

Introdução à Análise de Dados

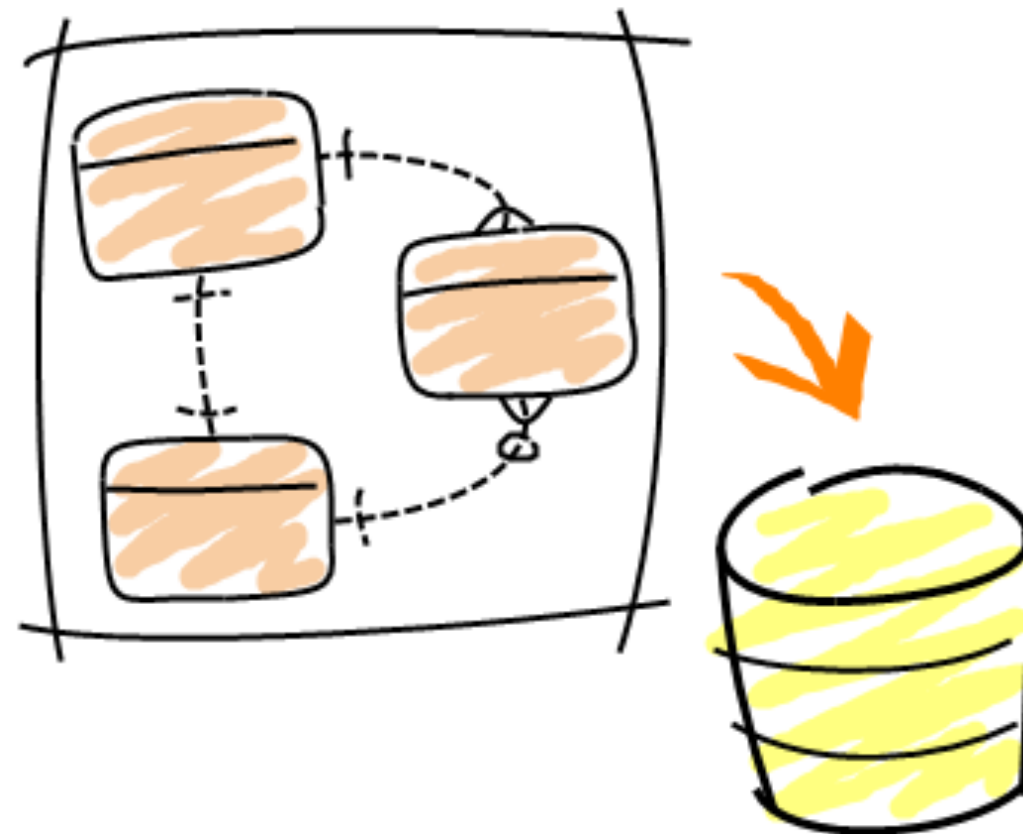
Utilizando o Tableau

Conceitos básicos da Visualização e Análise de Dados

Tipos de Dados

Dados Estruturados

- Relacionados a um sistema de banco de dados
- Mantem a mesma estrutura de representação (rígida), previamente projetada (esquema)
- É possível prever o que será inserido em um campo de uma tabela (texto, número, o tamanho máximo)



