



WBA0750_v1.0

Data Discovery, Olap e Visualização de Dados



A organização visual (visualização de dados e *Big Data analytics*)

Bloco 1

Washington H. C. Almeida





Conceitos fundamentais de *Big Data*

- Não há um consenso quando se trata do conceito básico de *Big Data*.
- O McKinsey Global Institute diz que o termo Big Data refere-se a um conjunto de dados cujo crescimento é exponencial.
- Big Data pode ser entendido como um conjunto de novas soluções para processamento de grande volume de dados, em velocidade a mais variada possível desde processos em batch a processamento quase em tempo real e numa grande variedade de formatos.





Surgimento dos dados

- Em era de *Big Data*, percebeu-se que os dados surgem de todos os cantos:
 - Surgem dos milhares de websites existentes.
 - Dos mais de cem mil tuítes por minuto,
 - De bilhões de compartilhamentos por usuários do Facebook.
 - Dos sensores e câmeras espalhados pela cidades
 - E pelos bilhões de celulares existentes.





Analytics

- Entende-se como um termo amplo que engloba processos e tecnologias.
- O Analytics é o processo de extração e criação de informações a partir de dados brutos por filtragem, processamento, categorização, condensação e contextualização dos dados.
- É o tratamento aplicado no dado para transformá-lo em informação útil.





Usos diversos de *Big Data*

- Por ser considerado um fenômeno, o *Big Data* envolve o uso de diversos tipos de conceitos e tecnologias, como:
- Computação nas nuvens.
- Virtualização.
- Internet.
- Estatística.
- Infraestrutura.
- Armazenamento.
- Processamento, governança e gestão de projetos.





Conclusão

- *Big Data* é o fenômeno em que dados são produzidos em vários formatos e armazenados por uma grande quantidade de dispositivos e equipamentos.
- Pela velocidade que o mundo tem se tornado digital, o conceito de *Big Data* vem aumentando de forma exponencial.
- Com os avanços na tecnologia, as *Big Datas* passam a ser produzidas por fontes variadas como redes sociais, comunidades virtuais, blogs etc.



A organização visual (visualização de dados e *Big Data analytics*)

Bloco 2

Washington H. C. Almeida





A organização visual

- As visualizações podem ser estáticas, dinâmicas ou interativas.
- Visualizações estáticas são utilizadas quando têm-se os resultados da análise armazenados em servidores.
- Se a análise for apresentada com atualizações frequentes, por conta de uma dinâmica de atualização dos dados, deve-se apresentar visualizações dinâmicas com gráficos.
- Se as visualizações tiverem fonte de dados alimentados por usuários diversos, o ideal é apresentá-la de forma interativa.



Intenção das ferramentas de visualização

- O uso de ferramentas de visualização tem a intenção de ajudar a transmitir de maneira fácil e rápida alguma história representada pelos dados tratados.
- Para que o nosso cérebro faça uma leitura do que a imagem está querendo nos dizer, é preciso ter contato com a imagem apropriada para cada tipo de dados e de público-alvo.





Variáveis quantitativas

- Por quantitativa, entende-se o dado que pode assumir qualquer valor numérico dentro de um intervalo de valores.
- Tal conceito está diretamente relacionado com variável quantitativa contínua.
- A diferença para uma variável quantitativa discreta é que esta tem uma restrição ao assumir valores, pois só pode assumir valores inteiros, como aqueles obtidos por contagem ou frequência.



