

## **1. Data Zoom Amazônia**

---

Esta proposta tem como objetivo a ampliação do acesso aos microdados da região amazônica, através do portal *Data Zoom*.

O Data Zoom, desenvolvido no Departamento de Economia da Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), com financiamento original da Financiadora de Estudos e Projetos (FINEP), é um portal que disponibiliza, gratuitamente, ferramentas que facilitam o acesso de pesquisadores do Brasil e do exterior aos microdados das pesquisas domiciliares do Instituto Brasileiro de Geografia e Estatística (IBGE). Trata-se de um projeto gerador de bem público, que poupa tempo dos pesquisadores usuários dos microdados socioeconômicos brasileiros e potencializa a realização de estudos e ensino nas grandes áreas de Ciências Sociais. Caracteriza-se, desse modo, como um projeto de inovação. Sua identidade visual e marca foram registradas no Instituto Nacional de Propriedade Intelectual (INPI), em 2014.

O diagnóstico por trás da concepção original do *Data Zoom*, de um lado, é a de que há uma significativa subutilização dos microdados de bases socioeconômicas. Ainda são poucos os pesquisadores, no Brasil e no exterior, que têm conhecimento dos detalhes relacionados à utilização operacional desses dados. Em razão dessa falta de conhecimento, muitos estudiosos da área de Ciências Sociais se limitam à análise de tabulações já prontas, disponibilizadas pelo IBGE ou por outras instituições.

De outro lado, há um desperdício enorme de esforços por parte de pesquisadores experientes na construção dos programas e dos dicionários para uso de microdados. A própria limpeza dos microdados é feita, em geral, de forma separada por cada usuário. Essas tarefas podem ser, relativamente, simples nos casos em que se deseja apenas colocar a base em formato utilizável pelos pacotes estatísticos, ou mais complicada e demorada, nos casos que envolvem o desenvolvimento de algoritmos de construção de dados em painel, algoritmos que aglutinam informações espalhadas em uma série de perguntas (no caso da variável de anos de estudo completos, por exemplo), ou nos casos de compatibilização de dados de uma mesma pesquisa em anos diferentes ou de agregação entre regiões.

No Brasil, a norma sempre foi o desenvolvimento, de forma independente, de dicionários e de algoritmos por centenas de usuários, em geral, localizados nas

instituições com maior experiência no uso dos microdados. Esse processo configurava, de um lado, uma sobreposição de esforços e, de outro, uma considerável barreira à entrada para quem se aventurava na tarefa de conhecer melhor a realidade por trás dos microdados socioeconômicos do país.

O diagnóstico, portanto, era o de que os substanciais avanços recentes, na disseminação gratuita dos microdados, não foram acompanhados de iniciativas que ofertassem os dados prontos para uso.

O *Data Zoom* preencheu essa lacuna, ao disponibilizar programas de acesso a todos os microdados das principais pesquisas domiciliares do IBGE: os Censos Demográficos de 1970 a 2010; todas as PNADs<sup>1</sup> de 1981 a 2015; todas as PNADs Contínuas<sup>2</sup> desde 2012; todas as PMEs<sup>3</sup> de 1991 a 2016; as POFs<sup>4</sup> de 1995-1996, 2002-2003, 2008-2009 e 2017-2018; a PNS<sup>5</sup> de 2013; e as ECINFs<sup>6</sup> de 1997 e 2003. O portal disponibiliza, de forma gratuita, via Internet, programas estatísticos e dicionários que permitem a qualquer usuário, em qualquer lugar do mundo, manipular essas bases.

O portal teve uma enorme repercussão desde o seu lançamento, em meados de 2014. Os programas, disponíveis através do *software Stata*, foram instalados mais de 30.000 vezes (em média, 460 vezes por mês). Cada instalação representa uma base de microdados pronta para uso no computador de um usuário. Mais de 15% das instalações correspondem aos pacotes em inglês. Cerca de 40% dos acessos ao portal são do exterior, na sua maior parte dos Estados Unidos da América (EUA), França, Inglaterra, China e Rússia.

A fim de atender à demanda direta de pesquisadores não usuários do *software Stata*, especialmente os de outras áreas das Ciências Sociais e analistas de órgãos públicos e de pesquisa, o *Data Zoom* está desenvolvendo programas em *R*. Esse software tem a vantagem de ser gratuito e de ser amplamente utilizado por estudiosos de várias áreas e por funcionários de órgãos governamentais.

Destaca-se, ainda, o caráter interativo da plataforma. Os programas possuem simples caixas de diálogo, nas quais o usuário escolhe como deseja manipular os dados, como seleção de conteúdo, anos de interesse, desagregações e compatibilizações. Há uma comunicação entre a equipe do *Data Zoom* e os usuários, que dão sugestões de aperfeiçoamento e avisos de detecção de

---

<sup>1</sup> Pesquisas Nacionais por Amostra de Domicílios

<sup>2</sup> Pesquisas Nacionais por amostra de Domicílios Contínuas

<sup>3</sup> Pesquisas Mensais de Emprego

<sup>4</sup> Pesquisas de Orçamento Familiar

<sup>5</sup> Pesquisa Nacional de Saúde

<sup>6</sup> Economia Informal Urbana

problemas específicos à equipe responsável pelo projeto. Além disso, o usuário tem acesso, no portal do projeto, a toda documentação dos questionários, dicionários e material de apoio para cada base de dados. Todo conteúdo do portal, inclusive as documentações, são disponibilizados em português e inglês.

