juiz de Fora –UFJF:

“O *SAS* é indicado para desenvolvimento de pesquisas com necessidade de análise de grandes bancos de dados, bem como desenvolvimento e aplicação de ferramentas estatísticas avançadas.”

(<https://www.ufjf.br/estatistica/eventos-e-projetos/projeto-sas/o-que-e-o-sas/>)

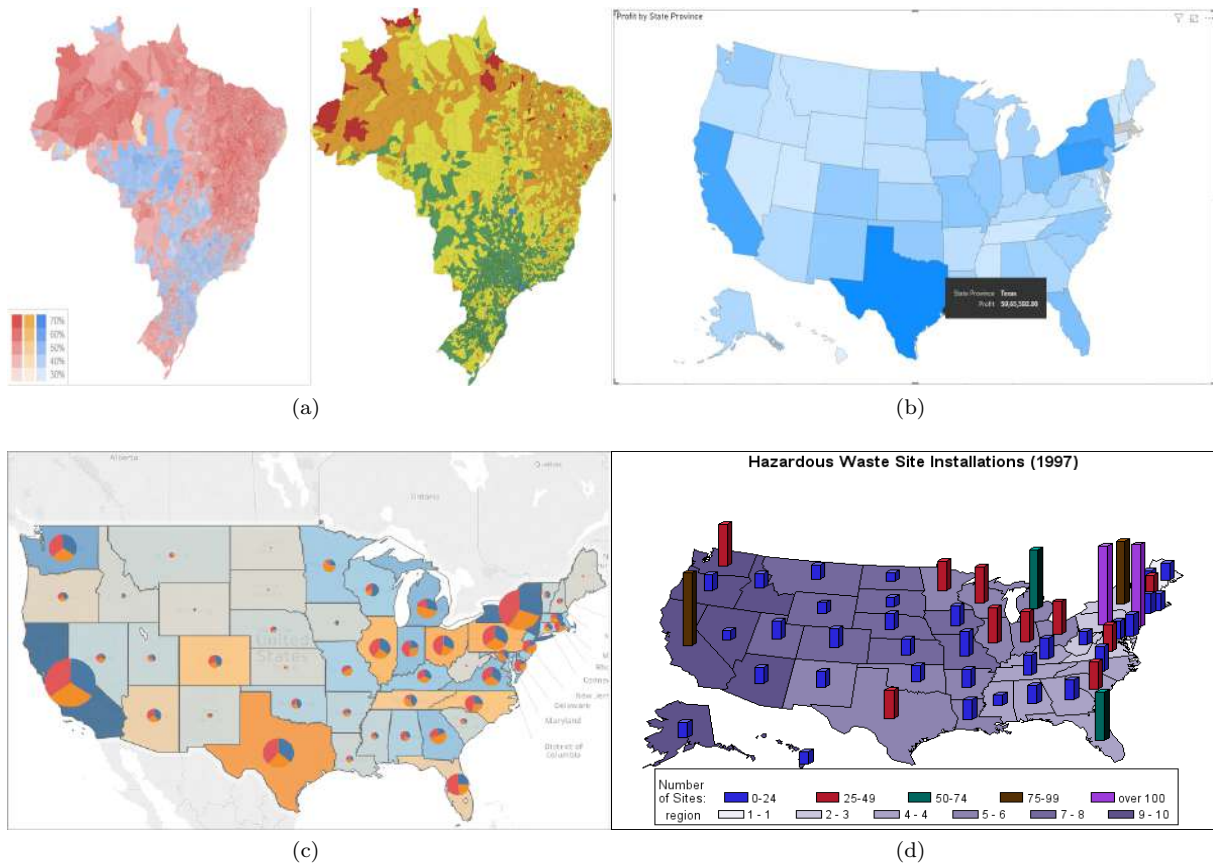


Figura 10: (a) Python (b) Power BI (c) Tableau (d) SAS

Imagens de cada ferramenta apresentada:

(a) Python - Fonte: <https://bastter.com/mercado/forum/865539/projeto-1-python-montar-um-mapa-coroletico-com-dados-eleitorais-parte-1>

(b) Power BI - Fonte: <https://nosqlba.org/uma-visao-geral-dos-tipos-de-grafico-no-power-bi/>

(c) Tableau - Fonte: https://help.tableau.com/current/pro/desktop/pt-br/maps_howto_filledpiechart.htm

(d) SAS - Fonte: https://support.sas.com/sassamples/graphgallery/PROC_GMAP.html

2 Metodologia

2.1 Tráfego de bicicletas em Brasília

2.1.1 Dados do +Bikes

A primeira visualização deste trabalho foi desenvolvida usando os dados do Sistema de Bicicletas Compartilhadas, o +BIKE. O projeto +BIKE era do Governo do Distrito Federal, operado pela empresa *Serttel* e era composto por estações inteligentes, alocadas em vários pontos de Brasília. O usuário retirava a bicicleta na estação por meio do aplicativo +BIKE baixado em seu celular, e entregava em qualquer estação depois do uso. O objetivo do projeto +BIKE era oferecer para a cidade de Brasília uma alternativa de transporte sustentável, saudável e não poluente. As estações funcionavam por meio de energia solar e eram ligadas a uma central via *wireless*, o que concedia um acompanhamento em tempo real das bicicletas e estações. O usuário podia consultar todas as informações sobre as estações, bicicletas e/ou vagas disponíveis por meio do site ou aplicativo do projeto. O projeto foi extinto em Abril de 2020.

Para melhor representar visualmente a movimentação das bicicletas *Serttel* entre as estações de Brasília e dessa forma compreender melhor esses deslocamentos, optou-se por aplicar um mapa de conexão aos dados do aplicativo +BIKE usando o *software* R. O gráfico exaltarà da melhor forma possível o fluxo de mobilidade entre as estações de bicicletas. Os dados foram obtidos no site de dados abertos do Governo de Brasília (www.dados.df.gov.br). Os arquivos eram separados por mês e são disponibilizados em formato *.ods*.

Os dados deste trabalho são datados de maio de 2019 até abril de 2019. Cada arquivo foi arrumado, corrigindo as inconsistências, analisando o tipo das variáveis, criando novas variáveis e limpando os ruídos contidos nos dados. O maior desafio nesta etapa de tratamento de dados foi corrigir os erros relacionados à digitação de nome e dados faltantes.

Após o tratamento dos dados, as variáveis mais importantes para a visualização foram selecionadas e todos os arquivos foram unificados para a criação de um banco de dados único, o que resultou em um arquivo de 7 colunas e 1.340.074 linhas chamado *dados.completo*. As viagens cuja estação de retirada é igual à estação de devolução não foram consideradas. A Tabela 1 mostra as 5 primeiras linhas do arquivo.

Tabela 1: Dados das Bicicletas.

Sexo	Nascimento	datahoradevolucao	datahoraretirada	EstacaoRetirada	EstacaoDevolucao
F	1988-04-01	2014-05-27 19:31:29	2014-05-27 16:24:56	Torre de TV	Memorial JK
F	1988-04-01	2014-05-27 20:12:02	2014-05-27 19:47:02	Memorial JK	Praça Buriti
M	1981-11-14	2014-05-27 20:12:02	2014-05-27 19:48:06	Memorial JK	Praça Buriti
F	1988-04-01	2014-05-27 23:31:26	2014-05-27 20:34:28	Praça Buriti	Centro de Convenções
F	1988-04-01	2014-05-27 23:33:01	2014-05-27 20:42:44	Praça Buriti	Centro de Convenções
F	1985-09-14	2014-05-28 12:53:02	2014-05-28 11:09:51	Memorial JK	Rodoviária

- Sexo: é o sexo do usuário, masculino e feminino são as opções disponíveis no arquivo;
- Nascimento: representa o ano de nascimento fornecido pelo usuário ao se cadastrar;
- datahoraretirada: essa coluna apresenta a data e o horário em que o usuário retirou a bicicleta da estação;
- datahoradevolucao: essa variável representa a data e o horário em que o usuário devolveu a bicicleta na estação;
- EstacaoRetirada: é o nome da estação em que a bicicleta foi retirada;
- EstacaoDevolucao: essa coluna retorna o nome da estação em que o usuário devolveu a bicicleta.

A Tabela 2 mostra a distribuição dos dados por sexo.

