



Graduação em Ciência da Computação

Felipe Santana de Luna

**AIRFAREVIS: UMA FERRAMENTA DE VISUALIZAÇÃO PARA A
ANÁLISE INTERATIVA DE DADOS DE TARIFA AÉREA**

Trabalho de Graduação



Universidade Federal de Pernambuco
secgrad@cin.ufpe.br
www.cin.ufpe.br/~secgrad

RECIFE
2016



Universidade Federal de Pernambuco
Centro de Informática
Graduação em Ciência da Computação

Felipe Santana de Luna

**AIRFAREVIS: UMA FERRAMENTA DE VISUALIZAÇÃO PARA A
ANÁLISE INTERATIVA DE DADOS DE TARIFA AÉREA**

Trabalho apresentado ao Programa de Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: *Nivan Roberto Ferreira Júnior*

RECIFE
2016

Agradecimentos

Agradeço primeiramente a minha mãe, ao meu pai, ao meu irmão e a toda minha família por todo o suporte que recebi durante todo esses anos, eles são parte fundamental da minha formação e sempre me apoiaram durante todos os momentos, principalmente os mais difíceis da graduação. Sem eles nada seria possível.

Meus agradecimentos seguem ao professor Nivan Ferreira pela orientação e por estar sempre solícito durante a elaboração desse projeto.

Agradeço aos amigos que fiz durante esse percurso, sempre juntos nos apoiando durante momentos de dificuldades, tornando a graduação um caminho mais prazeroso. Tenho certeza que levarei muitos deles para toda a vida.

Também quero agradecer em especial a minha namorada. Mesmo não sendo da área de computação e muitas vezes achar loucura o que eu estudava, esteve sempre presente me motivando.

*isso de querer ser
exatamente aquilo
que a gente é
ainda vai
nos levar além*

—PAULO LEMINSKI

Resumo

O mercado aéreo tem uma grande participação na economia atual, movimentando bilhões de passageiros, bens e dinheiro no cenário mundial. Este mercado é extremamente acirrado, no qual diversas companhias competem por passageiros a fim de aumentar seus lucros. Nesta competição, as companhias tomam complexas estratégias de mercado com o intuito de atrair consumidores, tornando este mercado um ambiente bastante complexo. A fim de entender este complexo cenário, faz-se necessário uma análise dos dados para compreender padrões de comportamentos das companhias no que diz respeito as suas estratégias de mercado o que pode auxiliar não só as companhias, mas também os consumidores diretos (passageiros).

Os dados gerados no mercado de passagens aéreas são complexos envolvendo componentes espaciais e temporais. Dada esta complexidade, técnicas de visualização interativa de dados são reconhecidas como ferramentas eficientes que auxiliam na exploração e identificação de padrões em coleções de dados complexos. Estas técnicas tomam proveito do grande poder do sistema visual humano e dos computadores modernos, para criar representações visuais de dados que possibilitam aumentar a capacidade de cognição e, portanto, auxiliam na análise destes dados.

Neste trabalho foi desenvolvido um sistema de visualização de dados, chamado AirFareVis com o propósito de explorar dados de tarifas aéreas. Este sistema permite a exploração interativa de centenas de milhares de viagens. Para transpor a inerente complexidade dos dados, AirFareVis emprega diversas técnicas de visualização de dados como como vistas múltiplas coordenadas, *scented widgets*, *Gantt Chart*, dentre outras. O sistema foi desenvolvido usando tecnologias web, com o intuito de possibilitar mobilidade e facilidade de compartilhamento da aplicação.

Para validar a eficácia do sistema casos de uso são apresentados usando dados de promoções de passagens aéreas mineradas do site Melhores Destinos. Estes casos de uso mostram o poder da visualização interativa de dados na identificação de padrões de comportamento de mercado das companhias aéreas no que diz respeito a preço de passagens. Dessa forma, constatou-se que a ferramenta cumpriu com seu propósito, uma vez que foi possível fazer análises de casos com diferentes parâmetros, bem como pontuar alguns comportamentos, quais sejam a periodicidade de postagem de promoções e concorrência entre companhias.

Palavras-chave: Visualização de dados, mineração de dados, dados espaço-temporais, dados de origem e destino

Abstract

The air market has a large share in the current economy, moving billions of passengers, goods and money in a global scenario. This market is extremely fierce, in which several companies compete for passengers in order to increase their profits. In this competition, companies take complex market strategies in order to attract consumers, making this market a very complex environment. In order to understand this complex scenario, it is necessary to analyze the data to understand patterns of behavior of the companies with respect to their strategies of market which can help not only the companies but also the direct consumers (passengers).

The data generated in the airline ticket market are complex involving spatial and temporal components. Given this complexity, interactive data visualization techniques are recognized as efficient tools that assist in the exploration and identification of patterns in complex data collections.

In this work a data visualization system was developed, called AirFareVis for the purpose of exploring airfares data. This system allows the interactive exploration of hundreds of thousands of trips. To bridge the inherent complexity of the data, AirFareVis employs a variety of data visualization techniques, such as multiple coordinate views, *scented widgets*, *Gantt Chart*, and so on. The system was developed using web technologies, in order to allow mobility and ease of application sharing.

To validate the effectiveness of the system use cases are presented using airline ticket deals data mined from the Melhores Destinos website. These use cases show the power of interactive data visualization in identifying airlines' market behavior patterns with regard to ticket prices. In this way, it was verified that the tool fulfilled its purpose, since it was possible to analyze cases with different parameters, as well as punctuate some behaviors, such as the periodicity of posting promotions and competition between companies.

Keywords: Data Visualization, Data Mining, Spatio-Temporal Data, Origin-Destination Data

Lista de Figuras

4.1	Interface do AirFareVis	18
4.2	Primeira tentativa de visualização proposta usando série temporal	20
4.3	Segunda tentativa de visualização proposta	20
4.4	Visualização usando transparência para identificação de valores	21
4.5	Representação do Box-Plot de promoções de Recife para Aracajú	22
4.6	Gráfico da duração de promoções com Filtros	22
5.1	Gráficos da frequência de postagem usando todo o <i>dataset</i>	25
5.2	<i>Box-plot</i> comparando preços de todas as promoções do <i>dataset</i>	26
5.3	Comparação entre promoções de Recife para Teresina	26
5.4	Comparação entre promoções de Recife para Fortaleza	27
5.5	Comparação entre promoções de Recife para Aracaju	27
5.6	Comparação dos preços das promoções entre Recife e Miami	28
5.7	Comparação dos preços das promoções entre Recife e Cidade do México	28
5.8	Comparação dos preços das promoções entre Recife e Frankfurt	28
5.9	Comparação dos preços das promoções entre Recife e Buenos Aires	29
5.10	Frequência de postagem das promoções entre Recife e Santiago	29

Lista de Tabelas

2.1	Comparação de ferramentas de exploração entre websites de venda de passagens aéreas.	12
3.1	Uma amostra do conjunto de dados minerados do site Melhores Destinos. . . .	14
5.1	Quantidade de promoções das principais companhias aéreas.	24

Lista de Acrônimos

SVG	<i>Scalable Vector Graphics</i>
IATA	<i>International Air Transport Association</i>
JSON	<i>JavaScript Object Notation</i>
HTML	<i>HyperText Markup Language</i>
DOM	<i>Document Object Model</i>

