

Coletando, processando e visualizando dados abertos

Usando python para coletar, processar e visualizar dados disponibilizados pelo poder público.

Adolfo Guimarães

@profadolfoguimaraes



Quem sou eu?



🌐 Brasileiro / Nordestino / Sergipano.

📖 Mestrado em Ciência da Computação.

💻 Python.

♥️ Inteligência Artificial.

👊 Acredito que a tecnologia pode e deve ser utilizada em prol da sociedade.

<https://github.com/adolfoguimaraes>

<https://twitter.com/adolfoguimaraes>

<https://instagram.com/profadolfoguimaraes>



Não é de hoje que a humanidade produz dados ...



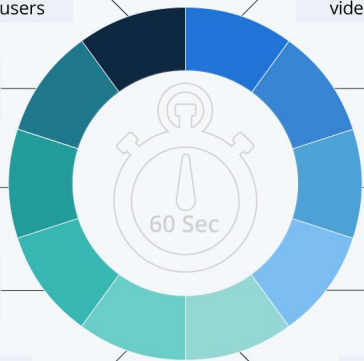
... nos mais variados formatos ...



... e continuamos produzindo MUITO dado.

A Minute on the Internet in 2020

Estimated amount of data created on the internet in one minute



Source: Visual Capitalist



statista

DATA.GOV | DATA TOPICS - RESOURCES STRATEGY DEVELOPERS CONTACT

DATA CATALOG

Search datasets...

Datasets ordered by Popular

Filter by location: Enter location...

193,205 datasets found

Department for the Aging (DFTA) Geriatric Mental Providers

City of New York - Listing of DFTA Geriatric Mental Health Centers of Operation.

Low-altitude aerial imagery (UAS) flights over Black E...

dados.gov.br

PORTAL BRASILEIRO DE DADOS ABERTOS

Dados | Organizações | Aplicativos | Inventários | Concursos | INDA | Perguntas frequentes

Conjuntos de dados

Organizações

- Banco Central do Br... (3305)
- Instituto Nacional ... (1386)
- Agência Nacional de... (637)

Buscar conjunto de dados...

10.709 conjuntos de dados encontrado(s)

Ordenar por:

Brasil.io

Datasets

COVID-19

Sistema Informativo e casos de coronavírus por município por dia

Estado da balneabilidade da costa baiana

Condição de balneabilidade das praias da costa baiana, retirado dos boletins semanais do PREMA.

EU Open Data Portal

Access to European Union open data

EUROPA > EU Open Data Portal > European Union Open Data Portal

Home | Data | Applications | Linked data | Visualisations > | Developers' corner | About

Search datasets...

Show results with:

