

Universidade Estadual de Campinas - UNICAMP
Instituto de Matemática, Estatística e Computação Científica
Departamento de Estatística

Análise de Áreas com Solo Contaminado em Campinas

Ana Flávia Polisel 157672
Bruna Nascimento Marques 135143
Melissa Pelermo 140830
Welligton Takao Tanaka 093259

Outubro/2017

1 Introdução

O objetivo do trabalho é verificar se o número de áreas com solo contaminado em Campinas tende a aumentar nos próximos anos. Para isso, utilizaremos os dados da CETESB, que contém registros de 2011 a 2016 com informações sobre as áreas contaminadas. Faremos apenas uma análise descritiva dos dados (devido ao curto prazo de entrega do projeto), para começar a entender melhor como eles se comportam.

2 Material

O banco de dados utilizado nesse projeto foi disponibilizado pela CETESB, e contém 832 registros de 2011 a 2016 sobre as áreas com solo contaminado na cidade de Campinas. A base de dados bruta tem informações como endereço, coordenadas geográficas, ramo de atividade, classificação do solo, fonte de contaminação, *status*, ano, contaminante, identificação e outras. Como a base de dados foi enviada em arquivos separados por ano, foi necessário juntar tudo em um arquivo só e a partir disso toda a análise precisou ser feita por ano, uma vez que temos medidas repetidas. Além disso, foi preciso várias correções e agrupamento no banco de dados antes de iniciar o trabalho, pois havia muitos erros de digitação e muitas variáveis tinham mais de dez categorias, inviabilizando a análise.

Para a realização da análise descritiva foi conversado com a cliente e ela selecionou as variáveis de interesse, que poderiam explicar o número de áreas contaminadas. Sendo assim, o banco de dados se reduziu em: bairro (bairros de Campinas); atividade (posto de combustível, indústria, comércio, resíduo); ano (2011 a 2016); fonte_cont (fontes de contaminação).

Foi criada a variável Região Central, pois tínhamos muitos bairros e a cliente sugeriu agrupá-los em duas categorias, uma vez que ela acredita que tenha diferença entre elas. Consideramos os seguintes bairros como central: Botafogo, Cambuí, Centro, Guanabara, Vila Industrial, Vila Itapura, e como não central os demais, totalizando 86 bairros registrados. Adicionalmente, a cliente disponibilizou somente o número de registros de 2002 a 2016, sem informações detalhadas sobre cada registro.

3 Análise Descritiva

A análise descritiva é a fase inicial do estudo dos dados coletados e tem como objetivo começar a entender o comportamento dos dados. Pela Figura 1, vemos que o número de registros de áreas contaminadas em Campinas vem aumentando ano após ano.