Sumário

1	Introdução	10
1.1	Objetivos	10
1.2	Estrutura do trabalho	11
2	Trabalhos Relacionados	12
3	Análise e tratamento dos dados	14
3.1	Conjunto de dados	14
3.1.1	Latitude e Longitude	15
4	AirFareVis	17
4.1	Tarefas Analíticas	17
4.2	AirFareVis	17
4.2.1	Mapa	18
4.2.2	Gráficos de Barras Frequência	19
4.2.3	Gráficos de Barras de Duração	19
4.2.4	Box-Plot	21
4.3	Interação	22
4.3.1	Filtro temporal	23
4.3.2	Filtro de Preço	23
4.3.3	Companhias Aéreas	23
5	Casos de Uso	24
5.1	Dados Usados nos Casos de Uso	24
5.2	Casos de Uso	25
5.3	Recife e cidades do Nordeste	26
5.4	Recife e cidades no Exterior	27
6	Conclusão	30
6.1	Melhorias e Futuros Trabalhos	30
	Referências	32

1

Introdução

De acordo com a *International Air Transport Association* (IATA), no ano de 2016 são esperados um total 36.8 milhões de voos em todo mundo, movimentando um total de 3.6 bilhões de passageiros e gerando um lucro total que ultrapassa a casa dos 700 bilhões de dólares americanos (IATA). O mercado de passagens aéreas tem, portanto, uma grande participação na economia atual. A fim de sobreviver neste mercado extremamente competitivo, as diversas companhias aéreas empregam complexas estratégias de negócio para ganhar a preferência dos passageiros MIOTTO; SOUZA; DIEHL (2008). Em contra partida, usuários de linhas aéreas tentam entender o mercado para identificar oportunidades de compra de passagens. A fim de entender este complexo cenário, faz-se necessário uma análise dos dados para compreender padrões de comportamentos das companhias no que diz respeito as suas estratégias de mercado o que pode auxiliar não só as companhias, mas também os consumidores diretos (passageiros). Devido a grande informatização deste mercado, grandes quantidades de dados são produzidas. Um dos principais tipo de dados produzido diz respeito ao preço das passagens aéreas. Os dados de preço de passagens aéreas são complexos envolvendo componentes espaciais e temporais, isto é os preços variam de acordo com o tempo e também com a origem e destino dos vôos. A análise deste tipo de dados será o foco deste trabalho.

1.1 Objetivos

Dada esta complexidade, técnicas de visualização interativa de dados são reconhecidas como ferramentas eficientes que auxiliam na exploração e identificação de padrões em coleções de dados complexos. Estas técnicas são consideradas ferramentas indispensáveis neste cenário: elas usam o poder do sistema visual humano e ferramentas gráficas para possibilitar a exploração de dados complexos e também formulação e teste de hipóteses de maneira interativa (MUNZNER, 2014).

O propósito desse trabalho foi construir uma aplicação de visualização interativa, chamada AirFareVis (descrita em detalhes no Capítulo 4), para análise de dados sobre o mercado de passagens aéreas. Este sistema permite a exploração interativa de centenas de milhares de

viagens. Para transpor a inerente complexidade dos dados, AirFareVis emprega diversas técnicas de visualização de dados como como vistas múltiplas coordenadas, *scented widgets*, *Gantt Chart*, dentre outras. O sistema foi desenvolvido usando tecnologias web, com o intuito de possibilitar mobilidade e facilidade de compartilhamento da aplicação. Para a implementação, foram usadas as bibliotecas D3.js e Leaflet.

Para validar a eficácia do sistema casos de uso são apresentados usando dados de promoções de passagens aéreas mineradas do site Melhores Destinos (descritos em mais detalhes na Seção 3.1). Estes casos de uso mostram o poder da visualização interativa de dados na identificação de padrões de comportamento de mercado das companhias aéreas no que diz respeito a preço de passagens. Dessa forma, constatou-se que a ferramenta cumpriu com seu propósito, uma vez que foi possível fazer análises de casos com diferentes parâmetros, bem como pontuar alguns comportamentos, quais sejam a periodicidade de postagem de promoções e concorrência entre companhias.

1.2 Estrutura do trabalho

A organização do resto deste documento está descrita a seguir. O Capítulo 2 apresenta trabalhos relacionados na literatura, bem como ferramentas comerciais disponíveis para a compra de passagens aéreas. O Capítulo 3 discute a base de dados utilizada e descreve os tratamentos necessários para a base ser utilizada no AirFareVis. O capítulo 4 descreve o design do AirFareVis, discutindo as tarefas analíticas a serem executadas, bem como as técnicas de visualização utilizadas. O capítulo 5 apresenta dois casos de uso que mostram o AirFareVis em ação. Finalmente, no capítulo 6, são feitas considerações sobre o que o projeto alcançou e quais serão os possíveis trabalhos futuros.

2

Trabalhos Relacionados

Este capítulo discute trabalhos relacionados ao tema de análise visual de dados de tarifas aéreas.

Existem várias ferramentas destinadas ao consumidor final dos serviços aéreos que visam dar suporte ao processo de compra. De fato, é comum o uso de aplicações web tais como *skiplagged.com*, *kayak.com*, *google flights*, cujo objetivo é proporcionar a localização de passagens aéreas para a compra. Para tanto, elas apresentam alguns aspectos de exploração de dados como filtros e visualizações simples como mapas (utilizados para mostrar destinos populares e preços dada uma origem) e gráficos de linha (que mostram a variação de preço no tempo dada uma origem e um destino), como mostra a Tabela 2.1. Apesar de essas ferramentas serem eficientes no seu objetivo, elas não viabilizam o estudo e análise do mercado de passagens aéreas de uma maneira mais geral e, portanto, não são apropriadas para o objetivo proposto. Entretanto, alguns dos aspectos de exploração de dados discutidos acima são interessantes em contexto mais gerais do que somente no auxílio de compra de passagens e por isso serão utilizados no AirFareVis (Capítulo 4).

Tabela 2.1: Comparação de ferramentas de exploração entre websites de venda de passagens aéreas.

Website	Gráficos	Mapa	Calendário	Filtro Companhias	Filtro Preço
skiplagged.com	Sim	Não	Não	Sim	Sim
kayak.com	Não	Sim	Não	Sim	Sim
google flights	Não	Sim	Não	Sim	Sim

No estudo desse trabalho, se tem interesse em um típico específico de dados espaço-temporal chamado dados de origem e destino (FERREIRA, 2015). Abstratamente, tais dados, conhecidos como *OD-data* (*origin and destination data*) (WOOD; DYKES; SLINGSBY, 2010), correspondem a registros de movimentação no qual são armazenados somente informações sobre o local de partida e o local de chegada do movimento. A análise visual tem sido bastante investigada pela comunidade acadêmica. Técnicas gerais de visualização como OD Maps WOOD; DYKES; SLINGSBY (2010) e Flow Maps PHAN et al. (2005) foram desenvol-

vidas especificamente para visualizar diferentes aspectos deste tipo de dados. Adicionalmente, sistemas de visualização interativa foram desenvolvidos para analisar dados de origem e destino em diversos contextos como *bikesharing* (OLIVEIRA, 2015), corridas de taxi (FERREIRA et al., 2013), movimento populacional (BOYANDIN et al., 2011), dentre outros.

Diferentemente dos dados explorados nos trabalhos citados acima, os dados a serem explorados nesta monografia apresentam um carácter multi-variável, isto é, diversas companhias aéreas podem fazer o mesmo trajeto de origem destino. Por este motivo, as técnicas e sistemas mencionados não são adequados para o nosso propósito e daí a necessidade do desenvolvimento do AirFareVis.

3

Análise e tratamento dos dados

Este capítulo descreve os dados que motivaram a realização deste trabalho, bem como o processamento feito nos mesmos para que pudessem ser visualizados no AirFareVis.

A base de dados utilizada neste trabalho foi obtida através de um processo de mineração de dados do site *Melhores Destinos*. Este site é um dos principais sites de compra de passagens aéreas do Brasil e, portanto, é uma fonte valiosa de dados no que diz respeito aos preços (promocionais) de passagens aéreas. Os dados foram minerados pelo professor Luciano de Andrade Barbosa do Centro de Informática da UFPE e foram cordialmente cedidos para a realização deste trabalho.