- all of these words
- any of these words
- the exact phrase

Total datasets available: 15735

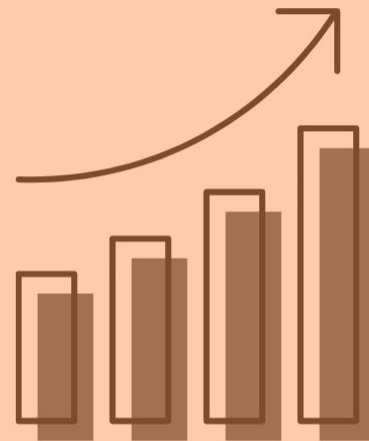
COLETA



PROCESSAMENTO E ANÁLISE

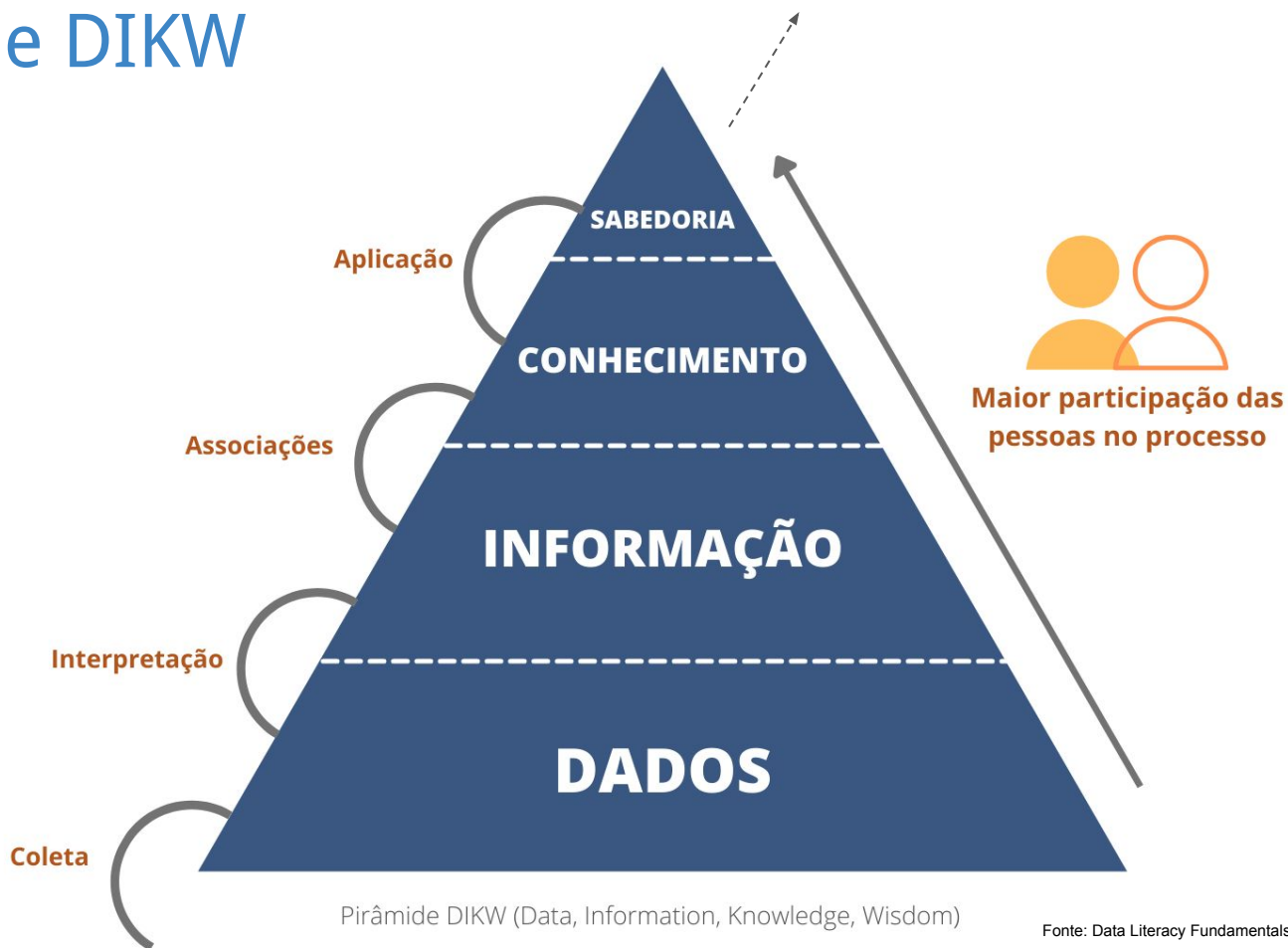


VISUALIZAÇÃO



Pirâmide DIKW

wisdom: the ability to use your knowledge and experience to make good decisions and judgments.

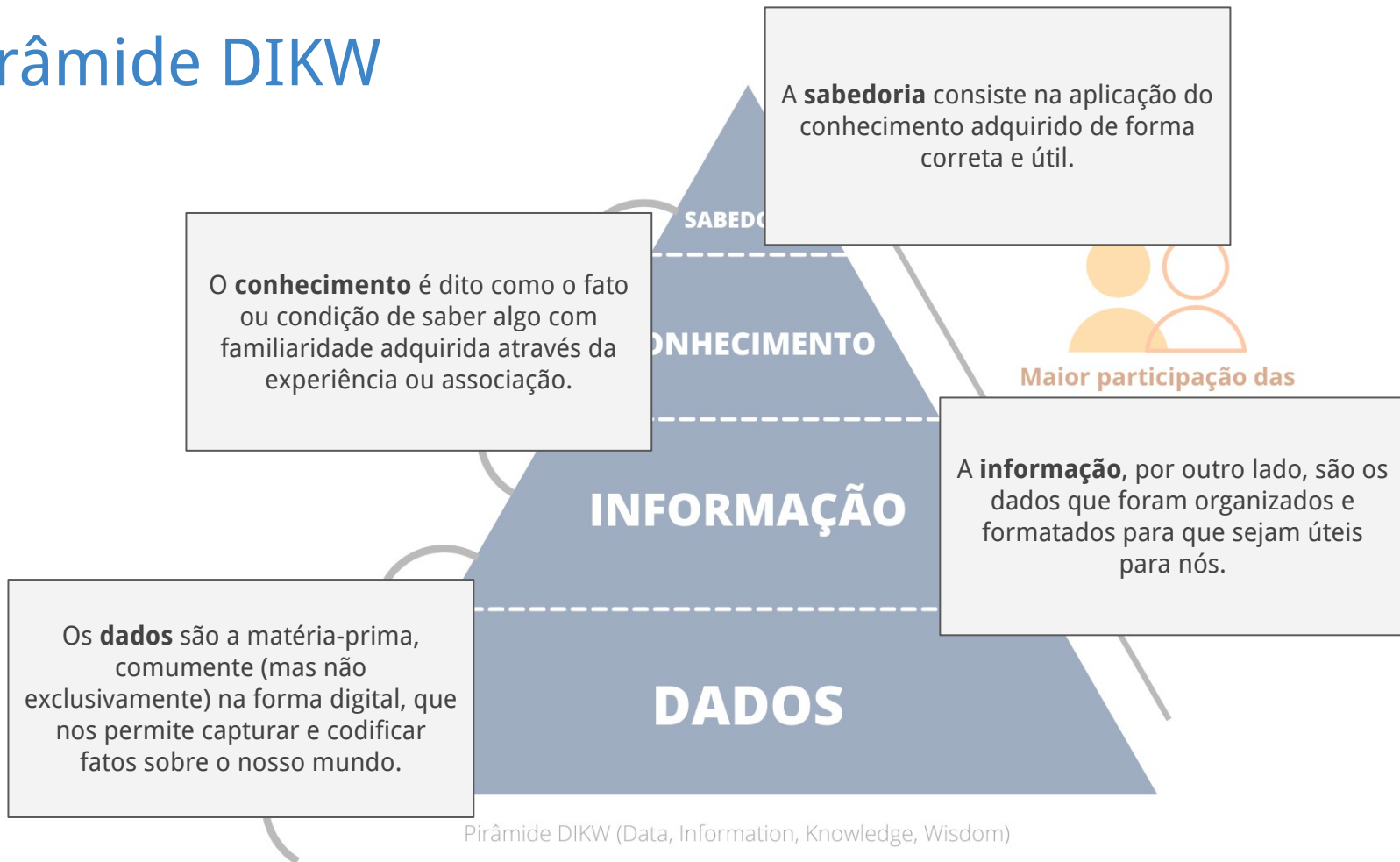


Pirâmide DIKW (Data, Information, Knowledge, Wisdom)

