

## Orientações práticas para jornalistas investigativos em dados ambientais

Com base nos problemas e soluções apontados no relatório, seguem 10 orientações práticas para jornalistas que desejam usar bases públicas abertas em investigações ambientais. As dicas focam em contornar limitações frequentes – como falta de metadados, dificuldade de download, formatação inadequada – e aproveitar ao máximo os dados disponíveis:

**1. Verifique a atualidade dos dados e busque atualizações** – Sempre observe a data de atualização indicada na base (último ano/mês disponível). Se os dados estiverem desatualizados, procure no site oficial se há relatórios mais recentes ou bases suplementares. Use a Lei de Acesso à Informação (LAI) para solicitar atualizações quando necessário, ou contate diretamente o órgão responsável perguntando se há previsão de novos dados. Essa atitude proativa evita usar informações defasadas que possam comprometer sua apuração. Lembre-se: dados ambientais mudam rapidamente (desmatamentos ocorrem todo mês, novos garimpos surgem, etc.), então trabalhar com dado velho pode levar a conclusões erradas sobre a situação atual.

**2. Avalie a completude e cruzar fontes para preencher lacunas** – Analise se a base contém todas as colunas e registros esperados. Muitas vezes, bases públicas podem excluir informações sensíveis ou ter “buracos” (p.ex., uma tabela de multas sem o nome do infrator ou sem coordenadas). Para contornar incompletudes, combine diferentes fontes: por exemplo, se a lista de embargos estaduais não traz coordenadas, cruze com dados de satélite (MapBiomas, INPE) para inferir localização; se o cadastro ambiental (CAR) não traz nomes de proprietários (por privacidade), use bases de registro de imóveis ou empresas para tentar identificar quem está por trás de um código. Outra estratégia é utilizar fontes não governamentais confiáveis (ONGs, pesquisas acadêmicas) que tenham compilado informações similares – elas podem servir de referência para checar se algo está faltando. Documente no roteiro da apuração quais lacunas existem e como você as supriu ou quais limitações permanecem, para transparência com o leitor.

**3. Utilize técnicas de scraping ou script para obter dados em lote** – Diante da falta de opção de download único, não desista: ferramentas de raspagem de dados podem ajudar a automatizar a coleta fragmentada. Por exemplo, se um portal só permite consulta por município ou por processo, escreva um pequeno script em Python (usando bibliotecas como requests ou selenium) para iterar sobre todos os municípios ou IDs e agregar os resultados em um banco local. Se não souber programar, avalie usar um serviço no-code de web scraping (como WebScraper, ParseHub ou similares) que grave as interações e extraia os dados em lote. Outra dica: procure se existe uma API oculta monitorando as requisições de rede no site – às vezes os portais têm endpoints JSON não documentados que podem ser acessados diretamente. Uma vez obtidos todos os pedaços, reúna-os em uma única planilha ou banco de dados para facilitar a análise. Essa abordagem contorna a fragmentação e garante que você tenha a visão completa do dado.

**4. Converta arquivos para formatos analisáveis e padronizados** – Quando os dados vêm em formatos pouco usuais ou não legíveis por máquinas, faça um esforço inicial de conversão. Se for um PDF com tabelas, utilize ferramentas como Tabula ou a função de importação de tabelas do Adobe Acrobat para extrair os dados para CSV/Excel. Para PDFs escaneados (imagem), empregue OCR (reconhecimento ótico de caracteres) – por exemplo, o Tesseract (via Python) ou serviços online – para extrair o texto, ciente de que pode ser necessário corrigir erros de reconhecimento. Caso receba dados em formato proprietário (p. ex., planilhas .xlsx com muitas abas, ou shapefiles do ArcGIS), considere convertê-los para formatos abertos: salve as planilhas em CSV separados (um por aba), exporte shapefiles para GeoJSON ou CSV com colunas de latitude/longitude se possível. Esses formatos abertos facilitam o uso em diversos softwares (R, Python, QGIS, etc.) e também tornam sua investigação mais reproduzível e compartilhável. Crie um repositório organizado dos dados já convertidos/padronizados antes de começar a análise profunda.

**5. Cheque as licenças de uso e aja com transparência** – Verifique se a base traz alguma informação de licença. Se houver uma licença aberta (como CC-BY-SA, ODbL, etc.), você tem liberdade para usar desde que dê crédito – faça isso explicitamente nas suas publicações (cite a origem dos dados no rodapé da matéria ou gráfico). Se não houver licença mencionada, tecnicamente no Brasil os dados governamentais deveriam ser públicos, mas para segurança jurídica, atue como se estivessem sob CC-BY:

atribua claramente a fonte e o órgão. Evite usar dados que contenham informações pessoais sensíveis sem avaliar as implicações (por exemplo, lista de CPF de infratores – dados pessoais – pode exigir anonimização dependendo do contexto; já nomes de empresas ou coordenadas de crimes são de interesse público). Na dúvida, consulte a equipe jurídica do veículo ou professores de jornalismo de dados sobre até onde você pode expor certos dados. Em suma, seja transparente com o público sobre a origem dos dados e atente para eventuais restrições, mas não deixe a falta de uma etiqueta de licença impedir o uso de dados que são claramente de domínio público.

**6. Busque ou construa a documentação dos dados** – Dado que muitas bases carecem de dicionário de dados, invista tempo em entender o que cada campo significa antes de fazer análises. Leia com atenção o contexto no site de origem: às vezes explicações importantes estão em notas técnicas, portarias ou mesmo em FAQs no portal. Procure por metadados escondidos – por exemplo, arquivos README, manuais em PDF anexos, ou se for um sistema geoespacial, arquivos .xml de metadados. Quando não encontrar, tente contato direto: mande e-mail para o órgão ou responsável (muitos datasets trazem o e-mail do mantenedor) pedindo esclarecimentos sobre colunas obscuras. Outra abordagem é comparar com bases similares: por exemplo, se a coluna “tp\_inf” do dataset de multas do Ibama não está clara, compare com uma coluna equivalente na base de multas do ICMBio para inferir pelo contexto (às vezes os valores ou códigos se repetem). Em último caso, construa você mesmo um glossário durante a apuração – anote o que descobriu de cada variável e em que unidade está medindo. Assim, ao escrever a reportagem, você pode incluir explicações para o leitor ou pelo menos ter segurança de que interpretou corretamente os dados.

**7. Aproveite ferramentas adequadas a cada tipo de dado** – Nem sempre o Excel será suficiente para lidar com dados ambientais, que podem ser geoespaciais, volumosos ou complexos. Portanto, escolha as ferramentas certas: para bases com coordenadas ou polígonos (como desmatamento, reservas, propriedades rurais), utilize um SIG (Sistema de Informações Geográficas) como o QGIS (que é gratuito) para visualizar e cruzar mapas. Para analisar grandes tabelas (milhões de registros, como alertas de desmatamento pixel a pixel), considere usar Python (pandas/GeoPandas) ou R – que lidam melhor com volume – ou plataformas de big data abertas (Google Earth Engine para imagens de satélite, por exemplo). Em casos de séries temporais extensas (por exemplo, evolução mensal de multas, dados diários de desmate), aprenda a usar planilhas dinâmicas ou tabelas pivô no Excel para sumarizar, ou scripts Python para agregar e plotar gráficos rapidamente (bibliotecas Matplotlib/Seaborn). Se não for da

sua expertise programar, recorra a ferramentas no-code dedicadas: por exemplo, o Flourish ou Datawrapper para mapas e gráficos interativos, que aceitam upload de CSVs e geram visualizações ricas sem uma linha de código. Em resumo, moldar a ferramenta ao dado, e não o contrário – isso poupa tempo e revela padrões que talvez fiquem ocultos em ferramentas genéricas.

**8. Tenha cuidado extra com dados espaciais e de coordenadas** – Muitos dados ambientais vêm em forma de mapas ou coordenadas (lat/long). Ao trabalhar com eles, preste atenção a detalhes técnicos: projeção cartográfica (coordinate reference system), datum, unidade de medida (graus, metros). Use softwares especializados (QGIS, ArcGIS) para qualquer operação geoespacial, pois eles lidam com essas questões automaticamente. Quando cruzar camadas (ex: sobrepor polígonos de desmatamento com polígonos de fazendas do CAR), assegure-se de que estão no mesmo sistema de coordenadas e refine a análise – por exemplo, calcule a intersecção geométrica para obter a área sobreposta exata entre um desmate e uma propriedade, em vez de apenas verificar “cai dentro” de forma booleana. Ferramentas como GeoPandas (Python) permitem fazer isso programaticamente. Além disso, ao apresentar resultados de mapas, simplifique e confira nomes: converta coordenadas em nomes de municípios ou regiões para facilitar a compreensão do leitor. E mantenha sempre um fator de verificação visual – por exemplo, abra o shapefile resultante de um cruzamento e confira manualmente alguns casos no mapa para ver se faz sentido (às vezes um erro de projeção ou de join pode posicionar dados em locais absurdos). Esse rigor com dados espaciais evita cair em armadilhas como atribuir um desmatamento à fazenda errada por causa de um deslocamento de coordenadas.

**9. Documente todas as etapas de tratamento dos dados** – Durante a apuração, crie o hábito de registrar o que você fez com os dados: filtros aplicados, colunas criadas, correções efetuadas, fontes complementares utilizadas. Essa documentação pode ser numa planilha de log ou mesmo comentando seu código (se usar Python/R). Isso é vital por dois motivos: (a) Reprodutibilidade – caso alguém questione a matéria, você consegue mostrar de onde veio cada número ou mapa (e o próprio veículo pode republicar a análise posteriormente). (b) Colaboração – muitas investigações são feitas em equipe ou viram projetos contínuos; ter claro o passo a passo permite que outro jornalista ou parceiro (por ex., um professor orientando alunos) entenda e valide seu processo. Inclua também nos seus apontamentos detalhes sobre qualidade dos dados: “Campo X tinha 20% valores nulos, desconsidere”, “Filtrei apenas os autos de infração com descrição contendo ‘flora’ para pegar desmatamento”, etc. – isso mostra

consciência das limitações. Ao final, você pode até publicar uma metodologia resumida junto com a reportagem (ou em um blog pessoal) para dar transparência ao leitor avançado. No contexto de dados ambientais, onde metodologias podem ser complexas, essa prestação de contas enriquece a credibilidade do trabalho.