Tabela 2: Distribuição por Sexo

Sexo	Quantitativo
F	248747
M	824894
Outro	1289
Total	1074930

2.1.2 Dados das Rotas

Além do banco de dados baixado do site do Governo do Distrito Federal, foi necessário adquirir as coordenadas das estações e de todos os pontos percorridos pelo usuário durante a viagem (estação de retirada até estação de devolução). Essas informações são muito importantes para a realização do mapa. Para a obtenção das coordenadas geográficas das estações, um pedido por e-mail foi feito para a empresa *Serttel* que é responsável pelo projeto. Como as bicicletas do +BIKE não possuem um GPS, as trajetórias fornecidas pelo *Google Maps* foram usadas para traçar os pontos percorridos pelo usuário. Vale lembrar que o objetivo não é a representação exata do percurso de cada usuário, mas um possível itinerário da viagem para responder futuros questionamentos relacionados ao projeto +BIKE, tais como horários de maior fluxo, rotas mais frequentes, estações mais usadas, etc.

Uma consulta foi feita no *Google Maps* por meio de uma API própria do *site* para coletar as coordenadas dos pontos das trajetórias fornecidas pelo *Google*. A Interface de Programação de Aplicativos (API) é um grupo de aplicativos que permite a realização de uma interface inteligente na qual dois sistemas ficam em comunicação. Ela opera como uma base de ligação de duas entradas. A API do *Google Maps* retorna as rotas mais eficientes ao calcular as direções. O tempo de viagem é um dos fatores mais otimizados mas a API considera também outros fatores para decidir qual rota é a mais eficiente como distância e número de curvas.

É possível fazer a consulta no *software* R usando o pacote *Googleway*. O pacote possibilita o acesso às APIs do *Google Maps* e oferece a vantagem de fazer um mapa interativo do *Google* sobreposto com múltiplas camadas e formas. O pacote fornece a possibilidade de sobrepor informações de trânsito e rotas de ciclismo. Para essa tarefa foi necessário criar um *dataframe* com o nome de todas as estações junto com as suas respectivas coordenadas. A função *expand.grid()* foi usada para criar todas as combinações possíveis entre as estações. Assim, a partir da variável Estação foram geradas duas colunas nomeadas de “Origem” e “Destino”. Os casos em que a estação de origem é igual à estação de destino foram desconsiderados. Além disso, um número foi atribuído a cada estação (“CodigoRetirada” e “CodigoDevolucao”).

Para facilitar o processo de consulta no *Google Maps*, foram criados:

- Um *dataframe* chamado rotas para receber as informações das estações (origem e destinos), suas respectivas coordenadas e o código atribuído a cada viagem;
- Uma Lista chamada de consultas para receber as informações que a API vai retornar;
- Outra Lista nomeada “rotas.detalhada” que contenha as latitudes e longitudes das rotas que o *Google Maps* fornecerá.

A consulta das rotas foi feita no R através da função *google_directions()*. Ela recebeu como argumento a lista de coordenadas das estações (origem e destino), o modo(*bicycling*), o idioma no qual os resultados serão retornados(pt-BR) e a senha de acesso à API. A função retorna as rotas das bicicletas e um conjunto de informações que não é de interesse do nosso estudo. Para extrair as rotas aproximadas (suavizadas) dentro dos dados retornados pela API, foi necessário usar os comandos *routes\$overview_polyline\$points*. Vale lembrar que o resultado da operação precedente é uma *string* representada por sequência de pares lat/lon, codificada usando um algoritmo de compressão com perdas, que permite armazenar a série de coordenadas como uma única *string*. A decodificação da *string* foi feita por meio da função *decode_pl()*.

2.1.3 Desenvolvimento

- Mapa

Após finalizar o tratamento dos dados, iniciou-se o desenvolvimento do mapa. As colunas “CodigoRetirada”, “CodigoDevolucao” e “Codigo.rota” foram acrescentadas ao arquivo “dados.completo”. A coluna “Codigo.rota” foi criada unindo as colunas “CodigoRetirada” e “CodigoDevolucao”, e representa o código atribuído a cada viagem (estação de origem para estação de destino). Os dados foram agrupados por “Codigo.rota” e as quantidades de viagens foram calculadas. O resultado obtido foi um arquivo chamado “aux”, que contém o código das rotas e as quantidades de viagens.

Algumas mudanças foram feitas no arquivo “dados.completo” para facilitar a manipulação dos dados e uma concatenação com o arquivo “aux” foi realizada. Durante essa concatenação as linhas duplicadas não foram consideradas e a saída foi salva em um *dataframe* chamado viagem. Uma função foi criada no R para contar o número de trechos contidos em cada rota. Para a tarefa ser feita, a função usou as informações presentes nos arquivos “viagem” e “rotas.detalhada”. O retorno da função foi agrupado pelas coordenadas dos trechos (“lat”, “lon”, “lat.dest” e “lon.dest”) e a quantidade de trecho somada; dando assim a quantidade de viagem por trecho.

Com os dados finalmente prontos, a produção da visualização foi feita usando a função *leaflet()* do pacote *leaflet*. O pacote disponibiliza várias ferramentas para a realização de mapa interativo. A função recebeu as latitudes e longitudes de cada trecho (“lat”, “lon”, “lat.dest” e “lon.dest”), e as espessuras das linhas foram desenhadas em função da quantidade de viagem que cada trecho teve. Como o mapa obtido era interativo, foi utilizado o argumento *setView()* da função para fixar o mapa no centro de Brasília e facilitar a visualização. O Plano Piloto é o centro da cidade de Brasília e por isso as suas coordenadas foram passadas para o *setView()* e o *zoom* fixado no valor de 13.25.

Para melhorar a estética do mapa como título, legenda e cores, um mapa vazio (sem nenhuma rota desenhada) de Brasília foi criado e exportado para o *software Photoshop*. No *Photoshop*, a imagem foi tratada, títulos e legendas foram inseridos e a cor do lago de Brasília foi trocada pela cor azul. A imagem produzida no *Photoshop* foi importada no R e concatenada com o mapa contendo as rotas através da função *image_composite* do pacote *magic*. O mapa final foi exportado do R nas dimensões 1920×1080 por meio da função *mapshot()*.

- Vídeo

A realização do vídeo seguiu basicamente 3 passos: a reestruturação do arquivo “dados.completo”, a geração das imagens e a produção do vídeo.

– Reestruturação do arquivo:

antes de iniciar a produção do vídeo foi preciso fazer algumas alterações na estrutura do arquivo “dados.completo” e incluir mais informações. Foram criadas variáveis que informam os dias da semana da retirada e da devolução de cada bicicleta. Foi feito também a separação e concatenação de diversas colunas a fim de apresentar informações mais interpretativas. Além disso, criou-se uma variável para representar as datas de ocorrências. Essa data foi colocada para unificar as datas contidas no arquivo e facilitar os filtros. Com a criação das variáveis data e hora da retirada e devolução da bicicleta foi possível estabelecer o intervalo de tempo de utilização de cada bicicleta. Assim, a criação de um vetor de intervalo de 2020-01-01 00:00:00 até 2020-01-02 00:00:00 variando por minuto, foi o último passo antes da geração das imagens.

– Geração de imagens:

a proposta nesta etapa era gerar uma imagem por intervalo de 60s, então, 24×60 imagens por dia e 24×60×7 para os 7 dias da semana. Ao total, 10080 imagens foram geradas e juntadas para produzir o vídeo. Para efetuar essa tarefa, uma função e alguns *loops* foram criados. A função recebe como argumento o dia da semana e um intervalo de tempo. Assim, ela filtra os dados segundo os parâmetros, realiza os procedimentos para a criação do mapa, gera a imagem e a exporta para o computador. Os *loops* alimentam os parâmetros da função para que todos os 10080 mapas sejam gerados e exportados para o computador.

– Produção do vídeo:

primeiramente, a criação de um contador é essencial para facilitar o entendimento do vídeo. O contador é uma legenda que ajuda a ter informações sobre a sequência de imagens que está passando. Ele informa o dia e o horário de cada mapa do vídeo. Para isso, foi gerado um *dataframe* contendo os 7 dias da semana (Segunda até Domingo), as 24 horas do dia (00h até 24h) e 60 minutos (00 min até 59 min). Uma imagem de um relógio foi baixada na internet para ilustrar o contador.