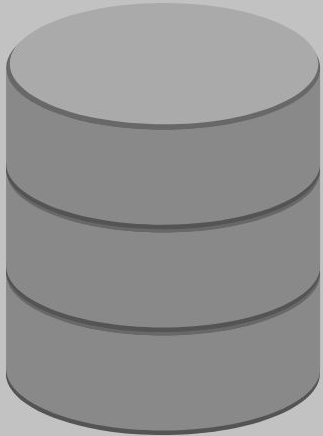
Fonte: Data Literacy Fundamentals: Understanding the Power & Value of Data
Ben Jones

<https://www.amazon.com.br/Data-Literacy-Fundamentals-Understanding-English-ebook>

Pirâmide DIKW



COLETA



Onde conseguir os dados?

Nosso caminho começa com a LAI



A **LAI (Lei de Acesso à Informação)** foi editada em **2011** e regula o direito de **acesso à informação** já previsto na constituição.

A lei permite que **nós cidadãos façamos o controle social** a partir do conhecimento e da participação da administração pública.

Nosso caminho começa com a LAI



A **LAI (Lei de Acesso à Informação)** foi editada em **2011** e regula o direito de **acesso à informação** já previsto na constituição.

A lei permite que **nós cidadãos façamos o controle social** a partir do conhecimento e da participação da administração pública.

A transparência pode ser classificada como **transparência ativa** e **transparência passiva**.

divulgação proativa de informações públicas

divulgação de informações mediante solicitação

Quem está sujeito a lei?

Qualquer instituição que receba dinheiro público deve responder à LAI:

- Executivo, Judiciário, Legislativo da União, estados e municípios;
- Sociedades de economia mista e demais entidades controladas direta e indiretamente pelo governo (incluindo estatais e o Sistema S);
- Entidades sem fins lucrativos beneficiadas por recursos públicos.

Conhecendo um pouco a lei.

Art. 8º: É dever dos órgãos e entidades públicas promover, independentemente de requerimentos, a divulgação em local de fácil acesso, no âmbito de suas competências, de informações de interesse coletivo ou geral por eles produzidas ou custodiadas.

§ 1º Na divulgação das informações a que se refere o caput, deverão constar, no mínimo:

I - registro das competências e estrutura organizacional, endereços e telefones das respectivas unidades e horários de atendimento ao público;

II - registros de quaisquer repasses ou transferências de recursos financeiros;

III - registros das despesas;

IV - informações concernentes a procedimentos licitatórios, inclusive os respectivos editais e resultados, bem como a todos os contratos celebrados;

V - dados gerais para o acompanhamento de programas, ações, projetos e obras de órgãos e entidades; e

VI - respostas a perguntas mais frequentes da sociedade.

Conhecendo um pouco a lei.

§ 3º Os sítios de que trata o § 2º deverão, na forma de regulamento, atender, entre outros, aos seguintes requisitos:

I - conter ferramenta de pesquisa de conteúdo que permita o acesso à informação de forma objetiva, transparente, clara e em linguagem de fácil compreensão;

II - possibilitar a gravação de relatórios em diversos formatos eletrônicos, inclusive abertos e não proprietários, tais como planilhas e texto, de modo a facilitar a análise das informações;

III - possibilitar o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina;

IV - divulgar em detalhes os formatos utilizados para estruturação da informação

V - garantir a autenticidade e a integridade das informações disponíveis para acesso;

VI - manter atualizadas as informações disponíveis para acesso;

VII - indicar local e instruções que permitam ao interessado comunicar-se, por via eletrônica ou telefônica, com o órgão ou entidade detentora do sítio; e

VIII - adotar as medidas necessárias para garantir a acessibilidade de conteúdo para pessoas com deficiência, nos termos do art. 17 da Lei nº 10.098, de 19 de dezembro de 2000, e do art. 9º da Convenção sobre os Direitos das Pessoas com Deficiência, aprovada pelo Decreto Legislativo nº 186, de 9 de julho de 2008.

Atenção!!!

A lei aborda mais tópicos além dos que foram citados aqui e deve ser de conhecimento de todo cidadão, mas principalmente para quem vai trabalhar com esse tipo de dados.

Acesse em: http://www.planalto.gov.br/ccivil_03/ato2011-2014/2011/lei/l12527.htm

O que são dados abertos?