#### **10. Explore cruzamentos criativos e contextuais, mas valide os achados –**

Seguindo o espírito do relatório, busque combinar bases distintas para encontrar histórias novas. Por exemplo, cruze dados de desmatamento com cadastros de terra para achar os maiores desmatadores, ou sobreponha locais de multas ambientais com dados socioeconômicos (IBGE) para ver impacto em comunidades. Essas conexões entre bases muitas vezes revelam insights que isoladamente não apareceriam. Entretanto, ao obter um resultado intrigante de um cruzamento, valide antes de publicar: contate fontes locais, use imagens de satélite de alta resolução para checagem pontual (por ex., se os dados indicam garimpo em certa coordenada, veja no Google Earth se há mesmo uma mina visível), ou procure o órgão fiscalizador para comentar. Dados abertos podem conter erros de cadastro, coordenadas imprecisas ou nomenclaturas confusas (fazendas homônimas, empresas do mesmo grupo, etc.), então toda correlação encontrada deve passar pelo crivo jornalístico tradicional – entrevistas, verificação in loco quando possível, ouvir o “outro lado” (proprietários, acusados) se for uma matéria-denúncia. Em resumo, use a criatividade para enriquecer a pauta com dados, mas mantenha o rigor para não tirar conclusões indevidas apenas porque “o dado mostrou”.

Seguindo essas orientações, o jornalista estará melhor preparado para lidar com as inevitáveis limitações das bases ambientais abertas e, ainda assim, extrair delas pautas relevantes. A chave é persistência técnica e curiosidade investigativa: driblar obstáculos como um verdadeiro hacker cívico, ao mesmo tempo em que conta histórias de interesse público escondidas nos números.

# Exemplos de pautas jornalísticas usando cruzamento de dados disponíveis

A seguir, apresentamos 10 exemplos factíveis de pautas investigativas que podem ser desenvolvidas com os dados abertos atualmente disponíveis nas bases citadas pelo relatório

(<https://www.abraji.org.br/noticias/estudo-revela-falhas-na-transparencia-de-bases-de-dados-para-apuracao-de-crimes-ambientais-no-brasil>). Para cada exemplo, detalhamos: o tema, quais bases de dados seriam utilizadas (com links para acesso), as técnicas de análise envolvidas, ferramentas sugeridas (abrangendo opções com Python, Excel e soluções no-code), e um possível passo a passo da apuração. Esses exemplos ilustram como, na prática, diferentes conjuntos de dados podem ser combinados para revelar histórias ocultas sobre crimes ambientais.

## Exemplo 1: Quem são os donos do desmatamento ilegal?

Tema da investigação: Identificar quais proprietários ou empresas estão por trás das áreas com desmatamento ilegal, cruzando dados de alertas de desmate com cadastros de terras. O objetivo é revelar “quem ganha com a derrubada da floresta” ao vincular polígonos de desmatamento a terrenos registrados.

Bases envolvidas:

- Dados de desmatamento anual PRODES do INPE – polígonos de áreas desmatadas por ano (por exemplo, shapefiles disponíveis no portal [TerraBrasilis](#) do INPE).
- Cadastro Ambiental Rural (CAR) – base de polígonos de propriedades rurais declaradas, disponível no [Portal de Dados Abertos do CAR](#) (os dados podem ser baixados por estado).
- (Opcionalmente: SNCR – dados do Sistema Nacional de Cadastro Rural, acessíveis via [consulta pública do SNCR](#) – embora não haja download completo, pode ser usado para buscar informações pontuais sobre proprietários, complementando o CAR.)\*

Técnicas usadas: Geoprocessamento (sobreposição de polígonos), análise espacial para calcular intersecções área-a-área, filtragem de dados por ano e município, agregação de áreas desmatadas por propriedade. Também técnicas investigativas tradicionais para verificar quem são os donos legais das áreas identificadas (por exemplo, checando CNPJs ou CPFs em outras fontes).

Ferramentas sugeridas: - QGIS (ou outro GIS) para carregar os shapefiles do PRODES e do CAR, e realizar a operação de intersect (interseção espacial) entre desmatamento e propriedades rurais. - Python (GeoPandas) como alternativa programática para fazer a sobreposição e calcular métricas (útil se quiser automatizar para muitos estados). - Excel/LibreOffice Calc para trabalhar a tabela resultante (por exemplo, ordenar as propriedades por área desmatada, fazer filtros por município). - Google Earth ou Google Earth Engine (no-code via interface) para visualização pontual – verificar manualmente algumas áreas desmatadas e conferir se correspondem a pastos, cultivos, etc., após a identificação. -

Ferramentas no-code de visualização como Datawrapper (mapas coropléticos) para eventualmente apresentar no formato de mapa as propriedades com maiores desmates.

Passo a passo da apuração:

1. Obtenção dos dados: Baixar o shapefile do PRODES Amazônia Legal correspondente ao ano de interesse (ex: desmatamento de 2022) no TerraBrasilis. Baixar também os dados do CAR para o estado ou região investigada (alguns estados disponibilizam o CAR completo via dados.gov.br; caso contrário, usar shapefiles estaduais do Serviço Florestal).
2. Preparação no SIG: Importar ambos shapefiles no QGIS. Garantir que estejam no mesmo sistema de coordenadas. Usar a ferramenta de Intersect (Interseção) para cruzar os polígonos de desmatamento com os polígonos de CAR – isso vai gerar uma camada nova contendo apenas as porções de desmatamento dentro de cada propriedade registrada, herdando atributos de ambos (incluindo o identificador do imóvel rural).
3. Cálculo de áreas: Na camada de interseção, calcular a área desmatada dentro de cada imóvel (QGIS permite adicionar uma coluna de área em hectares). Agregar/somar as áreas por código do imóvel rural (ID do CAR) – isso pode ser feito exportando a tabela e usando uma Tabela Dinâmica no Excel ou via Python (GeoPandas groupby). Identificar as top 20 propriedades com maior área desmatada ilegal recente.
4. Identificação dos proprietários: Usando os códigos ou nomes dos imóveis do CAR, consultar a base do CAR (que geralmente traz nome do declarante ou CNPJ da empresa) – pode estar num arquivo CSV anexo ou via portal do Serviço Florestal. Complementar, se preciso, com consultas no SNCR (buscando pelo município e nome da propriedade) para obter CPF/CNPJ do dono oficial. Nesse ponto, temos uma lista de proprietários ou empresas e a área de desmate não autorizado em suas terras.
5. Verificação e contexto: Para cada grande desmatador identificado, verificar se houve

autorização ou se é de fato ilegal. Isso envolve checar se aquela área tem autorização no SINAFLOR (ver Exemplo 2) ou se recebeu multa/embargo do Ibama (cruzar com base de autuações do Ibama pelo nome do proprietário ou coordenada). Também vale verificar o histórico: a empresa/indivíduo já apareceu em notícias? Possui outras áreas rurais? Está vinculada a algum político? Essa etapa dá profundidade à pauta.

6. Entrevistas e contrapontos: Entrar em contato com os proprietários ou empresas líderes em desmatamento para pedir esclarecimentos (eles podem alegar desconhecimento, fraude por terceiros, ou apontar supostas licenças). Ouvir autoridades (Ibama, MPF) sobre por que aqueles desmates ocorreram sem prevenção ou punição.

7. Publicação: Apresentar os achados com infográficos – um mapa mostrando as áreas mais críticas, e uma tabela/infográfico listando os “campeões do desmatamento” com área desmatada e alguma característica (ex: um é político, outro é fornecedor de gado, etc.). Incluir metodologia explicando que foi um cruzamento PRODES + CAR. Esse tipo de pauta evidencia nomes por trás da estatística do desmatamento e expõe possíveis responsáveis legais pelo crime ambiental, algo de grande interesse público.

## **Exemplo 2: Desmatamento autorizado vs. desmatamento ilegal**

Tema da investigação: Quantificar e exemplificar a diferença entre desmatamento autorizado e não autorizado. A pauta busca detectar onde está ocorrendo desmatamento sem licença ambiental cruzando dados de autorizações oficiais com os polígonos de desmate real. Isso ajuda a mostrar a magnitude do desmatamento ilegal e identificar casos específicos de corte raso sem permissão.

Bases envolvidas:

- SINAFLOR – Autorizações de desmatamento: especialmente as bases do Ibama sobre Autorização de Supressão de Vegetação e Autorização de Uso Alternativo do Solo, disponíveis no [portal de dados abertos do Ibama](#) (contêm registros de licenças emitidas, possivelmente com polígonos ou coordenadas e área autorizada).

- Alertas de desmatamento DETER (INPE) – dados frequentes (mensais) de áreas desmatadas detectadas por satélite, acessíveis via [TerraBrasilis](#) (formato vetor ou raster). Alternativamente, usar os polígonos consolidados do PRODES anual mais recente para simplificar.

- (Possivelmente: Bases estaduais de autorização se existirem – por ex., Mato Grosso e Pará emitem autorizações estaduais de desmate. Contudo, para abrangência nacional, a base do SINAFLOR consolidada já deve englobar autorizações, inclusive delegadas a estados.)\*



Técnicas usadas: Georreferenciamento e comparação de camadas espaciais, filtros por data (para casar o período de autorizações com período de desmate observado), cálculo de overlay de polígonos (para ver quanto do desmatamento detectado coincide com áreas que tinham licença). Também cálculo de estatísticas: porcentagem de desmate ilegal vs legal. E investigação individual de casos extremos (maior desmatamento ilegal identificado, etc.).

Ferramentas sugeridas: - QGIS para sobrepor os polígonos de autorização (SINAFLOR) aos polígonos de desmate (DETER/PRODES) e assim classificar desmatamentos em “dentro de área autorizada” ou “fora de área autorizada”. - Python (GeoPandas) se preferir automatizar o cruzamento e cálculo em lote (especialmente útil se trabalhando com DETER que tem milhares de polígonos). - PostGIS (base de dados geográfica) como alternativa robusta para quem conhece SQL, podendo cruzar tabelas por geometrias e sumarizar áreas. - Excel para compilar resultados finais em tabelas (por exemplo, total de hectares desmatados com licença vs sem licença, por estado). - Flourish/Datawrapper para criar visualizações, como um gráfico de barras comparando desmate legal x ilegal por estado, ou um mapa destacando áreas de desmate ilegal.

