

mgcvUI: uma interface interativa para o pacote mgcv (modelos aditivos generalizados)

Tradução para o português do Brasil, elaborada com assistência de IA e revisada pela redação; a

versão em inglês é a autorizada.

William Bert Craytor

Volume 1, Número 1 · Verão de 2026

Publicado em 1 de julho de 2026

DOI: [10.67330/xtv1pa25](https://doi.org/10.67330/xtv1pa25)

© 2026 William Bert Craytor. Published by Pacific Vista Net.

Este artigo de acesso aberto é distribuído sob os termos da [licença Creative Commons Atribuição 4.0](#).

Resumo. mgcvUI é uma interface gráfica de usuário para o pacote de R mgcv, que ajusta modelos aditivos generalizados (GAM) por meio de splines de regressão penalizada com seleção automática do grau de suavização. Oferece três modos de uso — modelagem preditiva geral, avaliação de imóveis e análise de região de mercado — e conduz o usuário pela importação de dados, pela especificação de termos suaves, pelo ajuste do modelo, pelos gráficos de diagnóstico e de efeitos e pelos relatórios passíveis de download. Este artigo documenta os requisitos de formato de dados do mgcvUI, o fluxo de trabalho de modelagem, as telas de saída e a referência completa de funcionalidades.

Palavras-chave: modelos aditivos generalizados, GAM, splines penalizadas, suavização, mgcv, R, interface gráfica de usuário, avaliação de imóveis

Nota do editor (julho de 2026): Pouco antes da publicação, o mgcvUI foi consolidado dentro do earthUI (versão 0.11.0 e posteriores), que agora oferece o fluxo de trabalho de GAM aqui descrito ao lado dos fluxos de trabalho de MARS e de elastic net. O mgcvUI continua disponível em sua forma final publicada (0.3.0) em <https://github.com/wcraytor/mgcvUI>; o desenvolvimento futuro prossegue no earthUI. Este artigo se refere ao mgcvUI tal como foi publicado.

Sumário

1	Introdução	4
1.1	Histórico e antecedentes	4
1.1.1	Modelos aditivos generalizados	4
1.1.2	O pacote mgcv	4
1.1.3	Comparação com MARS (earth) e elastic net (glmnet)	5
1.1.4	O fluxo earthUI-mgcvUI	5
1.1.5	Fluxo de trabalho recomendado	6
1.2	O que é o mgcvUI?	6
1.3	Três modos de uso	6
1.4	Recursos específicos do setor imobiliário	7
1.5	Primeiros passos	7
2	Requisitos dos dados de entrada do MLS	9
2.1	Formato e estrutura do arquivo	9
2.2	Colunas necessárias no modo de avaliação	9
2.3	Convenções especiais de nomenclatura de colunas	10
2.4	Qualidade e completude dos dados	10
2.5	Posicionamento da linha do avaliando	10

3	Modo General	11
3.1	Visão geral	11
3.2	O fluxo de trabalho da barra lateral	11
3.3	Importar do earthUI (opcional)	12
3.4	Abas do painel principal	12
3.5	Persistência das configurações	13
3.6	Modo escuro	13
3.7	Localidade e configurações regionais	13
4	Modo Appraisal	15
4.1	Tratamento da linha do avaliando	15
4.2	Data de referência e idade da venda	15
4.3	Designações especiais de colunas	15
4.4	Visão geral dos ajustes de RCA	15
5	Modo Market Area Analysis	16
5.1	Diferenças em relação ao modo Appraisal	16
5.2	Quando usar o modo de mercado	16
6	Seleção de variáveis	17
6.1	Variável alvo	17
6.2	Transformação da resposta	17
6.3	A tabela de preditores	17
6.4	Detecção de tipos de dados e substituições	18
6.5	Referência dos tipos especiais de coluna	18
7	Seleção de parâmetros	19
7.1	Predefinições de parâmetros	19
7.2	Family	19
7.3	Method	19
7.4	Gamma (multiplicador de suavização)	20
7.5	Validação cruzada	20
7.6	Select (encolhimento)	20
7.7	Base padrão	20
7.8	k padrão (dimensão da base)	20
7.9	Tipo de tensor	21
7.10	Interações permitidas	21
7.11	Interações fator por termo suave	21
7.12	Parâmetros avançados	21
8	Ajuste do modelo	23
8.1	O botão de ajuste	23
8.2	O que acontece durante o ajuste	23
8.3	Ajuste em segundo plano e progresso	23
9	Abas de resultados	24
9.1	Data	24
9.2	Equation	24
9.3	Summary	24
9.4	Variable Importance	24
9.5	Gráficos de contribuição	25
9.5.1	Termos suaves univariados — $s(x)$	25

9.5.2	Termos fator por suave — $s(x, by = factor)$	25
9.5.3	Termos de interação — $ti(x_1, x_2)$ e $te(x_1, x_2)$	25
9.5.4	Termos paramétricos — lineares e de fator	25
9.6	Correlation	25
9.7	Diagnostics	26
9.8	RCA Adjustments	26
9.9	Anova	26
9.10	Mgev Output	26
9.11	Sign Check	26
9.12	Concurvity	26
9.13	Basis Check	27
10	Download dos dados	28
10.1	Colunas de saída	28
10.2	Escores de CQA	28
11	Cálculos de RCA e download	29
11.1	Abertura da caixa de diálogo do RCA	29
11.2	Interpolação do escore de CQA	29
11.3	Colunas de saída	29
12	Grade de comparação de vendas	30
12.1	Caixa de diálogo de seleção de comparáveis	30
12.2	Estrutura da pasta de trabalho	30
12.3	Proteção da planilha e células residuais	30
13	Download de relatórios	31
13.1	Etapas 1: gerar o relatório Quarto	31
13.2	Etapas 2: converter o relatório Quarto	31
13.3	Conteúdo do relatório	31
14	Exportação das funções do modelo	33
14.1	Funções em R	33
14.2	Funções em Python	33
14.3	Cabeçalho C++	33
14.4	Banco de dados SQLite	33
15	Comparação com earthUI e glmnetUI	34
15.1	Principais diferenças	34
15.2	Recursos compartilhados	34
15.3	Fluxo de trabalho recomendado	34
16	Conjunto de dados de demonstração: Appraisal_1.csv	36
16.1	Descrição	36
16.2	Colunas	36
16.3	Roteiro rápido sugerido	37
17	Requisitos de sistema e solução de problemas	38
17.1	Plataformas suportadas	38
17.2	Versão do R	38
17.3	Pacotes de R necessários	38
17.4	Dependências opcionais	39
17.5	Observações por plataforma	39

17.5.1 macOS	39
17.5.2 Windows	39
17.5.3 Linux	39
17.6 Degradação graciosa	40
17.7 Solução de problemas	40
17.7.1 “PDF option is not available”	40
17.7.2 “Settings will not persist”	40
17.7.3 “Diagnostics plot unavailable”	41
17.7.4 “Report error: number of items to replace...”	41
17.7.5 Porta 7880 já em uso	41
17.7.6 Falha na instalação de pacotes no Linux	41

1 Introdução

1.1 Histórico e antecedentes

1.1.1 Modelos aditivos generalizados

Os **modelos aditivos generalizados** (GAM) foram formalizados por Trevor Hastie e Robert Tibshirani em sua monografia de 1990. Os GAM estendem o modelo linear substituindo cada termo linear $\beta_j x_j$ por uma função suave $f_j(x_j)$, de modo que o modelo fica assim:

$$y = \beta_0 + f_1(x_1) + f_2(x_2) + \cdots + f_p(x_p) + \varepsilon$$

Cada f_j é estimada a partir dos dados por meio de uma spline de regressão penalizada. As funções suaves permitem que o modelo capte relações não lineares sem que o usuário precise especificar de antemão uma forma funcional. A penalização controla a ondulação de cada termo suave, evitando o sobreajuste sem deixar de permitir a flexibilidade necessária para acompanhar os padrões reais.

1.1.2 O pacote mgcv

O pacote de R `mgcv` (Mixed GAM Computation Vehicle) foi desenvolvido por Simon Wood na Universidade de Bath (Wood 2017). Lançado pela primeira vez em 2001, é hoje a implementação padrão dos GAM em R e acompanha toda instalação de R. Seus métodos subjacentes são descritos em uma série de artigos que abrangem a estimação estável de múltiplos parâmetros de suavização (Wood 2004), as splines de regressão thin-plate (Wood 2003), a estimação REML rápida e estável (Wood 2011) e a seleção geral de modelos suaves (Wood, Pya e Säfken 2016). Entre seus recursos principais estão:

- **Seleção automática do grau de suavização:** o parâmetro de suavização de cada termo é estimado como parte do ajuste do modelo, por meio de REML (máxima verossimilhança restrita), GCV (validação cruzada generalizada) ou ML (máxima verossimilhança).
- **Múltiplos tipos de base:** splines de regressão thin-plate (tp), splines de regressão cúbica (cr), P-splines (ps) e B-splines (bs), cada uma com propriedades distintas.
- **Termos suaves de produto tensorial:** `te()` e `ti()` permitem interações suaves entre variáveis sem incorrer na maldição da dimensionalidade.
- **Seleção de variáveis:** a opção `select = TRUE` acrescenta uma penalização adicional que pode encolher termos suaves até zero, realizando de fato uma seleção de variáveis.
- **Diagnósticos de concurvidade:** medem o análogo da multicolinearidade para os termos suaves.

1.1.3 Comparação com MARS (earth) e elastic net (glmnet)

O aplicativo earth e seus dois módulos de pós-processamento (glmnet e mgcv) usam motores de modelagem diferentes. A tabela a seguir resume as principais diferenças:

Característica	mgcv (GAM)	earth (MARS)	glmnet (elastic net)
Tipo de modelo	Não linear suave (splines penalizadas)	Linear por partes (splines adaptativas)	Linear (com regularização)
Não linearidade	Automática, por funções suaves (thin plate, cúbica etc.)	Automática, por funções de dobradiça com nós determinados pelos dados	Somente por meio de interações ou expansão de base
Seleção de variáveis	Opcional, por encolhimento de dupla penalização (<code>select=TRUE</code>)	Nativa, por passos progressivos/regressivos com poda por GCV	Nativa, pela penalização L_1 (Lasso)
Interações	Termos suaves de produto tensorial (<code>te()</code> , <code>ti()</code>)	Produtos baseados em dobradiças (interpretáveis, até grau 3)	Produtos cruzados aos pares (difíceis de interpretar)
Interpretabilidade	Curvas suaves de efeito parcial; muito intuitivas	As funções g mostram efeitos parciais lineares por partes para cada variável	Os coeficientes são pesos lineares; fáceis de explicar individualmente
Mais indicado para	Relações não lineares suaves, modelagem flexível	Deteção automática de não linearidade e de interações	Dados de alta dimensão, seleção de variáveis, amostras pequenas
Ajustes de RCA	Ajustes suaves por efeitos parciais	Ajustes lineares por partes por meio das funções g	Ajustes lineares por variável
Controle do sobreajuste	Estimação do parâmetro de suavização por REML/GCV	Poda baseada em GCV	Lambda validado por validação cruzada

O **MARS** (Multivariate Adaptive Regression Splines) foi apresentado por Jerome Friedman em 1991. Ele constrói modelos lineares por partes selecionando de forma adaptativa funções de dobradiça e as posições de seus nós a partir dos dados. A implementação em R é o pacote **earth**, de Stephen Milborrow.