"Por dados abertos governamentais entendem-se os dados públicos, produzidos, coletados ou custodiados por órgãos estatais e disponibilizados na web em formato aberto (não proprietário), estruturado e compreensível logicamente por máquina, de modo que qualquer um possa livremente acessar, reutilizar, modificar e redistribuir para quaisquer finalidades, estando sujeito a, no máximo, exigências de creditar a sua autoria e compartilhar sob a mesma licença"

8 princípios dos dados abertos

Completos

todos os dados públicos devem ser disponibilizados, ao menos que tenham alguma restrição regulamentada por estatuto e publicamente justificada;

Atuais

os dados são atualizados periodicamente e rapidamente para manter o seu valor e utilidade;

Acessíveis

os dados são disponibilizados gratuitamente para alcançar o maior número de usuários e propósitos, sem restrições;

Primários

os dados são pré-processados, sem agregação ou modificação. Salvo exceções relacionadas ao sigilo de informações pessoais ou qualquer outro tipo de sigilo regulamentado por estatuto e publicamente justificado;

Não proprietários

os dados não estão sob o controle de nenhuma entidade e em formatos que necessitem uso de ferramentas exclusivas de uma determinada organização;

Compreensíveis por máquina

os dados são disponibilizados gratuitamente para alcançar o maior número de usuários e propósitos, sem restrições;

Não discriminatórios

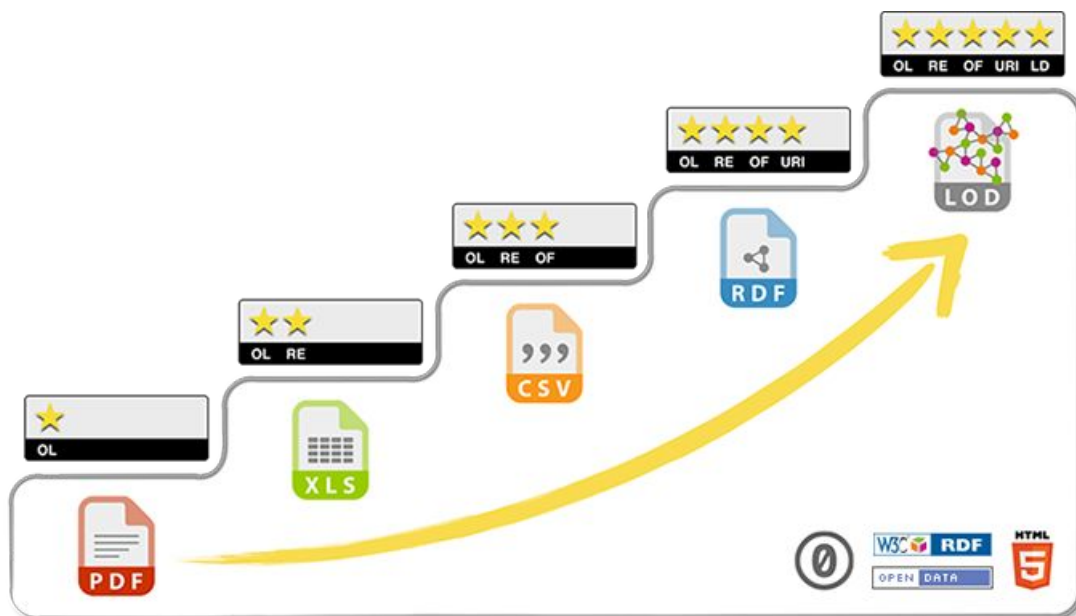
os dados são atualizados periodicamente e rapidamente para manter o seu valor e utilidade;

Livres de licença

os dados não estão sujeitos a regulações de direitos autorais, marcas, patentes ou segredo industrial. Algumas restrições relacionadas à privacidade, segurança, controle de acesso e outros privilégios são permitidos desde que sejam transparentes, justificadas e regulamentadas por estatuto.

As 5 estrelas dos dados abertos

Tim Berners-Lee, o inventor da Web e o primeiro a pensar nos Dados Ligados, sugeriu um esquema de implementação das 5 estrelas para Dados Abertos. Aqui, damos exemplos para cada etapa das estrelas e explicamos custos e benefícios que vêm junto com eles.



Onde buscar os dados?

<http://www.portaltransparencia.gov.br/>

The screenshot shows the main page of the Portal da Transparência. At the top, there is a navigation bar with the logo and the text 'Portal da Transparência' and 'CONTROLE GERAL DA UNIÃO'. Below this is a search bar with the text 'Todos' and a search icon. A large green banner for 'PAINEL CORONAVÍRUS' is prominent, with the text 'Dados agregados sobre despesas e transferências do Governo Federal destinadas ao combate da pandemia'. Below the banner is a grid of 12 categories, each with a thumbnail and a 'NOVO' tag: 'AUXÍLIO EMERGENCIAL', 'BENEFÍCIOS AO CIDADÃO', 'CARTÕES DE PAGAMENTO', 'CONVÊNIOS', 'DESPESAS E RECEITAS', 'EMENDA PARLAMENTAR', 'ESTADOS E MUNICÍPIOS', 'LICITAÇÕES E CONTRATOS', 'ORÇAMENTO', 'RECURSOS TRANSFERIDOS', 'SERVIDORES E PENSIONISTAS', and 'VIAGENS A SERVIÇO'. On the left side, there are sections for 'PANORAMA', 'Consulte os documentos diários de despesa', and 'Sanções'. At the bottom, there are two sections: 'Destques - Rede de Transparência' and 'Destques - Combate ao Coronavírus'.