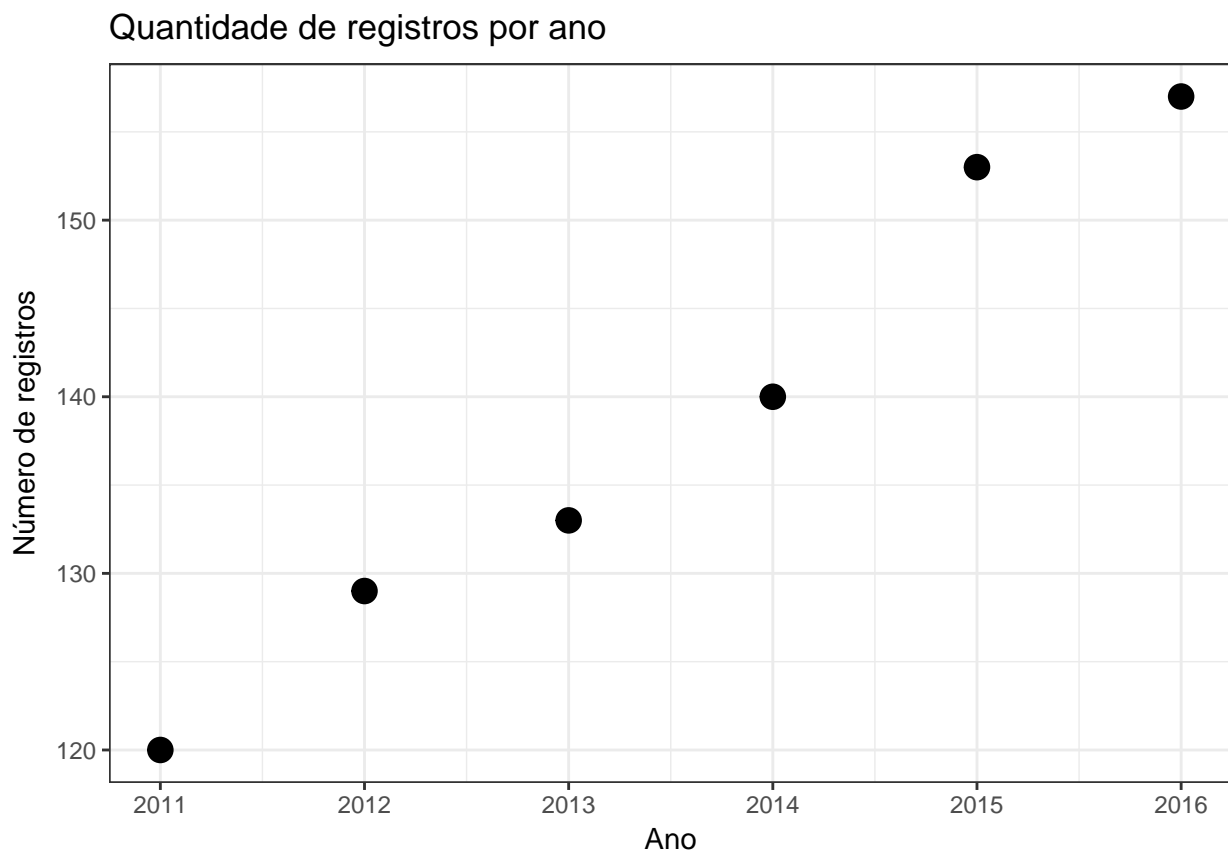


Figura 1: Gráfico de dispersão em que no eixo x temos o ano e no eixo y o número de registros. Note que o número de registros de áreas contaminadas em Campinas vem aumentando ano a ano.

Temos um total de 832 registros em 6 anos, note pela Figura 2 que a maioria desses registros são postos de combustível.

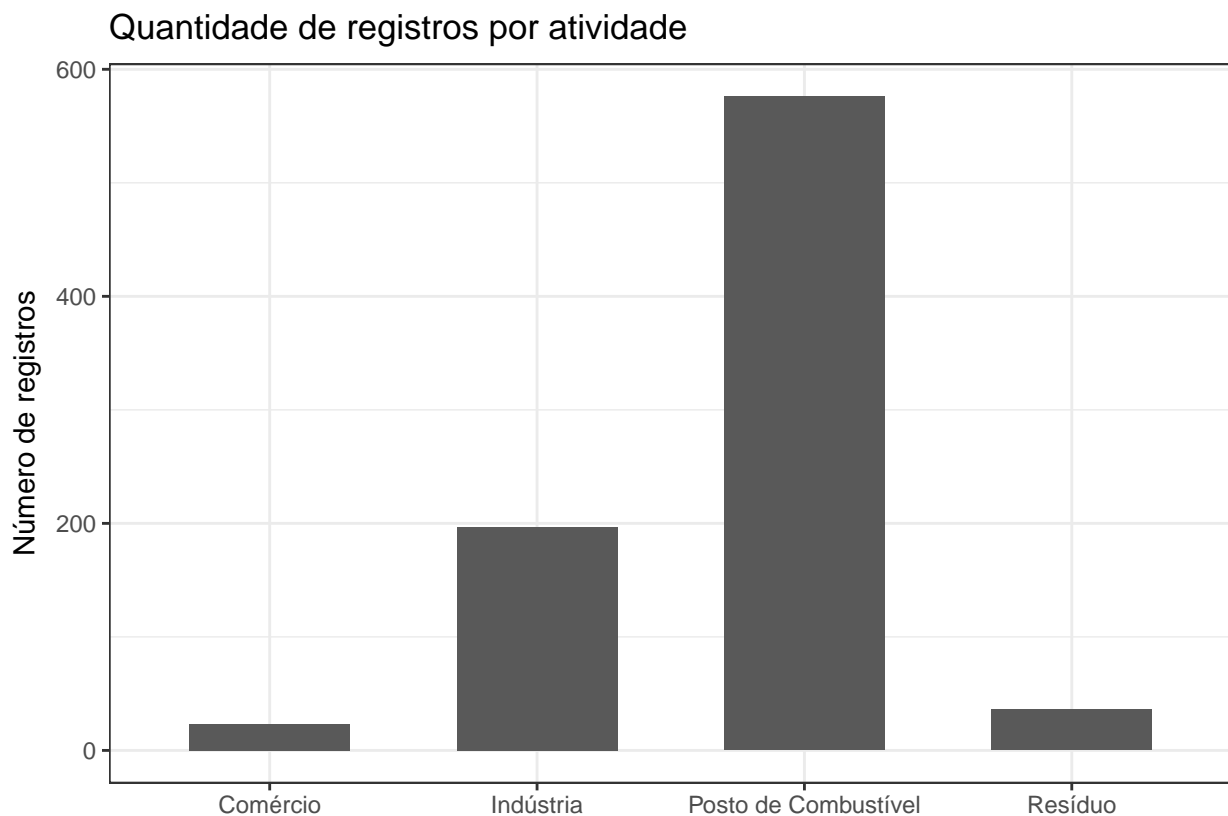


Figura 2: Quantidade de áreas contaminadas registradas nos seis anos, por atividade. O maior número de registros foram em postos de combustível, seguido de indústria.