Bubble chart e pictogramas

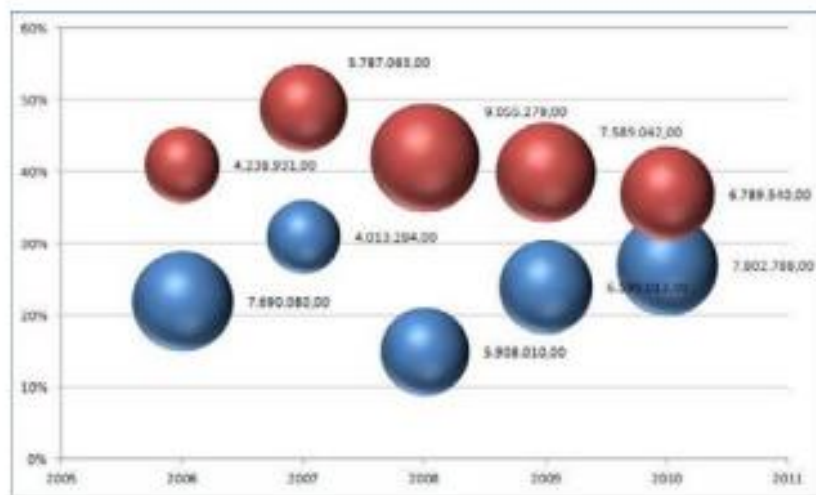
- O gráfico de bolhas está, geralmente, relacionado com variáveis quantitativas contínuas.
- É um tipo de gráfico próprio para representar grandezas com valores de áreas ou volumes.
- Pictogramas também são representações visuais que precisam de atenção na hora de serem utilizados para representação de informações, apesar de serem apropriados para representarem grandezas numéricas, podem confundir seus leitores.



Exemplo de gráfico de bolhas e pictograma

Figura 1 – Exemplo de gráfico de bolhas e pictograma

Gráfico de bolhas



Pictograma

BOTELLAS RECOGIDAS EN UN FIN DE SEMANA



15 Metros
Ciudad A



30 Metros
Ciudad B

Fonte: Lima (2019, p. 36).

Teoria em Prática

Bloco 3

Washington H. C. Almeida





Refleta sobre a seguinte situação

- *Big Data* trata de soluções as mais variadas possíveis, mas, existe os 5 V que nos ajudam a entender melhor o que são essas soluções: velocidade, variedade, volume, veracidade e valor. A velocidade trata do processamento dos dados que podem ser até em tempo real; a variedade versa dos múltiplos tipos de dados tanto estruturados como não estruturados. Já o volume, conforme o nome, representa uma quantidade gigantesca. A veracidade na confiabilidade da informação e, por fim, o valor é aquilo que é derivado do uso de *Big Data*. Dito isto, como você classificaria dentro dos 5 V os sistemas de recomendação que são soluções que usam *Big Data*, muito utilizadas na área de vendas e compras on-line para indicar itens em uma cesta de acordo com o perfil do usuário?



► Norte para a resolução...

- Um sistema de recomendação utiliza de *Big Data* da seguinte maneira:
- V de volume: podemos dizer que processam um volume de dados enorme devido às inúmeras compras e associações realizadas pelos usuários, usando, muita vezes, regras de associação com mineração de dados.
- V de variedade: pode ser usado tanto estruturados, que venham dos sistemas transacionais das instituições, quanto dados não estruturados retirados dos cookies dos usuários nos seus browsers.
- V de velocidade: podemos falar que trabalham em quase tempo real pois, ao clicar em um produto, outro produto é indicado automaticamente, e todas as vendas são analisadas em tempos muito curtos para ver quais produtos mais se associam.
- V de veracidade: se utiliza da base de compras e informações precisas para ter confiabilidade nos dados e indicações realizadas.
- V de valor: correta indicação de produtos e aumento de vendas devido à boa utilização dos dados.



Dica do(a) Professor(a)