3.1 Conjunto de dados

O *dataset* é composto por informações sobre promoções de passagens aéreas entre o início de outubro de 2013 e o final de dezembro de 2014, contendo 211496 entradas. Os atributos deste conjunto de dados incluem data de postagem da promoção, origem e destino, data de início da viagem, data de retorno, preço da promoção e, finalmente, o nome da companhia aérea responsável pelo voo, como pode ser visto na Tabela 3.1.

Tabela 3.1: Uma amostra do conjunto de dados minerados do site Melhores Destinos.

Post	Origin	Destination	Price	Start	End	Carrier
11/08/2014	RiodeJaneiro	Bologna	1836	23/12/2014	14/01/2015	Alitalia
11/08/2014	RiodeJaneiro	Bologna	1836	23/12/2014	15/01/2015	Alitalia
11/08/2014	RiodeJaneiro	Bologna	1836	23/12/2014	16/01/2015	Alitalia
11/08/2014	RiodeJaneiro	Bologna	1836	30/12/2014	07/01/2015	Alitalia
11/08/2014	RiodeJaneiro	Bologna	1836	30/12/2014	08/01/2015	Alitalia
11/08/2014	RiodeJaneiro	Bologna	1836	30/12/2014	14/01/2015	Alitalia
11/08/2014	RiodeJaneiro	Bologna	1836	30/12/2014	15/01/2015	Alitalia

3.1.1 Latitude e Longitude

Como visto na Tabela 3.1, os dados contém somente o nome da cidade de origem e destino. A fim de ser possível visualizar a distribuição espacial destes elementos, foi necessário realizar um pre-processamento para obter as coordenadas geográficas destas cidades, e para isso foi utilizado a API do Google Maps para conseguir esses dados. Para isso, foi construído um script em Node.JS que, para cada cidade, uma requisição era feita para a API a cada dois segundos, retornando um objeto no formato *JavaScript Object Notation* (JSON), que entre as informações recebidas continha a latitude e longitude da cidade requisitada. Infelizmente, há uma má interpretação devido à semelhança de nomes de cidades ao redor do mundo: onde claramente sabemos que a cidade dos dados se refere a uma cidade brasileira, a API retorna uma cidade de outro país. Nesses casos, o tratamento teve que ser realizado à mão. Após essa requisição, havia um array de objetos, cada objeto referente a uma cidade, como visto no trecho de código 3.1.

Trecho de Código 3.1: "Exemplo de Objeto contendo Geolocalização de uma cidade".

```
{ "city": "Recife",
  "location": {
    "lat": -8.0578381,
    "lng": -34.8828969
  }
}
```

Depois disso, foi feita uma outra etapa de pré-computação com o objetivo de agregar dados referentes viagens para cada cidade. Esta agregação tem o objetivo de otimizar a quantidade de dados a ser carregados e também otimizar estes dados para a visualização. Um exemplo do resultado desta etapa de agregação, pode ser visto em 3.2. Cada cidade contém um objeto neste formato, contendo além da informação de latitude e longitude, a quantidade de promoções para cada companhia quando a cidade é a origem da viagem e também quando é o destino.

Trecho de Código 3.2: "Exemplo do objeto completo usado para povoar o mapa".

```
{ var cities = [
  {
    "city": "Recife",
    "location": {
      "lat": -8.0578381,
      "lng": -34.8828969
    },
    "flightsOrigin": [
      {
        "carrier": "Azul",
```

```
    "flights": 154
  },
  {
    "carrier": "Gol",
    "flights": 382
  },
  {
    "carrier": "Tam",
    "flights": 428
  },
  {
    ...
  }
],
"totalOrigin": 1577,
"flightsDestination": [
  {
    "carrier": "Azul",
    "flights": 154
  },
  {
    "carrier": "Gol",
    "flights": 373
  },
  {
    "carrier": "Tam",
    "flights": 373
  },
  {
    ...
  }
],
"totalDestination": 961
}
```

4

AirFareVis

Este capítulo apresenta o design do AirFareVis. Este processo de design se deu de forma iterativa, no qual tarefas analíticas foram elicitadas (Seção 4.1) e estas foram usadas na escolha das visualizações e interações implementadas no sistema.

4.1 Tarefas Analíticas

O sistema AirFareVis foi desenvolvido para suportar as seguintes tarefas analíticas:

- (T1) Já que o tema refere-se a promoções de passagens aéreas, foi considerado que a visualização do espaço geográfico de origem e destino de cada promoção seria de grande importância para análise do comportamento das promoções. Dessa forma, seria possível ver quais regiões têm mais promoções, quais são os destinos e origens com mais promoções, e quais são os destinos com mais promoções para cada origem;
- (T2) Entender o padrão temporal das postagens de preço (promoções) de passagens em relação ao dia da semana e mês do ano para diferentes origens e destinos;
- (T3) Entender a competição entre as companhias em especial no que diz respeito à relação entre o dia de postagem e o dia de saída de um voo (duração da promoção);
- (T4) Entender a variabilidade do preço de promoções para as diferentes companhias, dada uma origem e um destino;
- (T5) Promover filtros capazes de auxiliar o usuário na exploração destes dados.

4.2 AirFareVis

Nesta seção é descrito o design do AirFareVis para suportar as tarefas analíticas descritas na seção anterior. A Figura 4.1 apresenta a interface do sistema desenvolvida para suportar

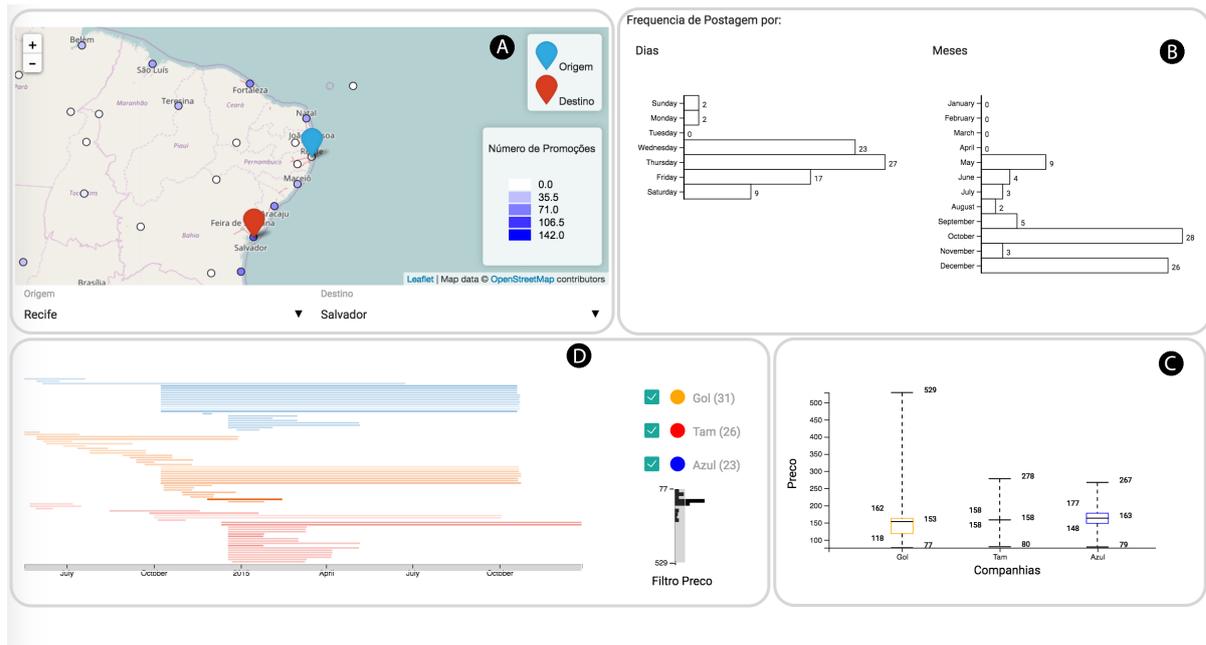


Figura 4.1: Interface do AirFareVis.

as tarefas analíticas descritas na seção anterior. Esta interface é composta de quatro vistas coordenadas. A primeira consiste em um mapa que é utilizado para a visualização do aspecto geográfico dos dados. A segunda consiste em dois gráficos de barra para visualizar a distribuição temporal das datas de postagem dos preços de passagens aéreas. A terceira é um *box plot* utilizado para sumarizar o preço das passagens aéreas visualizadas no sistema. Finalmente, a última vista consistem em *Gantt-Chart* para a visualização do período entre a postagem do preço de uma passagem e a partida do voo correspondente. Estas vistas bem como as interações do sistema são descritas nas seções que seguem.

