

# Visualização de Anúncios Classificados de Automóveis

Giovanni Paolo Santos de Carvalho (gpsc@cin.ufpe.br)  
Centro de Informática, Universidade Federal de Pernambuco

**Abstract—Relatório de projeto da disciplina de Visualização de Dados, edição 2018.1 no curso de Pós-graduação em Ciência da Computação do Centro de Informática @ Universidade Federal de Pernambuco.**

## I. INTRODUÇÃO

Pessoas compram e vendem veículos o tempo todo. Sites de anúncios facilitam a exposição de vendedores a uma gama maior de possíveis compradores em comparação com os outros meios tradicionais, como jornais, devido à facilidade de publicação e número de leitores.

Seja qual for o motivo da venda ou compra, o usuário tem acima de todo o objetivo de maximizar o seu lucro e minimizar os seus gastos. Muitos hábitos dos motoristas e características socio-econômicas pesam diretamente no valor de mercado praticado para um veículo, como o quanto o automóvel roda anualmente ou qual o poder de compra na região onde está sendo anunciado.

O objetivo deste trabalho é proporcionar um ambiente para análise dessas características associadas com os veículos anunciados, bem como estas informações podem ser utilizadas para propiciar que compradores e vendedores logrem seus objetivos. As seções a seguir descrevem os dados, metodologia utilizada, sistema desenvolvido, conclusões extraídas com o auxílio da ferramenta e possíveis direções a serem seguidas em trabalhos futuros.

## II. DADOS

Os dados foram obtidos a partir da extração de informações do site TrueCar, pelo usuário Evan Payne e disponibilizado sob licença CC0: Public Domain no site de competições online Kaggle. Os dados extraídos são de anúncios de 1.2 milhões de automóveis usados nos Estados Unidos. Entre os dados coletados de cada anúncio estão:

- **Preço:** valor do anúncio em \$USD;
- **Ano:** ano de fabricação do veículo;
- **Milhagem:** distância percorrida pelo veículo em milhas;
- **Cidade:** cidade onde está sendo anunciado;
- **Estado:** estado onde está sendo anunciado;
- **VIN:** identificador de veículo utilizado nos EUA;
- **Marca:** fabricante do veículo;
- **Modelo:** modelo do veículo.

Naturalmente, nem todas essas variáveis são imediatamente interessantes como objeto de estudo nas visualizações, como por exemplo o identificador de veículo, que não possui relação entre dois veículos.

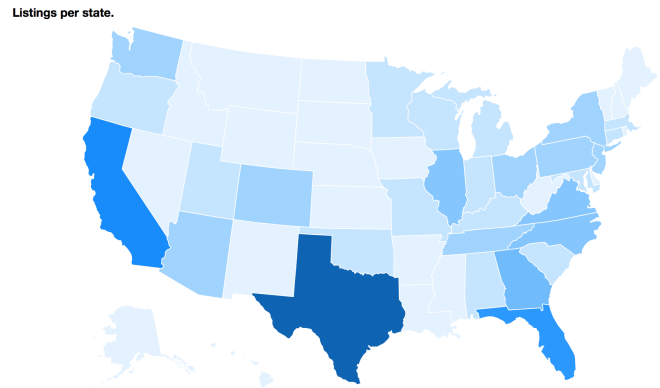


Fig. 1. Anúncios por estado.

## III. METODOLOGIA

A metodologia de estudo desenvolvida neste trabalho foi focada na análise exploratória dos dados através da construção de múltiplas visualizações que relacionassem itens de interesse entre os automóveis anunciados, agrupadas por localização e exibindo as distribuições das características numéricas de cada anúncio.

As tecnologias utilizadas foram as bibliotecas JavaScript D3.js<sup>1</sup>, Crossfilter<sup>2</sup> e dc.js<sup>3</sup>.

## IV. SISTEMA

O sistema final desenvolvido conta com 4 visualizações interativas sobre o conjunto de dados. Cada uma delas arranja os dados de uma maneira específica e os filtros aplicados em uma são reproduzidos nas outras (ex: filtragem por ano atualiza o histograma de preços).

As 4 visualizações são descritas nas sub-seções a seguir.

### A. Mapa de anúncios por estado.

Esta visualização mostra um gráfico com os estados e com a cor representando o número de anúncios realizados na localização. Quanto mais escuro, maior o número de localizações.

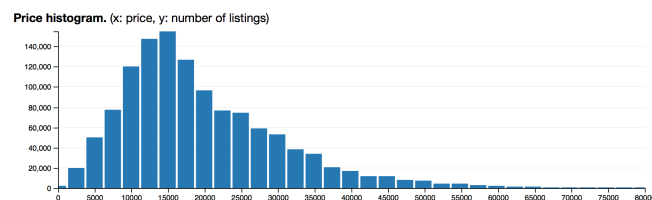
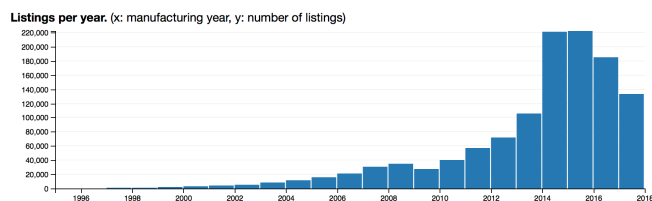
### B. Histograma de anúncios por ano de fabricação.