<http://dados.gov.br/>

The screenshot shows the dados.gov.br website. At the top, there is a navigation bar with the logo and the text 'Portal Brasileiro de Dados Abertos'. Below this is a search bar with the text 'Pesquisar conjuntos de dados por nome, órgão, área de governo...' and a search icon. A large green banner for 'CORONAVÍRUS (COVID-19)' is prominent, with the text 'Dados agregados sobre despesas e transferências do Governo Federal destinadas ao combate da pandemia'. Below the banner is a grid of 12 categories, each with a thumbnail and a 'NOVO' tag: 'AUXÍLIO EMERGENCIAL', 'BENEFÍCIOS AO CIDADÃO', 'CARTÕES DE PAGAMENTO', 'CONVÊNIOS', 'DESPESAS E RECEITAS', 'EMENDA PARLAMENTAR', 'ESTADOS E MUNICÍPIOS', 'LICITAÇÕES E CONTRATOS', 'ORÇAMENTO', 'RECURSOS TRANSFERIDOS', 'SERVIDORES E PENSIONISTAS', and 'VIAGENS A SERVIÇO'. On the left side, there are sections for 'PANORAMA', 'Consulte os documentos diários de despesa', and 'Sanções'. At the bottom, there are two sections: 'Destques - Rede de Transparência' and 'Destques - Combate ao Coronavírus'.

Portal da Transparência de estados e municípios

The screenshot shows the Portal da Transparência de estados e municípios website. At the top, there is a navigation bar with the logo and the text 'Portais da Transparência - Município de Aracaju'. Below this is a large purple banner for 'TRANSPARÊNCIA COVID-19' with the text 'Informações sobre o combate ao novo coronavírus' and a logo for 'Aracaju contra o Coronavírus'. Below the banner is a section for 'Portal da Transparência' with the text 'O que é o portal da transparência' and 'Acesso às Informações' with a mail icon. Below this is a section for 'Acompanhe aqui todas as informações para a maior transparência de nosso município.' with a grid of 6 boxes, each containing a logo and the name of a municipality: 'ARACAJU Prefeitura Municipal de Aracaju', 'ARACAJU Aracaju Previdência', 'FUNCAJU FUNCAJU', 'EMURB EMURB', 'ARACAJU SMTT', and 'EMSURB EMSURB'.

Onde buscar os dados?

<https://brasil.io/>

The screenshot shows the homepage of Brasil.io. At the top, there is a navigation bar with the logo and links for 'Sua conta', 'Dados', 'Sobre', 'Blog', and 'APOIE'. The main heading is 'O Brasil em dados libertos', followed by the subtitle 'Repositório de dados públicos disponibilizados em formato acessível'. Below this, there are sections for 'Últimas atualizações' and 'ESPECIAL COVID-19: DADOS POR MUNICÍPIO'. A news item dated 24/10/2021 is visible. At the bottom, there are four colored boxes: 'Manifesto' (green), 'Datasets' (yellow), 'Sugira um Dataset' (blue), and 'Colabore' (white).

<https://basedosdados.org/>

The screenshot shows the homepage of Basedosdados.org. The navigation bar includes 'Dados', 'Documentação', 'Contato', 'Sobre', 'Newsletter', 'APOIE', and 'Entrar'. The main heading is 'Um único lugar para buscar e acessar os dados que você precisa'. Below this is a search bar with the text 'Pesquisar palavras-chave, insira' and a search icon. There are three columns of content with icons: a magnifying glass, a database icon, and a code icon. At the bottom, there is a section for 'Novidades no datalake' and a featured article titled 'Relação Anual de Informações Sociais (RAIS)'.

E tantos outros sites de organizações civis e projetos que trabalham com dados abertos. Vale visitar:

- Open Knowledge Brasil:
<https://ok.org.br/>

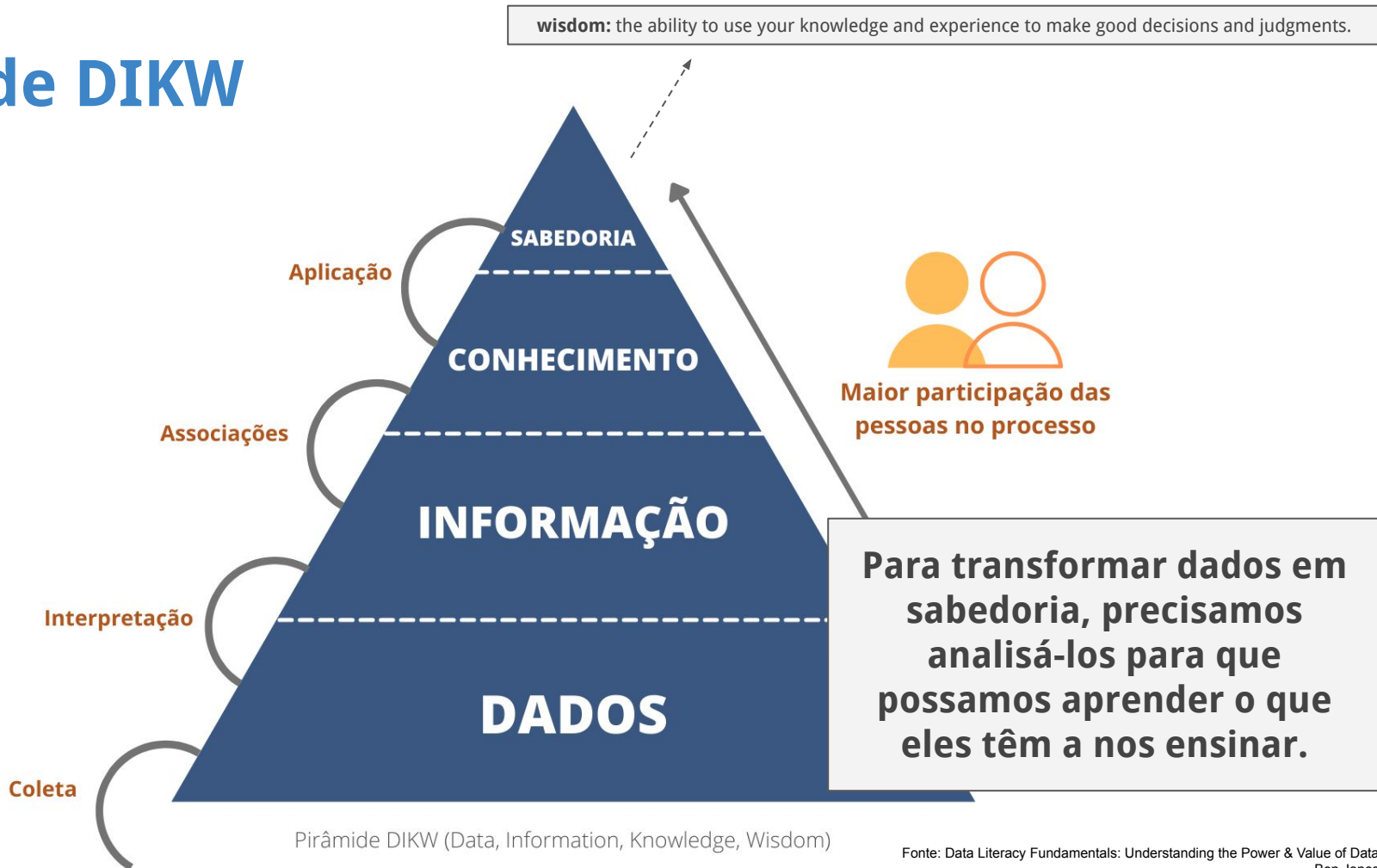
- Escola de Dados:
<https://escoladedados.org/>

- Querido diário:
<https://queridodiario.ok.org.br/>



**Já tenho os dados.
O que fazer com eles?**

Pirâmide DIKW



5 formas diferentes para analisar os dados



Descritiva



Inferencial



Diagnóstica



Preditiva



Prescritiva

5 formas diferentes para analisar os dados

Análise Descritiva

O que aconteceu no passado?

O foco é descobrir o que os dados podem contar sobre Quem, O quê, quando e onde sobre o tópico de interesse.

5 formas diferentes para analisar os dados

Análise Inferencial

E sobre o resto?

Esse tipo de análise aparece quando é impraticável ou impossível coletar dados de cada membro da população analisada.

5 formas diferentes para analisar os dados

Análise Diagnóstica

O que tem sob a superfície?

Esse tipo de análise envolve olhar além das observações iniciais, detalhando as categorias, procurando por anomalias, ou comparando os dados obtidos com outros dados não incluídos no início.

5 formas diferentes para analisar os dados

Análise Preditiva

O que acontece em seguida?

Esse tipo de análise envolve usar dados históricos para previsões sobre resultados futuros. Os dados geralmente revelam tendências e padrões que podem ser utilizados para a construção de modelos que podem "prever o futuro".

5 formas diferentes para analisar os dados

Análise Prescritiva

O que deveríamos fazer?

Esse tipo de análise concentra-se na tarefa de quais ações podemos tomar após as análises prévias apresentadas. Em resumo, permite escolher a melhor alternativa para transformar decisões em decisões baseadas em dados.

Analogia a um paciente visitando um médico



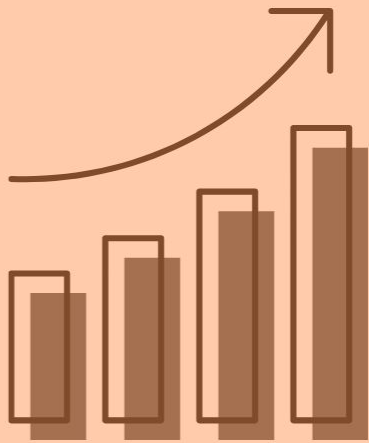
Análise descritiva: o médico começa observando e verificando os sintomas do paciente.

Análise inferencial: O médico pode fazer algumas inferências a outros pacientes ou estudos.

Análise diagnóstica: O médico pode investigar um pouco mais para entender melhor a causa dos sintomas.

Análise preditiva: A partir do que foi observado, o médico pode fazer algumas previsões de como a doença pode evoluir ao longo do tempo se não for feita alguma intervenção.

Análise prescritiva: E finalmente, o médico prescreve medicações ou procedimento para aumentar as chances de recuperação do paciente.



Dados coletados e analisados. Onde entra a visualização nisso tudo?