4.2.1 Mapa

Para dar suporte a tarefa analítica (T1), a primeira visualização construída foi o mapa, nele é possível ver todas as cidades presentes no conjunto de dados, tanto de origem como de destino. Elas são apresentadas como pequenos círculos sobre o mapa. Cada cidade é colorida de acordo com uma escala referente a quantidade de promoções em que a cidade é a origem, isso é feito para que, se existirem padrões de características espaciais, como por exemplo quais regiões que tem o maior número de promoções, estes possam ser vistos apenas usando o mapa. Como visto na Figura 4.1 A

Também foi implementado o comportamento de mudar a coloração a partir da interação do usuário. Quando o mesmo escolhe uma origem para filtrar os dados, o mapa reflete essa seleção pintando todos os destinos em promoções onde apenas a cidade selecionada existe como origem, permitindo que o escopo da análise seja alterado para a seleção.

A implementação do mapa foi feita utilizando a biblioteca javascript *Leaflet.js*, dire-

cionada à construção de aplicações web com mapas interativos. Além de proporcionar uma interação de *drag and zoom* YOU et al. (2007) já estabelecido como comportamento padrão em aplicações web, a biblioteca proporciona facilidades na composição dos elementos no mapa, como em adicionar marcadores indicando posições no mapa e criar camadas sobre o mapa que permitam a inserção de elementos em *Scalable Vector Graphics* (SVG) e *HyperText Markup Language* (HTML)..

Além do Leaflet, foi também utilizada uma biblioteca para a manipulação da *Document Object Model* (DOM), SVG HTML, elementos essenciais para a criação de visualizações interativas BOSTOCK; OGIEVETSKY; HEER (2011), e, no presente trabalho, todas as visualizações foram criadas com o auxílio da biblioteca *D3.js*, que permite a construção dos elementos visuais em um mapeamento o com os dados.

4.2.2 Gráficos de Barras Frequência

Foi dado suporte à tarefa analítica (T2) com a construção de mais dois *widgets*. A fim analisar a frequência de postagem das promoções no website, foram utilizados gráficos de barras, um método de visualização antigo FRIENDLY (2006) e simples, mas que consegue representar efetivamente a distribuição das postagens de acordo com dia da semana e mês do ano. Os gráficos, inicialmente, exibem a frequência de todas as promoções, no entanto, cada filtragem refletem a frequência do novo escopo selecionado. Figura 5.1

4.2.3 Gráficos de Barras de Duração

Esta seção descreve o *widget* destinado a dar suporte a tarefa (T3). Devido à maiores complexidade da mesma, o design desta visualização passou por várias tentativas. A ideia era que, ao usuário escolher uma origem e um destino, a visualização informasse não só quais companhias tinham as promoções, mas também informar a duração e em que período do ano as passagens ficaram/ficariam disponíveis para compra.

Sendo assim, uma solução de caráter temporal parecia ser essencial para essa visualização, por isso inicialmente foi construída uma série temporal, Figura 4.2, que plotava a data em que as promoções eram postadas. Geralmente, séries temporais são usadas em tarefas de análises que envolvem encontrar tendências, como variações em períodos de tempo. MUNZNER (2014).

Porém esta visualização não refletia com clareza a informação que se queria passar. Apesar de conseguir mostrar a ideia de tempo passado e a relação com o preço, não era possível tirar nenhuma conclusão sobre o intervalo tempo entre a data da postagem da promoção e a data da viagem, assim, outras alternativas foram testadas.

Depois da implementação da série temporal, foi testado um gráfico diferente, o eixo *x* continuava sendo o tempo e o eixo *y* os preços, porém desta vez ao invés de plotar apenas as datas de postagem, foram plotadas, juntamente, as datas do voo e ligados entre si por uma

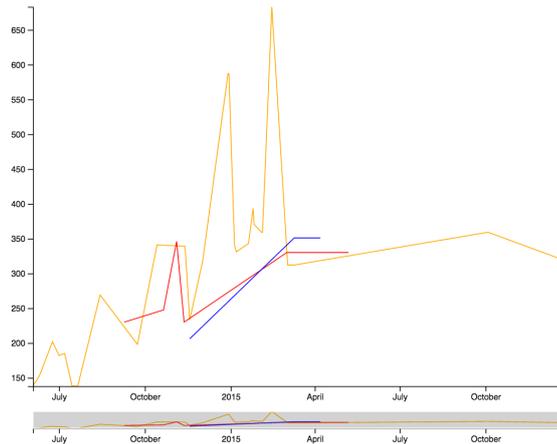


Figura 4.2: Primeira tentativa de visualização proposta usando série temporal.

linha horizontal. A posição da linha no eixo x indicava a localização da promoção no espaço temporal e o comprimento da linha indicava a relação de distância entre a data da postagem e a data da viagem. A posição da linha no eixo y continuava representando o preço da promoção. O problema é que como muitas promoções possuem preços próximos, a visualização ficava bastante poluída, como podemos ver na Figura 4.3, não tornando fácil a análise, uma vez que em muitos casos fica impraticável a comparação entre duas ou mais promoções que estejam muito próximas no eixo y . Além disso, em muitos casos havia a sobreposição quando o preço era o mesmo.

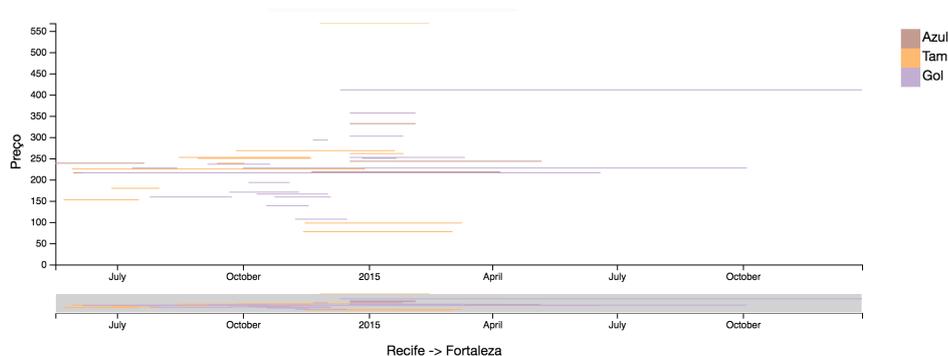


Figura 4.3: Segunda tentativa de visualização proposta.

Na terceira abordagem, a visualização construída da Figura 4.3 foi modificada, o eixo y que antes se referia aos preços passou a identificar as companhias. O posicionamento no eixo x continuava igual à visualização anteriormente descrita, tendo o início e o fim posicionados de acordo com as datas de postagem e de início da viagem. Como havia várias passagens de uma mesma companhia, uma sobreposição ficou evidente, mostrando apenas uma única barra, não sendo possível identificar quantas passagens, nesse intervalo de tempo, pertenciam a uma mesma companhia, o que deixou a visualização ineficaz, não atendendo ao objetivo.