A regressão **elastic net**, implementada no pacote **glmnet** por Friedman, Hastie e Tibshirani (2010), combina as penalizações Lasso (L_1) e ridge (L_2) para realizar simultaneamente a seleção de variáveis e o encolhimento dos coeficientes.

1.1.4 O fluxo earthUI–mgcvUI

Um fluxo de trabalho poderoso combina a descoberta automática de nós do earthUI com a estimação suave do mgcvUI. Quando o earthUI exporta um arquivo de resultados `.rds` e o mgcvUI o importa:

1. **O earth descobre os nós** — o MARS encontra posições de dobradiça (pontos de mudança) determinadas pelos dados para cada variável.

2. **O mgcvUI as usa como âncoras** — os nós do earth passam a ser a base para as splines de regressão cúbica (base `cr`) no GAM, dando aos termos suaves uma vantagem inicial, com posições de nós que refletem a estrutura real dos dados.
3. **O mgcv refina o ajuste** — a estimação por splines penalizadas suaviza o ajuste linear por partes do earth, produzindo curvas suaves de efeito parcial e preservando a forma geral descoberta pelo earth.

Esse fluxo faz a ponte entre a modelagem exploratória (earth) e a modelagem confirmatória (GAM), reunindo o melhor das duas abordagens.

1.1.5 Fluxo de trabalho recomendado

Para avaliação de imóveis e aplicações semelhantes que exigem modelos interpretáveis e defensáveis:

1. **Comece pelo earthUI** para descobrir as variáveis importantes, as relações não lineares e as interações presentes em seus dados. A seleção automática de base e as funções `g` do earth fornecem excelentes indicações iniciais.
2. **Refine com o mgcvUI** se quiser efeitos parciais mais suaves. Importe os nós do earthUI para o mgcvUI e faça uma transição contínua de modelos lineares por partes para modelos suaves.
3. **Use o glmnetUI** quando precisar de uma seleção de variáveis agressiva (Lasso), quando o número de preditores se aproximar ou ultrapassar o número de observações, ou quando quiser um modelo puramente linear com regularização, em nome da defensabilidade.

1.2 O que é o mgcvUI?

O mgcvUI é uma interface gráfica de usuário para o pacote de R `mgcv`. Ele é executado como um aplicativo Shiny local — não há login, não há servidor e não há contas. Você o inicia a partir do R, importa um conjunto de dados (CSV ou Excel), configura seu modelo e ajusta um GAM de forma interativa.

Na versão atual, o mgcvUI já não é um pacote instalado separadamente: ele é distribuído *dentro* do pacote earthUI, como seu módulo de pós-processamento de GAM, iniciado com `earthUI::launch_mgcv()`. O aplicativo em si — e tudo o que se descreve neste artigo — permanece inalterado; apenas a instalação e o comando de início são diferentes. As três rotinas (earth, glmnet, mgcv) compartilham uma mesma árvore de projetos, um mesmo banco de dados de configurações e uma mesma cópia da infraestrutura de apoio.

O aplicativo oferece um fluxo de trabalho completo: importação de dados, configuração de variáveis com especificação de termos suaves, ajuste do modelo com processamento em segundo plano, gráficos de diagnóstico, curvas suaves de efeito parcial e relatórios passíveis de download em formato Word, PDF ou HTML.

Os **modelos aditivos generalizados** substituem a relação linear βx por uma função suave $f(x)$ para cada preditor. Isso significa:

- **Não é preciso especificar formas funcionais** — os dados determinam a forma de cada relação.
- **Efeitos parciais suaves** — a contribuição de cada preditor é uma curva suave, e não uma reta, o que torna fácil ver como uma variável afeta a resposta em diferentes valores.
- **Seleção automática do grau de suavização** — o mgcv determina o quanto cada curva deve ondular, por REML ou GCV, equilibrando ajuste e sobreajuste.

1.3 Três modos de uso

Todo projeto tem um **purpose** (propósito ou modo de uso), escolhido ao criá-lo na Seção 1 da barra lateral (**Project**) e fixado a partir de então. O propósito do projeto ativo determina quais

ferramentas e elementos da interface aparecem; para trabalhar em outro modo, crie ou abra um projeto com esse propósito. Os três modos são:

- **General** (geral) — regressão GAM para qualquer tipo de população ou conjunto de dados. Esse é o modo padrão. Oferece o fluxo de trabalho completo de GAM sem acréscimos específicos de nenhum domínio.
- **For Appraisal** (para avaliação) — regressão GAM adaptada à avaliação de imóveis. Acrescenta recursos específicos para a avaliação de um único imóvel, incluindo o tratamento do imóvel avaliando, as designações especiais de colunas e o Residual Constraint Approach (RCA).
- **Market Area Analysis** (análise de região de mercado) — regressão GAM adaptada aos estudos de região de mercado. Acrescenta recursos para analisar grupos de imóveis dentro de um mercado definido, com tratamento opcional da linha do avaliando.

Nos três modos o motor de modelagem subjacente é idêntico — você sempre ajusta um GAM por meio de `mgcv::gam()`. A definição do propósito controla quais ferramentas e elementos de interface adicionais ficam disponíveis.

1.4 Recursos específicos do setor imobiliário

Quando se seleciona **For Appraisal** ou **Market Area Analysis**, o `mgcvUI` ativa vários recursos concebidos para a análise imobiliária:

- **Designações especiais de colunas** — cada preditor pode ser marcado com um papel especial, como `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions` ou `display_only`. Essas designações controlam como a coluna é tratada durante o ajuste e nas saídas.
- **Coluna Sale Age** — quando uma coluna é designada como `contract_date` e é informada uma data de referência, o `mgcvUI` calcula uma coluna `sale_age` (dias entre a data da venda e a data de referência) e a substitui como preditor.
- **Cálculos de RCA (somente em avaliação)** — no modo de avaliação, após ajustar o modelo, o `mgcvUI` pode calcular a saída do Residual Constraint Approach (RCA), que produz ajustes por comparável, resumos de ajuste líquido/bruto e um preço de venda ajustado para o imóvel avaliando.

***Dica:** O modo General funciona com qualquer conjunto de dados — financeiro, científico, de engenharia ou imobiliário. Os modos de avaliação e de mercado apenas acrescentam recursos de conveniência para os profissionais do setor imobiliário.*

1.5 Primeiros passos

Para usar o `mgcvUI`:

1. **Instale o earthUI no R:** `install.packages("earthUI")` — a rotina `mgcv` está incluída. (Veja “Installing earthUI and vProlog” no artigo sobre o `earthUI` nesta edição para a versão de desenvolvimento e os componentes opcionais.)
2. **Inicie o aplicativo:** execute `earthUI::launch_mgcv()` no R. O aplicativo abre em seu navegador na porta 7880.
3. **Crie ou abra um projeto** na Seção 1 da barra lateral. Um projeto define o propósito (General, For Appraisal ou Market Area Analysis) e o local de trabalho (país, estado, município, cidade), além de suas pastas de entrada e saída. Os projetos são armazenados sob uma raiz `regProj` compartilhada e são compartilhados com os aplicativos irmãos `earthUI` e `glmnetUI`. Abrir um projeto com outro propósito é a forma de alternar entre os modos.

4. **Importe seus dados** na Seção 2, selecionando um arquivo CSV ou Excel da pasta de entrada do projeto ativo. Acrescente arquivos a essa pasta e clique em **Refresh** para listá-los.
5. **Importe do earthUI (opcional)** — a Seção 3 permite importar um arquivo `.rds` de resultados do earthUI para semear o GAM com os nós descobertos pelo earth. Pule esta etapa se você não tiver um resultado do earthUI.
6. **Configure as variáveis** — escolha o alvo e os preditores, defina os tipos de dados, atribua eventuais papéis especiais de coluna e marque as variáveis como fator ou linear.
7. **Defina os parâmetros do mgcv** — família, método, gamma, tipo de base, k, tipo de tensor, interações e demais opções.
8. **Ajuste o modelo** — clique em “Fit Mgcvc GAM Model” e examine os resultados no painel principal.
9. **Exporte** — baixe as predições em Excel, gere e converta um relatório Quarto ou (no modo de avaliação) calcule os ajustes de RCA e uma grade de vendas.

As configurações são automaticamente persistidas em um banco de dados SQLite e restauradas quando você recarrega o mesmo arquivo de entrada.

2 Requisitos dos dados de entrada do MLS

Para os fluxos de trabalho de avaliação de imóveis e de análise de mercado, seus dados de entrada normalmente vêm de uma exportação de um Multiple Listing Service (MLS). Este capítulo descreve a estrutura de arquivo esperada e as colunas que o mgcvUI é capaz de utilizar.

2.1 Formato e estrutura do arquivo

O mgcvUI aceita arquivos **CSV** e **Excel** (.xlsx, .xls). Na importação, os nomes das colunas são automaticamente convertidos para **snake_case** — por exemplo, “Living SqFt” passa a **living_sqft**, “Contract Date” passa a **contract_date** e “Sale Price” passa a **sale_price**. Essa normalização garante referências de coluna consistentes ao longo de todo o fluxo de trabalho. O separador de CSV e o marcador decimal usados na importação são determinados pelas configurações de localidade (veja o Capítulo 3, “Localidade e configurações regionais”).

Seu arquivo de dados deve ser uma tabela plana, com uma linha por imóvel e uma coluna por atributo. A primeira linha do arquivo deve conter os cabeçalhos das colunas.

2.2 Colunas necessárias no modo de avaliação

Embora o mgcvUI funcione com qualquer conjunto de colunas, o fluxo de trabalho completo de avaliação se beneficia da presença das seguintes colunas:

Coluna	Tipo especial	Finalidade
Sale Price	(alvo)	Variável resposta do modelo
Contract Date	contract_date	Usada para calcular sale_age (dias desde a data de referência)
Listing Date	listing_date	Usada com a data do contrato para calcular o DOM caso não exista coluna de DOM
Days on Market	dom	Dias no mercado; exibida nas exportações
Concessions	concessions	Concessões na venda; SP líquido = preço de venda – concessões
Living Area (SF)	living_area	Habilita resíduos por SF (residual_sf , cqa_sf)
Lot Size	lot_size	Coluna de tamanho do terreno
Site Dimensions	site_dimensions	Agrupada com o tamanho do terreno
Latitude	latitude	Arredondada para 3 casas decimais
Longitude	longitude	Arredondada para 3 casas decimais
Area ID	area	Identificador de região de mercado / bairro
Actual Age	actual_age	Idade do imóvel
Effective Age	effective_age	Idade efetiva do imóvel
Address	display_only	Exibida nas exportações; excluída do modelo

Os nomes das colunas da planilha podem estar em outro idioma — os nomes “especiais” estão em inglês para que o programa em R possa lhes dar tratamento especial. No mais, os nomes de coluna informados aparecem nos modelos de regressão, nos gráficos e, em trabalhos de avaliação, nos relatórios de saída.

