

CURSO SOBRE MANUSEIO DE BASES DE DADOS DO GOVERNO FEDERAL NO SOFTWARE “R”

Competências desejáveis

Professores:

Marcelo Prudente

Rafael Giacomini

Competências desejáveis

- Habilidade com ferramentas de análise de dados
- Conhecimento da política, objeto de análise
- Conhecimento da base de dados que vai trabalhar e de onde encontrar dados e informações complementares

Habilidade com ferramentas de análise de dados

The logo for GOV DATA, featuring the word "GOV" in a light blue sans-serif font and "DATA" in a darker blue sans-serif font, with a small green and blue icon between them.

Hue (Hadoop User Experience) é uma interface web, de código aberto, que suporta o Apache Hadoop e seu ecossistema, licenciado sob a licença Apache v2.



O **RStudio** é uma IDE, de código aberto, para a linguagem de programação de gráficos e cálculos estatísticos R.



Conhecimento da política, objeto de análise

- Legislação que regulamenta a política
 - Leis
 - Decretos
 - Portarias
 - Instruções normativas
- Publicações
- Artigos
- **Experiência na área (consultar quem trabalha na área)**



Conhecimento da base de dados que vai trabalhar

- Importante ler/estudar toda a documentação, notas técnicas e dicionários das bases de dados, quando disponível

Exemplo:

PNAD 2015 - Atualizado em 17/05/2017 às 10:00 horas, incluindo as pesquisas Básica, Suplementar de Acesso à Internet e à Televisão e Posse de Telefone Móvel Celular Para Uso Pessoal da PNAD 2015, Aspectos dos Cuidados das Crianças de Menos de 4 Anos de Idade 2015, Aspectos das Relações de Trabalho e Sindicalização 2015 - PNAD e Prática de Esporte e Atividade Física 2015.

- Dados
- Dicionários e input
- Leia-me PNAD 2015
- Leitura em R
- Marca de imputação
- Metodologia
- Questionário

- Quando a documentação não estiver disponível, buscar pelo questionário da pesquisa ou formulário que abastece o cadastro ou registro administrativo

Onde encontrar dados e informações complementares?

➤ Fontes

• Tabelas prontas

Agregados (dados obtidos por estatísticas sobre os microdados)

Compõem o plano de divulgação dos resultados de uma pesquisa, observando:

✓ Critérios de agregação:

- Apesar dos dados serem coletados com grande precisão espacial, os dados são agregados em áreas mínimas que não permitam a revelação do informante;
- São observadas as classificações utilizadas na pesquisa para que categorias de baixa frequência não levem a revelação do informante.

✓ Precisão:

- A expansão de pesquisas amostrais é feita no mínimo sobre as unidades espaciais consideradas no projeto da amostra para que os valores resultantes apresentem estimativas consistentes.



Onde encontrar dados e informações complementares?

➤ Fontes

- Tabelas prontas

Planos Tabulares – Conjunto de tabelas com informações agregadas, apresentados sob a forma de planilhas

- ✓ São utilizados para divulgação de resultados de pesquisas e Censos.
- ✓ Apresentam estrutura temática cobrindo os principais pontos de conhecimento obtido numa investigação.
- ✓ Apesar de abrangentes, os planos tabulares não esgotam o potencial de informação de uma investigação. Em uma investigação, apenas uma parte dos possíveis cruzamentos de dados são apresentados.
- ✓ A possibilidade de explorar dados não cobertos pelos planos tabulares leva os pesquisadores a demandar o acesso aos microdados.

Exemplo: [PNAD Contínua de educação](#)



Onde encontrar dados e informações complementares?

➤ Fontes

- Ferramentas online

- Business Intelligence – BI

- Ferramenta de acesso pela internet que permite realizar algumas seleções de variáveis, realizar filtros e gerar lista e tabelas que podem ser baixadas no computador.



Onde encontrar dados e informações complementares?

➤ Fontes

- **Microdados**

- Arquivos que contém a menor fração de dados de pesquisas, cadastros ou registros administrativos. A partir da agregação de microdados é construída a informação.
- São dados característicos de unidades de uma população, como dados de indivíduos, de domicílios, de estabelecimentos etc., coletados em procedimentos administrativos, operacionais ou através de pesquisas estatísticas.
- As bases de microdados estão organizadas de forma a serem compreendidas por softwares específicos, o que agiliza o processo de tratamento e cálculos estatísticos.