Para contornar isso, valores de transparência foram inseridos, podendo assim identificar que regiões com as cores mais transparentes teriam menor ocorrência de promoções naquele

período e regiões da barra mais escuras teriam maior ocorrência, como pode ser visto na Figura 4.4. Infelizmente, essa abordagem de transparência não se revelou muito eficiente, já que após um certo número de barras sobrepostas, o elemento de transparência se esvai, deixando a região opaca. Já que existe um limite de transparência, fica impossível de analisar a quantidade de promoções sobrepostas, seja a o índice de transparência de cada barra, e $n \cdot a = 1$, qualquer número de promoções maior ou igual a n será representada com uma região opaca.

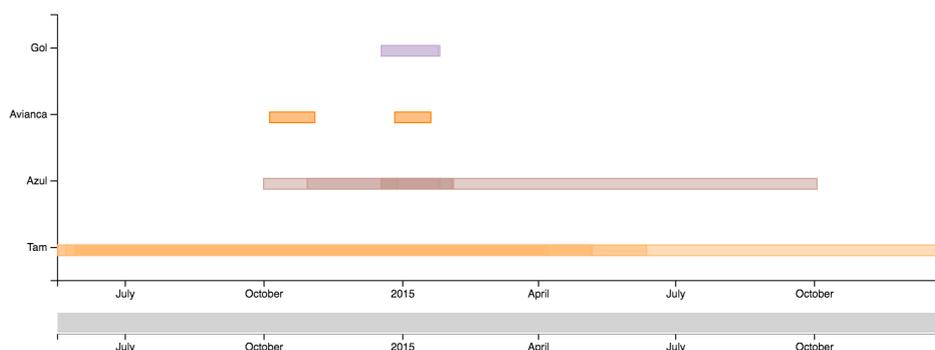


Figura 4.4: Visualização usando transparência para identificação de valores

Novamente, modificações foram realizadas. Como o objetivo era mostrar a relação de periodicidade das promoções, foi resolvido basear-se no *Gantt Chart*, descrito primeiramente em GANTT (1919), técnica visual utilizada para indicar as datas iniciais e finais de elementos, mostrando assim o intervalo entre a data de postagem da promoção e a data do embarque. Já que eixo x indica a variação de tempo ele permaneceu inalterado.

O eixo y , por outro lado foi removido completamente, a identificação das companhias aéreas ficou exclusivamente relacionada às cores das barras que agora são mostradas empilhadas. A altura das barras é calculada de acordo com a quantidade de promoções no subconjunto que quer ser representado. O preço fica sendo representado pela escala de cores que varia da cor da companhia, alterando de um tom mais claro para um mais escuro, representando o tom mais claro as promoções mais baratas e o mais escuro as mais caras. O resultado final pode ser visto na Figura 4.6, juntamente com os filtros de tempo sobre o eixo x , de companhias ao lado da legenda, e o de preço logo abaixo das legendas.

4.2.4 Box-Plot

Com o intuito de dar a suporte a tarefa (T4), foi resolvido construir um *box-plot* Figura 4.5. Box-plots são geralmente usados para resumir e comparar grupos de dados, usando mediana, primeiro e terceiro quartis, e os valores mínimo e máximo, sendo assim facilmente utilizados para encontrar valores discrepantes WILLIAMSON; PARKER; KENDRICK (1989). No projeto o box-plot exibe os valores referentes às companhias aéreas e sofre mudanças a partir do escopo dos dados observados no momento.

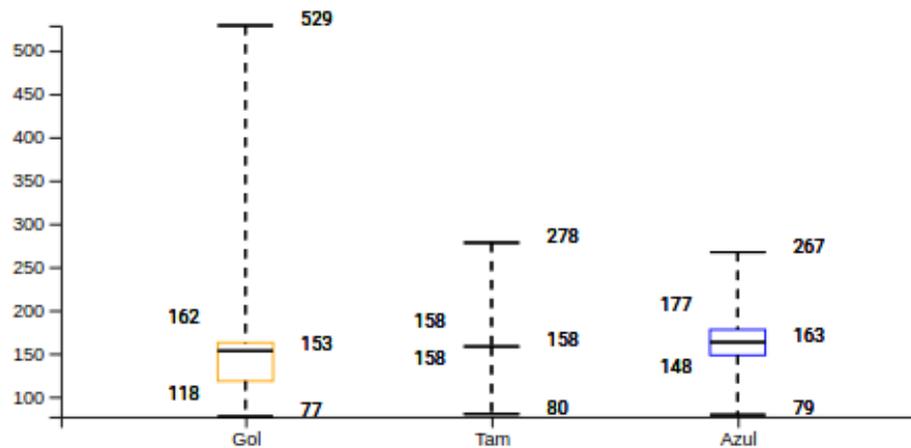


Figura 4.5: Representação do Box-Plot de promoções de Recife para Aracajú.

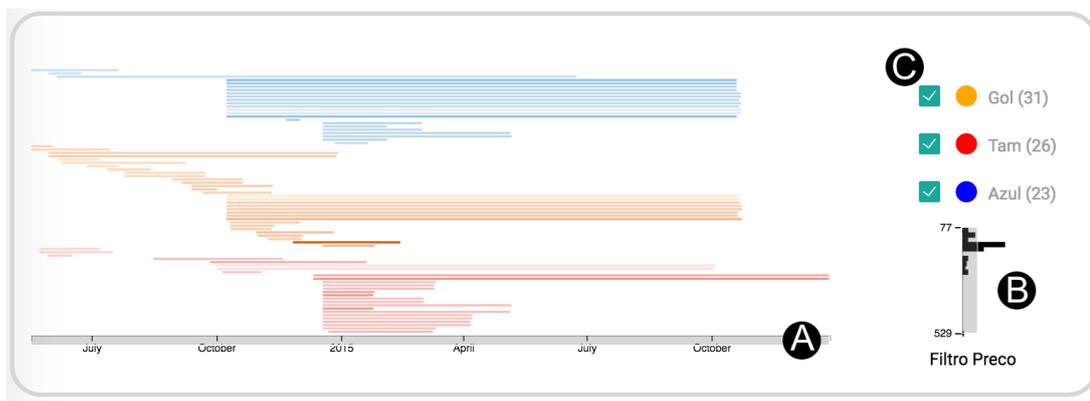


Figura 4.6: Gráfico da duração de promoções com Filtros.

4.3 Interação

Como descrito anteriormente, o projeto foi construído com o objetivo de permitir a interação do usuário a fim de que o mesmo possa descobrir tendências por meio da exploração, como dito na tarefa (T5). Para isso foram empregadas técnicas de vistas múltiplas coordenadas, partindo-se da premissa de que os usuários entendem melhor os dados quando interagem com eles ROBERTS (2007).

Por isso, a fim de alterar os escopo dos dados e que este fosse refletido nas visualizações, filtros foram criados. A cada mudança nos filtros, os dados eram atualizados e refletidos nas três visualizações. Abaixo são descritos os filtros utilizados.

4.3.1 Filtro temporal

O filtro temporal está diretamente relacionado ao caráter quantitativo representado pela escala de tempo utilizada no gráfico descrito em 4.2.3. O controle do escopo está diretamente relacionado à seleção de uma área horizontal posicionada em cima da escala no eixo x , essa localização permite que o usuário tenha controle do filtro enquanto observa as mudanças visuais acontecendo na mesma região. Figura 4.6 A.