De acordo com a Figura 3, o número de registros em postos de combustível aumenta todo ano. Porém, vale ressaltar que na indústria aumentou mais que nas outras atividades em 2013 para 2015. Isso causou uma mudança de patamar, uma vez que de 2011 a 2013 o número de registros estava em torno de 25 e passou para mais de 40 em 2015. Atenção também para o aumento de registros nos comércios.

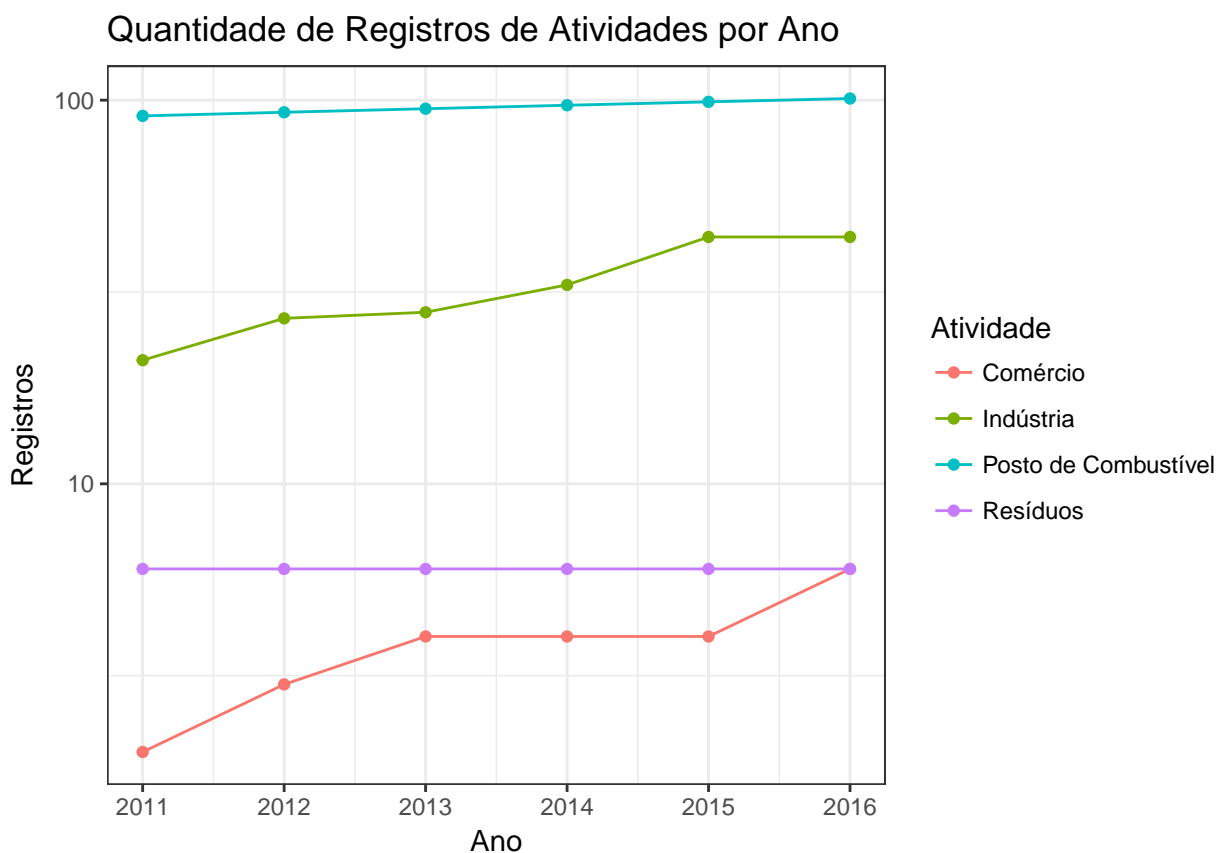


Figura 3: Quantidade de registros por atividade para cada ano. Note que o número de registros em Postos de combustível aumenta todo ano, porém atenção para o número de registros da indústria que teve um salto de 2013 para 2015.