Passo a passo da apuração:

1. Reunir dados de autorizações: Baixar do portal do Ibama as planilhas ou shapefiles das autorizações SINAFLOR. Provavelmente são dados tabulares com coordenadas ou polígonos. Confirmar o período coberto (ex: autorizações dos últimos 5 anos). Preparar esses dados no QGIS, unificando camadas se houver separação por tipo (uso alternativo do solo vs supressão).

2. Obter dados de desmatamento observado: Para casar com as autorizações, selecionar um período. Ex: pegar todos os alertas DETER de 2021-2022. No TerraBrasilis, é possível baixar shapefiles por mês ou ano consolidado.

Alternativamente, usar PRODES 2021 e 2022 (mas PRODES pega tudo anual, não distingue legal/ilegal). DETER tem vantagem de ser mais “em tempo real” e capturar dinâmica mensal.

3. Sobreposição espacial: No QGIS, usar a função Join Attributes by Location ou semelhante: comparar cada polígono de desmatamento com a camada de autorizações. Adicionar um campo “autorizado (sim/não)” ao polígono do DETER, indicando se ele cai dentro de algum polígono licenciado (e talvez trazendo ID da licença se sim). Para rigor, considerar também a data: uma autorização é válida dentro de certo período. Então, se possível, cruzar também por datas – ex.: desmate em março/2022 e licença emitida válida até março/2022. Isso pode exigir fazer o join espacial e depois filtrar manualmente correspondências onde o ano coincide.

4. Cálculo de áreas ilegais vs legais: Somar a área de todos os polígonos DETER marcados como “fora de autorização” – isso dá o total de desmatamento ilegal no período. Comparar com o total de desmatamento (tudo DETER) para calcular

porcentagem ilegal. Fazer isso por estado ou município para identificar onde o desmate clandestino é mais prevalente. Exemplo hipotético: “No período X, 80% do desmatamento na Amazônia ocorreu sem qualquer autorização ambiental”.

5. Identificação de casos emblemáticos: A partir do conjunto “desmatamento ilegal”, destacar os maiores polígonos ou clusters. Por exemplo, se houver um polígono de 500 ha não autorizado, isso é muito grande e merece atenção. Obter a localização (município, coordenadas) desses casos e investigar mais: usar o CAR (Exemplo 1) para ver de quem é a área, ou checar se o Ibama embargou o local (cruzando com Embargos do Ibama dataset pelo ponto). Selecionar 2 ou 3 casos para serem estudos de caso na matéria.

6. Checagem in loco (remota): Para os casos escolhidos, olhar imagens de satélite (Google Earth, Planet) para confirmar o desmatamento e talvez identificar o que ocorreu depois (virou pasto? plantação?). Isso ajuda a descrever na matéria e questionar autoridades sobre a finalidade daquele desmate ilegal.

7. Busca de explicações oficiais: Com os números apurados, procurar o Ibama e secretarias estaduais para comentar. Perguntar, por exemplo: “Por que X hectares foram desmatados sem licença no seu estado? Esses casos foram multados? Há dificuldade de fiscalizar?”. Isso enriquece a reportagem com contexto e resposta do poder público.

8. Escrita e visualização: Na redação da pauta, apresentar quanto do desmatamento é ilegal e destacar que existe um sistema de licenças mas que muitos ignoram. Usar um gráfico comparativo (legal vs ilegal) e um mapinha mostrando pontos vermelhos (ilegais) sobre o mapa da Amazônia. Relatar os casos específicos, tipo: “No município Tal, encontramos 300 ha derrubados sem permissão dentro de uma fazenda, cujo dono não tinha nenhuma autorização registrada – caracterizando possivelmente o maior desmatamento ilegal do ano. O Ibama informou que blablabla.” Essa pauta evidencia a eficácia (ou não) do sistema de licenciamento e fiscalização, mostrando quantitativamente a porção clandestina do desmate.

### **Exemplo 3: Rota da madeira: analisando fluxos de transporte de produtos florestais**

Tema da investigação: Acompanhar o caminho da madeira desde a origem até o destino, identificando possíveis esquemas de “esquentamento” de madeira ilegal. A ideia é usar os dados de transporte de produtos florestais para descobrir padrões suspeitos, como volumes incompatíveis, rotas inusitadas ou empresas reiteradamente envolvidas, levantando indícios de extração ilegal sendo disfarçada de legal.

Bases envolvidas:

- DOF – Documento de Origem Florestal (dados de transporte): disponível em dataset aberto no portal do Ibama, por exemplo o conjunto [DOF – Transportes de Produtos Florestais](#). Contém registros de cada remessa de madeira: origem (CPF/CNPJ, possivelmente coordenada ou município), destino (empresa compradora), espécie, volume, data, etc.
- DOF – Autorizações de Exploração (Autex): outro dataset do DOF que lista autorizações concedidas para exploração, útil para comparar o volume autorizado x volume efetivamente transportado. Disponível como [DOF – Autorizações de Exploração Florestal](#).
- Dados de desmatamento ou concessões florestais: para contexto, pode-se usar o PRODES ou dados do Imazon sobre áreas de extração, mas o essencial vem do próprio DOF. Possivelmente também Embargos do Ibama para ver se alguma origem está embargada (embargo proíbe exploração naquela área).

Técnicas usadas: Análise de redes (quem vende para quem), agrupamento e ordenação de dados (top exportadores de madeira, top compradores), detecção de outliers (ex.: empresa pequena transportando volume muito grande), cálculo de distâncias geográficas (para ver rotas estranhas). Também cruzamento simples: verificar se as coordenadas de origem do DOF batem com áreas autorizadas ou se incluem áreas proibidas. Pode envolver scripts para tratar texto (nomes de espécies, etc.) e cálculos de estatística descritiva (média, max, etc., por ator).

Ferramentas sugeridas: - Excel/LibreOffice para filtrar e fazer tabelas dinâmicas iniciais (ex.: total de madeira por estado de origem, por espécie, por empresa). - Python (pandas) para análises mais sofisticadas – por exemplo, agrupar por combinação origem-destino e somar volume, ou identificar automaticamente valores extremos. - Gephi ou \*\* Flourish (Chord Diagram) para visualizar a rede de fluxos: Gephi permite desenhar um grafo onde nós são municípios ou empresas e arestas são volumes de madeira; já um diagrama de cordas (chord) ou Sankey (fluxo) pode ser feito no Flourish para mostrar, por exemplo, volume de madeira saindo de estado A para estado B. - QGIS se quiser mapear origens e destinos geograficamente (por ex., plotar pontos de origem proporcional ao volume extraído). - OpenRefine\*\* para padronizar nomes de empresas (no caso de pequenos erros de digitação nos dados DOF, útil para consolidar CNPJs duplicados com grafia diferente).

Passo a passo da apuração:

1. Download e familiarização com DOF: Baixar os datasets de Transportes DOF e Autex DOF do [dadosabertos.ibama.gov.br](https://dadosabertos.ibama.gov.br). Esses podem vir em CSV grande ou múltiplos arquivos (talvez segmentados por ano). Carregar no Excel ou outro software

para entender as colunas disponíveis: identificar campos de origem (pode ser CPF/CNPJ do vendedor e possivelmente coordenadas ou local de extração), destino (CNPJ comprador ou UF destino), espécie de madeira, quantidade (m3), data, etc.

2. Limpeza e padronização: Se o dataset for muito grande (milhões de linhas), optar por trabalhar com Python/pandas. Caso seja manejável (dezenas de milhares), usar Excel para filtros. Padronizar nomes de espécies ou categorias se necessário, e nomes de empresas: usar CNPJ como identificador é ideal (se disponível) para evitar duplicatas.

3. Análises básicas: Calcular estatísticas gerais: total de volume de madeira transportado no período; quais são as principais espécies; listar os maiores remetentes (origens) e maiores receptores (destinos). Por exemplo, descobrir quais empresas madeireiras aparecem movendo mais volume – isso já pode render uma sub-pauta (“Empresa X lidera transporte de madeira na Amazônia”). Montar tabelas: top 10 origens (pode ser por município ou por nome de autorizados) e top 10 destinos (serrarias, exportadoras).

4. Identificação de padrões suspeitos: Procurar situações anômalas:

- Ordenar a lista de transportes por volume decrescente e observar se há cargas muito grandes em um único DOF (Documento). Se um único documento registra milhares de m3, vale checar se é permitido ou se há erro.

- Verificar rotas geográficas: se a base dá municípios, notar se há madeira sendo transportada de locais improváveis (ex: grande volume saindo de município sem tanta floresta – pode indicar que a origem declarada é falsa). Cruzar rapidamente: usar PRODES para ver se aquele município teve desmatamento compatível com o volume.

- Cruzar com Embargos: filtrar origens de DOF por coordenada ou nome e ver se coincidem com áreas embargadas pelo Ibama (talvez usar a base de “Termos de Embargo”). Se madeira está saindo de área embargada, isso é flagrante de irregularidade.

- Comparar Autex vs Transporte: para grandes remetentes, pegar a autorização (Autex) correspondente e ver quanto volume foi autorizado x quanto foi efetivamente transportado. Se um plano de manejo tinha autorização de 5.000 m3 mas aparecem 8.000 m3 transportados daquela origem, há excedente indicando superexploração.

- Observar empresas receptoras: alguma serraria recebe madeira de muitos lugares diferentes, possivelmente “esquentando” origem ilegal misturada com legal. Destacar empresa que compra de fornecedores em vários estados (sinal de rede ampla).

5. Focus em um caso: Escolher um caso emblemático para aprofundar – por exemplo, um município despontou como origem suspeita. Focar nele: qual empresa local extraiu? Tem histórico de ilegalidades (buscar nome em autos de infração do Ibama ou notícias)? Ou uma empresa desproporcionalmente grande: checar seu histórico, propriedade, sócios (via Receita/ junta comercial). Esse caso servirá como narrativa central na matéria, ilustrando o esquema maior.