A implementação foi feita por meio de uma barra onde o usuário pode arrastar mudando o início e o fim na escala de tempo, ou selecionar um área de *brush*. Essa mudança é refletida em tempo real nas barras que representam as promoções.

Assim, o filtro manipula duas variáveis que representam o limite inferior e o superior de datas de postagem, se uma promoção não foi postada no site no intervalo selecionado ela irá sumir da visualização.

4.3.2 Filtro de Preço

Outra importante maneira de reduzir o escopo dos dados é através dos preços das passagens, sendo assim, também foi construído um *widget* para a redução do escopo. Foi utilizado o mesmo princípio do filtro temporal descrito em 4.3.1, ou seja, uma área de seleção onde o usuário delimita o preço mínimo e máximo na qual as promoções mostradas acontecem nesse intervalo.

Essa técnicas de filtros por si só não são suficientes, as sugestões de navegação visual apropriadas podem ajudar os usuários guiando e refinando sua exploração. WILLETT; HEER; AGRAWALA (2007).

Assim, foi implementado um *Scented Widget* juntamente com o filtro de preço. Um histograma é posicionado ao lado da área de filtragem com o intuito de guiar os usuários para a região com o maior número de promoções. Figura 4.6 B.

4.3.3 Companhias Aéreas

Além dos atributos quantitativos dos dois filtros mencionados anteriormente, a aplicação possui o atributo categórico das companhias aéreas. Presume-se que durante a análise o usuário queira filtrar a informação que é mostrada pelas visualizações por alguma companhia específica.

Juntamente ao gráfico descrito na seção 4.3.1 é fornecido ao usuário uma legenda informando quais companhias aéreas fornecem as promoções de acordo com a seleção de origem e destino. Essa legenda também serve como filtro, como pode ser visto Figura 4.6 C, podendo o usuário selecioná-la individualmente, alternando a visualização para mostrar ou esconder as promoções que já estão sendo mostradas no gráfico.

5

Casos de Uso

5.1 Dados Usados nos Casos de Uso

A fim de demonstrar a utilidade do AirFareVis, utilizamos um subconjunto dos dados originais. Esta decisão tem como base a grande quantidade de companhias aéreas presentes nos dados (mais de 50 companhias aéreas diferentes). Uma análise simples mostra que as seis companhias que têm o maior número de promoções são: Tam, Gol, Combo, Delta, Copa e Azul.

Tabela 5.1

Tabela 5.1: Quantidade de promoções das principais companhias aéreas.

Companhia	% de Promoções	Qtd. de Promoções
Tam	18.77%	39710
Gol	13.67%	28913
Combo	9.71%	20545
Delta	9.68%	20488
Copa	8.72%	18446
Azul	5.09%	10772
Total	100%	211495

Observa-se que, logo após a Tam e a Gol, o terceiro lugar fica com **Combo**, liderando as promoções. Infelizmente, não se encontrou nenhuma companhia aérea com esse nome, esse termo, provavelmente, é usado quando a promoção refere-se a mais de uma companhia. Como não é possível decidir quais companhias estão envolvidas, esse conjunto de dados foi descartado, já que não representa, claramente, uma competição entre companhias.

Logo em seguida, nos quarto e quinto lugares, encontram-se a Delta e a Copa. Ao verificar os principais destinos destas duas, conclui-se que a maioria das promoções são para o exterior do Brasil. Seleccionando os 17 destinos mais requisitados do mundo, dentre as companhias analisadas, nota-se que: Delta e Copa possuem 0% dos destinos para estados brasileiros; Tam com 35%, seguida da Gol, com cerca de 64%, e finalmente dos 17 destinos com mais promoções da Azul, apenas 1 deles era pro exterior, sendo no total 94% dos destinos brasileiros.

Sendo assim foram escolhidas apenas promoções que pertenciam as companhias Gol, Tam e Azul, que juntas representavam quase 40% das promoções dos dados originais. Com esse escopo reduzido e definido, é feita uma melhor análise do mercado nacional.

5.2 Casos de Uso

O intuito da aplicação é usar visualizações para realizar análises comparativas intuitivas entre as companhias aéreas. Nesta seção serão apresentados alguns casos de uso, no qual serão feitas análises por meio da utilização da aplicação.

Tendo em vista o caráter de análise exploratório, os casos de uso foram testados usando a aplicação e seus filtros a fim de encontrar padrões no comportamento dos dados.

A princípio, apenas observando os gráficos de barra referentes à periodicidade de postagem das promoções, vê-se que os dados de todas as promoções são exibidos, além de que que, no geral, as publicações das promoções acontecem no final da semana, sendo 21298 publicações nas sextas e 14444 no sábado. Já em relação aos meses, percebe-se que os meses em que se têm a maior quantidade de promoções são os de julho, seguidos de maio e agosto. Figura 5.1

Frequencia de Postagem por:

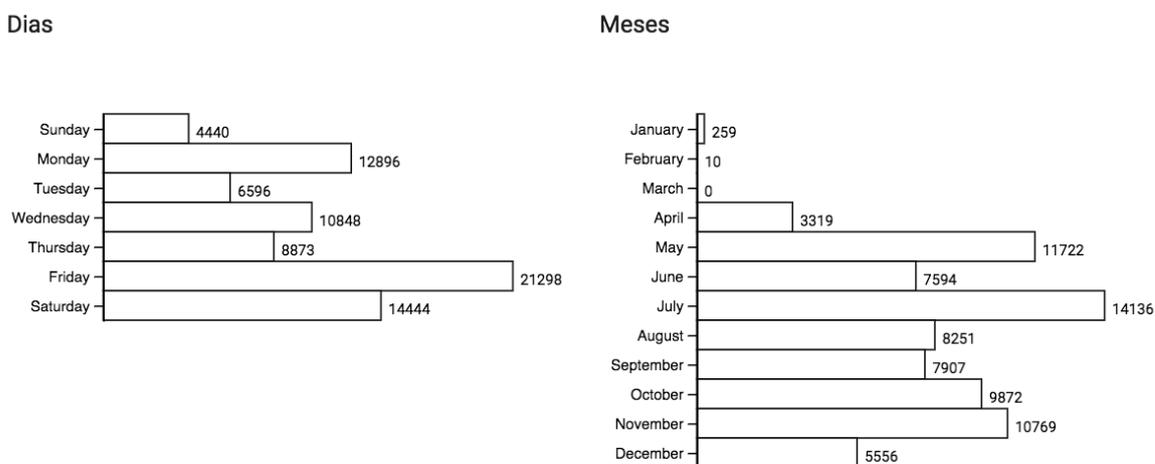


Figura 5.1: Gráficos da frequência de postagem das promoções de todo o *dataset*.

Além disso, é possível fazer uma análise comparativa, relacionando as medidas de preço das companhias exibidas pelo box-plot. Com todas as promoções sendo exibidas, observa-se que as companhias Gol e Azul possuem medidas de preços parecidas, ao contrário da concorrente Tam, que não apenas possui o máximo valor discrepante, como também o grande intervalo entre o primeiro e terceiro quartil, sendo estes maiores do que das outras duas companhias. Figura 5.2

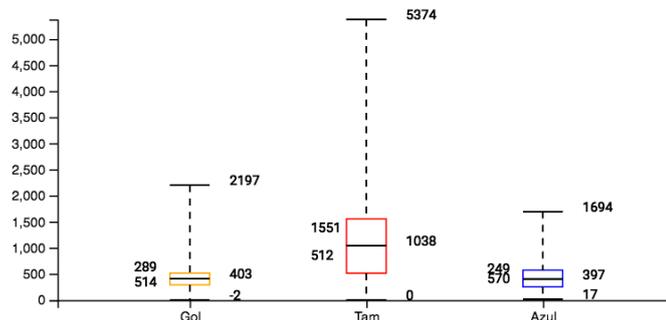


Figura 5.2: Box-plot comparando preços de todo o dataset.