Tabela 1: Porcentagem de registros nas regiões central e não central por ano. Observamos que proporção de registros por região variou muito pouco ao longo dos anos.

	Central(%)	Não Central(%)
2011	21	79
2012	20	80
2013	19	81
2014	19	81
2015	17	83
2016	17	83

Na Figura 4, vemos que a fonte de contaminação por armazenagem e outros é a que mais se destaca, pois possui maior número de registros, além de crescer conforme os anos. Para tratamento de efluentes, ocorreu apenas um registro em 2016 em uma indústria. Para fonte de contaminação por produção e outros houve um crescimento no número de registros a partir de 2013.

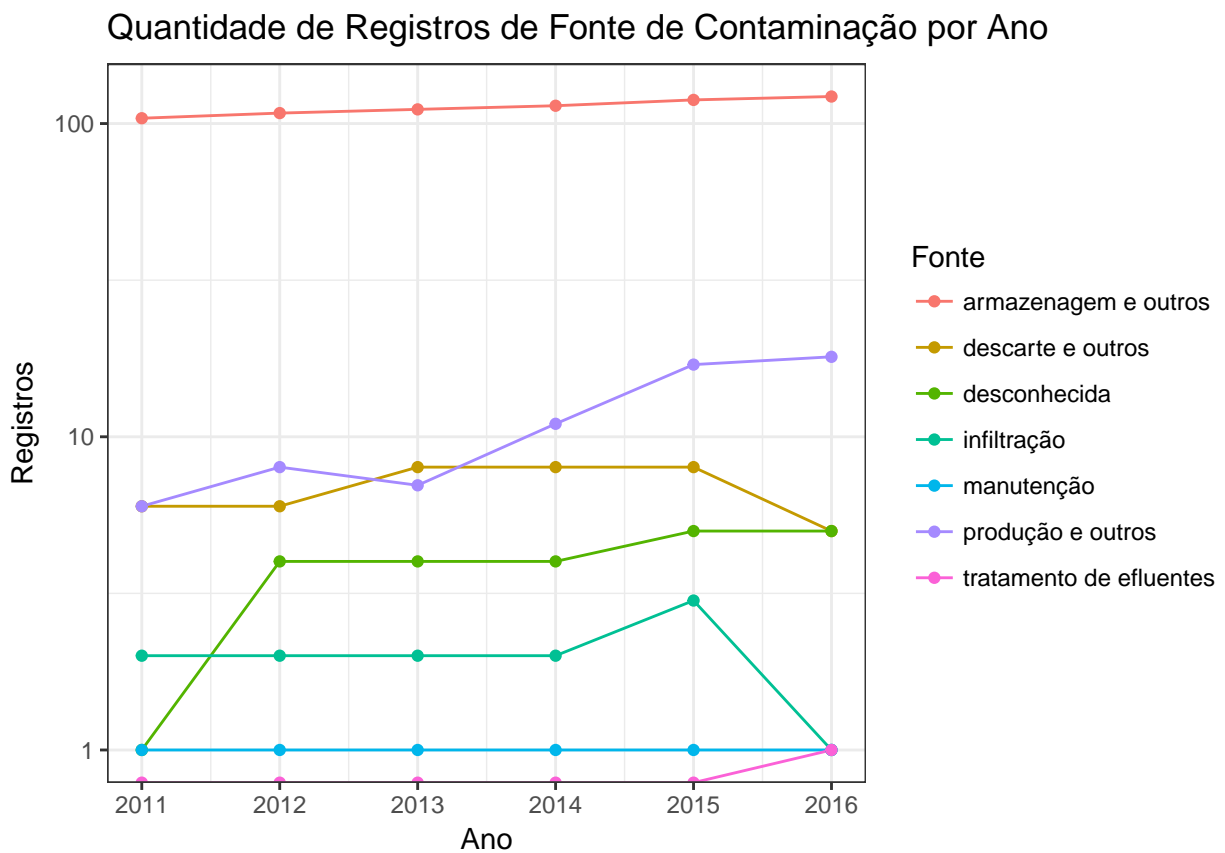


Figura 4: Quantidade de registros por fonte de contaminação, por ano. Armazenagem e outros é a fonte de comunicação que mais se destaca e o número de registros tem aumentado a cada ano.

O gráfico de tendência é simplesmente um gráfico de dispersão com as observações ligadas por retas. É uma ferramenta utilizada para estudar a variação de um processo ao longo do tempo. Na Figura 5 vemos que o número de ocorrências vem crescendo desde 2002, porém como o registro só é realizado se a área está contaminada, não temos o total de áreas, assim não é possível inferir sobre a tendência dos próximos anos.