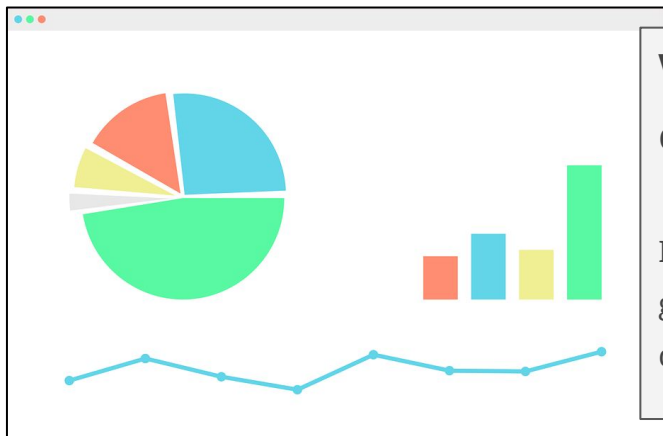


Dados coletados e analisados. Onde entra a visualização nisso tudo?

A visualização é inerente a diversas etapas do processo de análise de dados. Desde do seu uso para entender os dados a visualização das análises e apresentação dos resultados.

O que é visualização?

Consiste no estudo e na criação de representações visuais para os dados com o intuito de **ajudar as pessoas a realização tarefas como mais eficácia.**



Visual Literacy: capacidade de expor e interpretar dados de forma visual.

Isso vai além de construir gráficos. É preciso saber escolher o melhor gráfico para um determinado tipo de dados e entender a mensagem que os dados querem transmitir.

Ainda sobre a definição ...

- A visualização é adequada quando há a **necessidade de aumentar as capacidades humanas** ao invés de substituí-las por métodos computacionais de tomada de decisão;
- A maioria das possibilidades de design para construção das visualizações são ineficazes, portanto **avaliar a eficácia** de um design **é muito importante**, portanto **não é** uma tarefa **simples**;
- Deve-se levar em consideração 3 limitações: os **computadores**, os **humanos** e os **meios de visualização**;
- O uso de uma visualização deve ser analisada em termos de **por que o usuário precisa dela, quais dados vão ser exibidos e como será projetada**.

Ainda sobre a definição ...

"A escolha do melhor design para as **visualizações** de dados, vai muito **além de questões técnicas**.
O **fator humano**, o **público** e os **objetivos** são determinantes na seleção da visualização que será adotada".

Não é só um gráfico ...

Construir uma visualização não é só construir um gráfico. É pensar nos diversos aspectos que influenciam como percebemos os objetos, as cores e os estímulos visuais.

Alguns conceitos importantes

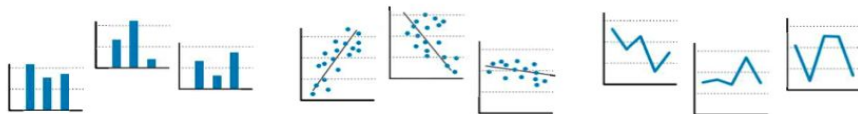
Classificação de Percepção

Permite estimativas precisas

Position along common scales



Position along identical, nonaligned scales



Length



Direction/slope



Angle



Parts of a whole



Area



Volume



Shading and saturation



Color hue



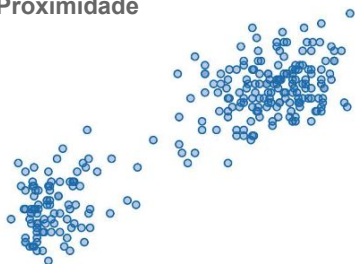
Permite estimativas mais gerais

Alguns conceitos importantes

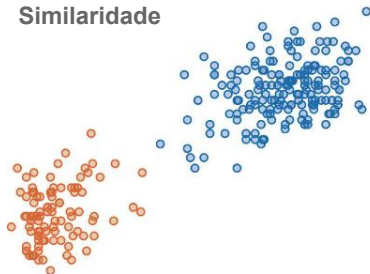
Princípios de Gestalt

A teoria da Gestalt foi desenvolvida no início do século XX por psicólogos alemães e se refere a como tendemos a organizar os elementos visuais em grupos.

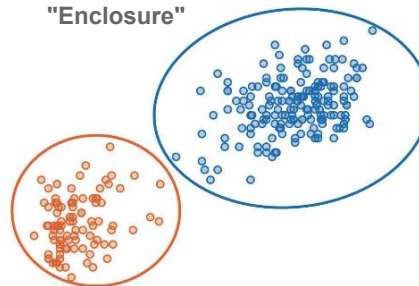
Proximidade



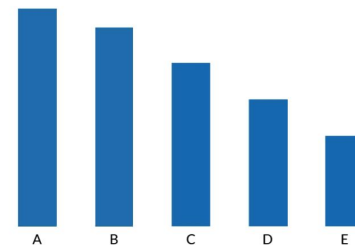
Similaridade



"Enclosure"