6. Apuração adicional: Se identificou possíveis fraudes, tentar ouvir fontes na área: contatar Ibama regional, ICV, Imaflora ou outras ONGs que monitoram madeira para validar achados (“estamos vendo que empresa tal, que era de pequeno porte, de repente transportou 5000 m3 de ipê – isso faz sentido?”). Talvez alguém confirme que é esquema já conhecido de esquematismo via DOF.

7. Confronto e respostas: Procurar as empresas citadas para comentar. Perguntar sobre volumes e origens – às vezes, só de perguntar, se for ilícito, elas negam ou não respondem, o que também é indicativo. Procurar também a autoridade ambiental sobre falhas no sistema DOF que permitiriam tais irregularidades.

8. Publicação: Estruturar a reportagem mostrando “como a madeira ilegal viaja”: comece explicando o DOF, depois revele os fluxos (ex: “Madeira extraída no Pará está sendo enviada para serrarias fictícias no Maranhão para burlar fiscalização”). Usar um gráfico de fluxo ou mapa para ilustrar as rotas principais. Destacar o caso-exemplo com narrativa (entrevistas, etc.). Concluir mencionando se autoridades prometeram apertar o controle. Essa pauta alia análise de dados volumosa com investigação de campo, expondo brechas na cadeia legal da madeira.

#### **Exemplo 4: Garimpo ilegal vs. concessões legais de mineração**

Tema da investigação: Detectar e quantificar atividades de mineração ilegal (garimpo) em áreas onde não há concessões ou permissões legais, cruzando dados de títulos minerários oficiais com evidências de extração (desmatamento, pistas de garimpo). O foco pode ser, por exemplo, garimpo de ouro em áreas protegidas ou terras indígenas, ou simplesmente garimpo fora dos polígonos autorizados.

Bases envolvidas:

- SIGMINE – Cadastros de processos minerários: disponível no [dados.gov.br](http://dados.gov.br) (fornece polígonos ou coordenadas de concessões, requerimentos, autorizações de lavra garimpeira, etc., com status). Havia dois conjuntos: um de processos minerários em geral (68% abertura) e outro de reservas garimpeiras (66%). Esses dados indicam onde há mineração legalizada.

- Alertas de desmatamento em áreas de garimpo: O DETER do INPE tem classificação de alertas por causa; possivelmente alguns alertas são marcados como mineração. Se disponível, usar shapefiles de alertas especificamente de mineração (ex.: DETER garimpo para Amazônia, se existente). Caso não, usar PRODES/DETER normal mas filtrar por áreas conhecidas de garimpo (ex. perto de rios de ouro).
- Dados de Terras Indígenas e Unidades de Conservação: para cruzar e ver se o garimpo está dentro de

área protegida (base disponível via Funai/ICMBio; por exemplo, shapefiles de TIs e UCs no [portal do ICMBio](#)).

- Autos de infração do Ibama (filtrados por infrações de mineração, se possível) para ver se confirmam a presença ilegal onde não deveria.

Técnicas usadas: Geoespacial – sobreposição de camadas: comparar polígonos de áreas tituladas (SIGMINE) com polígonos de desmatamento ou sedimentos de mineração. Uso de filtro por atributo: pegar do SIGMINE apenas títulos ativos (concedidos) ou requeridos vs áreas sem nada. Uso de análises de localização: se um ponto de garimpo detectado fica X km fora da concessão mais próxima. Também agregado por região: ex, quantos hectares de garimpo ativo (pelo desmate) estão fora de áreas legais, e onde.

Ferramentas sugeridas: - QGIS para manusear todos os layers: SIGMINE (provavelmente shapefile dos polígonos de concessão), Terras Indígenas/UCs (shapefile), e raster ou vetor de áreas de garimpo detectadas (se usar MapBiomias, eles têm uma camada de garimpo). QGIS servirá para fazer seleções espaciais (selecionar polígonos de garimpo que não interceptam nenhum polígono de concessão legal, e/ou que intersectam TIs). - Google Earth Engine (plataforma de satélite) para quem souber usar: pode identificar automaticamente assentos de garimpo (visualmente, leitos de rios barrentos) e cruzar com camadas vectoriais – mas requer habilidade de programação JavaScript/Python no cloud. Como no-code, pode-se usar interface do MapBiomias Alerta. - Python (GeoPandas) para replicar as operações espaciais de intersecção e distância de forma programática, caso QGIS fique pesado (depende da escala). - Excel para registrar os resultados por território (ex: montar tabela de Terras Indígenas com número de garimpos ilegais detectados em cada).

Passo a passo da apuração:

1. Mapear concessões legais: Baixar a base SIGMINE – possivelmente um shapefile grande de polígonos ou pontos de minerações. Importar no QGIS e filtrar: selecionar apenas títulos aprovados (exploração mineral ativa) e talvez requerimentos em trâmite. Remover áreas já abandonadas ou negadas (se tiver campo de status). O resultado é a cartografia de onde pode haver mineração legal.

2. Identificar sinais de garimpo: Existem várias abordagens:

- Baixar do INPE se disponível um shapefile de DETER classificados como garimpo (por exemplo, o INPE lançava dados de garimpo no Rio Madeira, etc.).
- Ou usar a coleção do MapBiomias que tem mapeamento de garimpos (MapBiomias

mineração). Talvez eles tenham shapefiles de água barrenta e clareiras de garimpo anualmente. - Na falta, usar PRODES: PRODES tem categoria de desmatamento “mineração” quando é mineração industrial de grande porte – mas garimpo ilegal muitas vezes aparece como desmatamento convencional perto de rios. Então possivelmente complementar com sobrevoo de satélite manual: pegar imagens Landsat em locais suspeitos. Contudo, para um trabalho de dados, podemos supor que MapBiomas ou DETER tenha algo. - Vamos supor que se consiga um layer de pontos ou polígonos de garimpo identificados (por exemplo, coordenadas de balsas ou clareiras em rios).

3. Cruzamento concessão vs garimpo: Com as camadas prontas, usar QGIS para selecionar as ocorrências de garimpo que não estão dentro de nenhum polígono SIGMINE ativo. Isso pode ser feito invertendo a seleção: primeiro fazer Spatial Join do layer de garimpos adicionando um campo indicando o ID de concessão que cobre ele (se nenhum, fica nulo). Ou simplesmente selecione todos que não intersectam a camada de concessões. Esse subconjunto são garimpos ilegais prováveis. Também fazer o inverso: quantos garimpos estão dentro de concessões (garimpos “legais”).

4. Cruzamento com áreas protegidas: Agora, pegar os garimpos ilegais identificados e sobrepor com Terras Indígenas e UCs. Selecionar quantos estão dentro de TIs, quantos dentro de Unidades de Conservação. Isso pode ser resumido em números e nomes – ex: “Encontramos garimpo ilegal ativo em 5 Terras Indígenas, incluindo TI Munduruku (Pará) e TI Yanomami (Roraima), e em 3 Unidades de Conservação federais”. Cada interseção relevante anotada.

5. Quantificação e localização: Gerar métricas: área total minerada ilegalmente (se os dados permitirem área), ou número de pontos de garimpo ilegal, por região. Identificar hotspots: por exemplo, vale do Rio X tem dezenas de pontos fora de concessões, aquilo é um polo de garimpo ilegal.

6. Investigação de campo (secundária): Para um hotspot ou TI específica, complementar com contexto: buscar relatórios ou notícias de invasão garimpeira ali, conversar com especialistas (Ibama, Funai, líderes indígenas) para confirmar a situação em terra. Isso serve para adicionar voz humana e confirmação aos dados geoespaciais.

7. Busca de responsáveis: Consultar a base do SIGMINE se há pedidos de concessão exatamente nas áreas identificadas de garimpo ilegal – às vezes garimpeiros ilegais usam alguém para dar entrada num requerimento para simular legalidade. Se encontrar, citar esses nomes. Verificar também autos de infração do Ibama filtrando por município ou coordenadas, para ver se já houve flagrantes (Ibama Autos dataset permite busca textual; se tiver coordenada, melhor). Isso dá ideia se o governo sabe e multou.

8. Contatos oficiais: Enviar questionamento à Agência Nacional de Mineração (ANM) com seus achados: quantos garimpos ilegais identificados e onde, e perguntar sobre

fiscalização e sobre morosidade em liberar concessões comunitárias (muitas vezes garimpeiros dizem que atuam ilegalmente por falta de licença). Também contatar Ibama ou PF sobre operações (eles podem revelar se planejam algo ou culpam falta de apoio).

9. Publicação: Estruturar a matéria destacando o número de garimpos ilegais e onde ocorrem. Por exemplo: “Levantamento de dados indica pelo menos 150 frentes de garimpo em operação fora de áreas autorizadas na Amazônia, incluindo invasões em terras indígenas”. Mapas serão impactantes aqui – um mapa da região com pontos de garimpo, usando cores para ilegais vs legais, e sombreados as terras indígenas. Inserir declarações, e possivelmente fotos (se conseguir de agências, ou usar captura de tela do Google Earth de um garimpo). A reportagem mostrará claramente a sobreposição (ou falta dela) entre o mapa legal e a realidade, evidenciando as falhas na contenção do garimpo ilegal.

### **Exemplo 5: Tráfico de fauna: mapeando pontos quentes e rotas**

Tema da investigação: Analisar os registros de crimes contra a fauna para identificar onde ocorrem mais apreensões de animais silvestres e quais são os possíveis trajetos do tráfico, desde a captura no interior até a venda em centros urbanos. A investigação busca desenhar um panorama nacional do tráfico de fauna com base nos dados de fiscalização.