Conceitos básicos da Visualização e Análise de Dados

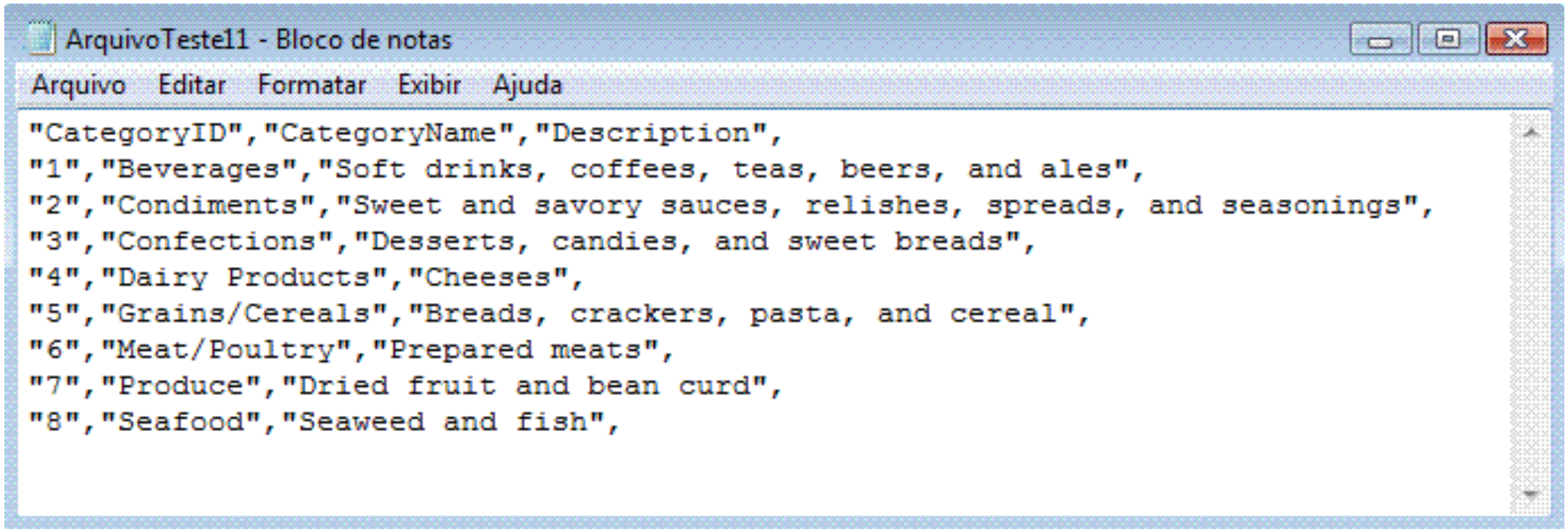
Formatos de Bases de Dados em Arquivo

CSV/TXT

- Comma-separated values
- Compacto e adequado para transferir grandes conjuntos de dados com a mesma estrutura
- Os dados podem ser inúteis sem uma documentação precisa de cada campo
- É essencial que a estrutura do arquivo seja respeitada, já que a simples omissão de um campo pode perturbar a leitura do restante dos dados
- Codificação de caracter, delimitador de colunas e qualificador de texto



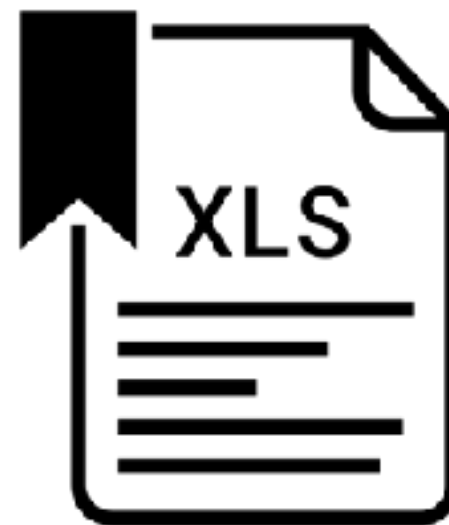
CSV/TXT

A screenshot of a Windows Notepad window titled "ArquivoTeste11 - Bloco de notas". The window has a menu bar with "Arquivo", "Editar", "Formatar", "Exibir", and "Ajuda". The text area contains a CSV file with 8 rows of data. The first row is a header with three columns: "CategoryID", "CategoryName", and "Description". The following rows contain numerical IDs, category names, and their descriptions, all enclosed in double quotes and separated by commas.

```
"CategoryID", "CategoryName", "Description",  
"1", "Beverages", "Soft drinks, coffees, teas, beers, and ales",  
"2", "Condiments", "Sweet and savory sauces, relishes, spreads, and seasonings",  
"3", "Confections", "Desserts, candies, and sweet breads",  
"4", "Dairy Products", "Cheeses",  
"5", "Grains/Cereals", "Breads, crackers, pasta, and cereal",  
"6", "Meat/Poultry", "Prepared meats",  
"7", "Produce", "Dried fruit and bean curd",  
"8", "Seafood", "Seaweed and fish",
```