Nem todas as colunas são obrigatórias. O mgcvUI se adapta — se uma coluna estiver ausente, o recurso correspondente é simplesmente omitido. Ainda assim, para modelos de precificação imobiliária, certas colunas são altamente recomendáveis:

1. **Sale Age** — o número de dias entre a data do contrato de venda e a data de referência. Quando se usa um histórico de vendas de vários anos, **sale_age** costuma desempenhar papel central.

2. **Living Area** — também aparece sob nomes como “Living Sqft” ou “GLA” (gross living area, área bruta habitável).
3. **Total Bath Count** — o número total de banheiros completos, de um quarto, de meio banheiro e de três quartos.
4. **Garage Bays** ou **Garage Area** — vagas de garagem ou área em pés quadrados.
5. **Lot Size** — a área de terreno do imóvel.
6. **Longitude**, **Latitude** e, se disponível, **Area ID**, para a variação geográfica de preços.

2.3 Convenções especiais de nomenclatura de colunas

O mgcvUI identifica as colunas pela **designação de tipo especial**, e não pelo nome da coluna. Você pode dar às suas colunas os nomes que quiser na exportação do MLS — o que importa é atribuir o tipo especial correto na tabela de configuração de variáveis (Capítulo 6).

2.4 Qualidade e completude dos dados

- **Valores ausentes (NA)**: linhas com valores NA em qualquer coluna de preditor ou de alvo são removidas automaticamente antes do ajuste. O mgcvUI restringe os dados apenas às colunas usadas no modelo, de modo que NAs em colunas não utilizadas não causam a exclusão de linhas.
- **Colunas numéricas**: preço de venda, área habitável, tamanho do terreno e campos semelhantes devem ser numéricos.
- **Colunas de fator**: colunas de texto, TRUE/FALSE e de fator do R são tratadas automaticamente como categóricas — só podem entrar no modelo como fatores. Uma coluna **numérica** (por exemplo, um código numérico de região ou de estilo) só se torna fator quando você marca sua caixa de seleção **Factor**; o mgcvUI nunca infere natureza categórica a partir da faixa de valores de uma coluna numérica, de modo que preditores numéricos discretos como `bath_count` permanecem contínuos (suavizados). O papel especial (Special) é independente da decisão sobre fator.

***Dica:** Examine a coluna NA na tabela de preditores após a importação. Colunas com muitos valores ausentes podem provocar a exclusão de linhas. Colunas com $\geq 50\%$ de NAs são marcadas em vermelho e automaticamente excluídas.*

2.5 Posicionamento da linha do avaliando

No modo **Appraisal**, a **linha 1 deve ser o imóvel avaliando**. Todas as demais linhas são vendas comparáveis. A linha do avaliando é excluída do ajuste do modelo. Após o ajuste, o modelo ainda assim gera previsões para a linha do avaliando.

No modo **Market Area Analysis**, colocar o avaliando na linha 1 é opcional.

No modo **General** não há tratamento especial de linhas — todas as linhas são tratadas igualmente.

3 Modo General

3.1 Visão geral

O modo General é o padrão quando você inicia o mgcvUI. Ele oferece o fluxo de trabalho completo de GAM para qualquer conjunto de dados — não apenas imobiliário. Você pode usar o mgcvUI para dados científicos, análise financeira, estudos de engenharia ou qualquer problema de regressão em que se esperem relações não lineares suaves.

No modo General, a interface omite os recursos específicos do setor imobiliário (colunas especiais, idade da venda, RCA). A barra lateral é simplificada para concentrar-se na seleção de variáveis, na configuração de parâmetros, no ajuste do modelo e na exportação.

3.2 O fluxo de trabalho da barra lateral

A barra lateral é organizada em seções numeradas e recolhíveis que o conduzem da seleção do projeto até a exportação do relatório. A maioria das seções aparece somente depois que há um projeto ativo aberto, e várias delas são específicas do propósito do projeto. A pasta de saída deriva do projeto ativo — não há um campo separado de pasta de saída.

1. Project — crie ou abra um projeto. Um projeto define o propósito e o local de trabalho (país, estado, município, cidade), além de suas pastas de entrada (`<os>_in/`) e de saída. Os projetos são armazenados sob uma raiz `regProj` compartilhada (por padrão `~/regProj`, substituível pela variável de ambiente `REGPROJ_ROOT` ou pelo menu suspenso `Settings`) e são compartilhados com os aplicativos irmãos `earthUI` e `glmnetUI`, de modo que os modelos dos três convivem lado a lado. Clique em **Close Project** para trocar.

2. Import Data — selecione um arquivo CSV ou Excel da pasta de entrada do projeto ativo. Acrescente arquivos à pasta pelo Finder/Explorer e clique em **Refresh** para listá-los. Para arquivos Excel com várias planilhas, aparece um seletor de planilha. Os nomes das colunas são automaticamente convertidos para `snake_case`.

3. Import from earthUI (opcional) — importe um arquivo `.rds` de resultados do `earthUI` para semear o GAM com as posições de nós descobertas pelo `earth` (e para semear as seleções de variáveis). Esta etapa é opcional — pule-a se você não tiver um resultado do `earthUI` para importar.

4. Variable Configuration — seletor da variável alvo, transformação da resposta (`none`, `log`, `log10`) e tabela de preditores com caixas de seleção para `Include`, `Factor` e `Linear`. A coluna `Special` aparece somente nos modos `Appraisal` e `Market`. Veja o Capítulo 6 para todos os detalhes.

5. MgcV Call Parameters — toda a configuração do modelo: predefinições de parâmetros, `family`, `method`, `gamma`, validação cruzada, `select`, tipo de base, `k`, tipo de tensor, matriz de interações e parâmetros avançados. Veja o Capítulo 7 para a referência completa de parâmetros.

6. Fit MgcV GAM Model — o botão que executa o modelo. Os resultados aparecem nas abas à direita.

7. Download Output — exporta predições, resíduos, escores de CQA e contribuições por variável como um arquivo Excel para a pasta de saída do projeto. Disponível em todos os modos de uso, depois que o modelo é ajustado.

8. Calculate RCA Adjustments & Download — calcula os ajustes de RCA por variável para cada comparável e interpola o resíduo do avaliando por meio do CQA. Aparece somente nos modos `Appraisal` e `Market`, depois do ajuste.

9. Generate Sales Grid & Download — gera a grade de comparação de vendas, recomendando comparáveis pelo percentual de ajuste bruto. Aparece somente no modo `Appraisal`.

10. Generate Quarto Report — grava um pacote de relatório Quarto (`.qmd`) autocontido — com todos os gráficos e tabelas — na pasta de saída do projeto. Disponível depois que o modelo é ajustado.

11. Convert Quarto Report — renderiza o pacote `.qmd` gerado em HTML, Word ou PDF. O PDF exige uma instalação de LaTeX; se nenhuma for detectada, a opção de PDF fica oculta. Veja o Capítulo 13.

As etapas específicas de cada propósito são omitidas (e as demais etapas renumeradas) nos outros modos: no modo Market a etapa da grade de vendas está ausente e, no modo General, tanto a etapa de RCA quanto a da grade de vendas estão ausentes.

3.3 Importar do earthUI (opcional)

A Seção 3 da barra lateral — **Import from earthUI** — permite importar um arquivo `.rds` de resultados do earthUI. Isso habilita o fluxo earth-mgcv: as posições de nós determinadas pelos dados no earth tornam-se pontos de ancoragem para os termos suaves do GAM.

Quando um resultado do earthUI é importado, o mgcvUI:

- Exibe um resumo com a variável alvo, o R quadrado, o número de termos e as posições de nós por variável.
- Seleciona automaticamente a predefinição de parâmetros “Earth Pipeline” (base de spline de regressão cúbica, $\gamma = 1.4$).
- Define a variável alvo de modo a coincidir com o alvo do earth.
- Marca previamente Include para as variáveis preditoras do earth.
- Marca previamente Factor para as variáveis categóricas do earth e Linear para os preditores lineares do earth.
- Preenche previamente as interações detectadas pelo earth (destacadas em amarelo na matriz de interações e bloqueadas contra alterações manuais).
- Oferece um botão **Export Knots CSV** para baixar os nós do earth como arquivo CSV.

A tabela Variable Configuration passa então a exibir uma nota explicando que as seleções de variáveis (Include, Linear, Factor) e os nós das splines foram *semeados* a partir do modelo earth. Diferentemente do glmnetUI, em que o ajuste glmnet fica atrelado à base do earth, o mgcvUI ajusta o GAM a partir de suas seleções atuais — o earth apenas dá uma vantagem inicial, e todos os controles permanecem editáveis.

Um botão **Clear** remove os dados importados do earth e restaura o modo autônomo.

3.4 Abas do painel principal

Após a importação dos dados, o painel principal oferece as seguintes abas (as abas que dependem do modelo são preenchidas depois do ajuste):

Aba	Conteúdo
Data	DataTable interativa dos dados importados. Nos modos de avaliação e de mercado, dividida em tabelas de imóvel avaliando e de vendas comparáveis.
Equation	Fórmula completa do GAM renderizada em MathJax, mais uma tabela com as definições das funções suaves.
Summary	Métricas principais (R^2 , R^2 de VC, desviância explicada, AIC, n, número de termos suaves) e tabelas de termos suaves e paramétricos.
Variable Importance	Gráfico de barras horizontais das estatísticas F (termos suaves) e dos valores $ t $ (paramétricos), mais tabela ordenada.
Contribution	Curvas suaves de efeito parcial interativas em plotly, com faixas de confiança e marcadores dos nós do earth.
Correlation	Mapa de calor das correlações entre preditores numéricos (disponível antes do ajuste).
Diagnostics	Resíduos versus ajustados, gráfico Q-Q, histograma, resposta versus ajustados (via gratia). Dispersão de observados versus preditos.
RCA Adjustments	Histogramas da análise de RCA (modo de avaliação/mercado, após o cálculo).
Anova	Tabela de ANOVA combinando termos paramétricos e suaves.
Mgcv Output	Texto bruto: tempos de execução, fórmula, resumo do modelo, família, concurvidade.
Sign Check	Verificação da consistência de direção entre earth e GAM (quando há importação do earth).
Concurvity	Matrizes de concurvidade global e aos pares, com codificação por cores.
Basis Check	Testes de adequação da dimensão da base por <code>gam.check()</code> .

3.5 Persistência das configurações

O mgcvUI salva automaticamente sua configuração em um banco de dados SQLite, indexado pelo nome do arquivo de entrada. Quando você recarrega o mesmo arquivo, todas as configurações são restauradas: seleção do alvo, caixas de seleção dos preditores, tipos de dados, parâmetros do mgcv e matriz de interações. O propósito é uma propriedade do projeto ativo, e não uma configuração global, e o último arquivo de entrada utilizado é lembrado por projeto e carregado automaticamente quando o projeto é reaberto.

3.6 Modo escuro

Clique no ícone de lua/sol no canto superior direito para alternar entre os temas Nord Light e Nord Dark. A preferência de tema é salva no localStorage e persiste entre as sessões. Todos os elementos da interface se adaptam ao tema selecionado.