Bases envolvidas: - Autos de infração do Ibama – fauna: filtrar o dataset de autos de infração ambientais do Ibama para extrair apenas as ocorrências relacionadas à fauna silvestre (ex.: pela descrição da infração ou código de tipo). O conjunto [Autos de infração – Ibama](#) contém todos os tipos, então será preciso selecionar aqueles pertinentes (caça ilegal, comércio de animais, criação ilegal, etc.). - SISFAUNA: o Sistema de Controle de Fauna (Ibama) tinha 66% de abertura, possivelmente indicando que existe algum dado acessível – talvez registros de criadores autorizados, transporte de animais ou algo do tipo. Se disponível, usar dados de transporte de fauna autorizados ou lista de criadouros comerciais registrados. Isso pode servir para confrontar legal vs ilegal. - Dados estaduais: alguns estados (São Paulo, por exemplo) têm bases de ocorrências de fauna apreendida (pela polícia ambiental). Contudo, mantemos foco nas bases citadas: talvez algum estado amazônico tenha informações via seus autos de infração (pouco provável, dado baixos índices). - Dados complementares: Lista dos animais ameaçados (ICMBio) só para contexto; e talvez bases de comércio internacional (CITES) se coubesse, mas aí é outro escopo.



Técnicas usadas: Filtragem textual e categorização (garimpar nos autos do Ibama as infrações de fauna), geocodificação ou agregação por município/estado para mapear distribuição, análise de rede se houver dados de origem-destino (por ex., animal apreendido em local X mas comprador em Y). Caso possua dados de transporte legal (SISFAUNA), comparar volumes e espécies com os ilegais. Também análise temporal para ver se há sazonalidade (ex.: picos de apreensão em determinados meses).

Ferramentas sugeridas: - Python (pandas) para filtrar os autos: como são milhares de registros, usar pandas para selecionar linhas cujo campo “Descrição” contenha palavras-chave (ex: “fauna”, “animal silvestre”, nomes de espécies) ou códigos de infração específicos. Depois agrupar por UF ou município. - Excel se o volume não for enorme – o autos de infração dataset pode ter, digamos, uns 30 mil registros; Excel pode lidar se filtrando manualmente, mas Python seria mais assertivo. - QGIS para criar um mapa de calor ou marcadores nas coordenadas das infrações (se a base tiver latitude/longitude; não está claro se tem, mas muitos autos Ibama vêm com coordenadas da ocorrência). Se não tiver coordenadas, usar colunas de município/UF para juntar a alguma base geográfica de Brasil e plotar. - Kepler.gl (ferramenta no-code de mapear pontos) para rapidamente fazer um mapa interativo das apreensões se tiver coordenadas. - OpenRefine para padronizar nomes de espécies se isso aparecer nos dados (às vezes as descrições citam animais, poderia extrair principais espécies traficadas).

Passo a passo da apuração:

1. Extração de infrações de fauna: Abrir o CSV de autos de infração do Ibama. Identificar a coluna que indica o tipo de infração (pode ser um código de legislação, por ex: Art. 29 da Lei de Crimes Ambientais refere-se a fauna). Se houver, filtrar por esse código. Se não, usar busca textual: filtrar linhas cujo texto mencione termos ligados a fauna (“animais”, “passeriformes”, “papagaio”, “pássaro”, “fauna”, etc.). Criar uma sub-tabela só com essas infrações.
2. Georreferenciar ocorrências: Ver quais campos de localização existem – normalmente há município e UF, às vezes coordenadas. Se tiver coordenadas, melhor: importar essa lista filtrada no QGIS e visualizar os pontos para ver distribuição. Se não, usar o campo município+UF: pode agregar o número de ocorrências por município ou por estado para ter uma ideia. (Se for por estado, ainda é válido – ex: “São Paulo lidera apreensões de fauna” pode indicar destino final, enquanto “Amazonas tem poucas” pode indicar baixa fiscalização ou origem oculta).

3. Identificar hotspots: A partir do mapa ou da tabela agregada, destacar locais com mais incidentes. Exemplos possíveis: grandes centros urbanos (São Paulo, Rio) costumam aparecer com muitas apreensões – isso indica destino (feiras, etc.); algumas cidades do interior (no Nordeste ou Amazônia) podem aparecer por serem origem ou rota de coleta (ex: cidade X no Pará onde muitos pássaros são capturados). Anotar os top 5 estados e top 10 municípios em número de casos.

4. Análise de espécies e volumes: Se as descrições trazem espécie ou quantidade (por ex: “apreensão de 50 aves silvestres”), extrair essa informação. Isso pode exigir ler manualmente amostra de descrições ou usar regex via Python para pegar números e nomes. Listar quais espécies surgem com mais frequência (papagaio, arara, jabuti, etc.). E volumes: qual foi a maior apreensão registrada (em número de animais). Esses detalhes ajudam a contar histórias concretas.

5. Cruzar com SISFAUNA (opcional): Se conseguirmos dados de criadouros legais (por ex, lista de criadores comerciais de passeriformes em SP), comparar com os ilegais: por exemplo, ver se há estados com muitos criadouros legais mas também muitos casos ilegais – talvez indicando fraude (criador legal vendendo ilegalmente). Ou se há espécies que praticamente só existem no tráfico e não têm quase criação legal (destacando demanda clandestina). Esse cruzamento pode ser complexo mas enriquece insights.

6. Rastrear rotas: A partir dos dados coletados, tentar inferir rotas: por exemplo, se muitos autos no Pará e muitos no SP, conclui-se que animais do Pará são vendidos em SP. Procurar nos autos se há indicação de origem nos autos feitos no destino (às vezes a polícia ambiental informa “animais vindos de tal estado”). Se não nos dados, complementar com pesquisa qualitativa: reportagens e especialistas que digam “os pássaros apreendidos em SP vêm principalmente do Nordeste”. Mapear no papel algumas setas de fluxo prováveis.

7. Investigar casos notórios: Ver se nos dados há casos únicos chamativos – ex: uma apreensão em aeroporto, ou um traficante reincidente (se o mesmo nome ou CPF aparece em mais de um auto, significaria reincidência). Isso pode virar personagem central. Pesquisar no Google pelo nome se ele foi preso, etc.

8. Entrevistas: Falar com policiais ambientais ou agentes do Ibama especializados em fauna para comentar os padrões observados. Eles podem confirmar rotas (“sim, geralmente o pessoal coleta no Nordeste e vende em SP e RJ”), ou destacar desafios (ex: falta de integração de sistemas entre estados). Tentar também contato com algum centro de triagem de animais silvestres (CETAS), onde os bichos apreendidos vão – eles têm dados do volume recebido, que pode reforçar o achado.

9. Publicação: A matéria pode ser estruturada assim: “Mapa do tráfico de animais: Sudeste é o maior mercado, Nordeste é o fornecedor”. Apresentar um mapa do Brasil com setas, ou dois mapas (um de origem, outro de destino), ou mesmo um infográfico estilo “caminho de um papagaio – do Pará até a feira clandestina em SP”. Citar números: quantas ocorrências, quantos animais. Destacar espécies mais visadas. Incorporar fala de especialista alertando para o impacto nas populações silvestres. E possivelmente humanizar com um caso de alguém preso ou de um animal emblemático resgatado. Ao final, incluir resposta do Ibama (ex: “Ibama afirma que reforçou operações, mas admite que fiscalização esbarra na falta de pessoal”). Essa pauta traz à luz um crime ambiental menos visível que o desmatamento, usando dados para dar concretude a ele.

### **Exemplo 6: Agrotóxicos proibidos na lavoura: onde e quem usa**

Tema da investigação: Investigar a utilização de agrotóxicos ilegais ou em desacordo com a legislação – por exemplo, produtos banidos no Brasil que ainda são encontrados em fazendas, ou agrotóxicos aplicados de forma irregular (contrabando, uso fora da cultura autorizada). A ideia é identificar focos de uso clandestino de venenos agrícolas e as empresas envolvidas.

Bases envolvidas:

- Empresas autuadas por agrotóxicos ilegais: dataset citado no relatório (24% de abertura) que lista empresas multadas pelo Ministério da Agricultura ou Ibama pelo uso irregular de defensivos. Se disponível, isso daria nomes de empresas/fazendas, local (UF) e motivo (produto X proibido encontrado). Pode estar nos portais do MAPA (talvez em formato de relatórios).
- Sistema Agrofit (MAPA): base aberta de agrotóxicos registrados [Agrofit](#), contendo quais produtos são permitidos para quais culturas, etc. Útil para checar se um princípio ativo encontrado é permitido ou não.
- Autos de

infração Ibama: filtrar por casos envolvendo agrotóxicos (há operações do Ibama, ex: Operação Deriva, que multou uso irregular). Buscar no dataset de autos termos como “agrotóxicos”, “agroquímicos”, ou códigos específicos.

- (Opcional: Dados estaduais de saúde ou vigilância, ex: casos de intoxicação por agrotóxicos por região – mas vamos nos ater a ambientais).\*

Técnicas usadas: Cruzamento de listas de substâncias com listas de permissões (para identificar se autuação foi por produto banido), agrupamento geográfico (quais estados têm mais ocorrências), e linkagem de empresas a setores (ex: se as empresas autuadas são principalmente algodozeiras de MT ou fruticultores do CE). Também alguma mineração de texto para extrair dos autos os nomes dos produtos envolvidos.

Ferramentas sugeridas: - Excel para organizar a lista de empresas autuadas: se vier em PDF ou DOC, colar no Excel e estruturar colunas (Empresa, Cidade, Produto, Ano, etc.). - OpenRefine para limpar nomes de empresas (remover variações, CNPJ etc.) e consolidar duplicatas. - Python (Pandas) para combinar com Agrofit: por exemplo, se temos nome do princípio ativo apreendido, Pandas pode buscar no dataset Agrofit se ele consta como registrado ou proibido. - QGIS para mapear casos por estado (por ex., um mapa do Brasil com intensidade de ocorrências por UF). - Power BI ou Tableau (no-code) se quiser criar um painel mostrando, por exemplo, filtros por ano, por tipo de cultura (mas isso pode ser overkill; uma visualização estática bem feita basta).

Passo a passo da apuração:

1. Obter lista de autuados: Encontrar a fonte da lista de empresas autuadas. Talvez no site do MAPA haja um relatório anual ou planilha. Se não achar facilmente, usar o dado do relatório como base e fazer pedido via LAI ao MAPA para fornecer a relação (como é uma base citada, supõe-se que exista publicamente). Uma vez com ela, estruturar os dados em planilha, incluindo: nome da empresa ou pessoa autuada, local (município/UF), ano, substância ou motivo da autuação.