Onde encontrar dados e informações complementares?

➤ **Vantagens** do microdado em relação ao BI

- ✓ Geralmente estão disponíveis **mais variáveis** que no BI
- ✓ **Sem limitações** quanto a combinações de variáveis
- ✓ Pode ser **cruzado** com outro microdado
- ✓ **Atualização** mais rápida

Onde encontrar dados e informações complementares?

➤ **Desvantagens** do microdado em relação ao BI

- ✓ Em geral, a “qualidade” do microdado é inferior ao do BI, **requer limpeza e transformação**
- ✓ É mais **trabalhoso**, quando não for uma tarefa rotineira
- ✓ Requer maior **conhecimento da base de dados**

Por que aprender a lidar com ferramentas de análise de dados?

Área	Analista de BI	Cientista de Dados
Foco	Relatórios, KPI's, Tendências	Padrões, Correlações, Modelos Preditivos
Processo	Estático, Comparativo	Exploratório, Experimental, Visual
Fontes de Dados	Data Warehouses, Bancos Transacionais	Big Data, Dados Não-Estruturados, Bancos Transacionais e NoSQL, Dados Gerados em Tempo Real
Qualidade dos Dados na Fonte	Alta	Baixa ou Média (requer processo de limpeza e transformação)
Modelo de Dados	Esquema de dados bem definido na fonte	Esquema de dados definido no momento da consulta
Transformações nos Dados	Pouca ou nenhuma (dados já organizados na fonte)	Transformação sob demanda, necessidade de complementar os dados
Análise	Descritiva, Retrospectiva	Preditiva, Prescritiva
Responde à pergunta:	O que aconteceu?	O que pode acontecer?

Novas tendências em análise de dados

➤ Computação cognitiva

- Big Data

O termo amplamente utilizado na atualidade para nomear conjuntos de dados muito grandes ou complexos, principalmente, dados não estruturados, como logs de acesso, dados de sensores, frases escritas por funcionários e clientes, dados de redes sociais, posições geográficas, e-mails, arquivos no servidor de arquivos, entre outros.



Novas tendências em análise de dados

➤ Computação cognitiva

- Inteligência artificial

Tecnologias cognitivas como o aprendizado de máquina (Machine Learning, ML) e aprendizado profundo (Deep Learning, DL).

Máquina que consegue aprender sem o uso de um código específico - capacidade de se auto modificar e melhorar o próprio desempenho sem intervenção humana.

Utilizada para gerar informações relevantes, a partir de grandes bases de dados (big data)



Explorando algumas bases de governo federal – Cadastros e registros administrativos

➤ **Ministério do Desenvolvimento Social e Agrário - MDS**

- Acesso a tabelas prontas: Área de imprensa > [Dados](#)
- Ferramenta – [TabCad](#): dados agregados do Cadastro Único
- Ferramenta - [Relatórios de Informações Sociais](#): dados agregados
- Ferramenta - [VIS DATA](#) (Visualizador de Dados Sociais)
- [Microdados](#)

Explorando algumas bases de governo federal – Cadastros e registros administrativos

➤ **Ministério do Desenvolvimento Social e Agrário - MDS**

- [Microdados](#) do Cadastro Único

➤ Tipo	Exemplos	Acesso	Cessão e/ou utilização
Dados que identificam famílias e pessoas cadastradas	Nome, documentos pessoais, Número de Identificação Social (NIS), Código Familiar, endereço e telefone.	Sigilosos	Só podem ser fornecidos para a gestão de políticas públicas e para a realização de estudos e pesquisas.
Dados que caracterizam pessoas, famílias e domicílios cadastrados	Características específicas do domicílio de uma família, total de membros de uma família, despesas mensais, rendimentos, escolaridade, situação no mercado de trabalho, etc.	Sigilosos	Só podem ser fornecidos para a gestão de políticas públicas e para a realização de estudos e pesquisas.
Dados não identificados	Aqueles que possam ser agregados, conformando números absolutos e porcentagens, bem como aqueles que, quando correlacionados, não identificam uma pessoa ou família.	Públicos	Podem ser consultados, por meio de diversas ferramentas disponibilizadas pelo MDS.