Para importar todas as fotos, juntar e gerar o vídeo, a criação de uma função foi de grande utilidade. Essa função realiza o seguinte procedimento:

- importa a imagem do relógio;

-
- importa a imagem feita no *Photoshop*;
 - lê a imagem do mapa com as rotas;
 - concatena as imagens do mapa com a imagem feita no *Photoshop* usando a função *image_composite()* do R;
 - redimensiona e centraliza a imagem resultante;
 - concatena o resultado com a imagem do relógio;
 - coloca o contador na imagem;
 - exporta o resultado para o computador.

Dois *loops* foram realizados para que todo o procedimento fosse feito para as 10080 imagens. Por fim, as imagens resultantes do processo foram usadas para produzir o vídeo. O vídeo foi feito por meio da função *av_encode_video()* do pacote *magic*.

O processo era muito custoso computacionalmente, por isso foi decidido selecionar uma amostra de 10000 viagens e produzir o vídeo considerando somente as segundas feiras presente na amostra. Então para o TCC, o vídeo realizado foi baseado em 1440 imagens das Segundas feiras da amostra e o processo foi paralelizado para melhorar o desempenho dos algoritmos.

2.2 Chegada de Estrangeiros à UnB

2.2.1 Dados do sistema da UnB

Considerando o número importante de estrangeiros na UnB, e a questão da internacionalização da Universidade de Brasília, decidiu-se exaltar o fluxo de entrada de estrangeiros na UnB. Nessa intenção de produzir uma visualização adequada ao fluxo de entrada dos alunos e professores estrangeiros na UnB, foi feito um mapa de conexão.

Depois de ter feito um pedido via e-mail à reitoria da Universidade de Brasília, 3 arquivos foram recebidos. Um arquivo com os dados dos alunos estrangeiros, outro com os dados dos funcionários da UnB, e um terceiro arquivo contendo as informações do corpo docente aposentado. Os dados dos estudantes estrangeiros estavam em um arquivo de formato *.xlsx*, com 6120 linhas e 9 colunas. O arquivo dos funcionários estava em formato *.csv* e tinha 2836 linhas e 50 colunas. Os dados do corpo docente aposentado estavam apresentados em um arquivo de 2444 linhas e 49 colunas e em formato *.csv*.

Foi observado que, no arquivo dos estudantes, havia a coluna nacionalidade invés do nome do país de origem do aluno, por isso foi criada uma nova coluna preenchida com o nome do país de origem de cada aluno. A limpeza e o tratamento foram feitos individualmente nos três arquivos. Foram somente considerados as variáveis comuns aos três arquivos e aquelas mais relevantes para a construção do mapa.

Por fim, os três arquivos já tratados foram juntados e um arquivo de 6454 linhas e 6 colunas foi obtido. Para fazer um mapa de conexão, as informações de latitude e de longitude do ponto de origem e de destino são muito importantes. Essas informações, permitem que os pontos sejam representados no mapa. Por esse motivo, duas colunas a mais foram criadas. Em uma delas foram colocadas as coordenadas geográficas da Universidade de Brasília e na outra as coordenadas dos países dos estrangeiros.

Ao longo do projeto, foi observado que seria mais inteligível apresentar cores de linhas diferentes para cada continente, por esse motivo foi introduzido no arquivo mais uma variável contendo o nome do continente a que cada país pertence. Segue a baixo a Tabela 3 com as 5 primeiras linhas do arquivo final usado para realizar o mapa.

Tabela 3: Dados dos Estrangeiros

país	Continente	lat	lon	ano_ingresso	brasiliat	brasialon
Afganistão	Ásia	33.93911	67.70995	2006	-15.7941	-47.8825
África do Sul	África	-30.55950	22.93750	2005	-15.7941	-47.8825
África do Sul	África	-30.55950	22.93750	2002	-15.7941	-47.8825
África do Sul	África	-30.55950	22.93750	2002	-15.7941	-47.8825
África do Sul	África	-30.55950	22.93750	2005	-15.7941	-47.8825
África do Sul	África	-30.55950	22.93750	2011	-15.7941	-47.8825

Segue abaixo a lista das variáveis usadas para a construção do mapa

- país: essa variável representa o país de origem do estrangeiro;
- Continente: essa variável representa continente em que o país de origem está localizado;
- lat: latitude do centro do país, obtida no site www.adistanciaentre.com;
- lon: longitude do centro do país e foi obtida no site www.adistanciaentre.com;
- ano_ingresso: indica o ano em que o estrangeiro ingressou na Universidade de Brasília. Os anos variam de 1962 até 2019;
- brailiat: nessa variável temos a latitude da UnB;
- brasialon: essa coluna retorna a longitude da Universidade.

A Tabela 4 mostra a distribuição dos dados por ano.

Tabela 4: Distribuição por Ano

Ano	Quantitativo
1962	1
1964	1
1965	2
1967	2
1968	2
1969	1
1970	2
1971	4
1972	4
1973	4
1974	4
1975	5
1976	6
1977	6
1978	4
1979	6
1980	5
1981	10
1982	16
1983	23
1984	35
1985	85
1986	106
1987	126
1988	101
1989	108
1990	85
1991	79
1992	109
1993	10
1994	125
1995	127
1996	137
1997	155
1998	127
1999	108
2000	135
2001	127
2002	168
2003	170
2004	151
2005	159
2006	178
2007	159
2008	178
2009	220
2010	234
2011	277
2012	308
2013	352
2014	330
2015	357
2016	324
2017	260
2018	283
2019	259
Total	6454

2.2.2 Desenvolvimento

- Mapa

Antes de apresentar todas as ligações presentes no conjunto de dados, é essencial entender como funciona a visualização de conexão geográfica em um mapa. É uma prática comum de ligar 2 pontos usando a rota mais curta entre eles em vez de uma linha reta. Essa técnica é chamada de Grande Círculo. A propriedade mais relevante dessa técnica é que a distância mais curta entre dois pontos em uma esfera é o arco menor de um Grande Círculo. Quando há ausência de interferência das correntes e do vento, os navios e as aeronaves usam rotas de Grande Círculo.

O projeto de visualização se baseou nessa propriedade de Grande Círculo. Para realizar esse mapa, é indispensável dispor principalmente das coordenadas geográficas do ponto de origem e do ponto de destino. Com todas as informações reunidas, o conjunto de dados pode ser utilizado para ilustrar em um mapa a chegada dos estrangeiros na UnB usando o *software* R. Para descomplicar e agilizar o processo, em cada etapa foi criada uma função no R para realizar as tarefas.

Por causa do tamanho do arquivo, optou-se por agrupar as conexões que possuem os mesmos pontos de origem e contar as viagens considerando o ano de ingresso do estrangeiro e do continente. Além disso, para destacar os padrões mais importantes e ocultar os ruídos de conexões não relevantes, as conexões com altos volumes foram apresentadas no topo do gráfico.

Uma vez de posse de todas as viagens, para desenhar as conexões foi necessário obter os pontos intermediários entre o ponto de origem e o ponto de destino. É possível determinar os pontos intermediários usando a função *gcIntermediate()* do pacote *geosphere*. Ela recebe a latitude e longitude dos dois pontos de conexão e gera as coordenadas dos pontos do Grande Círculo. O parâmetro *n* da função, indica quantos pontos intermediários a função vai retornar. Quanto mais pontos, mais suave será a linha resultante, mas a partir de certo ponto, a diferença é quase inexistente.

Neste projeto, foi considerado *n* = 50. O argumento *addStartEnd* indica se os pontos inicial e final serão inseridos nas coordenadas do Grande Círculo. Ele recebe *TRUE* ou *FALSE* dependendo da necessidade. Nesse caso, o valor *TRUE* foi atribuído, e os pontos de origem e de destino foram inseridos nas coordenadas do Grande Círculo.

Com o objetivo de realizar um mapa de fácil entendimento, as conexões foram organizadas em um *dataframe* da seguinte maneira apresentada na Tabela 5.