2. Analisar distribuição geográfica: Agregar a quantidade de autuações por UF. Provavelmente estados agrícolas terão mais (Mato Grosso, Paraná, etc.). Notar se algum estado foge à regra (por ex, “Bahia tem muito mais casos relativos à área cultivada do que São Paulo”, indicando fiscalização direcionada ou problema local).

Mapear esses dados.

3. Identificar substâncias ilícitas: Da coluna de “motivo”, extrair os nomes dos produtos químicos citados. Cruzar com o dataset do Agrofit: marcar quais estão proibidos no Brasil (Agrofit tem lista de produtos registrados; se não consta, ou se registro cancelado, então é contrabando). Por exemplo, substâncias como Paraquat e Endosulfan são banidas – se aparecem nas autuações, evidenciar. Quantificar: “Foram encontrados X casos envolvendo agrotóxicos totalmente proibidos no país e Y casos de uso irregular de produtos restritos”.

4. Perfil das empresas autuadas: Ver os nomes – são produtores rurais individuais? São grandes fazendas (muitas vezes dá para identificar pelo nome, ex: “Fazenda Santa Luzia” ou “Usina Tal”)? Ou cooperativas/agroindústrias? Classificar manualmente os maiores casos. Se houver CNPJ, usar Receita (via Google ou base CNPJ aberta) para descobrir ramo (cana, soja, frutas?). Isso ajuda a dizer “setor sucroalcooleiro lidera infrações” ou “muitas autuações envolvendo produção de soja em MT e algodão em BA”, etc.

5. Casos exemplares: Escolher um ou dois casos para aprofundar. Exemplo: uma grande fazenda de soja em MT multada por usar um inseticida chinês proibido – pesquisar se já saiu na mídia local, se reincidente, tentar contato. Ou uma cooperativa de hortaliças em SP usando agrotóxico veterinário em verduras (caso hipotético) – isso ilustraria risco à saúde. Coletar detalhes: quantia apreendida, destino (as vezes apreensão indica se era para exportar ou para mercado interno).

6. Entrevista autoridades e especialistas: Falar com fiscal do MAPA responsável por agrotóxicos ou Ibama para comentar as dificuldades de controle – contrabando vindo via fronteira, etc. Também um especialista em saúde pública ou ambiental para dimensionar o dano desses ilegais (por ex, “Tal produto causa câncer e foi banido, mas está sendo usado clandestinamente – risco para trabalhadores e consumidor”).

7. Contrapontos: Se identificar empresas específicas de grande porte, procurá-las para comentar. Muitas podem alegar desconhecimento (ex: “um funcionário usou por conta própria, já demitimos”) ou contestar a autuação. Registrar as respostas.

8. Publicação: Estruturar mostrando onde e o quanto se usa veneno ilegal. Talvez título: “O mapa dos agrotóxicos proibidos: [Estado] lidera uso clandestino de venenos agrícolas”. Usar infográfico: um mapa do Brasil com barras ou cores indicando autuações por estado. Listar as substâncias mais encontradas e por que são perigosas (ex: “5 casos de Paraquat, banido por toxicidade”). Citar exemplos: “Em Primavera do Leste (MT), uma fazenda foi flagrada usando pesticida importado ilegalmente – a substância tem uso vetado aqui desde 2018.” Colocar falas de especialistas alertando

para falta de fiscalização e falhas na política de controle (pode mencionar que o índice de abertura dessa base era baixo, sugerindo pouca transparência também). Concluir com o que governo diz que fará para melhorar (talvez ações de cooperação com aduanas, etc.). Essa pauta combina dados e saúde pública, com apelo tanto ambiental quanto de consumo.

### **Exemplo 7: Grilagem digital: registros sobrepostos indicando fraude fundiária**

Tema da investigação: Detectar indícios de grilagem de terras a partir de dados públicos de cadastro – especificamente, procurando sobreposição indevida de registros ou registros suspeitos em áreas públicas. A pauta revela como territórios inteiros podem estar sendo apropriados ilegalmente “no papel”, ao encontrar múltiplos cadastros (CAR, registros) sobre a mesma área ou cadastro em terra que deveria ser da União.

Bases envolvidas:

- Cadastro Ambiental Rural (CAR) – polígonos de imóveis rurais declarados. Especialmente, interesse nos CARs inscritos dentro de áreas que não poderiam ser privatizadas (florestas públicas não designadas, terras devolutas, áreas protegidas).
- SIGEF / SNCI (Incra) – base de imóveis rurais titulados oficialmente (SIGEF teve 0% abertura, mas talvez dados fragmentados via SIGEF web; SNCR 45% com consulta pública). Talvez não seja possível obter shapefiles facilmente, mas há como verificar se determinada área tem título.
- Limites de áreas públicas: shapefiles de Florestas Públicas Federais (serviço do Serviço Florestal Brasileiro), de terras devolutas estaduais (variável, talvez disponíveis em geoportais estaduais) e limites de terras indígenas e unidades de conservação (via Funai e ICMBio, conforme citado antes). Também dados de assentamentos rurais (Incra) – que são áreas públicas destinadas – facilmente obtido via Incra.
- Embargos do Ibama: pode ser transversalmente útil, pois área embargada por desmatamento em terra pública sugestionaria grilagem.

Técnicas usadas: Geoespacial intensiva – detectar sobreposição de polígonos do CAR entre si (múltiplos CAR sobre mesma área, conhecido como “overlay” indicando disputa) e sobreposição de CAR em camadas de áreas públicas (ex: CAR dentro de Floresta Nacional ou sobre terra devoluta). Também verificar tamanhos anômalos (CARs enormes, acima do que a lei permite para particulares). Necessário cuidado de processamento, pois CAR tem milhões de registros; talvez focar por região de interesse ou pegar amostra (ex: escolher um município conhecido por grilagem para zoom).

Ferramentas sugeridas: - QGIS ou ArcGIS Pro – idealmente ArcGIS Pro pela capacidade de geoprocessar milhões de polígonos com mais eficiência (se disponível), mas QGIS com bons plugins pode dar conta se dividir por estado. Usar ferramentas de geoprocessamento como Union ou Intersection para encontrar interseções entre polígonos CAR (mas isso é pesado). Em vez disso, usar aproximações: densidade de CARs (pode gerar raster de count of overlaps) ou inspeção visual. - PostGIS – uma abordagem avançada seria carregar os dados de CAR em um banco geográfico e rodar consultas de sobreposição. - Python (GeoPandas) pode funcionar para subconjuntos (por ex, por município). - Google Earth Engine – não lida bem com vetores complexos, mas poderia talvez processar rasterização de CARs. Complexo. - Excel – para análises complementares: ex, extrair lista de CARs que caem em terra indígena (via join espacial) e depois filtrar duplicatas para ver quantos sobrepõem e quantos declarantes distintos – Excel aqui como ferramenta para manusear a lista resultante.

Passo a passo da apuração:

1. Recorte da área de estudo: Devido ao volume de dados do CAR nacional (~6 milhões de imóveis), defina um recorte prioritário. Ex: “Amazônia Legal” ou um estado específico com histórico de grilagem (Pará, Amazonas, RO). Baixar os shapefiles do CAR desse recorte (o Serviço Florestal fornece por estado).
2. CAR sobre áreas protegidas: No QGIS, carregar as camadas de Terras Indígenas, Unidades de Conservação de proteção integral e Florestas Públicas não destinadas. Usar a ferramenta de Select by Location para selecionar todos CARs que intersectam essas áreas (ou, inversamente, recortar a parte do CAR que fica dentro). O resultado: quantos e quais imóveis privados foram cadastrados indevidamente dentro de florestas públicas ou TIs – um claro sinal de grilagem (registro indevido). Tabular: por exemplo, “Encontramos 2000 cadastros rurais sobrepostos a Terras Indígenas, o que é ilegal”. Destacar casos: ex: Terra Indígena Apyterewa tem 50 CARs inscritos sobre seu território.
3. CAR sobre terras devolutas: Essa é mais complexa porque falta base consolidada de terras devolutas (terras públicas sem destinação). Uma proxy: áreas fora de qualquer propriedade registrada no SIGEF (mas SIGEF não temos). Alternativa: pegar mosaico de CAR e ver grandes sobreposições entre si ou áreas de florestas públicas federais não destinadas (existe shapefile dessas florestas públicas na Amazônia Legal). Selecionar CARs dentro dessas florestas públicas. Isso dará casos de grilagem em terras da União.
4. Sobreposição CAR vs CAR: Identificar áreas de conflito (dois ou mais CARs sobrepostos indicando disputa pela mesma área). No QGIS, uma forma: dissolver

todos CARs em um único layer sem unir atributos e contar sobreposição (difícil direto). Outra: converter CARs em raster de baixa resolução e somar, pixels com valor >1 indicam sobreposição. Ou dividir por municípios e manualmente checar visivelmente regiões onde vários polígonos se sobrepõem (um sinal comum: formas idênticas duplicadas ou um polígono grande cobrindo vários pequenos). Para exemplificar, talvez pegue caso conhecido: p.ex., em Altamira-PA, onde grileiros registram várias vezes áreas no CAR. Identificar um ou dois locais e quantificar: “na região tal, uma mesma área de 50 mil ha foi registrada por 3 diferentes pessoas no CAR – todos alegando posse”.

5. Tamanhos suspeitos: Ordenar a lista de CARs por área declarada. Imóveis acima de certo tamanho (ex: maior que 15 mil ha, que é latifúndio imenso) merecem atenção – legalmente, dependendo do estado, pode até exceder o limite de concessão. Anotar esses grandões e ver se por acaso caem em área pública (muitas vezes sim – grileiro registra área gigante dentro de floresta pública).

6. Investigação de casos: Selecionar 2 casos emblemáticos descobertos: - Um CAR dentro de terra indígena (tentar contato com lideranças ou Funai para comentar, e com o suposto dono do CAR via contato no Sistema, se disponível). - Um caso de sobreposição múltipla (falar com Incra ou ouvidoria do CAR para saber se têm processo de anulação, etc.). - Talvez um caso de CAR em nome de pessoa famosa ou político (checar nomes se aparecem – às vezes políticos locais registram terras, se o nome consta e é reconhecível, vira notícia).