Bloco 4

Washington H. C. Almeida



Parâmetros importantes

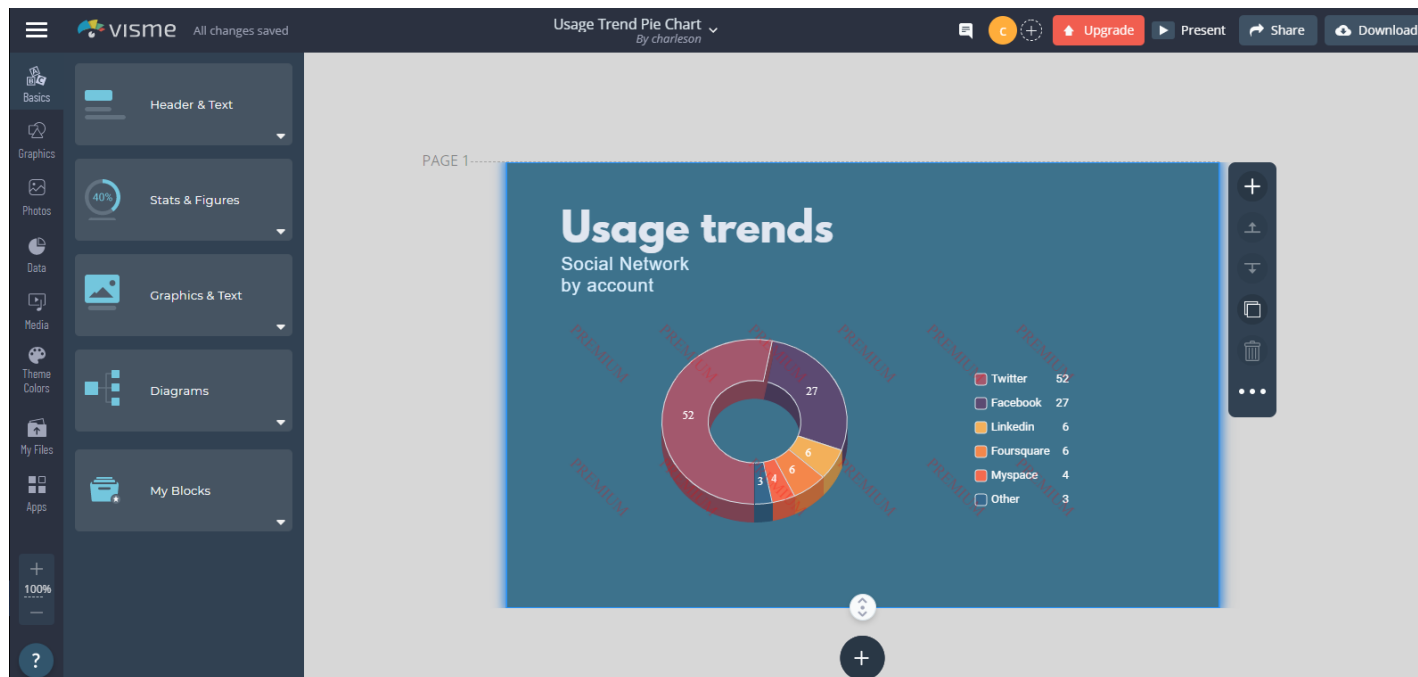
- Pesquisar mais sobre os parâmetros importantes para a elaboração adequada de uma imagem com dados.
- Entender melhor as regras para especificar as cores ideais para ser utilizada em uma imagem.



Ferramentas de criação – Visme

- Visme é uma plataforma de criação e colaboração de conteúdo visual baseada em nuvem, capacitando qualquer pessoa a criar conteúdo profissional de marca, independentemente de seu histórico de design.

Figura 2 – Tela de criação do Visme



Fonte: captura de tela de Visme.



Referências

AMARAL, Fernando. **Introdução a ciência de dados:** mineração de dados e Big Data. Rio de Janeiro: Alta Books, 2016. Kindle.

BAHGA, Arshdeep; MADISETTI, Vijay. **Big Data science & analytics:** a hands-on approach. Arshdeep Bahga & Vijay Madisetti, 2019.

FRANCO, G. Quando usar box plot. **SOS Estatística**, [s.l.], 1 de setembro de 2012. Disponível em: <https://sosestatistica.com.br/quando-usar-box-plots/>. Acesso em: 5 ago. 2019.

GRANT, Robert. **Data visualization:** charts, maps and interactive graphics. Boca Raton: CRC Press, 2019.

LIMA, Marcelo T. de. **Data Discovery, Olap e visualização de dados.** Londrina: Editora e Distribuidora Educacional S.A., 2019,

TAURION, Cezar. **Big Data.** Rio de Janeiro: Brasport, 2013. Disponível em: <https://bv4.digitalpages.com.br/#/legacy/epub/160676>. Acesso em: 9 jul. 2019.



Bons estudos!