Tabela 5: Pontos a serem Representados

lon	lat	group	Continente	ano_ingresso	pais	n
-47.88250	-15.79410	1	América do Sul	1975	Argentina	121
-48.14023	-16.25121	1	América do Sul	1975	Argentina	121
-48.39917	-16.70800	1	América do Sul	1975	Argentina	121
-48.65934	-17.16447	1	América do Sul	1975	Argentina	121
-48.92080	-17.62060	1	América do Sul	1975	Argentina	121
-49.18359	-18.07639	1	América do Sul	1975	Argentina	121

- lon e lat: Representa as coordenadas de todos os pontos (ponto de origem, ponto de destino e os pontos intermediários entre eles) a serem plotados;
- Group: Para conter o número atribuído a cada viagens (conexão entre ponto de origem e destino);
- Continente: O continente do país de origem do estrangeiro;
- ano_ingresso: O ano de ingresso do estrangeiro.
- pais: país de origem;
- n: representa o número de vezes em que a viagem (conexão) apareceu no banco de dados.

Com todos os recursos prontos, o processo de construção do mapa foi iniciado. Para realizar essa tarefa, o pacote utilizado foi o *ggplot* disponível no software R. Esse pacote oferece a possibilidade de personalizar os gráficos e mapas de várias maneiras. Com base em uma das propriedades do *ggplot* foi possível construir um mapa de conexão personalizado utilizando uma imagem do mapa mundi de luzes noturnas publicada em 2016 no site da NASA (earthobservatory.nasa.gov). Segue na Figura 11 a imagem baixada

Figura 11: Imagem da NASA



A Figura 11 está carregada de informações desnecessárias como cores, luzes, etc. Para resolver isso foi necessário, vetorizar a imagem no editor de imagens vetoriais *Adobe Illustrator*, fazer ajustes nos gradientes de cores até encontrar uma combinação de cores neutras e por fim exportar em formato *.JPG* com resolução 150 *dpi*. A Figura 12 foi a imagem obtida.

Figura 12: Imagem Vetorizada



A imagem da Figura 12 foi carregada no R com a função *readJPEG* do pacote *JPEG* e foi renderizada usando *rasterGrob()* do pacote *grid*. A função *annotation_custom()* do *ggplot* foi usada para colocar a imagem no pano de fundo. A função recebe como argumento a imagem renderizada e as coordenadas (*xmin*, *xmax*, *ymin* e *ymax*) dando a localização horizontal e vertical da imagem.

Na representação do mapa mundi, as latitudes vão de -90 até 90 e as longitudes vão de -180 até 180 . Por esse fato, $xmin = -180$, $xmax = 180$, $ymin = -90$ e $ymax = 90$ foram as coordenadas consideradas na função *annotation_custom()*. No mapa, as longitudes são representadas pelo eixo X e as latitudes pelo eixo Y.

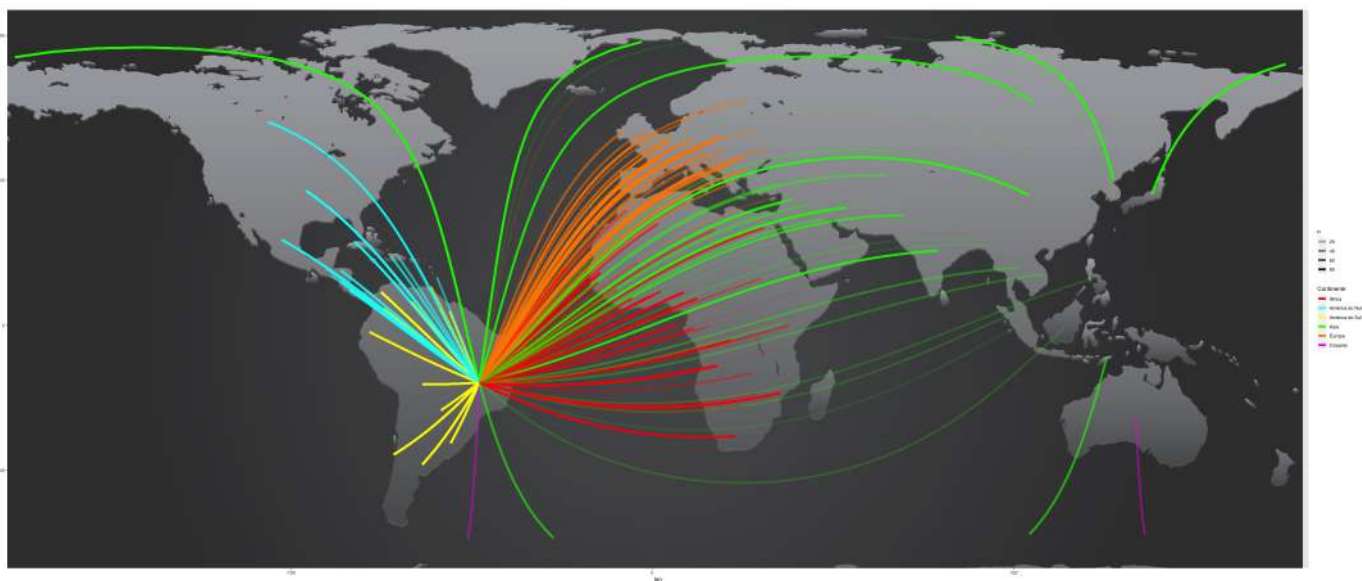
Tendo a imagem pronta, a função `geom_point()` foi utilizada para plotar cada ponto na imagem segundo suas coordenadas e as ligações entre os pontos foram feitas usando a função `geom_line()`. As latitudes e longitudes dos pontos foram passadas para as duas funções. Na intenção de facilitar a compreensão do mapa, as cores das linhas foram definidas por continente e a espessura das linhas pela quantidade de estrangeiros naquela conexão (ponto de origem até o ponto de destino). O seguinte resultado foi obtido:

Figura 13: Mapa dos Estrangeiros



Com base nesse resultado, foi possível perceber que vários pontos não ficaram bem ajustados e várias linhas ultrapassaram os limites da imagem. Para resolver esses problemas foi necessário redefinir os limites dos eixos X e Y usando as funções `xlim()` e `ylim()`. Os limites definidos foram $-180, 180$ para o eixo X e $-60, 80$ para o eixo Y. Graça a função `scale_x_continuous()`, a escala do eixo x foi também redefinida e usando o argumento `expand`, os pontos foram colocados a alguma distância do eixo X. Por fim, a função `coord_equal()` foi usada para garantir que os eixos X e Y tivessem escalas iguais.

Figura 14: Mapa_Melhorado dos Estrangeiros



Houve um progresso na apresentação das informações visto que, com o ajuste feito nas linhas, é possível ter uma melhor visualização do mapa. Já as legendas feitas automaticamente pelo software R não ficaram muito explícitas necessitando assim de mais alguns ajustes. Por isso as legendas automáticas foram excluídas e um espaço da imagem foi usado para criar manualmente as legendas. Uma legenda de texto foi feita e as cores dos

textos foram definidas de acordo com as cores das linhas.

Esse procedimento foi repetido para realizar a visualização de cada continente. Só que, nesse caso, para manter as mesmas dimensões para os 6 mapas, foi necessário usar 2 vezes a função `geom_line()`. Na primeira vez, as coordenadas de todos os pontos foram passadas para a função e foram utilizados os argumentos `alpha = 0.0` e `size = 0.0` para que as linhas ficassem invisíveis. Por fim, a função foi chamada de novo, mas dessa vez ela recebeu somente as coordenadas dos pontos do continente de interesse e as visualizações foram feitas seguindo os critérios do primeiro mapa.

- Vídeo

A partir dos dados e do mapa criado no R, optou-se por criar uma animação utilizando o pacote `gganimate` para ter uma visualização mais dinâmica das conexões. Com o `gganimate`, é possível aproveitar dos códigos feitos no `ggplot` e usando algumas linhas adicionais de códigos realizar uma animação.