Explorando algumas bases de governo federal – Cadastros e registros administrativos

➤ Saúde

- Sistema de Informações de Agravos de Notificação – SINAN
 - Sistema de Informações de Nascidos Vivos – SINASC
 - Sistema de Informações de Mortalidade – SIM
 - Sistema de Informações Hospitalares do SUS – SIH/SUS
 - Sistema de Informações Ambulatoriais do SUS – SIA/SUS
 - Cadastro Nacional de Estabelecimentos de Saúde – CNES
 - Sistema de Cadastramento de usuários do SUS – CADSUS
 - Sistema de Avaliação do Programa de Imunização – PNI/API
 - Sistema de Informação de Atenção Básica – SIAB
 - Registro de Câncer de Base Populacional – RCBP
 - Sistema de Informações sobre Orçamentos Públicos em Saúde – SIOPS
 - Sistema de Informações de Beneficiários – SIB
 - Sistema de Cadastro de Operadoras – CADOP
 - Pesquisa Nacional sobre Demografia e Saúde – PNDS
 - Inquérito domiciliar sobre comportamento de risco e morbidade referida de doenças e agravos não transmissíveis
 - Vigilância de Doenças Crônicas por Inquérito Telefônico – VIGITEL
- Informações de Saúde ([TABNET](#))
 - MS não disponibiliza microdados

Explorando algumas bases de governo federal – Cadastros e registros administrativos

➤ Educação

Fontes INEP ([microdados](#)):

- **Censo Escolar – Educação Básica:** Principais Variáveis: A) Escola: situação e local de funcionamento; caracterização da infraestrutura (abastecimento de água, energia elétrica, etc.); disponibilidade de equipamentos didático/pedagógicos; existência de dependências físicas; fornecimento de alimentação escolar; oferta de atendimento educacional especializado; disponibilidade de materiais didáticos de atendimento à diversidade sociocultural. B) Turma: horário de funcionamento; tipo de atendimento; existência de atividade complementar; atendimento educacional especializado; tipo de disciplinas. C) Alunos: idade; sexo; cor ou raça; uso de transporte escolar; necessidade educacional especial; rendimento escolar (aprovado/reprovado); D) Docentes: idade; sexo; cor ou raça; escolaridade; função que exerce na escola; turma que atua; disciplina que leciona;
- **Censo da Educação Superior:** Principais variáveis: vagas, candidatos e ingressantes por área, curso, habilitação, modalidade, por turno, por tipo de processo seletivo e período; alunos matriculados; portadores de necessidades especiais; perdas de alunos; concluintes; alunos com financiamento educacional; alunos-bolsistas; servidores técnico-administrativos por grau de formação, tipo de contrato; número de docentes por grau de formação, regime de trabalho e por sexo; dados financeiros; infraestrutura e instalações, equipamentos e outros recursos institucionais

Explorando algumas bases de governo federal – Cadastros e registros administrativos

➤ Educação

Fontes INEP ([microdados](#)):

- Exame Nacional de Desempenho dos Estudantes – ENADE: instituições, cursos e habilitações; resultados alunos; dados do questionário socioeconômico; médias e os conceitos das instituições de ensino superior.
- Exame Nacional do Ensino Médio – ENEM: variáveis de controle da escola, da prova objetiva, da prova de redação; questionário socioeconômico; características da família e do trabalho do estudante; características do estudo, do estudante e seus valores.
- Prova Brasil e Sistema Nacional de Avaliação da Educação Básica – SAEB*: dados da turma; dados da escola; dados da proficiência em língua portuguesa e matemática; dados do aluno; dados do professor e dados do diretor.

* Prova Brasil: Censitária - Escola localizada na área urbana da rede pública

SAEB: Amostragem probabilística - Escolas localizadas nas áreas urbana e rural da rede pública e da rede privada

Explorando algumas bases de governo federal – Cadastros e registros administrativos

➤ Educação

Fontes MEC (microdados):

- [FIES](#) - Detalhamento dos financiamento a estudantes por mês, contendo UF; município; IES; curso e valor da mensalidade. Detalhamento do perfil dos beneficiários do FIES por sexo; raça/cor; data de nascimento. Sem identificador.
- [ProUni](#) - Detalhamento das bolsas concedidas pelo Prouni por ano, contendo a região; UF; município; IES; curso; modalidade de ensino (presencial ou EAD); turno e tipo de bolsa. Detalhamento do perfil dos beneficiários do Prouni por sexo; raça/cor; data de nascimento e pessoas com deficiência. Identificador: CPF camuflado