XLS/XLSX

- Formato de arquivo do Microsoft Excel
- Permite organizar os dados em colunas e linha
- Excel pode realizar cálculos com os dados inseridos nas células
- Permite a criação de tabelas e gráficos a partir de dados armazenados em arquivos XLS
- Várias bases são disponibilizadas em XLS
- O Excel ainda é muito utilizado para realizar análise de dados



XLS/XLSLX

The screenshot shows a Microsoft Excel spreadsheet titled '1.xls [Compatibility Mode] Microsoft Excel'. The spreadsheet contains the following data:

ROLL	NAME	SSC_PEF	HSC_PEF	F3_TOT	ENG	BAK	3_KV	B_ENG	MO	G_TOTAL	MERIT
00002	SALMAKASHAN SAKSA	21.55	37.00	15.56	18.50	13.50	21.00	0.00	55.10	111.46	3096
00008	KOLANAKAD M. S ALAM	27.00	42.00	20.00	15.00	13.50	40.50	0.00	65.50	124.60	365
00009	KOLANAKAD SHAHIDUL ISLAM	20.25	37.00	17.28	21.00	15.00	20.56	0.00	60.50	123.88	1993
00010	KOLANAKAD NUR NAZI	25.50	47.50	20.00	15.00	12.00	41.40	0.00	65.30	142.00	266
00020	ABDUL BAKIR	19.14	37.00	15.14	9.00	16.50	22.00	0.00	58.50	114.64	3490
00021	KARIM CHANDRA DAS	24.00	35.00	15.00	12.50	12.56	22.16	0.00	57.50	113.60	2433
00028	TRIPATI BOHMAN	24.00	44.00	15.00	22.50	17.40	22.70	0.00	72.50	143.60	32
00029	VEHER NIGAR	19.55	34.00	15.86	19.50	12.00	28.50	0.00	60.00	113.86	3315

XML

- Linguagem de marcação (possui marcadores e dados)
- Facilidade de compartilhamento de informações por intermédio da internet
- Separação do conteúdo da formatação
- Simplicidade e legibilidade, tanto para humanos quanto para computadores
- Possibilidade de criação de tags sem limitação
- Interligação de bancos de dados distintos