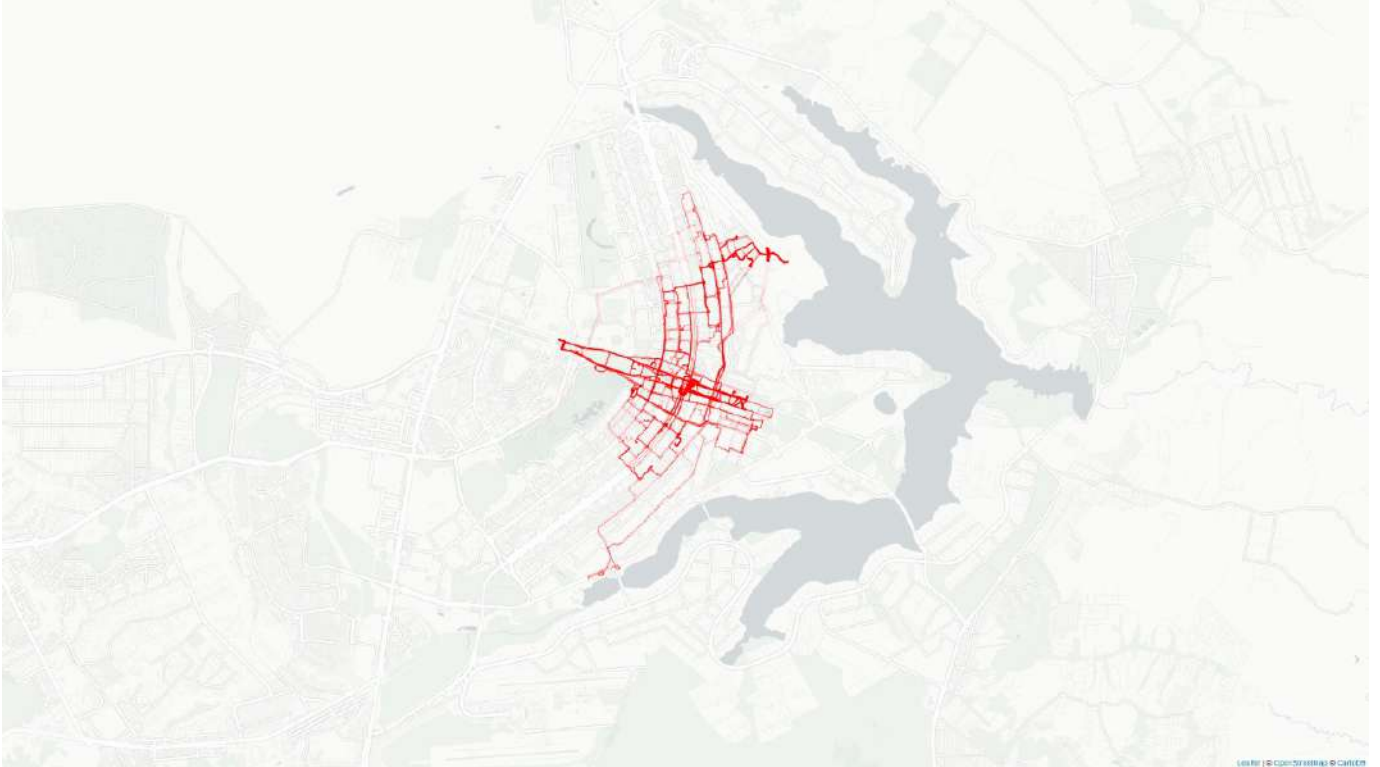
Assim, foram usadas as funções `transition_reveal()`, `shadow_wake()` e `animate()` para produzir a animação. A função `transition_reveal()` recebe a variável de transição, no caso deste projeto, ano de ingresso foi a variável de transição. A função `transition_reveal()` permite ter uma animação que vai revelando os dados conforme o tempo, para observar a velocidade entre as transições. A função `Shadow_wake()` deixa um rastro atrás dos pontos conforme eles se movem no tempo. Os argumentos da função permitem o ajuste da duração do rastro, o tamanho do rastro e assim por diante. E por fim, `animate()` serviu para renderizar o objeto `gganim` em uma animação. A função oferece a opção de mudar o as dimensões e duração da animação.

3 Resultados

3.1 Tráfego de bicicletas em Brasília

- Mapa

Figura 15: Mapa das Bicicletas



O mapa da Figura 15 foi a primeira visualização obtida depois do processo. Para a melhoria estética e da compreensão da informação, foram incluídos a legenda, o título e a cor do lago foi alterada. Depois da melhoria feita na estética do mapa, a Figura 16 foi obtida.

Figura 16: Mapa_Final das Bicicletas



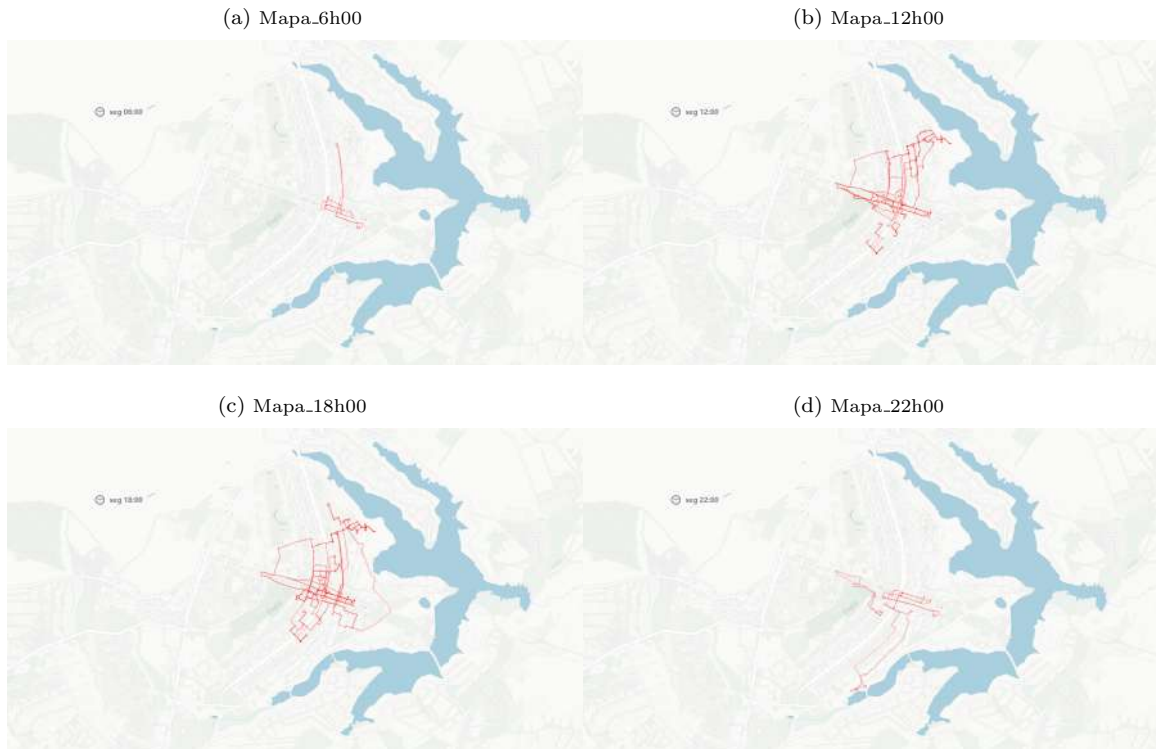
Com base nessa visualização algumas informações básicas podem ser extraídas. Observando este mapa da Figura 16, é possível perceber que as linhas são mais grossas na parte do eixo monumental. Então o fluxo é bem maior nessa região. Esse fato poderia ser explicado pelas movimentações devido ao número alto de servidores da administração públicas que trabalham na Esplanada. Os funcionários provavelmente usam as bicicletas para se locomoverem até um certo ponto para pegarem o transporte público adequado.

O trecho entre o Centro olímpico e a Universidade de Brasília apresenta também linhas grossas. Esse fluxo alto poderia ser explicado pela quantidade de moradores da Casa de Estudante da UnB, que usa esse meio de transporte para se deslocar.

Alguns estudantes da Faculdade de Educação Física também vão para as aulas no Centro Olímpico de bicicleta. Vale lembrar que a região onde o Centro Olímpico e a Casa de Estudante da UnB estão localizados é muito difícil de acesso e quase não existe transporte público lá. O projeto +BIKE facilitou muito o acesso à localidade. Há igualmente, uma movimentação importante dentro da Universidade de Brasília. Lugares como: Eixão, as quadras de Asa Norte e Asa Sul apresentam um fluxo de movimentação relativamente baixo.