O público-alvo do projeto tem sido o de investigadores em temas socioeconômicos no Brasil e no exterior; usuários em geral dos microdados domiciliares do IBGE, como pesquisadores, alunos de graduação e de pós-graduação, membros do terceiro setor e jornalistas; e analistas de macroeconomia, desenvolvimento econômico, economia do trabalho, finanças públicas e mercado financeiro em geral.

No âmbito do projeto Amazônia 2030, pretendemos desenvolver o portal *Data Zoom* Amazônia. Partimos da premissa de que o diagnóstico por trás da criação do *Data Zoom* se apresenta, também, em relação aos microdados da região amazônica. Há uma significativa subutilização dessas bases de dados, muitas já distribuídas de forma gratuita. Além disso, os substanciais avanços na disseminação dos microdados não foram acompanhados por iniciativas que disponibilizem os dados prontos para uso. Na verdade, da mesma forma que em relação aos microdados das pesquisas domiciliares do IBGE, ainda são poucos os pesquisadores, no Brasil e no exterior, que têm conhecimento dos detalhes relacionados à utilização operacional dos dados da região amazônica.

A ideia do *Data Zoom* Amazônia é disponibilizar ferramentas de facilitação do uso dos microdados: socioeconômicos e demográficos, dados sobre queimadas, desmatamentos (fotos de satélite), cobertura e uso do solo, comércio exterior, entre outros. O correto uso das fotos de satélite, por exemplo, requer não apenas alto conhecimento técnico para uso do microdado original, mas também um esforço não trivial de agregação espacial (por grids, municípios ou qualquer outra decisão de georreferenciamento) e temporal. O *Data Zoom* Amazônia propõe-se, portanto, a construir algoritmos que permitam ao usuário, por meio de simples caixas de diálogo, definir o nível de agregação desejado.

Além desse primeiro objetivo do *Data Zoom* Amazônia, que pretende harmonizar os processos de limpeza dos dados e disseminar o seu uso, o produto tem como segundo objetivo construir uma plataforma geradora de informações interativas e de fácil absorção por parte de qualquer usuário. A ideia é oferecer tabulações e gráficos interativos que permitam ao usuário visualizar a evolução de indicadores selecionados a partir de recortes, tais como período, nível de agregação e tamanho das unidades. Os gráficos interativos serão desenvolvidos seguindo uma estrutura semelhante à utilizada na plataforma *Gapminder*.

## 2. Estudos sobre a região amazônica

---

Além do desenvolvimento do *Data Zoom* Amazônia, tem-se como objetivo gerar dois estudos sobre a região.

No primeiro estudo, pretende-se fazer um amplo diagnóstico da natureza e do funcionamento do mercado de trabalho na Amazônia. A principal base de dados utilizada será a PNAD-Contínua. A diferença fundamental da amostra mestra da PNAD-Contínua em relação à PNAD e à PME foi a realização da estratificação dos setores censitários em vários níveis (divisão administrativa; estratos geográficos e espaciais de setores de acordo com um conjunto de características; regiões urbanas versus rurais; e renda do responsável pelo domicílio). Tal processo de estratificação ampliou, consideravelmente, o número de domicílios cobertos pela pesquisa na região amazônica.

Os microdados da PNAD-Contínua permitirão:

- i) medir a evolução dos principais indicadores de ocupação no mercado de trabalho (taxa de desemprego, taxa de ocupação, taxa de participação, taxa de informalidade, distribuição dos trabalhadores por ocupação, setor de atividade, e posição na ocupação – empregados com e sem carteira, trabalhadores domésticos com e sem carteira, trabalhadores por conta-própria que contribuem e não contribuem para a Previdência Social, funcionários públicos e militares, etc.) na região, no período recente, para o conjunto dos estados da Amazônia e de acordo com várias desagregações (gênero, raça, idade, estado, urbano *versus* rural, etc.);
- ii) analisar, com base nos dados longitudinais da PNAD-Contínua, transições ao longo do tempo entre as situações de emprego – desemprego - fora da força de trabalho; e entre os diversos tipos de emprego (caracterizados pelas situações de informalidade e formalização mencionados na alínea i);
- iii) analisar a evolução dos rendimentos reais, da desigualdade de renda e pobreza de acordo com as várias formas de participação no mercado de trabalho citadas anteriormente, bem como a evolução dinâmica desses rendimentos procurando medir a exposição dos trabalhadores ao risco de perda de emprego e/ou mudança de *status* ocupacional;

- iv) relacionar o grau de informalidade e de fragilidade socioeconômica dos trabalhadores à estrutura produtiva da região em que estão inseridos (por exemplo, Zona Franca de Manaus, outras regiões metropolitanas, regiões rurais, entre outras);

O segundo estudo terá seu escopo definido ao longo do projeto, dependendo dos resultados obtidos e das prioridades identificadas pela equipe ampla do Amazônia 2030. Esse poderá ser uma análise mais aprofundada de alguma questão identificada no primeiro estudo ou mesmo uma análise detalhada da estrutura produtiva da Amazônia com base nas pesquisas econômicas do IBGE.