5.3 Recife e cidades do Nordeste

Nessa análise, Recife foi fixada como cidade de origem e se analisou o comportamento das promoções para outras cidades do nordeste, tais como Aracaju, Teresina e Fortaleza.

Com a interação de escolher através do mapa os destinos, alguns fatos podem ser observados, como por exemplo:

- Com destino a Teresina a maioria das promoções são da Azul e Tam, sendo apenas três, das trinta e oito promoções, da Gol. Como pode ser visto na Figura 5.3 a maioria das promoções da Azul acontecem em dezembro, ao passo que as da Tam acontecem geralmente em Maio e Junho. Também se pode perceber que as promoções da Tam que acontecem no final do ano têm preços maiores do que os da Azul.

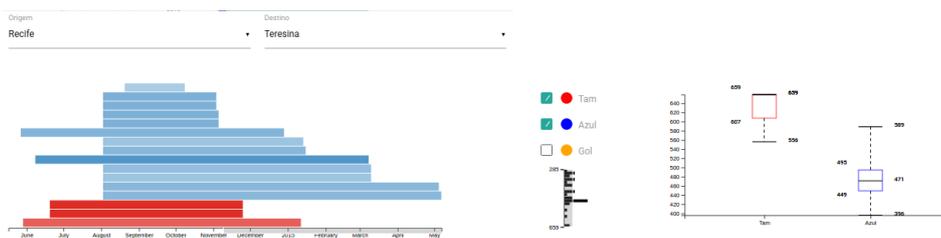


Figura 5.3: Comparação entre promoções de Recife para Teresina.

- Já quando o destino é Fortaleza, as companhias que mais competem no que diz respeito à quantidade de promoções postadas são Tam e Gol. Há uma quantidade de postagens maior no final do ano, apesar de a Tam possuir um valor discrepante no geral, as duas possuem preços semelhantes. Isto por ser visto na Figura 5.4

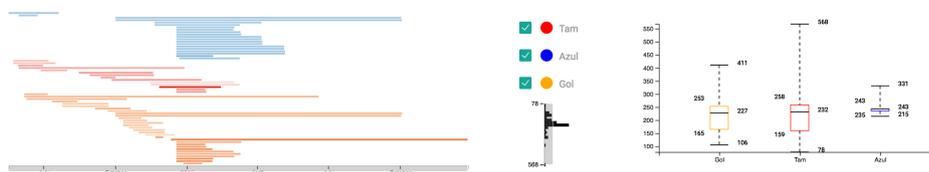


Figura 5.4: Comparação entre promoções de Recife para Fortaleza.

- Pode-se notar um comportamento semelhante quando o destino é Aracaju. Por meio do gráfico que mostra a distribuição temporal das promoções Figura 5.5, é possível constatar que a quantidade de promoções da Tam é muito superior à das outras duas, contabilizando cerca de 80% de todas as promoções. Estas ocorrem em grande parte do ano, sendo com maior frequência no final. A Gol, mesmo com poucas promoções nessa época, oferece passagens com custo maior que as concorrentes.

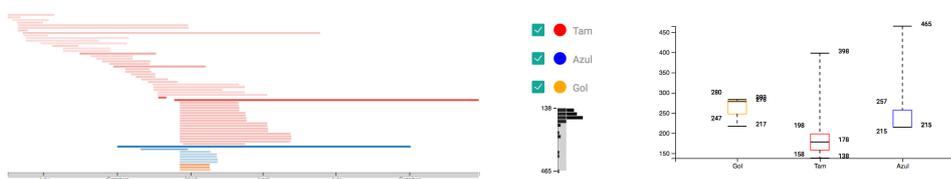


Figura 5.5: Comparação entre promoções de Recife para Aracaju.

5.4 Recife e cidades no Exterior

Se for analisar, por outro lado, a relação de promoções com destino ao exterior, partindo de Recife, pode-se ver no mapa que as cidades que possuem mais promoções partindo de Recife são:

- Na América do Norte: Miami e Cidade do México.
- Na América do Sul: Santiago, Buenos Aires e Montevidéu.
- Na Europa: Frankfurt e Londres.

Constata-se que as promoções para Miami e Cidade do México são apenas da Tam. Nesse sentido, é possível reparar, pela Figura 5.6, que promoções para Miami tem uma grande concentração de passagens postadas no final do ano. Enquanto para o México existe uma grande

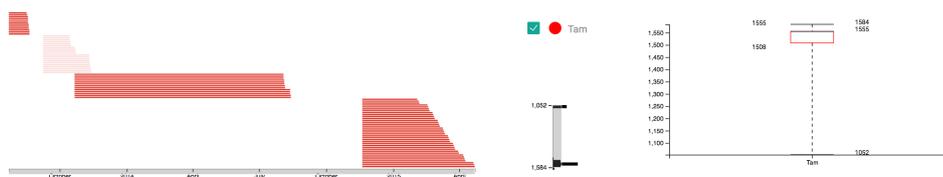


Figura 5.6: Comparação dos preços das promoções entre Recife e Miami.

concentração no mês de agosto, Figura 5.7, porém a data das viagens são de Abril/Março do ano seguinte.

Ademais, verifica-se que a maioria das passagens para Miami se concentram na região de maior preço da escala, já a do México na de menor.

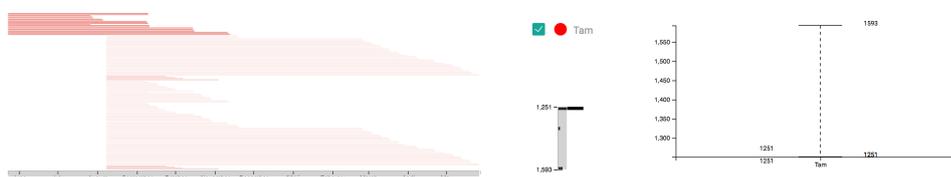


Figura 5.7: Comparação dos preços das promoções entre Recife e Cidade do México.

Analisando as promoções com destino aos dois países da Europa, percebe-se que, novamente, a Tam se mostra a única a conter promoções, estas são postadas entre Abril e Julho com datas de viagem para o final do ano. Figura 5.8



Figura 5.8: Comparação dos preços das promoções entre Recife e Frankfurt.

Nos países da América do sul, nota-se que as promoções para Buenos Aires existem com as companhias Gol e Tam, essas, porém, demonstram-se bem isoladas, tanto em relação à data de postagens quanto aos preços, como pode-se ver em Figura 5.9. As mais baratas são da Gol, que tem suas postagens no meio do ano para voos de datas entre outubro e dezembro de 2014. Já a Tam tem suas promoções postadas no final do ano de 2014, com data de início para o meio do ano de 2015.

Pode-se analisar, ainda, que houve uma alternância no mercado, antes as promoções da Gol para essa cidade são agora ofertadas pela Tam.

Em Santiago, praticamente todas as passagens são postadas na metade do ano, como se pode observar em Figura 5.10, para viagens de embarque em Novembro e Dezembro, sendo o mercado exclusivo da Tam.

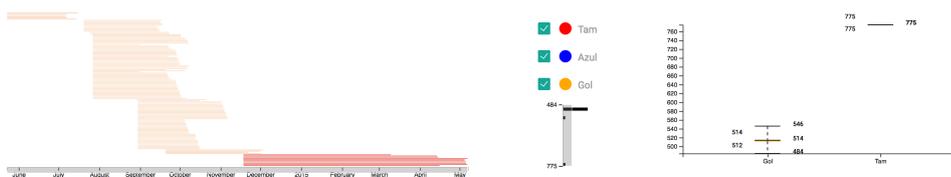


Figura 5.9: Comparação dos preços das promoções entre Recife e Buenos Aires.