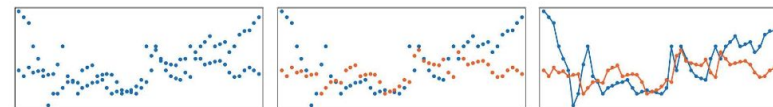
Continuidade



Continuidade

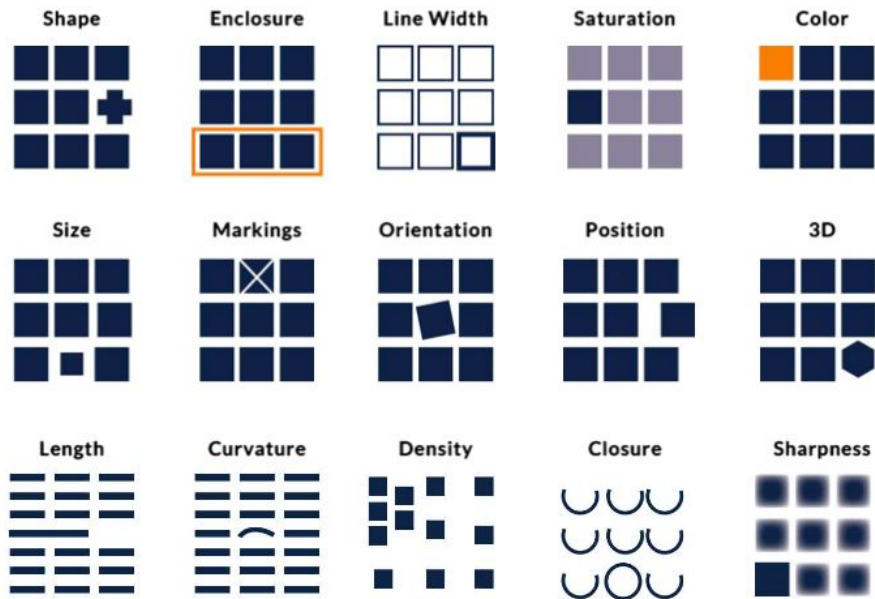


Conexão



Alguns conceitos importantes

Atributos pré-atenciosos



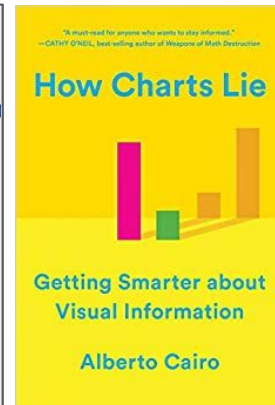
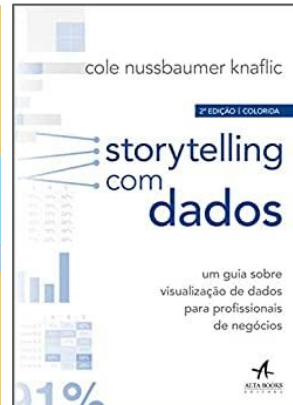
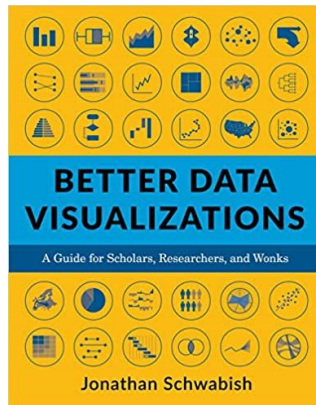
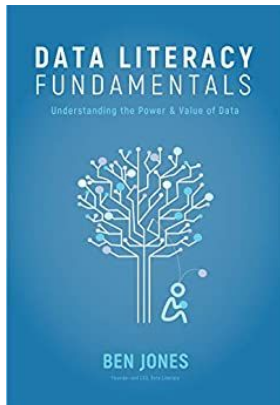
Exemplos de atributos de pré-atenciosos que podemos utilizar em nossas visualizações para direcionar a atenção do leitor

Vamos ver um pouco disso tudo na prática :)

Acesse o repositório:

www.github.com/adolfoguimaraes/minicurso_dadosabertos

Algumas dicas de material ...



ESCOLA DE DADOS
@EscolaDeDados

Ajudamos pessoas e organizações a transformarem dados em histórias, evidências e mudanças sociais. Junte-se a nós pela democratização dos dados no Brasil.

📍 Brasil escoladedados.org/membresia 📅 Ingressou em junho de 2013

DataViz Brasil
@datavizbr Segue você

Publicação sobre visualização de dados em português: medium.com/datavizbr
Curadoria por Rodrigo Medeiros (@medeiros_rodr) #dataviz

📍 Brasil linktr.ee/datavizbr 📅 Nascido(a) em 1 de dezembro
📅 Ingressou em dezembro de 2017

DataViz Lisboa
@DataVizLisboa

Data Visualization Lisboa is both a Meetup and a passion. We see great potential in making data more accessible and usable. #vislis #DataVis #lisboa

Traduzir bio

📍 Lisbon, Portugal youtube.com/c/DataVizLisboa
📅 Ingressou em fevereiro de 2019

PEIXE BABEL
#tecnologia, diversidade e sociedade
📍 🇧🇷 @mialarjanjeira | @avivimota
toda segunda-feira, às 18h

Canal Peixe Babel
@canalpeixebabel Segue você

Transmitindo conhecimento um bit de cada vez! Tecnologia na voz de profissionais diversas 🗣️ Por: @mialarjanjeira e @avivimota, e membro do @svbroficial

📍 Brasil youtube.com/CanalPeixeBabel 📅 Ingressou em setembro de 2018

E para finalizar ...



... isso não é mágica ...

E para finalizar ...



... isso não é
ficção científica ...

E para finalizar ...



... isso é computação!

Coletando, processando e visualizando dados abertos

Usando python para coletar, processar e visualizar dados disponibilizados pelo poder público.

Adolfo Guimarães

@profadolfoguimaraes



Obrigado :)