- Vídeo

Figura 17: Sequências do Vídeo



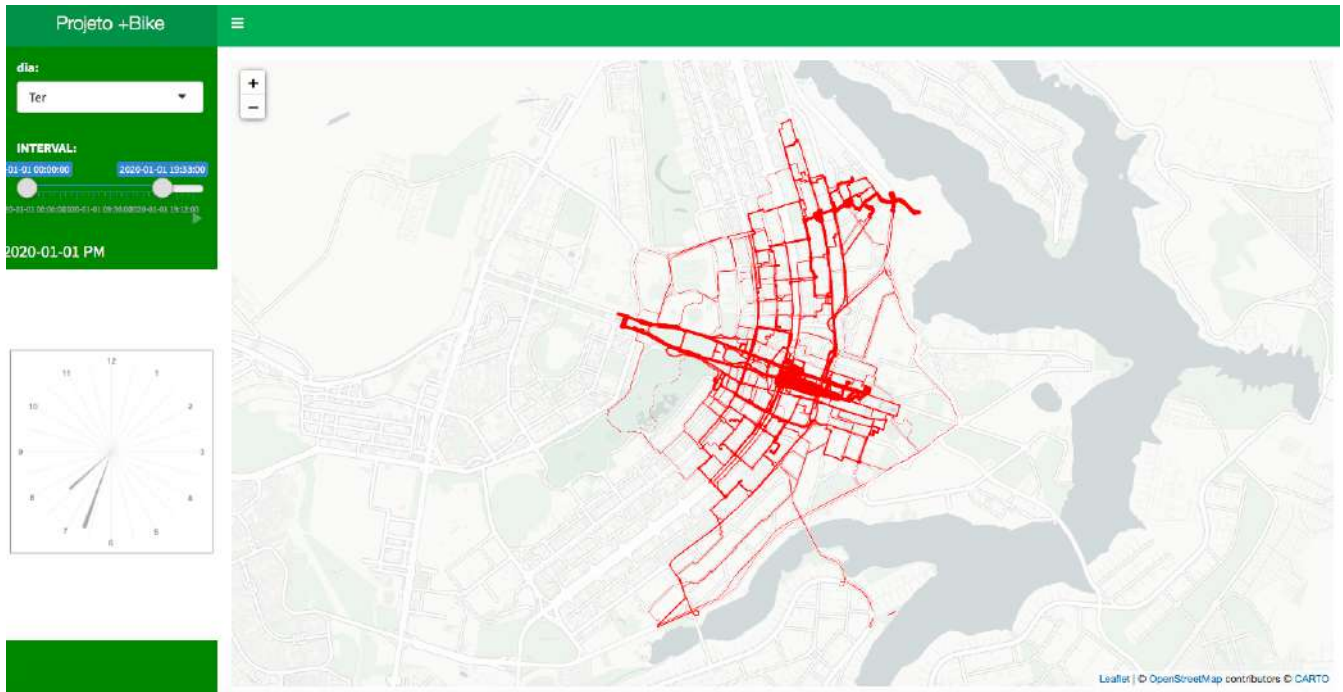
O vídeo permite visualizar com detalhes a evolução do fluxo das bicicletas ao longo do dia. Entre 6h e 7h, o fluxo é relativamente baixo, aumenta depois das 8h e cai depois das 10h. Entre 10h e 16h, a movimentação fica relativamente moderada. Depois das 16h, ela começa a aumentar e apresenta um pico entre 17h e 19h. Antes das 20h o fluxo cai e fica baixo até 23h59 min.

É possível afirmar que as bicicletas são mais usadas o início do dia depois entre 7h-9h e no final do dia entre 17h-19h. Nesses horários os cidadãos entram ou saem do trabalho, por isso que o fluxo nesses períodos é tão alto. É importante lembrar que as bicicletas podem somente ser usadas entre 06h e 23h59, foi por motivo que entre 00h e 06h não tem nenhum ponto representado no mapa. O vídeo pode ser acessado no link youtu.be/GQC1K9XTrYo.

- Shiny

Para ter uma visualização mais interativa e permitir o controle manual dos filtros, surgiu a necessidade de realizar um *dashboard*. É possível realizar um *dashboard* no R usando o *shiny*. Segue na Figura 18 a representação do painel :

Figura 18: Dashboard das Bicicletas



Os filtros definidos para a produção do vídeo foram igualmente definidos na realização do *dashboard*. As cores do painel foram escolhidas de acordo com as cores da identidade visual do Projeto +BIKE. Todo o procedimento de construção do painel foi feito somente no R.

Com esse *dashboard*, o usuário tem a possibilidade de controlar a visualização por meio dos filtros. Assim, ele pode ficar à vontade para escolher visualizar o mapa que é do seu interesse. O painel apresenta o mapa segundo o filtro escolhido pelo usuário. Para uma data e horário definidos, o usuário tem um mapa correspondente. O *dashboard* fornece a liberdade de visualização que o vídeo não proporciona.

3.2 Chegada de Estrangeiros à UnB

- Mapa

Depois de ter feito a legenda e todos os ajustes, o mapa na Figura 19 foi obtido:

Figura 19: Mapa_Final dos estrangeiros

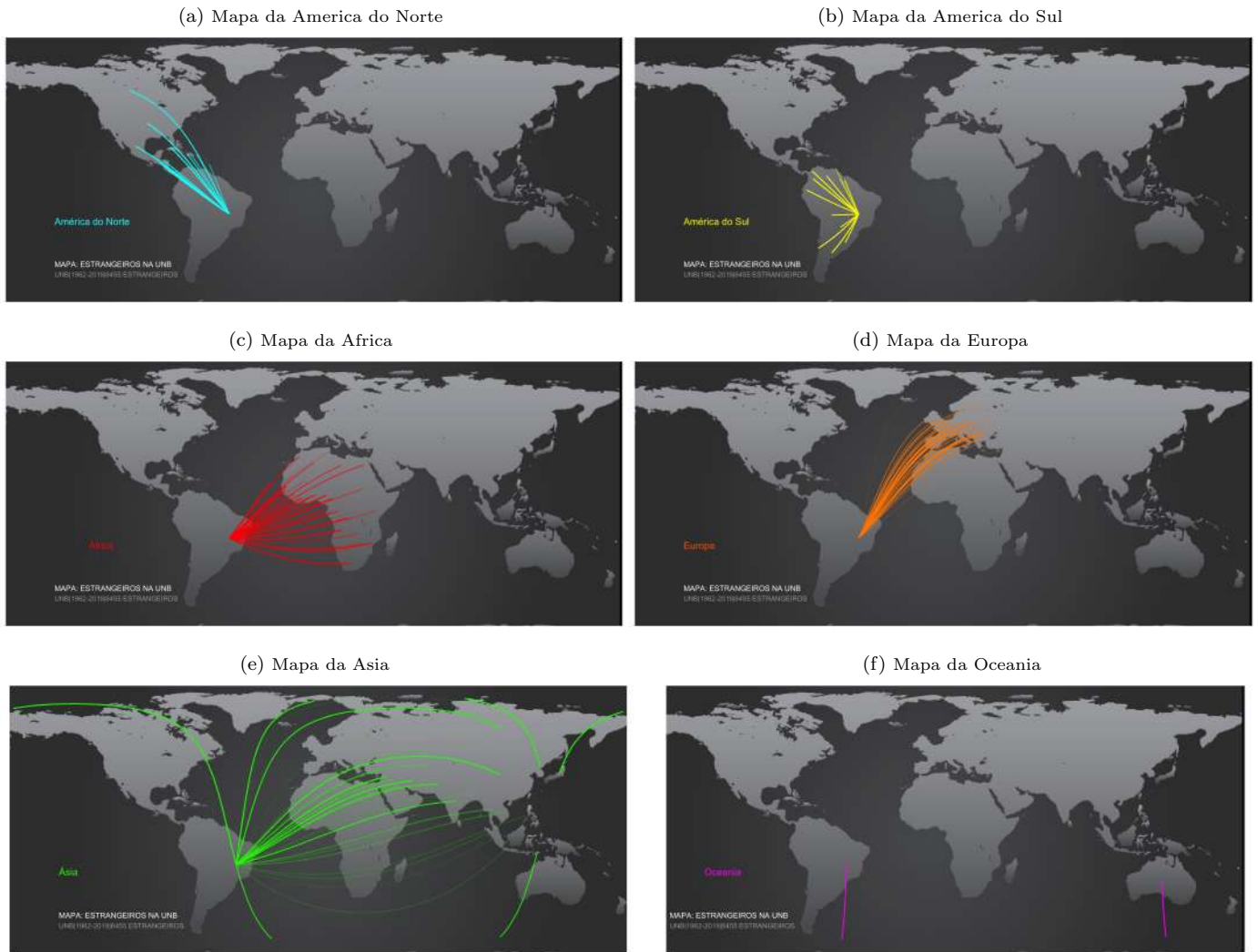


Com essa visualização, é possível extrair, de forma geral, muitas informações. Os 6 continentes estão muito bem representados. A quantidade de países presentes na Figura 19 é impressionante: quase todos os países do mundo estão no mapa. A Universidade de Brasília é realmente uma universidade aberta ao mundo. Somente a América do Sul e do Norte apresentaram poucas linhas finas. Isso indica que os países dos dois continentes são muito bem representados na UnB.