7. Autoridades: Ouvir o Serviço Florestal/Incra: o que estão fazendo sobre CARs fraudulentos? Existe filtro? Verificar se já cancelaram quantos. Trazer dados oficiais se existirem (ex: “já invalidaram X cadastros sobrepostos a TIs”).

8. Publicação: Título algo como: “Registros conflitantes expõem grilagem: milhares de falsas propriedades reivindicam áreas públicas na Amazônia”. Incluir mapas: um mapinha mostrando um exemplo (pode ser um detalhe de um TI salpicada de polígonos de CAR – visualmente forte). Uma tabela ou gráfico dizendo: quantos CAR em terras indígenas, quantos em UCs, quantos sobrepostos a si. No texto: “Nos dados do CAR, deveriam constar apenas propriedades legítimas, mas encontramos numerosos registros duvidosos: X% da área cadastrada na região Y está em território proibido ou duplicada. Isso sugere uso do CAR para tentar legitimar posses irregulares – a conhecida grilagem digital.” Complementar com falas do órgão (“Estamos auditando, mas é difícil...”) e de especialistas (“isso enfraquece o CAR como instrumento ambiental”). Essa pauta revela um problema estrutural usando puramente dados públicos.

## **Exemplo 8: Os reincidentes ambientais: quem mais comete infrações?**



Tema da investigação: Identificar os maiores reincidentes em crimes ambientais – pessoas físicas ou jurídicas que acumulam múltiplas autuações – e examinar por que continuam a infringir a lei. A matéria revelaria “os piores infratores do meio ambiente”, questionando a efetividade das punições.

Bases envolvidas: - Autos de infração do Ibama: lista de infrações ambientais federais com nome do autuado (às vezes CPF/CNPJ ou razão social) e detalhes. Já que a base é nacional e inclui desde desmate até pesca ilegal, é a fonte principal para contabilizar reincidências. - Autos de infração estaduais (quando disponíveis, ex: Mato Grosso, Pará) para verificar se alguns infratores também têm multas estaduais – mas pode complicar, foquemos no federal. - Embargos do Ibama: complementa, pois muitas vezes o mesmo infrator terá áreas embargadas. Pode usar para verificar se os reincidentes tiveram áreas fechadas e ainda assim continuaram.

Técnicas usadas: Data mining de texto ou identificação de entidade – padronizar nomes de autuados (pois “João da Silva” pode aparecer em caixa alta, minúscula, ou com CPF). Agregação por nome/CPF para contar número de autos e valor total de multas. Ordenar para achar top 10. Investigação qualitativa de cada top (setor de atuação, local, status das multas).

Ferramentas sugeridas: - Python com pandas para agrupar por autuado. Ideal se a base tiver campo CNPJ/CPF separado – aí é fácil somar por esse identificador. Se só tiver nome, usar Python + fuzzy matching (library fuzzywuzzy) ou heurística para juntar grafias similares, ou OpenRefine manual clusterizando nomes. - OpenRefine de fato é bom aqui: carregar coluna de “Autuado” e usar facet de similaridade (metaphone) para unir variações. Mas cuidado para não juntar pessoas distintas com nome comum. Talvez focar em pessoas jurídicas (empresas) que têm CNPJ claro – essas são mais fáceis de consolidar e muitas vezes figuram entre maiores infratores (grandes madeireiras, etc.). - Excel para finalizar lista consolidada e ordenar. - Sistemas externos: consultar a Dívida Ativa da União (via site da PGFN) para ver se esses infratores têm dívidas de multas ambientais – complementa dizendo “fulano não pagou”.

Passo a passo da apuração:

1. Consolidar autuações por infrator: Carregar o CSV de autos do Ibama no pandas.

Se há coluna de CPF/CNPJ, usar diretamente: groupby nesse campo, contar autos e somar valor das multas (havia campo de valor da multa?). Se não houver valor, só contagem. Se não há CPF/CNPJ, usar nome: primeiro normalizar (tudo maiúsculo sem acento) para minimizar variações. Agrupar por nome e contar. Extrair top 20 nomes com mais autos.

2. Refinar identificação: Ver manualmente esses top nomes: eliminar falsos agregados (ex: “NOME NÃO IDENTIFICADO” pode aparecer em vários, tirar). Se tem muitos “JOSE DA SILVA”, verificar se são a mesma pessoa (provavelmente não, comum demais; esses talvez não fiquem no topo a menos que um fiscal esqueça de por CPF). Empresas: nomes como “Fazenda XX Ltda” provavelmente aparecem. Priorizar esses porque são mais confiáveis de serem a mesma entidade.

3. Obter contexto dos reincidentes: Para cada dos top 10 selecionados, levantar: quantas infrações, de que tipo (pandas: filtrar registros daquele nome e ver as descrições mais comuns), valor total multado. Por exemplo: “Empresa Y recebeu 15 multas somando R\$50 milhões, principalmente por desmatamento ilegal e uso de trabalho escravo” – às vezes as autuações incluem todo tipo de infração correlata.

4. Verificar status das multas: Se possível, usar o site da Dívida Ativa ou ferramenta da CGU para ver se essas multas foram pagas. Provavelmente muitas não foram. Isso pode virar parte do texto (“dos top 10 infratores, 8 não pagaram as multas e recorrem na Justiça”).

5. Localização e setor: Pegar do dataset os campos de município/UF para ver onde cada infrator age (pode ser multiestado). Identificar setor econômico: pelo nome ou infração dá para inferir (ex: uma madeireira, um fazendeiro, uma mineradora). Também checar se algum é político ou servidor. Com nome/CNPJ, pesquisar notícias e processos jurídicos. Montar um breve perfil de cada: ex: “João da Silva – fazendeiro no Pará, responde por 10 autos de infração por desmatamento em área embargada, já foi alvo de operação do MPF”.

6. Checar embargos: Cruzar cada nome com a lista de embargos do Ibama (que talvez tenha nome do responsável). Se não possível, pelo menos ver se nas descrições das multas consta embargo ou crime continuado. Isso reforça reincidência: “mesmo com área embargada, desmatou de novo”.

7. Ouvir os citados: Essencial – entrar em contato com as empresas ou pessoas listadas. Provavelmente muitas não responderão ou vão dizer que estão recorrendo, que discordam das multas. Incluir essas falas ou menção de que não retornaram.

8. Oficial: Pedir comentário do Ibama: Por que esses reincidentes continuam? Há dificuldade em cobrar multas? Existe lista de infratores prioritários? À Justiça: por que demora julgar?

9. Publicação: Formato tipo “Os 10 campeões de multas ambientais do Brasil”. Listar eles num quadro, com nome, no de infrações, valor total, principal local e crime. O texto explora 2 ou 3 casos principais em detalhe e discute a falha do sistema de punição – o

infrator contínuo indica que multa não está sendo suficiente dissuasória. Pode citar que das X multas desses 10, poucas foram pagas (pegar dado do TCU ou CGU se existir). Essa matéria aponta para impunidade: dados comprovam que alguns destroem o meio ambiente reiteradamente sem consequências efetivas.

### **Exemplo 9: Fiscalização sob lupa: comparação entre atuação federal e estadual**

Tema da investigação: Avaliar se os estados da Amazônia estão cumprindo seu papel na fiscalização ambiental ou se dependem principalmente do Ibama. Em outras palavras, comparar quantidade de autos de infração estaduais vs. federais em cada estado e identificar discrepâncias. Isso revela onde a fiscalização local é fraca (“deixa tudo para o Ibama”), possivelmente contribuindo para mais crimes ambientais.

Bases envolvidas: - Autos de infração estaduais: dados disponíveis em alguns estados – conforme o relatório, Mato Grosso, Pará, Amazonas, Rondônia, Tocantins tinham informações abertas (embargos e autos). Por exemplo:

- MT: Portal SEMA/Indea ([Geoportal SEMA MT](#)) – lista de autos de infração ambientais do estado.
- PA: Portal Semas ([Lista de Desmatamento Ilegal – LDI](#) inclui autos/embargos estaduais).
- AM: IPAAM ([IPAAM Geoportal](#)) – dados de autos/embargos.
- RO, TO possivelmente via transparência (links nos achados do relatório). - Autos de infração do Ibama: filtrados por estado (há coluna UF do fato).

Técnicas usadas: Agregação simples: contar número de autos do Ibama em cada estado vs número de autos do órgão estadual no mesmo estado, no mesmo período. Cálculo de proporções ou índices (por exemplo, quantos autos por km<sup>2</sup> de desmatamento ou por 1000 km<sup>2</sup> de área – para contextualizar). Comparação por gráfico para evidenciar diferença.

Ferramentas sugeridas: - Planilha Excel para compilar dados estaduais (provavelmente vêm em planilhas/PDFs separadas) e somar quantidades. - Python para filtrar a base do Ibama por UF e contar, garantindo precisão. - Datawrapper para gráficos comparativos (barra dupla: Ibama vs Estado para cada UF; ou gráfico de linhas se tiver

séries temporais). - QGIS (opcional) para mapa coroplético mostrando onde a fiscalização estadual é praticamente inexistente vs ativa (mas tabela pode bastar).