Ressaltando que para fazer uma previsão sobre o número de áreas contaminadas, precisamos de mais informações, como por exemplo registros de locais **não** contaminados, pois assim é possível avaliar as características que podem estar relacionadas com o aumento da chance de contaminação nos próximos anos.



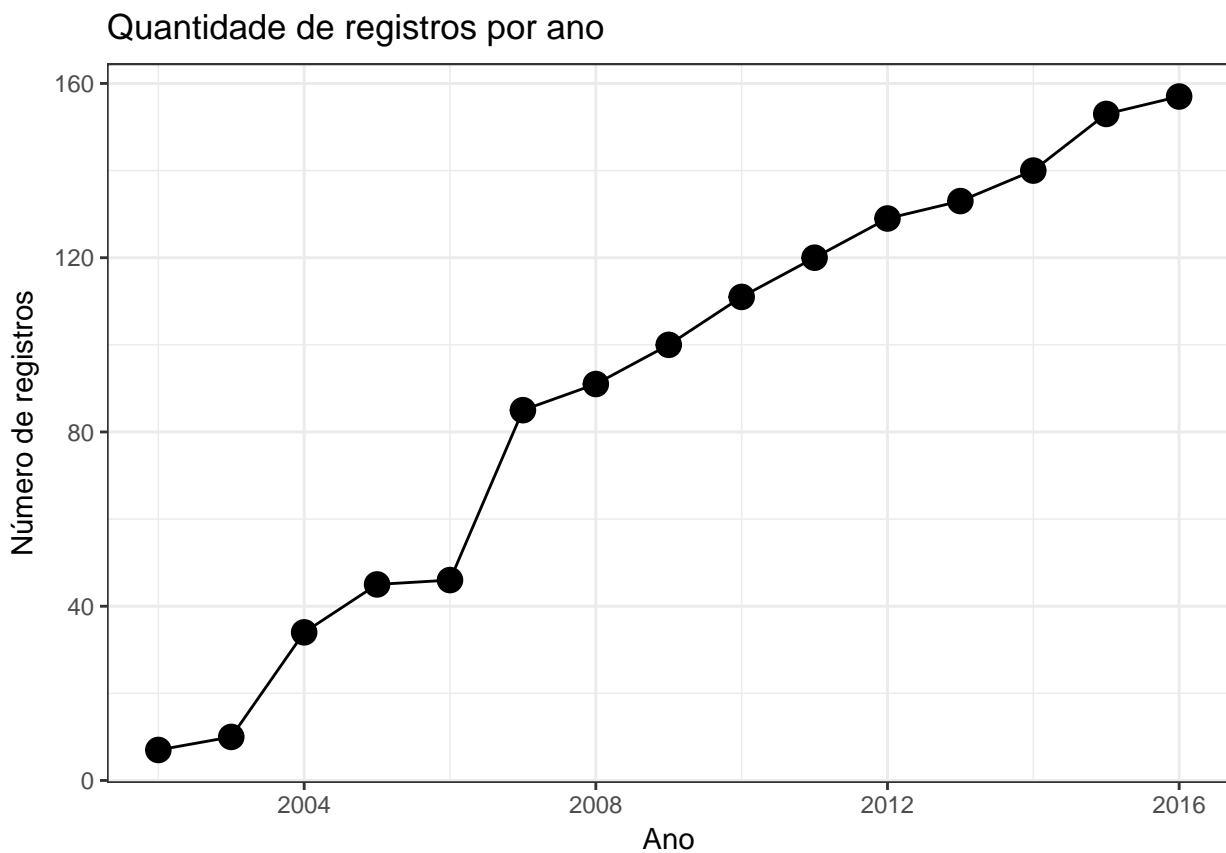


Figura 5: Gráfico de tendência que mostra o aumento no número de registros desde 2002, complementando a Figura 1.

4 Conclusão

Notamos que o número de registros é maior em postos de combustível e tivemos um crescimento notável para o número de registros em indústrias também. Temos um número muito maior de registros de áreas contaminadas por armazenagem do que por outras fontes, nota-se também um salto no número de registros cuja fonte de contaminação é produção e outros em 2014.

Para cada ano, a proporção de registros na região não central é maior do que na região central, mas não houve aumento considerável ao longo do tempo, vale lembrar que Campinas é uma cidade urbanizada e que isso não necessariamente significa que a taxa de ocorrência seja menor no centro, pois não sabemos o número total de estabelecimentos.

Por meio das análises descritivas, temos indícios que o número de registros de áreas contaminadas