Vários países do continente africano e europeu estão presentes, mas alguns têm um fluxo muito baixo de estrangeiros. Não é fácil identificar todos os países que têm uma representação baixa na UnB, mas parece que a maioria está localizada no continente asiático. Para ter uma visão mais detalhada da situação, foi necessária uma visualização por continente.

Veremos na Figura 20 como ficaram as visualizações por continente:

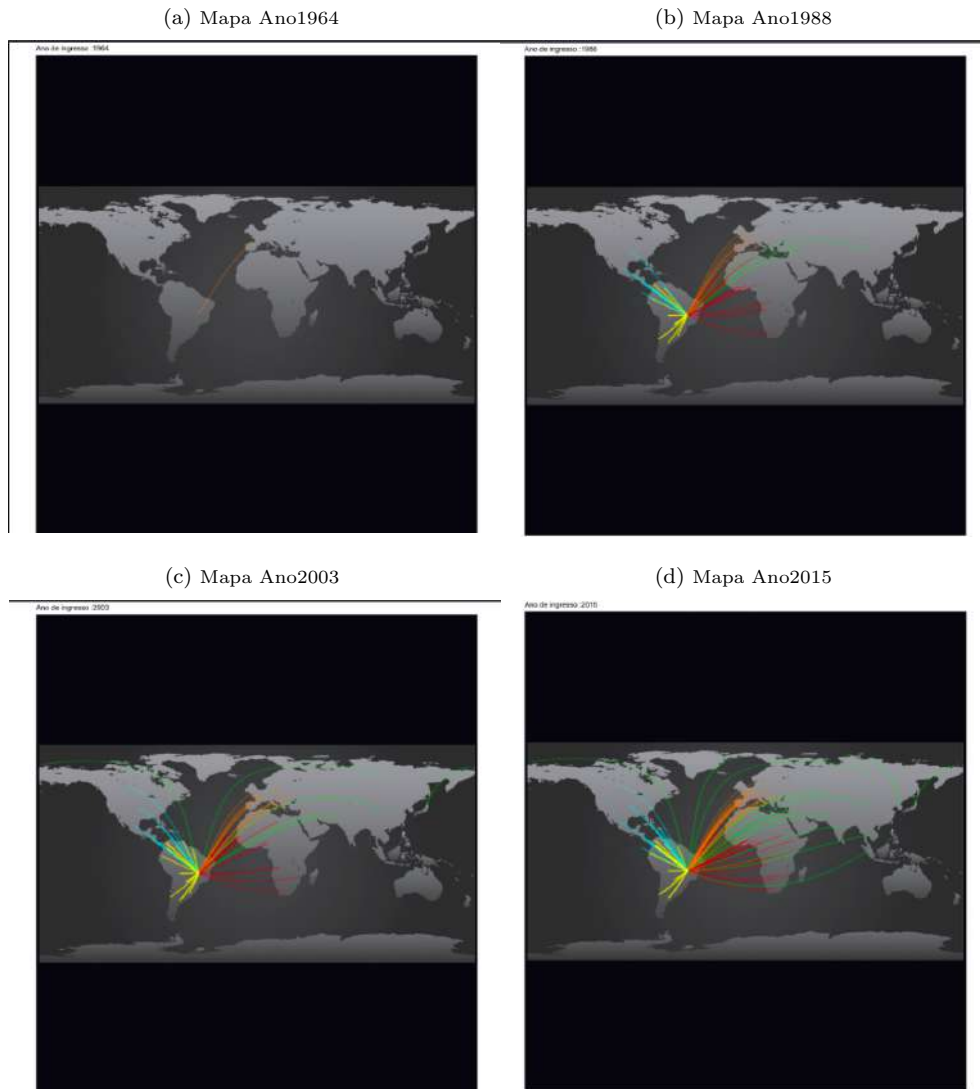
Figura 20: Visualização por Continente



Todos os 6 continentes estão presentes na Figura 20. Observa-se a ocorrência de alguns países no continente africano que estão com baixo fluxo, ou seja, têm as linhas finas: Tunísia, Mauritânia, Etiópia e Botsuana. Os países africanos mais representados são os lusófonos.

É possível perceber que houveram poucas conexões entre o continente asiático e a Universidade de Brasília. No geral, países como Myanmar, Tailândia, Vietnã, Filipinas e Indonésia apresentam linhas muito finas, o que indica que eles são poucos representados na Universidade. Observando o mapa da Europa, a ligação entre a Islândia e a UnB é extremamente fina por isso entende-se que dos europeus, os islandeses são os mais ausentes na universidade de Brasília. A Austrália, um país da Oceania, também é muito bem representada.

Figura 21: Sequências do Gif



O vídeo mostra basicamente a evolução do fluxo de estrangeiros na Universidade de Brasília ao longo dos anos. Interessante perceber no vídeo que o primeiro estrangeiro que chegou na UnB era francês e foi em 1962. Nos primeiros anos vieram muitos europeus. Ao longo dos anos vieram nessa ordem, asiáticos, norte-americanos, sul americanos, africanos e australianos. O primeiro africano que chegou na Universidade de Brasília era do Egito. O vídeo pode ser acessado no link youtu.be/j6gqEPAdpE.

4 Conclusão

Na primeira parte deste trabalho, foi apresentada uma visão global da área de visualização de dados e discutiu-se sobre a sua importância para os negócios e para a compreensão dos dados brutos. Foram expostas também algumas etapas do processo de projeto de visualização de dados e alguns cuidados que precisam ser tomados na realização desses projetos. Além disso, foram listadas as principais ferramentas utilizadas para visualização de dados e suas respectivas funcionalidades e características mais importantes.

Com os dados do aplicativo +BIKE, foi feito um mapa de conexão usando o pacote *leaflet*. Depois disso, o objetivo era fazer um vídeo usando todos os dados obtidos, mas como o processo era muito custoso e por falta de recursos computacionais adequados, decidiu-se por realizar o vídeo com base em uma amostra aleatória de 10000 viagens considerando somente os dados das segundas-feiras. Porém, mesmo reduzindo os dados, o processo teve que ser paralelizado para acelerar os resultados. Para auxiliar na realização desse vídeo foi utilizado o pacote *magic*.

Após essa etapa, foi criado um *dashboard* interativo utilizando o pacote *shiny* para facilitar a compreensão e a interação com o mapa. Com base nessas visualizações, foi possível perceber que as rotas mais usadas estão na região do eixo monumental e o período de maior fluxo é entre às 17h e às 19h.

No segundo projeto, o mapa de conexão foi feito com os dados dos estrangeiros da Unb. O procedimento para o desenvolvimento dessa visualização foi realizado com o emprego do pacote *ggplot2*. Para ter uma visualização mais detalhada e de fácil de compreensão, os mapas foram gerados por continente, assim, foi mais simples enxergar o fluxo de cada continente e país e para exaltar alguns detalhes que não foram possíveis perceber nos mapas, foi produzido um vídeo. Para o vídeo utilizou-se o pacote *gganimate*.

Com as visualizações sobre os estrangeiros da UnB concluídas, foi possível extrair algumas informações tais como os seis continentes estão presentes no mapa mas alguns países tem um fluxo muito baixo de estrangeiros, os países mais representados do continente africano são os lusófonos e no continente europeu, o país menos representado é a Islândia. É possível observar também que o primeiro estrangeiro a chegar na Universidade de Brasília era um francês e foi em 1962.

Os códigos e arquivos usados durante esse trabalho podem ser acessados no github.com/Guilherme-Souza-Rodrigues/TCC-DataViz.