3.7 Localidade e configurações regionais

O mgcvUI dá suporte a convenções internacionais de formatação de números e de CSV por meio de um sistema de localidades baseado em país. O menu suspenso **Settings** na barra de título oferece os seletores **Country** e **Paper** para mais de 30 países. Cada predefinição configura:

- **Separador de CSV** — vírgula (,) para EUA/Reino Unido/Japão ou ponto e vírgula (;) para a maior parte da Europa.
- **Marcador decimal** — ponto (.) ou vírgula (,).
- **Separador de milhares** — vírgula (EUA/Reino Unido/Japão), ponto (Alemanha/Itália/Espanha), espaço (Finlândia/França/Polônia/países bálticos) ou apóstrofo (Suíça).

- **Tamanho do papel** — Letter (EUA/Canadá/México) ou A4 (nos demais países).

Clique em **Save** para gravar suas preferências de localidade globalmente. A seção Import Data também tem um seletor **Locale** por arquivo, para importações pontuais que usem outras convenções. O mesmo menu suspenso Settings contém o campo **regProj root folder**, que realoca a árvore de projetos compartilhada usada por mgcvUI, earthUI e glmnetUI.

***Dica:** A formatação numérica nos eixos dos gráficos e nos rótulos de inclinação se adapta automaticamente à sua localidade. A localidade alemã usa pontos para os milhares (200.000), a finlandesa usa espaços (200 000) e a suíça usa apóstrofes (200'000). Nenhum símbolo monetário é exibido — o mgcvUI é agnóstico quanto à moeda.*

4 Modo Appraisal

Quando o propósito do projeto ativo é **For Appraisal**, o mgcvUI se configura para a avaliação de um único imóvel. Todos os recursos descritos no Capítulo 3 continuam disponíveis; este capítulo trata apenas dos acréscimos específicos da avaliação.

4.1 Tratamento da linha do avaliando

No modo de avaliação, **a linha 1 do seu conjunto de dados é o imóvel avaliando** e todas as demais linhas são vendas comparáveis. Seu arquivo de entrada deve estar organizado de acordo (veja o Capítulo 2). O preço de venda do avaliando pode ficar em branco ou receber qualquer valor — o mgcvUI automaticamente o trata como NA durante o ajuste.

Após o ajuste, o modelo gera previsões para a linha do avaliando, o que é a base do fluxo de trabalho de ajustes de RCA.

4.2 Data de referência e idade da venda

Nos modos de avaliação e de mercado, um campo **Effective Date** aparece na seção Variable Configuration (com a data de hoje como padrão). Se você designar uma coluna como **contract_date** no menu suspenso Special, o mgcvUI calcula uma coluna **sale_age** — o número inteiro de dias entre a data do contrato de cada venda e a data de referência. Essa coluna é acrescentada como preditor.

Quando a data de referência muda, **sale_age** é recalculada automaticamente.

4.3 Designações especiais de colunas

Nos modos de avaliação e de mercado, um menu suspenso **Special** aparece para cada preditor na tabela Variable Configuration. Veja o Capítulo 6 para a referência completa dos tipos especiais e de seus efeitos.

4.4 Visão geral dos ajustes de RCA

O botão **Calculate RCA Adjustments & Download** (visível nos modos de avaliação e de mercado após o ajuste) calcula ajustes derivados do mercado para cada comparável em relação ao avaliando. O fluxo de trabalho completo do RCA está descrito no Capítulo 11.

***Dica:** O score de CQA que você atribui ao avaliando controla quanto da distribuição de resíduos é atribuído ao avaliando. Um score de 5,00 coloca o avaliando na mediana dos comparáveis.*

5 Modo Market Area Analysis

Quando o propósito do projeto ativo é **Market Area Analysis**, o mgcvUI oferece os mesmos recursos específicos do setor imobiliário que o modo de avaliação (colunas especiais, idade da venda), mas é orientado à análise de um grupo de imóveis, e não à avaliação de um único avaliando.

5.1 Diferenças em relação ao modo Appraisal

- **O tratamento da linha do avaliando é flexível** — no modo de mercado, a linha 1 não é automaticamente tratada como imóvel avaliando.
- **Não há grade de vendas** — a etapa “Generate Sales Grid & Download” não está disponível.

5.2 Quando usar o modo de mercado

O modo Market Area Analysis é apropriado quando você está:

- Construindo um modelo GAM para um bairro ou região de mercado a fim de compreender os determinantes de valor não lineares
- Analisando como variáveis como área construída, idade, tamanho do terreno e localização afetam os preços de venda em um grupo de imóveis
- Preparando fundamentação para uma análise de condições de mercado ou para a delimitação de um bairro
- Trabalhando com um conjunto de dados em que você quer os recursos de colunas especiais (idade da venda), mas não precisa do fluxo de trabalho da grade de vendas

***Dica:** O modo de mercado também é útil para regressões imobiliárias em geral, quando você quer os recursos de colunas especiais mas não precisa do fluxo de trabalho de avaliação de um único imóvel.*

6 Seleção de variáveis

A Seção 4 da barra lateral — **Variable Configuration** — é onde você escolhe quais colunas participam do modelo e como elas são tratadas.

6.1 Variável alvo

O menu suspenso **Target (response) variable**, no alto da Seção 4, lista todas as colunas numéricas do seu conjunto de dados. Selecione uma coluna como variável resposta (por exemplo, o preço de venda). A coluna alvo é automaticamente excluída da lista de preditores.

6.2 Transformação da resposta

Abaixo do seletor de alvo, um menu suspenso **Response Transform** oferece três opções:

- **None (raw)** — usa a resposta como está. Adequado quando a relação entre preditores e resposta é aproximadamente aditiva na escala original.
- **Log (natural)** — aplica a transformação logarítmica natural antes do ajuste. Ao ser retrotransformada, ela torna os ajustes proporcionais (%) em vez de absolutos (\$). É particularmente útil para preços de venda, em que um ajuste de \$10.000 significa coisas muito diferentes em distintos patamares de preço.
- **Log10** — aplica a transformação logarítmica de base 10.

***Dica:** A transformação logarítmica torna os ajustes temporais proporcionais (%) em vez de absolutos (\$). Em modelos imobiliários que abrangem uma ampla faixa de preços, as transformações logarítmicas costumam produzir resíduos mais bem comportados.*

Quando se seleciona uma transformação logarítmica, os valores ≤ 0 na resposta são automaticamente filtrados. Todas as saídas (predições, contribuições, resíduos) são automaticamente retrotransformadas para a escala original.

6.3 A tabela de preditores

Abaixo do seletor de alvo, uma tabela rolável lista todas as colunas restantes. As colunas de caixas de seleção usam cabeçalhos verticais girados, por compacidade. A ordem das colunas é:

Coluna	Descrição
Variable	Nome da coluna (fonte monoespçada, truncado com reticências). Célula com borda.
Type	Menu suspenso de tipo de dado: numeric, integer, character, factor, logical, Date, POSIXct. Menu suspenso com borda.
Include	Caixa de seleção — incluir esta coluna como preditor no modelo. Cabeçalho vertical.
Factor	Caixa de seleção — tratar como fator categórico (cria um coeficiente por nível). Cabeçalho vertical.
Linear	Caixa de seleção — forçar uma relação linear (sem termo suave, apenas βx). Cabeçalho vertical.
Special	Menu suspenso (apenas nos modos de avaliação/mercado) — veja a referência de tipos especiais de coluna adiante. Menu suspenso com borda.
NAs	Contagem e percentual de valores ausentes, com código de cores: vermelho ($\geq 50\%$), laranja ($\geq 20\%$), cinza (> 0). Célula com borda.

Fator vs. suave vs. linear: por padrão, as variáveis numéricas incluídas recebem um termo suave $f(x)$. Marcar **Factor** cria um termo categórico (um coeficiente por nível). Marcar **Linear** cria um termo linear simples βx em vez de um termo suave. Variáveis marcadas como Factor e Linear ao mesmo tempo são tratadas como Factor.

6.4 Detecção de tipos de dados e substituições

O mgcvUI detecta automaticamente os tipos de dados na importação. Colunas numéricas, inteiras, lógicas, de fator e de data são reconhecidas. Você pode substituir qualquer detecção alterando o menu suspenso **Type**. Alterar os tipos afeta o modo como a coluna é tratada no modelo.

6.5 Referência dos tipos especiais de coluna

Nos modos de avaliação e de mercado, o menu suspenso **Special** oferece as seguintes opções:

Ponderação:

- **weight** — coluna de peso das observações (apenas uma é permitida; linhas com peso = 0 são excluídas do ajuste)

Tipos de data e hora:

- **contract_date** — dispara o cálculo automático de **sale_age** a partir da data de referência
- **sale_age** — identifica uma coluna de idade da venda já calculada (dias desde a data de referência), usada como está
- **listing_date** — usada como alternativa para calcular os dias no mercado
- **dom** — identifica a coluna de dias no mercado

Tipo de registro:

- **sale_type** — identifica a coluna de tipo de registro que distingue o avaliando das vendas comparáveis

Tipos monetários:

- **concessions** — identifica as concessões na venda (créditos do vendedor, incentivos ao comprador etc.)

Tipos de tamanho e localização:

- **latitude** — valores automaticamente arredondados para 3 casas decimais
- **longitude** — mesmo tratamento de arredondamento da latitude
- **area** — identificador de região de mercado ou de bairro
- **living_area** — habilita os cálculos de resíduo por pé quadrado (**residual_sf** e **cqa_sf**)
- **lot_size** — coluna de tamanho do terreno
- **site_dimensions** — agrupada com o tamanho do terreno

Tipos de idade:

- **actual_age** — coluna de idade do imóvel
- **effective_age** — idade efetiva do imóvel

Tipos de exibição:

- **display_only** — a coluna é incluída nas exportações para Excel, mas fica inteiramente excluída do ajuste do modelo. Use este tipo para campos de endereço, números de MLS, identificadores de matrícula ou outros dados de referência.

***Dica:** As colunas designadas como **display_only** permanecem em seu conjunto de dados e aparecem na exportação para Excel, mas ficam inteiramente fora da tabela de preditores. Use-as para colunas de identificação, endereços ou outros campos de referência.*

7 Seleção de parâmetros

A Seção 5 da barra lateral — **Mgcv Call Parameters** — dá acesso a todas as opções de configuração do GAM. Cada parâmetro tem um ícone azul de ajuda (?) com uma explicação em dica de tela.

7.1 Predefinições de parâmetros

Um menu suspenso no alto oferece duas predefinições que configuram valores padrão sensatos para fluxos de trabalho comuns:

Predefinição	Base	Configurações
Standalone (descoberta)	tp (thin plate)	k = auto (10), gamma = 1.2, select = TRUE
Earth Pipeline (refinamento)	cr (regressão cúbica)	k = auto (10), gamma = 1.4, select = TRUE

A predefinição **Earth Pipeline** é selecionada automaticamente quando se importam nós do earthUI. A base de spline de regressão cúbica (cr) é necessária para a integração dos nós do earth, porque permite especificar posições exatas de nós.

7.2 Family

Escolha a família de distribuição para sua variável resposta:

Família	Caso de uso
gaussian	Respostas contínuas (por exemplo, preço de venda). A mais comum.
Gamma	Respostas contínuas positivas com variância proporcional à média.
poisson	Dados de contagem (por exemplo, número de vendas).
binomial	Desfechos binários.
inverse.gaussian	Dados positivos com forte assimetria à direita.