XML

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<receita nome="pão" tempo_de_preparo="5 minutos" tempo_de_cozimento="1 hora">
  <titulo>Pão simples</titulo>
  <ingredientes>
    <ingrediente quantidade="3" unidade="xícaras">Farinha de Trigo</ingrediente>
    <ingrediente quantidade="7" unidade="gramas">Fermento</ingrediente>
    <ingrediente quantidade="1.5" unidade="xícaras" estado="morna">Água</ingrediente>
    <ingrediente quantidade="1" unidade="colheres de chá">Sal</ingrediente>
  </ingredientes>
  <instrucoes>
    <passo>Misture todos os ingredientes, e dissolva bem.</passo>
    <passo>Cubra com um pano e deixe por uma hora em um local morno.</passo>
    <passo>Misture novamente, coloque numa bandeja e asse num forno.</passo>
  </instrucoes>
</receita>
```

Arquivos PDF

- Portable Document Format
- Criado pela empresa Adobe Systems
- Capacidade de manter a qualidade do arquivo original
- A fonte de dados são as tabelas contidas no documento



Conceitos básicos da Visualização e Análise de Dados

Bases de Dados Públicas

Portal Brasileiro de Dados Abertos

- Lei de Acesso a Informação Pública (Lei 12.527/2011)
- Contém dados e informações públicas sem restrição de acesso
- Bases em CSV, JSON, XML e outros

The screenshot displays the homepage of the Brazilian Open Data Portal. At the top, there is a navigation bar with links for 'Serviços', 'Simplifique!', 'Participe', 'Acesso à informação', 'Legislação', and 'Canais'. Below this is a green header with the 'dados.gov.br' logo and the text 'PORTAL Brasileiro de Dados Abertos'. A search bar is located in the center, with the text 'Pesquisar conjuntos de dados...' and a 'PESQUISAR' button. Below the search bar, it indicates 'em 5412 conjuntos de dados com 28129 recursos (o que é isto?)'. The main content area is divided into two columns. The left column features 'Conjuntos de dados em destaque' with a list of featured datasets, including 'Serviços Públicos Federais para o Cidadão' and 'Compras públicas do governo federal'. The right column displays 'Publicações mais recentes' with a table of recent publications.

Conjunto de dados	Data	Responsável
Formulários (desclassificação a recurso referente a pedido de desclassificação)	12/05/2018	Universidade Federal de Roraima - UF...
Informações sobre o Serviço de Informação ao Cidadão (SIC)	12/05/2018	Universidade Federal de Roraima - UF...
Relatório de	12/05/2018	Universidade

Portal da Transparência

- Lançado em 2004
- Iniciativa da CGU
- Informações sobre Transferências de Recursos
- Informações sobre Gastos Diretos do Governo Federal
- Informações diárias sobre a execução orçamentária e financeira
- Informações sobre cargo, função e situação funcional dos Servidores e agentes públicos

The screenshot displays the official website of the Portal da Transparência (Transparency Portal) for the Brazilian Government. The page features a green header with the title 'Portal da Transparência GOVERNO FEDERAL' and the Ministry of Transparency and Control Administration. A navigation bar includes links for 'Perguntas frequentes', 'Contato', 'Glossário', 'Links', and 'Manual de navegação'. The main content area is divided into several sections: 'CONSULTAS' (with sub-sections like Despesas, Receitas, Convênios, etc.), 'GRÁFICOS E DOWNLOADS', and 'INFORMAÇÕES'. The central part of the page shows data for 'Despesas - Empenho, liquidação e pagamento' and 'Despesas - Pagamento', with a search filter for 'Transferências de Recursos' set to 'Brasil'. On the right side, there is a sidebar with various services and links, including 'Painel de Municípios', 'Reciba Informações de Liberação de Convênios', 'Banco de Projetos', 'Jogos Rio 2016', 'Rede de Transparência', 'Portal de Acesso à Informação', and 'Páginas de Transparência Pública'.

Instituto Brasileiro de Geografia e Estatística

- Sistema Cartográfico Brasileiro
- Censo Demográfico
- Dados Econômicos
- Bases de apoio para algumas análises (municípios, população)