Frequencia de Postagem por:

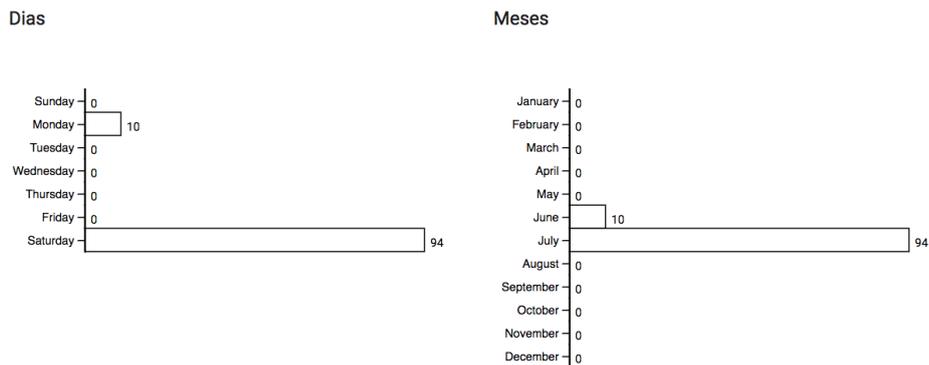


Figura 5.10: Frequência de postagem das promoções entre Recife e Santiago.

Para Montevideú, as passagens são postadas em julho e agosto, para voos no final do ano e todas são pertencentes à Gol.

Com a breve análise feita acima, é possível notar diversos resultados ao combinar destinos e origens diferentes. Além disso, com uma análise mais detida, constata-se que a maior parte dos voos ocorrem entre os meses de dezembro/ janeiro ou julho/agosto, que são épocas consideradas de alta temporada para viagens, uma vez que coincidem com as férias. Assim, pela aplicação, pode o próprio usuário fazer diversas pesquisas e observar diferentes padrões e comportamentos.

6

Conclusão

Este trabalho propôs um design de uma ferramenta de visualização interativa de dados (AirFareVis) para a exploração de dados de passagens aéreas. Esta ferramenta foi construída usando tecnologias web a fim de obter uma aplicação independente de sistemas. Aproveitando-se da ubiquidade da internet, uma aplicação de visualização de dados voltada a web pode ser utilizada e compartilhada sem a necessidade de nenhuma dificuldade. AirFareVis emprega várias técnicas de visualização de dados para auxiliar o usuário em tarefas analíticas em tempo real. Apesar de ser um design simples e inicial, acreditamos que o AirFareVis atingiu o objetivo esperado de ser uma ferramenta que possibilita o rápido reconhecimento de padrões e tendências de atributos qualitativos e quantitativos, tais como medidas de tempo e preço.

A próxima seção deste capítulo descreve dificuldades enfrentadas, sugestões de possíveis futuros trabalhos e possíveis melhorias.

6.1 Melhorias e Futuros Trabalhos

Por falta da interação com um especialista da área de transporte e mercado de passagens aéreas, o trabalho foi construído baseado em perguntas que o autor acreditava serem essenciais para uma compreensível análise do mercado aéreo. O autor acredita que o sistema desenvolvido tem características gerais que seriam interessantes para uma análise conduzida por um especialista na área. Entretanto, uma futura interação com um especialista seria de vital importância para elicitar tarefas analíticas e gerar visualizações para suportá-las.

Uma limitação importante do AirFareVis é quanto a quantidade de companhias aéreas e passagens que podem ser visualizadas ao mesmo tempo no gráfico de barras de duração das promoções (Capítulo 4). As legendas correspondentes às companhias e as barras que correspondem às passagens aéreas exaurir o espaço visual na tela. Um desenvolvimento futuro é desenvolver melhores estratégias de agrupamento visual para evitar esse tipo de problema.

Além de outras visualizações, outro ponto de melhoria seria na construção da aplicação, nenhum *framework* de desenvolvimento foi utilizado como arcabouço para construção da aplicação, esta foi seguida sem padrão de projeto, o que a deixou um pouco frouxa e desordenada.

Uma modularização tornaria a adição de outras visualizações uma tarefa fácil, pois não seria necessária uma mudança no código toda vez que um teste fosse realizado, isso ocorreu constantemente uma vez que à medida que o projeto evoluía, várias refatorações foram necessárias para adaptação do código existente com as novas alterações.

Referências

- BOSTOCK, M.; OGIEVETSKY, V.; HEER, J. D3: data-driven documents. **IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)**, [S.l.], 2011.
- BOYANDIN, I. et al. Flowstrates: an approach for visual exploration of temporal origin-destination data. In: COMPUTER GRAPHICS FORUM. **Anais...** [S.l.: s.n.], 2011. v.30, n.3, p.971–980.
- FERREIRA, N. **Visual analytics techniques for exploration of spatiotemporal data**. 2015. Tese (Doutorado em Ciência da Computação) — POLYTECHNIC INSTITUTE OF NEW YORK UNIVERSITY.
- FERREIRA, N. et al. Visual exploration of big spatio-temporal urban data: a study of new york city taxi trips. **IEEE Transactions on Visualization and Computer Graphics**, [S.l.], v.19, n.12, p.2149–2158, 2013.
- FRIENDLY, M. A Brief History of Data Visualization. In: CHEN, C.; HÄRDLE, W.; UNWIN, A. (Ed.). **Handbook of Computational Statistics: data visualization**. Heidelberg: Springer-Verlag, 2006. v.III, p.15–56. (In press).
- GANTT, H. L. **Organizing for work**. [S.l.]: Harcourt, Brace and Howe, 1919.
- (IATA), I. A. T. A. **Fact Sheet Industry Statistics**. 2016.
- MIOTTO, G. R.; SOUZA, M. A. de; DIEHL, C. A. Reflexos das Mudanças na Concorrência do Setor de Aviação Comercial Brasileiro: um estudo no âmbito da desregulamentação governamental e da entrada da companhia gol. **ABCustos**, [S.l.], v.3, n.2, 2008.
- MUNZNER, T. **Visualization Analysis and Design**. [S.l.]: CRC Press, 2014.
- OLIVEIRA, G. d. N. **Ordered Stacks of Time Series for Exploratory Analysis of Large Spatio-Temporal Datasets**. 2015. Tese (Doutorado em Ciência da Computação) — UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL.
- PHAN, D. et al. Flow map layout. In: IEEE SYMPOSIUM ON INFORMATION VISUALIZATION, 2005. INFOVIS 2005. **Anais...** [S.l.: s.n.], 2005. p.219–224.
- ROBERTS, J. C. State of the Art: coordinated multiple views in exploratory visualization. In: FIFTH INT. CONF. COORDINATED AND MULTIPLE VIEWS IN EXPLORATORY VISUALIZATION (CMV 2007). **Proceedings...** [S.l.: s.n.], 2007. p.61–71.
- WILLETT, W.; HEER, J.; AGRAWALA, M. Scented Widgets: improving navigation cues with embedded visualizations. **IEEE Transactions on Visualization and Computer Graphics**, Piscataway, NJ, USA, v.13, n.6, p.1129–1136, Nov. 2007.
- WILLIAMSON, D. F.; PARKER, R. A.; KENDRICK, J. S. The box plot: a simple visual method to interpret data. **Annals of internal medicine**, [S.l.], v.110, n.11, p.916–921, 1989.
- WOOD, J.; DYKES, J.; SLINGSBY, A. Visualisation of origins, destinations and flows with OD maps. **The Cartographic Journal**, [S.l.], v.47, n.2, p.117–129, 2010.

YOU, M. et al. A usability evaluation of web map zoom and pan functions. **International Journal of Design**, [S.l.], v.1, n.1, 2007.