7.3 Method

O método de estimação do parâmetro de suavização:

Método	Descrição
REML	Máxima verossimilhança restrita (padrão, recomendado). O mais robusto contra o sobreajuste.
GCV.Cp	Validação cruzada generalizada. Tende a selecionar modelos um pouco mais complexos.
ML	Máxima verossimilhança. Semelhante ao REML, mas pode subsuavizar.
P-REML	Variante REML de Pearson.
P-ML	Variante ML de Pearson.
GACV.Cp	Validação cruzada aproximada generalizada.

7.4 Gamma (multiplicador de suavização)

O parâmetro gamma multiplica os graus de liberdade efetivos no critério de suavização, favorecendo ajustes mais suaves (menos ondulados). O padrão é 1.2.

Valor	Efeito
1.0	Suavização padrão
1.2–1.4	Um pouco mais suave que o padrão (recomendado)
2.0+	Muito mais suave — bom para dados ruidosos ou amostras pequenas

Valores mais altos de gamma protegem contra o sobreajuste ao penalizar a complexidade mais fortemente. A predefinição Earth Pipeline usa 1.4 para obter suavização adicional ao refinar os nós do earth.

7.5 Validação cruzada

Quando marcada (padrão), o mgcvUI executa validação cruzada de 10 dobras após o ajuste para calcular um R quadrado de validação cruzada. Isso fornece uma estimativa honesta do poder preditivo fora da amostra.

7.6 Select (encolhimento)

Quando marcada (padrão), acrescenta a cada termo suave uma penalização extra que pode encolhê-lo até zero. Isso habilita a seleção automática de variáveis — os termos suaves sem importância são efetivamente removidos do modelo. Implementado por meio de `select = TRUE` em `gam()`.

7.7 Base padrão

O tipo de função de base spline para os termos suaves:

Base	Nome completo	Descrição
tp	Spline de regressão thin plate	Padrão. Isotrópica (não é preciso posicionar nós). Boa escolha de uso geral.
cr	Spline de regressão cúbica	Permite o posicionamento explícito de nós. Necessária para a integração dos nós do earth.
ps	P-spline	Base de B-splines igualmente espaçadas com penalização por diferenças. Computacionalmente eficiente.
bs	B-spline	Base de B-splines flexível.

7.8 k padrão (dimensão da base)

A dimensão da base k controla a ondulação máxima de cada termo suave. O valor 0 (o padrão) significa “automático” — o mgcvUI usa $k = 10$ ou o número de nós do earth, o que for apropriado.

- **k pequeno (3–5)**: curvas muito suaves, quase lineares.
- **k médio (8–15)**: flexibilidade moderada (em geral suficiente).
- **k grande (20+)**: muito flexível — risco de sobreajuste sem regularização forte.

Dica: Os graus de liberdade efetivos (EDF) na aba Summary indicam quantos graus de liberdade cada termo suave de fato usa. Se o EDF estiver próximo de $k - 1$, a base pode ser pequena demais — aumente k . Se o EDF for bem menor que k , o termo suave está bem penalizado.

7.9 Tipo de tensor

Para interações entre variáveis contínuas, há dois tipos de produto tensorial disponíveis:

Tipo	Função	Descrição
ti	ti(x1, x2)	Interação tensorial — modela apenas o efeito de interação, com os efeitos principais tratados separadamente por termos <code>s()</code> univariados. Preferível pela interpretabilidade.
te	te(x1, x2)	Produto tensorial — modela todo o efeito conjunto, inclusive os efeitos principais. Mais difícil de decompor para os ajustes de RCA.

7.10 Interações permitidas

Uma seção recolhível **Allowed Interactions** exibe uma matriz triangular superior de caixas de seleção para todos os preditores suaves incluídos (não lineares e não fatores). Cada caixa de seleção habilita um termo suave de produto tensorial entre o par de variáveis correspondente.

- **Allow All / Clear All** — botões de alternância em massa.
- **Clicar no nome de uma variável** — alterna todas as interações daquela variável.
- **Clicar com o botão direito no nome de uma variável** — “Block from main effect” — marca a variável para uso apenas em interações (sem termo suave univariado).

Quando se importam nós do earthUI, as interações detectadas pelo earth ficam previamente marcadas e bloqueadas (destacadas com fundo amarelo). Se o modelo earth tiver usado `degree = 1` (sem interações), exibe-se uma mensagem informativa.

Aviso

Habilitar interações tensoriais aumenta substancialmente a complexidade do modelo. Use interações quando houver conhecimento de domínio ou evidência detectada pelo earth que as sustente. Verifique se os ajustes resultantes são razoáveis e explicáveis.

7.11 Interações fator por termo suave

Uma matriz separada, **Factor-by-Smooth Interactions**, aparece quando há variáveis de fator e variáveis suaves incluídas. Cada caixa de seleção cria um termo `s(x, by = factor)` — uma curva suave distinta para cada nível do fator.

7.12 Parâmetros avançados

Uma seção recolhível “Advanced” expõe configurações adicionais:

- **Optimizer** — algoritmo de estimação do parâmetro de suavização: `outer/newton` (padrão), `outer/bfgs` ou `efs` (extended Fellner-Schall).

- **Scale** — parâmetro de escala. 0 (padrão) = estimar a partir dos dados. Defina um valor conhecido para modelos de escala fixa.
- **Discrete covariate method** — ajuste discretizado rápido para conjuntos de dados grandes (`discrete = TRUE`).
- **Threads** — número de threads paralelas (1–32). Só tem efeito quando `discrete = TRUE`.

8 Ajuste do modelo

8.1 O botão de ajuste

A Seção 6 da barra lateral contém o botão **Fit MgcV GAM Model**. Clicar nele executa o modelo com sua configuração atual.

8.2 O que acontece durante o ajuste

Quando você clica em Fit, o mgcvUI:

1. **Valida a configuração** — verifica se há pelo menos um preditor e pelo menos 10 linhas completas. Avisa se menos de 50% das linhas estiverem completas (listando as colunas responsáveis pela maior parte dos NAs).
2. **Prepara os dados** — restringe às colunas relevantes para o modelo, remove linhas com NAs, aplica os pesos (se designados), exclui a linha do avaliando (no modo de avaliação) e aplica as transformações da resposta.
3. **Constrói a fórmula do GAM** — monta termos suaves $s(x, bs="cr", k=10)$ para os preditores numéricos, termos de fator para os categóricos, termos lineares para as variáveis marcadas como Linear e produtos tensoriais para as interações selecionadas. Os nós do earth são incorporados como posições explícitas de nós nos termos de base **cr**.
4. **Trata os casos-limite** — descarta variáveis constantes, converte preditores binários em fatores, limita k ao número de valores únicos menos 1 e complementa os nós do earth com quantis dos dados quando são necessários mais nós.
5. **Ajusta o GAM** — chama `mgcv::gam()` com a fórmula construída, a família, o método, gamma, select e os demais parâmetros.
6. **Executa a validação cruzada** (se habilitada) — VC de 10 dobras, calculando o R quadrado fora da amostra.
7. **Atualiza todas as abas de resultados** — Summary, Equation, gráficos de Contribution, Diagnostics e todas as demais abas são preenchidos.

8.3 Ajuste em segundo plano e progresso

Quando o pacote `callr` está disponível, o ajuste do modelo é executado em um processo em segundo plano, de modo que o aplicativo permanece responsivo. Uma sobreposição de progresso em estilo de terminal, com tema escuro, mostra:

- Um **cronômetro** exibindo os segundos/minutos decorridos.
- Um **registro de execução com código de cores**: azul para o estado, amarelo para o progresso das dobras da VC, vermelho para erros e verde para resultados.
- Um botão **Abort** para encerrar o processo de ajuste, se necessário.
- Um botão **Close**, que aparece depois da conclusão.

Se `callr` não estiver disponível, o ajuste é executado de forma síncrona (o aplicativo congela até o ajuste terminar).

Após um ajuste bem-sucedido, aparece uma marca de verificação branca no botão Fit e uma linha de estado mostra: “R-sq = X, CV R-sq = X, Dev = X%, AIC = X, n = X.”

9 Abas de resultados

9.1 Data

Mostra os dados importados como uma DataTable interativa, com rolagem horizontal e 15 linhas por página. Nos modos de avaliação e de mercado, a aba é dividida em uma tabela **Subject Property** (linha 1) e uma tabela **Comparable Sales** (as demais linhas); no modo General, exibe-se uma única tabela.

Cada célula é mantida em uma única linha e truncada na largura da coluna, de modo que campos largos de texto livre (como observações sobre o imóvel) não estiquem uma linha página abaixo. Um clique simples seleciona a linha; **clique duas vezes em qualquer célula** para abrir uma janela suspensa com seu conteúdo completo, sem truncamento.

9.2 Equation

Exibe o modelo GAM ajustado em duas partes:

Equação do modelo (renderizada em MathJax):

- O lado esquerdo mostra a resposta com a transformação apropriada (log, log10) e a função de ligação.
- O lado direito mostra o valor do intercepto, os termos suaves como $f_i(\text{variável})$, os termos de fator agrupados por variável com a contagem de níveis e os termos lineares com os valores dos coeficientes.
- A família e o método são exibidos acima da equação.

Tabela **Smooth Function Definitions**:

- Texto da fórmula do termo (por exemplo, `s(living_sqft, bs="cr", k=8)`)
- Nome da variável
- Tipo (s, te, ti ou linear)
- Tipo de função de base
- Valor de k
- Posições dos nós do earth (se importados)

9.3 Summary

Seis cartões de métricas exibidos na parte superior:

- **R-squared** — proporção da desviância explicada
- **CV R-squared** — R quadrado de validação cruzada (se a VC estiver habilitada)
- **Dev. Explained %** — percentual de desviância explicada
- **AIC** — critério de informação de Akaike
- **n** — número de observações utilizadas
- **Smooths** — número de termos suaves

Abaixo dos cartões:

- **Tabela Smooth Terms** — termo, EDF (graus de liberdade efetivos), Ref.df, estatística F, valor-p
- **Tabela Parametric Terms** — termo, estimativa, erro padrão, valor t, valor-p

9.4 Variable Importance

Um gráfico de barras horizontais mostrando a importância relativa de cada termo do modelo:

- **Termos suaves**: ordenados pela estatística F (barras azuis)
- **Termos paramétricos**: ordenados pelo valor $|t|$ (barras verdes)

Abaixo do gráfico, uma DataTable mostra termo, tipo, EDF, estatística e valor-p, ordenados pela estatística em ordem decrescente.

9.5 Gráficos de contribuição

Todo termo do modelo que contribui para o valor predito tem um gráfico de contribuição interativo em **plotly**, exibido em um cartão com borda. A aba mostra gráficos para todos os tipos de termo:

9.5.1 Termos suaves univariados — $s(x)$

- **Linha azul** com faixa de confiança de 95%
- **Pontos cinza de dispersão** mostrando os resíduos parciais (para modelos sem transformação logarítmica)
- **Linhas verticais vermelhas tracejadas** nas posições dos nós do earth (se o earth tiver sido importado)
- **Dica de tela ao passar o cursor:** valor da variável, contribuição (\$), IC de 95%, taxa (inclinação por unidade)

Nos **modelos com transformação logarítmica**, o eixo y é retrotransformado para contribuições em dólares pela fórmula $\bar{y} \times (e^{f(x)} - 1)$, o que torna as curvas diretamente interpretáveis em termos monetários.