Explorando algumas bases de governo federal – Cadastros e registros administrativos

➤ Trabalho

Dados Agregados	BI	Microdados	
		Público	Sigiloso
Síntese do Emprego Formal INSPER Anuário Estatístico da RAIS	Programa de Disseminação das Estatísticas do Trabalho - PDET	Microdados da RAIS desidentificados	Microdados da RAIS
		Microdados da CAGED desidentificados	Microdados da CAGED

Diferença entre microdados e PDET:

- ✓ nos microdados estão disponíveis variáveis, como por exemplo, Indicador de Pessoa com Deficiência, Tipo de Deficiência, Atividade Econômica (Subclasse), Ocupação e Raça/Cor que devem ser utilizadas com certa cautela, pois, algumas são relativamente novas e outras, dado o nível de desagregação, podem apresentar algumas inconsistências. Contudo, em termos gerais, sua utilização em nível mais agregado não está comprometida.

Explorando algumas bases de governo federal – Cadastros e registros administrativos

➤ Trabalho

ITEM	RAIS	CAGED
Levantamento	Anual	Mensal
Tipo de vínculo	qualquer tipo: estatutários, celetistas, temporários, avulsos;	apenas os celetistas
Abrangência	todos os empregados do ano base em 31/12 e movimentação dos admitidos e desligados mês a mês	apenas a movimentação de admitidos e desligados
Obrigatoriedade	todos os estabelecimentos, mesmo aqueles que não apresentaram movimentação	apenas os estabelecimentos que apresentaram movimentação
Útil para estudos	estruturais do mercado de trabalho formal	de análise de conjuntura do mercado de trabalho formal

Vantagens da RAIS e CAGED

- ✓ Abrangência, atingindo até o município, e à diversidade de cruzamento em nível ocupacional e setorial.
- ✓ Disponibilização dos dados em período relativamente pequeno no caso do CAGED.

Limitações da RAIS e CAGED

- ✓ Omissão de declaração dos estabelecimentos e erro de preenchimento.

Explorando algumas bases de governo federal – Cadastros e registros administrativos

➤ Previdência

Dados Agregados	BI	Microdados	
		Público	Sigiloso
Boletim Estatístico da Previdência Social – BEPS	Infologo	N/D	Maciça
Anuário Estatístico da Previdência Social – AEPS		N/D	CNIS

Previdência Social (Regime Geral / INSS): inclui grandes números da Previdência Social, tais como quantitativo de benefícios concedidos, dados do atendimento do INSS, arrecadação previdenciária, valores pagos aos beneficiários, entre outros.

Previdência no Serviço Público (Regimes Próprios / RPPS): inclui informações relativas aos RPPS, provenientes do Sistema de Informações dos Regimes Públicos de Previdência Social – CADPREV.

Previdência Complementar: inclui informações sobre as Entidades Fechadas de Previdência Complementar (EFPC) e seus planos de benefícios.

Saúde e Segurança do Trabalhador: inclui informações sobre acidentes de trabalho, benefícios por incapacidade e motivos de afastamento do trabalho.

Nota: Atentar para diferença entre benefícios e beneficiários na análise das estatísticas da Previdência Social.

Explorando algumas bases de governo federal – Cadastros e registros administrativos

➤ Portal da transparência ([download de microdados](#))

- **Bolsa Família** – Pagamentos e Saques. Contém o valor da parcela paga ou sacada e o mês de referência. Identificador: NIS e Nome
- **Seguro Defeso Pescador Artesanal** – Contém, entre outros, nome do município; portaria defeso ibama; data início e fim defeso; data emissão parcela; número parcela; data saque parcela; valor parcela; data restituição. Identificador: CPF camuflado, PIS e nome
- **Servidores Civis e Militares do Executivo Federal**. Identificador: CPF camuflado e nome
 - Cadastro – Dados como cargo; função; órgão de lotação e órgão de exercício; tipo e situação de vínculo; regime jurídico; data de ingresso e nomeação, etc;
 - Remuneração – remunerações, gratificações, verbas indenizatórias, férias, deduções;
 - Observações – demais observações como: serviço no exterior, valores incluídos na remuneração, etc;
 - Honorários (Jetons) – empresa e valores recebidos
 - Honorários Advocatícios – valores recebidos
 - Afastamentos – data de início e fim de afastamentos