Este gráfico mostra o histograma de anos de fabricação dos veículos anunciados. Clicar em algum estado na visualização descrita na seção anterior filtra e atualiza o histograma incluindo apenas os estados selecionados.

<sup>1</sup><http://d3js.org>

<sup>2</sup><http://square.github.io/crossfilter/>

<sup>3</sup><https://dc-js.github.io/dc.js/>



### C. Histograma de preços dos anúncios.

Este gráfico mostra a distribuição de preços dos automóveis anunciados. Similar ao gráfico anterior, é possível filtrar por estados e o histograma é atualizado para incluir apenas os estados que foram selecionados.

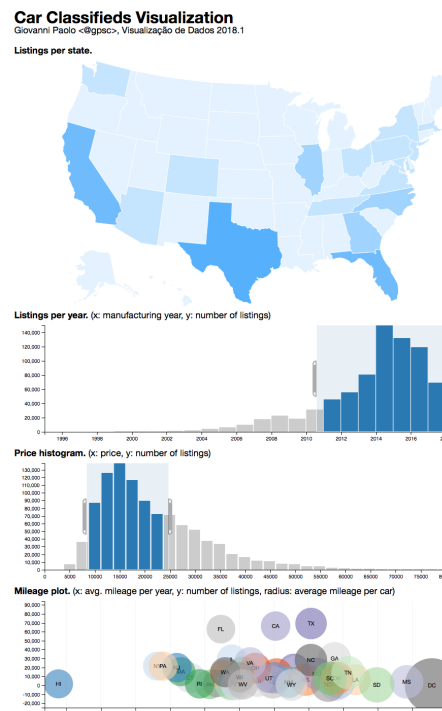
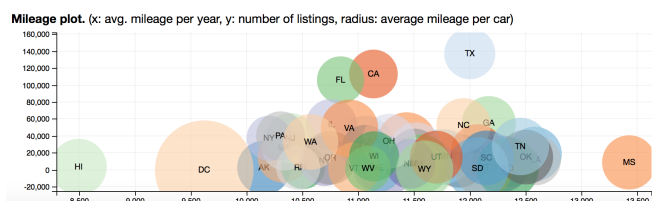
D. Gráfico de milhagem dos veículos anunciados (por estado).

Este gráfico permite analisar as características de direção dos motoristas que anunciam veículos em cada estado. O eixo x representa a milhagem anual média dos veículos anunciados, o raio a milhagem média dos veículos e o eixo y o número de automóveis anunciados para o estado.

## V. CONCLUSÃO

Neste trabalho foi desenvolvido um sistema que permite visualizar e interagir com um grande volume de dados de anúncios classificados de automóveis. As visualizações evidenciam algumas características específicas de cada região dos Estados Unidos firmemente relacionadas com os hábitos de direção dos motoristas e do panorama socioeconômico.

Por exemplo, no Texas, possivelmente devido às dimensões do estado, o número de veículos anunciados é muito grande, visto que automóveis são um meio de locomoção essencial. O mesmo pode ser observado para a Califórnia, onde o sistema de transporte público mais deficitário e questões de poder aquisitivo permitem que uma boa parcela da população possua veículos próprios.



Alguns estados como o Havái possuem um número de anúncios de veículos mais antigos muito maior em comparação com o quadro geral dos outros estados. O estado ocupa as últimas posições em termos de renda per capita, que pode estar relacionada com essa observação.

É também possível observar que Mississipi está bem a frente em termos de milhas anuais percorridas em média por veículo, ao mesmo tempo em que possui uma quantidade anúncios de veículos mais novos superior aos outros estados, o que sugere que há algum fator que contribui com a alta rotatividade de veículos e com os padrões de usos (em termos de distância percorrida) acima da média. Enquanto isso, no Havaí - um estado ilhado e de pequenas dimensões, veículos rodam muito menos anualmente e possuem uma rotatividade inferior.

De maneira geral, o trabalho produzido foi satisfatório e permite realizar observações interessantes sobre os dados. Alguns recursos que podem ser adicionados para tornar a ferramenta mais completa em possíveis trabalhos futuros são o *breakdown* por Marca e Modelo e a visualização das tendências de preço para um modelo específico ao longo dos anos.

O sistema completo, com filtros aplicados pode ser observado na Figura 5, e está disponível online no endereço <https://giovannipcarvalho.github.io/carclassifieds-visualizacao-2018-1/>.