Para as variáveis de **latitude/longitude**, a inclinação é mostrada por 0,001 grau, e não por grau, já que esse é um incremento geográfico mais significativo.

9.5.2 Termos fator por suave — $s(x, \text{by} = \text{factor})$

Uma linha colorida por nível do fator, sobrepostas em um único gráfico. Útil para ver como o efeito de um preditor varia entre grupos (por exemplo, entre bairros diferentes).

9.5.3 Termos de interação — $ti(x_1, x_2)$ e $te(x_1, x_2)$

Mapa de calor mostrando a superfície bidimensional de contribuição. Escala de cores: vermelho (negativo) → branco (zero) → azul (positivo). Ao passar o cursor, exibem-se os valores das duas variáveis e o montante da contribuição.

9.5.4 Termos paramétricos — lineares e de fator

- **Termos lineares numéricos:** gráfico de dispersão dos pontos de dados com a reta ajustada sobreposta
- **Termos de fator:** gráfico de barras mostrando a contribuição média por nível do fator

9.6 Correlation

Uma matriz em mapa de calor mostrando as correlações de Pearson entre todos os preditores numéricos. Disponível imediatamente após a importação dos dados — não é preciso ajustar o modelo.

- Gradiente de cores: azul (-1) → branco (0) → vermelho ($+1$)
- Valores de correlação impressos em cada célula
- Útil para identificar concurvidade (o análogo suave da multicolinearidade) antes do ajuste

9.7 Diagnostics

Dois conjuntos de visualizações de diagnóstico:

Painel de diagnóstico (via `gratia::appraise`): quatro gráficos — resíduos versus ajustados, gráfico Q-Q, histograma dos resíduos e resposta versus valores ajustados.

Gráfico de dispersão **Actual vs Predicted**: valores observados versus preditos, com uma reta de referência tracejada a 45 graus. Nos modelos com transformação logarítmica, os valores são retrotransformados para a escala original.

9.8 RCA Adjustments

Três histogramas exibidos após o cálculo do RCA (somente no modo de avaliação/mercado):

- **Residual Adjustment %** — distribuição dos ajustes de resíduo como percentual do preço de venda
- **Net Adjustment %** — distribuição dos ajustes líquidos
- **Gross Adjustment %** — distribuição dos ajustes brutos

Cada histograma mostra média, mediana e desvio padrão no subtítulo, com linhas de referência tracejadas para a média e a mediana.

9.9 Anova

Uma tabela de ANOVA combinada, reunindo os termos paramétricos e suaves do resumo do GAM. Mostra o termo, o tipo (paramétrico ou suave) e todas as estatísticas associadas.

9.10 Mgcvc Output

Saída em texto bruto mostrando:

- Tempos de execução do ajuste
- A fórmula do GAM
- A saída de impressão do modelo
- A saída completa de `summary(model)`
- A família e a função de ligação
- As medidas globais de concurvidade

9.11 Sign Check

Quando se importam nós do earthUI, esta aba compara a direção de cada variável no modelo earth com sua direção no GAM:

- **earth_direction**: crescente, decrescente ou mista (com base nos sinais das funções de dobradiça)
- **gam_direction**: crescente, decrescente ou mista (estimada a partir da derivada do termo suave em uma grade de 100 pontos)
- **consistent**: TRUE/FALSE
- **warning**: descrição de eventual divergência

Uma linha de estado mostra “All directions consistent” ou o número de variáveis com inconsistências.

9.12 Concurvity

A concurvidade é o análogo suave da multicolinearidade. Esta aba mostra:

- **Tabela Overall Concurvity:** linhas para as medidas worst/observed/estimate e colunas para cada termo suave. Com código de cores: branco ($<0,5$), amarelo ($0,5-0,8$), vermelho ($>0,8$).
- **Tabela Pairwise Concurvity:** matriz aos pares no pior caso, com o mesmo código de cores.

Valores acima de $\sim 0,8$ na linha “worst” indicam que um termo suave pode ser bem aproximado pelos demais termos suaves, sugerindo redundância.

9.13 Basis Check

Saída de `mgcv::gam.check()`: testes de adequação da dimensão da base para cada termo suave. Se o k de um termo suave for pequeno demais, este teste o sinalizará com um valor-p baixo, sugerindo que você aumente k .

10 Download dos dados

Após o ajuste, o botão **Download Output (Excel)** (seção 7 da barra lateral) exporta um arquivo Excel com predições e diagnósticos para a pasta de saída do mgcv do projeto ativo.

10.1 Colunas de saída

Coluna	Descrição
basis	Valor do intercepto retrotransformado (o mesmo para todas as linhas)
<var>_contribution	Contribuição de cada termo suave para a predição. Nos modelos logarítmicos, alocada proporcionalmente a valores em dólares.
est_<target>	Predição do modelo (por exemplo, <code>est_sale_price</code>), retrotransformada da escala logarítmica quando aplicável
residual	Observado – predito
calc_residual	Verificação: observado – soma das contribuições (retrotransformada)
cqa	Escore de Condition-Quality-Appeal (CQA), em escala de 0 a 10
residual_sf	Resíduo / área habitável (se <code>living_area</code> estiver designada)
cqa_sf	CQA calculado pela ordenação segundo <code>residual_sf</code>

No modo de avaliação, a linha 1 (avaliando) tem `residual/cqa/cqa_sf` definidos como NA. As linhas são ordenadas por `residual_sf` em ordem decrescente quando há uma coluna `living_area` designada.

Uma marca de verificação branca aparece no botão de download após a conclusão bem-sucedida.

10.2 Escores de CQA

O CQA ordena o resíduo de cada linha em relação a todos os demais, em uma escala de 0 a 10:

- **CQA alto (~9–10):** vendeu por muito mais do que o predito — provavelmente condição/qualidade/atratividade superiores
- **CQA baixo (~0–1):** vendeu por muito menos do que o predito — provavelmente condição inferior ou venda em situação de desconforto
- **CQA ~5:** próximo da mediana

11 Cálculos de RCA e download

O fluxo de trabalho do RCA (Residual Constraint Approach) está disponível nos modos **Appraisal** e **Market Area Analysis**, depois do ajuste do modelo. O imóvel avaliando deve estar na linha 1.

11.1 Abertura da caixa de diálogo do RCA

Clique no botão **Calculate RCA Adjustments & Download** na seção 8 da barra lateral. Aparece uma caixa de diálogo modal com:

- **Tipo de escore:** CQA ou CQA por SF (se `living_area` estiver designada). A opção por SF escala o resíduo do avaliando pela área habitável.
- **Subject CQA Score:** aceita valores de 0,00 a 10,00. O campo começa em branco; quando um modelo do earthUI tiver sido importado, o tipo e o valor de escore que você usou no earthUI são trazidos das configurações compartilhadas do projeto.

Clique em **Generate** para calcular e baixar. O arquivo Excel é gravado na pasta de saída do mgcv do projeto ativo (`<os>_out_mgcv`).

11.2 Interpolação do escore de CQA

1. Os escores de CQA e os resíduos dos comparáveis são ordenados
2. A interpolação linear (`approx()`) mapeia seu valor de CQA para um resíduo
3. Valor do avaliando = predição do modelo + resíduo interpolado
4. As linhas com peso 0 são tratadas como se tivessem o mesmo resíduo do avaliando

11.3 Colunas de saída

Coluna	Descrição
<code>subject_value</code>	Predição do modelo + resíduo interpolado
<code>subject_cqa</code>	Escore de CQA informado pelo usuário
<code><var>_adjustment</code>	Contribuição do avaliando – contribuição do comparável (em dólares)
<code>residual_adjustment</code>	Resíduo do avaliando – resíduo do comparável
<code>net_adjustments</code>	Soma de todos os ajustes
<code>gross_adjustments</code>	Soma dos ajustes em valor absoluto
<code>residual_adj_pct</code>	Ajuste de resíduo como % do preço de venda
<code>net_adj_pct</code>	Ajustes líquidos como % do preço de venda
<code>gross_adj_pct</code>	Ajustes brutos como % do preço de venda
<code>adjusted_sale_price</code>	Preço de venda observado + ajustes líquidos

Uma marca de verificação branca aparece no botão após o cálculo bem-sucedido.

12 Grade de comparação de vendas

A grade de comparação de vendas está disponível somente no modo **Appraisal**, depois do cálculo dos ajustes de RCA. Ela gera uma pasta de trabalho Excel formatada, adequada para inclusão em laudos de avaliação.

12.1 Caixa de diálogo de seleção de comparáveis

Clicar no botão **Generate Sales Grid & Download** abre uma caixa de diálogo modal listando todas as vendas comparáveis. Os comparáveis são divididos em dois grupos:

- **Comparáveis recomendados** (previamente marcados): ajuste bruto < 25% do preço de venda, ordenados pela idade da venda (o mais recente primeiro). No máximo 30 comparáveis.
- **Outros comparáveis** (desmarcados por padrão): os comparáveis restantes, ordenados pelo percentual de ajuste bruto.

Selecione seus comparáveis e clique em **Generate Sales Grid** para criar a pasta de trabalho.

Tipos especiais necessários: `contract_date` e `living_area` devem estar designados.

Tipos especiais recomendados: `latitude`, `longitude` e `lot_size` — a grade funciona sem eles, mas alguns campos ficarão em branco.

12.2 Estrutura da pasta de trabalho

A pasta de trabalho gerada contém várias planilhas, cada uma com 3 comparáveis lado a lado:

- **Linhas de cabeçalho:** endereços do avaliando e dos comparáveis, preços de venda, datas de venda e dias no mercado
- **Linhas agrupadas:** localização (latitude, longitude, região), tamanho do terreno (lot size, site dimensions) e idade (idade real, idade efetiva), com cabeçalhos de grupo e subitens
- **Linhas das variáveis do modelo:** uma linha por termo suave, mostrando a contribuição do avaliando, as contribuições dos comparáveis e os montantes de ajuste
- **Linhas de características residuais:** células vazias e desbloqueadas para preenchimento pelo avaliador quanto a características não captadas pelo modelo (por exemplo, condição, qualidade, vista)
- **Linhas de resumo:** % de ajuste líquido, % de ajuste bruto e preço de venda ajustado, com fórmulas ativas do Excel

12.3 Proteção da planilha e células residuais

Cada planilha é protegida para evitar edições acidentais, mas as células de valor das características residuais ficam explicitamente desbloqueadas. Isso permite que os avaliadores informem valores para características ausentes do modelo, preservando os cálculos de ajuste orientados por fórmulas.

13 Download de relatórios

A geração de relatórios é um fluxo de trabalho Quarto em duas etapas, que substitui o relatório de etapa única das versões anteriores.