Passo a passo da apuração:

1. Coletar dados estaduais: Ir em cada portal: - MT: possivelmente baixar uma planilha de autos de infração (Mato Grosso foi bem avaliado então deve ter dados consolidados, talvez CSV). - PA: a LDI do Pará talvez tenha número de embargos/autos, mas se não, entrar em contato com Semas PA ou extrair de relatórios. - AM, RO, TO: usar links do relatório que apontam para transparência (e possivelmente páginas HTML com listas, pode ser necessário scraping ou manual). Se algum estado não disponibiliza abertamente (ex: Maranhão e Acre sabemos que zero), podemos imputar zero para fins de análise comparativa. - Organizar numa tabela: colunas Estado, Autos pelo Estado (no no último ano, ou média anual), Autos pelo Ibama (no mesmo período).
2. Extrair dados do Ibama por estado: Usar o dataset do Ibama e fazer um pivot: contar quantos autos com UF = AC, AM, AP etc. Se for dos últimos anos, limitar ao mesmo período dos dados estaduais (ex: 2020-2023). Isso dá a coluna Ibama.
3. Análise: Calcular para cada estado a diferença ou razão: ex: “No Amazonas, Ibama fez 1200 autos em 3 anos, enquanto o IPAAM fez 300 – Ibama fez 4 vezes mais autuações que o governo estadual.” Em Mato Grosso, talvez o estado fez até mais que Ibama (chute, dado eles têm boa estrutura).
4. Identificar casos extremos: Provavelmente Acre, Maranhão, Roraima – estados com zero ou quase nenhuma autuação estadual, versus o Ibama atuando (se o Ibama atuou lá, pois se nem Ibama muito, pode estar sem fiscalização nenhuma!). Esses casos extremos merecem destaque. Ex: “No Maranhão, de 2018 a 2022, não se registrou nenhum auto de infração pelo órgão estadual de meio ambiente, deixando toda fiscalização a cargo da União.”
5. Contextualizar com desmatamento: Seria interessante ver se a falta de fiscalização estadual coincide com altas taxas de desmatamento ou outros crimes. Por exemplo, Roraima e Maranhão tiveram aumento de desmate nos últimos anos; sem estado atuando, isso casa. Mencionar dados do PRODES por estado para mostrar o paradoxo: “Maranhão teve 2o maior desmate do Cerrado mas zero atuação estadual.”
6. Questionar os estados: Enviar perguntas aos órgãos de meio ambiente estaduais, especialmente os com baixos números. Perguntar por que, se falta pessoal, se confiam no Ibama, etc. Alguns podem não responder ou dizer que fazem “educação ambiental” em vez de multa. Incluir seus posicionamentos ou o silêncio.
7. Ibama: Pegar declaração do Ibama se possível, sobre esse desequilíbrio – eles talvez diplomaticamente não critiquem estados, mas podem dizer que intensificam presença onde órgãos locais não atuam.

8. Publicação: Formato de reportagem: “Fiscalização ambiental capenga nos estados: Ibama sozinho em boa parte da Amazônia”. Apresentar um gráfico ou tabela comparativa (ex: Estado X – Ibama: 100 autos, Estado: 5 autos). No texto, evidenciar exemplos: “No Acre, nenhum infrator ambiental foi multado pelo órgão estadual em 2021, enquanto o Ibama aplicou 50 multas no mesmo período. Já em Mato Grosso, a situação é diferente: o estado lavrou mais autos (200) que o Ibama (150), indicando estrutura local forte.” Tentar explicar as causas das disparidades – orçamento, vontade política, etc. Colher opinião de especialista (ex: TI/Abraji autores do relatório podem comentar implicações disso – ex: crime migra para onde fiscalização é fraca). Essa matéria, ao cruzar dados de diversas fontes, faz um cheque de accountability nos estados, estimulando melhorias.

### **Exemplo 10: Quando crimes se sobrepõem: trabalho escravo em áreas de desmatamento**

Tema da investigação: Verificar a hipótese de que crimes ambientais andam junto com crimes trabalhistas, especificamente se há incidência de trabalho escravo em fazendas ou madeiras envolvidas em desmatamento ilegal. A ideia é cruzar a “Lista Suja” do trabalho escravo (empregadores flagrados explorando trabalho análogo à escravidão) com dados de desmatamento e infrações ambientais para achar conexões – revelando redes criminosas multifacetadas.

Bases envolvidas:

- Cadastro de Empregadores (Lista Suja do Trabalho Escravo): disponível no Portal de Dados Abertos (formato CSV) listando nomes de empregadores, CPF/CNPJ, e geralmente município/ UF, além de setor de atividade. É atualizada periodicamente pelo Ministério do Trabalho. - Autuações/ Embargos ambientais (Ibama): base de autos de infração e embargos, para cruzar pelo nome do empregador ou localização. - Dados de desmatamento (PRODES/DETER): para ver se as fazendas na lista suja correspondem a áreas com desmatamento recente. - CAR: para localizar a propriedade do empregador (se ele tiver CAR em seu nome ou de sua empresa, se disponíveis publicamente via cruzamento de CNPJ). - (Opcional: SIGMINE/ANM se for garimpos com trabalho escravo – houve casos de escravidão em garimpo. Mas talvez foco agropecuária/madeira que são mais comuns.)\*

Técnicas usadas: Correspondência de entidades – pegar nome/CNPJ da lista suja e buscar nas bases ambientais se aparecem. Isso requer padronização e busca textual.

Em caso de não achar por nome exato (pode estar diferente), usar CNPJ se constar nas autuações (Ibama nem sempre divulga CNPJ, mas às vezes nome fantasia). Geográfico: usar município como indicador – se certa fazenda de lista suja está em município X, ver se esse município é hotspot de desmatamento ou tem embargos. Pode também cruzar coordenadas: alguns casos de escravidão vêm com coordenadas da fazenda (não no CSV, mas em relatórios do MPT).

Além disso, pode se tentar cruzar indireto: pegar coordenadas de embargos e ver se algum embargo cai perto de local de resgate de escravo (difícil sem dados georef de escravo). Principal é link por nome/ CNPJ: ex: “Fazenda Boa Esperança Ltda” está na lista suja; ver se nos autos do Ibama existe multa para “Fazenda Boa Esperança Ltda” – se sim, bingo, temos uma conexão concreta.

Ferramentas sugeridas: - Python (pandas) para leitura do CSV da lista suja e do CSV de autos do Ibama, e tentativa de merge por CNPJ (se ambos tiverem) ou join parcial por nome (pode usar fuzzy matching para pegar empresas com nomes similares). - OpenRefine para normalizar nomes de empregadores da lista suja e nomes de autuados do Ibama, e tentar encontrar matches manualmente. - Excel para consolidar resultados e marcar correspondências encontradas. - QGIS for mapping if needed (e.g., show municipalities with both illegal deforestation and slave labor cases). Mas talvez não necessário, pode ser mais textual/analítico.

Passo a passo da apuração:

1. Obter a Lista Suja: Baixar o CSV atualizado do Cadastro de Empregadores – Trabalho Escravo. Compreender colunas: deve ter nome, CPF/CNPJ, município, UF, data de inclusão. Filtrar para casos na Amazônia Legal (ou nacional, dependendo do foco).
2. Extrair CNPJs e nomes: Separar os registros de pessoas jurídicas (CNPJ) e pessoas físicas. Pessoas jurídicas (fazendas, empresas) são mais fáceis de cruzar com autuações do Ibama, pois Ibama costuma multar o CNPJ da fazenda/empresa.
3. Cruzamento com autuações do Ibama: Usando pandas, carregar autos de infração do Ibama. Se houver campo de CPF/CNPJ do autuado, ótimo – fazer merge com o CNPJ da lista suja. Se não, usar nome: normalizar ambos (caixa alta, remover acentos). Fazer um inner join ou loop comparando cada nome da lista suja com autuados do Ibama no mesmo estado (para evitar coincidir nomes iguais de estados diferentes). Isso pode achar alguns matches exatos. Para ampliar, usar fuzzy matching

para nomes que são parecidos (OpenRefine cluster pode ajudar, ou Python fuzzy). Identificar, digamos, uma dúzia de coincidências interessantes. Exemplo fictício: “Agropecuária ABC Ltda” está na lista suja (usou trabalho escravo) e nos autos do Ibama (multada por desmatamento).

4. Análise dos resultados: Para cada match encontrado, reunir informações: quantos autos do Ibama aquela empresa teve (e por quais crimes: checar descrição – desmate, queimada, etc.), e quantos trabalhadores foram resgatados (a lista suja não dá número, mas notas do MPT ou notícias sim – pode pesquisar empresa + trabalho escravo).

Assim construímos mini-casos. - Se os matches forem poucos, expandir busca: ver se o município tem correlação – ex: Município de São Félix do Xingu lidera tanto desmate quanto escravos. Ainda que não seja mesma fazenda, vale mencionar contextualmente.

5. Casos para detalhar: Escolher um caso emblemático encontrado. Ex: Fazenda X, no Pará, constou na lista suja por escravizar trabalhadores em carvoaria; essa mesma fazenda foi embargada por desmatamento ilegal em APA. Contatar autoridades locais ou sindicato para falar das condições etc. Tentar falar com proprietário (provavelmente não dará retorno, mas tentar).

6. Dados agregados: Além dos casos específicos, ver o macro: “Encontramos Y empregadores presentes tanto na lista de trabalho escravo quanto nos registros de infração ambiental”. Se Y for pequeno ( $< 5$ ), ampliar análise qualitativa. Se for razoável ( $> 10$ ), pode quantificar: setorialmente, eram sobretudo fazendas de gado ou carvoarias etc.

7. Fontes externas: Falar com um procurador do trabalho ou da república que atuou nesses casos integrados, para comentário sobre a “convergência de ilícitos” (há estudos mostrando ligação entre desmatamento e trabalho escravo). Isso adiciona autoridade à pauta.

8. Publicação: Abordagem: “Quadrilha ambiental: fazendeiros que desmatam também escravizam” (claro, suavizar título, mas essa é a ideia). Começar com um caso narrativo (um trabalhador resgatado contando que derrubava floresta e era explorado – se conseguir depoimento de arquivo). Depois mostrar dados: “Levantamento do Reportagem cruzou a lista de trabalho escravo com multas ambientais e identificou X coincidências. Especialistas afirmam que na expansão ilegal da fronteira agrícola, é comum usar mão de obra análoga à escrava para reduzir custos e avançar sobre a floresta.” Inclui também um parágrafo sobre metodologia do cruzamento. Mapas ou gráficos podem ser simples: talvez um diagrama de Venn visualizando a intersecção entre o conjunto de infratores ambientais e infratores trabalhistas (só se tiver número significativo). Listar 2 ou 3 casos com nome e o que fizeram. E claro, pedir resposta do governo ou dos mencionados (Ibama e MTE poderiam destacar ações integradas, etc.). Essa pauta traz um ângulo novo conectando duas agendas normalmente tratadas separadamente, reforçando a gravidade das infrações.