Para um trabalho futuro os pontos seguintes poderiam ser levados em consideração:

- Realizar o vídeo do projeto das bicicletas usando todos os dados e um computador mais potente.
- Realizar o mapa dos estrangeiros na UnB usando o pacote *leaflet*.
- Criar o vídeo dos estrangeiros usando o pacote *magic*.
- Desenvolver o *dashboard* interativo utilizando os dados dos estrangeiros da Unb.
- Encontrar uma alternativa para o *shiny*, pois ele é muito lento e o desempenho cai quando tem muitos dados.
- Tentar colocar alguns efeitos visuais e *feedbacks* sonoros nos vídeos para deixar a visualização mais atraente.

Referências

- [1] ALVIM B. D. ; LAZARO D. , 2014. Disponível em: https://www.accenture.com/_acnmedia/PDF-45/Accenture-Entendido-De-Dados.pdf. Acesso em 17 de Novembro de 2020.
- [2] ASSIS A. F. , 2019. Disponível em: <https://www.igti.com.br/blog/importancia-da-visualizacao-de-dados-no>. Acesso em 22 de Novembro de 2020.
- [3] CAIRO A. **Revista do Programa de Pós-graduação em Comunicação Universidade Federal de Juiz de Fora**: Por que a visualização de dados falha se não pensamos sobre seu propósito, , mai./ago. 2019
- [4] CERTO. Samuel C. **Tomada de decisões. Administração moderna**. 9. ed. São Paulo: Pearson, 2005. cap. 7, p. 123.
- [5] CHIAVENATO, Idalberto. **Comportamento organizacional: a dinâmica do sucesso das organizações**. 2. ed. Rio de Janeiro: Elsevier, 2010. 539 p.
- [6] COMPUTER WORD, 2019. Disponível em: <https://computerworld.com.br/carreira/python-10-motivos-para-aprender-a-linguagem-em-2019/:.text=Python%20%C3%A9%20usado%20para%20%E2%80%9CScripting,como%20um>. Acesso em 12 de Outubro de 2020.
- [7] CUNHA, Luis. Rock Content, 2018. Disponível em: <https://comunidade.rockcontent.com/subtitulos/>. Acesso em: 08 de Outubro de 2020.
- [8] DECICINO Ronald. Uol, 2014. Disponível em: <https://educacao.uol.com.br/disciplinas/geografia/legendas-convencoes-graficas-ajudam-a-compreender-os-mapas.htm>. Acesso em: 08 de Outubro de 2020.
- [9] DEPARTAMENTO DE ESTATÍSTICA - UNIVERSIDADE FEDERAL DE JUIZ DE FORA. Disponível em: <https://www.ufjf.br/estatistica/eventos-e-projetos/projeto-sas/o-que-e-o-sas/>. Acesso em 12 de Outubro de 2020.
- [10] DURCEVIC, Sandra. Datapine, 2019. Disponível em: <https://www.datapine.com/blog/best-data-visualization-books/>. Acesso em: 30 de Outubro de 2020.
- [11] FIVE ACTS, 2016. Disponível em: <https://www.fiveacts.com.br/tableau-software-como-funciona-e-quais-sao-os-beneficios/>. Acesso em 12 de Outubro de 2020.
- [12] FONTES, Marcos Vinícius. Dataficação, 2017. Disponível em: <https://datafizacao.wordpress.com/2017/02/21/criando-mapa-brasil-r/>. Acesso em: 13 de Setembro de 2020.
- [13] GLOVER, Marissa. Psicologia Online, 2019. Disponível em: <https://br.psicologia-online.com/processos-cognitivos-basicos-e-superiores-exemplos-e-tipos-212.html>. Acesso em: 05 de Outubro de 2020.
- [14] MAXIMIANO, Antonio Cesar Amaru. **Introdução à Administração**. Ed. Compacta. São Paulo: Atlas, 2009. 294p.
- [15] MERITE, Gabrielle. Visme, 2020. Disponível em: <https://visme.co/blog/best-data-visualizations/>. Acesso em: 20 de Junho de 2020.
- [16] MICROSOFT, 2020. Disponível em: <https://docs.microsoft.com/pt-br/power-bi/fundamentals/power-bi-overview>. Acesso em 12 de Outubro de 2020.
- [17] MILLER, Sergio. Nosql BA, 2020. Disponível em: <https://nosqlba.org/uma-visao-geral-dos-tipos-de-grafico-no-power-bi/>. Acesso em 20 de Setembro de 2020.
- [18] MINARD, Charles Joseph. Sciencenorway, 2019 . Disponível em: <https://sciencenorway.no/blog-blog-from-numbers-to-graphics-statistics/charles-joseph-minards-map-of-napoleons-flawed-russian-campaign-an-ever-current-classic/1618695>. Acesso em: 30 de outubro de 2020.
- [19] NEOMOVE , 2019. Disponível em: <https://www.neomove.com.br/data-visualization/visualizacao-de-dados>. Acesso em 20 de Novembro de 2020.

[20] OLIVEIRA et tal. **Revista Destaques Acadêmicos: Técnicas de Visualização de Informação como Apoio á Gestão Estratégica**, 2018.

[21] QUOOS, João Henrique. Laboratório de cartografia. Disponível em: <http://coral.ufsm.br/cartografia/index.php?option>
Acesso em: 01 de julho de 2020.

[22] ROBBINS, Stephen; JUDGE, Timothy; SOBRAL, Filipe. **organizacional: teoria e prática no contexto brasileiro**. 14. ed. São Paulo: Pearson, 2010. 633 p.

[23] SIILVA F.C.C. **Visualização de dados: passado, presente e futuro**. Lincem Revista, Rio de Janeiro, v.15, n.2, 2019. 205-223 p.

[24] SANCHIS, Sara. Psicologia Online, 2020. Disponível em: https://br.psicologia-online.com/como-funciona-o-cerebro-humano-430.htmlanchor_3. Acesso em: 02 de Outubro de 2020.

[25] SAS. Disponível em: https://support.sas.com/sassamples/graphgallery/PROC_GMAP.html. Acesso em 20 de Setembro de 2020.

[26] SIGNIFICADOS, 2015. Disponível em: <https://www.significados.com.br/estetica/>. Acesso em 05 de outubro de 2020.

[27] TABLEAU. Disponível em: https://help.tableau.com/current/pro/desktop/pt-br/maps_howto_filledpiechart.htm. Acesso em 20 de Setembro de 2020.

[28] THE R FOUNDATION. R PROJECT. Disponível em: <https://www.r-project.org/about.html>. Acesso em 15 de Setembro.

[29] THIAGODV. Bastter, 2019. Disponível em: <https://bastter.com/mercado/forum/865539/projeto-1-python-montar-um-mapa-coropletico-com-dados-eleitorais-parte-1>. Acesso em 20 de Setembro de 2020.

[30] VARIG. Disponível: <https://www.varig-airlines.com/pt/rotas.htm>. Acesso em: 13 de Setembro de 2020.

[31] ZEVIANI W. , 2019. Disponível em: <http://leg.ufpr.br/~walmes/talks/visualizacao-dados.pdf>. Acesso em 15 de Novembro de 2020.

[32] ZEVIANI W. , 2018. Disponível em: <http://leg.ufpr.br/~walmes/cursor/data-vis/slides/datavis.pdf>. Acesso em 17 de Novembro de 2020.

•

Links Utilizados:

https://www.accenture.com/_acnmedia/PDF-45/Accenture-Entendodo-De-Dados.pdf
<https://www.datapine.com/blog/best-data-visualization-books/>
<https://r4ds.had.co.nz/index.html>
<http://www.dados.df.gov.br/dataset> // www.maisbikecompartilhada.com.br/
<https://cran.r-project.org/web/packages/googleway/vignettes/googleway-vignette.html>
<https://developers.google.com/maps/documentation/directions/overview>
https://www.rdocumentation.org/packages/googleway/versions/2.7.1/topics/google_directions
https://developers.google.com/public-data/docs/canonical/countries_csv
<https://www.vectorstock.com/royalty-free-vector/alarm-clock-icon-vector-19516879>
https://www.nasa.gov/specials/blackmarble/2016/globalmaps/BlackMarble_2016_01deg.jpg
<https://www.dataplanes.org/notes/2018/01/27/flight-routes-night-lights>
<https://flowingdata.com/2011/05/11/how-to-map-connections-with-great-circles/>
<https://www.gis-blog.com/flight-connection-map-with-r/>
<https://flowingdata.com/2011/05/05/where-do-major-airlines-fly-in-the-united-states/>