13.1 Etapa 1: gerar o relatório Quarto

O botão **Generate Quarto Report** (disponível depois que o modelo é ajustado) grava um pacote `.qmd` autocontido na pasta de saída do `mgecv` do projeto ativo. O pacote contém o código-fonte Quarto, arquivos de gráficos pré-renderizados (PNG e PDF), uma carga de dados `report_data.rds` com todas as tabelas e os metadados do modelo, e um modelo de estilo `reference.docx` para a saída em Word. Como o pacote é autocontido, você pode editar o código `.qmd` ou combiná-lo com outros documentos Quarto antes de renderizar.

Os gráficos são pré-renderizados quando o pacote é gerado; se um gráfico individual falhar, grava-se uma imagem de espaço reservado e o restante do pacote é gerado normalmente.

13.2 Etapa 2: converter o relatório Quarto

A seção **Convert Quarto Report** renderiza um arquivo `.qmd` em um ou mais formatos de saída, via `quarto::quarto_render()`. Por padrão, usa o pacote gerado mais recentemente, mas o botão **Browse** permite selecionar qualquer arquivo `.qmd`. Há três formatos disponíveis:

- **HTML** — documento HTML autocontido. Funciona em todas as plataformas, sem software adicional. Padrão recomendado.
- **Word (.docx)** — renderizado com o modelo de estilo `reference.docx` incluído no pacote. Adequado para edição e distribuição.
- **PDF** — **exige uma instalação de LaTeX** (veja os requisitos de sistema adiante). Se nenhum LaTeX for detectado, a opção de PDF é automaticamente ocultada nas caixas de seleção de formato.

Os renderizadores legados de etapa única, `render_gam_report()` e `export_gam_docx()`, continuam disponíveis para uso programático.

13.3 Conteúdo do relatório

O relatório gerado inclui:

- **Conjunto de dados e modelo:** arquivo de dados, tamanho da amostra, família, método e informações sobre o ajuste
- **Resumo dos resultados:** R^2 , R^2 de VC, desviância explicada, AIC e métricas de ajuste relacionadas
- **Equação do modelo:** equação do modelo com a notação f_i para os termos suaves e coeficientes explícitos para os termos lineares, além das definições das funções suaves. As equações longas se quebram em várias linhas no PDF.
- **Tabelas de resumo:** termos suaves (EDF, F, valor-p) e termos paramétricos (estimativa, erro padrão, t, valor-p)
- **Interações habilitadas:** uma matriz dos termos de interação tensorial e de fator por suave incluídos no modelo (exibida somente quando há interações)
- **Importância das variáveis:** gráfico de barras ordenado pela estatística F (termos suaves) ou por $|t|$ (paramétricos)
- **Efeitos suaves** (todos os tipos de termo):
 - Gráficos de termos suaves univariados (versões estáticas em `ggplot2` dos gráficos interativos)

- Gráficos multilinhas de fator por suave
- Mapas de calor de interação (termos ti e te)
- Gráficos de termos paramétricos (dispersão para numéricos, barras para fatores)
- Mapa de calor de **correlações entre preditores**
- **Diagnósticos:** diagnósticos de resíduos e dispersão de observados versus preditos
- **Resumo do mgcv:** a saída de texto completa do modelo ajustado
- **Apêndice de notação:** uma seção final não numerada explicando a notação das funções de contribuição de valor usada no relatório

Uma marca de verificação branca aparece no botão Generate Quarto Report após a geração bem-sucedida.

14 Exportação das funções do modelo

O `mgcvUI` pode exportar as funções suaves do GAM ajustado como código autônomo em várias linguagens. Cada termo suave é avaliado em uma grade de 200 pontos e exportado como uma tabela de consulta com interpolação linear, o que permite usar o modelo fora do R sem qualquer dependência de `mgcv`.

Na versão atual, a exportação de funções é programática. Os auxiliares exportados recebem um objeto de resultado ajustado (produzido por `fit_gam()`) e gravam os arquivos descritos a seguir:

- `generate_r_code()`, `generate_python_code()` e `generate_cpp_code()` — arquivos de código-fonte por linguagem
- `export_functions_sqlite()` — banco de dados SQLite com as tabelas de consulta
- `export_functions_zip()` — reúne qualquer combinação dos quatro formatos em um arquivo zip

14.1 Funções em R

Arquivo: `gam_functions.R`

- Comentários de cabeçalho com a variável resposta, R^2 , R^2 de VC e a fórmula.
- Uma constante `INTERCEPT`.
- Um `approxfun()` por variável suave, com os vetores `x/y` pré-calculados.
- Os coeficientes dos termos lineares como constantes nomeadas.
- Uso: `some` o intercepto, cada `g_var(value)` e cada `coef_var * value`.

14.2 Funções em Python

Arquivo: `gam_functions.py`

- Importação de `numpy`.
- Constante `INTERCEPT`.
- Por variável: arranjos `_x_var` e `_fx_var` e função `g_var(x)` usando `np.interp()`.
- Os coeficientes dos termos lineares como constantes `COEF_*`.

14.3 Cabeçalho C++

Arquivo: `gam_functions.hpp`

- Somente de cabeçalho, com `#pragma once`.
- Espaço de nomes `gam_functions`.
- Auxiliar `interp()` usando `std::lower_bound`.
- Por variável: função inline com arranjos `static const`.
- Os coeficientes dos termos lineares como `constexpr double`.

14.4 Banco de dados SQLite

Arquivo: `gam_functions.sqlite`

- Tabela `smooth_grids`: colunas (`variable`, `x`, `fx`) — as tabelas de consulta de 200 pontos.
- Tabela `model_info`: `response`, `r_squared`, `cv_rsqr`, `dev_explained`, `aic`, `n_obs`, `family`, `method`, `formula`.
- Tabela `linear_terms`: `variable`, `coefficient`.
- Tabela `intercept`: valor do intercepto.

15 Comparação com earthUI e glmnetUI

O aplicativo earth e seus módulos de pós-processamento formam um conjunto de ferramentas para modelagem por regressão. Eles compartilham o mesmo formato de dados, os mesmos tipos especiais de coluna, o mesmo fluxo de trabalho de RCA e os mesmos conjuntos de dados de demonstração, mas usam motores de modelagem diferentes.

15.1 Principais diferenças

Característica	mgcvUI	earthUI	glmnetUI
Método	GAM (mgcv)	MARS/Earth (earth)	Elastic net (glmnet)
Relações	Não lineares suaves	Lineares por partes	Lineares
Seleção de variáveis	Encolhimento com <code>select=TRUE</code>	Poda por passos	Penalização Lasso
Interações	Produtos tensoriais (te/ti)	Produtos de dobradiças (grau 2-3)	Produtos aos pares
Efeitos parciais	Curvas suaves com IC	Funções g (por partes)	Inclinações lineares
Fluxo com o earth	Importa nós do earth	Fonte dos nós	Importa a base de dobradiças
Exportação de código	R, Python, C++, SQLite	R, Python, C++, SQLite	—
Formatos de relatório	Word, PDF, HTML	Word, PDF, HTML	Word, PDF, HTML

Vale registrar um diferencial exclusivo do earthUI: ele inclui um recurso experimental em Prolog/DCG que extrai atributos estruturados de imóveis a partir das observações em texto livre do MLS.

15.2 Recursos compartilhados

As três ferramentas oferecem:

- Uma árvore de projetos `regProj` compartilhada (por padrão `~/regProj`) — projetos, arquivos de entrada e saídas por método são compartilhados entre os aplicativos irmãos
- Importação de CSV e Excel com conversão dos nomes de coluna para `snake_case`
- Suporte a localidades de mais de 30 países, com separador de CSV, marcador decimal e tamanho de papel
- Designações especiais de colunas (`contract_date`, `living_area`, `area`, `latitude`, `longitude` etc.)
- Cálculo automático de `sale_age` a partir da data de referência
- Alternância entre modo claro e escuro (tema Nord)
- Persistência das configurações (SQLite e/ou `localStorage`)
- Mapa de calor de correlações, importância das variáveis, diagnósticos
- Escores de CQA e fluxo de trabalho de ajustes de RCA
- Grade de comparação de vendas com seleção de comparáveis e saída formatada em Excel
- Compatibilidade com os mesmos conjuntos de dados de demonstração (`Appraisal_1.csv`)

15.3 Fluxo de trabalho recomendado

- **Comece pelo earthUI** para descobrir as variáveis importantes, as relações não lineares e as interações. A construção automática de modelos do earth oferece o melhor ponto de partida.

- **Refine com o mgcvUI** se quiser efeitos parciais mais suaves. Importe os nós do earthUI para uma transição contínua.
- **Use o glmnetUI** quando precisar de um modelo puramente linear com seleção agressiva de variáveis ou restrições de sinal dos coeficientes.
- **Compare os três** sobre os mesmos dados para avaliar se a complexidade adicional dos modelos não lineares se justifica.

16 Conjunto de dados de demonstração: Appraisal_1.csv

16.1 Descrição

O conjunto de dados de demonstração é compartilhado com o earthUI e o glmnetUI. Ele está na pasta `demo_mls/` do código-fonte do mgcvUI ou, se o earthUI estiver instalado:

```
demo_file <- system.file("extdata", "Appraisal_1.csv", package = "earthUI")
```

O arquivo contém cerca de 1.500 registros de vendas residenciais de uma exportação simulada de MLS, com o imóvel avaliando na linha 1 e as vendas comparáveis nas demais linhas. Os dados representam vendas de residências unifamiliares em um mercado com várias regiões, com uma variedade de tamanhos, idades e localizações de imóveis.

Não se trata de dados reais, mas eles se baseiam em um bairro realista do norte da Califórnia. Todas as informações de identificação foram alteradas ou removidas.