The screenshot shows the IBGE website homepage. At the top, there is a navigation bar with links for 'BRASIL', 'Serviços', 'Participar', 'Acesso à Informação', 'Legislação', and 'Canais'. Below this is the IBGE logo and a main navigation menu with links for 'INSTITUCIONAL', 'RESPONDENTES', 'LOJA', 'ATENDIMENTO', and 'PORTAL IBGE ANTIGO'. A search bar is located below the navigation menu. The main content area features a horizontal menu with links for 'Estatísticas', 'Geociências', 'Biblioteca', 'Agência de Notícias', 'Calendário', and 'Nossos sites'. Below this, there are social media icons for Facebook, YouTube, Instagram, and Twitter. The main content area is divided into three columns. The left column features the 'IBGE educa' logo and the text 'Portal do IBGE voltado para a educação'. The middle column contains a news article titled 'Vendas no varejo variam 0,3% em março' with a sub-headline 'Em março de 2018, o volume de vendas do comércio varejista nacional variou 0,3% frente a fevereiro, na série com ajuste sazonal, após recuar 0,2% em...'. The right column contains a section titled 'Estimativa da População' with the value '208.998.033' and a table of price indices.

Índice	Variação	Período
Preços IFCA - mensal	0,22%	abr 2018
Preços INPC	0,27%	abr 2018
Preços IPCA-E	0,29%	abr 2018

RAIS

- Relação Anual das Informações Sociais
- Empregos em 31 de dezembro e admitidos e desligados segundo gênero, faixa etária, grau de escolaridade, tempo de serviço e rendimentos, desagregados em nível ocupacional, geográfico e setorial

The screenshot displays the website for the Programa de Disseminação das Estatísticas do Trabalho (PDET). The header includes the logo of the Ministério do Trabalho and the text 'PDET PROGRAMA DE DISSEMINAÇÃO DAS ESTATÍSTICAS DO TRABALHO'. A search bar is located in the top right corner. The main content area features a navigation menu on the left with options like 'CAGED', 'RAIS', 'O PDET', and 'CONSULTAS'. The central part of the page is titled 'O que é RAIS?' and includes a sub-header 'RAIS - RELAÇÃO ANUAL DE INFORMAÇÕES SOCIAIS' and a section for 'Objetivos'. There are also social media sharing buttons for Google+, Twitter, and Facebook.

CAGED

- Cadastro Geral de Empregados e Desempregados
- Relevante instrumento à reciclagem profissional e à recolocação do trabalhador no mercado de trabalho
- Possibilita o cálculo do índice de emprego, taxa de rotatividade e a flutuação de emprego, desagregados em nível geográfico, setorial e ocupacional

The screenshot shows the website for the PDET (Programa de Disseminação das Estatísticas do Trabalho) portal. The header includes the text 'MINISTÉRIO DO TRABALHO' and 'PDET PROGRAMA DE DISSEMINAÇÃO DAS ESTATÍSTICAS DO TRABALHO'. A search bar is present with the text 'Buscar no portal'. The main content area features a large teal banner with the text 'CAGED' in large white letters, followed by 'DADOS DE MARÇO/2018'. To the left of the banner is a navigation menu with links for 'CAGED', 'RAIS', 'O PDET', 'O Programa', 'Registros Administrativos', 'O que é RAIS?', 'O que é CAGED?', and 'CONSULTAS'.

Conceitos básicos da Visualização e Análise de Dados

A Ferramenta

Ferramentas Tradicionais

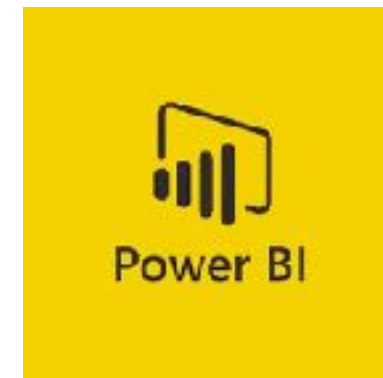


- Elevada dependência da TI
- Algumas não possuem funcionalidades análitas
- Ferramentas tradicionais e a escrita de código podem não ser a melhor maneira de extrair informações de um conjunto de dados complicado



Novas Plataformas

- Hoje, há oferta de novas plataformas para visualização dinâmica de dados
- Possuem funcionalidades analíticas
- Menos dependência da TI



Exemplos

- QlikView
 - <https://www.siop.planejamento.gov.br/siop/>
 - <http://paineldecompras.planejamento.gov.br/>
- Tableau
 - <http://brasiliaemnumeros.codeplan.df.gov.br/>
 - <http://ifceemnumeros.ifce.gov.br>
- PowerBI
 - <http://www.cjf.jus.br/observatorio/>

Por que Tableau?

- Requer estatística básica e um treinamento rápido em laboratório
- Dá conta de bases complicadas com facilidade e possui recursos sofisticados
- Apesar da sofisticação dos recursos, baseia-se na simplicidade da visualização de dados (análise visual e intuitiva)



Hands-On

Obtenção de bases de dados públicas

brunopontes.com.br/cursos