16.2 Colunas

Coluna	Tipo	Tipo especial	Descrição
<code>sale_type</code>	character	<code>sale_type</code>	Código de tipo de registro que distingue o avaliando das vendas comparáveis
<code>weight</code>	numeric	<code>weight</code>	Peso da observação (0 = excluir do ajuste)
<code>id</code>	numeric	<code>display_only</code>	Identificador interno do registro
<code>property_id</code>	numeric	<code>display_only</code>	Identificador do imóvel no MLS
<code>listing_id</code>	character	<code>display_only</code>	Número do anúncio no MLS
<code>parcel_number</code>	character	<code>display_only</code>	Número da matrícula no cadastro do condado (APN)
<code>street_address</code>	character	<code>display_only</code>	Endereço do imóvel
<code>city_name</code>	character	<code>display_only</code>	Cidade
<code>postal_code</code>	character	<code>display_only</code>	Código postal (ZIP)
<code>county_name</code>	character	<code>display_only</code>	Condado
<code>contract_date</code>	Date	<code>contract_date</code>	Data do contrato de venda (calcula <code>sale_age</code>)
<code>sale_age</code>	numeric	—	Dias entre a data do contrato e a data de referência (pré-calculado)
<code>coe_date</code>	Date	<code>display_only</code>	Data de encerramento do escrow
<code>listing_status</code>	character	<code>display_only</code>	Situação do anúncio (por exemplo, "Sold")
<code>sale_price</code>	numeric	(alvo)	Preço de venda — variável resposta
<code>rent</code>	numeric	—	Aluguel mensal (para modelos com vários alvos)
<code>list_price</code>	numeric	<code>display_only</code>	Preço de anúncio
<code>original_list_price</code>	numeric	<code>display_only</code>	Preço de anúncio original
<code>living_sqft</code>	numeric	<code>living_area</code>	Área bruta habitável em pés quadrados
<code>ppsf</code>	numeric	<code>display_only</code>	Preço por pé quadrado
<code>beds_total</code>	integer	—	Número de dormitórios
<code>baths_total</code>	numeric	—	Número total de banheiros (por exemplo, 2,5 = 2 completos + 1 lavabo)
<code>lot_size</code>	numeric	<code>lot_size</code>	Tamanho do terreno em pés quadrados
<code>area_id</code>	integer	<code>area</code>	Identificador de região do MLS
<code>area_text</code>	character	<code>display_only</code>	Nome da região
<code>age</code>	numeric	<code>actual_age</code>	Idade do imóvel em anos
<code>year_built</code>	integer	<code>display_only</code>	Ano de construção
<code>latitude</code>	numeric	<code>latitude</code>	Latitude (arredondada para 3 casas decimais no modelo)
<code>longitude</code>	numeric	<code>longitude</code>	Longitude (arredondada para 3 casas decimais no modelo)
<code>latitude6</code>	numeric	<code>display_only</code>	Latitude com precisão total
<code>longitude6</code>	numeric	<code>display_only</code>	Longitude com precisão total
<code>garage_spaces</code>	integer	—	Número de vagas de garagem

Coluna	Tipo	Tipo especial	Descrição
<code>fp_count</code>	integer	—	Número de lareiras
<code>no_of_stories</code>	numeric	—	Número de pavimentos
<code>style</code>	character	—	Estilo arquitetônico
<code>view</code>	character	—	Tipo de vista (por exemplo, “Neighborhood”, “Hills”)
<code>days_on_market</code>	integer	<code>dom</code>	Dias no mercado
<code>listing_date</code>	Date	<code>listing_date</code>	Data do anúncio
<code>sale_concessions</code>	numeric	<code>concessions</code>	Concessões do vendedor

16.3 Roteiro rápido sugerido

1. Inicie a rotina mgcv: `earthUI::launch_mgcv()`
2. Crie um novo projeto (Seção 1) com o propósito **For Appraisal**
3. Coloque `Appraisal_1.csv` na pasta de entrada do projeto e selecione-o na Seção 2 (clique em **Refresh** se necessário)
4. Selecione `sale_price` como alvo
5. Opcionalmente, defina Response Transform como **Log (natural)** para obter ajustes proporcionais
6. Atribua os tipos especiais conforme a tabela acima
7. Inclua os preditores: `sale_age`, `living_sqft`, `baths_total`, `lot_size`, `area_id` (como fator), `age`, `latitude`, `longitude`, `garage_spaces`
8. Mantenha os padrões: REML, $\gamma = 1.2$, base tp, $k = \text{auto}$, `select = TRUE`
9. Clique em **Fit Mgcv GAM Model**
10. Examine as abas Summary, Contribution (gráficos suaves) e Diagnostics
11. Baixe a saída (etapa 7) e examine a ordenação por CQA
12. Calcule os ajustes de RCA (etapa 8) com um escore de CQA de ~5,00
13. Gere uma grade de comparação de vendas (etapa 9) com os comparáveis recomendados

17 Requisitos de sistema e solução de problemas

17.1 Plataformas suportadas

O mgcvUI roda em **macOS**, **Windows** e **Linux**. O aplicativo é desenvolvido e testado principalmente em macOS com o RStudio. As observações específicas de cada plataforma vêm a seguir.

17.2 Versão do R

É necessário $R \geq 4.1.0$. Recomenda-se fortemente o RStudio Desktop (2023.06 ou posterior) — ele já inclui o **Quarto** e o **pandoc** (necessários para renderizar relatórios) e oferece um ambiente conveniente para iniciar o aplicativo.

17.3 Pacotes de R necessários

Os pacotes a seguir são instalados automaticamente como dependências:

Pacote	Finalidade
shiny	Arcabouço de aplicações web
mgcv	Ajuste dos modelos GAM
valengrCore	Núcleo compartilhado de avaliação (papéis especiais, transferência a partir do earthUI)
ggplot2, plotly	Gráficos de contribuição estáticos e interativos
gratia	Extração de termos suaves e diagnósticos
DT	Tabelas de dados interativas
quarto, rmarkdown	Geração e renderização de relatórios Quarto
readxl	Importação de arquivos Excel
shinyFiles	Navegadores de arquivos e pastas
RSQLite, DBI	Bancos de dados de configurações e de projetos (SQLite)
jsonlite	Serialização do estado das variáveis

17.4 Dependências opcionais

Componente	Quando é necessário
LaTeX (TinyTeX, MiKTeX ou MacTeX)	Somente para gerar relatórios em PDF. Se nenhuma instalação de LaTeX for detectada, a opção de PDF é automaticamente ocultada no menu de formatos. Instale com: <code>tinytex::install_tinytex()</code>
sysfonts + showtext	Fonte Roboto Condensed para os gráficos em ggplot2. Se estiver indisponível ou o computador estiver off-line, o aplicativo recorre automaticamente à fonte sem serifa do sistema.
earth ($\geq 5.3.0$)	Necessário apenas para o fluxo de importação do earthUI. Não é exigido no uso autônomo.
callr	Ajuste do modelo em segundo plano (mantém o aplicativo responsivo). Sem ele, o ajuste é executado de forma síncrona.
writexl, openxlsx	Downloads em Excel (saída do modelo, RCA, grade de vendas). O aplicativo avisa se estiverem ausentes.
bslib	Tematização Bootstrap 5 (Nord Light/Dark).
officer	Exportação programática legada para Word (<code>export_gam_docx()</code>).

17.5 Observações por plataforma

17.5.1 macOS

É onde há menos problemas. Recomenda-se o Homebrew para as bibliotecas de sistema. Se forem necessários relatórios em PDF:

```
# In R:
install.packages("tinytex")
tinytex::install_tinytex()
```

17.5.2 Windows

Funciona bem com o RStudio Desktop. Observações principais:

- **Máquinas corporativas ou restritas:** se o diretório temporário ou o diretório de dados do usuário estiver restrito, a persistência das configurações e a geração de relatórios podem falhar. O aplicativo exibirá um aviso no console do R, mas continuará funcionando sem persistência.
- **LaTeX:** instale o TinyTeX a partir do R (mesmo comando do macOS acima) ou instale o MiKTeX em miktex.org.
- **Caminhos de arquivo longos:** o Windows tem um limite de 260 caracteres para caminhos. Mantenha curto o caminho do seu diretório de trabalho.

17.5.3 Linux

É a plataforma mais variável entre distribuições. Usuários de Ubuntu/Debian podem precisar de bibliotecas de sistema antes que os pacotes de R sejam compilados:

```
sudo apt install libcurl4-openssl-dev libssl-dev \
  libxml2-dev libsqlite3-dev libfontconfig1-dev
```

Para relatórios em PDF: `tinytex::install_tinytex()` ou instale `texlive-full` pelo gerenciador de pacotes do sistema.

Em servidores headless (sem monitor), os gráficos interativos em plotly são renderizados como alternativas estáticas nos relatórios. O próprio aplicativo Shiny exige uma conexão por navegador web.

17.6 Degradação graciosa

O `mgcvUI` foi concebido para funcionar mesmo quando componentes opcionais estão ausentes:

Componente ausente	Comportamento
Sem LaTeX	A opção de PDF fica oculta no menu de formatos de relatório. HTML e Word continuam disponíveis.
Sem internet / falha nas fontes	A fonte Roboto Condensed é substituída pela fonte sem serifa do sistema. Registra-se uma mensagem no console.
Sem RSQLite / sistema de arquivos somente leitura	As configurações não persistem entre sessões. Exibe-se um aviso no console. No mais, o aplicativo funciona normalmente.
Falha de <code>gratia::appraise</code>	A aba Diagnostics e os relatórios exibem uma mensagem de espaço reservado. Use <code>mgcv::gam.check()</code> no console como alternativa.
Diretório temporário sem permissão de escrita	A geração de relatórios falha com uma mensagem de erro clara. Resolva definindo a variável de ambiente <code>TMPDIR</code> para um local com permissão de escrita.

17.7 Solução de problemas

17.7.1 “PDF option is not available”

Nenhuma instalação de LaTeX foi detectada. Execute no R:

```
install.packages("tinytex")
tinytex::install_tinytex()
```

Depois reinicie o aplicativo. A opção de PDF aparecerá.

17.7.2 “Settings will not persist”

O diretório do banco de dados SQLite não pode ser criado ou gravado. Verifique as permissões em:

- **macOS/Linux:** `~/Library/Application Support/R/mgcvUI/` ou `~/.local/share/R/mgcvUI/`
- **Windows:** `%APPDATA%/R/data/mgcvUI/`

17.7.3 “Diagnostics plot unavailable”

A função `gratia::appraise()` ocasionalmente falha em modelos complexos (muitas interações tensoriais, famílias não padronizadas). É um problema conhecido a montante. Solução alternativa: execute `mgcv::gam.check(model)` no console do R para ver diretamente os gráficos de diagnóstico.

17.7.4 “Report error: number of items to replace...”

Este foi um problema conhecido em versões anteriores, relacionado ao pacote `gratia`. Certifique-se de ter a versão mais recente do `mgcvUI` instalada. A versão atual captura esse erro e produz o relatório com um espaço reservado na seção de diagnósticos.

17.7.5 Porta 7880 já em uso

Se o aplicativo não iniciar porque a porta 7880 está ocupada (por uma sessão anterior), execute:

```
# macOS/Linux:
```

```
lsof -ti:7880 | xargs kill
```

```
# Windows (PowerShell):
```

```
Stop-Process -Id (Get-NetTCPConnection -LocalPort 7880).OwningProcess
```

17.7.6 Falha na instalação de pacotes no Linux

Faltam bibliotecas de sistema. Instale os cabeçalhos de desenvolvimento listados na seção sobre Linux acima e repita `install.packages()`.

Referências

- Wood, S. N. (2003). “Thin-plate regression splines”. Em: *Journal of the Royal Statistical Society (B)* 65.1, pp. 95–114. DOI: [10.1111/1467-9868.00374](https://doi.org/10.1111/1467-9868.00374).
- (2004). “Stable and efficient multiple smoothing parameter estimation for generalized additive models”. Em: *Journal of the American Statistical Association* 99.467, pp. 673–686. DOI: [10.1198/016214504000000980](https://doi.org/10.1198/016214504000000980).
- (2011). “Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models”. Em: *Journal of the Royal Statistical Society (B)* 73.1, pp. 3–36. DOI: [10.1111/j.1467-9868.2010.00749.x](https://doi.org/10.1111/j.1467-9868.2010.00749.x).
- (2017). *Generalized Additive Models: An Introduction with R*. 2ª ed. Chapman e Hall/CRC.
- Wood, S. N., N. Pya e B. Säfken (2016). “Smoothing parameter and model selection for general smooth models (with discussion)”. Em: *Journal of the American Statistical Association* 111, pp. 1548–1575. DOI: [10.1080/01621459.2016.1180986](https://doi.org/10.1080/01621459.2016.1180986).